



**HAL**  
open science

# Modalités de régulation d'ER $\alpha$ 36 et leurs conséquences sur la physiopathologie de la glande mammaire

Charlène Thiebaut

## ► To cite this version:

Charlène Thiebaut. Modalités de régulation d'ER $\alpha$ 36 et leurs conséquences sur la physiopathologie de la glande mammaire. Cancer. Université de Lorraine, 2019. Français. NNT : 2019LORR0090 . tel-02387078

**HAL Id: tel-02387078**

**<https://hal.univ-lorraine.fr/tel-02387078v1>**

Submitted on 29 Nov 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : [ddoc-theses-contact@univ-lorraine.fr](mailto:ddoc-theses-contact@univ-lorraine.fr)

## LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

[http://www.cfcopies.com/V2/leg/leg\\_droi.php](http://www.cfcopies.com/V2/leg/leg_droi.php)

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>



UNIVERSITÉ  
DE LORRAINE

BioSE



**Ecole Doctorale BioSE (Biologie-Santé-Environnement)**

**Thèse**

**Présentée et soutenue publiquement pour l'obtention du titre de**

**DOCTEUR DE L'UNIVERSITE DE LORRAINE**

**Mention : « Sciences de la Vie et de la Santé »**

par **Charlène THIEBAUT**

**Modalités de régulation d'ER $\alpha$ 36 et leurs conséquences sur  
la physiopathologie de la glande mammaire**

**Le 30 Septembre 2019**

**Membres du jury :**

**Rapporteurs :** Mme Muriel LE ROMANCER-CHERIFI

Directrice de Recherche INSERM,  
CRCL UMR INSERM 1052 CNRS 5286, Lyon

M. Farzad PAKDEL

Directeur de recherche CNRS,  
IRSET INSERM UMR 1085, Rennes

**Examineurs :** Mme Marianne CLAUSEL

Professeur,  
IECL UMR 7502, Vandœuvre-lès-Nancy

Mme Virginie ROUILLER-FABRE

Professeur,  
CEA INSERM UMR 967, Fontenay-aux-Roses

Mme Hélène DUMOND

Maître de Conférence, **Directrice de thèse**  
CRAN UMR CNRS 7039, Vandœuvre-lès-Nancy

M. Taha BOUKHOBZA

Professeur, **Co-directeur de thèse**  
CRAN UMR CNRS 7039, Vandœuvre-lès-Nancy

---

**CRAN BioSIS, UMR 7039 CNRS,  
Faculté des Sciences et Technologies, Université de Lorraine, 54506 Vandœuvre-lès-Nancy**

# Remerciements

« Aucun de nous, en agissant seul, ne peut atteindre le succès ». Nelson MANDELA

C'est grâce à vous tous que j'ai réussi à mener à bien mon projet et à finaliser ce travail de thèse.

C'est pourquoi je tiens à vous remercier !

Avant tout, je souhaite remercier l'ensemble des membres du jury d'avoir accepté d'évaluer mon travail de thèse. Merci à Madame **Muriel LE ROMANCER-CHERIFI**, Directrice de Recherche INSERM et Monsieur **Farzad PAKDEL**, Directeur de Recherche CNRS, d'avoir accepté d'être rapporteurs de ce travail de thèse. Je tiens également à remercier le Professeur **Virginie ROUILLER-FABRE** et le Professeur **Marianne CLAUSEL** d'avoir participé à mon comité de suivi de thèse et d'avoir accepté de juger ce travail en tant qu'examinatrices.

Je remercie également la direction du laboratoire et plus particulièrement le Professeur **Muriel BARBERI-HEYOB** et le Professeur **Alain RICHARD**, d'avoir permis le financement de cette thèse et de m'avoir accueillie au sein du département BioSIS.

Je tiens à remercier tout particulièrement ma directrice de thèse, le Dr **Hélène DUMOND** qui m'a accueilli au sein du laboratoire il y a 6 ans déjà, dans le cadre de mon stage de DUT. Depuis, tu m'as toujours accordé ta confiance et aujourd'hui c'est en grande partie grâce à toi que j'en suis arrivée là. Merci pour tes conseils toujours avisés, ta bienveillance et ta grande disponibilité. Merci d'avoir supporté mon stress et de m'avoir soutenu et encouragé pendant la rédaction mais aussi avant chaque congrès ou présentation ! J'ai beaucoup appris à tes côtés ! Enfin, merci d'avoir mis sur ma route, mon petit chat, **Oligo**, sans qui la rédaction de la thèse aurait sûrement été un peu plus paisible...!

Je remercie également mon co-directeur de thèse, le Professeur **Taha BOUKHOBZA**, d'avoir permis d'enrichir ce travail de thèse grâce à des modèles mathématiques qui garderont toujours une petite part de mystère pour moi ! Un grand merci pour ta disponibilité, ta bonne humeur et surtout ta patience quand il a fallu m'expliquer le fonctionnement de « R », « Matlab » etc. Je te dois beaucoup et notamment en ce qui concerne les postes ATER ! Enfin, merci de m'avoir fait découvrir le gâteau Thé vert/Framboise et le gâteau à la guimauve !

Un grand merci également au Dr **Amand CHESNEL**, qui depuis mon arrivée au laboratoire m'a transmis sa rigueur scientifique et de précieuses connaissances surtout concernant la biologie moléculaire. Je me rappellerai toujours des nombreuses extractions d'ADN que nous avons réalisé

ensemble et de tous les conseils avisés que tu m'as donnés. Merci également d'avoir partagé avec moi ton bureau et un petit peu de ta culture bretonne ! Je retiendrai notamment le gâteau breton (dont je garde précieusement la recette !), le kouign-amann et les bonnes crêpes de **Zabeth**, que j'ai eu le plaisir de déguster ! Encore merci pour tout et je te souhaite, ainsi qu'à Zabeth, une retraite pleine de joies, de jardinage, de balades en vélo et de nouvelles découvertes en Bretagne !

Je tiens également à remercier tous les membres du laboratoire avec qui j'ai de près ou de loin pu collaborer : **Alexandra, Audrey T., Catherine, Elia, Emilie, Vanessa** mais aussi les Dr **Stéphanie GRANDEMANGE, Isabelle GRILLIER-VUISSOZ, Sandra KUNTZ, Sabine MAZERBOURG** et **Nadège TOUCHE**. Je vous remercie pour votre sympathie et vos encouragements notamment lors de la dernière ligne droite de la rédaction. Je pense également aux différents stagiaires passés par le laboratoire et plus particulièrement à **Annaïck** qui nous a apporté beaucoup de bonne humeur ! Enfin, je tiens à souhaiter une bonne continuation pour sa thèse à **Marine L.**, alias « petite Marine ».

Je n'oublie pas les « anciens » du labo. Notamment les anciens doctorants **Clémence, Audrey B., Marine G. et Alexis** avec qui j'ai pu partager de bons moments. Merci pour votre bonne humeur et vos encouragements ! Je pense également à **Chloé M.**, avec qui j'ai partagé mon stage de M1. Notre collaboration n'a duré que 2 mois mais cela a suffi à développer une belle amitié. J'espère venir te voir à Bordeaux très bientôt !

Merci également à tous ceux qui ont contribué à ce travail de thèse. Je pense notamment à **Romain** qui a participé au développement du projet bioinfo et à **Henri-Philippe** qui m'a accueillie à Lyon et qui m'a formée à la technique de PLA.

Mille mercis pleins d'émotions à mes collègues, qui sont aujourd'hui devenus mes amis : Claire, Florian, Guillaume et Maelle. Grâce à vous j'ai versé beaucoup de larmes de rire et de bonheur au cours de cette thèse ! Vous m'avez apporté tellement de choses au niveau professionnel et personnel que je ne sais pas comment vous remercier !

**Claire** (Je n'ai pas osé dire ton surnom...), partager le bureau de la rédaction avec toi a été un vrai plaisir et une source de réconfort dans les moments de doute. Merci de m'avoir sensibilisé à la problématique bio, équitable, local et de saison (Je crois que je n'ai rien oublié...) ! Je te souhaite bon courage pour la dernière ligne droite et je serai là pour te soutenir !

Mon **Floflo**, je me souviens encore de notre première conversation en salle WB. Merci pour ta bonne humeur communicative, ton écoute attentive et tes techniques de relaxation plutôt originales. Merci d'avoir absolument TOUT partagé avec nous (même ce que j'aurais parfois aimé que tu gardes pour toi... !). Je te souhaite le meilleur pour la suite, tu le mérites !

**Guigui**, ou plutôt Dr Drouot, ton humour incontrôlable manque au laboratoire depuis ton départ (même si Floflo a repris le flambeau !). Je voulais te remercier d'avoir tenté (en vain !) de me transmettre ton savoir-faire pour la ChIP et pour tes précieux conseils au moment de la rédaction de ce manuscrit. Un grand merci également à **Amélie** pour son sourire, sa bonne humeur et pour avoir pris soin de ma grand-mère ! Je vous souhaite beaucoup de bonheur à tous les 2 !

Enfin, **Mama**, ou plutôt ma bichette, je n'oublierais jamais notre complicité dans les manips et ton aide précieuse au cours des 2 premières années de ma thèse. Ton départ pour la Bretagne a laissé un grand vide au laboratoire mais nos après-midi Skype et toutes tes petites attentions m'ont permis de tenir le coup dans la dernière ligne droite ! Aujourd'hui, tu es devenue ma confidente et une amie sincère grâce à laquelle ma philosophie de vie ne cesse d'évoluer !

Merci aussi à mes amis et plus particulièrement à ma **Caro**, **Justine** et **Pauline**, qui au travers de restos, de sorties shoppings ou de petits week-ends m'ont permis de m'évader du labo et de la thèse !

Merci à **Mélanie**, **Steve** et leur 2 p'tits loups, **Noham** et **Mahdi** d'égayer la plupart de mes samedis matin !

Mille mercis à ma **Maman** et mon **p'tit Frère** qui m'ont toujours soutenue et encouragée quels que soient mes choix. Je sais que ça n'a pas dû être facile tous les jours mais finalement ça en valait la peine ! En ce jour si particulier de ma soutenance de thèse, je ne trouve pas de mot assez fort pour vous dire à quel point je suis heureuse de vous avoir à mes côtés !

Un grand merci également à mon **Tonton**, ma **Tatie** ainsi qu'à **Tételle** et **Mimi** ! Merci d'avoir toujours suivi mon parcours, de m'avoir soutenue dans les moments difficiles et pour tous vos petits mots d'encouragements !

Merci à ma **Grand-mère Ginette**, qui je l'espère sera toujours aussi fière de « ses beaux yeux bleus » et une pensée émue pour ma **Grand-mère Jeannine**...

# Sommaire

<u>Remerciements</u>	1
<u>Liste des figures</u>	9
<u>Liste des tableaux</u>	13
<u>Abréviations</u>	14
<u>Introduction bibliographique</u>	21

## Chapitre 1 : La signalisation œstrogénique ..... 22

I. Les œstrogènes : .....	22
A) Structure et étapes de la biosynthèse des œstrogènes .....	22
B) Transport et métabolisme des œstrogènes .....	23
C) Rôles physiologiques des œstrogènes .....	24
i) Rôle dans la fonction de reproduction .....	24
ii) Rôles cardio-vasculaires .....	25
iii) Rôles sur le système squelettique .....	26
iv) Rôles sur le système immunitaire .....	26
II. Les récepteurs aux œstrogènes .....	27
A) Les récepteurs nucléaires des œstrogènes : ER $\alpha$ 66 et ER $\beta$ 1 .....	27
i) Organisation génomique des gènes ESR1 et ESR2 .....	28
ii) Organisation protéique d'ER $\alpha$ 66, ER $\beta$ 1 et leurs isoformes .....	29
iii) La signalisation génomique des œstrogènes .....	32
B) Les récepteurs aux œstrogènes membranaires .....	38
i) Localisation membranaire de la forme canonique ER $\alpha$ 66 .....	38
ii) Les variants ER $\alpha$ 36 et ER $\alpha$ 46 .....	38
iii) Le récepteur membranaire GPER .....	39
iv) La signalisation non génomique des œstrogènes .....	40
C) Les récepteurs apparentés aux récepteurs aux œstrogènes : les ERRs .....	41

## **Chapitre 2 : Le développement normal et pathologique de la glande mammaire ..... 43**

I.	Organisation générale de la glande mammaire .....	43
	A) L'épithélium mammaire .....	44
	B) Le stroma mammaire .....	45
II.	Morphogénèse de la glande mammaire.....	45
	A) Le développement embryonnaire de la glande mammaire .....	45
	B) Le développement post-natal de la glande mammaire.....	47
	i) La puberté .....	47
	ii) Le cycle gestation-lactation-involution .....	48
III.	Le développement pathologique de la glande mammaire : le cancer du sein .....	49
	A) Epidémiologie du cancer du sein .....	50
	B) Les facteurs de risques du cancer du sein .....	51
	i) Les facteurs de risques internes .....	51
	ii) Les facteurs de risques externes .....	52
	C) La carcinogénèse mammaire.....	53
	D) La classification des tumeurs mammaires .....	54
	i) La classification histologique des tumeurs mammaires .....	55
	ii) La classification anatomo-pathologique des tumeurs mammaires .....	56
	iii) La classification moléculaire des tumeurs mammaires .....	56
	iv) Développement de nouvelles classifications grâce à des outils dérivés de l'intelligence artificielle .....	58
	E) Les stratégies thérapeutiques dans le cancer du sein.....	60
	i) Les traitements à visée locale .....	60
	ii) Les traitements à visée systémique .....	61
	iii) Les thérapies ciblées .....	62



## **Chapitre 3 : Modulation de la signalisation œstrogénique dans la physiopathologie de la glande mammaire ..... 65**

I.	La régulation du métabolisme des œstrogènes.....	65
II.	La régulation de l'expression d'ER $\alpha$ 66.....	66
	A) Amplification du gène ESR1 .....	66
	B) Mécanismes de régulation épigénétique de l'expression d'ER $\alpha$ 66 .....	67
	C) Mécanismes de régulation transcriptionnelle de l'expression d'ER $\alpha$ 66.....	69
	i) Les facteurs de transcription de la famille AP2.....	69
	ii) Les facteurs de transcription de type Forkhead .....	71
	iii) Les autres facteurs de transcription .....	72
	D) Mécanismes de régulation post-transcriptionnelle de l'expression d'ER $\alpha$ 66.....	74
III.	La Régulation de l'activité d'ER $\alpha$ 66 .....	76
	A) Les mutations activatrices ciblant ER $\alpha$ 66.....	76
	B) Mécanismes de régulation de l'activité transcriptionnelle d'ER $\alpha$ 66 .....	77
	C) Les modifications post-traductionnelles d'ER $\alpha$ 66.....	78
	D) Mécanismes de régulation de l'activité d'ER $\alpha$ 66 par d'autres récepteurs hormonaux.....	79
IV.	Le cas du variant ER $\alpha$ 36.....	80
	A) Les données cliniques .....	80
	B) Les voies de signalisation dépendantes d'ER $\alpha$ 36 dans les tumeurs mammaires.....	81
	i) La voie EGFR/Src/STAT3/5 .....	82
	ii) La voie ER $\alpha$ 36/HER2/EGFR.....	82
	iii) La voie Ras/MEK/ERK1/2 .....	82
	iv) La voie dépendante de la PKC.....	83
	v) La voie PI3K/AkT.....	83
	C) Mécanismes de régulation de l'expression d'ER $\alpha$ 36.....	84
	D) Mécanismes de régulation de l'activité d'ER $\alpha$ 36.....	86
	E) Molécules régulant l'expression et l'activité d'ER $\alpha$ 36.....	86

Contexte et objectifs de la thèse 88

Résultats 91

- I. Quel est l'impact d'ER $\alpha$ 36 sur le développement physiologique de la lignée germinale mâle et de la glande mammaire ? ..... 94
- A) Impact d'ER $\alpha$ 36 sur le développement de la lignée germinale mâle murine ..... 95
- B) Impact d'ER $\alpha$ 36 sur le développement de la glande mammaire murine ..... 97
- i) Etudes in vivo sur le modèle de souris transgénique MMTV-ER $\alpha$ 36 ..... 97
- ii) Etudes in vitro et in silico sur les cellules épithéliales mammaires saines MCF-10A..98

*Publication N°1* 104

C. Thiebaut, C. Chamard-Jovenin, A. Chesnel, C. Morel, E.H. Djermoune, T. Boukhobza, H. Dumond. 2017. "Mammary epithelial cell phenotype disruption in vitro and in vivo through ERalpha36 overexpression". PLoS ONE, 12, pp.e0173931.

*Publication N°2* 105

C. Chamard-Jovenin, C. Thiebaut, A. Chesnel, E. Bresso, C. Morel, M. Smaïl-Tabbone, M.D. Devignes, T. Boukhobza, H. Dumond. 2017. "Low-dose alkylphenol exposure promotes mammary epithelium alterations and transgenerational developmental defects, but does not enhance tumorigenic behaviour of breast cancer cells". Frontiers in Endocrinology, 8, pp.272.

- II. Quels sont les mécanismes qui régulent l'expression d'ER $\alpha$ 36 ?..... 106
- A) Régulation de l'expression d'ER $\alpha$ 36 par des facteurs trans-régulateurs ..... 106
- B) Régulation de l'expression d'ER $\alpha$ 36 par le statut de méthylation de son promoteur 109
- C) Régulation post-transcriptionnelle de l'expression d'ER $\alpha$ 36 par des micro-ARNs .. 110
- D) Impact de la régulation de l'expression d'ER $\alpha$ 36 par la DAC sur la résistance au tamoxifène..... 112
- E) Développement d'une approche bioinformatique permettant d'identifier de nouveaux facteurs impliqués dans la signalisation dépendante d'ER $\alpha$ 36 ..... 113

**C. Thiebaut**, A. Chesnel, J.L. Merlin, M. Chesnel, A. Leroux, A. Harlé, and H. Dumond. 2019. “Dual epigenetic regulation of ER $\alpha$ 36 expression in breast cancer cells”. International Journal of Molecular Sciences, 20(11), 2637.

B. Bastien, T. Boukhobza, H. Dumond, **C. Thiebaut**, H. Chakir, A. Gegout-Petit, A. Muller-Gueudin, Y. Shi. “A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology”. Soumis le 04/07/2019 dans le journal Biostatistics.

Discussion et Perspectives

- I. L’expression d’ER $\alpha$ 36 dans les cellules épithéliales mammaires saines est-elle nécessaire et/ou suffisante pour induire l’apparition d’un phénotype néoplasique ? ..... 122
- II. Quels sont les signaux et les mécanismes responsables de la régulation de l’expression d’ER $\alpha$ 36 ? ..... 125
- III. Quelle est la part de l’implication d’ER $\alpha$ 36 dans l’apparition de résistances intrinsèques ou acquises aux thérapies anti-cancéreuses ? ..... 129

# Liste des figures

## Introduction bibliographique

**Figure 1 :** Structures chimiques des œstrogènes naturels.

**Figure 2 :** Les étapes de la biosynthèse des œstrogènes.

**Figure 3 :** Les étapes du catabolisme des œstrogènes.

**Figure 4 :** Le rôle physiologique des œstrogènes.

**Figure 5 :** Le cycle menstruel.

**Figure 6 :** L'organisation génomique du gène ESR1.

**Figure 7 :** L'organisation génomique du gène ESR2.

**Figure 8 :** Structure générale des récepteurs nucléaires.

**Figure 9 :** Structure des formes canoniques et des différents variants d'ER $\alpha$ 66 et d'ER $\beta$ 1.

**Figure 10 :** La signalisation génomique des œstrogènes.

**Figure 11 :** Mécanismes d'action des coactivateurs d'ER $\alpha$ 66.

**Figure 12 :** Mécanismes d'action des corépresseurs d'ER $\alpha$ 66.

**Figure 13 :** La voie ERE-indépendante *via* le facteur de transcription NF- $\kappa$ B.

**Figure 14 :** La signalisation non-génomique des œstrogènes.

**Figure 15 :** La signalisation non-génomique dépendante du récepteur GPER.

**Figure 16 :** Structure des récepteurs apparentés aux récepteurs aux œstrogènes (ERR).

**Figure 17 :** Anatomie des glandes mammaires humaines et murines à l'âge adulte.

**Figure 18 :** Histologie des canaux mammaires murins.

**Figure 19 :** Modèle hypothétique de ségrégation des lignages épithéliaux de la glande mammaire.

**Figure 20 :** Développement embryonnaire de la glande mammaire humaine chez le fœtus féminin.

**Figure 21 :** Développement embryonnaire de la glande mammaire murine chez le fœtus femelle.

**Figure 22 :** Les différents stades postnataux de différenciation de la glande mammaire murine.

**Figure 23 :** Epidémiologie du cancer de sein.

**Figure 24 :** Les différentes étapes de la carcinogénèse.

**Figure 25** : La transition épithélio-mésenchymateuse.

**Figure 26** : Les 2 modèles explicatifs du concept d'hétérogénéité intra-tumorale.

**Figure 27** : La classification histologique des tumeurs mammaires.

**Figure 28** : La classification moléculaire des tumeurs mammaires.

**Figure 29** : Micro-ARNs associés à différentes caractéristiques phénotypiques des cellules cancéreuses mammaires.

**Figure 30** : Le drainage lymphatique du sein.

**Figure 31** : Les thérapies ciblées impactant la signalisation œstrogénique.

**Figure 32** : Statut de méthylation de la région promotrice du gène ESR1.

**Figure 33** : Métabolisme intracellulaire de la 5-Azacytidine et de la Décitabine.

**Figure 34** : Stabilisation de l'ARNm d'ER $\alpha$ 66 de façon dépendante des éléments MEMSS.

**Figure 35** : La biosynthèse des micro-ARNs.

**Figure 36** : Micro-ARNs ciblant la région 3'UTR de l'ARNm codant ER $\alpha$ 66.

**Figure 37** : Les mutations activatrices affectant le domaine E/F d'ER $\alpha$ 66.

**Figure 38** : La signalisation dépendante d'ER $\alpha$ 36.

**Figure 39** : Organisation de la région promotrice d'ER $\alpha$ 36.

## **Résultats**

**Figure 40** : Impact de la surexpression d'ER $\alpha$ 36 dans le testicule fœtal de souris sur les processus biologiques « Cell cycle », « Differentiation » et « Meiosis ».

**Figure 41** : Impact des alkylphénols associés à la surexpression d'ER $\alpha$ 36 dans le testicule fœtal de souris sur les processus biologiques « Cell cycle », « Differentiation » et « Meiosis ».

**Figure 42** : Analyse quantitative de l'arbre mammaire des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 au sevrage.

**Figure 43** : Etude de l'histologie des canaux mammaires des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 au sevrage.

**Figure 44** : Etude de l'histologie des canaux mammaires des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 à l'âge adulte.

**Figure 45** : Stratégie d'enrichissement des données à partir du transcriptome des cellules MCF-10A/ER $\alpha$ 36 et MCF-10A/Zeo.

**Figure 46 :** Impact d'ER $\alpha$ 36 sur l'expression de marqueurs associés à la transition épithélio-mésenchymateuse.

**Figure 47 :** Implication des facteurs STAT3, NF- $\kappa$ B et PTEN dans la signalisation induite par la surexpression d'ER $\alpha$ 36 dans les cellules MCF-10A.

**Figure 48 :** Impact de la surexpression d'ER $\alpha$ 36 sur la réponse apoptotique des cellules MCF-10A.

**Figure 49 :** Impact de la surexpression d'ER $\alpha$ 36 sur les capacités migratoires des cellules MCF-10A.

**Figure 50 :** Impact de la surexpression d'ER $\alpha$ 36 sur la prolifération des cellules MCF-10A.

**Figure 51 :** Effet de la surexpression d'ER $\alpha$ 36 sur la localisation intracellulaire d'ER $\beta$ .

**Figure 52 :** Impact du mélange M4 sur la prolifération, la migration et la survie des cellules MCF-10A.

**Figure 53 :** Impact du mélange M4 sur la prolifération, la migration et la survie des cellules MCF-10A/ER $\alpha$ 36.

**Figure 54 :** Etude des effets multigénérationnels et transgénérationnels des perturbateurs endocriniens.

**Figure 55 :** Etude des effets transgénérationnels du mélange M4 sur l'histologie des canaux mammaires des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 au sevrage.

**Figure 56 :** Etude des effets transgénérationnels du mélange M4 sur l'histologie des canaux mammaires des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 à l'âge adulte.

**Figure 57 :** Système rapporteur de l'activité du promoteur d'ER $\alpha$ 36.

**Figure 58 :** Impact de l'activation d'Ahr par son ligand, le BaP, sur l'expression d'ER $\alpha$ 36.

**Figure 59 :** Représentation schématique de la localisation des 4 îlots CpGs susceptibles de réguler l'expression d'ER $\alpha$ 36.

**Figure 60 :** Etude de la corrélation entre le statut de méthylation du promoteur d'ER $\alpha$ 36 et le niveau d'expression de ce variant dans des biopsies de tumeurs mammaires.

**Figure 61 :** Impact de la décitabine sur la méthylation du promoteur d'ER $\alpha$ 36 et l'expression de ce variant.

**Figure 62 :** Implication d'ER $\alpha$ 66 dans la déméthylation de la région promotrice d'ER $\alpha$ 36 induite par la décitabine.

**Figure 63 :** Etude de l'interaction entre hsa-miR-136-5p et la région 3'UTR de l'ARNm d'ER $\alpha$ 36.

**Figure 64 :** Effet de la décitabine sur l'expression d'hsa-miR-136-5p.

**Figure 65 :** Effet de la décitabine sur l'expression du long ARN non codant CRNDE.

**Figure 66 :** Impact de la décitabine sur la résistance au tamoxifène dépendante de l'expression d'ER $\alpha$ 36.

**Figure 67 :** Schéma récapitulatif de la stratégie développée pour identifier de nouveaux facteurs impliqués dans la signalisation dépendante d'ER $\alpha$ 36.

**Figure 68 :** Validation biologique des résultats obtenus grâce l'algorithme R « ARMADA » .

**Figure 69 :** Détermination des différentes classes d'expression d'ER $\alpha$ 36 dans chaque jeu de données.

# *Liste des tableaux*

## *Introduction bibliographique*

**Tableau I :** Classification des récepteurs nucléaires.

**Tableau II :** Utilisation différentielle des différents promoteurs du gène ESR1 dans les tissus cibles des œstrogènes.

**Tableau III :** La classification des stades de développement pubertaire de la glande mammaire selon Tanner.

**Tableau IV :** La classification TNM des tumeurs mammaires.

**Tableau V :** Le grade histopronostique de Scarff-Bloom-Richardson (SBR).

**Tableau VI :** Régulateurs transcriptionnels du gène ESR1.

**Tableau VII :** Principales modifications post-traductionnelles ciblant ER $\alpha$ 66.

**Tableau VIII :** Principales modifications post-traductionnelles d'ER $\alpha$ 66 dans le contexte des tumeurs mammaires.

## *Résultats*

**Tableau IX :** Caractéristiques des 62 tumeurs du jeu de données « Luminal ».

**Tableau X :** Résultats de l'analyse d'enrichissement des données.

**Tableau XI :** Répartition des tumeurs du jeu de données « Luminal » dans les différentes classes d'expression d'ER $\alpha$ 36.



# *Abréviations*

## **1-9**

3 $\beta$ -HSD	3 $\beta$ – HydroxySteroid Dehydrogenase
4-OHT	4- Hydroxy-tamoxifène
5-Aza	5-azacytidine
5, 15 DPP	5,15-DiPhenylPorphyrin
17 $\beta$ -HSD	17 $\beta$ – HydroxySteroid Dehydrogenase

## **A**

ADN	Acide DésoxyriboNucléique
ADNc	Acide DésoxyriboNucléique complémentaire
Adora-1	Adenosine A1 receptor
AF-1, 2	Activation Function 1, 2
AhR	Aryl hydrocarbon Receptor
AJCC	American Joint Committee on Cancer
Akt	Protein kinase B
ALDH1	ALdehyde DeHydrogenase 1
AMPc	Adénosine MonoPhosphate cyclique
AP-1	Activator Protein-1
AR	Androgen Receptor
ARN	Acide Ribonucléique
ARNm	Acide Ribonucléique messenger
ATM	Ataxia Telangiectasia Mutated
ATP	Adenosine TriPhosphate
AUF1p45	A + U-rich RNA-binding Factor 1 protein isoform p45

## **B**

BaP	Benzo[a]Pyrène
BCDB	Breast Cancer DataBase
bHLH-PAS	basic Helix-Loop-Helix- PER-ARNT-SIM
BLG	$\beta$ -LactoGlobuline
BLIMP-1	B Lymphocyte-Induced Maturation Protein-1
BPA	BisPhénol A
BRCA1, 2	BReast Cancer Associated 1, 2
BRE	TFIIB Recognition Element
BSA	Bovine Serum Albumin

## C

C57BL/6J	C57 Black 6J mouse strain
C/EBPs	CCAAT-Enhancer-Binding Proteins
CAF	Cancer-Associated Fibroblast
CARM1	Coactivator-associated Arginine Methyltransferase 1
CBP/p300	CREB-Binding Protein / p300
CCAR1	Cell Cycle and Apoptosis Regulator 1
CD24, 44	Cluster of Differentiation
CDH1, 2	CaDHerin 1, 2
CDK	Cyclin-Dependent Kinase
CGP	Cancer Genome Project
CHEK2	CHEckpoint Kinase 2
CICLADES	CIRCuLATING Dna ESR1 Gene Mutations Analysis
CINC-1	Cytokine-Induced Neutrophil Chemoattractant 1
CircRNA	Circular RNA
CoCoA	Coiled-coil CoActivator
COMT	Catéchol-O-MéthylTransférase
CREB	c-AMP Response Element-Binding protein
CRNDE	ColoRectal Neoplasia Differentially Expressed
CSC	Cancer Stem Cells
CSMa	Cellules Souches Mammaires
CtBP	Carboxyl-terminal Binding Protein
CXCL-8	C-X-C motif Chemokine Ligand-8
CYP	Cytochrome P450

## D

DAC	Décitabine / 2'-deoxy-5-azacytidine
DAG	DiAcylGlycerol
DAVID	Database for Annotation, Visualization and Integrated Discovery
DBD	DNA Binding Domain
DES	DiEthylStilbestrol
DFS	Disease Free Survival
DLK1-DIO3	Delta-LiKe homolog 1 - type III iodothyronine DeIOdinase
DNMT	DNA Methyltransferase
DMSO	Diméthylsulfoxyde

## **E**

E <sub>1</sub>	œstrone
E <sub>2</sub>	17β-œstradiol
E <sub>3</sub>	œstriol
E <sub>4</sub>	œstétrol
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
eNOS	endothelial Nitric Oxide Synthase
ERα	Estrogen Receptor alpha
ERβ	Estrogen Receptor bêta
ERBB2	Erythroblastic leukaemia viral oncogene homologue 2
ERBF-1	Estrogen Receptor promoter B associated Factor-1
ERE	Estrogen Response Element
ERF-1	Estrogen Receptor Factor-1
ERK1/2	Extracellular signal-Regulated Kinase 1/2
EROS1	Estrogen Receptor One Study 1
ERR	Estrogen-Related Receptor
ERRE	Estrogen-Related Response Element
ESR1, 2	Estrogen Receptor 1, 2
Ets	Avian erythroblastosis virus E26 homolog-1

## **F**

FGF	Fibroblast Growth Factor
FHRE	ForkHead Responsive Element
FISH	Fluorescence In Situ Hybridization
FKH	ForKHead
Fli-I	Flightless-I
Foxc2	Forkhead box protein C2
FSH	Follicle-Stimulating Hormone
FZD7	Frizzled-7

## **G, H**

GDP	Guanosine DiPhosphate
GLUT1	GLUcose Transporter type 1
GO	Gene Ontology
GPER	G-Protein coupled Estrogen Receptor
GTP	Guanosine TriPhosphate

GWAS	Genome Wide Association Study
HB-EGF	Heparin Binding - Epidermal Growth Factor
HDAC	Histone deACetylase
HDL	High Density Lipoprotein
HER2	Human Epidermal growth factor Receptor 2
HMT	Histone MéthylTransférase
HOTAIR	HOX Transcript Antisense RNA
Hsa-miRNA	Human-miRNA
HSP	Heat Shock Protein
HuR	HUman antigen R

## I

I $\kappa$ B	Inhibitor of NF- $\kappa$ B
ICGC	International Cancer Genome Consortium
IFN $\gamma$	Interferon gamma
IG-DMR	InterGenic-Differentially Methylated Region
IGF	Insulin-like Growth Factor
IKK	I $\kappa$ B kinase
IL-4, 6, 10, 12	InterLeukine-4, 6, 10, 12
INCa	Institut National du Cancer
IP3	Inositol trisPhosphate

## J

JAK2	Janus kinase 2
JARID1B	Jumonji AT-Rich Interactive Domain 1B
JNK	c-Jun Kinase
jpc	jour post-coïtum

## K, L

KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Knock Out
LBD	Ligand Binding Domain
LDL	Low Density Lipoprotein
LEF-1	Lymphoid Enhancer-binding Factor 1
LH	Luteinizing Hormone
LHRH	Luteinizing Hormone Releasing Hormone
lncRNA	Long non-coding RNA

## M

m-TOR	mammalian Target Of Rapamycin
MAPK	Mitogen-Activated Protein Kinases
MEG3-DMR	Maternally Expressed Gene 3 - Differentially Methylated Region
MEK1	MAP/ERK Kinase-1
MEMSS	Minimal Estradiol-Modulated Stability Sequences
MKP3	Mitogen-activated protein kinase phosphatase 3
MMP	Matrix MetalloProteinase
MMTV	Mouse Mammary Tumor Virus
MNAR	Modulator of Non-genomic Activity of estrogen Receptor
MRTF-A	Myocardin-Related Transcription Factor A
MSigDB	Molecular Signatures DataBase

## N, O

N-Cor	Nuclear Receptor Corepressor
NES	Nuclear Export Signal
NF-κB	Nuclear Factor kappa B
NGS	Next Generation Sequencing
NLS	Nuclear Localization Sequence
NO	Nitric oxide
NRD	N-terminal Repressor Domain
OCLN	Occludine
Oct4	OCTamer-binding transcription factor 4
ORF	Open Reading Frame

## P

p90-RSK	p90 Ribosomal S6 Kinase
PADA-1	PAlbociclib and Circulating Tumor DNA for ESR1 Mutation Detection
PAK1	p21-Activated Kinase 1
Panther	Protein ANalysis THrough Evolutionary Relationships
PAT	Palmitoyl AcylTransferase
PAX-2	Paired BoX-2
PBX1	Pre-B cell leukemic homeobox 1
PDX	Patient-Derived Xenograft
PELP1	Proline-, glutamic acid- and leucine-rich Protein 1
PGC-1	Peroxisome proliferator-activated receptor-Gamma Coactivator 1
PI3K	PhosphoInositide 3-Kinase

PIC	PreInitiation Complex
PIP2	PhosphatidyInositol-4,5-bisPhosphate
PKA	Protéine Kinase A
PKC	Protéine Kinase C
PLC	Phospholipase C
PND	PostNatal Day
PR	Progesterone Receptor
PR/SET	PRDF1-RIZ /Su(var)3-9, Enhancer-of-zeste and Trithorax
PRDM1	PR domain containing protein1
PRMT1	Protein arginine N-MethylTransferase 1
PTEN	Phosphatase and TENsin homolog
PTM	Post-Translational Modification

## **R, S**

RIP140	Receptor Interacting Protein of 140 kDa)
RISC	RNA-Induced Silencing Complex
RT-qPCR	Reverse Transcription – Réaction de Polymérisation en Chaîne quantitative
RTL1	RetroTransposon Like 1
SBR	Scarff-Bloom-Richardson
SERD	Selective Estrogen Receptor Degradation
SERM	Selective Estrogen Receptor Modulators
SH3	SRC Homology 3 Domain
SHBG	Sex Hormone Binding Globulin
SIRT1	Sirtuin 1
SMRT	Silencing Mediator of Retinoid and Thyroid receptors
SP1	Stimulating Protein 1
SPM	Structural Perturbation Method
Src-1	Steroid receptor coactivator-1
STAT3, 5	Signal Transducer and Activator of Transcription 3,5
SUT	SUIfoTransférases
SWI/SNF	SWITch/Sucrose Non-Fermenting

## **T**

TAF	TBP-Associated Factor
TBP	TATA Binding Protein
TCF7LD	Transcription Factor 7
TCGA	The Cancer Genome Atlas

TDLU	Terminal Ductal-Lobular Unit
TEB	Terminal End Bud
TEM	Transition Epithélio-Mésenchymateuse
Tg	Transgénique
TGF- $\beta$	Transforming Growth Factor – Beta
TNF- $\alpha$	Tumor Necrosis Factor alpha
TNM	Tumor Node Metastasis
TP53	Tumor Protein p53
TSS	Transcription Start Site
TUNEL	Terminal deoxynucleotidyl transferase dUTP Nick End Labeling

## **U, V, W**

UGT	UDP-GlucuronosylTransférases
UICC	Union Internationale Contre le Cancer
UTR	UnTranslated Region
UV	Ultra Violet
VEGF-A	Vascular Endothelial Growth Factor A
WAP	Whey Acidic Protein
wt	wild-type
WT1	Wilms' Tumor 1

## **Z**

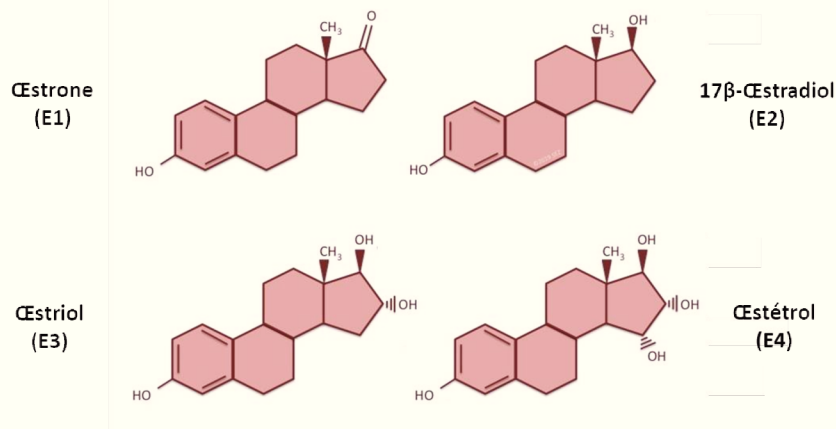
ZEB1	Zinc-finger E-box Binding homeobox 1
ZIC1	Zinc finger of the Cerebellum 1
ZO1	Zonula Occludens 1

---

# *Introduction bibliographique*

---

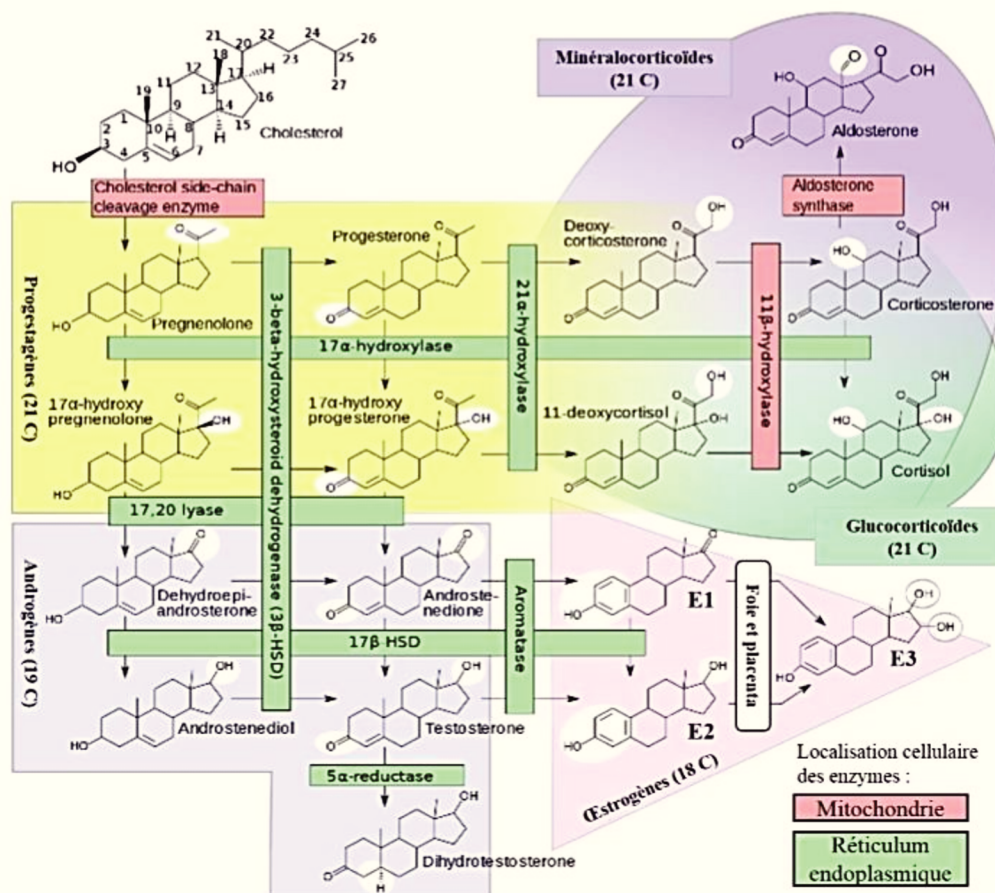




**Figure 1 : Structures chimiques des œstrogènes naturels.**

La structure chimique des œstrogènes correspond à un noyau œstrane, constitué de quatre cycles hydrocarbonés, pour un total de 18 atomes de carbone. Différentes modifications, comme des substitutions de groupements hydroxyles, permettent de différencier les molécules de cette famille.

(D'après Zimmerman, 2019)



**Figure 2 : Les étapes de la biosynthèse des œstrogènes.**

Les hormones stéroïdiennes, sont formées à partir du cholestérol qui subit une série de clivages et de modifications pour former dans un premier temps les progéstagènes (21 Carbones, zone jaune) qui sont les précurseurs des gluco-corticoïdes (21C, zone verte), des minéralo-corticoïdes (21C, zone violette) et des androgènes (19C, zone bleue). L'œstrone (E<sub>1</sub>) et le 17β-œstradiol (E<sub>2</sub>) sont ensuite synthétisés respectivement à partir de l'androstenedione et de la testostérone via l'activation de l'enzyme cytochrome P450 aromatasé. L'enzyme 17β-hydroxystéroïde déshydrogénase (17β-HSD) assure la conversion réversible E<sub>1</sub>/E<sub>2</sub>. Enfin, au niveau du foie et du placenta, une 16α-hydroxylase assure la conversion réversible de l'E<sub>1</sub> ou de l'E<sub>2</sub> en œstriol (E<sub>3</sub>). La biosynthèse de l'œstétrol (E<sub>4</sub>) n'est pas représentée ici.

(D'après Häggström et Richfield, 2014)

# Chapitre 1 : La signalisation œstrogénique

*C'est en juin 1905 que le mot « hormone » fut prononcé pour la première fois par Ernest Starling, professeur de physiologie à l'University College of London. Ce mot tiré du grec « hormôn » qui signifie « exciter » a été défini à cette époque comme désignant « les messagers chimiques qui se déplacent d'une cellule à l'autre grâce au flux sanguin et qui coordonnent les activités et le développement de différentes parties du corps ». Aujourd'hui, les hormones sont au cœur du fonctionnement du système endocrinien, qui est un des grands systèmes de communication de l'organisme et la compréhension du mécanisme d'action de ces hormones a permis de grandes avancées scientifiques notamment dans le domaine de la santé humaine.*

## I. Les œstrogènes :

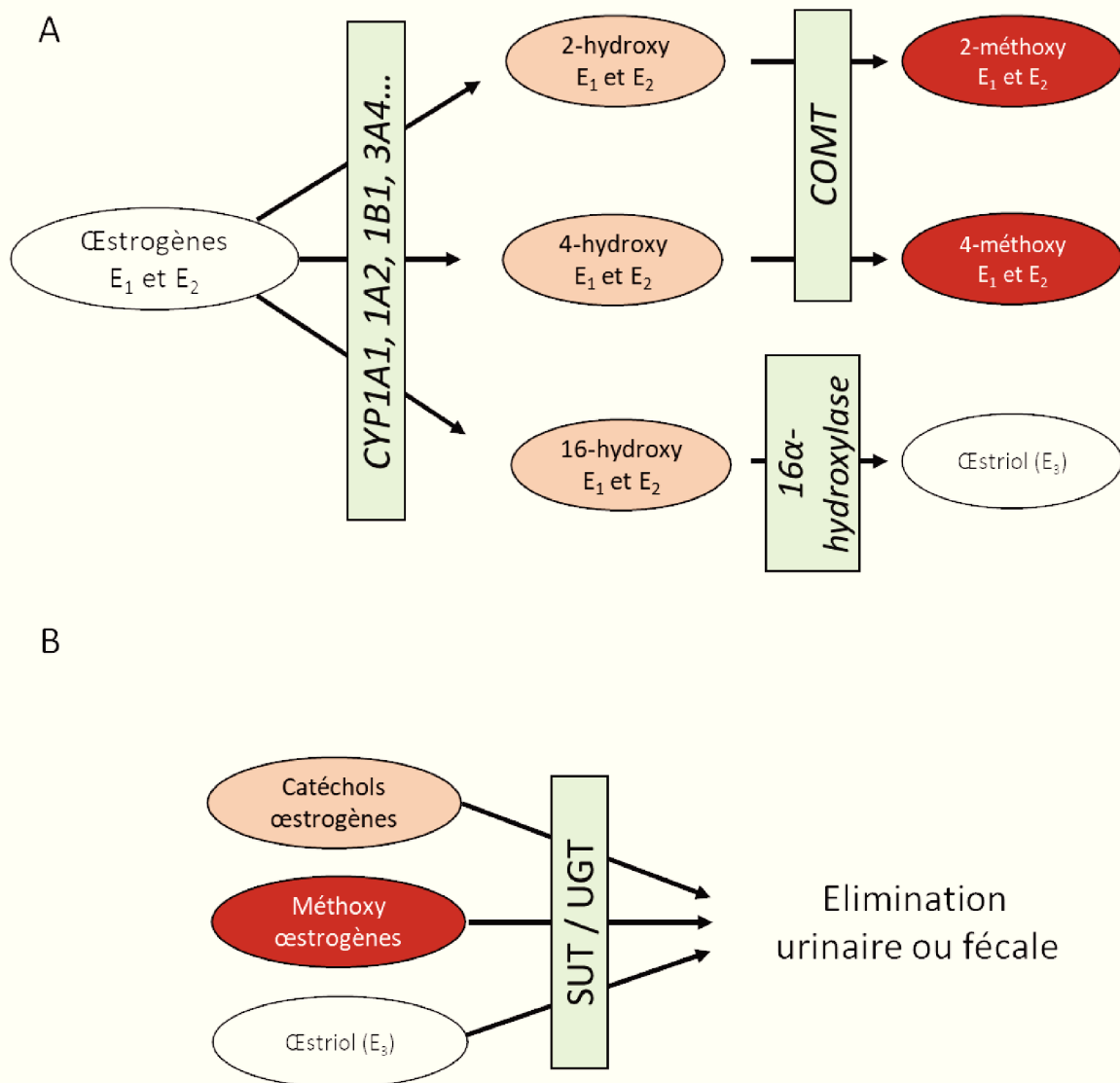
*C'est dans les années 1920-1930 que Adolf Butenandt, Tadeus Reichstein et Edward Adelbert Doisy, ont purifié et caractérisé les premières hormones stéroïdiennes. Parmi ces hormones lipophiles qui dérivent du cholestérol on trouve la testostérone, la progestérone mais aussi les œstrogènes.*

### A) Structure et étapes de la biosynthèse des œstrogènes

Le terme « œstrogène » désigne un ensemble de 4 hormones stéroïdiennes naturellement produites par l'organisme : l'œstrone (E<sub>1</sub>), le 17β-œstradiol (E<sub>2</sub>), l'œstriol (E<sub>3</sub>) et l'œstétrol (E<sub>4</sub>).

Au niveau structurelle, les œstrogènes sont tous constitués de 18 atomes de carbone qui définissent un squelette carboné à 4 cycles. Les différences entre les hormones E<sub>1</sub>, E<sub>2</sub>, E<sub>3</sub> et E<sub>4</sub> se localisent au niveau du cycle D qui peut comporter un nombre variable de groupements hydroxyles ou un groupement cétone (Figure 1).

Comme toutes les hormones stéroïdiennes, les œstrogènes sont synthétisés à partir du cholestérol (27 carbones) par l'intermédiaire d'une voie de biosynthèse enzymatique qui se déroule principalement au niveau des ovaires chez la femme non ménopausée et au niveau placentaire au cours de la grossesse. Lorsque la fonction ovarienne cesse au moment de la ménopause les œstrogènes sont produits en moindre quantité par les tissus périphériques, comme les glandes corticosurrénales, le foie, le cerveau et le tissu adipeux.



**Figure 3 : Les étapes du catabolisme des œstrogènes.**

- (A) La métabolisation de l'œstrone ( $E_1$ ) et du  $17\beta$ -œstradiol ( $E_2$ ) débute par une étape d'hydroxylation, catalysée par différentes enzymes Cytochromes P450 (CYP), qui conduisent à la formation des catéchols œstrogènes (2-, 4-, 16- hydroxy  $E_1$  ou  $E_2$ ). Les 2- et 4- catéchols œstrogènes sont ensuite métabolisés en 2- et 4- Méthoxy-œstrogènes par l'enzyme catechol-O-méthyltransferase (COMT). Les 16- catéchols œstrogènes subissent une nouvelle hydroxylation qui conduit à la synthèse d'œstriol ( $E_3$ ).
- (B) Les catéchols œstrogènes, les méthoxy-œstrogènes et l'œstriol peuvent ensuite être conjugués à l'acide glucuronique ou à un groupement sulfate via les enzymes sulfotransférases (SUT) et UDP-glucuronosyltransférases (UGT) afin de permettre leur élimination urinaire ou fécale.

(D'après Samavat et Kurzer, 2015)

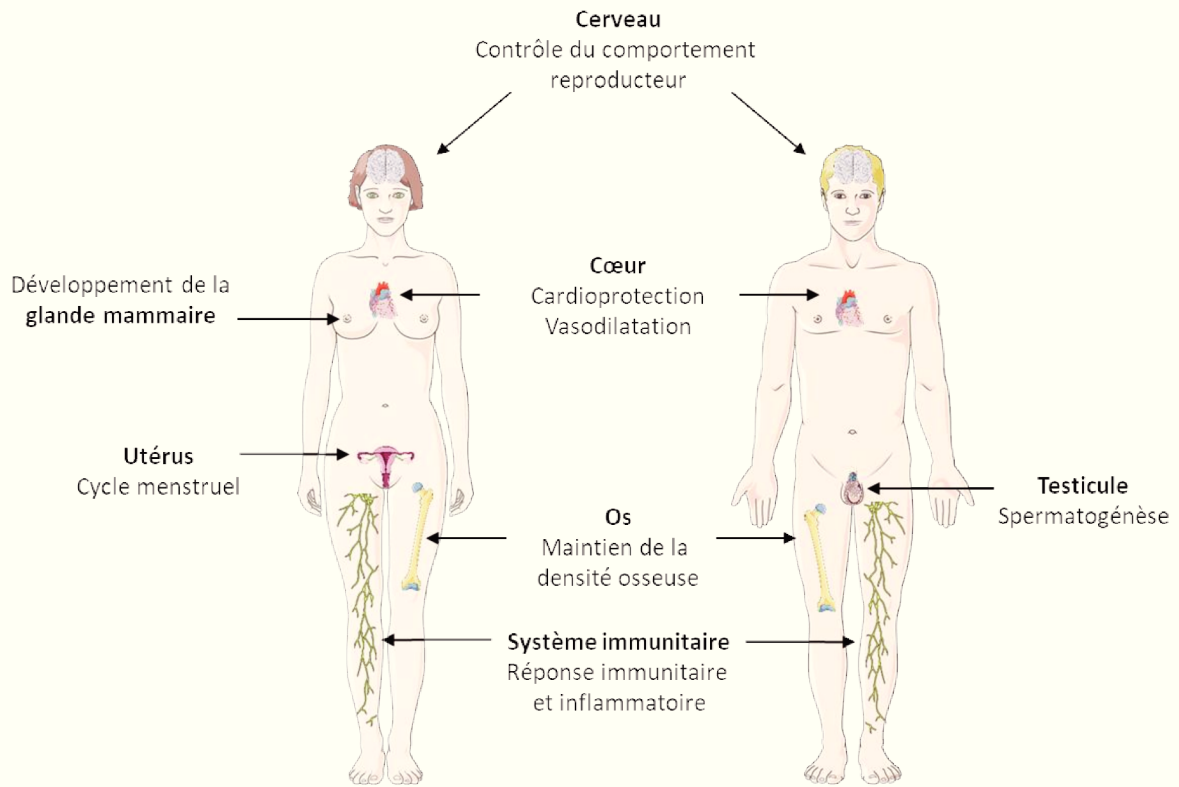
Chez l'homme la stéroïdogénèse se déroule principalement au niveau testiculaire. Elle est initiée dans les cellules de Leydig et elle se termine par l'aromatation des androgènes dans les cellules de Sertoli (Josso et Rey, 1995 ; Carreau et al., 2011). Il existe également une source secondaire d'œstrogènes, produits par les tissus périphériques.

La stéroïdogénèse débute par le clivage de la chaîne latérale du cholestérol par l'enzyme cytochrome P450<sub>sc</sub> (side chain clivage) localisée au niveau de la membrane interne mitochondriale. Cette première étape conduit à la formation de la prégnénolone qui va ensuite rejoindre le réticulum endoplasmique où elle sera transformée en progestérone sous l'action de l'enzyme 3 $\beta$ -HSD (3 $\beta$ -hydroxysteroid dehydrogenase). Suite à l'intervention de 2 oxydoréductases et d'une lyase (17 $\alpha$ -hydroxylase, 17-20 lyase, 17  $\beta$ -HSD) la progestérone est convertie en androgènes, androsténone et testostérone, qui seront transformés par l'enzyme cytochrome P450 aromatasase respectivement en œstrone (E<sub>1</sub>) et en 17 $\beta$ -œstradiol (E<sub>2</sub>). Concernant l'œstriol (E<sub>3</sub>), il est obtenu à partir de l'E<sub>1</sub> ou de l'E<sub>2</sub> grâce à l'action d'enzymes cytochrome P450 (CYP3A4, CYP3A5) exprimées uniquement au niveau du foie et du placenta. Enfin, l'œstérol (E<sub>4</sub>) est synthétisé à partir de l'E<sub>2</sub> et l'E<sub>3</sub> dans le foie fœtal par l'action enzymatique des 15 $\alpha$ - et 16 $\alpha$ -hydroxylases (Coelingh Bennink et al., 2008) (Figure 2).

### B) Transport et métabolisme des œstrogènes

Immédiatement après leur synthèse les œstrogènes sont libérés dans la circulation sanguine. Pour permettre leur transport vers les tissus cibles, ces hormones lipophiles se lient de façon réversible à une globuline présentant une forte affinité pour les hormones stéroïdiennes : la SHBG (Sex Hormone Binding Globulin). Elles peuvent également se lier de façon non spécifique à l'albumine ou rester sous forme libre (2 à 3% de l'œstradiol circulant).

Concernant le catabolisme de ces hormones, le 17 $\beta$ -œstradiol et l'œstrone peuvent être hydroxylés en position C2, C4 ou C16 par différentes enzymes cytochrome P450. Cela conduit à la synthèse de dérivés appelés catéchols œstrogènes (2-,4- ou 16-hydroxyestrone et hydroxyestradiol). Ces composés peuvent ensuite être métabolisés en méthoxy-œstrogènes par l'action de l'enzyme catechol-O-méthyltransferase (COMT). Afin de permettre leur élimination dans les urines ou les fèces l'ensemble de ces composés (catéchols œstrogènes et méthoxy-œstrogènes) peuvent être conjugués à l'acide glucuronique ou à un groupement sulfate par des enzymes hépatiques : les UDP-glucuronosyltransférases et les sulfotransférases (Figure 3) (Samavat et Kurzer, 2015).



**Figure 4 : Le rôle physiologique des œstrogènes.**

Les œstrogènes jouent un rôle ubiquitaire en intervenant dans la régulation d'un certain nombre de processus physiologiques chez la femme mais également chez l'homme. Ces effets sont prédominants chez la femme qui présente une concentration plasmatique en œstrogènes plus élevée que chez l'homme.

### C) Rôles physiologiques des œstrogènes

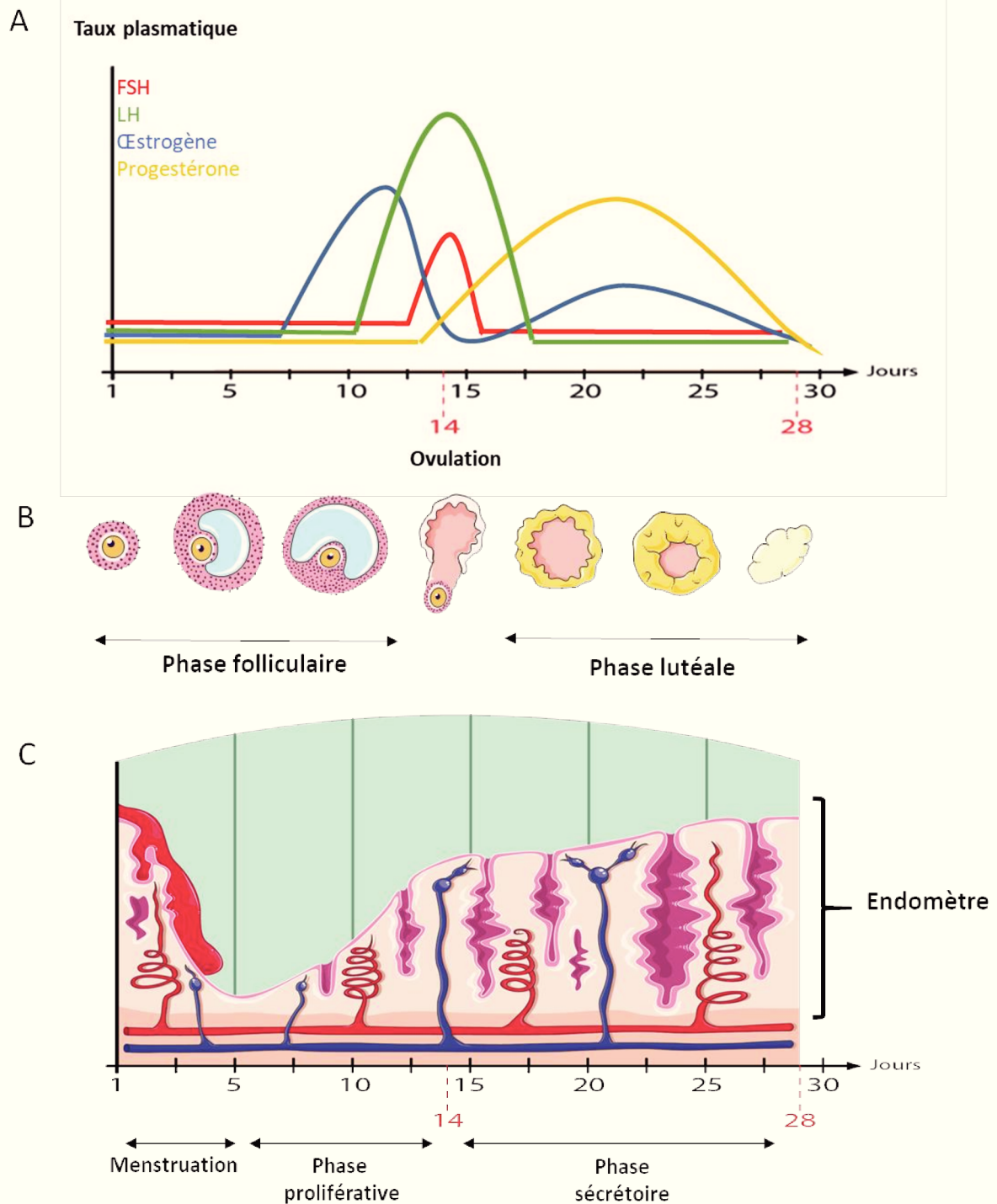
Les œstrogènes interviennent dans la régulation d'un grand nombre de processus physiopathologiques chez la femme mais également chez l'homme. Ces hormones sont surtout connues pour leur rôle dans le fonctionnement de l'appareil reproducteur mais elles ont également un effet sur les systèmes neuroendocrinien, cardiovasculaire, immunitaire et squelettique (Figure 4).

#### *i) Rôle dans la fonction de reproduction*

##### ➤ La fonction de reproduction chez la femme

La proportion relative des différentes hormones qui constituent le groupe des œstrogènes dans l'organisme varie au cours de la vie. Chez la femme non ménopausée les œstrogènes sont majoritairement présents sous la forme de  $17\beta$ -œstradiol alors qu'après la ménopause c'est l'œstrone qui est majoritaire. Quant à l'œstriol et l'œstétrol, qui sont produits respectivement par le placenta et le foie fœtal, ils sont prédominants au cours de la grossesse (Holinka et al., 2008).

Les œstrogènes jouent un rôle central lors du cycle menstruel. Lors de la phase folliculaire qui précède l'ovulation, les œstrogènes sont sécrétés par les follicules ovariens en réponse à la production de LH (Luteinizing Hormone) et de FSH (Follicle-Stimulating Hormone) par l'hypophyse. Cette sécrétion de LH/FSH est elle-même induite par la production de LHRH (Luteinizing Hormone Releasing Hormone) par l'hypothalamus. Durant cette phase folliculaire, l'augmentation progressive du taux d'œstrogènes ( $17\beta$ -œstradiol) est responsable de l'épaississement de l'endomètre et de la mise en place d'un rétrocontrôle positif sur la sécrétion de LHRH, de LH et de FSH par l'axe hypothalamo-hypophysaire. Lorsque le taux de LH atteint un seuil de concentration suffisamment élevé, le follicule mature se rompt afin de libérer l'ovocyte puis il dégénère et se transforme en corps jaune. Lors de cette phase lutéale qui suit l'ovulation, le corps jaune sécrète principalement de la progestérone ainsi qu'une faible quantité d'œstrogènes. La combinaison de ces hormones permet de préparer l'endomètre à l'implantation de l'œuf fécondé et conduit à un rétrocontrôle négatif de l'axe hypothalamo-hypophysaire qui se traduit par une diminution du taux de LH. En absence de fécondation de l'ovocyte, le taux de progestérone diminue ce qui conduit à une desquamation de la muqueuse utérine se traduisant par la survenue des règles, et à une levée de l'inhibition de l'axe



**Figure 5 : Le cycle menstruel.**

Le taux plasmatique des hormones hypophysaires (LH et FSH), des œstrogènes et de la progestérone (A) conditionne le déroulement du cycle ovarien (B) et du cycle utérin (C). En moyenne, le cycle menstruel dure 28 jours et l'ovulation survient le 14<sup>ème</sup> jour de ce cycle. Le cycle ovarien se compose de 2 phases successives durant chacune 14 jours : la phase folliculaire et la phase lutéale. Le cycle utérin, qui correspond à un remodelage cyclique de l'endomètre se divise en 3 phases : La phase de menstruation également dite phase de desquamation, la phase proliférative de régénération de l'endomètre et la phase sécrétoire.

(D'après Hawkins et Matzuk, 2008)

hypothalamo-hypophysaire. Le cycle peut alors recommencer (Figure 5)(Hawkins et Matzuk, 2008).

Les œstrogènes interviennent également dans le développement post-pubertaire de la glande mammaire en permettant l'élongation des canaux au moment de la puberté ainsi que le développement alvéolaire au moment de la gestation. Cette glande est sensible aux variations cycliques du taux d'œstrogènes circulants avec une expression des récepteurs aux œstrogènes qui est plus élevée lors de la phase folliculaire que lors de la phase lutéale (Gruber et al., 2002).

Bien que les œstrogènes soient indispensables au bon déroulement de la fonction de reproduction chez la femme, ces hormones sont impliquées dans la progression des cancers hormono-dépendants comme le cancer du sein ou de l'endomètre.

➤ La fonction de reproduction chez l'homme

Même si les œstrogènes sont plutôt considérés comme une hormone féminine, ils jouent également un rôle important dans la régulation de la fonction de reproduction chez l'homme.

En effet, un déficit congénital en aromatasase et donc une baisse du taux d'œstrogènes chez l'homme est associée à une diminution du nombre, de la mobilité et du pouvoir fécondant des spermatozoïdes (Rochira et Carani, 2009). Ce constat souligne l'importance de la signalisation œstrogénique dans la régulation de la spermatogénèse et de la maturation fonctionnelle des spermatozoïdes (capacitation et réaction acrosomiale)(Carreau et al., 2011).

Enfin, chez l'homme comme chez la femme, les œstrogènes ont également un effet au niveau cérébral. En effet, on observe au niveau de l'hypothalamus antérieur et de l'aire préoptique médiane une colocalisation de l'expression de l'aromatase, des récepteurs aux œstrogènes et des récepteurs à la sérotonine. Ces régions du cerveau sont décrites pour leur rôle dans le contrôle du comportement reproducteur (Schulster et al., 2016).

*ii) Rôles cardio-vasculaires*

Les œstrogènes semblent avoir un effet protecteur sur le système cardio-vasculaire. En effet, en contribuant à diminuer les taux sériques de LDL-cholesterol tout en augmentant le taux de HDL-cholestérol et de triglycérides, ils permettent de limiter le risque d'apparition d'athérosclérose. De plus, les œstrogènes ont un effet hypotenseur et vasodilatateur en induisant la production d'oxyde nitrique et de prostacyclines dans les cellules endothéliales et en diminuant le tonus des muscles lisses vasculaires (White, 2002).



Ces effets bénéfiques des œstrogènes pourraient expliquer pourquoi les femmes sont moins sujettes que les hommes aux maladies cardiovasculaires. Cependant une étude de 2012, suggère que des taux élevés d'œstradiol chez les femmes ménopausées augmentent de façon significative le risque d'infarctus du myocarde et d'accident vasculaire cérébral (Scarabin-Carré et al., 2012).

### *iii) Rôles sur le système squelettique*

Les œstrogènes ont un effet protecteur sur le système squelettique. Par l'intermédiaire de leurs récepteurs qui sont exprimés dans les cellules osseuses ces hormones favorisent l'activité des ostéoblastes qui contribuent à la formation osseuse, tout en limitant celle des ostéoclastes qui sont impliqués dans la résorption osseuse. Une diminution du taux d'œstrogènes circulants notamment chez la femme ménopausée est associée à une perte de masse osseuse et à l'apparition d'ostéoporose (Khosla et al., 2012).

Chez l'homme, ce sont également les œstrogènes qui sont les principaux acteurs du maintien de l'homéostasie osseuse.

### *iv) Rôles sur le système immunitaire*

Les récepteurs aux œstrogènes sont exprimés dans la plupart des cellules qui participent à la réponse immunitaire innée (les neutrophiles, les macrophages, les cellules Natural Killer et les cellules dendritiques) et adaptative (les lymphocytes T et les lymphocytes B) (Kovats, 2015). Les œstrogènes sont capables de réguler différents aspects de la réponse immunitaire et inflammatoire en modulant notamment la production de cytokines (IFN  $\gamma$ , IL-4-6-10-12, TGF $\beta$ , TNF $\alpha$ ) et de chimiokines (CINC-1, CXCL-8)(Khan et Ansar Ahmed, 2016).

L'implication des œstrogènes dans la réponse immunitaire pourrait expliquer pourquoi les femmes sont plus sensibles que les hommes aux maladies auto immunes comme la sclérose en plaque ou le lupus érythémateux systémique. Cependant, ce biais hormonal n'explique pas à lui seul la prévalence de ces maladies chez les femmes. En effet, cela pourrait également être lié à une dérégulation de certains mécanismes de contrôle épigénétiques qui conduirait à l'expression pathologique de gènes portés par le chromosome X inactivé, favorisant l'auto-immunité (Brooks et Renaudineau, 2015 ; Khan et Ansar Ahmed, 2016).

**Tableau I : Classification des récepteurs nucléaires.**

Classe	Récepteurs associés
NR0	Récepteurs DAX-1 et SHP
NR1	Récepteurs aux hormones thyroïdiennes (TR) Récepteurs à la vitamine D (VDR) Récepteurs à l'acide rétinoïque (RAR) Récepteurs PPAR Récepteurs aux xénobiotiques (CAR, PXR) Récepteurs aux cholestérols et dérivés (ROR, LXR)
NR2	Récepteurs à l'acide rétinoïque (RXR) Récepteurs testiculaires (TR) Récepteurs orphelins (HNF-4, COUP-TF I et II)
NR3	Récepteurs aux œstrogènes (ER) et les récepteurs apparentés (ERR) Récepteurs à la progestérone (PR) Récepteurs aux androgènes (AR) Récepteurs aux glucocorticoïdes (GR) et minéralocorticoïdes (MR)
NR4	Récepteurs orphelins (NGFIB, NURR1, NOR1)
NR5	Récepteurs orphelins SF1 et LHR-1
NR6	Récepteurs orphelins de type GCNF

Cette classification a été établie en 1997 par Laudet et al., et validée en 1999 par le comité de nomenclature des récepteurs nucléaires. Celle-ci repose sur des analyses phylogénétiques et sur la comparaison des modes de liaison à l'ADN et des capacités de dimérisation des récepteurs. Elle ne prend pas en compte la nature des ligands.

Les facteurs de la classe NR0 ont la particularité de ne pas posséder de domaine de liaison à l'ADN. Les récepteurs orphelins sont des récepteurs pour lesquels aucun ligand n'a, à l'heure actuelle, été identifié.

**DAX-1** = Dosage sensitive sex reversal, Adrenal hypoplasia critical region, on chromosome X, gene 1 ; **SHP** = Small Heterodimer Partner ; **RAR** = Retinoic Acid Receptor ; **PPAR** = Peroxisome Proliferator-Activated Receptors ; **ROR** = Retinoic acid receptor-related Receptors ; **CAR** = Constitutive Androstane Receptor ; **PXR** = Pregnane X Receptor ; **LXR** = Liver X Receptor ; **RXR** = Retinoid X Receptors ; **HNF-4** = Hepatic Nuclear Factor-4 ; **COUP-TF I et II** = Chicken Ovalbumin Upstream Promoter-Transcription Factor I and II ; **ERR** = Estrogen-Related Receptors ; **NURR1** = Nuclear Receptor Related 1 ; **NOR1** = Neuron-derived Orphan Receptor 1 ; **NGFIB** = Nerve Growth Factor IB ; **SF1** = Steroidogenic Factor 1 ; **LHR-1** = Liver Receptor Homolog-1 ; **GCNF** = Germ Cell Nuclear Factor-like.

## II. Les récepteurs aux œstrogènes

Classiquement, les effets des hormones peuvent être médiés par 2 types de récepteurs : les récepteurs membranaires et les récepteurs nucléaires.

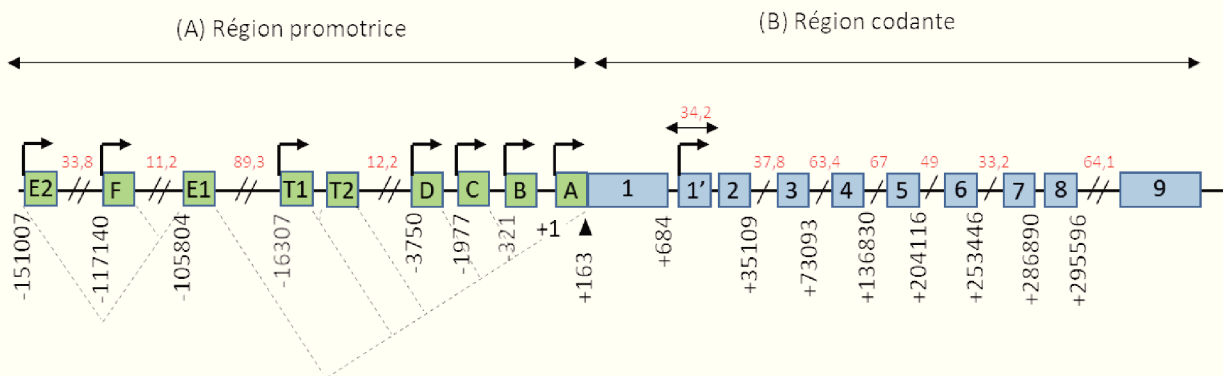
Les récepteurs membranaires sont plutôt associés à l'action des hormones polypeptidiques (hormone de croissance, insuline...), des neurotransmetteurs et de la plupart des facteurs de croissance (EGF, VEGF, IGF...). L'interaction spécifique d'un ligand avec son récepteur membranaire permet de déclencher des cascades de signalisation impliquant un certain nombre de seconds messagers et d'effecteurs qui sont à l'origine d'un effet rapide (Tata, 2005).

Les récepteurs nucléaires quant à eux sont plutôt associés à l'action des hormones stéroïdiennes et d'autres messagers lipophiles, comme la vitamine D et l'acide rétinoïque, qui sont capables de traverser la membrane plasmique par diffusion ou transport facilité. Lorsqu'ils sont activés par leur ligand, ces récepteurs nucléaires agissent comme des facteurs de transcription capables de réguler directement l'expression de leurs gènes cibles (Puzianowska-Kuznicka et al., 2013).

### A) Les récepteurs nucléaires des œstrogènes : ER $\alpha$ et ER $\beta$

C'est dans les années 1960 que l'existence des récepteurs nucléaires a été évoquée pour la première fois suite aux travaux de Jensen et Jacobson. En suivant le devenir du 17 $\beta$ -œstradiol radio-marqué ([<sup>3</sup>H]-œstradiol) dans le tissu utérin, ils ont mis en évidence la formation d'un complexe intracellulaire constitué du 17 $\beta$ -œstradiol et d'une protéine qu'ils nommeront estrophiline (Jensen, 1962). Ce fut le point de départ de plusieurs décennies de travaux de recherche qui aboutiront au clonage du récepteur aux œstrogènes alpha (ER $\alpha$ ) en 1986 à partir d'une banque d'ADNc issus des cellules cancéreuses mammaires MCF-7. Dix ans plus tard, ces travaux seront complétés par le clonage du récepteur aux œstrogènes bêta (ER $\beta$ ) à partir d'une banque d'ADNc issus de prostate de rat (Greene et al., 1986 ; Kuiper et al., 1996).

De manière plus générale, les récepteurs aux œstrogènes ER $\alpha$  et ER $\beta$  font partis des 48 membres de la famille des récepteurs nucléaires actuellement identifiés chez l'Homme. Ils appartiennent à la classe NR3 qui regroupe l'ensemble des récepteurs aux hormones stéroïdiennes (Tableau I).



**Figure 6 : L'organisation génomique du gène ESR1.**

(A) La région promotrice du gène ESR1 contient 9 exons non codants (cadres verts), annotés de A à F - T1 et T2, et 7 promoteurs alternatifs possédant chacun leur propre TSS (flèches). On trouve également un site accepteur d'épissage localisé en position +163 pb (▲). Les lignes pointillées symbolisent les différentes possibilités d'épissage des exons non codants décrites dans la littérature. Une petite particularité concerne les exons E1 et T2 dont l'expression dépend respectivement des promoteurs E2 et T1.

(B) La région codante du gène ESR1 se compose de 10 exons (cadres bleus), numérotés de 1-1' à 9, séparés par 9 introns. On trouve également un 8<sup>ème</sup> promoteur alternatif localisé au niveau de l'exon 1'.

Sur cette représentation schématique, les échelles ne sont pas respectées mais certaines positions estimées sont indiquées en dessous des exons, en paires de bases (pb). La position +1 correspond au premier site d'initiation de la transcription (TSS) décrit par Greene et al., en 1986. Les indications en rouge correspondent à la longueur des longues régions introniques en kilo paires de bases (kpb).

(D'après Kos et al., 2001, Wang Z. et al., 2005)

**Tableau II : Utilisation différentielle des différents promoteurs du gène ESR1 dans les tissus cibles des œstrogènes.**

Tissus cibles des œstrogènes	Promoteur(s) du gène ESR1 associé(s)
Glande mammaire	Promoteurs A et C
Utérus	Promoteur C
Testicule	Promoteurs T1 et T2
Os	Promoteur F
Foie	Promoteurs E1 et E2
Ovaires	Non déterminé

(D'après Kos et al., 2001 ; Pennolazzi et al., 2004)

### *i) Organisation génomique des gènes ESR1 et ESR2*

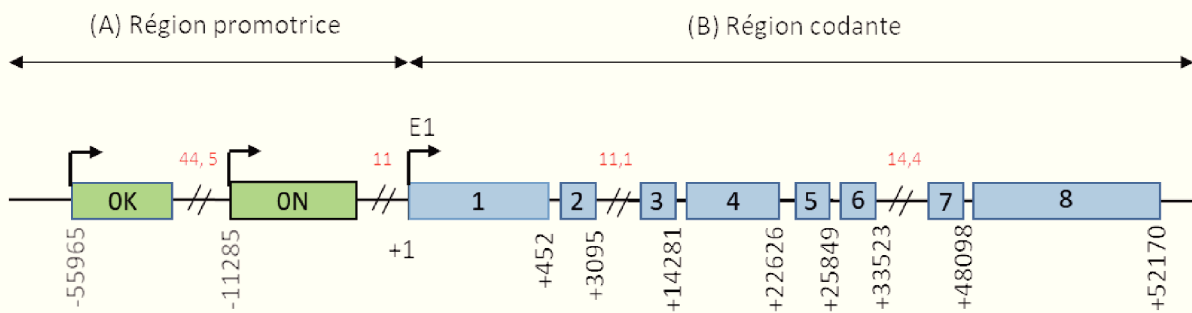
Les récepteurs ER $\alpha$  et ER $\beta$  sont codés respectivement par le gène ESR1, localisé sur le chromosome 6 (6q25.1), et le gène ESR2 localisé sur le chromosome 14 (14q23.2)(Menasce et al., 1993 ; Mosselman et al., 1996).

#### ➤ Le gène ESR1

La région promotrice du gène ESR1 qui occupe un peu plus de 150 kb se compose d'au moins 9 exons non codants et de 7 promoteurs alternatifs (Figure 6) qui permettent de réguler de façon différentielle l'expression d'ER $\alpha$  dans les différents tissus cibles des œstrogènes (Tableau II). Le premier site d'initiation de la transcription (TSS), localisé au niveau de l'exon non codant « A », a été identifié par Greene et al., en 1986. Depuis, il a été mis en évidence que la majorité des exons non codants possèdent leur propre TSS (Kos et al., 2001). De plus, Keaveney et al., (1991), ont démontré l'existence d'un site accepteur d'épissage localisé 163 pb en aval du TSS de l'exon A. Ce site, très conservé entre les espèces, est essentiel puisqu'il constitue le point d'ancrage pour les différents exons non codants de la région promotrice.

Afin de faciliter la compréhension et la comparaison des données obtenues par les différentes équipes travaillant sur la région promotrice d'ESR1, Kos et al., (2001) ont proposé une nomenclature unifiée. Les exons non codants sont identifiés par une lettre capitale allant de A à F et ceux dont l'utilisation est restreinte au testicule ont été nommés T1 et T2. Avant cette nouvelle nomenclature officielle il y avait beaucoup de confusion autour de la dénomination des promoteurs du gène ESR1. En effet, les premières études réalisées avaient uniquement mis en évidence l'existence d'un promoteur proximal (également appelé P1 ou A) et d'un promoteur distal (également appelé P0 ou B). Suite à la découverte des autres promoteurs alternatifs, le promoteur distal, qui été appelé promoteur B, est devenu dans la nouvelle nomenclature le promoteur C. Dans ce manuscrit, nous utiliserons la nomenclature officielle de Kos et al., (2001) ce qui nous a conduit à actualiser le nom des promoteurs cités dans les publications datant d'avant 2001.

Concernant la région codante du gène ESR1, elle s'étend sur environ 300kb et se compose de 10 exons et de 9 introns qui sont à l'origine de 4 isoformes : ER $\alpha$ 66, ER $\alpha$ 46, ER $\alpha$ 36 et ER $\alpha$ 30. On trouve également un 8<sup>ème</sup> promoteur alternatif localisé en amont de l'exon 1' (Kos et al., 2001 ; Wang Z. *et al.*, 2005 ; Gibson et Saunders, 2012).



**Figure 7 : L'organisation génomique du gène ESR2.**

(A) La région promotrice du gène ESR2 contient 2 exons non codants (cadres verts), nommés OK et ON, associés à 2 promoteurs alternatifs possédant chacun un site d'initiation de la transcription TSS (flèches).

(B) La région codante du gène ESR1 se compose de 8 exons (cadres bleus), numérotés de 1 à 8, séparés par 7 introns. On trouve également un troisième promoteur alternatif (E1) associé à l'exon codant n°1.

Sur cette représentation schématique, les échelles ne sont pas respectées mais certaines positions sont indiquées en dessous des exons, en paires de bases (pb). La position +1 correspond au site d'initiation de la transcription (TSS) du promoteur E1. Les indications en rouge correspondent à la longueur des longues régions introniques en kilo paires de bases (kpb).

(D'après Haldosén et al., 2014 ; Al-nakhle et al., 2013)

➤ Le gène ESR2

La région promotrice du gène ESR2 qui s'étend sur environ 55 kb est constituée de 3 promoteurs alternatifs localisés en position distale (0K), centrale (0N) et proximale (E1). Les promoteurs 0K et 0N sont chacun associés à un exon non codant portant le même nom, tandis que le promoteur E1 est associé au premier exon codant du gène ESR2 (Figure 7) (Al-nakhle *et al.*, 2013 ; Haldosén *et al.*, 2014). La présence d'autres promoteurs alternatifs dans cette région est fortement suggérée par l'existence de transcrits ER $\beta$  pour lesquels la région 5'UTR est différente de celle des transcrits obtenus à partir des promoteurs 0K, 0N et E1. Toutefois ces promoteurs n'ont, à l'heure actuelle, pas été identifiés (Zhao *et al.*, 2008).

Concernant la région codante du gène ESR2, elle s'étend sur environ 60 kb et se compose de 8 exons et 7 introns qui sont à l'origine de 5 variants d'épissage : ER $\beta$ 1, ER $\beta$ 2, ER $\beta$ 3, ER $\beta$ 4 et ER $\beta$ 5.

*ii) Organisation protéique d'ER $\alpha$ 66, ER $\beta$ 1 et leurs isoformes*

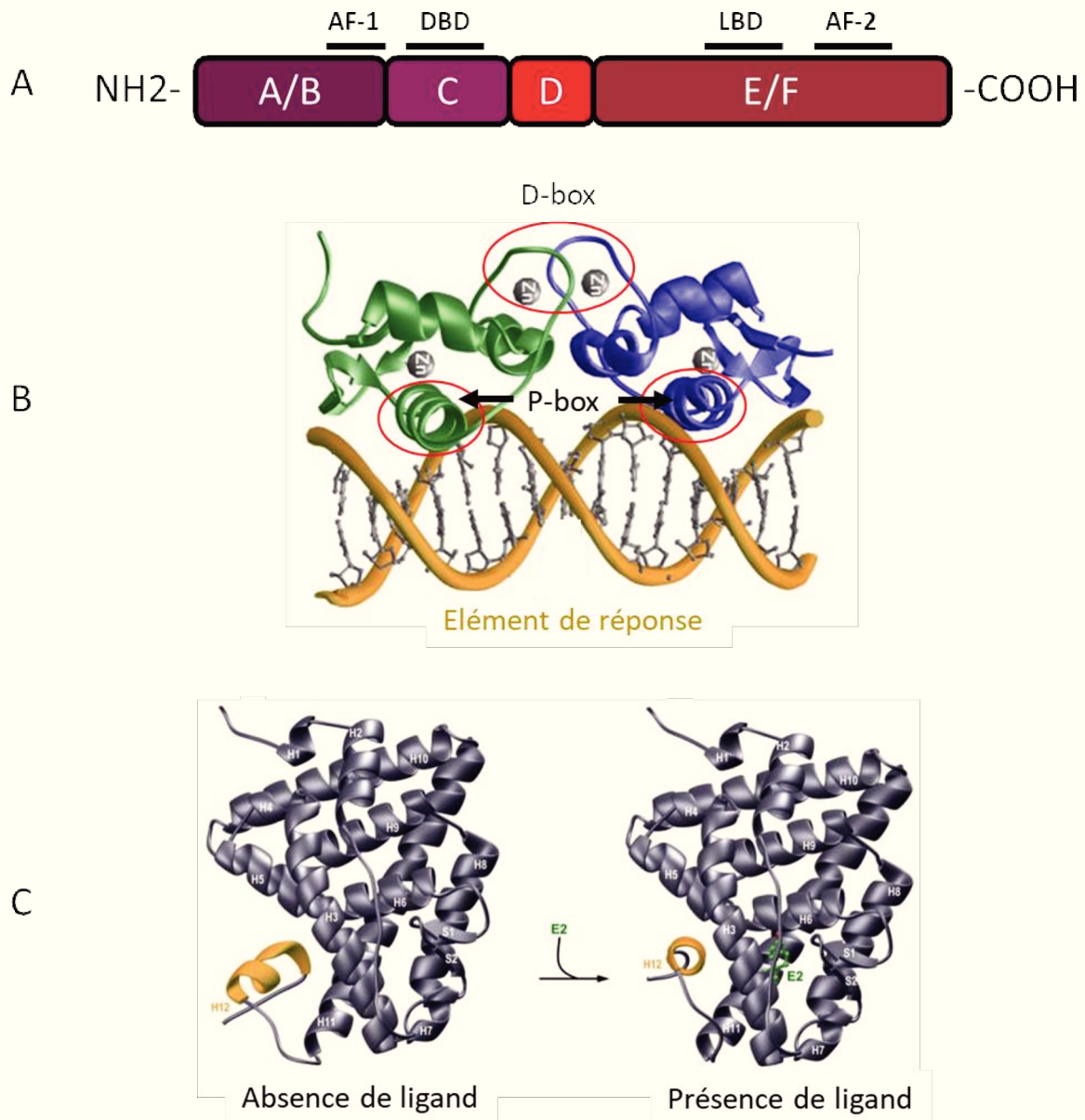
➤ Structure générale des récepteurs nucléaires

Tous les membres de la superfamille des récepteurs nucléaires présentent une grande homologie structurale avec une organisation en 6 domaines fonctionnels désignés par une lettre capitale allant de A à F. (Figure 8A)

Depuis l'extrémité N-terminale (-NH<sub>2</sub>) vers l'extrémité C-terminale (-COOH) on trouve :

✓ **Le domaine A/B de transactivation indépendante du ligand**

Le domaine A/B localisé à l'extrémité N-terminale est le domaine le moins conservé entre les membres de la superfamille des récepteurs nucléaires, que ce soit en terme de longueur ou d'homologie de la séquence en acides aminés. Cette région contient le domaine de transactivation AF-1 qui permet d'activer la transcription de façon ligand-indépendante. De plus, le domaine A/B possède plusieurs sites de phosphorylation et de sumoylation qui permettent le recrutement de coactivateurs ou de corépresseurs pouvant moduler la liaison des récepteurs nucléaires à leurs éléments de réponse spécifiques (McEwan, 2016).



**Figure 8 : Structure générale des récepteurs nucléaires.**

- (A) Les récepteurs nucléaires sont caractérisés par leur structure en 6 domaines. De l'extrémité NH<sub>2</sub>- à l'extrémité COOH-terminale, on distingue (i) le domaine A/B, avec la région AF-1 (Activation Function 1) qui permet la transactivation indépendante du ligand, (ii) le domaine C avec la région DBD (DNA Binding Domain) de liaison à l'ADN, (iii) le domaine D charnière, (iv) le domaine E/F qui possède le domaine de liaison au ligand LBD (Ligand Binding Domain) et la région AF-2 (Activation Function 2) pour la transactivation ligand-dépendante.
- (B) Structure cristallographique du DBD d'un homodimère de récepteur nucléaire en interaction avec l'ADN au niveau d'un élément de réponse spécifique. Le DBD vert et le DBD bleu proviennent chacun d'un monomère de récepteur nucléaire. Les D-box impliquées dans la dimérisation et les P-box impliquées dans la reconnaissance de l'élément de réponse sont entourées en rouge.
- (C) Structure cristallographique du LBD du récepteur aux œstrogènes en absence et en présence d'un ligand (E<sub>2</sub>). Les hélices constituant ce domaine sont numérotées de H1 à H12. L'interaction avec le ligand conduit à une modification conformationnelle de la protéine. On observe notamment un mouvement de l'hélice H12.

(D'après Farooq et al., 2015)



✓ **Le domaine C de liaison à l'ADN**

Le domaine C, également appelé DBD (DNA Binding Domain), est un domaine très conservé qui permet la fixation des récepteurs nucléaires au niveau d'éléments de réponse spécifiques localisés dans la région promotrice des gènes cibles (Figure 8B). Ce domaine se compose de 2 motifs en doigt de zinc chacun constitués de 4 cystéines entourant un ion  $Zn^{2+}$ . A la base du premier doigt de zinc se trouve un motif P-box (boîte proximale) impliqué dans la reconnaissance des éléments de réponse, tandis que le second doigt de zinc est associé à un motif D-box (boîte distale) qui est lui nécessaire à la dimérisation du récepteur (Ponglikitmongkol et al., 1988 ; Kumar et Thompson, 1999 ; Aranda et Pascual, 2001).

✓ **Le domaine D**

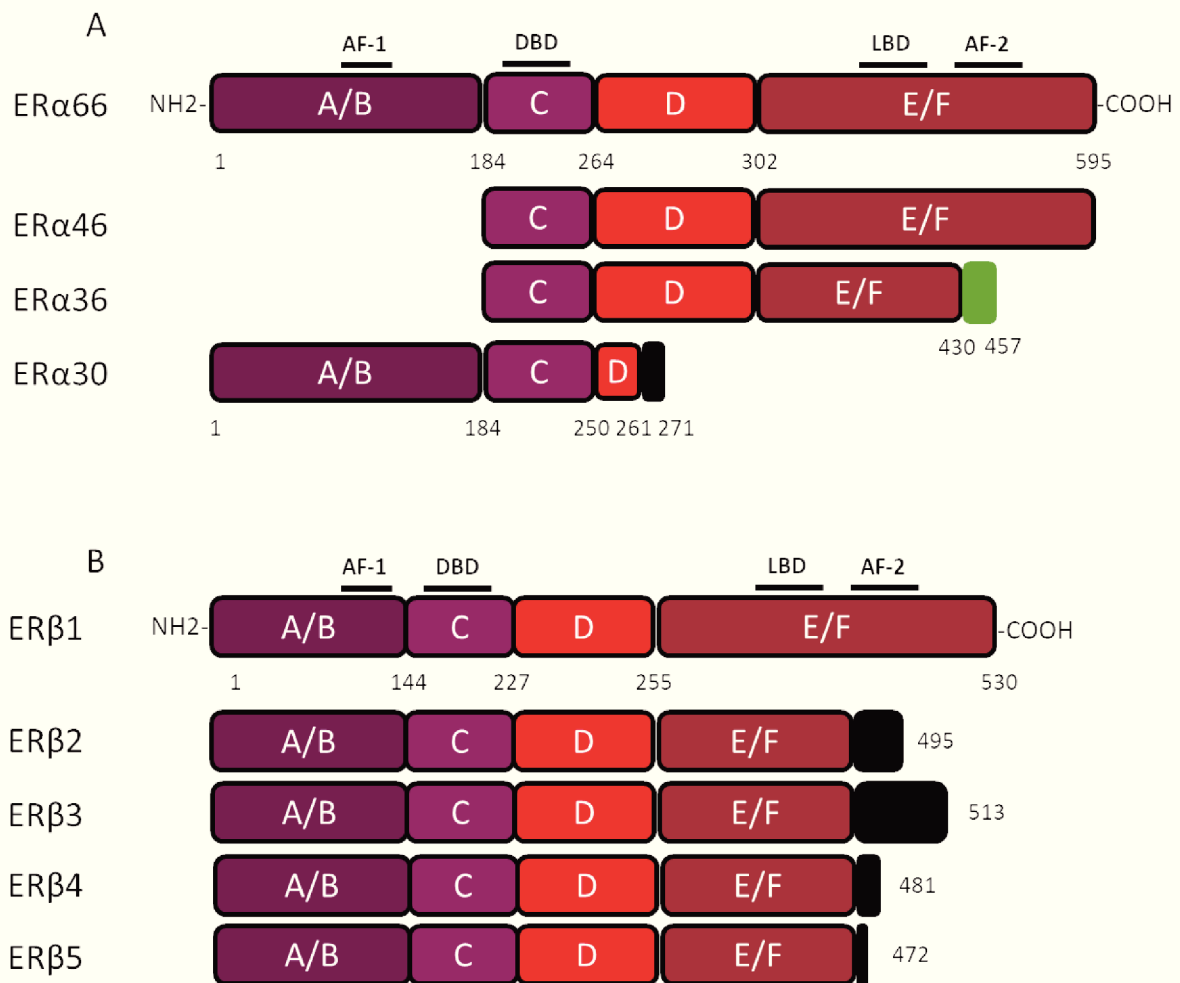
Le domaine D constitue une région charnière flexible qui permet la rotation du domaine de liaison à l'ADN afin de faciliter la fixation du récepteur sur son élément de réponse. Dans cette région qui est peu conservée, on trouve la séquence de localisation nucléaire qui permet la translocation du récepteur dans le noyau. Cette séquence est exposée suite au changement de conformation du domaine D induit par la fixation du ligand (Kumar et al., 2011).

✓ **Le domaine E/F de dimérisation, de liaison au ligand et de transactivation dépendante du ligand**

Le domaine E/F également appelé LBD (Ligand Binding Domain) correspond au domaine de liaison au ligand et participe à la dimérisation du récepteur. Pour la plupart des récepteurs nucléaires ce domaine E/F se compose de 12 hélices alpha et d'un feuillet bêta antiparallèle qui forment la poche de liaison pour les ligands hydrophobes (Figure 8C). Cette région contient également le domaine de transactivation AF-2 qui est constitué des hélices H3, H4 et H12 et qui permet d'activer la transcription de façon ligand-dépendante. En effet, en fonction de la nature du ligand fixé (agoniste ou antagoniste), l'orientation de l'hélice H12 sera différente et cela aura un impact sur le recrutement des cofacteurs nécessaires à la régulation de l'activité transcriptionnelles des ER $\alpha$  et  $\beta$  (Bourguet *et al.*, 2000 ; Heldring et al., 2007).

➤ Les formes canoniques : ER $\alpha$ 66 et ER $\beta$ 1

La forme canonique des ER $\alpha$ , appelée ER $\alpha$ 66, correspond à une protéine de 595 acides aminés ayant un poids moléculaire de 66 kDa. Concernant les ER $\beta$ , la forme canonique ER $\beta$ 1 est une protéine de 530 acides aminés ayant un poids moléculaire de 59,2 kDa. Ces 2 récepteurs qui résultent de la duplication d'un gène ancestral commun possèdent l'ensemble des domaines



**Figure 9 : Structure des formes canoniques et des différents variants d'ERα66 et d'ERβ1.**

Les formes canoniques ERα66 et ERβ1 se composent respectivement de 595 et 530 acides aminés.

(A) Les variants ERα46 et ERα36, ne possèdent pas le domaine A/B qui est impliqué dans la transactivation indépendante du ligand. De plus, le variant ERα36, possède un domaine E/F tronqué et une petite séquence spécifique, de 27 acides aminés (boîte verte), à l'extrémité C-terminale. Le variant ERα30 a conservé uniquement le domaine A/B, le domaine C et une partie d'un domaine D. En C-terminale ce variant possède une courte séquence spécifique de 10 acides aminés (boîte noire).

(B) Les 4 variants d'ERβ1, présentent tous un domaine E/F tronqué et une petite séquence spécifique de taille variable à l'extrémité C-terminale (boîte noire). De manière caractéristique d'ERβ2, les 26 acides aminés spécifiques situés à l'extrémité C-terminale sont codés par un exon appelé « exon 9 » ou « exon cx » localisé en aval de l'exon 8 du gène ESR2.

(D'après Le Romancer et al., 2011 ; Jia et al., 2015)

fonctionnels caractéristiques des récepteurs nucléaires (Thornton, 2001). En revanche, le degré d'homologie des différents domaines d'ER $\alpha$ 66 et d'ER $\beta$ 1 est variable.

Les domaines les plus conservés entre ces 2 récepteurs sont le domaine C de liaison à l'ADN et le domaine E/F de liaison au ligand qui présentent un pourcentage de similarité de séquence de 95 % et de 55 %, respectivement. Cette homologie est de 17% pour le domaine A/B et de 36% pour le domaine D.

Ces différences structurales ont un impact sur le fonctionnement de ces 2 récepteurs. Tout d'abord, l'activité transactivatrice du domaine AF-1 d'ER $\alpha$ 66 est significativement plus importante que celle d'ER $\beta$ 1. Concernant la poche de liaison au ligand d'ER $\beta$ 1, elle est plus petite que celle d'ER $\alpha$ 66 ce qui a un impact sur la spécificité et l'affinité de liaison des ligands potentiels. Enfin, au niveau du domaine D, ER $\beta$ 1 ne possède qu'une seule séquence de localisation nucléaire alors qu'on en retrouve plusieurs chez ER $\alpha$ 66.

#### ➤ Les variants d'ER $\alpha$ 66

Il existe 3 variants d'ER $\alpha$ 66 à savoir, ER $\alpha$ 46, ER $\alpha$ 36 et ER $\alpha$ 30, qui sont issus des mécanismes d'épissage alternatif mais également de l'utilisation de promoteurs alternatifs (Figure 9A).

A la différence de la forme canonique ER $\alpha$ 66 qui est codée à partir des exons 1 à 8, le variant ER $\alpha$ 46 est issu de l'utilisation des promoteurs E et F et il est codé à partir des exons 2 à 8. Au niveau protéique, ce variant de 46 kDa ne possède pas le domaine de transactivation AF-1 (Flouriot et al., 2000).

Concernant le variant ER $\alpha$ 36, il est produit à partir d'un promoteur alternatif localisé dans le premier intron du gène ESR1 et il est codé par les exons 1', 2 à 6, et 9. D'un point de vue fonctionnel, ER $\alpha$ 36, qui a conservé le domaine de liaison à l'ADN ainsi qu'une partie du domaine de liaison au ligand ne possède pas les domaines de transactivation AF-1 et AF-2. En revanche, à l'extrémité C-terminale, ER $\alpha$ 36 possède une région spécifique de 27 acides aminés codée par l'exon 9 (Wang Z. et al., 2005).

Plus récemment, un 3<sup>ème</sup> variant de 30 kDa, appelé ER $\alpha$ 30, a été identifié par Zhu H. et al., (2018). Ce variant est codé par les exons 1, 2, 3 et 8 mais également par une partie des exons 4 et 6. En effet, lors de l'épissage alternatif, les 24 premiers nucléotides de l'exon 4 sont directement liés au 44 derniers nucléotides de l'exon 6. Au niveau protéique, cela se traduit par un domaine charnière D tronqué ainsi que par l'absence du domaine de liaison au ligand et du domaine de transactivation AF-2. En revanche, ER $\alpha$ 30 possède en C-terminale une petite région

spécifique de 10 acides aminés dont la fonction et l'impact sur la structure de la protéine n'est pas décrite.

➤ Les variants d'ERβ1

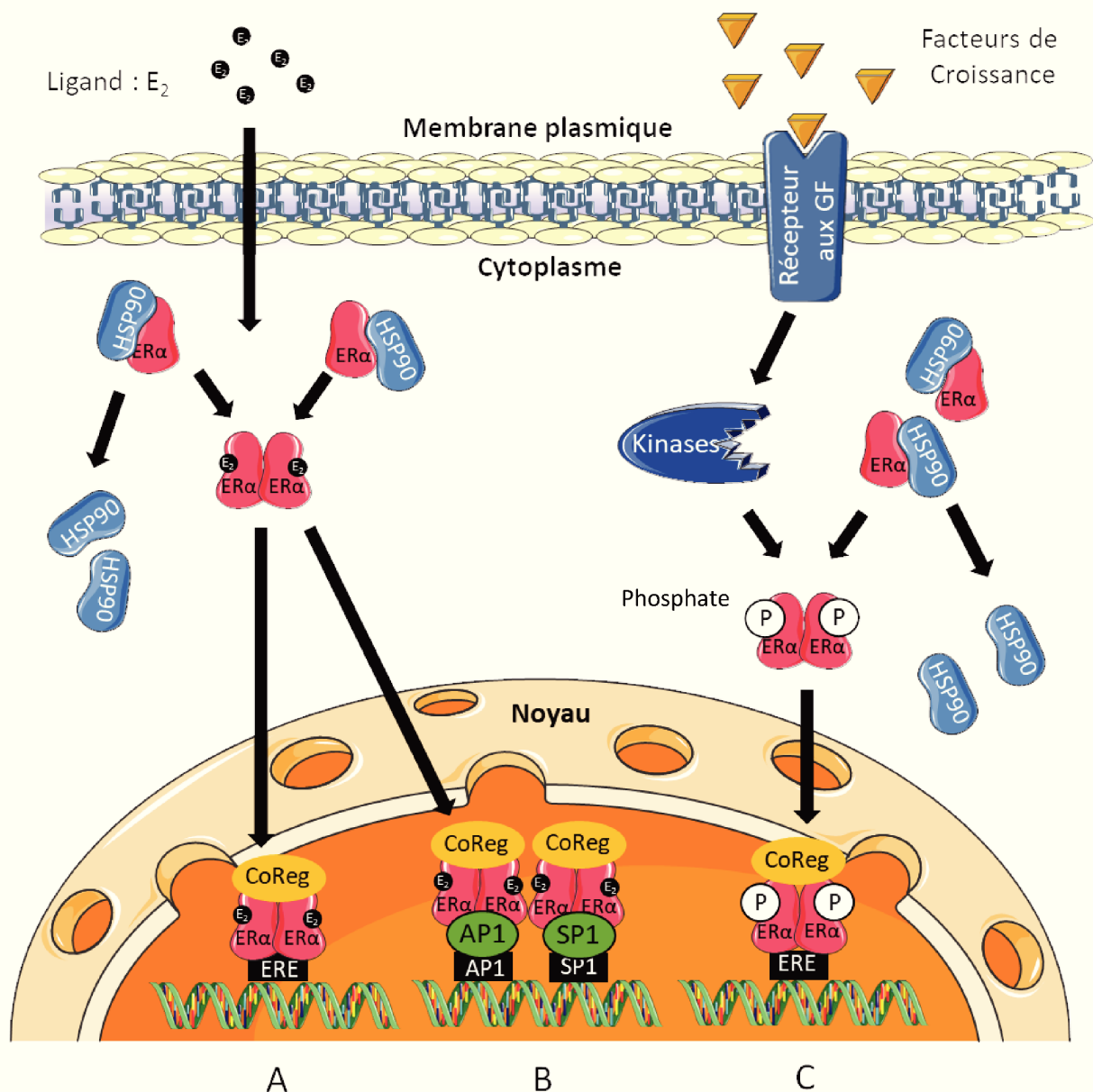
La forme canonique ERβ1 qui est codée par les exons 1 à 8 partage avec l'ensemble de ses variants la région correspondant aux exons 1 à 7. Les différents variants sont issus de l'épissage alternatif de l'exon 8 qui dans le cas du variant ERβ2 est remplacé par l'exon 9, également appelé exon cx, identifié en 1998 par Ogawa et al.

Au niveau protéique, les différents variants d'ERβ1 se composent d'un domaine E/F de longueur et de séquence variable (Figure 9B). Les isoformes ERβ2, 3, 4 et 5 présentent un domaine de liaison au ligand tronqué ainsi qu'une absence du domaine de transactivation AF-2 (Haldosén et al., 2014). Cela implique qu'ils ne sont pas capables d'induire à eux seuls une transcription ligand-dépendante. Toutefois, ces variants peuvent s'hétérodimériser avec la forme canonique ERβ1 qui possède un domaine de liaison au ligand complet ainsi que le domaine de transactivation AF-2. Cette association conduit à une potentialisation de l'activité transcriptionnelle d'ERβ1 (Zhao et al., 2008). Une hétérodimérisation est également possible avec ERα66 mais cette association a plutôt un effet répresseur sur la transcription des gènes cibles (Zhao et al., 2007).

*iii) La signalisation génomique des œstrogènes*

*Les récepteurs nucléaires aux œstrogènes constituent les principaux médiateurs de la signalisation œstrogénique. Plus précisément, ces récepteurs, qui sont des facteurs de transcription capables de réguler directement l'expression de leurs gènes cibles, sont impliqués dans la signalisation dite génomique des œstrogènes. On distingue 3 types de voies génomiques qui sont toutes impliquées dans la régulation de la plupart des processus cellulaires (prolifération, différenciation, migration et mort cellulaire): (i) la voie classique correspond à la fixation des ERs au niveau de leurs éléments de réponse spécifiques appelés ERE, (ii) la voie ERE-indépendante dépend de l'interaction des ERs avec d'autres facteurs de transcription et (iii) la voie indépendante des œstrogènes correspond à une activation des ERs induite par la signalisation dépendante des facteurs de croissance (Yaşar et al., 2017).*

Dans cette partie nous nous intéresserons uniquement à la signalisation génomique des œstrogènes qui a été décrite dans des conditions physiologiques. Les mécanismes de régulation



**Figure 10 : La signalisation génomique des œstrogènes.**

La signalisation génomique des œstrogènes comprend 3 voies différentes :

- (A) La voie classique, ERE-dépendante
- (B) La voie ERE-indépendante
- (C) La voie ligand-indépendante

Avant dimérisation, les ER $\alpha$ 66 monomériques sont maintenus sous forme inactive dans le cytoplasme par des protéines chaperonnes comme HSP90. En présence du ligand ( $E_2$ ), les monomères se dimérisent et sont importés au niveau nucléaire. Dans le noyau, ils peuvent interagir (A) directement avec l'ADN, au niveau de leurs éléments de réponse spécifiques (ERE), ou (B) avec d'autres facteurs de transcription comme AP-1 et SP-1, qui sont déjà liés au niveau de leurs éléments de réponse respectifs.

(C) L'activation d'ER $\alpha$ 66 peut également être déclenchée en réponse à des facteurs de croissance qui activent des kinases intracellulaires responsables de la phosphorylation d'ER $\alpha$ 66.

**AP1** = Activator Protein 1, **CoReg** = Corégulateurs, **ERE** = Estrogen Response Element (Élément de réponse aux œstrogènes), **GF** = Growth Factor (Facteurs de croissance), **HSP** = Heat Shock Protein (Protéine de choc thermique), **SP1** = Specificity protein 1.

(D'après Le Romancer et al., 2011)

de cette signalisation qui ont été mis en évidence dans le contexte du cancer du sein seront abordés dans le Chapitre 3.

➤ Mécanismes de maintien des ERs dans un état inactif

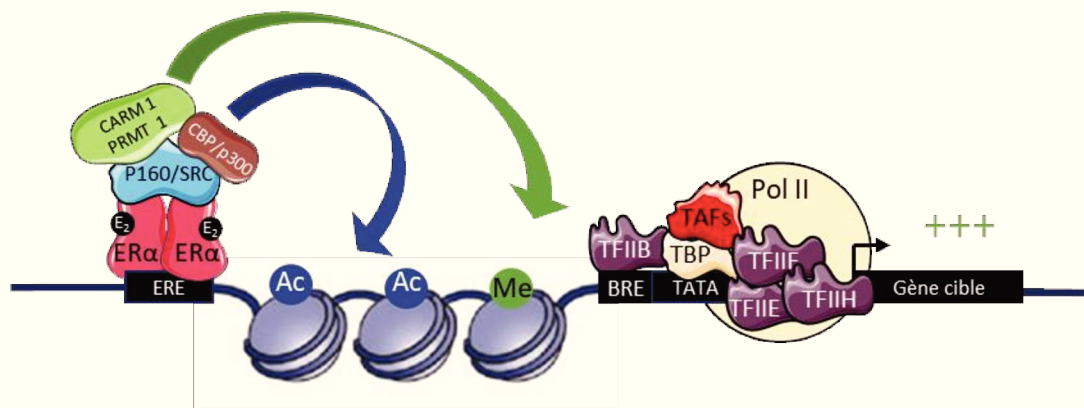
En absence de ligand les récepteurs aux hormones stéroïdiennes sont séquestrés dans le cytoplasme par des protéines chaperonnes, comme les protéines de choc thermique (HSP). Ces protéines très conservées jouent un rôle majeur en régulant notamment le repliement, le transport, la dégradation des protéines ainsi que l'assemblage des complexes multi-protéiques. Plus particulièrement, ER $\alpha$  est maintenu dans le cytoplasme sous forme monomérique et dans une conformation favorable à la liaison du ligand par les protéines chaperonnes HSP40, HSP70 et HSP90. Dans cet état, la demi-vie de ce récepteur est d'environ 4 à 5 heures avant d'être dégradé par le système ubiquitine-protéasome (Pratt et Toft, 2003 ; Reid et al., 2003 ; Dhamad et al., 2016).

➤ La voie génomique classique

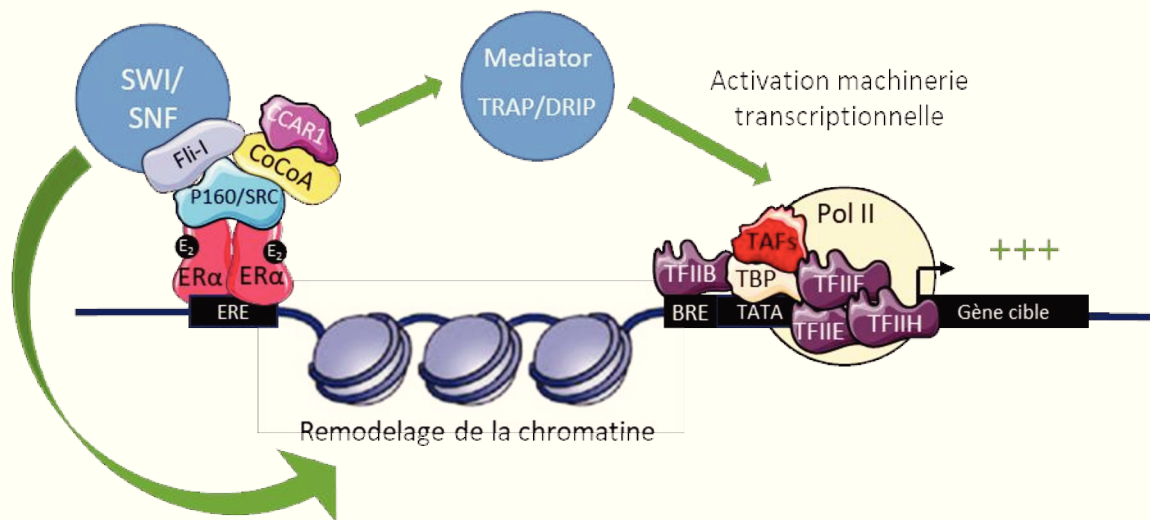
La fixation du ligand au niveau du LBD induit un changement de conformation du récepteur qui aboutit à la dissociation des protéines chaperonnes et à sa dimérisation. Cela conduit également à l'exposition de la séquence de localisation nucléaire localisée au niveau du domaine charnière D permettant ainsi la translocation nucléaire du récepteur. Dans le noyau, les ERs sous forme dimérisée sont capables d'interagir avec les EREs que l'on peut retrouver au niveau d'éléments enhancers localisés à distance des gènes cibles ou directement dans la région promotrice proximale de ces gènes (Figure 10) (Carroll et al., 2005). Ces ERE correspondent à une séquence palindromique de 15 paires de bases comprenant 3 nucléotides centraux non spécifiques entourés par 2 demi-sites hexanucléotidiques (5'-AGGTCAnnnTGACCT-3') (Driscoll et al., 1998). Toutefois, cette séquence consensus n'est pas retrouvée dans la région promotrice de tous les gènes cibles des œstrogènes. En effet, elle est fréquemment associée ou remplacée (i) par des ERE imparfaits non palindromiques, dont la séquence diffère de la séquence consensus au niveau d'un ou plusieurs nucléotides ou (ii) par des demi-ERE pour lesquels la séquence d'un des demi-sites hexanucléotidiques est variable (Augereau et al., 1994 ; Kim *et al.*, 2000).

L'affinité de liaison des ERs à l'ADN ainsi que l'activité transcriptionnelle qui en découle sont directement impactées par la séquence de l'ERE. En effet, l'affinité des ERs pour les EREs imparfaits et demi-ERE est moins forte que pour la séquence consensus ce qui conduit à une

A



B



### Figure 11 : Mécanismes d'action des coactivateurs d'ER $\alpha$ 66.

En présence d'œstrogènes, les dimères d'ER $\alpha$ 66 fixés au niveau des EREs sont capables de recruter différents coactivateurs afin d'activer la transcription des gènes cibles.

L'activation transcriptionnelle dépendante des coactivateurs de la famille p160/SRC résulte :

- (A) Du recrutement d'histones méthyltransférases (CARM1/PRMT1) et d'histones acétyltransférases (CBP/p300)
- (B) Du recrutement du complexe de remodelage de la chromatine SWI/SNF et du complexe Mediator (TRAP/DRIP) via les facteurs Fli-I et CoCoA/CCAR1.

**Ac** = Groupement Acétyle, **CARM1/PRMT1** = Arginines Methyltransférases, **CBP/p300** = Protéine de liaison au facteur CREB, **CCAR1** = Cell division Cycle and Apoptosis Regulator 1, **CoCoA** = Coiled-coil coactivator, **BRE**= Elément de reconnaissance du facteur TFIIB, **ERE** = Elément de réponse aux œstrogènes, **Fli-I** = Flightless-I, **Me** = Groupement Méthyle, **Pol II** = ARN Polymérase II, **TBP** = Protéine de liaison à la boîte TATA, **TAF** = Facteurs associés à la TBP, **TFII(B,E,E,H)** = Facteurs généraux de la transcription.

plus faible modulation de l'activité transcriptionnelle du gène cible (Klinge, 2001). De plus, le nombre d'ERE présents dans la région promotrice d'un même gène cible est un paramètre crucial car la présence d'ERE adjacents à un effet synergique sur l'activité transcriptionnelle induite par ER $\alpha$  en réponse à l'E<sub>2</sub> (Klein-Hitpass et al., 1988 ; Yi et al., 2002).

Lorsqu'ils sont fixés sur un ERE, les ERs sont capables de recruter différents corégulateurs transcriptionnels (coactivateurs et corépresseurs) qui, pour la majorité, possèdent au moins un motif NR-box ayant pour séquence consensus : LxxLL (Leucine, 2 acides aminés, Leucine, Leucine) (Heery et al., 1997). Ce motif est reconnu spécifiquement par une interface constituée d'acides aminés provenant des hélices H3, H4, H5 et H12 du LBD du récepteur mais cette interaction est influencée par la nature du ligand (agoniste ou antagoniste). En effet, la liaison d'un agoniste aura pour effet de stabiliser l'hélice H12 et de favoriser le recrutement de cofacteurs tandis que la fixation d'un antagoniste déstabilise le positionnement des hélices H11 et H12 ce qui empêche le recrutement de ces cofacteurs (Bourguet et al., 2000 ; Savkur et Burris, 2004).

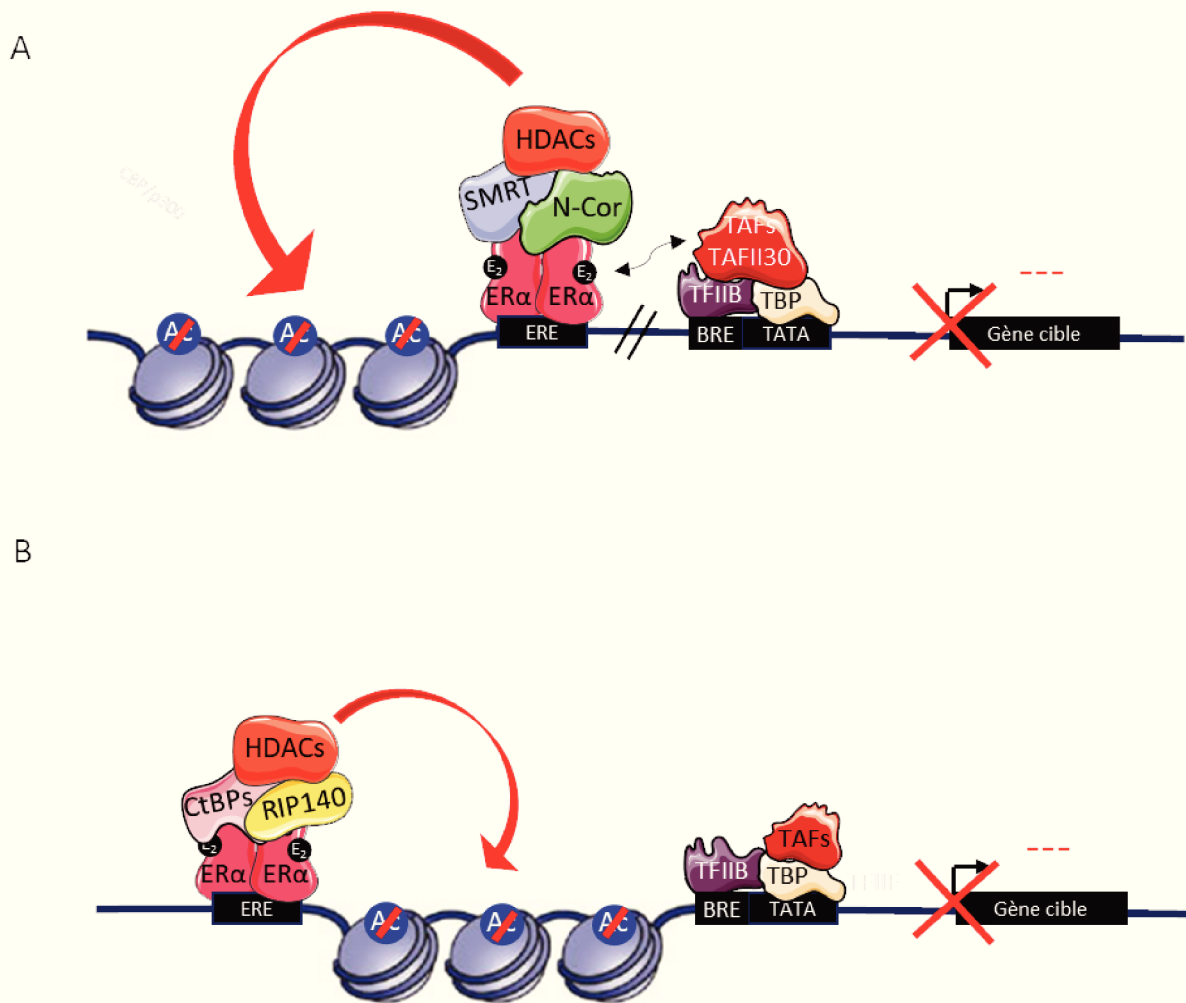
#### ✓ **Les coactivateurs transcriptionnels**

L'activation de la transcription des gènes cibles des œstrogènes dépend d'une interaction directe ou indirecte entre les ERs fixés sur leurs EREs et le complexe de pré-initiation (PIC) de la transcription qui comprend l'ARN polymérase II et 6 facteurs généraux de la transcription (TFIIA, TFIIB, TFIID, TFIIE, TFIIF et TFIIH) (Figure 11).

L'assemblage du PIC se déroule de façon séquentielle. Cela débute par le recrutement du facteur TFIID qui est constitué de la protéine TBP (TATA binding protein) et d'un nombre variable de sous unités appelées TAF (TBP-Associated Factor) qui sont spécifiques de certains promoteurs. La fixation de TFIID, au niveau des séquences cibles (TATA-Box) localisées en amont du site d'initiation de la transcription, est stabilisée par le facteur TFIIA et permet le recrutement du facteur TFIIB au niveau des séquences BRE (TFIIB Recognition Element). Ce facteur TFIIB permet de définir précisément le site d'initiation et l'orientation de la transcription. La dernière étape de l'assemblage du PIC correspond au recrutement de l'ARN polymérase II et des facteurs TFIIE, TFIIF et TFIIH qui favorisent l'élongation de la transcription (Lee et Young, 2000 ; Coin et Egly, 2000 ; Gangloff *et al.*, 2001).

Plusieurs études ont démontré l'existence d'interactions directes entre les ERs et certains facteurs constituant le PIC. Par exemple, les ERs sont capables d'interagir avec la protéine TBP, par l'intermédiaire des domaines de transactivation AF-1/AF-2 et avec le facteur TFIIB via leur





### Figure 12 : Mécanismes d'action des corépresseurs d'ERα66.

En présence d'œstrogènes, les dimères d'ERα66 fixés au niveau des EREs sont capables de recruter différents corépresseurs afin d'inhiber la transcription des gènes cibles.

Le recrutement des histones désacétylases (HDACs) conduit à une condensation de la chromatine qui empêche le recrutement de l'ARN polymérase II et donc réprime l'expression des gènes cibles. Ce recrutement des HDACs dépend :

(A) De l'interaction du dimère d'ERα66 avec le facteur TAFII30 qui conduit au recrutement des corépresseurs N-Cor et SMRT.

(B) Du recrutement de RIP140 et des CtBPs par ERα66.

**Ac** = Groupement Acétyle, **BRE** = Elément de reconnaissance du facteur TFIIB, **CtBP** = C-terminal-binding protein 1, **ERE** = Estrogen Response Element (Élément de réponse aux œstrogènes), **HDAC** = Histone désacétylase, **N-Cor/SMRT** = Nuclear Receptor Co-repressor, **RIP140** = Receptor Interacting Protein, **TBP** = Protéine de liaison à la boîte TATA, **TAF** = Facteurs associés à la TBP, **TFIIB** = Facteurs généraux de la transcription.

région C-terminale qui contient le domaine de liaison à l'ADN. Ces interactions qui peuvent avoir lieu lorsque les ERs sont liés à leurs EREs conduisent à une modification de la structure du complexe TBP/TFIID-TFIIB-ADN qui permet de faciliter le recrutement des facteurs TFIIF et TFIIH. Ce mécanisme est directement corrélé à l'augmentation de l'activité transcriptionnelle induite par les ERs (Sabbah et al., 1998)

En parallèle, les ERs peuvent recruter successivement différents complexes coactivateurs qui permettent la régulation indirecte de la machinerie transcriptionnelle. Dans un premier temps, les ERs liés aux ERE peuvent recruter les coactivateurs de la famille p160/Src (p160/Steroid receptor coactivator) qui comprend 3 membres : Src-1, Src-2 et Src-3. Chacun de ces facteurs possède 3 motifs LxxLL qui permettent l'interaction avec le domaine AF-2 des ERs, ainsi que 2 domaines de transactivation AD-1 et AD-2 qui permettent de recruter respectivement des histones acétyltransférases (CBP/p300) et des histones méthyltransférases (PRMT1 et CARM1). En modifiant certains acides aminés de la queue des histones, ces enzymes participent au remodelage de la chromatine et augmentent l'accessibilité du promoteur pour la machinerie transcriptionnelle.

Les facteurs SRC sont également capables de recruter via leur domaine bHLH-PAS, localisé dans la partie N-terminale, d'autres corégulateurs comme Fli-I et CoCoA. Fli-I est nécessaire pour le recrutement du complexe SWI/SNF qui participe également au remodelage de la chromatine par l'intermédiaire de sa sous-unité ATPase, tandis que le facteur CoCoA associé à la protéine CCAR1 permet le recrutement du complexe Mediator qui conduit à l'activation de la machinerie transcriptionnelle (Rakesh, 2008 ; York et O'Malley, 2010).

#### ✓ **Les corépresseurs transcriptionnels**

Les ERs peuvent également réguler négativement l'expression de leurs gènes cibles en recrutant des corépresseurs transcriptionnels (Figure 12). Il a par exemple été montré que les ERs peuvent interagir via leur domaine de transactivation AF-2 avec une des protéines TAF constituant le facteur TFIID appelée TAFII30. Cette interaction permet l'association des ERs avec les corépresseurs SMRT (Silencing mediator of retinoid and thyroid receptors) et N-Cor (Nuclear receptor corepressor) qui sont responsables du recrutement d'histones désacétylases (HDAC). Ces enzymes conduisent à une condensation de la chromatine qui est associée à une diminution de l'expression génique (Jacq et al., 1994 ; Hao et al., 2007).

Un autre corépresseur appelé RIP140 a également été identifié. Au niveau structural, ce facteur possède 9 motifs LxxLL permettant l'interaction avec les ERs et 4 domaines répresseurs notés

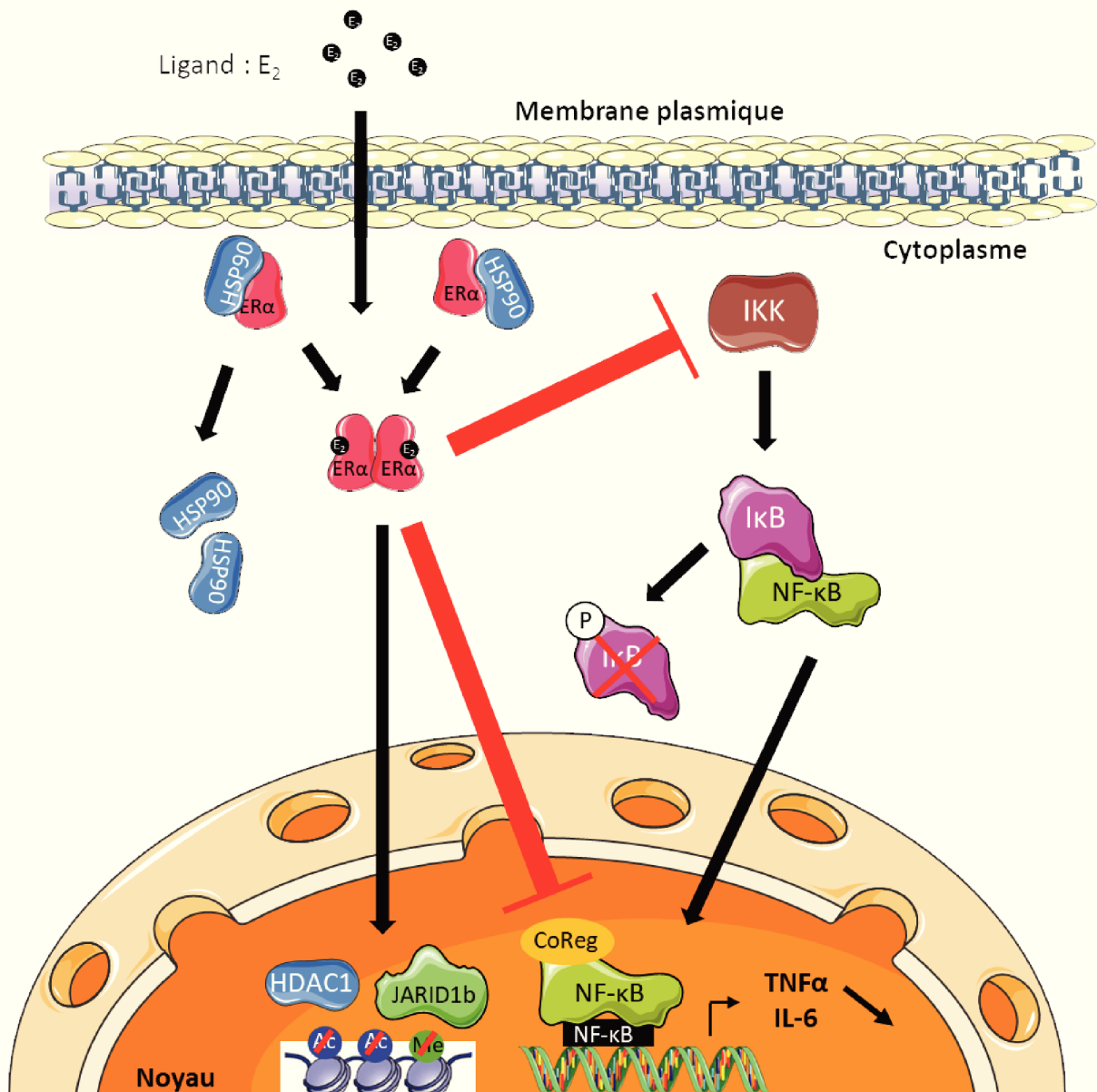
RD1 à RD4. Le domaine RD1 localisé dans la partie N-terminale permet le recrutement des HDAC et le domaine RD2 contient 2 motifs conservés (PIDLS et PINLS) qui constituent les sites de fixation des protéines CtBP1 et CtBP2, également impliquées dans le recrutement des HDAC. A l'heure actuelle, les facteurs associés à l'effet répresseur des domaines RD3 et RD4 de RIP140 n'ont pas été identifiés (Augereau et al., 1994 ; Castet et al., 2004 ; Castet et al., 2005).

➤ La voie génomique ERE-indépendante

En parallèle de la voie classique de réponse aux œstrogènes décrite précédemment, les ERs sont capables de réguler l'expression de gènes cibles sans interagir avec les EREs (Figure 10). Cela permet notamment la régulation œstrogène-dépendante de l'expression de gènes qui ne possèdent pas d'ERE dans leur région promotrice. Dans cette voie dite ERE-indépendante, l'interaction du ligand avec les ERs conduit comme dans la voie classique à la dimérisation et à la translocation nucléaire du récepteur. En revanche dans le noyau, les ERs dimérisés s'associent avec d'autres facteurs de transcription qui reconnaissent leurs propres éléments de réponse sur l'ADN.

Le principal facteur de transcription impliqué dans cette voie est le facteur SP1 qui reconnaît des séquences promotrices riches en GC. L'interaction ER $\alpha$ 66/SP1 a été démontrée pour la première fois par l'étude de la région promotrice du gène de la cathepsine D qui ne possède pas d'ERE mais dont l'activité est induite par une exposition à l'E<sub>2</sub> (Krishnan et al., 1995). Parmi les autres cibles du complexe ER $\alpha$ 66/Sp1 on trouve différents gènes impliqués notamment dans le contrôle de la prolifération cellulaire comme les gènes codant EGFR, c-fos ou la cycline D1 (O'Lone et al., 2004).

Les ERs sont également capables de moduler l'activité du complexe de facteur de transcription AP-1 qui est composé des proto-oncogènes c-Fos et c-Jun. Cette interaction est dépendante des coactivateurs de la famille p160 décrits précédemment qui sont recrutés par les domaines de transactivation AF-1 et AF-2 des ERs. Parmi les gènes cibles du complexe AP-1 dont l'expression est induite par une exposition à l'E<sub>2</sub> on trouve des gènes codant des protéines impliquées dans la prolifération, la différenciation et la motilité cellulaire comme la fibronectine 1, la tropomyosine 1 ou encore le proto-oncogène c-Myc (Webb et al., 1999 ; DeNardo et al., 2005 ; DeNardo et al., 2007 ; Safe et Kim, 2008).



**Figure 13 : La voie ERE-indépendante via le facteur de transcription NF- $\kappa$ B.**

La signalisation correspondant à cette figure est détaillée dans le paragraphe concernant la voie de signalisation ERE-indépendante.

**Ac** = Groupement Acétyle, **CoReg** = Corégulateurs, **HDAC**= Histone désacétylase, **JARID1b** = Histone demethylase, **IKK** =  $I\kappa$ -B kinase, **IL-6** = Interleukine 6, **Me** = Groupement Méthyle, **TNF $\alpha$**  = Tumor Necrosis Factor.

Pour finir, les ERs peuvent également réprimer l'expression de gènes régulés par le facteur de transcription NF- $\kappa$ B via plusieurs mécanismes (Figure 13). Classiquement, NF- $\kappa$ B qui se compose de 2 sous-unités variables en fonction du contexte cellulaire (RelA(p65), RelB ou c-Rel) est séquestré dans le cytoplasme par le facteur I $\kappa$ B. La phosphorylation d'I $\kappa$ B par le complexe IKK (I $\kappa$ B Kinase) permet la libération et la translocation nucléaire de NF- $\kappa$ B qui peut alors se lier à des éléments de réponse spécifiques au niveau de ses gènes cibles et recruter différents coactivateurs. Au niveau cytoplasmique, les ERs peuvent agir en inhibant l'activité du complexe IKK et ainsi empêcher la dégradation d'I $\kappa$ B. Au niveau nucléaire, les ERs peuvent empêcher le recrutement de certains coactivateurs comme la protéine CBP (cAMP response element-binding protein (CREB)-binding protein). Ils peuvent également permettre le recrutement d'histones déméthylases (JARID1b) ou d'histones désacétylases (HDAC1) qui conduisent à une condensation de la chromatine. Ces mécanismes ont surtout un impact sur la réponse inflammatoire en diminuant l'expression du gène codant le TNF $\alpha$ , mais aussi sur le système squelettique où les œstrogènes contribuent à limiter le mécanisme de résorption osseuse en diminuant l'expression de l'IL-6 induite par NF- $\kappa$ B (Kalaitzidis et Gilmore, 2005 ; Cvorovic et al., 2006 ; Nettles et al., 2008 ; Kovacs, 2015).

➤ La voie génomique ligand indépendante

Grâce au domaine de transactivation ligand-indépendant AF-1, les ERs sont capables de moduler l'expression de nombreux gènes cibles en absence d'œstrogènes (Figure 10). Cette voie œstrogène-indépendante est déclenchée en réponse à des facteurs de croissance, comme l'EGF et l'IGF, qui activent des kinases intracellulaires responsables de la phosphorylation du domaine AF-1 des ERs non liés aux œstrogènes (apo-ER). Par exemple dans le cas de la signalisation induite par l'EGF, les apo-ERs sont phosphorylés au niveau des Sérine S118 et S106 par la voie des MAPK et au niveau de la sérine S167 par Akt (également appelée PKB) ou m-TOR. Ces modifications post-traductionnelles activent les apo-ERs qui peuvent alors se fixer au niveau de certains EREs et recruter des corégulateurs transcriptionnels comme par exemple SRC-1 et CBP/p300 via la S118. De façon intéressante, les gènes régulés par la voie ligand-indépendante sont majoritairement impliqués dans le contrôle du développement et de la différenciation cellulaire (Caizzi et al., 2014 ; Bennesch et Picard, 2015).

## B) Les récepteurs aux œstrogènes membranaires

Comme nous venons de le voir, les récepteurs aux œstrogènes sont des récepteurs nucléaires responsables de l'activation de voies de signalisation génomiques qui conduisent à la régulation de l'expression de gènes cibles. Cependant, de nombreuses études ont montré l'existence d'effets dépendants des œstrogènes dont la survenue est trop rapide pour être médiée par la régulation transcriptionnelle des gènes. Ces effets dépendent de l'activation de voies de signalisation non génomiques qui sont initiées à partir de récepteurs aux œstrogènes localisés au niveau membranaire.

### *i) Localisation membranaire de la forme canonique ER $\alpha$ 66*

Différentes études portant sur la localisation cellulaire d'ER $\alpha$ 66 indiquent qu'une partie du pool de ce récepteur est localisée à proximité de la membrane plasmique malgré l'absence du peptide signal nécessaire à la localisation membranaire des protéines (Norfleet et al., 1999 ; Clarke et al., 2000). Cette association entre ER $\alpha$ 66 et la membrane plasmique dépend de la modification post-traductionnelle de ce récepteur et de l'interaction avec des protéines adaptatrices. ER $\alpha$ 66 peut en effet être palmitoylé par l'enzyme PAT (Palmitoyl AcylTransferase) au niveau de la Cystéine 447. Cela conduit à une augmentation de son hydrophobicité qui facilite l'interaction avec la membrane plasmique (Acconcia et al., 2004). De plus, la palmitoylation d'ER $\alpha$ 66 permet son interaction avec la cavéoline-1 et la striatine qui sont des protéines adaptatrices localisées au niveau des cavéoles (Lu et al., 2004). D'autres études indiquent que la sérine 522 du domaine E jouerait un rôle majeur pour la localisation membranaire d'ER $\alpha$ 66 probablement en permettant l'interaction entre ce récepteur et la cavéoline-1 (Razandi et al., 2004).

Bien que peu de données soient disponibles dans la littérature concernant les voies de signalisation non génomiques dépendantes d'ER $\beta$ , il a été démontré que ce récepteur peut se localiser au niveau des cavéoles et activer l'enzyme eNOS (endothelial Nitric Oxide Synthase) responsable de la production d'oxyde nitrique en réponse à l'E<sub>2</sub> dans les cellules endothéliales. (Chambliss et al., 2002). Cela suggère que ce récepteur est également impliqué dans le déclenchement des voies de signalisation non génomiques en réponse aux œstrogènes.

### *ii) Les variants ER $\alpha$ 36 et ER $\alpha$ 46*

Comme nous l'avons vu précédemment, le variant ER $\alpha$ 36 qui a été mis en évidence par Wang Z. et al., en 2005 ne possède pas les domaines de transactivation AF-1 et AF-2 qui sont

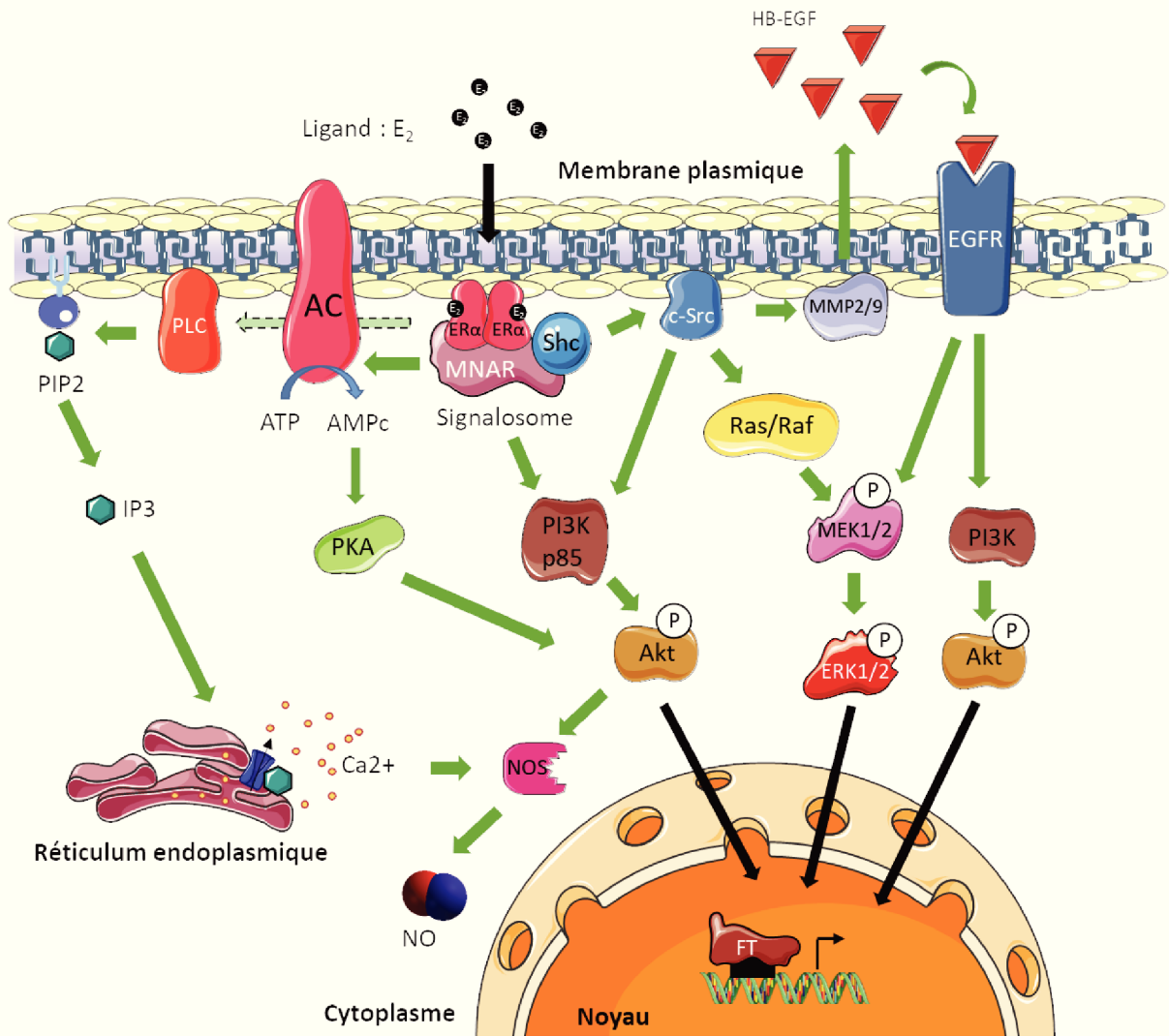
indispensables pour l'activation de la signalisation génomique des œstrogènes. En revanche, ce variant possède comme ER $\alpha$ 66 deux séquences de localisation nucléaire et un domaine de liaison à l'ADN qui lui permettent d'agir comme un dominant négatif sur l'activité transactivatrice d'ER $\alpha$ 66 et d'ER $\beta$ 1. En effet, ER $\alpha$ 36 est capable de se lier de façon compétitive sur les ERE ou de former un hétérodimère avec ER $\alpha$ 66 ce qui conduit à une inhibition de l'activité du domaine AF-1 de la forme canonique (Zou et al., 2009).

De plus, ER $\alpha$ 36 possède au niveau des domaines C et E, 3 sites potentiels de myristoylation qui permettent l'ancrage de ce variant au niveau de la face interne de la membrane plasmique. Ainsi, 50% du pool d'ER $\alpha$ 36 est localisé au niveau membranaire, 40% au niveau cytoplasmique et seulement 10% au niveau nucléaire (Wang Z. et al., 2006). Cette localisation majoritairement membranaire permet l'intervention de ce variant dans les voies de signalisation oestrogéniques non génomiques.

Concernant le variant ER $\alpha$ 46, il peut, comme la forme canonique ER $\alpha$ 66, être palmitoylé par l'enzyme PAT et ainsi être recruté au niveau des cavéoles. Les effets de ce variant sur l'activation des voies de signalisation non génomiques sont décrits comme plus importants que ceux médiés par ER $\alpha$ 66 (Li L. et al., 2003).

### *iii) Le récepteur membranaire GPER*

En 1997, Carmeci et al., ont cloné à partir de la lignée de cellules cancéreuses mammaires MCF-7, un récepteur aux œstrogènes membranaire appartenant à la superfamille des récepteurs couplés aux protéines G. Ce récepteur appelé GPR30 ou GPER (G-protein coupled estrogen receptor) est codé par le gène GPER1 et correspond à une protéine de 375 acides aminés qui se compose de 7 domaines transmembranaires. Concernant sa localisation, GPER est principalement retrouvé au niveau de la membrane plasmique et de la membrane du réticulum endoplasmique. Toutefois, des études immunohistologiques récentes indiquent que ce récepteur peut également être retrouvé au niveau nucléaire dans les fibroblastes associés au cancer (CAFs). Cette translocation nucléaire, qui est dépendante des importines, serait possible grâce à la présence d'une séquence de localisation nucléaire localisée à l'extrémité C-terminale de GPER (Pupo et al., 2013). De plus, il a été démontré que l'un variant de GPER pour lequel la proline en position 16 est remplacée par une Leucine (P16L) serait préférentiellement localisé au niveau nucléaire. En effet, cette mutation est responsable d'un changement de conformation de la protéine qui conduit à un défaut de N-glycosylation et qui permet ainsi à GPER d'être transloqué dans le noyau (Qian et al., 2016 ; Pupo et al., 2017).



**Figure 14 : La signalisation non-génomique des œstrogènes.**

La signalisation correspondant à cette figure est détaillée dans le paragraphe concernant les voies de signalisation non génomiques.

FT = Facteurs de transcription variables en fonction des voies de signalisation activées  
 AC = Adénylate cyclase, AMPc = Adénosine monophosphate cyclique, HB-EGF : Heparin-binding EGF-like growth factor, IP3 = Inositol trisphosphate, MMP = Métalloprotéases Matricielles, MNAR = Modulator of Non-genomic Activity of estrogen Receptor (Modulateur de l'activité non-génomique des récepteurs aux œstrogènes), NO = Oxyde Nitrique, NOS = Oxyde Nitrique Synthase, PKA = Protéine Kinase A, PLC = Phospholipase C, PIP2 = Phosphatidylinositol-4,5-bisphosphate, Shc = Src homology 2-containing protein.

(D'après Le Romancer et al., 2011)

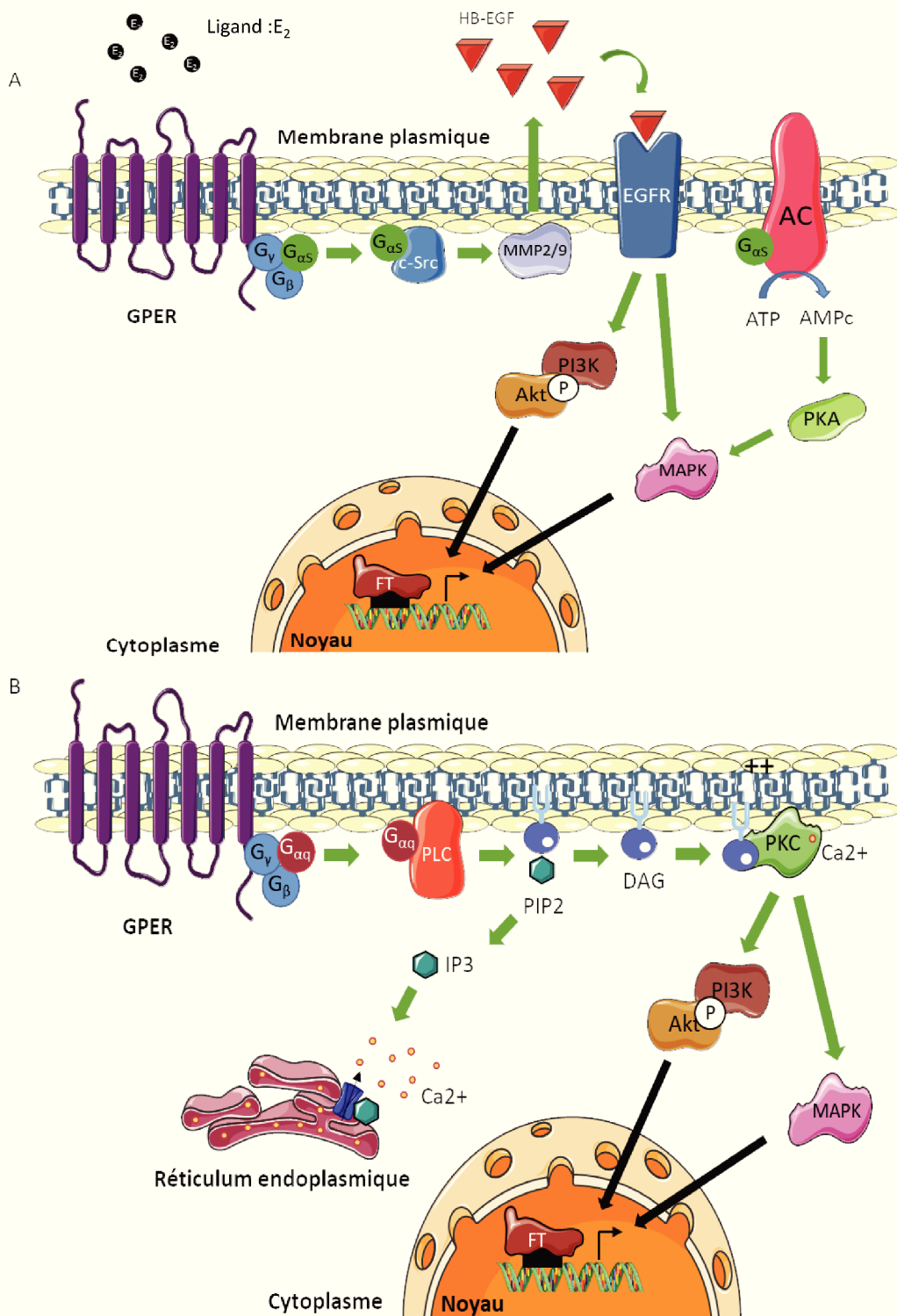


Initialement GPER était considéré comme un récepteur orphelin pour lequel aucun ligand n'avait été identifié. C'est seulement en 2005 que Thomas et al., ont démontré que les œstrogènes sont capables d'interagir avec le récepteur GPER, qui peut également lier avec une forte affinité des phytoestrogènes comme la Génistéine ainsi que des xœstrogènes comme le Bisphénol A (BPA). En effet, le BPA présente une affinité pour GPER qui est 8 à 50 fois plus importante que son affinité pour les récepteurs nucléaires aux œstrogènes (Thomas et Dong, 2006 ; Filardo et Thomas, 2012).

#### *iv) La signalisation non génomique des œstrogènes*

Les effets non-génomiques qui sont déclenchés quelques secondes à quelques minutes après l'exposition aux hormones stéroïdiennes ont été évoqués pour la première fois en 1942 par Hans Selye qui a observé qu'une injection intrapéritonéale de progestérone pouvait avoir un effet anesthésiant rapide chez le rat (Selye, 1942 ; Moriarty et al., 2006). Concernant les œstrogènes, la première démonstration de leurs effets non génomiques date de 1967 avec la détection d'une augmentation du taux d'AMPc qui est un second messager dans l'utérus d'une ratte ovariectomisée, 15 secondes après l'injection intraveineuse d'une concentration physiologique d'E<sub>2</sub> (Szego et Davis, 1967). Cela a également été confirmé en 1975 par Pietras et Szego qui ont mis en évidence in vitro le déclenchement d'un flux calcique 2,5 minutes après addition d'E<sub>2</sub> dans une culture de cellules endométriales murines obtenues à partir d'utérus de rattes ovariectomisées.

Les récepteurs nucléaires aux œstrogènes ne possédant pas d'activité kinase intrinsèque, l'activation des voies non-génomiques dépend de l'interaction de ces récepteurs avec plusieurs partenaires protéiques qui permettent d'activer différentes voies de signalisation intracellulaires (Figure 14). Dans ces complexes protéiques sous-membranaires appelés « signalosome » on retrouve en association avec les ERs des partenaires protéiques comme la protéine MNAR (Modulator of Non genomic activity of the ER), la protéine adaptatrice Shc ou encore la protéine de choc thermique Hsp90 (Moriarty et al., 2006). La protéine MNAR également appelée PELP1, est capable d'interagir avec les ERs via ses 10 motifs « LxxLL » et de recruter des protéines à domaine SH3 via ses motifs riches en proline « PxxP ». Parmi les protéines recrutées par MNAR, on peut citer la Tyrosine kinase c-Src ou la sous-unité régulatrice p85 de la PI3K (Balasenthil et Vadlamudi, 2003 ; Brann et al., 2008). Ces « signalosomes » sont responsables de l'activation de 2 voies de signalisation majoritaires, la voie PI3K/Akt et la voie des MAPK, mais aussi de l'activation de récepteurs aux facteurs de croissance comme EGFR,



**Figure 15 : La signalisation non-génomique dépendante du récepteur GPER.**

La signalisation correspondant à cette figure est détaillée dans le paragraphe concernant les voies de signalisation non génomiques.

(A) Voies de signalisation induites par l'activation des protéines G hétérotrimériques de type **G $\alpha$ s** (sous-unité verte).

(B) Voies de signalisation induites par l'activation des protéines G hétérotrimériques de type **G $\alpha$ q** (sous unité rouge).

**AC** = Adénylate cyclase, **AMPc** = Adénosine monophosphate cyclique, **DAG** = Diacylglycérol, **FT** = Facteurs de transcription variables en fonction des voies de signalisation activées, **HB-EGF** : Heparin-binding EGF-like growth factor, **IP3**= Inositol trisphosphate, **PIP2** = Phosphatidylinositol-4,5-bisphosphate, **PKA**= Protéine Kinase A, **PKC**= Protéine Kinase C, **PLC** = Phospholipase C.

de la mobilisation du calcium intracellulaire, de la production d'oxyde nitrique (NO) ou de second messagers comme l'AMPc. L'ensemble de ces processus aboutit à des effets physiologiques rapides (Moriarty et al., 2006).

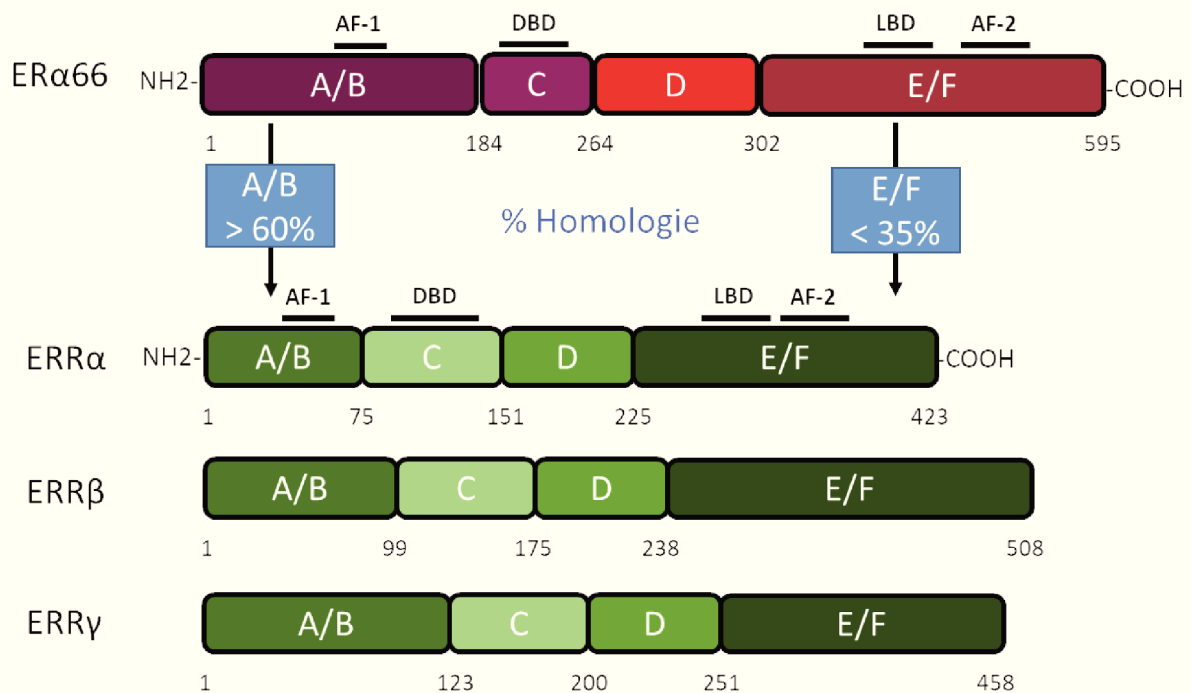
Concernant la signalisation non génomique dépendante du récepteur membranaire GPER, elle met en jeu des partenaires protéiques différents mais elle conduit à l'activation d'une signalisation intracellulaire similaire à celle décrite précédemment (Figure 15). Plus précisément, la stimulation de GPER par les œstrogènes conduit à un changement de conformation du récepteur permettant une interaction directe entre GPER et les protéines G hétérotrimériques de type G $\alpha$ s. Suite à l'hydrolyse du GTP en GDP, les protéines G $\alpha$ s activent différents effecteurs intracellulaires comme l'adénylate cyclase qui conduit à la production d'AMPc et/ou la protéine c-Src qui contribue au déclenchement de la signalisation dépendante d'EGFR. D'autres études suggèrent également que GPER pourraient activer les protéines G de type G $\alpha$ q et Gai/o afin de permettre la libération du calcium intracellulaire via l'activation de la phospholipase C et d'activer les voies de signalisation PI3K/Akt et MAPK. (Thomas et al., 2005 ; Revankar et al., 2005 ; Vrtačnik et al., 2014 ; Pupo et al., 2016 ; Gonzalez de Valdivia et al., 2017).

### C) Les récepteurs apparentés aux récepteurs aux œstrogènes : les ERRs

Les récepteurs apparentés aux récepteurs aux œstrogènes (ERR) sont des récepteurs orphelins appartenant à la classe NR3 des récepteurs nucléaires. Cette famille qui a été mise en évidence dans les années 1990 compte 3 membres : ERR $\alpha$ , ERR $\beta$  et ERR $\gamma$  (Figure 16)(Giguère et al., 1988 ; Hong et al., 1999).

Ces ERRs possèdent une structure caractéristique des récepteurs nucléaires et présentent une forte homologie avec les ERs au niveau du domaine de liaison à l'ADN. Cela leur permet de se fixer au niveau des EREs et de réguler l'expression des mêmes gènes cibles que les ERs. En parallèle les ERRs peuvent également se fixer au niveau d'éléments de réponse spécifiques appelés ERREs dont la séquence consensus est : 5' TnAAGGTCA 3' (Giguère, 2002). Du point de vue de la régulation transcriptionnelle, les ERRs recrutent préférentiellement les coactivateurs PGC-1a et 1b mais ils sont également capables d'interagir avec les mêmes corégulateurs que les ERs (p160/Src, RIP140) (Giguère, 2008 ; Misawa et Inoue, 2015).

Le domaine de liaison au ligand est peu conservé entre les ERs et les ERRs. Cela pourrait expliquer que ces récepteurs ne soient pas capables de lier les œstrogènes (Horard et Vanacker,



**Figure 16 : Structure des récepteurs apparentés aux récepteurs aux œstrogènes (ERR).**

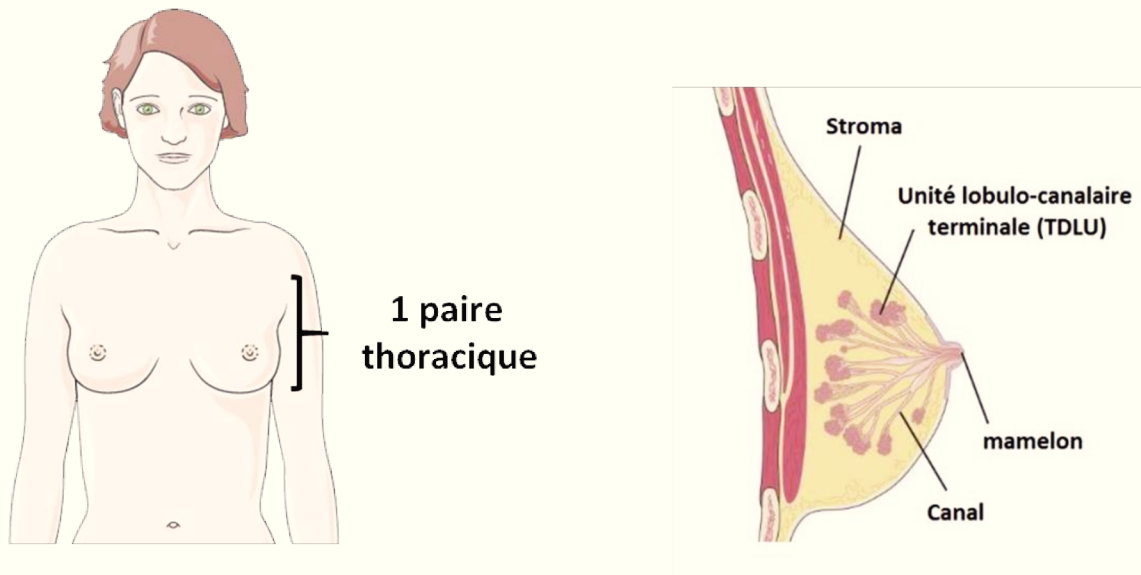
Les 3 isoformes des ERRs (ERRα, ERRβ, ERRγ) possèdent une structure similaire à celle de la forme canonique des récepteurs aux œstrogènes ERα66. Le pourcentage d'homologie entre les domaines A/B et E/F d'ERα66 et des ERRs est indiqué dans les cadres bleus.

(D'après Misawa et al., 2015 ; Horard et Vanacker, 2003)

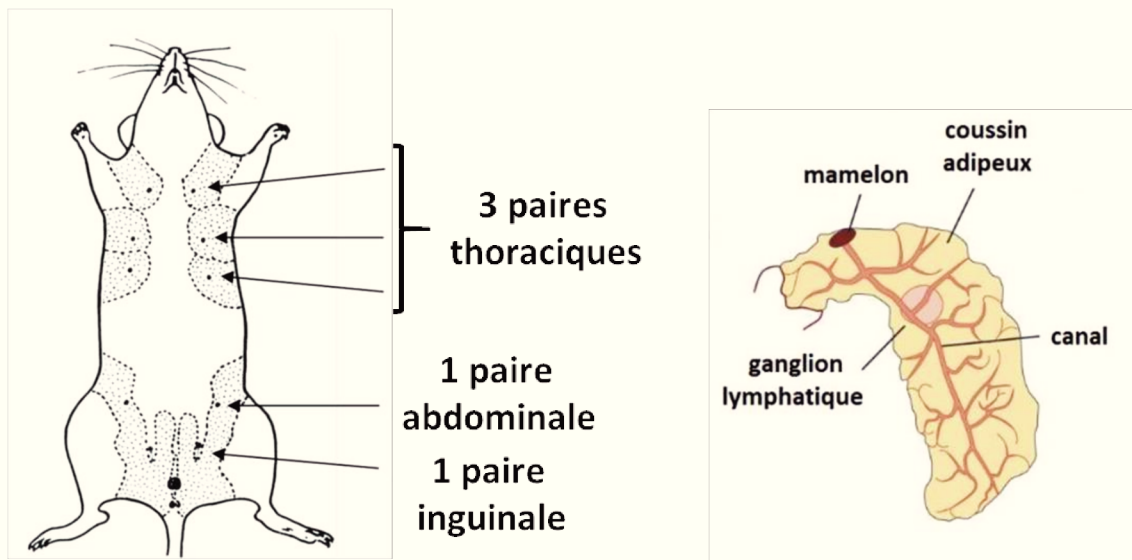
2003). Bien qu'aucun ligand endogène n'ait à l'heure actuelle été identifié pour ces récepteurs, plusieurs études basées sur la structure cristallographique du LBD suggèrent qu'ils pourraient interagir avec des phytoestrogènes, comme par exemple la genistéine ou la daidzéine, qui auraient des effets agonistes ou avec des molécules de synthèse, comme le diéthylstilbestrol (DES) ou le 4-hydroxytamoxifène (4-OHT = métabolite actif du tamoxifène), qui auraient plutôt des effets antagonistes. De plus, ces études structurales indiquent que le domaine de transactivation AF-2 se trouve en conformation active en absence de ligand ce qui permet aux ERRs de fonctionner de manière constitutive. (Tremblay et al., 2001 ; Suetsugi et al., 2003 ; Misawa et Inoue, 2015).

Bien que peu d'études soient disponibles dans la littérature, il semble que les ERRs pourraient stimuler des voies de signalisation non génomiques. Gong et al., (2016), ont par exemple montré que ERR $\gamma$  est capable d'activer la voie PI3K/AkT dans le contexte des cellules épithéliales glomérulaires. Enfin, il a été montré que l'expression des ERRs peut être régulée par la voie des MAPKs dans les cellules cancéreuses mammaires (Heckler et al., 2014).

A



B



**Figure 17 : Anatomie des glandes mammaires humaines et murines à l'âge adulte.**

- (A) Chez la femme, on compte une paire de glandes mammaires localisée en position thoracique. La glande mammaire humaine se compose d'unités lobulo-canaulaires terminales (TDLU), qui correspondent à des grappes de petites alvéoles débouchant sur un vaste réseau de canaux ramifiés qui convergent vers le mamelon.
- (B) Chez la souris femelle, on compte 5 paires de glandes mammaires. La glande mammaire murine comprend un canal primaire qui aboutit au mamelon ainsi que des ramifications secondaires et tertiaires qui envahissent le stroma aussi appelé coussin adipeux.

(D'après Visvader, 2009 ; Chiche, 2012)

## Chapitre 2 : Le développement normal et pathologique de la glande mammaire

La glande mammaire est une glande exocrine *d'origine ectodermique* caractéristique des mammifères dont la fonction principale est de produire et sécréter le lait nécessaire à la nutrition et à l'apport d'une protection immunitaire au nouveau-né. Le développement séquentiel de cette glande débute durant la vie embryonnaire, se poursuit à la puberté et s'achève à la fin de la première gestation. Dans certains cas, ce développement peut être perturbé par *une combinaison de facteurs qui peuvent conduire à l'apparition d'un cancer du sein*.

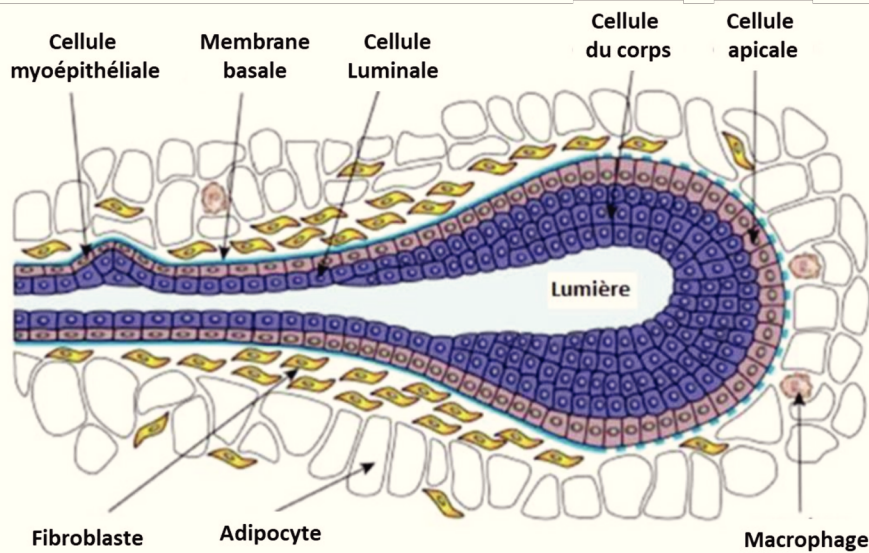
*En fonction de l'espèce*, le nombre et la morphologie de la glande mammaire peuvent varier. Dans ce manuscrit, nous nous intéresserons particulièrement aux glandes mammaires humaines et murines sur lesquelles ces travaux de thèse ont été réalisés.

### I. Organisation générale de la glande mammaire

Dans l'espèce humaine, on compte une paire de glandes mammaires localisées de part et d'autre du sternum en avant des muscles pectoraux, tandis que chez la souris on dénombre 3 paires de glandes mammaires en position thoracique, une quatrième paire en position abdominale et enfin une cinquième paire en position inguinale (Figure 17).

La glande mammaire se compose d'un réseau épithélial canalaire plus ou moins complexe en fonction de l'espèce qui se développe et se ramifie au sein d'un stroma mammaire riche en adipocytes aussi appelé coussin adipeux. Chez la femme, l'arbre mammaire se compose de 15 à 20 lobes qui sont drainés par des canaux galactophores convergeant vers le mamelon. Chacun de ces lobes est constitué de plusieurs lobules composés d'alvéoles (ou acini) agencées en grappes appelées unités lobulo-canaliaires terminales (TDLU). Ce sont ces structures qui permettront la sécrétion du lait au moment de la lactation (Figure 17).

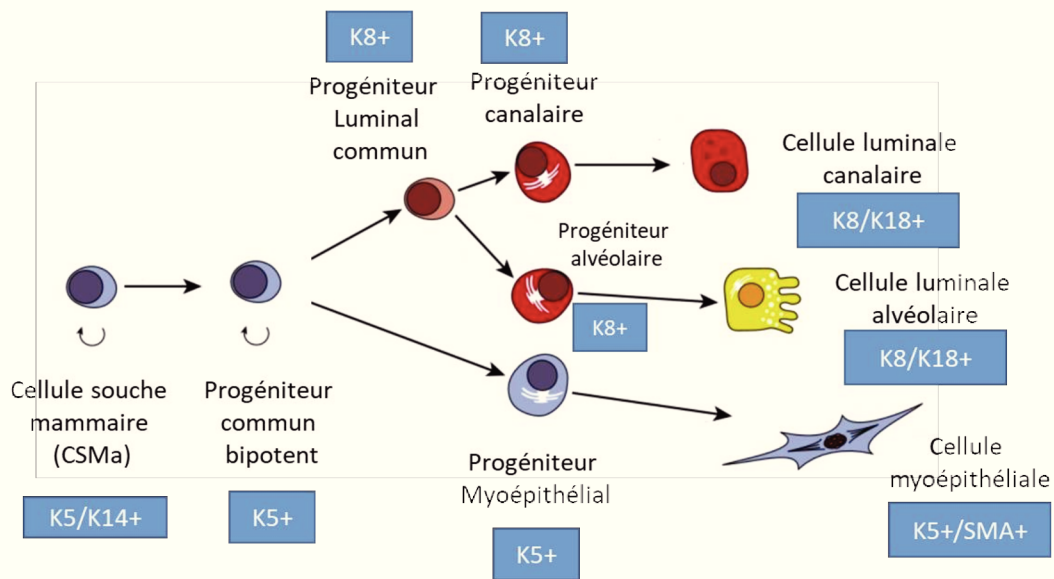
Chez la souris, l'organisation de l'arbre mammaire se résume à un canal primaire débouchant au mamelon et présentant plusieurs ramifications secondaires et tertiaires qui envahissent le coussin adipeux au cours de la puberté. Aux extrémités de ces ramifications on trouve des bourgeons alvéolaires qui se développent à chaque cycle oestrien et qui permettront la sécrétion du lait au niveau des bourgeons terminaux (TEB) au moment de la lactation (Figure 17) (Visvader, 2009).



**Figure 18 : Histologie des canaux mammaires murins.**

Les canaux mammaires sont constitués de cellules luminales qui tapissent la lumière des canaux et de cellules myoépithéliales qui reposent sur une membrane basale. Aux extrémités des canaux se trouvent les bourgeons terminaux (TEB) qui se composent de plusieurs couches de cellules du corps, entourées par une monocouche de cellules apicales. Concernant le stroma, qui joue un rôle de soutien et qui régule le développement du tissu mammaire, il contient majoritairement des adipocytes, des fibroblastes et des macrophages.

(D'après Visvader., 2009)



**Figure 19 : Modèle hypothétique de ségrégation des lignages épithéliaux de la glande mammaire.**

Les différents types cellulaires constituant l'épithélium mammaire, à savoir, les cellules luminales canales, les cellules luminales alvéolaires et les cellules myoépithéliales, proviennent d'une population de cellules souches mammaires (CSMa) capables de s'autorenouveler et de générer des progéniteurs communs bipotents. Ces derniers, qui sont également capables d'autorenouvellement, donneront des progéniteurs engagés dans la voie de différenciation luminaire ou myoépithéliale. Chaque population cellulaire du lignage épithélial est caractérisée par l'expression de marqueurs indiqués dans les encadrés bleus.

**K**= Cytokératine, **SMA** = Actine de muscle lisse (Smooth Muscle Actin).

(D'après Visvader et Stingl, 2014 ; Yang X. et al., 2016)



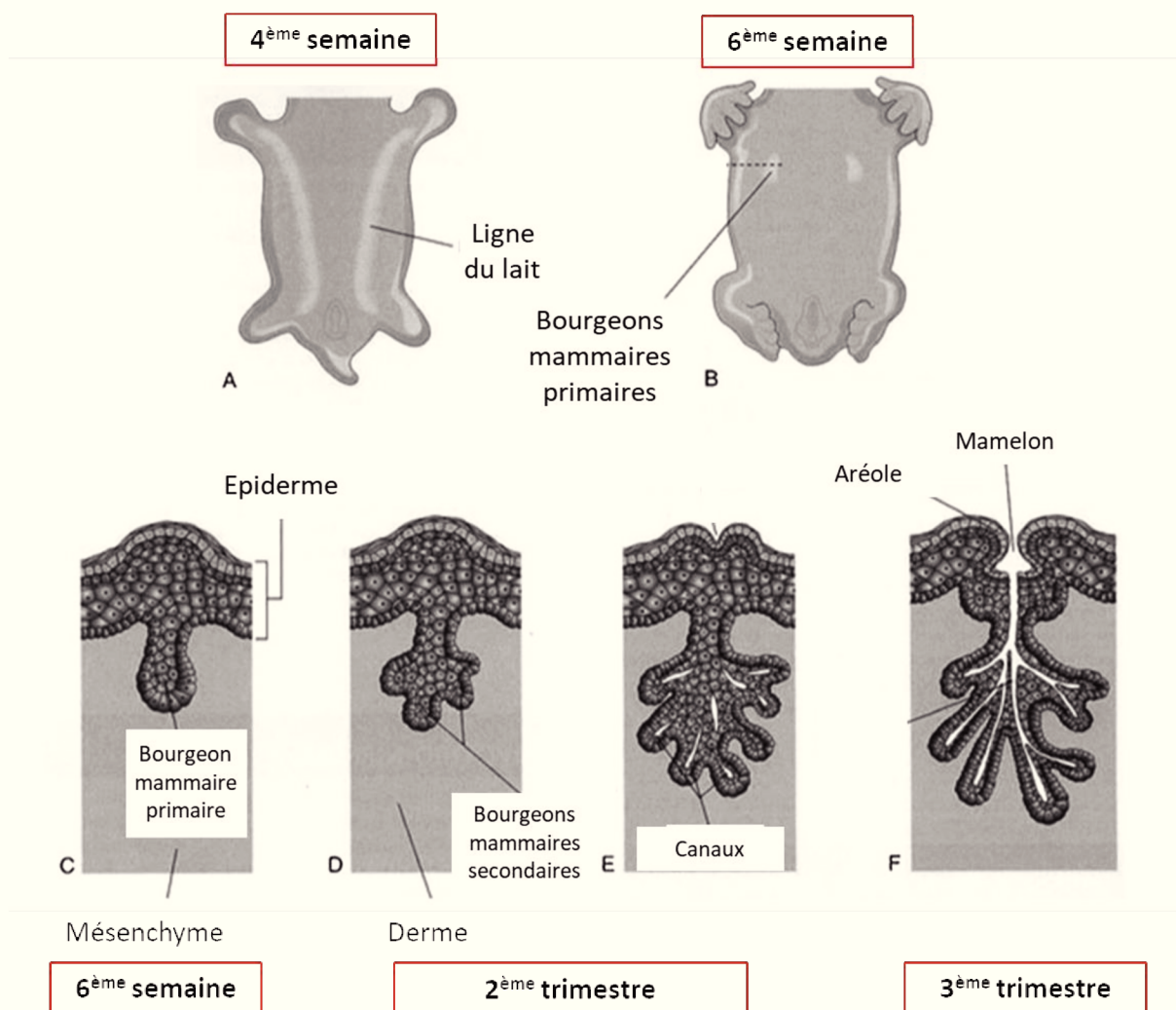
### A) L'épithélium mammaire

L'épithélium mammaire est pluristratifié. La lumière des canaux et des alvéoles mammaires est tapissée par une couche de cellules luminales, canalaire ou alvéolaires, de forme cubique, polarisées et reliées les unes aux autres par des jonctions serrées. Ces cellules reposent sur une monocouche de cellules basales myoépithéliales de forme aplatie et présentant des propriétés contractiles (Figure 18).

Ces 2 types cellulaires sont caractérisés par l'expression de différentes cytokératines et de protéines d'adhérence spécifiques. Les cellules luminales expriment les cytokératines 8 et 18 ainsi que des protéines impliquées dans les jonctions serrées comme les claudines (Naimi et al., 2009 ; Inman et al., 2015). Certaines de ces cellules expriment également des récepteurs hormonaux comme les récepteurs aux œstrogènes, à la progestérone ou à la prolactine (Briskin et O'Malley, 2010). Au moment de la lactation les cellules luminales alvéolaires synthétisent le lait maternel qui se compose majoritairement d'eau, de lipides, de glucides (lactose) et de protéines (WAP (Whey Acidic Protein),  $\beta$ -caséine et  $\beta$ -lactoglobuline) (Wagner et al., 2004).

Concernant les cellules myoépithéliales, elles sont caractérisées notamment par l'expression des cytokératines 5, 14 et 17 ainsi que par l'expression de la P-cadhérine et des cadhérines desmosomiques : desmocolline-3 et desmoglérine-3. Elles expriment également différents marqueurs associés aux muscles lisses comme l' $\alpha$ -actine et la myosine de muscle lisse (Gudjonsson et al., 2005 ; Moumen et al., 2011). Ces cellules myoépithéliales mammaires permettent le transport du lait des alvéoles vers le mamelon grâce à leurs propriétés contractiles.

Différentes expériences de transplantation et de clonogénicité menées chez la souris ont permis de mettre en place un modèle hypothétique de ségrégation des lignages épithéliaux de la glande mammaire. Dans ce modèle, toutes les cellules du compartiment épithélial sont issues d'une population de cellules souches multipotentes mammaires (CSMa) capables de se diviser de manière asymétrique pour s'autorenouveler et de générer une population de progéniteurs communs bipotents. A leur tour, ces progéniteurs communs donneront des progéniteurs myoépithéliaux à l'origine des cellules myoépithéliales canalaire et des progéniteurs luminaux qui seront à l'origine des cellules luminales canalaire mais aussi des cellules luminales alvéolaires au moment de la gestation (Figure 19) (Visvader, 2009; Visvader et Stingl, 2014 ; Yang X. et al., 2016).



**Figure 20 : Développement embryonnaire de la glande mammaire humaine chez le fœtus féminin (XX).**

Le développement de la glande mammaire humaine débute à la 4<sup>ème</sup> semaine de grossesse avec la mise en place de la ligne du lait (A) qui va ensuite régresser pour former une paire de bourgeons mammaires primaires, qui s'étendent dans le mésenchyme sous-jacent, aux alentours de la 6<sup>ème</sup> semaine (B et C).

Le 2<sup>ème</sup> trimestre de la grossesse, se caractérise par le développement et la ramification des bourgeons mammaires secondaires qui aboutit à la formation des canaux de l'arbre mammaire primitif (D et E).

Enfin, le 3<sup>ème</sup> trimestre correspond à la mise en place de l'aréole et du mamelon (F).

Chez le fœtus masculin (XY), le développement de la glande mammaire suit le même schéma temporel que la glande mammaire du fœtus féminin (XX) durant la vie embryonnaire. C'est seulement au moment de la puberté que le dimorphisme sexuel de la glande mammaire se met en place.

Les encadrés rouges indiquent le stade de la grossesse auquel ces étapes du développement de la glande mammaire se déroulent.

(D'après Javed et Lteif, 2013)

## B) Le stroma mammaire

Le compartiment épithélial de la glande mammaire est séparé du stroma par une membrane basale composée de laminines, de collagène, de protéoglycanes et de nidogène. Les cellules myoépithéliales qui reposent sur cette membrane basale sont directement exposées aux signaux provenant de la matrice extracellulaire sous-jacente et elles contribuent à la transmission de ces signaux aux cellules luminales qui n'établissent qu'un contact limité avec le stroma (Maller et al., 2010).

Concernant sa composition, le stroma mammaire est constitué principalement d'adipocytes mais aussi de fibres nerveuses, de vaisseaux sanguins, de vaisseaux lymphatiques et de cellules immunitaires entourés par de la matrice extracellulaire sécrétée par des fibroblastes (Figure 18). La proportion des différents constituants du stroma peut varier entre les espèces. Par exemple, le stroma de la glande mammaire murine contient proportionnellement plus d'adipocytes que celui de la glande mammaire humaine qui est plus fibreux (Visvader, 2009).

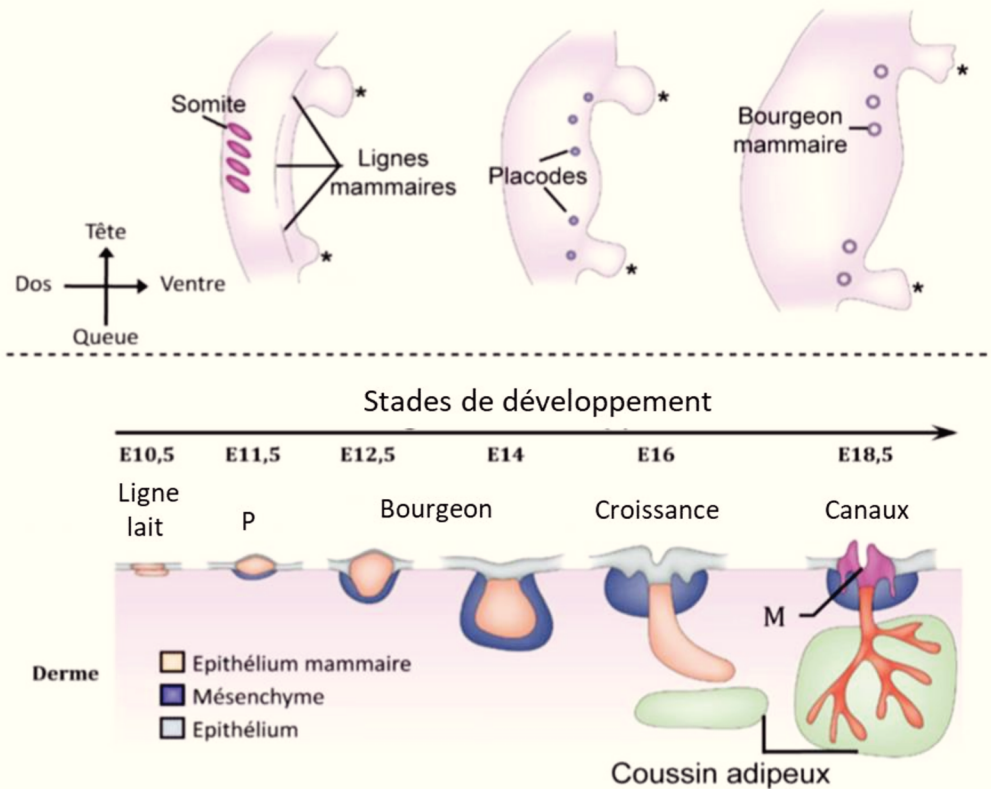
Ce compartiment stromal participe à la régulation du développement et de la différenciation du tissu mammaire grâce à la production de facteurs systémiques et paracrines comme le FGF-2 ou 7 et les métalloprotéinases matricielles MMP-2 et 14 qui peuvent être sécrétés par les différents types cellulaires contenus dans le stroma (Tanos et Briskin, 2008).

## II. Morphogénèse de la glande mammaire

Le développement de la glande mammaire humaine et murine se déroule de façon séquentiel. Il débute par une phase de développement embryonnaire qui est très limitée et qui aboutit à la formation d'un arbre rudimentaire qui restera quiescent jusqu'à la puberté où le développement postnatal reprendra sous l'influence de la signalisation hormonale.

### A) Le développement embryonnaire de la glande mammaire

Dans l'espèce humaine, le développement embryonnaire de la glande mammaire débute au cours du premier trimestre de gestation de façon totalement indépendante des signalisations hormonales induites par les œstrogènes, la progestérone ou la prolactine (Figure 20) (Hens et Wysolmerski, 2005). La 4<sup>ème</sup> semaine de vie fœtale se caractérise par la mise en place de 2 crêtes mammaires ventrales également appelées ligne du lait qui correspondent à un épaississement de l'ectoderme s'étendant des aisselles fœtales jusqu'à la région inguinale. Ces crêtes vont ensuite régresser aux alentours de la 6<sup>ème</sup> semaine pour former 2 bourgeons mammaires primaires localisés en position thoracique qui vont se développer dans le tissu



**Figure 21 : Développement embryonnaire de la glande mammaire murine chez le fœtus femelle (XX).**

Le développement embryonnaire de la glande mammaire murine débute à E10,5 c'est-à-dire à 10,5 jour de vie embryonnaire, avec la mise en place de la ligne du lait qui va ensuite régresser pour donner 5 paires de placodes mammaires (P) à E11,5. Ces dernières se différencient ensuite en bourgeons primaires qui s'étendent dans le mésenchyme sous-jacent entre les stades E12,5 et E14. Au stade E16, les bourgeons mammaires forment des cordons dans le mésenchyme, qui s'est densifié, pour rejoindre le coussin adipeux et former au stade E18,5 un arbre mammaire primitif. Chez le fœtus mâle (XY), le développement de la glande mammaire suit le même schéma temporel que la glande mammaire du fœtus femelle (XX) jusqu'au stade E14. Ensuite, les bourgeons mis en place régressent sous l'influence de la signalisation androgénique.

(D'après Robinson G.W. et al., 2007)

mésenchymateux sous-jacent. En parallèle, les cellules mésenchymateuses commencent à se différencier pour donner les adipocytes, les fibroblastes et les cellules endothéliales qui constituent le stroma mammaire. Lors du 2<sup>ème</sup> trimestre, des bourgeons épithéliaux secondaires se développent et se ramifient dans le mésenchyme qui entoure le bourgeon primaire pour former les canaux mammaires. Au sein de ces structures on observe une différenciation des cellules d'origine ectodermique en cellules luminales ou cellules myoépithéliales. A ce stade, l'arbre mammaire rudimentaire se compose de 15 à 20 lobes associés à des canaux galactophores. Peu de modifications surviendront lors du 3<sup>ème</sup> trimestre qui va surtout permettre la mise en place de la structure du mamelon et de l'aréole (Robinson G.W., 2007 ; Gusterson et Stein, 2012 ; Javed et Lteif, 2013). Bien que ces étapes du développement soient indépendantes des signaux hormonaux, les événements survenant aux cours des 2<sup>ème</sup> et 3<sup>ème</sup> trimestre peuvent être perturbés par l'exposition à des perturbateurs endocriniens (Briskin et O'Malley, 2010).

Concernant la glande mammaire murine, les crêtes mammaires se forment à partir de l'ectoderme au 10,5<sup>ème</sup> jour de vie embryonnaire (E10,5) puis elles régressent à E11,5 pour former 5 paires de placodes mammaires qui se différencieront en 5 paires de bourgeons primaires entre les stades E12,5 et E13,5. Ces bourgeons se développent en profondeur dans le tissu mésenchymateux sous-jacent qui se condense pour former au stade E14,5 un mésenchyme dense composé de glycoprotéines comme la tenascine et la fibronectine et exprimant les récepteurs aux androgènes et aux œstrogènes. En parallèle, le futur coussin adipeux commence à se développer à partir du mésenchyme dense profond. Au stade E16, sous l'influence de la signalisation induite par le peptide apparenté à l'hormone parathyroïdienne, l'épithélium du bourgeon mammaire s'allonge en cordon au sein du mésenchyme dense pour rejoindre le coussin adipeux et former au stade E18,5 un arbre mammaire primitif qui se compose d'un canal principal et de 10 à 15 branches (Figure 21) (Veltmaat et al., 2003 ; Macias et Hinck, 2012 ; Sakakura et al., 2013).

Jusqu'à ce stade de développement, le dimorphisme sexuel de la glande mammaire est peu marqué. En effet, à la naissance, la taille et la morphologie de la glande mammaire diffèrent peu entre les mâles et les femelles. Chez la souris mâle, on observe simplement une rupture de la connexion entre le canal primaire et le mamelon qui est due à l'activation des récepteurs hormonaux localisés dans le mésenchyme dense par les androgènes produits par les testicules aux alentours du stade E14 (Sakakura et al., 2013). C'est seulement au moment de la puberté

**Tableau III : La classification des stades de développement pubertaire de la glande mammaire selon Tanner.**

Stade	Description
I	Absence de développement mammaire L'aréole est petite et plate Élévation du mamelon
II	Apparition d'un petit bourgeon mammaire Augmentation du diamètre de l'aréole
III	Le volume du sein commence à augmenter mais son contour ne se dessine pas encore. L'aréole continue de s'élargir
IV	Développement maximal du sein (apparition d'un sillon sous-mammaire) L'aréole et le mamelon se surélèvent
V	Aspect adulte avec disparition de la saillie de l'aréole qui se retrouve dans le même plan que le sein

La classification de Tanner permet de décrire en 5 stades le développement pubertaire de la glande mammaire. Cette classification a également été adaptée pour la description du développement pubertaire des organes génitaux chez l'homme (Non présenté ici).

(D'après Marshall et Tanner, 1969)

que le dimorphisme sexuel de la glande mammaire sera mis en place sous l'influence des hormones sexuelles féminines.

### B) Le développement post-natal de la glande mammaire

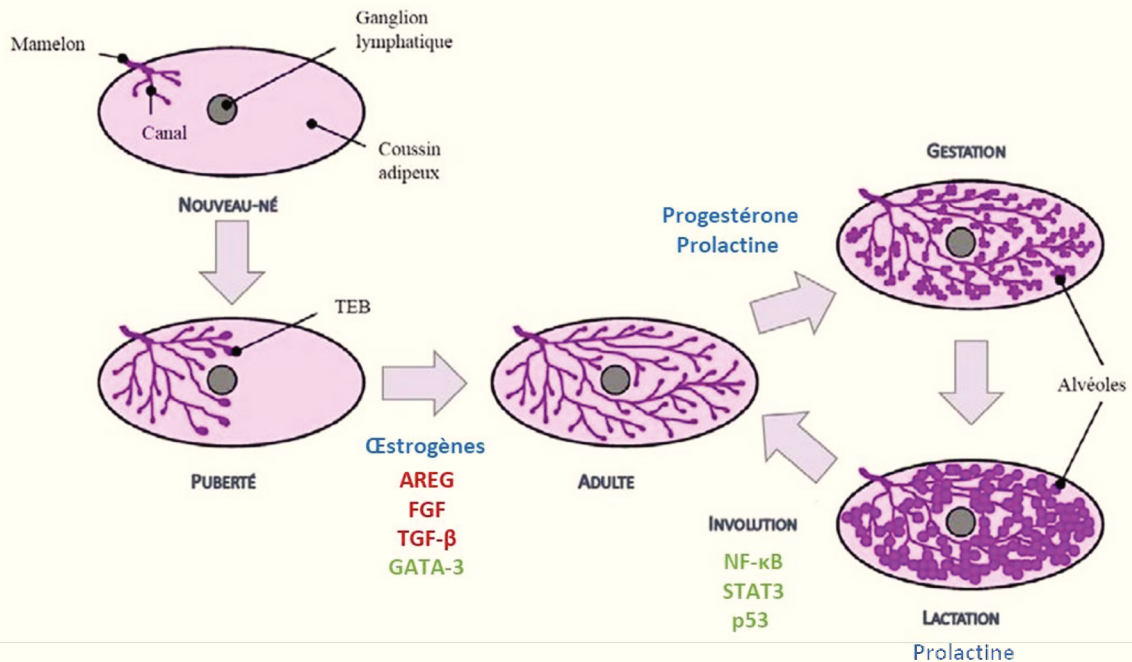
Au terme des étapes du développement embryonnaire, la glande mammaire reste quiescente jusqu'à la puberté qui correspond à la première phase de son développement post natal. Cette première étape qui se caractérise par le développement canalaire de la glande, est suivi par une seconde phase de développement post natal qui intervient au moment de la gestation et qui se caractérise par la différenciation alvéolaire et l'acquisition des propriétés fonctionnelles de la glande. L'ensemble de ce développement est contrôlé par les principales hormones sexuelles féminines à savoir les œstrogènes, la progestérone et la prolactine.

#### *i) La puberté*

Chez la femme, le développement pubertaire de la glande mammaire débute à un âge compris entre 8 ans ½ et 13 ans. L'acquisition de ce caractère sexuel secondaire se déroule en plusieurs étapes décrites par la classification de Tanner, énoncée en 1969 (Tableau III). Brièvement, il s'agit d'un développement en 5 étapes qui débute par une élévation du mamelon au stade 1 et qui se poursuit par le développement d'un bourgeon mammaire au stade 2. On observe ensuite une augmentation du volume du sein qui se dessine progressivement entre les stades 3 et 4 pour atteindre sa morphologie adulte au stade 5. La progression entre les stades 1 et 5 prend en moyenne 4 ans ½ (Marshall et Tanner, 1969 ; Javed et Lteif, 2013).

Cette évolution morphologique de la glande mammaire est liée à une augmentation importante du volume de tissu adipeux entourant les canaux épithéliaux qui, dans le même temps s'allongent et se ramifient sous l'influence majoritaire des œstrogènes. Aux extrémités de ces canaux, on observe l'apparition de bourgeons terminaux qui se différencieront en bourgeons alvéolaires au cours de la première gestation afin de constituer la partie sécrétrice de la glande (Javed et Lteif, 2013).

Chez la souris, le développement pubertaire de l'arbre mammaire qui débute au moment du sevrage c'est-à-dire à environ 21 jours après la naissance (PND21) se caractérise par la mise en place des premiers bourgeons terminaux qui sont le siège d'une prolifération cellulaire importante permettant la formation de ramifications canalaire secondaires et tertiaires qui envahissent le coussin adipeux. Le stade de croissance de l'arbre mammaire correspondant aux



**Figure 22 : Les différents stades postnatals de différenciation de la glande mammaire murine.**

A la naissance, la glande mammaire murine correspond à une structure canalaire primitive insérée dans un coussin adipeux et reliée au mamelon. Sous l'influence des hormones et plus particulièrement des œstrogènes, produits à partir de la puberté, l'arbre mammaire primitif va s'étendre pour envahir le coussin adipeux. Ces mécanismes d'élongation des canaux s'effectuent sous le contrôle d'un certain nombre de facteurs diffusibles, indiqués en rouge, et du facteur de transcription GATA-3.

A chaque gestation, sous l'influence majoritaire de la progestérone et de la prolactine, on observe un développement lobulo-alvéolaire de la glande mammaire qui acquiert des propriétés fonctionnelles nécessaires à la lactation. L'arrêt de la succion du mamelon, au moment du sevrage de la descendance, entraîne le déclenchement de la phase d'involution qui correspond à une apoptose importante des TEB (Bourgeons terminaux) et qui s'effectue notamment sous le contrôle des facteurs de transcription NF-κB, STAT3 et p53.

**AREG** = Amphiréguline, **FGF** = Fibroblast Growth Factor, **TEB** = Terminal End Bud (Bourgeon terminal), **TGF-β** = Transforming Growth Factor.

(D'après Watson et Neoh, 2008)



femelles vierges adultes est atteint aux alentours de la 12<sup>ème</sup> semaine. A cette période, la glande mammaire entame une nouvelle période de quiescence qui prendra fin au moment de la gestation (Figure 22). Ces mécanismes de croissance et de ramification des canaux mammaires qui surviennent chez la souris au moment de la puberté s'effectuent sous le contrôle des œstrogènes, dont les effets sont alors principalement médiés par le récepteur ER $\alpha$ 66 et de l'hormone de croissance qui est sécrétée par l'hypophyse et qui stimule la prolifération des cellules épithéliales mammaires (Mueller et al., 2002 ; Brisken et O'Malley, 2010). Ils sont également régulés par différents facteurs comme l'amphiréguline qui appartient à la famille de l'EGF, les métallo-protéinases matricielles MMP2, 3 et 14, le facteur de transcription GATA-3 ou encore des facteurs de croissance comme le FGF et le TGF $\beta$  (Wiseman et al., 2003 ; Kouros-Mehr et al., 2006 ; Ingman et Robertson, 2008 ; McBryan et al., 2008).

### *ii) Le cycle gestation-lactation-involution*

Chaque gestation s'accompagne d'un cycle de développement lobulo-alvéolaire de la glande mammaire qui comprend 3 phases : l'alvéologénèse, la lactogénèse et enfin l'involution (Figure 22).

Le premier trimestre de gestation se caractérise par une expansion importante de l'arbre mammaire qui acquiert de nouvelles ramifications secondaires et tertiaires. Cette phase de développement qui est principalement contrôlée par la progestérone est suivie, lors de la seconde moitié de la gestation, d'une phase de différenciation alvéolaire correspondant à l'acquisition des propriétés fonctionnelles de la glande. Sous l'effet de la prolactine, les cellules luminales localisées aux extrémités des canaux se différencient en cellules alvéolaires sécrétrices de lait (Brisken et O'Malley, 2010). Pendant cette phase d'alvéologénèse on observe également une réorganisation du stroma avec une infiltration par des cellules immunitaires, une perte du contenu lipidique des adipocytes et une augmentation de la vascularisation qui permet d'acheminer les nutriments nécessaires à la production du lait (Need et al., 2014). A la fin de la gestation, la glande mammaire a atteint son niveau maximal de différenciation fonctionnelle.

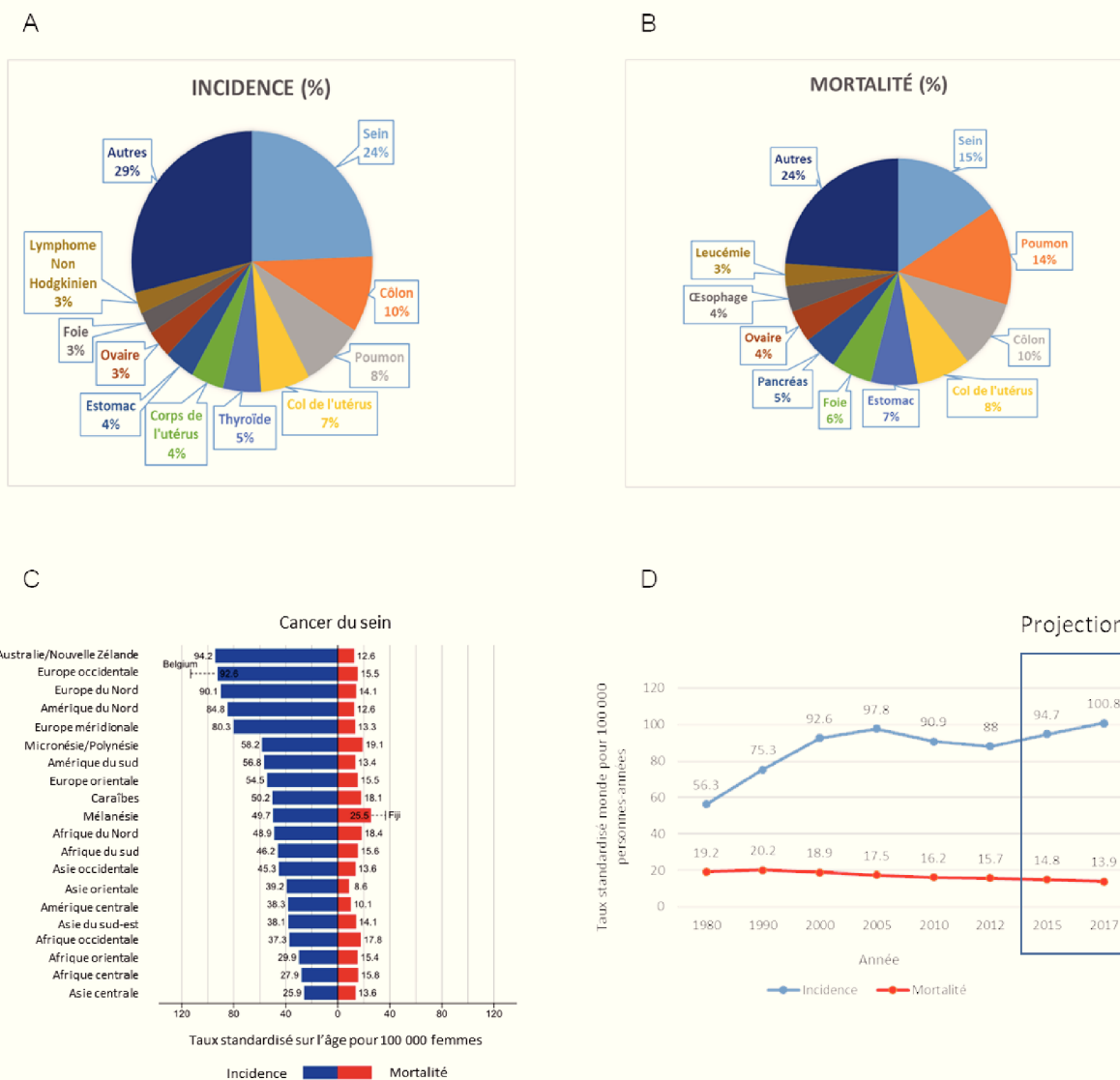
La lactogénèse débute en fin de gestation sous l'influence de la prolactine qui induit l'expression de gènes codant les protéines du lait comme la  $\beta$ -caseine et l' $\alpha$ -lactalbumine. Chez la souris, l'initiation de la lactogénèse est repérée par l'apparition de gouttelettes lipidiques dans l'épithélium mammaire tandis que chez la femme l' $\alpha$ -lactalbumine peut être détectée dans le sang (Neville et al., 2002). Juste après l'accouchement, on observe une chute importante du taux de progestérone qui s'accompagne d'un pic de prolactine permettant le déclenchement de

la seconde phase de la lactogénèse. Cela se traduit par la sécrétion du premier lait appelé colostrum qui est riche en protéines, sels minéraux, vitamines et immunoglobulines. La composition de ce premier lait va progressivement évoluer pour donner le lait dit mature qui s'est enrichi en lactose et en lipides (Rezaei et al., 2016). Durant cette phase, le lait est produit par les cellules alvéolaires puis sécrété dans la lumière des canaux mammaires. La succion du mamelon par le nouveau-né lors de l'allaitement stimule la production d'ocytocine qui, en induisant la contraction des cellules myoépithéliales, permet le transport du lait des alvéoles vers le mamelon.

Au moment du sevrage, c'est-à-dire lorsque la succion du mamelon par le nouveau-né cesse, le lait s'accumule dans les canaux mammaires et on observe une baisse du taux de prolactine et d'ocytocine. Cela conduit à une phase d'involution correspondant à une régression massive du tissu alvéolaire mammaire. L'âge du sevrage qui est d'environ 3 semaines chez la souris est extrêmement variable dans l'espèce humaine pour laquelle l'organisation mondiale de la santé (OMS) recommande de poursuivre l'allaitement au moins jusqu'aux 2 ans de l'enfant. Au niveau cellulaire, l'involution se traduit dans un premier temps par l'apoptose des cellules alvéolaires sécrétrices. Cette première phase, qui est réversible si la succion reprend, est contrôlée notamment par les facteurs de transcription STAT3 et NF- $\kappa$ B ainsi que par la protéine p53 qui induisent l'expression de gènes pro-apoptotiques dans les cellules alvéolaires (Matthews et Clarke, 2005 ; Connelly et al., 2010). Au-delà du troisième jour de sevrage, on observe la mise en place de mécanismes irréversibles qui conduisent à un remodelage important de l'ensemble du tissu mammaire. Lors de cette seconde phase, l'épithélium continue de régresser tandis que les adipocytes du stroma se développent pour combler l'espace libéré par l'épithélium mammaire (Watson, 2006). Au terme de l'involution, la glande mammaire retrouve une morphologie semblable à celle des femelles vierges adultes et entame une nouvelle phase de quiescence qui prendra fin à la prochaine gestation.

### III. Le développement pathologique de la glande mammaire : le cancer du sein

Selon la définition de l'OMS, le terme général de «cancer » s'applique à un grand groupe de maladies pouvant toucher n'importe quelle partie de l'organisme. L'une des caractéristiques principales du cancer correspond à la prolifération anarchique de cellules transformées qui peuvent ensuite se disséminer dans l'organisme pour former des métastases dans d'autres organes. Dans ce manuscrit nous nous intéresserons spécifiquement aux tumeurs malignes qui se développent au niveau du sein et qui sont regroupées sous le terme de « cancer du sein ».



**Figure 23 : Epidémiologie du cancer de sein.**

Répartition de l'incidence (A) et de la mortalité (B) des différents types de cancers en 2018, au niveau mondial, chez la femme. Au total, dans le monde, on recense 8,6 millions de nouveaux cas et 4,2 millions de décès liés au cancer chez la femme en 2018.

(C) Incidence et mortalité du cancer du sein dans les différentes régions du monde. Représentation du taux standardisé sur l'âge pour 100 000 femmes.

(D) Evolution de l'incidence et de la mortalité par cancer du sein entre 1980 et 2012 en France métropolitaine. Représentation du taux standardisé monde pour 100 000 personnes/an.

(D'après Bray et al., 2018 ; INCa 2017)

### A) Epidémiologie du cancer du sein

Au niveau mondial, le cancer du sein constitue la pathologie cancéreuse qui présente la plus forte incidence et la plus forte mortalité chez la femme devant le cancer colorectal et le cancer du poumon (Figure 23A et B). Les données récentes indiquent que 2,1 millions de nouveaux cas de cancer du sein ont été diagnostiqués en 2018 dans le monde ce qui correspond à environ  $\frac{1}{4}$  des cancers diagnostiqués chez la femme la même année. On recense également 630 000 décès liés au cancer du sein en 2018 ce qui représente 15% de la mortalité liée au cancer chez la femme. Toutefois, il est important de remarquer que l'incidence de ce cancer varie dans les différentes régions du monde. En effet, il faut distinguer les pays industrialisés qui présentent une forte incidence du cancer du sein et les pays en développement pour lesquels cette incidence est plus faible (Figure 23C). Concernant le taux de mortalité, les variations géographiques sont moins marquées même si dans certains pays d'Amérique du nord, d'Europe du nord et en Australie, la mortalité liée au cancer du sein est inférieure à celle liée au cancer du poumon chez la femme. Ces observations peuvent être liées à des différences de mode de vie mais aussi à l'existence de programmes et de techniques de dépistage précoces dans les pays industrialisés. (Bray et al., 2018).

En France aussi, le cancer du sein est celui qui présente la plus forte incidence et qui constitue la première cause de décès lié au cancer chez la femme (Figure 23D). Bien que son incidence n'a cessé d'augmenter entre 1980 et 2005, passant de 56,3 à 97,8 cas pour 100 000 femmes, une diminution progressive de l'incidence du cancer de sein a ensuite été observée jusqu'à atteindre un taux d'incidence de 88 cas pour 100 000 femmes en 2012. Depuis, ce taux d'incidence tend de nouveau à augmenter et les projections pour l'année 2017 estiment à 58 968 le nombre de nouveaux cas de cancer du sein. Cela correspond à une incidence de 100,8 cas pour 100 000 femmes.

Concernant la mortalité, elle ne cesse de diminuer depuis 1995, passant de 18,9 à 15,7 décès pour 100 000 femmes entre 2000 et 2012. Les projections pour l'année 2017 s'inscrivent dans la continuité de cette diminution avec une mortalité qui pourrait passer à 13,9 décès pour 100 000 femmes soit une estimation de 11 833 décès liés au cancer du sein en 2017. A l'heure actuelle, aucune étude récente n'est disponible pour confirmer la validité de ces projections. Toutefois, cette corrélation inverse entre l'incidence et la mortalité du cancer du sein depuis 2012 pourrait s'expliquer par le développement du dépistage organisé qui permet une prise en charge plus précoce de la maladie, mais aussi par l'amélioration de la prise en charge

thérapeutique notamment grâce au développement de thérapies plus personnalisées (Jéhannin-Ligier, 2017).

Malgré tout, le cancer du sein en France, reste de bon pronostic avec une survie nette standardisée sur l'âge à 10 ans de 76% (cas diagnostiqués entre 1989 et 2010).

### B) Les facteurs de risques du cancer du sein

*Le cancer du sein est une maladie multifactorielle résultant d'une combinaison de facteurs de risques internes, c'est à dire inhérents à l'individu et de facteurs de risques externes liés à l'environnement et au mode de vie.*

#### *i) Les facteurs de risques internes*

Le premier facteur de risque interne correspond au sexe de l'individu. En effet, les cancers du sein se développent dans 99% des cas chez la femme et seulement dans 1% des cas chez l'homme. L'âge a également son importance puisque la plupart des cancers du sein sont diagnostiqués après 55 ans. Ce constat a d'ailleurs conduit à la mise en place d'un programme de dépistage organisé qui s'adresse aux femmes âgées de plus de 50 ans et qui comprend un examen clinique des seins ainsi qu'une mammographie qu'il est recommandé de réaliser tous les 2 ans. Le suivi du dépistage organisé n'a pas démontré son utilité avant l'âge de 50 ans. Toutefois, un examen clinique des seins réalisé tous les ans est recommandé à partir de 25 ans surtout en cas de risque lié par exemple à l'existence d'antécédents familiaux.

On estime que 5 à 10% des cancers du sein seraient liés à une prédisposition génétique qui peut être recherchée s'il existe des antécédents familiaux. En effet, 20 à 30% des cancers du sein sont diagnostiqués chez des femmes pour lesquelles une parente au premier degré (mère, sœur, fille) a développé un cancer du sein avant l'âge de la ménopause (Feng et al., 2018). Cela peut être dû au hasard ou au mode de vie commun de ces individus mais la plupart du temps, des mutations génétiques qui se transmettent de génération en génération sont détectées au moment du diagnostic (INCa).

Dans le contexte du cancer du sein, la prédisposition génétique la plus connue concerne les gènes suppresseurs de tumeurs BRCA1 et BRCA2 (Breast Cancer Associated 1 et 2) qui codent des facteurs intervenant dans les systèmes de réparation de l'ADN. Ce sont des gènes de haute pénétrance, ce qui signifie que lorsqu'ils sont mutés, le risque de développer un cancer du sein augmente de façon importante. Il augmente de 65% pour une mutation BRCA1 et de 45% pour

une mutation BRCA2. En parallèle, d'autres gènes de prédisposition associés au développement du cancer du sein ont été mis en évidence comme TP53, ATM, CHEK2, PTEN ou CDH1 (Lee et Muller, 2010 ; Aloraifi et al., 2015 ; Feng et al., 2018).

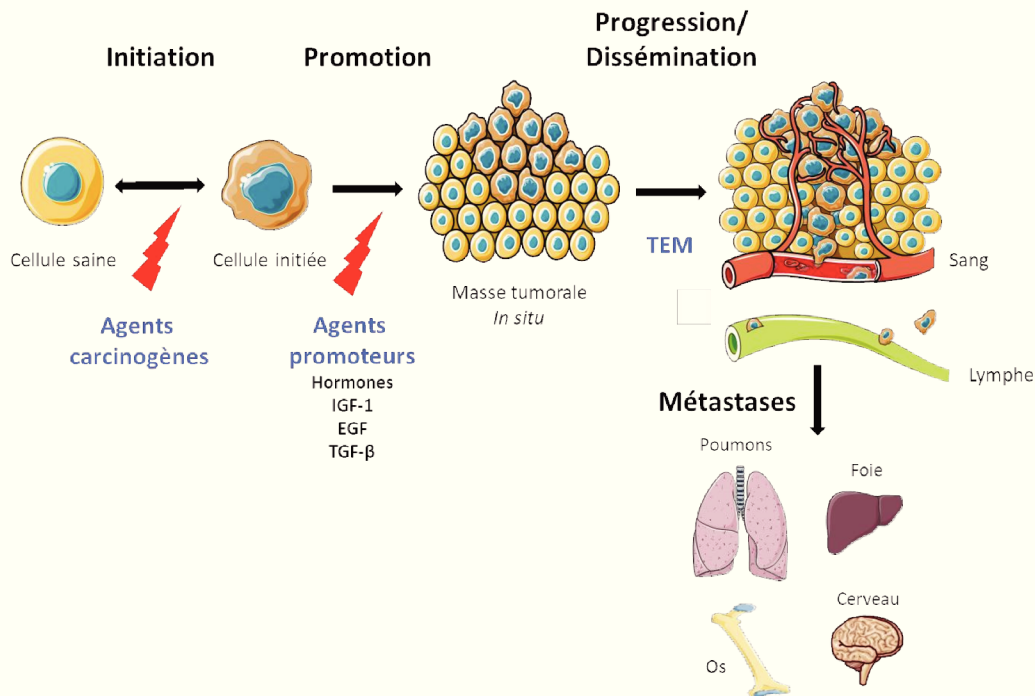
D'un point de vue plus individuel, l'histoire hormonale d'une femme peut également constituer un facteur de risque interne. En effet, plus la période d'exposition aux œstrogènes endogènes sera longue plus le risque de développer un cancer du sein sera important. Par exemple, l'allongement de la période reproductive d'une femme, notamment en lien avec une puberté précoce (avant 12 ans), une ménopause tardive (après 55 ans) ou une nulliparité, est associé à une augmentation de ce risque. A l'inverse, une première grossesse précoce, une multiparité ou un allaitement prolongé sont associés à une diminution du risque de cancer du sein (Feng et al., 2018).

### *ii) Les facteurs de risques externes*

On estime qu'environ 40% des cas de cancer du sein en France sont liés à des facteurs de risques externes dont l'impact pourrait être limité, voire même évité, grâce à la mise en place de mesures de prévention individuelles ou collectives (INCa, 2019).

Parmi les facteurs de risques liés au mode de vie des individus, on trouve la consommation de tabac et d'alcool, la surcharge pondérale liée à une alimentation déséquilibrée et à un manque d'activité sportive ainsi que l'exposition excessive aux UV naturels ou artificiels. De plus, la prise d'une contraception orale ou de traitements hormonaux substitutifs durant la ménopause contribue à augmenter de façon importante l'imprégnation hormonale de l'individu et donc le risque de cancer du sein. Toutefois, le risque associé à la prise d'une contraception orale devient négligeable 10 ans après l'arrêt du traitement (Feng et al., 2018).

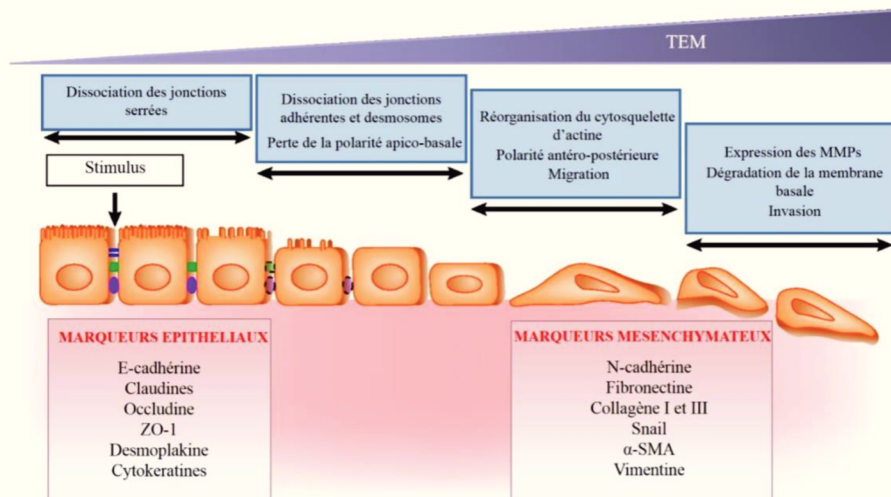
Au niveau environnemental, nous sommes également exposés à un certain nombre de perturbateurs endocriniens et de polluants chimiques comme les pesticides organochlorés qui sont susceptibles de favoriser le développement des tumeurs mammaires. Dans la catégorie des perturbateurs endocriniens, on peut notamment citer le BPA, les parabènes et les alkylphénols qui sont présents dans de nombreux produits à usage courant (alimentation, cosmétiques, contenants/ustensiles en plastique...). On peut également évoquer le DES qui a été prescrit entre les années 1940 et 1970 aux femmes enceintes pour prévenir les avortements spontanés et qui est aujourd'hui associé à des effets transgénérationnels avec un risque accru de cancer du sein pour la personne ayant reçu le traitement mais aussi pour sa descendance. (Charles et Darbre, 2013 ; Seachrist et al., 2016 ; He et al., 2017 ; Feng et al., 2018 ; Rodgers et al., 2018).



**Figure 24 : Les différentes étapes de la carcinogénèse.**

La première étape de la carcinogénèse correspond à l'initiation. Celle-ci est déclenchée après une exposition à des agents carcinogènes qui vont conduire à l'apparition d'altérations génétiques ou épigénétiques au sein d'une cellule saine qui devient alors une cellule initiée.

La seconde étape correspond à la promotion. Plus précisément des agents promoteurs induisent une prolifération anarchique et/ou une survie accrue des cellules initiées qui forment alors une masse tumorale in situ. La troisième et dernière étape correspond à la progression/dissémination. Suite à l'acquisition de propriétés migratrices et invasives, certaines cellules de la masse tumorale peuvent franchir la membrane basale afin de rejoindre les circulations sanguines et lymphatiques. Cela permet la dissémination des cellules cancéreuses qui peuvent alors aller s'implanter dans d'autres tissus pour former des métastases.



**Figure 25 : La transition épithélio-mésenchymateuse.**

Ce schéma décrit les différentes étapes de la transition épithélio-mésenchymateuse.

Au cours de ce phénomène, les cellules perdent l'expression de marqueurs épithélias au profit de marqueurs mésenchymateux :

Marqueurs épithélias : Protéines de jonctions serrées (Claudine, Occludine, ZO-1), protéines de jonctions adhérentes (E-cadhérine) et filaments intermédiaires (Cytokératines).

Marqueurs mésenchymateux : Protéines de surface (N-cadhérine), protéines du cytosquelette (Vimentine,  $\alpha$ -SMA), protéines de la matrice extracellulaire (Fibronectine, Collagène de type I et III) et des facteurs de transcription (Snail).

(D'après Aroeira et al., 2007)

### C) La carcinogénèse mammaire

La carcinogénèse mammaire se déroule classiquement en 3 étapes à savoir l'initiation, la promotion et la progression (Figure 24).

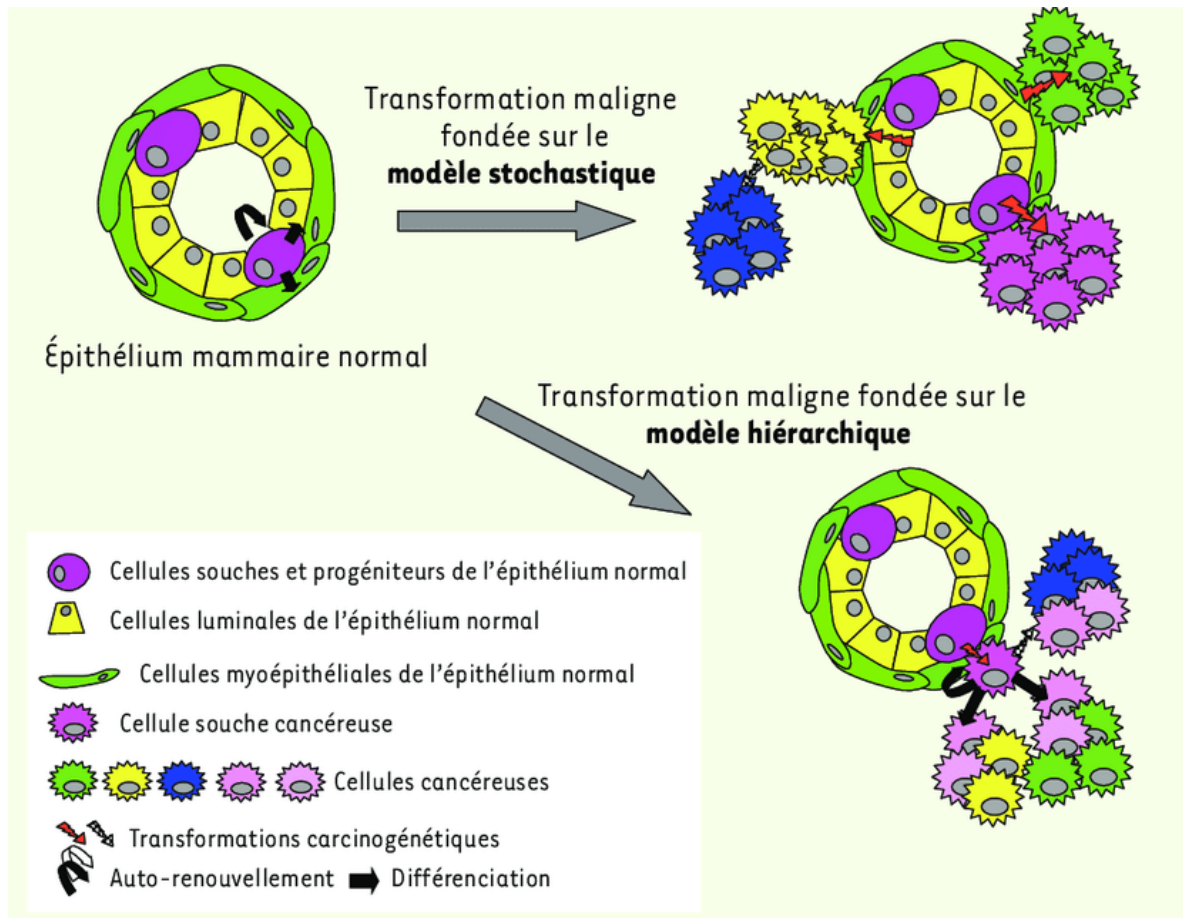
- L'étape d'initiation correspond à l'apparition, dans une cellule saine, d'altérations génétiques ou épigénétiques conduisant à (i) l'activation de proto-oncogènes inducteurs de la survie, de la prolifération et de la migration cellulaire, et à (ii) l'inactivation de gènes suppresseurs de tumeurs qui au contraire inhibent ces mécanismes cellulaires. Ces altérations de l'expression génique qui peuvent être en lien avec des prédispositions génétiques ou apparaître suite à l'exposition à des agents carcinogènes conduisent à la formation d'une cellule dite initiée. Toutefois, grâce à l'intervention des mécanismes de réparation de l'ADN, cette étape d'initiation peut être réversible.

- L'étape de promotion correspond à une phase de croissance caractérisée par une prolifération importante de la cellule initiée en réponse aux hormones stéroïdiennes et à différents facteurs de croissance comme l'IGF-1, le TGF- $\beta$  ou l'EGF qui favorisent en plus de la prolifération, l'échappement à l'apoptose. La promotion aboutit à la formation d'une masse tumorale in situ (Reznikov, 2015).

- L'étape de progression également appelée étape de dissémination correspond à la propagation des cellules de la masse tumorale qui sont transportées par voies sanguines ou lymphatiques afin d'aller coloniser et former des tumeurs secondaires, appelées métastases, dans d'autres tissus. Cette étape dépend de l'acquisition de propriétés migratrices et invasives qui est permise par le phénomène de transition épithélio-mésenchymateuse (TEM). Ce processus se caractérise par la perte d'expression de marqueurs épithéliaux au profit de l'acquisition de marqueurs mésenchymateux caractéristiques des cellules tumorales plus invasives. (Figure 25).

Parmi les caractéristiques majeures du cancer décrites par Hanahan et Weinberg en 2011, on trouve le concept d'hétérogénéité intra-tumorale. Cela signifie que toutes les cellules tumorales qui composent une même tumeur n'ont pas forcément la même morphologie, la même tumorigénicité, la même sensibilité aux traitements ou encore les mêmes propriétés migratrices et invasives. La prise en compte de cette hétérogénéité est donc primordiale pour affiner la stratégie thérapeutique adoptée (Welch, 2016 ; Turashvili et Brogi, 2017).





**Figure 26 : Les 2 modèles explicatifs du concept d'hétérogénéité intra-tumorale.**

Afin d'expliquer l'hétérogénéité intra-tumorale, 2 théories ont été développées :

Le modèle stochastique : la transformation maligne peut être initiée à partir de n'importe quel type cellulaire de l'épithélium mammaire. Chaque clone tumoral généré va ensuite pouvoir évoluer indépendamment des autres en fonction des nouvelles altérations oncogéniques qu'il accumule. Cela contribue à la formation d'une tumeur hétérogène.

Le modèle hiérarchique : la transformation maligne peut être initiée uniquement à partir des cellules souches et des progéniteurs de l'épithélium mammaire. Le caractère pluripotent de ces cellules leur permet de générer différents types de cellules cancéreuses mammaires qui constituent ainsi une masse tumorale hétérogène.

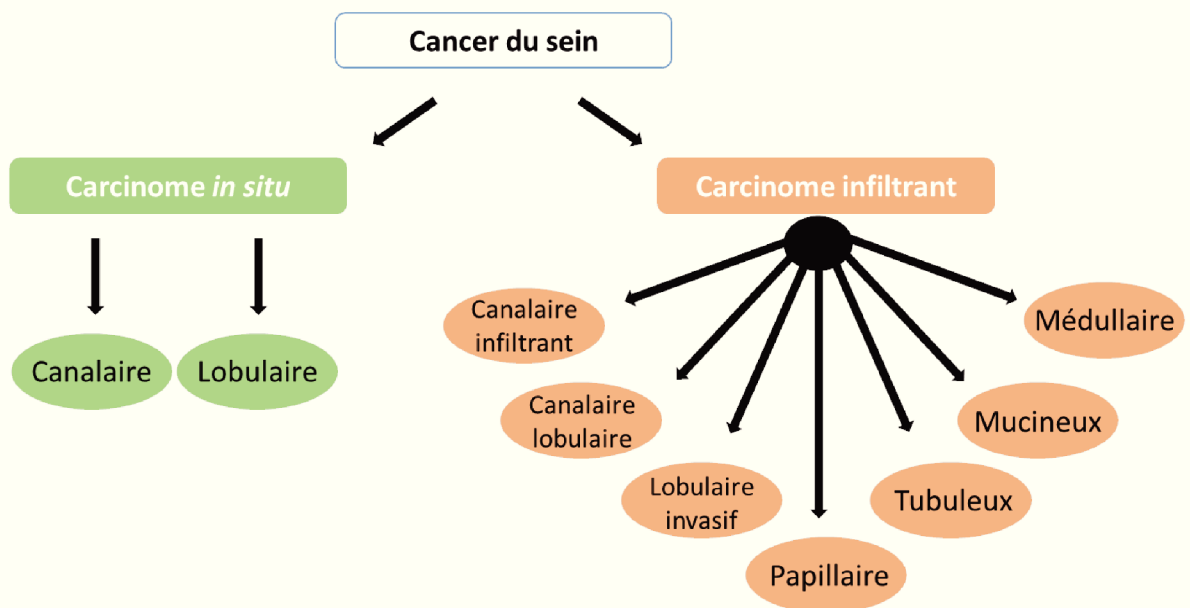
(D'après Ginestier et al., 2007)

Pour tenter d'expliquer cette hétérogénéité intra-tumorale, 2 modèles de carcinogénèse ont été proposés (Figure 26). Tout d'abord, le modèle stochastique propose que toutes les cellules qui composent un tissu, et cela quel que soit leur degré de différenciation, sont susceptibles d'initier la formation d'une tumeur lorsqu'elles accumulent des mutations génétiques ou épigénétiques. Si on applique ce modèle au cancer du sein cela signifie que toutes les cellules qui composent l'épithélium mammaire à savoir, les cellules souches, les différents progéniteurs ainsi que les cellules luminales et myoépithéliales différenciées seraient capables d'initier la formation d'un clone tumoral indépendant. Au fil du temps, les cellules qui constituent ces différents clones pourront évoluer de façon indépendante suite à l'accumulation d'autres mutations pour former une tumeur hétérogène (Ginestier et al., 2008).

Le second modèle de carcinogénèse est le modèle hiérarchique qui considère qu'une tumeur ne peut se développer qu'à partir des cellules d'un tissu qui possèdent des propriétés de cellules souches. Dans le contexte du cancer du sein, le concept des cellules souches cancéreuses mammaires (CSC) a émergé suite aux travaux de Al-Hajj et al., en 2003. En injectant des cellules tumorales mammaires humaines dans le coussin adipeux des glandes mammaires thoraciques de souris immunodéprimées, cette équipe a observé que seule une infime partie de ces cellules sont tumorigéniques c'est-à-dire capables de générer une nouvelle tumeur hétérogène. Ces cellules appelées cellules souches cancéreuses mammaires proviennent des cellules souches normales de la glande mammaire ou des progéniteurs communs bipotents qui ont accumulé des altérations oncogéniques. Les CSCs sont caractérisées par l'expression du marqueur de surface CD44, par la faible expression de CD24 et par l'absence d'expression de marqueurs de cellules différenciées (CD44<sup>+</sup>/CD24<sup>-low</sup>/Lineage-) (Al-Hajj et al., 2003 ; Dittmer, 2018). De plus, ces cellules présentent une forte activité ALDH1 (ALdehyde DesHydrogenase) qui joue un rôle dans le contrôle de la différenciation des cellules souches (Ginestier et al., 2007, 2008). D'un point de vue fonctionnel, les CSCs possèdent des capacités d'autorenouvellement et de différenciation qui leur permettent de reconstituer, à elles seules, une masse tumorale hétérogène (Ginestier et al., 2007 ; Dittmer, 2018).

#### D) La classification des tumeurs mammaires

*En plus de l'hétérogénéité intra-tumorale évoquée précédemment, il existe une grande hétérogénéité dite inter-tumorale entre les tumeurs mammaires provenant de différents patients. En effet, il n'existe pas un cancer du sein mais plusieurs types de cancers du sein. Afin d'orienter au mieux la stratégie thérapeutique à adopter, il existe plusieurs classifications*



**Figure 27 : La classification histologique des tumeurs mammaires.**

95% des tumeurs mammaires correspondent à des adénocarcinomes qui peuvent être localisés (in situ) ou se propager hors du tissu d'origine (infiltrant). Cet arbre représente les différents sous-types d'adénocarcinomes mammaires répertoriés.

(D'après Malhotra et al., 2010)

des tumeurs mammaires qui reposent sur des critères histologiques, anatomo-pathologiques ou moléculaires.

*i) La classification histologique des tumeurs mammaires*

Les adénocarcinomes de la glande mammaire représentent 95% des cancers du sein. Dans la majorité des cas ces tumeurs se développent à partir des cellules épithéliales des canaux mammaires mais elles peuvent également se développer à partir des cellules épithéliales des lobules dans 10% des cas. On parle alors respectivement, d'adénocarcinome canalaire ou lobulaire (Figure 27).

En fonction du degré d'infiltration de la tumeur on distingue les adénocarcinomes in situ non invasifs et les adénocarcinomes infiltrants. Les adénocarcinomes in situ correspondent à une tumeur localisée et bien délimitée qui n'a pas franchi la membrane basale. Ils représentent ¼ des cancers du sein et ces tumeurs sont de bon pronostic si elles sont diagnostiquées précocement. Toutefois, en absence de prise en charge thérapeutique adaptée les adénocarcinomes in situ peuvent progresser et devenir infiltrants. Dans ce cas les cellules tumorales franchissent la membrane basale et peuvent se propager vers les ganglions axillaires ou vers d'autres organes pour former des métastases. Les adénocarcinomes infiltrants qui représentent 75% des cancers du sein sont associés à un pronostic moins favorable que celui des adénocarcinomes canaux in situ.

Il existe également 4 formes rares d'adénocarcinomes infiltrants qui représentent chacun 1 à 2% des cancers du sein et qui sont caractérisés par un aspect histologique particulier : (i) Le carcinome médullaire touche surtout les femmes jeunes et il est souvent associé à une prédisposition génétique, (ii) Le carcinome mucineux ou colloïde touche les femmes entre 60 et 70 ans et se caractérise par la présence de cellules cancéreuses sécrétrices de mucus, (iii) Le carcinome tubuleux touche les femmes âgées de plus de 55 ans et se propage rarement aux ganglions axillaires, enfin, (iv) le carcinome papillaire touche les femmes âgées et n'est pas toujours invasif. Dans tous les cas, ces formes rares présentent un meilleur pronostic que les adénocarcinomes canaux infiltrants (Malhotra et al., 2010 ; Sharma et al., 2010).

Par ailleurs, outre les adénocarcinomes, d'autres tumeurs mammaires malignes ont été répertoriées comme la tumeur phyllode, le sarcome ou le lymphome.

**Tableau IV : La classification TNM des tumeurs mammaires.**

Stade	Description
0	<b>Tis</b> : Carcinome canalaire ou lobulaire in situ <b>N0</b> : Absence d’envahissement ganglionnaire régional <b>M0</b> : Absence de métastases à distance
I	<b>T1</b> : Tumeur ≤ 2 cm dans sa plus grande dimension <b>N0</b> : Absence d’envahissement ganglionnaire régional <b>M0</b> : Absence de métastases à distance
II	<b>T0</b> : Tumeur primitive non palpable à <b>T3</b> : Tumeur > 5 cm <b>N0</b> : absence d’envahissement ganglionnaire à <b>N1</b> : envahissement de 1 à 3 ganglions axillaires <b>M0</b> : Absence de métastases à distance
III	<b>T0</b> : Tumeur primitive non palpable à <b>T4</b> : Tumeurs de toutes les tailles avec extension à la paroi thoracique ou à la peau <b>N0</b> : absence d’envahissement ganglionnaire à <b>N3</b> : envahissement d’au moins 10 ganglions axillaires ou envahissement des ganglions sous-claviculaires ou des ganglions mammaires internes homolatéraux <b>M0</b> : Absence de métastases à distance
IV	<b>T0</b> : Tumeur primitive non palpable à <b>T4</b> : Tumeurs de toutes les tailles avec extension à la paroi thoracique ou à la peau <b>N0</b> : absence d’envahissement ganglionnaire à <b>N3</b> : envahissement d’au moins 10 ganglions axillaires ou envahissement des ganglions sous-claviculaires ou des ganglions mammaires internes homolatéraux <b>M1</b> : Présence de métastases à distance

La classification TNM prend en compte la taille de la tumeur (T), l’envahissement des ganglions (N), ainsi que la présence de métastases (M). Chacun de ces trois critères est évalué afin de déterminer le stade (0 à IV) des tumeurs mammaires. Cette classification a été établie par l’Union Internationale Contre le Cancer (UICC) et l’American Joint Committee on Cancer (AJCC).

**Tis** = Tumeur in situ

*ii) La classification anatomo-pathologique des tumeurs mammaires*

Afin de compléter la classification histologique en caractérisant plus précisément l'étendue et l'agressivité d'une tumeur mammaire, 2 classifications basées sur l'examen anatomo-pathologique des biopsies de tumeurs ont été mise en place.

Tout d'abord, la classification TNM (Tumor Node Metastasis) de l'Union Internationale Contre le Cancer (UICC) et de l'American Joint Committee on Cancer (AJCC), permet d'évaluer l'étendue d'une tumeur. Elle comprend 5 stades (0 à IV) qui sont définis en fonction de la taille et de l'infiltration de la tumeur (T), de l'atteinte ganglionnaire (N) et de la présence de métastases (M)(Tableau IV).

Concernant le degré d'agressivité d'une tumeur, il est évalué par le grade histopronostique de Scarff-Bloom-Richardson (SBR). Au cours de l'examen anatomo-pathologique, l'architecture tumorale, la taille et la forme du noyau des cellules tumorales ainsi que l'activité mitotique sont évalués et une note allant de 1 à 3 est attribuée à chacun des paramètres (Tableau V). La note globale obtenue permet de classer la tumeur selon plusieurs grades, allant du grade I pour les tumeurs les moins agressives au grade III pour les tumeurs les plus agressives.

Ces classifications anatomo-pathologiques permettent de déterminer l'état d'avancement de la maladie et d'affiner le pronostic. Toutefois, elles ne sont pas adaptées pour orienter la stratégie thérapeutique de façon personnalisée ou pour prédire l'efficacité des traitements qui pourront être utilisés. En effet, seule une classification plus fine basée sur le profil d'expression génique des tumeurs peut permettre de personnaliser la prise en charge thérapeutique.

*iii) La classification moléculaire des tumeurs mammaires*

L'hétérogénéité inter-tumorale ne repose pas uniquement sur des aspects morphologiques. En effet, elle révèle de grandes disparités moléculaires qui ont été mises en évidence suite au développement des techniques de transcriptomique dans les années 2000. Grâce à l'analyse de l'expression de 8102 gènes dans 65 biopsies de tumeurs mammaires, Perou et al., ont établi une première classification moléculaire qui a ensuite été affinée et validée grâce à des données complémentaires. Aujourd'hui, la classification moléculaire des tumeurs mammaires comprend 6 sous-types moléculaires établis en fonction de l'expression de la forme canonique du récepteur aux œstrogènes alpha (ER $\alpha$ 66), du récepteur à la progestérone (PR) et du récepteur aux facteurs de croissance épidermiques humain 2 (HER2) (Figure 28) (Eroles et al., 2012).

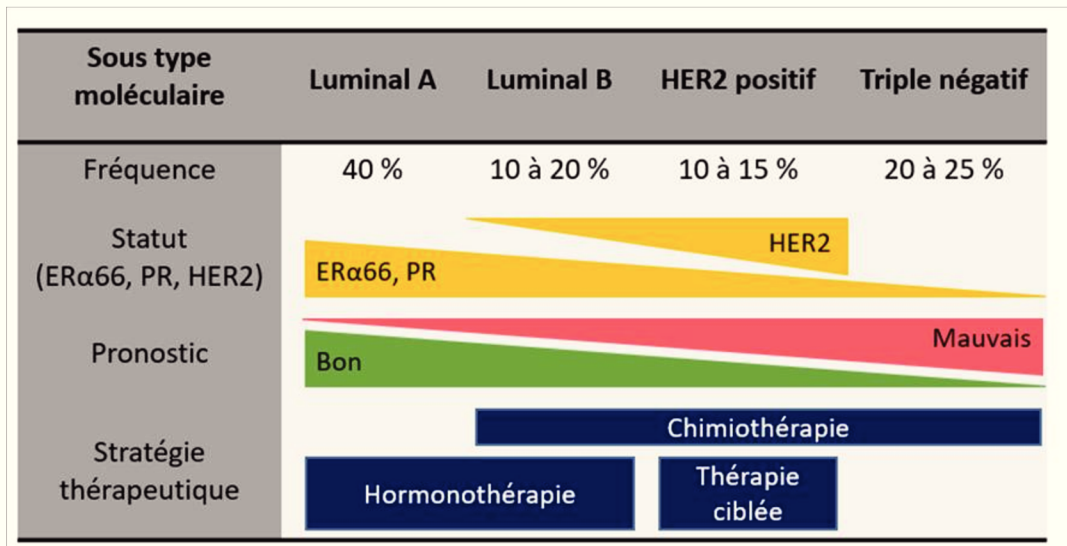
**Tableau V : Le grade histopronostique de Scarff-Bloom-Richardson (SBR).**

Critère	Note attribuée		
	1	2	3
Architecture cellulaire	La tumeur contient beaucoup de structures bien formées		La tumeur contient peu ou aucune structures bien formées
Forme du noyau	Les noyaux de la tumeur sont petits et uniformes		Les noyaux de la tumeur sont gros et leur taille et leur forme varient
Activité mitotique	Les cellules de la tumeur se divisent lentement : faible nombre de mitoses		Les cellules de la tumeur se divisent rapidement : grand nombre de mitoses

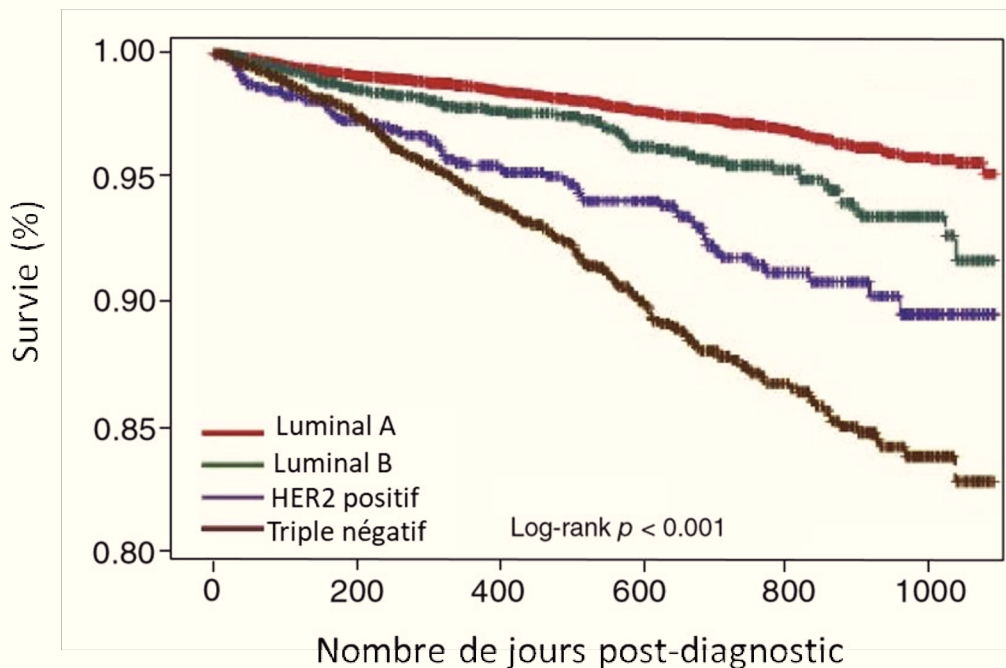
Grade	Descriptif	Score
I	Tumeurs les moins agressives	3,4 ou 5
II	Grade intermédiaire	6 ou 7
III	Tumeurs les plus agressives	8 ou 9

La détermination du grade SBR est basée sur 3 critères : l'architecture cellulaire, la forme du noyau et l'activité mitotique. L'addition des notes (1 à 3) attribuées à chaque critère permet de calculer un score qui détermine le grade (I à III) de la tumeur. La note 2 correspond à une observation intermédiaire entre les notes 1 et 3.

A



B



**Figure 28 : La classification moléculaire des tumeurs mammaires.**

- (A) La classification moléculaire des tumeurs mammaires repose sur l'expression de 3 biomarqueurs : ERα66, PR et HER2. La fréquence des différents sous-types moléculaires est indiquée. Cette classification permet principalement de déterminer le pronostic et d'orienter la stratégie thérapeutique.
- (B) Diagramme de Kaplan-Meier représentant la survie globale de patientes atteintes d'un cancer du sein en fonction du sous-type moléculaire. Etude réalisée sur une cohorte de 22 538 patientes diagnostiquées entre 2010 et 2012 (Données issues du registre du cancer de l'Ontario).

**ERα66** = Forme canonique des récepteurs aux œstrogènes α, **HER2** = Récepteur pour les facteurs de croissance épidermiques humains 2, **PR** = Récepteur à la progestérone.

(D'après Eroles et al., 2012 et Fallahpour et al., 2017)



➤ Les sous-types « Luminal A » et « Luminal B »:

Les tumeurs appartenant aux sous-types « Luminal A » ou « Luminal B » sont caractérisées par l'expression de certains marqueurs caractéristiques des cellules luminales de l'épithélium mammaire comme les cytokératines 8 et 18. Ces tumeurs sont qualifiées d'hormono-dépendantes car elles expriment les récepteurs hormonaux ER $\alpha$ 66 et PR.

Les tumeurs de type « Luminal A » qui représentent 40% des cancers du sein sont associées à un faible taux de prolifération et à un bon pronostic à 10 ans post-diagnostic. Ces tumeurs n'expriment pas le récepteur HER2, contrairement à certaines tumeurs de type « Luminal B » qui sont associées à une amplification du gène codant HER2. Ce sous-type « Luminal B » qui représente 10 à 20% des cancers du sein est associé à un grade SBR élevé et son pronostic est moins favorable que celui des tumeurs de type « Luminal A ».

➤ Le sous-type « HER2-positif »

Les tumeurs HER2 positives représentent 10 à 15% des cancers du sein et n'expriment ni les marqueurs de cellules luminales, ni les récepteurs hormonaux ER $\alpha$ 66 et PR. En revanche, elles sont caractérisées par l'amplification du gène codant le récepteur HER2 sur le chromosome 17 et par une forte expression d'un cluster de gènes impliqués dans l'induction de la prolifération cellulaire. Le sous-type « HER2-positif » est associé à un stade TNM élevé et donc à un mauvais pronostic (Fallahpour et al., 2017).

➤ Les sous-types « Basal-like » et « Claudin-low »

Les tumeurs mammaires de type « Basal-like » et « Claudin-low » appartiennent au groupe des tumeurs triples négatives qui n'expriment ni ER $\alpha$ 66, ni PR, ni HER2. Le sous-type « Basal-like » concerne 10 à 20% des cancers du sein et se caractérise par l'expression de marqueurs spécifiques aux cellules basales myoépithéliales de la glande mammaire comme les cytokératines 5 et 14. Concernant le sous-type « Claudin-low », il représente environ 10% des tumeurs mammaires et se distingue par la perte d'expression des protéines impliquées dans les adhérences intercellulaires comme les claudines 3/4/7, la E-cadhérine et l'occludine.

De manière générale, les tumeurs triples négatives sont très agressives et associées à un mauvais pronostic. En effet, elles présentent une signature d'expression génique caractéristique des cellules souches cancéreuses mammaires CD44+/CD24-/low et elles expriment fortement les gènes associés au processus de transition épithélio-mésenchymateuse.

Enfin, les tumeurs basal-like se développent généralement avant l'âge de 40 ans et sont souvent associées aux mutations BRCA1 et BRCA2.

➤ Le sous-type « Normal-like »

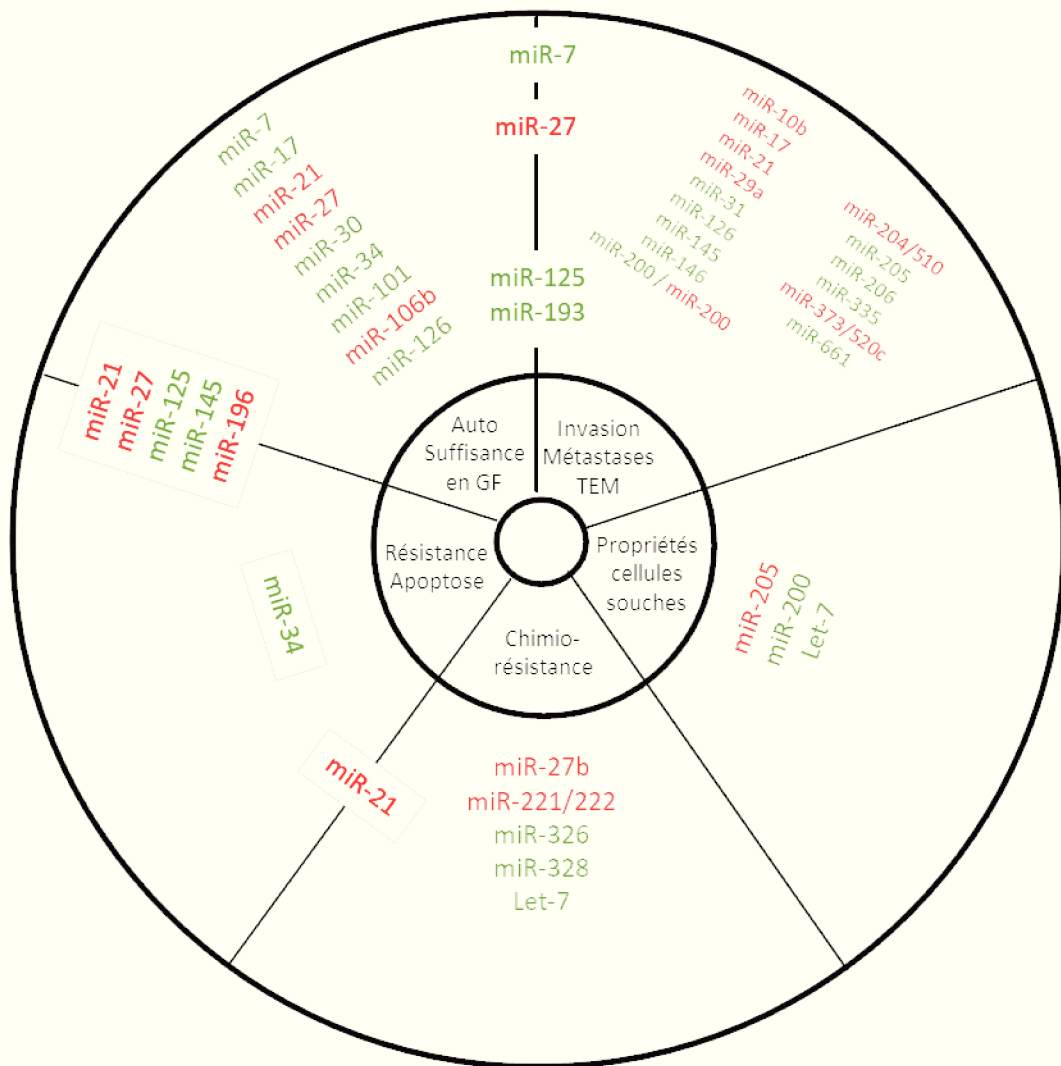
Le sous-type « Normal-like » présente des caractéristiques proches du sous-type « Luminal A » mais son pronostic est moins favorable. Toutefois, l'existence de ce sous-type est controversée. En effet, le profil d'expression génique des tumeurs de type « Normal-like » étant très proche de celui du tissu adipeux, il est possible que ce sous-type corresponde à un artefact résultant d'une contamination de l'échantillon tumoral utilisé pour l'analyse transcriptomique par du tissu mammaire sain (Perou et al., 2000).

*iv) Développement de nouvelles classifications grâce à des outils dérivés de l'intelligence artificielle*

La biologie computationnelle consiste à intégrer dans une démarche commune différents domaines d'expertise tels que la biologie, l'informatique et les mathématiques. L'avènement des techniques à haut débit, comme le séquençage de nouvelle génération (NGS) ou les études d'association pangénomique (GWAS), a permis de générer une masse d'informations biologiques hétérogènes importante. Afin de permettre l'analyse de ces informations, l'ensemble des données « -omiques » (génomique, transcriptomique ou protéomiques) sont stockées dans différents types de bases de données, pour la plupart accessibles en libre accès. (Low et al., 2018).

Dans le domaine du cancer, de nombreuses bases de données collaboratives permettent d'accéder à un descriptif détaillé des propriétés de différents types ou sous-types de tumeurs. On peut par exemple citer les bases de données CGP (Cancer Genome Project), ICGC (International Cancer Genome Consortium) et TCGA (The Cancer Genome Atlas) dans lesquelles on retrouve des informations génomiques, épigénétiques, transcriptomiques ou encore protéomiques (Pavlopoulou et al., 2015). Il existe également des bases de données plus spécifiques comme la base BCDB (Breast Cancer DataBase) qui regroupe uniquement des informations concernant les tumeurs mammaires (Mohandass et al., 2010).

A partir de la myriade de données disponibles, de nombreuses équipes cherchent à développer des méthodes génériques dérivées de l'intelligence artificielle afin d'identifier de nouvelles signatures moléculaires et de nouveaux biomarqueurs permettant d'affiner le diagnostic, d'orienter la stratégie thérapeutique et de suivre l'évolution des patients. Dans le contexte du



**Figure 29 : Micro-ARNs associés à différentes caractéristiques phénotypiques des cellules cancéreuses mammaires.**

Liste des micro-ARNs associés à l'autosuffisance en facteurs de croissance, à l'invasion/métastase/transition épithélio-mésenchymateuse, à l'acquisition de propriétés caractéristiques des cellules souches, à la chimio-résistance et à la résistance à l'apoptose des cellules cancéreuses mammaires.

Les micro-ARNs oncogéniques sont indiqués en rouge tandis que les micro-ARNs suppresseurs de tumeurs sont indiqués en vert.

**GF** = Growth Factor (Facteurs de croissance), **TEM** = Transition Epithélio-Mésenchymateuse.

(D'après Le Quesne et Caldas, 2010)

cancer du sein, 3 signatures moléculaires ont été identifiées. Il s'agit des signatures Prosigna<sup>®</sup>, Oncotype-DX<sup>®</sup> et MammaPrint<sup>®</sup> qui reposent sur le profil d'expression de 55, 21 et 70 gènes respectivement. A l'heure actuelle, ces 3 méthodes sont particulièrement utilisées en clinique pour prédire le risque de métastase et estimer le bénéfice de la chimiothérapie. Elles s'adressent à des patientes présentant une tumeur hormono-dépendante (ER $\alpha$ 66+ et/ou PR+), HER2-, avec une absence ou une faible atteinte ganglionnaire (Duffy et al., 2017).

En parallèle, depuis quelques années, la recherche se concentre sur une nouvelle classe de biomarqueurs correspondant aux micro-ARNs. En effet, il a été démontré à plusieurs reprises que leur expression peut être dérégulée et associée à l'évolution d'un certain nombre de pathologies. De plus, ces micro-ARNs sont pour la plupart détectés en quantité significative dans différents fluides biologiques. Cela présente un intérêt majeur pour la mise au point de méthodes de diagnostic non invasives. Pour faciliter leur identification, une étude récente de Zeng et al., (2018) propose une méthode générique optimisée à partir d'un algorithme SPM (structural Perturbation Method) permettant d'identifier avec une fiabilité de 80% des micro-ARNs associés à différentes maladies.

Dans le contexte du cancer du sein, plusieurs études ont permis d'identifier des micro-ARNs oncogéniques qui interviennent dans la régulation de différents hallmarks du cancer comme l'autosuffisance en facteurs de croissance, la résistance à l'apoptose, la prolifération anarchique ou encore l'acquisition de propriétés migratrices et invasives (Figure 29). Les effets de ces onco-micro-ARNs peuvent être contrebalancés par l'activité de micro-ARNs suppresseurs de tumeurs qui interviennent de façon concomitante dans la régulation de ces différentes caractéristiques des cellules cancéreuses (Le Quesne et Caldas, 2010).

Afin de *s'inscrire dans* le domaine de la biologie computationnelle qui est en pleine expansion, nous avons développé au laboratoire une approche pluridisciplinaire intégrant des données *obtenues à partir d'outils biologiques et mathématiques*. Plus précisément, dans le cadre de ma thèse nous avons souhaité développer un outil permettant de caractériser les réseaux de gènes faisant intervenir le variant *ER $\alpha$ 36* dans le contexte de la glande mammaire. Les résultats obtenus sont présentés dans la partie « *Développement d'une approche bioinformatique permettant d'identifier de nouveaux facteurs impliqués dans la signalisation dépendante d'ER $\alpha$ 36* » des résultats.

### E) Les stratégies thérapeutiques dans le cancer du sein

La stratégie de prise en charge thérapeutique du cancer du sein dépend principalement des caractéristiques de la tumeur à savoir son étendue, son agressivité et son statut moléculaire. Elle dépend également de l'âge, des antécédents et de l'état général de la patiente au moment de débiter le traitement. En fonction des cas, l'objectif du protocole thérapeutique sera d'éliminer la tumeur dans sa totalité ou de ralentir sa progression tout en empêchant la formation de métastases.

De manière générale, on peut distinguer les traitements à visée locale, les traitements à visée systémique et les traitements ciblés, qui peuvent être utilisés seuls ou en association.

#### *i) Les traitements à visée locale*

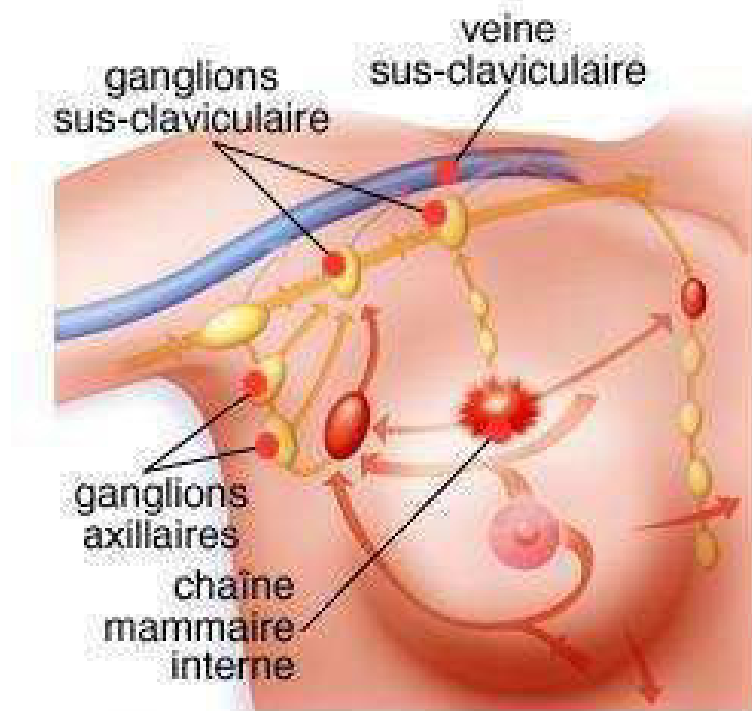
##### ➤ La chirurgie

La chirurgie est le traitement utilisé en première intention pour la prise en charge de la plupart des cancers du sein. Dans la majorité des cas, lorsque la tumeur est bien localisée et de petite taille, le chirurgien réalise une chirurgie conservatrice (tumorectomie) qui consiste à retirer uniquement la tumeur ainsi qu'une marge d'exérèse correspondant à une petite partie du tissu sain avoisinant. En revanche, lorsque la masse tumorale est trop importante ou s'il s'agit d'une tumeur infiltrante, le chirurgien doit pratiquer une chirurgie non conservatrice (mastectomie) qui correspond à l'ablation complète du sein. Ce type de chirurgie peut s'accompagner d'un curage ganglionnaire, c'est-à-dire d'une exérèse des ganglions lymphatiques axillaires. Quel que soit le type de chirurgie pratiquée, elle est souvent suivie de la mise en place de traitements adjuvants (radiothérapie, chimiothérapie et/ou hormonothérapie) qui ont pour objectif de diminuer le risque de récurrence.

Afin de privilégier la chirurgie conservatrice, un traitement néo-adjuvant peut être administré à la patiente avant l'opération afin d'essayer de réduire la taille de la tumeur.

##### ➤ La radiothérapie

La radiothérapie externe consiste à diriger précisément sur la zone à traiter des rayons ionisants afin d'endommager les cellules cancéreuses mammaires et de bloquer leur prolifération. Elle est principalement utilisée en seconde intention après une chirurgie et elle contribuerait à diminuer d'environ 70% le risque de récurrence (Clarke et al., 2005). Les rayonnements peuvent



**Figure 30 : Le drainage lymphatique du sein.**

Le drainage lymphatique de la glande mammaire implique :

- La région axillaire, qui constitue la voie de drainage principale
- La région mammaire interne
- La région sus-claviculaire

(D'après INFOCANCER, ARCAGY-GINECO)

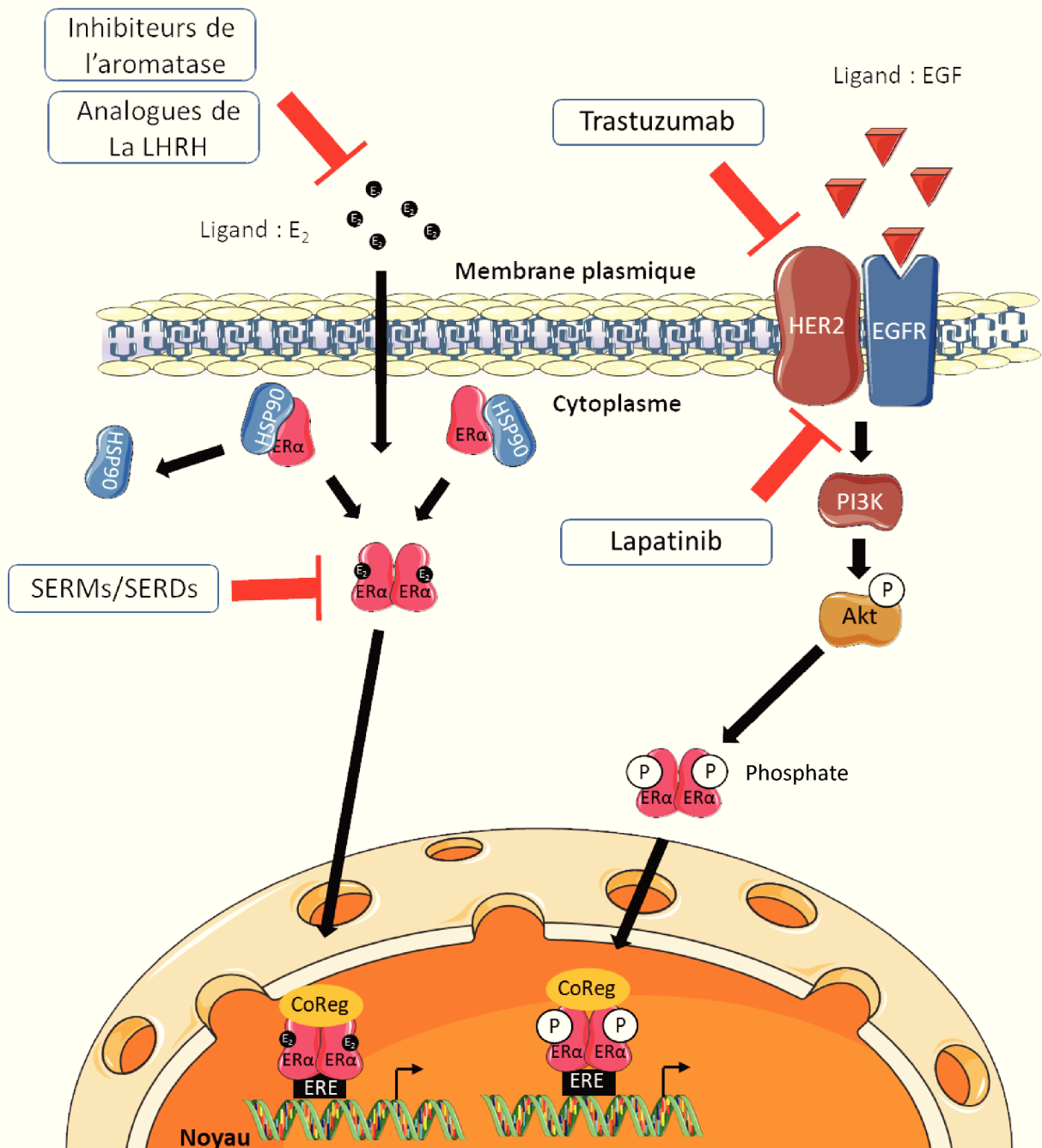
être dirigés vers le lit tumoral, c'est-à-dire vers la région de la glande mammaire où se trouvait la tumeur, afin d'éliminer les potentielles cellules cancéreuses résiduelles. Ils peuvent également être dirigés vers les ganglions sus-claviculaires ou vers la chaîne ganglionnaire mammaire interne qui peuvent contenir des cellules cancéreuses mais qui contrairement aux ganglions axillaires sont très rarement retirés lors de la chirurgie (Figure 30) (Hennequin et Fourquet, 2014). Enfin, ce type de traitement peut être utilisé pour traiter les métastases osseuses et cérébrales.

Un deuxième type de radiothérapie peut également être utilisé en complément de la radiothérapie externe pour traiter le cancer du sein. Il s'agit de la curiethérapie qui consiste à placer directement au contact de la zone à traiter une source radioactive composé d'iridium 192. Cependant, cette technique est moins utilisée car elle nécessite l'utilisation de matériel radioactif et une hospitalisation des patientes pendant 2 à 6 jours dans des chambres protégées (INCa).

#### *ii) Les traitements à visée systémique*

La chimiothérapie correspond à une méthode de traitement médicamenteuse qui agit dans l'ensemble de l'organisme et qui a pour but de détruire les cellules cancéreuses quelle que soit leur localisation. Ce type de traitement peut être administré par voie orale ou par voie intraveineuse à des fréquences variables. En effet, en fonction du type de cancer et des propriétés de la tumeur, les médecins établissent un protocole de soin qui correspond à une association de différents agents de chimiothérapie qui doivent être administrés à une fréquence et sur une durée bien définies.

De manière générale, on distingue 4 grandes familles d'agents de chimiothérapie à savoir les agents alkylants, les taxanes, les anthracyclines et les anti-métabolites, qui présentent des modes d'action différents mais qui aboutissent tous au blocage de la division cellulaire. Les effets de la chimiothérapie étant peu spécifiques, de nombreux effets secondaires sont observés : chute des cheveux, nausées et vomissements, modifications de la peau et des ongles... Cela peut même aller jusqu'à l'apparition d'une toxicité hématologique et neurologique (Gustave Roussy : <https://www.gustaveroussy.fr/fr/chimiotherapie>).



**Figure 31 : Les thérapies ciblées impactant la signalisation œstrogénique.**

Dans le contexte du cancer du sein, différents types de thérapies ciblant la signalisation œstrogénique ont été développées :

Les inhibiteurs de l'aromatase et les analogues de la LHRH permettent de bloquer la production d'œstrogènes chez les femmes non ménopausées.

Les SERMs et les SERDs régulent négativement l'activité d'ERα66.

Le Trastuzumab est un anticorps monoclonal bloquant l'activation de la signalisation dépendante du récepteur HER2. Le Lapatinib est un inhibiteur de Tyrosine Kinase qui peut bloquer l'activation d'HER2 mais également d'EGFR.

**CoReg** = Corégulateurs, **ERE** = Élément de réponse aux œstrogènes, **LHRH** = Luteinizing Hormone Releasing Hormone , **SERD** = Selective Estrogen Receptor Degradation, **SERM** = Selective Estrogen Receptor Modulators



### *iii) Les thérapies ciblées*

L'objectif de ce type de traitement est de cibler le plus spécifiquement possible les cellules cancéreuses afin d'avoir un impact limité sur les cellules saines. Le choix de la stratégie à adopter dépend principalement de la classification moléculaire de la tumeur, c'est-à-dire de son statut ER $\alpha$ 66, PR et HER2. Toutefois d'autres stratégies ciblant des acteurs de la prolifération cellulaire et de l'angiogenèse existent.

#### ➤ L'hormonothérapie

Les tumeurs dites hormono-sensibles, c'est-à-dire exprimant les récepteurs aux œstrogènes ou à la progestérone, peuvent être traitées par hormonothérapie. L'objectif de ce traitement est de bloquer l'effet des hormones féminines endogènes qui stimulent la croissance de la tumeur mammaire. Pour cela plusieurs stratégies ont été développées.

La première stratégie qui est non médicamenteuse consiste à réaliser une ablation ou une irradiation des ovaires afin d'empêcher la production des hormones féminines endogènes chez les patientes non ménopausées. L'ovariectomie peut également être réalisée de manière préventive chez les femmes présentant une prédisposition génétique BRCA1 ou BRCA2 et contribue à réduire de moitié le risque de développer un cancer du sein (Rebbeck et al., 2009). Bien que très efficaces, ces méthodes induisent une ménopause précoce chez des femmes jeunes en âge de procréer. Pour pallier à ce problème, des solutions visant à préserver la fertilité de la patiente en congelant ses ovocytes sont aujourd'hui proposées.

En parallèle, 3 stratégies d'hormonothérapie médicamenteuse ont été développées (Figure 31) :

- Les anti-œstrogènes ont pour objectif d'empêcher l'activation du récepteur aux œstrogènes ER $\alpha$ 66. On distingue les SERMs (Selective Estrogen Receptor Modulators) qui agissent comme des antagonistes de l'activité d'ER $\alpha$ 66 et les SERDs (Selective Estrogen Receptor Degradation) qui conduisent à la dégradation de ce récepteur.

Le représentant des SERMs le plus connu et le plus utilisé est le tamoxifène qui a été mis au point en 1970. On peut également citer le raloxifène et le torémifène. Ils peuvent être utilisés comme traitement adjuvant pour une durée allant de 5 ans à 10 ans, afin de prévenir une éventuelle rechute (Davies et al., 2013).

Concernant les SERDs, ils sont principalement prescrits en seconde intention lors d'une récurrence ou d'une progression de la maladie sous traitement. Le plus connu est le fulvestrant (Principe actif : ICI 182,780).

- Les inhibiteurs de l'aromatase sont des molécules inhibant l'enzyme cytochrome P450 aromatase qui catalyse la conversion des androgènes en œstrogènes. Ce type de traitement est principalement utilisé chez les femmes ménopausées pour lesquelles une petite quantité d'œstrogènes est produite au niveau des tissus périphériques (glandes corticosurrénales, foie...). Les agents les plus connus sont le létrozole, l'anastrozole et l'exémestan qui peuvent être administrés comme traitement adjuvant pendant 10 ans (Goss et al., 2016).

- Les analogues de la LHRH comme la goseréline bloquent la production d'œstrogènes en empêchant la stimulation des ovaires par l'axe hypothalamo-hypophysaire. Comme nous l'avons vu dans le Chapitre 1, la LHRH est une hormone produite par l'hypothalamus qui stimule la libération de LH et de FSH par l'hypophyse. La présence d'analogues de la LHRH dans l'organisme induit une hyperstimulation de l'hypophyse qui conduit à un rétrocontrôle négatif de la production de LH et de FSH. Les ovaires n'étant plus stimulés, la production d'œstrogènes est bloquée. Ce type de traitement peut être administré aux femmes non ménopausées sur une durée de 3 à 5 ans.

### ➤ Thérapies ciblant HER2

Depuis la fin des années 1990, différentes stratégies thérapeutiques ciblant le récepteur HER2 ont été développées. Le traitement de référence pour les tumeurs dites HER2-positives correspond au trastuzumab (Nom commercial : Herceptin®). Il s'agit d'un anticorps monoclonal humanisé qui empêche la signalisation intracellulaire dépendante d'HER2 en se liant au domaine extracellulaire de ce récepteur (Figure 31). Ce médicament est administré par voie intraveineuse et peut être couplé à de la chimiothérapie par anthracyclines et taxanes dans le cas des cancers du sein infiltrants (Merlin, 2008).

L'activité Tyrosine kinase d'HER2 peut également être inhibée au niveau intracellulaire par un agent thérapeutique appelé lapatinib (Nom commercial : Tyverb®). Ce médicament se fixe de manière compétitive sur les sites de liaisons de l'ATP au niveau du domaine intracellulaire d'HER2 bloquant ainsi la cascade de signalisation. Il est administré en association avec la

capécitabine (anti-métabolite) dans le contexte des cancers du sein métastatiques pour lesquels la poursuite d'un traitement par le trastuzumab n'est pas indiquée (Merlin, 2008).

➤ Autres types de thérapies ciblées

D'autres thérapies ciblant des acteurs impliqués dans la progression des tumeurs ont également été développées et sont fréquemment utilisées dans le contexte du cancer du sein. On peut par exemple citer le palbociclib (Nom commercial : Ibrance<sup>®</sup>), un inhibiteur des enzymes CDK4 et 6 (Cyclin Dependant Kinase) impliquées dans la régulation du cycle cellulaire et l'évérolimus (Nom commercial : Afinitor<sup>®</sup>), un inhibiteur de la protéine m-TOR (mammalian Target Of Rapamycin) qui joue un rôle clé dans l'induction de la prolifération et de la survie de certaines tumeurs (Figure 31).

Enfin, une dernière stratégie consiste à cibler le processus d'angiogenèse qui est impliqué dans la croissance mais aussi dans la dissémination métastatique des tumeurs. L'anti-angiogénique le plus connu est le bevacizumab (Nom commercial : Avastin<sup>®</sup>) qui correspond à un anticorps monoclonal ciblant le VEGF-A (Vascular Endothelial Growth Factor A). En empêchant l'interaction de ce facteur de croissance avec son récepteur, ce médicament empêche la néo-vascularisation de la tumeur qui ne reçoit alors plus suffisamment de nutriments et d'oxygène pour se développer.

# Chapitre 3 : Modulation de la signalisation œstrogénique dans la physiopathologie de la glande mammaire

*Comme l'indique la classification moléculaire, environ 70 % des cancers du sein sont liés à une dérégulation de la signalisation œstrogénique au sein de la tumeur. Dans un contexte où de nombreux cas de résistances adaptatives à l'hormonothérapie sont observés, la compréhension des mécanismes qui régulent de façon intrinsèque l'expression et l'activité des récepteurs aux œstrogènes dans la tumeur constitue un enjeu majeur pour pouvoir adapter les stratégies thérapeutiques utilisées. Dans ce chapitre, nous nous concentrerons principalement sur la régulation d'ER $\alpha$ 66 qui est l'isotype des récepteurs aux œstrogènes dont l'expression est actuellement utilisée pour le diagnostic clinique et l'orientation de la stratégie thérapeutique. Cela permettra d'exposer les différents éléments sur lesquels nous nous sommes appuyés pour étudier les modalités de régulation du variant ER $\alpha$ 36 au cours de ce travail de thèse.*

*Bien que l'implication d'ER $\beta$  dans la carcinogénèse mammaire et la résistance à l'hormonothérapie soit de plus en plus décrite (Fox et al., 2008 ; Haldosén et al., 2014) nous avons choisi d'évoquer les mécanismes de régulation d'ER $\beta$ , en lien avec ER $\alpha$ 66 et ER $\alpha$ 36, dans la partie discussion de ce manuscrit.*

## I. La régulation du métabolisme des œstrogènes

La croissance des tumeurs hormono-dépendantes (ER $\alpha$ 66+), peut être favorisée par l'induction de la synthèse des œstrogènes directement au sein de la tumeur. En effet, plusieurs études ont démontré que l'enzyme cytochrome P450 aromatase, qui convertit les androgènes en œstrogènes, est fortement exprimée dans certaines tumeurs mammaires et plus particulièrement dans les cellules adipeuses du stroma tumoral (Figure 2). De plus, via l'activation des voies de signalisation non génomiques, l'E<sub>2</sub> peut stimuler l'activité de cette enzyme en favorisant sa phosphorylation au niveau de l'acide aminé Y361. Cela se traduit par des taux de 17 $\beta$ -œstradiol significativement plus élevés dans le tissu tumoral mammaire que dans la circulation sanguine de certaines patientes (Sasano et al., 1994 ; Chetrite et al., 2000 ; Catalano et al., 2009 ; Bulun et al., 2012 ; Lipovka et Konhilas, 2016). De la même façon, l'E<sub>2</sub> peut stimuler l'activité de l'enzyme 17 $\beta$ -HSD qui est très exprimée notamment dans les carcinomes canauxaires et qui

permet la conversion de l'E<sub>1</sub> (œstrone) en E<sub>2</sub> (Figure 2). Cela conduit à une augmentation locale de la concentration en E<sub>2</sub> qui favorise la croissance tumorale (Shehu et al., 2011). Enfin, la présence de TNF $\alpha$  dans le microenvironnement tumoral peut également conduire à un rétrocontrôle positif de la voie de biosynthèse des œstrogènes. Cette cytokine peut notamment être sécrétée par les tumeurs mammaires ER $\alpha$ 66+ en réponse à l'E<sub>2</sub> ce qui amplifie encore le rétrocontrôle positif (To et al., 2014).

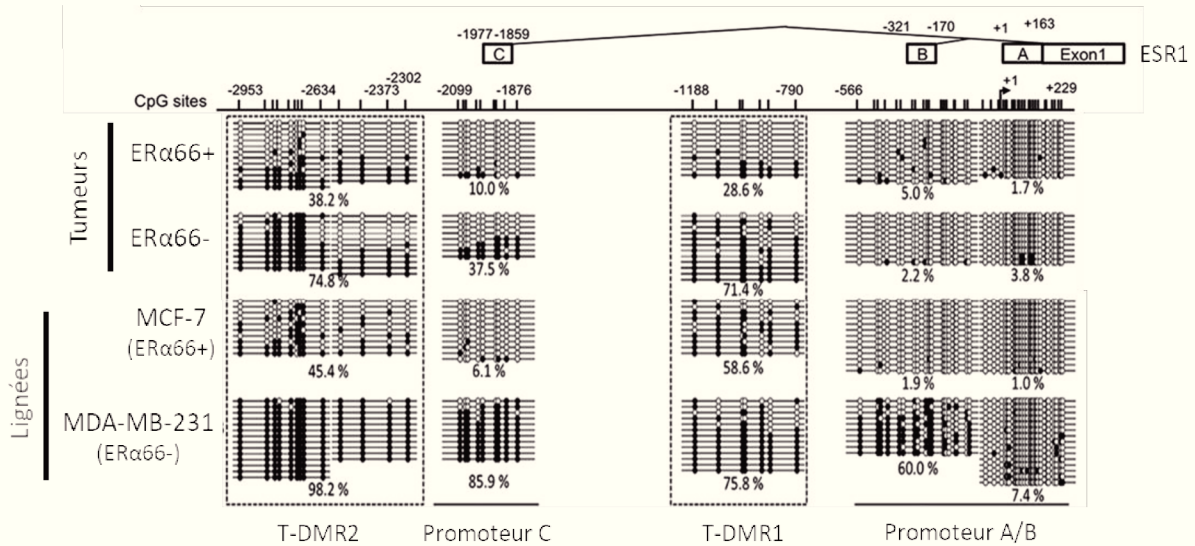
Cette augmentation du taux d'œstrogènes intra-tumoral conduit à l'accumulation de dérivés issus du métabolisme des œstrogènes qui jouent un rôle dans la carcinogénèse mammaire. En effet, lorsque les systèmes de détoxification (COMT, sulfotransférases, UDP-glucuronosyltransférases) sont dépassés, les 4-hydroxy cathécols œstrogènes sont convertis en semi-quinones puis en quinones responsables de la formation d'adduits à l'ADN et de sites apuriniques. De plus, la formation des quinones s'accompagne de la production d'espèces réactives de l'oxygène, comme l'anion superoxyde (O<sub>2</sub><sup>-</sup>), qui peuvent endommager l'ADN, les lipides et les protéines. Ce type de dommages contribue au développement et à la progression de la tumeur. Toutefois, d'autres dérivés comme le 2-hydroxyestrone, le 2-hydroxyestradiol et le 2-méthoxyestradiol jouent un rôle protecteur dans le contexte de la tumorigénèse mammaire en inhibant la prolifération et en induisant l'apoptose des cellules cancéreuses (Figure 3) (Lakhani et al., 2003 ; Manavathi et al., 2013).

## II. La régulation de l'expression d'ER $\alpha$ 66

Dans le contexte du cancer du sein, l'*expression des ERs* peut être régulée à différents niveaux. En effet, des mécanismes de régulation épigénétiques, transcriptionnels et post-transcriptionnels ont été décrits dans la littérature.

### A) Amplification du gène ESR1

L'augmentation du nombre de copies d'un gène est un mécanisme fréquemment observé dans les cellules cancéreuses, qui conduit à une surexpression du gène amplifié. Depuis quelques années, différentes études ont été menées pour déterminer si le gène ESR1 est amplifié dans certaines tumeurs mammaires et si cette amplification est associée au statut ER $\alpha$ 66+ de la tumeur. Cependant, les résultats obtenus sont très hétérogènes puisque le taux d'amplification du gène ESR1 dans les tumeurs mammaires varie de 0 à 75% en fonction des études (Holst, 2016).



**Figure 32 : Statut de méthylation de la région promotrice du gène ESR1.**

Dans cette étude, le statut de méthylation de 4 îlots CpG a été analysé, par séquençage au bisulfite, dans des tumeurs ou des lignées cancéreuses mammaires exprimant ou non ERα66 (ERα66+ ou ERα66-) :

- « Promoteur A/B » : -556 à +229 pb (paires de bases)
- « T-DMR1 » : -1188 à -790 pb
- « Promoteur C » : -2099 à -1876 pb
- « T-DMR2 » : -2953 à -2302 pb

Les sites CpG méthylés sont représentés par un rond noir tandis que les sites CpG déméthylés sont représentés par un rond blanc. Les pourcentages indiqués correspondent à la proportion de sites CpG méthylés dans la région étudiée. La position +1 correspond au site d'initiation de la transcription (TSS) du promoteur alternatif A.

Pour chaque condition, plusieurs réplicats biologiques sont présentés.

**DMR** = Differentially Methylated Region (Région différenciellement méthylée).

(D'après Maekawa et al., 2015)

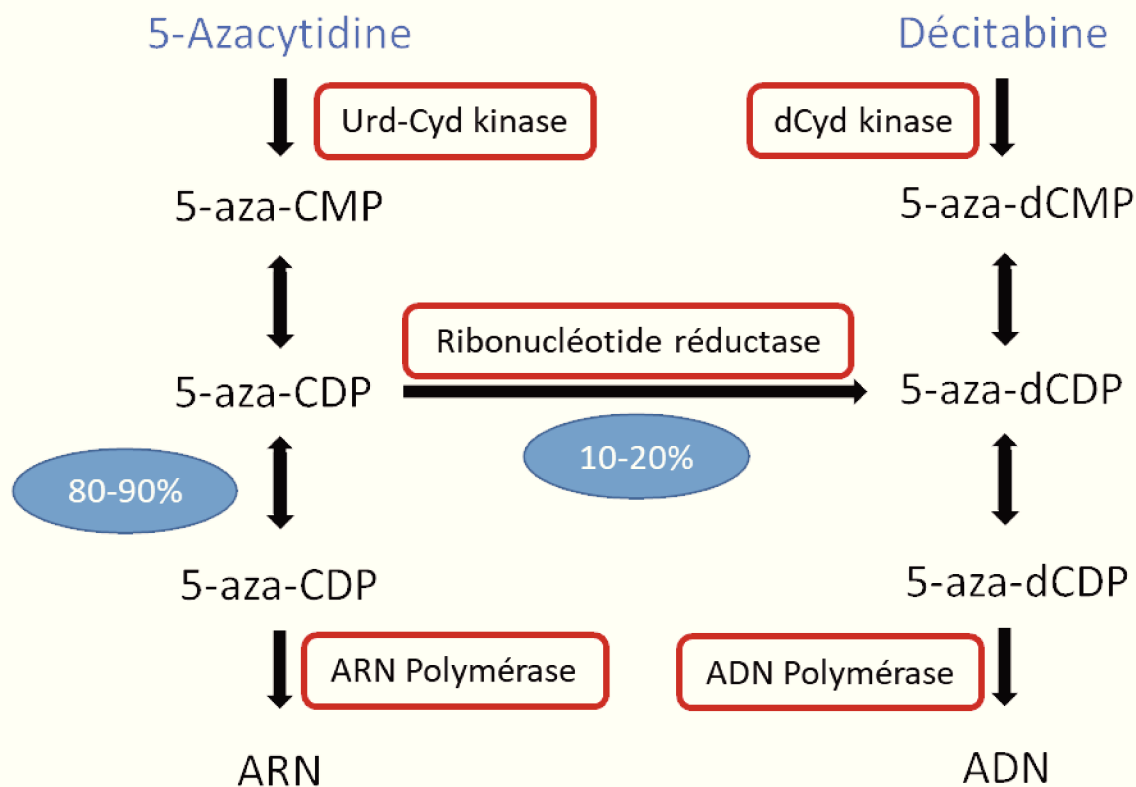
Récemment Nagel et al., (2019) ont montré que l'hétérogénéité des résultats obtenus pourrait être liée à la technique utilisée. En effet, dans cette étude, le statut d'amplification du gène ESR1 déterminé par qPCR n'est pas corrélé à celui obtenu par FISH. Cela peut s'expliquer par le fait que la qPCR permet de détecter des altérations plus fines du nombre de copies et des régions amplifiées du gène ESR1 par rapport à la technique FISH qui est basée sur l'utilisation d'une sonde unique de 395 000 bp. De plus, le lien entre cette amplification et le statut ER $\alpha$ 66 de la tumeur est difficile à établir car le taux d'amplification mesuré par qPCR est associé à une diminution de l'expression d'ER $\alpha$ 66 tandis que celui mesuré par FISH est associé à une augmentation de l'expression d'ER $\alpha$ 66.

A l'heure actuelle il est donc difficile de conclure sur le lien entre l'amplification du gène ESR1 et l'expression d'ER $\alpha$ 66 dans les tumeurs mammaires.

### B) Mécanismes de régulation épigénétique de l'expression d'ER $\alpha$ 66

Le terme « épigénétique » désigne l'ensemble des mécanismes qui permettent de moduler de façon réversible l'activité des gènes, sans modifier leur séquence ADN. Ces marques épigénétiques, qui sont transmissibles au cours des divisions cellulaires, peuvent correspondre à une *méthylation de l'ADN au niveau des sites CpG* ou à une modification post-traductionnelle des histones (*acétylation, méthylation...*). Depuis plusieurs années, il a été démontré qu'une dérégulation de ces mécanismes peut jouer un rôle déterminant dans l'initiation et la progression des cancers (Sharma et al., 2010 ; Chatterjee et al., 2018 ; Nebbioso et al., 2018).

La région promotrice du gène ESR1 comporte plusieurs îlots CpG qui participent à la régulation des différents promoteurs alternatifs. Une étude de Maekawa et al., (2016), s'est intéressée à la régulation des promoteurs alternatifs A, B et C, qui sont les principaux promoteurs activés dans les tissus présentant une forte expression du gène ESR1. Plus particulièrement, ils ont analysé le statut de méthylation de 4 régions riches en sites CpG, représentées sur la Figure 32. Les résultats de cette étude, indiquent que c'est principalement le statut de méthylation des régions « Promoteur C », « T-DMR 1 » et « T-DMR 2 » qui varie entre les tumeurs mammaires ER $\alpha$ 66+ et ER $\alpha$ 66-. En effet, ils ont mis en évidence que ces régions sont hyperméthylées dans les tumeurs ER $\alpha$ 66- étudiées tandis qu'elles sont hypométhylées dans les tumeurs ER $\alpha$ 66+. Bien que le même type de résultats soit observé entre les lignées de cellules cancéreuses mammaires MCF-7 (ER $\alpha$ 66+) et MDA-MB-231 (ER $\alpha$ 66-), il est à noter que dans certains cas, ces régions sont hypométhylées dans des tumeurs mammaires ER $\alpha$ 66-. Cela suggère donc que les mécanismes épigénétiques ne régulent pas, à eux seuls, l'expression d'ER $\alpha$ 66.



**Figure 33 : Métabolisme intracellulaire de la 5-Azacytidine et de la Décitabine.**

Après absorption intracellulaire, la 5-azacytidine et la décitabine, doivent être métabolisées en nucléotides actifs pouvant être incorporés au niveau de l'ARN ou de l'ADN par l'ARN polymérase ou l'ADN polymérase, respectivement. La métabolisation correspond à 3 phosphorylations successives. Afin de pouvoir être incorporée dans l'ADN, la 5-Azacytidine être déshydroxylée par une ribonucléotide réductase. Cette étape ne concerne que 10 à 20% de la 5-Azacytidine absorbée par la cellule.

**Urd-Cyd kinase** = uridine-cytidine kinase ; **dCyd kinase** = Deoxycytidine kinase.

**5-aza-CMP** = 5-Azacytidine monophosphate ; **5-aza-CTP** = 5-Azacytidine triphosphate ; **5-aza-dCMP** = 5-Aza-2'-déoxycytidine monophosphate ; **5-aza-dCTP** = 5-Aza-2'-déoxycytidine triphosphate ; **5-aza-dCDP** = 5-Aza-2'-déoxycytidine diphosphate ; **5-aza-dCTP** = 5-Aza-2'-déoxycytidine triphosphate.

(D'après Stresemann et al.,2008)



Ainsi, à l'heure actuelle, la corrélation entre le niveau d'expression d'ER $\alpha$ 66 et le niveau de méthylation de chacun des îlots CpG du gène ESR1 n'est pas clairement établie. Seule, une étude de Tsuboi et al., (2017) a mis en évidence, dans des cellules MCF-7 résistantes à une privation en œstrogènes et dans 36 biopsies de tumeurs mammaires, que le niveau de méthylation d'un site CpG, localisé dans un îlot en amont du promoteur C, est corrélé au niveau d'expression d'ER $\alpha$ 66. Toutefois, *in vitro*, cette corrélation semble dépendante du contexte cellulaire. En effet, le statut de méthylation de ce site CpG n'est pas corrélé à la différence d'expression d'ER $\alpha$ 66 entre les lignées de cellules cancéreuses mammaires ZR-75-1 (ER $\alpha$ 66+) et MDA-MB-231 (ER $\alpha$ 66-).

La mise en place de la méthylation de l'ADN dépend notamment du facteur ZEB1 (zinc-finger E-box binding homeobox 1) qui peut s'associer à l'ADN grâce à son domaine en doigts de zinc et recruter des ADN méthyltransférases, comme DNMT1 et des histones désacétylases, comme HDAC1. La méthylation de l'ADN associée à une désacétylation des histones conduit à une condensation de la structure chromatinienne qui empêche le recrutement des facteurs nécessaires pour la transcription du gène ESR1 (Zhang J. et al., 2017 ; Hua et al., 2018).

D'un point de vue clinique, une hyperméthylation de la région promotrice du gène ESR1 est corrélée à une diminution de la survie des patientes et à une augmentation du risque de progression métastatique. De plus, ces tumeurs qui n'expriment plus le récepteur ER $\alpha$ 66 ne peuvent pas être traitées par hormonothérapie. Dans ce contexte, l'objectif est de développer des stratégies thérapeutiques permettant de réinduire l'expression d'ER $\alpha$ 66 en ciblant les mécanismes de modifications épigénétiques (Tsuboi et al., 2017 ; Kirn et al., 2018). Parmi les différentes molécules développées, les analogues de la cytidine, à savoir la 5-Azacytidine et la 2'-déoxy-5-azacytidine (Décitabine), sont les plus prometteuses et ont déjà obtenues une autorisation de mise sur le marché par la FDA pour le traitement des syndromes myélodysplasiques. Concernant leur mécanisme d'action, ces analogues qui sont incorporés dans l'ADN au cours de la réplication, inhibent l'activité des DNMTs et conduisent à leur dégradation. Contrairement à la décitabine qui peut être incorporée directement dans l'ADN, la 5-Aza doit être déshydroxylée par une ribonucléotide réductase pour être prise en charge par la machinerie de réplication de l'ADN (Figure 33). Cette réaction est limitée puisqu'on estime que 80 à 90% de la 5-Aza est incorporée dans l'ARN (Stresemann et Lyko, 2008). Dans le contexte des tumeurs mammaires, il a été démontré que l'utilisation en combinaison d'un agent déméthylant et d'un inhibiteur de HDAC permet de réinduire, *in vitro*, l'expression d'ER $\alpha$ 66 dans des lignées de cellules cancéreuses mammaires ER $\alpha$ 66-. C'est par exemple le cas des

**Tableau VI : Régulateurs transcriptionnels du gène ESR1.**

Facteurs de transcription	Promoteur(s) du gène ESR1 associés	Corrélation avec ER $\alpha$ 66 dans les tumeurs mammaires	Pronostic associé dans les tumeurs mammaires
ERF-1/AP2 $\gamma$	Promoteur A (2 sites : entre +130 pb et +201 pb)	Non corrélé	Mauvais pronostic ↗ Expression de HER2
AP2 $\alpha$	Promoteur A (2 sites : entre +130 pb et +201 pb)	Corrélation positive	Bon pronostic ↘ Prolifération cellulaire
FOXO3a	Promoteur C (3 sites FHRE : entre -3174 pb et -2595 pb)	Corrélation positive faible	Bon pronostic
FOXM1	Promoteur A (3 sites FHRE : entre -385 pb et -84 pb)	Corrélation positive	Mauvais pronostic ↗ Prolifération cellulaire
ERBF-1	Promoteur C (1 site : -1904 pb à -1864 pb )	ND	ND
GATA-3	Promoteurs A, E et F (2 sites : entre -174 389 pb et -149 791 pb)	Corrélation positive	Bon pronostic Maintien de l'état différencié
BLIMP-1	Promoteur C (1 site : -2772 pb à -2754 pb)	Corrélation négative	Mauvais pronostic ↗ Migration et invasion

Chacun des facteurs de transcription, décrits dans la littérature pour leur implication dans la régulation transcriptionnelle de l'expression du gène ESR1, peut se lier à un ou plusieurs promoteurs cibles au niveau de sites de fixation dont la position est bien définie. La corrélation avec l'expression d'ER $\alpha$ 66 dans les tumeurs mammaires et le pronostic associé à l'expression de ces facteurs de transcription sont également indiqués.

ND : Non déterminé

combinaisons 5-Azacytidine/ Trichostatine A et Décitabine/Romidepsine (Pryzbylkowski et al., 2008 ; Von Roemeling et al., 2014).

### C) Mécanismes de régulation transcriptionnelle de l'expression d'ER $\alpha$ 66

La présence de promoteurs multiples au niveau du gène ESR1 permet de réguler de façon différentielle l'expression d'ER $\alpha$ 66 dans les différents tissus cibles des œstrogènes. Au niveau de la glande mammaire, la majorité des transcrits sont issues de l'activité du promoteur A. Toutefois, une utilisation différentielle des promoteurs du gène ESR1 dans le tissu mammaire sain ou cancéreux a été mise en évidence. En effet, on observe une plus forte activité des promoteurs A et C\* (\* = anciennement décrit comme étant le promoteur B) dans les lignées de cellules cancéreuses mammaires MCF-7 et T47D, que dans le tissu mammaire sain. Cela dépend notamment des facteurs de transcription qui sont exprimés au sein de la tumeur (Grandien et al., 1993 ; Amaral et al., 2009).

Dans cette partie, nous allons nous intéresser aux facteurs de transcription qui régulent de façon différentielle l'activité des différents promoteurs du gène ESR1 dans le contexte de la glande mammaire saine mais aussi cancéreuse (Tableau VI).

#### *i) Les facteurs de transcription de la famille AP2*

##### ➤ ERF-1/AP2 $\gamma$

Le facteur ERF-1 (Estrogen Receptor Factor 1) également appelé AP2 $\gamma$  appartient à la famille des facteurs de transcription AP2 qui sont surtout décrits dans la littérature pour leur intervention au cours du développement embryonnaire et de la tumorigenèse (McPherson et al., 1997). Il se caractérise par la présence d'un domaine de dimérisation et de liaison à l'ADN de type « hélice-coude-hélice » et d'un domaine de transactivation riche en proline et glutamine (Eckert et al., 2005).

Dans le tissu mammaire sain ERF-1/AP2 $\gamma$  est exprimé principalement dans les cellules myoépithéliales qui bordent les lobules et les canaux mammaires. Ce facteur est important pour le contrôle du développement physiologique post-pubertaire de la glande mammaire, notamment au moment de la gestation et de l'involution (Zhang J. et al., 2003).

Concernant le rôle et le niveau d'expression de ce facteur dans les tumeurs mammaires, les informations disponibles dans la littérature sont souvent contradictoires et controversées. Toutefois, il a été clairement établi que le gène codant ERF-1/AP2 $\gamma$  est fréquemment amplifié

dans le cancer du sein et il semble qu'une augmentation significative de son expression dans les tumeurs mammaires soit associée à un mauvais pronostic (Turner et al., 1998 ; Zhao et al., 2003).

D'un point de vue fonctionnel, on trouve au niveau du promoteur A du gène ESR1, 2 séquences consensus riches en GC qui constituent des sites de fixation pour le facteur ERF-1/AP2 $\gamma$  (DeConinck et al., 1995). D'après une étude de Schuur et al., (2001) la fixation d'ERF-1 sur ses sites induit une modification de la structure de la chromatine de cette région et conduit à l'activation de l'expression d'ER $\alpha$ 66. Cependant, les études cliniques indiquent qu'il n'y a pas de corrélation significative entre l'expression de ERF-1/AP2 $\gamma$  et le statut ER $\alpha$ 66 dans les tumeurs mammaires. En revanche il existe une corrélation positive entre ERF-1 et l'expression du récepteur HER2 qui pourrait expliquer le mauvais pronostic lié à l'expression de ce facteur dans les tumeurs (Pellikainen et al., 2004).

➤ AP2 $\alpha$

Le facteur AP2 $\alpha$  présente une similarité de séquence de 83% avec le facteur ERF-1/AP2 $\gamma$ . Cette homologie qui se concentre principalement au niveau du domaine de liaison à l'ADN suggère que AP2 $\alpha$  est également capable de se lier au niveau des éléments de réponse du promoteur A du gène ESR1 (McPherson et Weigel, 1999). De plus, contrairement à ce qui est décrit pour ERF-1/AP2 $\gamma$ , il existe dans les tumeurs mammaires une corrélation positive entre l'expression d'AP2 $\alpha$  et celle d'ER $\alpha$ 66.

Cette co-expression AP2 $\alpha$ /ER $\alpha$ 66 est associée à un bon pronostic. En effet, l'expression de ce facteur varie en fonction de l'agressivité de la tumeur. Dans le cas des carcinomes canauxaires in situ son expression est similaire à celle observée dans le tissu mammaire sain, alors qu'elle diminue de façon significative dans les carcinomes canauxaires invasifs (Friedrichs et al., 2005). Dans les tumeurs mammaires AP2 $\alpha$  semble donc agir comme un suppresseur de tumeur qui régule négativement la prolifération cellulaire notamment via l'activation de l'expression de l'inhibiteur du cycle cellulaire p21<sup>WAF1/CIP1</sup> (Wajapeyee et Somasundaram, 2003).

Ce mécanisme d'action est similaire à celui qui est observé dans le tissu mammaire sain où l'expression d'AP2 $\alpha$ , qui est principalement détectée dans le noyau des cellules épithéliales canauxaires, joue un rôle majeur dans le contrôle du développement lobulo-alvéolaire au moment de la gestation/lactation (Zhang J. et al., 2003).

*ii) Les facteurs de transcription de type Forkhead*

➤ FOXO3a

Le facteur FOXO3a appartient à la famille des facteurs de transcription de type Forkhead. D'un point de vue structural, il possède un domaine de liaison à l'ADN FKH, qui correspond à un dérivé du motif « hélice-tour-hélice » appelé « hélice ailée ». FOXO3a est également composé d'un domaine de transactivation, d'une séquence de localisation nucléaire (NLS) et d'une séquence d'export nucléaire (NES) (Obsil et Obsilova, 2008).

L'activité des facteurs de transcription de la famille FOXO dépend principalement de la voie de signalisation PI3K/Akt et joue un rôle majeur dans l'induction de l'apoptose et l'arrêt du cycle cellulaire. Lorsque la voie PI3K/Akt est activée, le facteur FOXO3a est phosphorylé par la protéine kinase Akt au niveau des résidus T32, S253 et S315. Cela conduit à l'inactivation de FOXO3a qui est séquestré par la protéine 14-3-3 dans le cytoplasme puis dégradé par le protéasome. A l'inverse, lorsque la voie PI3K/Akt n'est pas activée, le facteur FOXO3a, qui n'est pas phosphorylé par Akt, présente une localisation nucléaire et peut activer l'expression de ses gènes cibles (Zhang X. et al., 2011a).

Concernant la région promotrice du gène ESR1, FOXO3a peut se fixer sur 3 sites FHRE (Forkhead-responsive elements : S1, S2 et S4) localisés en amont du promoteur C\*. De façon intéressante, la forme active de FOXO3a est retrouvée au niveau nucléaire principalement dans les cellules cancéreuses mammaires ER $\alpha$ 66 positives (MCF-7 et ZR-75) (Guo et Sonenshein, 2004). Toutefois, il semble que FOXO3a ne soit pas le principal inducteur de l'expression d'ER $\alpha$ 66 dans ce modèle. En effet, dans une étude réalisée sur un ensemble de 16 lignées de cellules tumorales mammaires l'expression de ce facteur, qui est plutôt décrit comme un suppresseur de tumeur, est faiblement corrélée à l'expression d'ER $\alpha$ 66 (Madureira et al., 2006).

➤ FOXM1

Le facteur FOXM1 appartient également à la famille des facteurs de transcription de type Forkhead. Au niveau structural on retrouve le domaine de liaison à l'ADN FKH et le domaine de transactivation précédemment décrit pour FOXO3a. En revanche, FOXM1, qui ne possède pas les séquences NES et NLS, est composé d'un domaine répresseur (NRD) localisé à l'extrémité N-terminale. D'un point de vue fonctionnel, l'activité du facteur FOXM1 est plutôt

décrite comme oncogénique puisqu'il participe à la progression tumorale en régulant positivement le cycle cellulaire (Chakraborty et al., 2014).

Concernant la régulation de l'expression du gène ESR1, FOXM1 est capable de se lier de façon préférentielle à 3 sites FHRE (Forkhead-responsive elements : S6, S7 et S8) localisés au niveau du promoteur A. Contrairement à FOXO3a, l'expression de FOXM1 est fortement corrélée à l'expression d'ER $\alpha$ 66 dans les 16 lignées de cellules tumorales mammaires utilisées dans l'étude de Madureira et al. (2006) Cela suggère que parmi les facteurs de transcription Forkhead, FOXM1 est le principal inducteur de l'expression d'ER $\alpha$ 66 dans le contexte des cellules cancéreuses mammaires. Cependant, dans cette même étude, il a été montré que FOXO3A est également capable de se fixer sur les sites S6, S7, S8 et de former un complexe avec FOXM1. Cela suggère qu'il existe une coopération entre ces 2 facteurs pour induire l'expression d'ER $\alpha$ 66 dans les cellules cancéreuses mammaires.

### *iii) Les autres facteurs de transcription*

#### ➤ ERBF-1

En 1999, Tanimoto et al., ont identifié un élément enhancer riche en purines localisé au niveau du promoteur C\* du gène ESR1, entre les positions -1904 et -1864. Cet élément, constitue le site de fixation d'un facteur de transcription appelé ERBF-1. A l'heure actuelle, la nature de ce facteur n'est pas décrite dans la littérature mais il pourrait appartenir à la famille des facteurs de transcription oncogéniques Ets, qui ont la caractéristique de se lier spécifiquement à des éléments de réponse riches en purine. (GGAA/T).

Ce facteur est retrouvé lié à son élément enhancer dans les cellules cancéreuses mammaires ER $\alpha$ 66 positives qui présentent une forte activité du promoteur C\* (MCF-7, T47D), mais il n'est pas détectable dans celles qui présentent une forte activité du promoteur A (ZR-75-1). (Tanimoto et al., 1999 ; Yoshida et al., 2000).

#### ➤ GATA-3

Le facteur GATA-3 appartient à la famille des facteurs de transcription de type GATA. Au niveau structural il possède 2 domaines de transactivation et 2 domaines de liaison à l'ADN de type doigts de zinc qui reconnaissent la séquence consensus T/A GATA A/G (Adomas et al.,

2014). Il est surtout décrit dans la littérature pour son rôle essentiel dans la différenciation des lymphocytes T.

Dans la région promotrice du gène ESR1 on trouve 2 éléments cis-régulateurs localisés de part et d'autre du promoteur E2 qui constituent des sites de fixation pour GATA-3. La fixation de ce facteur au niveau de ces éléments enhancers, qui sont capables d'agir à de longues distances, permet d'activer les promoteurs A, E et F du gène ESR1 (Eeckhoute et al., 2007).

Physiologiquement, GATA-3 est impliqué dans le développement embryonnaire mais également post-pubertaire de la glande mammaire. Il est indispensable pour l'élongation des canaux mammaires au moment de la puberté et au début de la gestation. GATA-3 permet la différenciation des cellules progénitrices, qui composent le corps des bourgeons terminaux, en cellules luminales canalaire (Kouros-Mehr et al., 2006 ; Asselin-Labat et al., 2007).

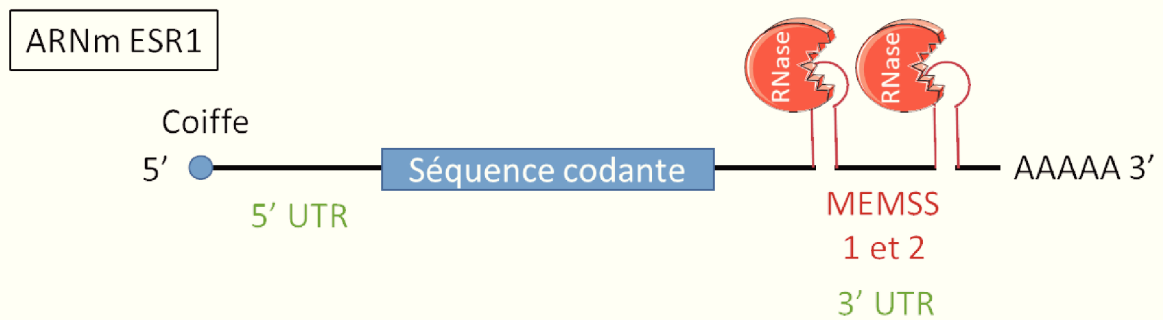
Dans les tumeurs mammaires de type « Luminal », l'expression de GATA-3, qui est positivement corrélée à l'expression d'ER $\alpha$ 66 et de PR, est associée à un bon pronostic avec une augmentation significative de la survie sans récurrence des patientes. En effet, ce facteur est impliqué dans le maintien de l'état différencié des tumeurs mammaires de bas grade qui présentent un faible potentiel métastatique (Mehra et al., 2005 ; Adomas et al., 2014 ; Guo Y. et al., 2017).

#### ➤ BLIMP1

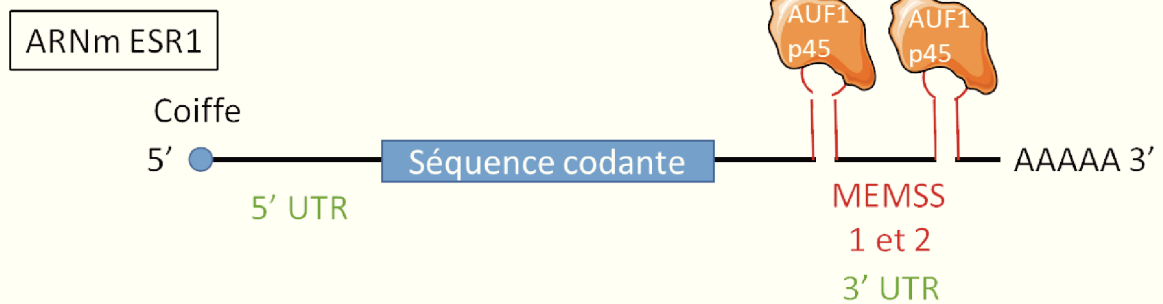
Le facteur BLIMP1, également appelé PRDM1, est un facteur de transcription qui est surtout décrit pour son rôle essentiel dans le développement des lymphocytes B et dans la spécification des cellules germinales primordiales (Chang et al., 2002). Il se compose d'une région centrale riche en proline/sérine et de 5 motifs de liaison à l'ADN de type doigts de zinc localisés en position C-terminale. De plus, BLIMP1 possède en position N-terminale, un domaine PR/SET qui permet le recrutement de facteurs de remodelage de la chromatine qui conduisent à une répression de l'expression des gènes cibles (Yu J. et al., 2000).

Des études menées à partir de biopsies de tumeurs mammaires et de différentes lignées cellulaires ont montré qu'il existe une corrélation inverse entre l'expression de BLIMP-1 et celle du récepteur ER $\alpha$ 66, dans le contexte du cancer du sein. En effet, l'expression de BLIMP-1 est significativement plus importante dans les cellules cancéreuses mammaires ER $\alpha$ 66- (Hs578T, MDA-MB-231) que dans les cellules cancéreuses mammaires ER $\alpha$ 66+ (MCF-7, ZR-75, T47D) (Wang X. *et al.*, 2009). Ces observations résultent de la présence d'un site de liaison

(A) En absence d' $E_2$



(B) En présence d' $E_2$



**Figure 34 : Stabilisation de l'ARNm d'ER $\alpha$ 66 de façon dépendante des éléments MEMSS.**

La région 3'UTR de l'ARNm codant ER $\alpha$ 66 s'étend sur environ 4306 bases et contient 2 éléments riches en « AU », nommés MEMSS. Comme cela a été prédit grâce au logiciel MFOLD, ces structures sont organisées en « tige-boucle ».

- (A) En absence d'œstrogènes ( $E_2$ ), les éléments MEMSS, qui sont découverts, sont la cible de RNases qui dégradent la région 3'UTR de l'ARNm et qui conduisent donc à une diminution de l'expression d'ER $\alpha$ 66.
- (B) En présence d'œstrogènes, le facteur AUF1p45 est recruté au niveau des éléments MEMSS qui sont alors protégés de l'action des RNases. Ainsi l'ARNm codant ER $\alpha$ 66 est stabilisé.

**MEMSS** = Minimal Estrogen Modulated Stability Sequence, 3' ou 5' **UTR** : UnTranslated Region (Région non traduite)

(D'après Menon et Goldstrohm, 2015)



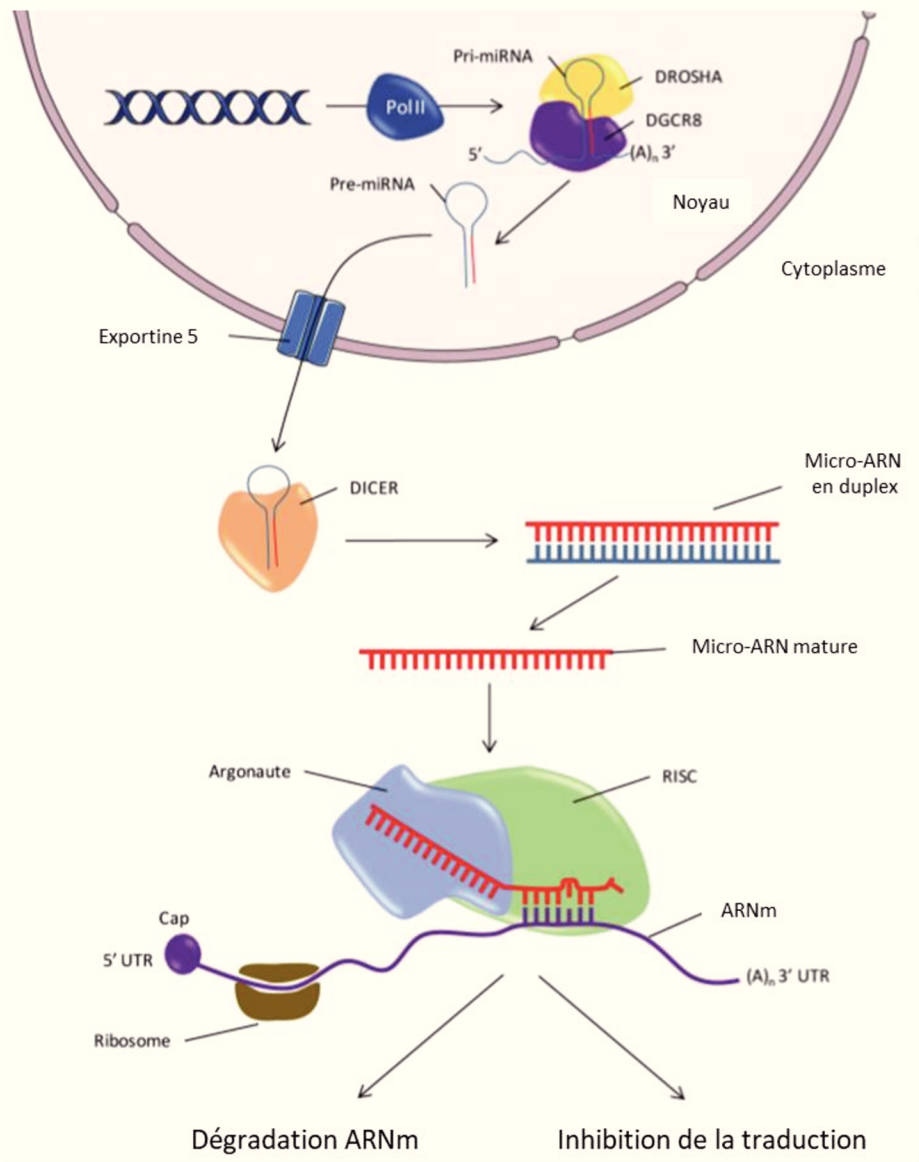
pour le facteur BLIMP1 localisé en amont du promoteur C\* du gène ESR1 et indiquent que BLIMP-1 agit comme un répresseur majeur de l'expression d'ER $\alpha$ 66. De plus, l'expression de ce facteur, qui est induite par une surexpression de la sous unité RelB de NF- $\kappa$ B dans les tumeurs mammaires ER $\alpha$ 66 négatives, est associée à une augmentation du potentiel migratoire et invasif de ces tumeurs (Wang X. et al., 2009).

#### D) Mécanismes de régulation post-transcriptionnelle de l'expression d'ER $\alpha$ 66

La transcription du gène ESR1 à partir de ces différents promoteurs conduit à la synthèse d'un ARNm d'environ 4,5 kb. Le région 3'UTR, qui est environ 3 fois plus longue que la région codante (ORF), joue un rôle essentiel dans la régulation de la stabilité de cet ARNm. En effet, elle contient de nombreux éléments riches en AU, également appelés MEMSS (Minimal Estradiol-Modulated Stability Sequences), qui permettent le recrutement de facteurs comme AUF1p45 et HuR qui stabilisent l'ARNm en le protégeant de l'action des RNases (Ing et al., 2008 ; Pryzbylkowski et al., 2008). Peu d'informations sont disponibles dans la littérature concernant le lien entre la stabilisation de l'ARNm d'ER $\alpha$ 66 et le cancer du sein. Toutefois, l'interaction entre AUF1p45 et l'ARNm d'ER $\alpha$ 66, qui est induite par les œstrogènes eux-mêmes, pourrait participer au maintien de la stabilité de l'expression protéique d'ER $\alpha$ 66 dans les tumeurs mammaires (Figure 34) (Menon et Goldstrohm, 2015).

En parallèle, la région 3'UTR est la principale cible d'un autre mécanisme de régulation post-transcriptionnelle correspondant à l'action des micro-ARNs. Cette nouvelle classe de régulateurs a été mise en évidence en 1993 chez *Caenorhabditis elegans* et on compte aujourd'hui plus de 2000 micro-ARNs décrits chez l'Homme (Sempere, 2019). On estime, que ces micro-ARNs réguleraient 60% des gènes codant des protéines chez l'Homme (Friedman et al., 2009).

Ces micro-ARNs sont codés par des gènes localisés dans des unités transcriptionnelles qui leur sont propres (régions intergéniques) ou dans les introns des gènes codant des protéines. La plupart du temps, ces gènes sont groupés en clusters codant plusieurs micro-ARNs différents. Ils sont produits via l'activité de l'ARN Polymérase II sous la forme d'un précurseur long, pri-miRNA, qui peut contenir la séquence d'un ou plusieurs micro-ARNs. La forme mature simple brin qui se compose de 18 à 25 nucléotides est obtenue après 2 étapes de maturation faisant intervenir les enzymes Drosha, au niveau nucléaire, et Dicer, au niveau cytoplasmique. Au terme de sa biosynthèse, le micro-ARN mature est pris en charge par une protéine de la famille Argonaute pour former la particule RISC (Figure 35)(Derghal et al., 2016).

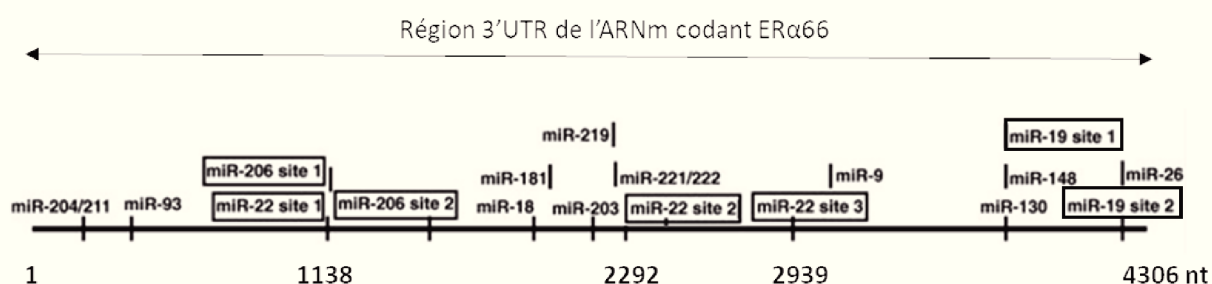


**Figure 35 : La biosynthèse des micro-ARNs.**

Cette voie de biosynthèse débute par la transcription, dépendante de l'ARN polymérase II (Pol II), des gènes codant les micro-ARNs. Le transcrit primaire produit (pri-miRNA) est ensuite pris en charge dans le noyau par le complexe DROSHA (RNase)/DGCR8 pour subir une première étape de clivage qui aboutit à la formation d'un précurseur appelé pré-miRNA. Celui-ci est ensuite exporté vers le cytoplasme, par l'intermédiaire de l'exportine 5, où il va subir une deuxième étape de maturation correspondant au clivage de la boucle terminale par DICER (RNase). Un duplex double brin est ainsi formé. Le micro-ARN mature est ensuite pris en charge par la protéine Argonaute et incorporé dans le complexe RISC, tandis que son complémentaire est dégradé. Il est admis que si la complémentarité entre le micro-ARN et sa cible (ARNm) est parfaite, l'ARNm sera dégradé tandis que si la complémentarité est imparfaite, une répression de la traduction de l'ARNm cible sera observée.

**Cap** = coiffe, **DGCR8** = DiGeorge Syndrome Critical Region 8, **RISC** = RNA-Induced Silencing Complex.

(D'après Derghal et al., 2016)



**Figure 36 : Micro-ARNs ciblant la région 3'UTR de l'ARNm codant ER $\alpha$ 66.**

Les barres verticales indiquent les sites de fixation potentiels des différents micro-ARNs susceptibles de cibler la région 3'UTR de l'ARNm codant ER $\alpha$ 66, qui s'étend sur 4306 nucléotides. Les micro-ARNs pour lesquels plusieurs sites de fixation potentiels ont été identifiés sont encadrés.

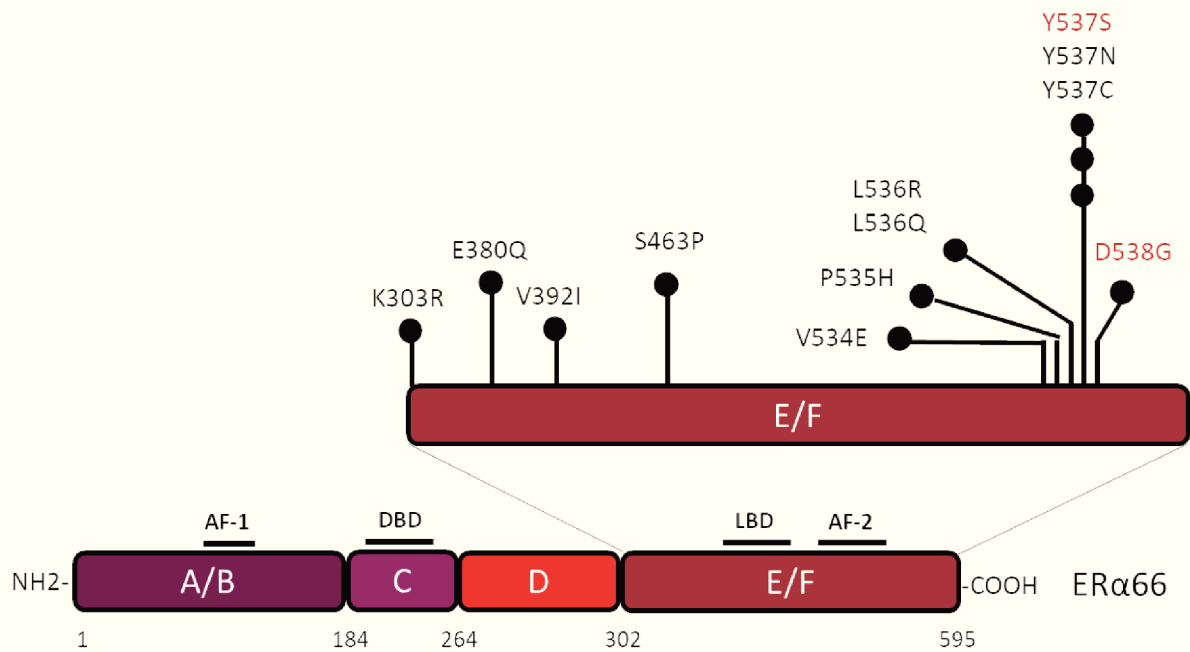
(D'après Pandey et al., 2009)

D'un point de vue fonctionnel, la séquence du micro-ARN mature permet de guider la particule RISC jusqu'à l'ARNm cible pour induire sa dégradation ou bloquer sa traduction. La complémentarité des bases, entre la séquence du micro-ARN et celle de la région 3'UTR de l'ARNm, peut être parfaite ou imparfaite. En effet, bien que l'extrémité 5' du micro-ARN, appelée région « seed », doit être parfaitement complémentaire à la séquence de l'ARNm, l'appariement de l'extrémité 3' peut être complet ou incomplet. Cela a une influence directe sur le mécanisme de répression mis en jeu. Dans le cas d'un appariement imparfait, la traduction de l'ARNm est réprimée tandis que dans le cas d'un appariement parfait l'ARNm cible est dégradé par un clivage endonucléolytique catalysé par la protéine Argonaute 2. De façon plus sporadique, quelques études ont montré que les micro-ARNs peuvent s'apparier à la région 5'UTR ou à la région codante de l'ARNm (Miranda et al., 2006 ; Tay et al., 2008).

Comme nous l'avons évoqué précédemment, l'étude des micro-ARNs associés à la progression des tumeurs mammaires est en pleine expansion (Tan et al., 2018). Plusieurs de ces micro-ARNs ont été décrits comme ciblant directement la région 3'UTR de l'ARNm codant ER $\alpha$ 66 (Figure 36). C'est notamment le cas des micro-ARNs miR-206, miR-22 ou encore miR-130a dont l'expression est inversement corrélée à celle d'ER $\alpha$ 66 dans les tumeurs mammaires (Adams et al., 2007 ; Castellano et al., 2009 ; Pandey et Picard, 2009). Ces micro-ARNs, qui ont souvent plusieurs cibles différentes, peuvent avoir un impact majeur sur l'évolution de la tumeur. C'est par exemple le cas de miR-22 qui en ciblant les ARNm d'ER $\alpha$ 66, du transporteur de glucose GLUT1 et de la désacétylase SIRT1, contribue à inhiber la prolifération et l'invasion des cellules cancéreuses mammaires (Chen B. et al., 2015 ; Zhang X. *et al.*, 2017).

En parallèle, la régulation de l'expression d'ER $\alpha$ 66 par les micro-ARNs peut être indirecte. En effet, ils peuvent cibler des ARNm codant des régulateurs de l'expression d'ER $\alpha$ 66 comme FOXM1 ou FOXO3a qui sont régulés négativement par miR-802 et miR-96, respectivement (Lin H. et al., 2010 ; Yuan et Wang, 2015).

Un autre point important concerne la régulation de l'activité de ces micro-ARNs. Premièrement, ils peuvent être séquestrés par des longs ARNs non codants (lncRNA) ou par des ARN circulaires (CircRNA) agissant comme des « éponges » qui empêchent l'interaction de ces micro-ARNs avec les ARNm cibles (Baulande et al., 2014 ; Panda, 2018). L'expression de ces éponges est donc à prendre en compte car elle peut impacter directement l'activité des micro-ARNs. On peut notamment citer l'exemple du lncRNA HOTAIR (HOX Transcript Antisense RNA) qui est décrit dans le contexte des tumeurs mammaires comme capable de séquestrer miR-206 (Ding et al., 2017). De plus, l'activité de la région promotrice des gènes codant les



**Figure 37 : Les mutations activatrices affectant le domaine E/F d'ERα66.**

Les différentes mutations activatrices touchant le domaine E/F d'ERα66, décrites dans la littérature, sont représentées. La position et la nature de la substitution sont indiquées (code à 1 lettre des acides aminés).

Les mutations Y537S et D538G, qui sont les plus fréquentes, sont indiquées en rouge.

(D'après Jeselsohn et al., 2015)

pri-miRNA peut être régulée par des mécanismes épigénétiques et notamment par la méthylation de l'ADN. Toutefois, la localisation de la région promotrice de ces gènes, qui peut être située plusieurs kilobases en amont du TSS, et la corrélation entre le niveau de méthylation de différents îlots CpG et l'expression d'un pri-miRNA d'intérêt, sont difficiles à établir expérimentalement. C'est pourquoi différents outils bioinformatiques, comme le « ProMIRNAtool » (<http://promirna.molgen.mpg.de>), commencent à se développer afin de faciliter l'étude de la régulation épigénétique des micro-ARNs (Morales et al., 2017). Plus précisément, cet outil s'appuie sur les données de la littérature correspondant à la cartographie des sites d'initiation de la transcription présents dans le génome, afin de prédire quelles sont les régions promotrices qui sont le plus susceptibles de réguler l'expression d'un gène codant un micro-ARN d'intérêt. Une fois ces séquences identifiées, on peut expérimentalement analyser leur statut de méthylation.

### III. La Régulation de l'activité d'ER $\alpha$ 66

*Les œstrogènes participent à la régulation d'un grand nombre de processus physiologiques mais aussi pathologiques. Dans le contexte du cancer du sein, on observe une dualité d'action des œstrogènes qui peuvent favoriser la progression tumorale ou au contraire avoir un effet suppresseur de tumeur.*

#### A) Les mutations activatrices ciblant ER $\alpha$ 66

Plusieurs études cliniques menées en parallèle ont permis de mettre en évidence l'existence de mutations du gène ESR1 dans plus de 20% des tumeurs mammaires ER $\alpha$ 66+ métastatiques, ayant préalablement été traitées par hormonothérapie. L'émergence de ces mutations résulterait de la sélection de cellules tumorales qui expriment une forme mutée d'ER $\alpha$ 66 leur permettant de résister aux traitements. Pour comparaison, la prévalence de ces mutations est seulement de 2% dans les tumeurs mammaires primaires (Jeselsohn et al., 2015, 2018)

Dans 80% des cas, ces mutations faux-sens touchant le gène ESR1 conduisent à la substitution d'un acide aminé par un autre au niveau du domaine E/F d'ER $\alpha$ 66 (Figure 37). Ces mutations sont dites activatrices car elles induisent un changement conformationnel qui est associé à une activité constitutive de ce récepteur. Cela permet de stimuler la croissance tumorale de façon ligand-indépendante et de conférer une résistance aux anti-aromatases. Une résistance partielle au tamoxifène, qui doit être utilisé à plus haute dose pour bloquer l'activité du récepteur muté, est également décrite (Jeselsohn et al., 2015). En particulier, la mutation D538G, qui est

associée à une augmentation des capacités migratoires des cellules tumorales qui l'expriment, contribue au développement de la résistance aux anti-œstrogènes (Robinson D.R. et al., 2013).

En parallèle, une autre mutation localisée à la jonction entre le LBD et le domaine charnière a été mise en évidence : K303R. Celle-ci pourrait contribuer au développement d'une résistance au tamoxifène en favorisant la signalisation dépendante des facteurs de croissance. En effet, in vitro cette mutation est associée à une augmentation de la stabilité et de l'activation du récepteur HER2 dans des cellules MCF-7 exprimant la forme K303R d'ER $\alpha$ 66 (Giordano et al., 2010).

Grâce aux progrès des techniques de séquençage, il est aujourd'hui possible de détecter les mutations du gène ESR1 au niveau de l'ADN tumoral circulant qui est obtenu par une simple prise de sang. Actuellement, différentes études cliniques sont en cours pour :

- Déterminer la prévalence et l'impact clinique (survie des patientes) des différentes mutations du gène ESR1 de l'ADN tumoral circulant dans des tumeurs mammaires ER $\alpha$ 66+ récidivantes ou métastatiques : Etudes EROS1 (Estrogen Receptor One Study 1) (Najim et al., 2019) et BOLERO-2 (Chandarlapaty et al., 2016).

- Evaluer l'intérêt du suivi des mutations ESR1 dans l'ADN tumoral circulant pour observer en « temps réel » l'efficacité de la stratégie thérapeutique adoptée : Etudes CICLADES (CIRCuLating Dna ESR1 Gene Mutations Analysis) et PADA-1 (PALbociclib and Circulating Tumor DNA for ESR1 Mutation Detection) (<https://clinicaltrials.gov/>).

### B) Mécanismes de régulation de l'activité transcriptionnelle d'ER $\alpha$ 66

Comme nous l'avons vu précédemment, la signalisation génomique des ERs dépend du recrutement d'un certain nombre de facteurs protéiques qui contribuent à induire ou à maintenir l'interaction des ERs avec l'ADN. Dans le contexte du cancer du sein, les facteurs de transcription FOXA1, GATA-3 ou encore PBX1, dont les séquences consensus de liaison à l'ADN sont retrouvées à proximité de régions reconnues par les ERs, semblent jouer un rôle essentiel. En effet, il a été démontré que l'inhibition de leur expression dans des cellules cancéreuses mammaires est associée à une perturbation de l'interaction des ERs avec l'ADN (Carroll, 2016 ; Eeckhoute et al., 2007 ; Magnani et al., 2011). Plus précisément, FOXA1 et PBX1 sont des facteurs dits pionniers capables de se lier de façon autonome au niveau de l'hétérochromatine afin d'entraîner sa décondensation et de permettre le recrutement d'autres facteurs comme les ERs. L'expression de FOXA1 est corrélée à celle d'ER $\alpha$ 66 dans les cancers

**Tableau VII : Principales modifications post-traductionnelles ciblant ER $\alpha$ 66**

Acide aminé	Type de modification post-traductionnelle	Fonctions associées
Y52	Phosphorylation	↗ Activité transcriptionnelle
S102	Phosphorylation	↗ Activité transcriptionnelle
S104/106	Phosphorylation	↗ Activité transcriptionnelle, ↗ Dimérisation
S118	Phosphorylation	↗ Activité transcriptionnelle, ↗ Dimérisation
S154	Phosphorylation	ND
S167	Phosphorylation	↗ Activité transcriptionnelle, ↗ Stabilité
S236	Phosphorylation	Empêche la dimérisation
Y219	Phosphorylation	↗ Activité transcriptionnelle, ↗ Dimérisation, ↗ Liaison à l'ADN, ↗ Stabilité
R260	Méthylation	↗ Signalisation non génomique
K266	Acétylation	↗ Activité transcriptionnelle
	Sumoylation	↗ Activité transcriptionnelle, ↗ Liaison à l'ADN
K268	Acétylation	↗ Activité transcriptionnelle
	Sumoylation	↗ Activité transcriptionnelle, ↗ Liaison à l'ADN
S282	Phosphorylation	↘ Activité transcriptionnelle
S294	Phosphorylation	↗ Activité transcriptionnelle
K299	Acétylation	↘ Activité transcriptionnelle
	Sumoylation	↗ Activité transcriptionnelle, ↗ Liaison à l'ADN
K302	Acétylation	↘ Activité transcriptionnelle
	Sumoylation	↗ Activité transcriptionnelle, ↗ Liaison à l'ADN
	Ubiquitinylation	↗ Dégradation par le protéasome
	Méthylation	↗ Stabilité
K303	Acétylation	↘ Activité transcriptionnelle
	Sumoylation	↗ Activité transcriptionnelle, ↗ Liaison à l'ADN
	Ubiquitinylation	↗ Dégradation par le protéasome
S305	Phosphorylation	↗ Activité transcriptionnelle, ↗ résistance aux anti-aromatase
T311	Phosphorylation	Localisation nucléaire
C447	Palmitoylation	Localisation sous-membranaire
Y537	Phosphorylation	↗ Liaison du ligand, Dimérisation, Liaison à l'ADN
S559	Phosphorylation	↘ Activité transcriptionnelle

ND = Non déterminé

(D'après Le Romancer et al., 2011)



du sein de type « Luminal » et pourrait favoriser l'apparition de résistances à l'hormonothérapie et à la formation de métastases.

Concernant le facteur GATA-3, il agit comme un régulateur du complexe ER $\alpha$ 66/FOXA1. Bien que son rôle physiologique soit de réguler la morphogénèse mammaire en inhibant la prolifération et en favorisant la différenciation cellulaire, dans le contexte du cancer du sein, GATA-3 semble impliquer dans le maintien du potentiel prolifératif des cellules cancéreuses ER $\alpha$ 66+ (Kong et al., 2011 ; Carroll, 2016).

En parallèle, d'autres facteurs capables de former un complexe transcriptionnel avec les ERs ont également été décrits comme favorisant la croissance tumorale. On peut par exemple citer les facteurs Ets-1 (Kalet et al., 2013), Adora-1 (Lin Z. *et al.*, 2010) et MRTF-A. En réponse à l'E<sub>2</sub>, ce dernier facteur est associé à l'acquisition de propriétés métastatiques dans les cellules cancéreuses mammaires ER $\alpha$ 66+ (Zhang C. et al., 2013). A l'inverse d'autres facteurs associés aux ERs ont plutôt un effet suppresseur de tumeur comme par exemple PAX-2 (Paired box 2) qui inhibe l'expression du gène ERBB2 codant le récepteur HER2 (Hurtado et al., 2008).

### C) Les modifications post-traductionnelles d'ER $\alpha$ 66

De manière générale, les modifications post-traductionnelles (PTM) permettent de moduler de façon réversible et rapide l'activité d'un grand nombre de protéines. Les principales PTMs qui régulent l'activité physiologique des ERs sont la phosphorylation, la méthylation, l'acétylation, la sumoylation et enfin l'ubiquitinylation. En fonction des acides aminés modifiés, les PTMs peuvent affecter la dimérisation, la localisation, la stabilité, la liaison à l'ADN mais aussi l'activité transcriptionnelle d'ER $\alpha$ 66 (Tableau VII)(Le Romancer et al., 2011).

Certaines des PTMs qui affectent ER $\alpha$ 66 peuvent être dérégulées dans les tumeurs mammaires sans que l'on sache vraiment s'il s'agit d'une cause ou d'une conséquence de la maladie. Toutefois, l'identification et la caractérisation de l'effet de ces PTMs constituent un enjeu majeur qui pourrait conduire au développement de nouvelles stratégies thérapeutiques ciblant certaines PTMs (Le Romancer et al., 2011).

Dans le contexte du cancer du sein, 3 sites de phosphorylation d'ER $\alpha$ 66 sont particulièrement intéressants. Il s'agit des sérines S118, S167, S305 qui lorsqu'elles sont phosphorylées, favorisent l'activité transcriptionnelle physiologique d'ER $\alpha$ 66 en permettant le recrutement de corégulateurs.

**Tableau VIII : Principales modifications post-traductionnelles d'ER $\alpha$ 66 dans le contexte des tumeurs mammaires.**

Acide aminé	Type de modification post-traductionnelle	Pronostic	Réponse au Tamoxifène
S104/106	Hyperphosphorylation	Bon	ND
S118*	Hyperphosphorylation	Bon	Sensibilité
	Hypophosphorylation	Bon	Sensibilité
S167	Hyperphosphorylation	Bon	Sensibilité
R260	Hyperméthylation	ND	ND
S282	Hyperphosphorylation	Bon	ND
S305	Hyperphosphorylation	Mauvais	Résistance
T311	Hyperphosphorylation	Mauvais	ND
S559	Hyperphosphorylation	Bon	ND

Pour chaque type de modification post-traductionnelle, le pronostic et la réponse au Tamoxifène associés sont indiqués. Les acides aminés cibles de ces modifications sont désignés par le code à une lettre des acides aminés suivi de leur position dans la séquence protéique d'ER $\alpha$ 66.

\*Concernant la S118, une étude de Yamashita et al., (2005) a montré que cette modification post-traductionnelle est associée à un mauvais pronostic et n'intervient pas dans la résistance au Tamoxifène.

ND = Non déterminé

(D'après Le Romancer et al., 2011)

La phosphorylation des sérines S118 et S167 est positivement corrélée à l'activation des kinases MAPK, Akt et p90-RSK. Ces PTMs sont principalement associées à des tumeurs mammaires de bon pronostic et bien que cela soit controversé dans le cas de la sérine 118, elles pourraient constituer des marqueurs de réponse à l'hormonothérapie. En effet, différentes études suggèrent que la présence d'une S118 ou d'une S167 phosphorylée au sein de la tumeur est associée à une réponse favorable à l'hormonothérapie et plus particulièrement à une bonne efficacité du tamoxifène (Yamashita et al., 2005 ; Kok et al., 2009).

Concernant la phosphorylation de la S305, elle résulte de l'activation des kinases PKA ou PAK1 et elle est très souvent associée à la mutation activatrice K303R décrite précédemment. Contrairement au cas des S118 et S167, la phosphorylation de la S305 est plutôt associée à un mauvais pronostic et constitue un marqueur de résistance au tamoxifène. En effet, cette PTM induit un changement de conformation du récepteur ER $\alpha$ 66 qui malgré la présence du tamoxifène garde sa capacité à se lier au coactivateur Src-1. L'inhibition de l'activité des kinases phosphorylant la S305 pourrait donc constituer une stratégie thérapeutique permettant de re-sensibiliser certaines tumeurs au tamoxifène (Zwart et al., 2007 ; Giordano et al., 2010 ; Kastrati et al., 2019).

En parallèle, d'autres PTMs associées aux tumeurs mammaires ont été mises en évidence et sont présentées dans le tableau VIII (Le Romancer et al., 2011).

#### D) Mécanismes de régulation de l'activité d'ER $\alpha$ 66 par d'autres récepteurs hormonaux

La croissance des tumeurs mammaires ER $\alpha$ 66+ dépend principalement de la signalisation œstrogénique. Toutefois, d'autres hormones stéroïdiennes présentes dans le microenvironnement tumoral peuvent influencer le développement de ces tumeurs. Différentes études ont en effet montré que l'activation des récepteurs à la progestérone (PR) ou aux androgènes (AR) peut avoir un impact sur l'activité transcriptionnelle des ERs. La compréhension de ces mécanismes de régulation croisée est particulièrement importante car les récepteurs PR et AR sont exprimés respectivement dans 75% et 90% des tumeurs ER $\alpha$ 66+ (Siersbæk et al., 2018).

En présence de leur ligand, les PRs peuvent recruter des corégulateurs transcriptionnels qui interviennent également dans la signalisation génomique d'ER $\alpha$ 66. Cette compétition pour l'utilisation de facteurs communs peut conduire à une modification du profil d'expression génique de la tumeur avec une augmentation de l'expression des gènes antiprolifératifs,

caractéristiques de la signalisation des PRs, aux dépend des gènes pro-prolifératifs dont l'expression est oestrogène-dépendante. Ce constat ouvre la voie à la mise au point d'une nouvelle stratégie d'hormonothérapie qui consisterait à activer PR afin de diminuer l'impact de la signalisation œstrogénique sur la progression tumorale (Mohammed et al., 2015 ; Carroll et al., 2017).

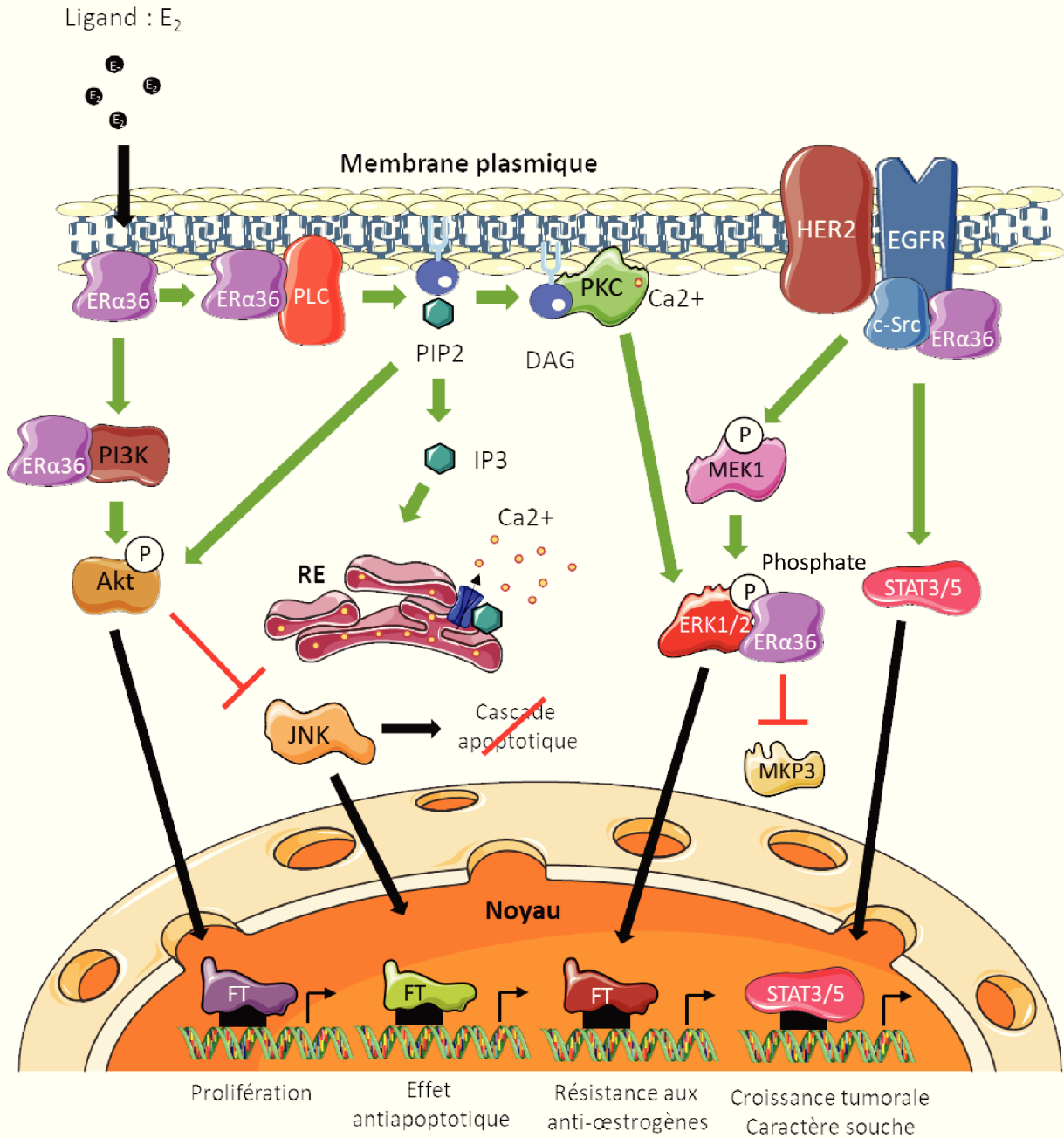
Concernant les ARs, ils peuvent impacter la signalisation œstrogénique grâce à leur capacité à se lier à ER $\alpha$ 66 et à stabiliser l'interaction ER $\alpha$ 66/ERE. Toutefois, l'impact de l'activité des ARs sur la croissance des tumeurs mammaires ER $\alpha$ 66+ est à l'heure actuelle controversée. En effet, en fonction des antagonistes de AR utilisés, les effets observés ne vont pas toujours dans le sens d'une diminution de la croissance tumorale (Peters et al., 2009 ; D'Amato et al., 2016).

#### IV. Le cas du variant ER $\alpha$ 36

*Depuis la mise en évidence d'ER $\alpha$ 36 par Wang Z. et al., en 2005, un nombre croissant d'études suggèrent que ce variant joue un rôle majeur dans la progression tumorale et dans le développement de résistances aux traitements anti-tumoraux. Cependant, son expression n'est pas prise en compte lors du diagnostic clinique des tumeurs mammaires. Au début de ma thèse, peu de données étaient disponibles dans la littérature concernant le niveau d'expression et le rôle d'ER $\alpha$ 36 dans la glande mammaire saine.*

##### A) Les données cliniques

Au laboratoire, Clémence Chamard-Jovenin a montré grâce à une étude rétrospective sur une cohorte de 118 patientes, qu'une forte expression d'ER $\alpha$ 36 dans les tumeurs mammaires est corrélée à l'expression de différents marqueurs métastatiques (Chamard-Jovenin et al., 2015). De plus, une étude clinique de Wang Q. et al., (2018) menée sur une cohorte de 1068 patientes présentant une tumeur mammaire a montré que l'expression d'ER $\alpha$ 36 est corrélée à la taille, au stade, au grade histologique, à l'atteinte ganglionnaire (risque de métastases) et au statut PR de la tumeur. Ces résultats sont différents de ceux obtenus par Shi L. et al., (2009) sur une cohorte de 710 patientes. En effet, dans cette étude antérieure l'expression d'ER $\alpha$ 36 apparaît corrélée uniquement à l'âge de la patiente et au statut HER2 et PR de la tumeur. Cela peut s'expliquer par un nombre plus important d'échantillons dans l'étude de Wang Q. et al., par une différence au niveau des critères d'inclusion des patientes, mais aussi par l'utilisation d'anticorps différents pour la détection immuno-histologique de l'expression d'ER $\alpha$ 36.



**Figure 38 : La signalisation dépendante d'ERα36.**

La signalisation correspondant à cette figure est détaillée dans le paragraphe concernant les voies de signalisation activées par ERα36 dans les tumeurs mammaires.

**DAG** = Diacylglycérol, **ERK**= Extracellular signal-Regulated Kinase, **FT** = Facteurs de transcription variables en fonction des voies de signalisation activées, **IP3** = Inositol trisphosphate, **JNK** = c-Jun N-terminal Kinase, **MEK** = MAP/ERK Kinase, **MKP3** = MAP Kinase Phosphatase 3, **PIP2** = Phosphatidylinositol-4,5-bisphosphate, **PKC**= Protéine Kinase C, **PLC** = Phospholipase C, **RE** = Réticulum endoplasmique.

De manière intéressante, l'analyse de Shi L. et al., apporte des informations complémentaires sur la survie des patientes. En effet, elle indique que l'expression d'ER $\alpha$ 36 est associée à une diminution de la survie sans récurrence (DFS) des patientes présentant une tumeur ER $\alpha$ 66+. De plus, dans cette cohorte, la survie à 10 ans post-traitement des patientes présentant une tumeur ER $\alpha$ 66+/ER $\alpha$ 36+ traitée au tamoxifène est significativement inférieure à celle des patientes avec une tumeur ER $\alpha$ 66+/ER $\alpha$ 36-. Ainsi, d'un point de vue clinique, l'expression d'ER $\alpha$ 36 dans les tumeurs mammaires est associée à un mauvais pronostic.

Il est à noter que le variant ER $\alpha$ 36 a également été décrit pour son rôle dans le développement et la progression d'autres types de cancer. En effet, l'expression de ce variant, qui est très élevée dans plusieurs lignées de cellules cancéreuses gastriques (SGC-7901, AGS, BGC-923 et MKN-45), est fortement corrélée au taux de métastases ganglionnaires dans ce type de cancer (Deng et al., 2010). De la même façon, une forte expression d'ER $\alpha$ 36 est corrélée à une diminution significative de la DFS dans le cancer de l'endomètre, les adénocarcinomes pulmonaires et les carcinomes rénaux (Sun et al., 2013 ; Zhang S. et al., 2014 ; Wang Q. et al., 2015). Enfin, contrairement à ce qui est décrit dans la plupart des cancers, ER $\alpha$ 36 est associé à un bon pronostic dans le cancer colorectal où son expression est inversement corrélée au taux de métastases ganglionnaires et au stade de la tumeur (Jiang H. et al., 2008 ; Sołtysik et Czekaj, 2015).

### B) Les voies de signalisation dépendantes d'ER $\alpha$ 36 dans les tumeurs mammaires

Les effets d'ER $\alpha$ 36 sur la progression et la résistance aux traitements des tumeurs mammaires sont liés à sa capacité à activer/médier les voies de signalisation non génomiques qui sont déclenchées au niveau membranaire (Figure 38). L'impact de différents agonistes (œstrogènes, perturbateurs endocriniens œstrogène-mimétiques) et antagonistes (tamoxifène ou fulvestrant) d'ER $\alpha$ 66 sur la modulation de l'activité d'ER $\alpha$ 36 demeure controversé (Lin A.H.Y. et al., 2013). En effet, la modélisation du LBD de ce variant, qui a été réalisée sur la base de la structure cristallographique du LBD de la forme canonique ER $\alpha$ 66, suggère que l'interaction entre ER $\alpha$ 36 et les différents ligands potentiels cités précédemment est possible. Cependant, les tests d'affinité de liaison entre ER $\alpha$ 36 et ces différentes molécules, réalisés in vitro, ne permettent pas de confirmer les interactions prédites par les modèles. Malgré tout, ces observations n'excluent pas que ces agonistes et antagonistes d'ER $\alpha$ 66 soient capables de moduler la signalisation dans laquelle ER $\alpha$ 36 est impliqué.

D'autre part, Kang et al., (2010), ont montré grâce à d'autres études de modélisation et à des tests d'affinité de liaison réalisés in vitro que l'agoniste de GPER, appelé G1, est capable d'interagir et d'activer directement ER $\alpha$ 36.

*i) La voie EGFR/Src/STAT3/5*

Les études menées par Zhang X. et al., (2011b ; 2012) dans différentes lignées de cellules cancéreuses mammaires ER $\alpha$ 66- (MDA-MB-231 et MDA-MB-436) ont montré qu'ER $\alpha$ 36 peut interagir physiquement avec Src. Cette interaction conduit à la phosphorylation du résidu Y416 de Src, qui est associé à son activation, et à la déphosphorylation du résidu Y527 qui est lui associé à son inactivation. Une fois activé, Src est à son tour responsable de l'activation d'EGFR via la phosphorylation du résidu Y845 de ce récepteur. L'activation de cette voie conduit à l'induction de l'effecteur STAT3/5 qui régule notamment l'expression de la cycline D1 et contribue ainsi à stimuler la croissance tumorale (Rao et al., 2011). En lien avec cette voie, il a également été démontré qu'ER $\alpha$ 36 pourrait stabiliser EGFR en empêchant sa dégradation (Zhang X. et al., 2011b).

*ii) La voie ER $\alpha$ 36/HER2/EGFR*

L'existence d'une interconnexion entre les voies de signalisation induites par ER $\alpha$ 36, HER2 et EGFR a été démontrée et semble jouer un rôle majeur dans l'acquisition de la résistance aux anti-œstrogènes. En effet, Chu et al., (2005) ont montré que l'utilisation en combinaison, du tamoxifène et du lapatinib, qui est un inhibiteur de l'activité de HER2 et d'EGFR, permet de restaurer la sensibilité des cellules cancéreuses mammaires MCF-7 au tamoxifène. De plus, cette boucle ER $\alpha$ 36/EGFR/HER2 est impliquée dans le développement d'une résistance à la chimiothérapie et plus particulièrement au Cisplatine, dans les cellules cancéreuses mammaires MCF-7, BT474 et MDA-MB-231 (Zhu L. et al., 2018).

Enfin, Kang et al., (2011a) ont démontré que ER $\alpha$ 36 et HER2 coopèrent pour le maintien et la régulation de la population de cellules souches cancéreuses mammaires ALDH1<sup>+</sup> dans les tumeurs.

*iii) La voie Ras/MEK/ERK1/2*

Wang Z. et al., (2006) ont montré, dans un modèle de cellules HEK293 exprimant ER $\alpha$ 36 suite à une transfection transitoire, que l'activation de la forme membranaire d'ER $\alpha$ 36 avec de l'E<sub>2</sub> couplé à de la BSA, conduit à la phosphorylation et donc à l'activation des kinases MEK1 et

ERK1/2. De plus, ils ont montré que le traitement de ces mêmes cellules par du tamoxifène ou du fulvestrant favorise la phosphorylation de ERK1/2 induite par ER $\alpha$ 36. Ces observations ont depuis été confirmées notamment dans une lignée de cellules cancéreuses mammaires ER $\alpha$ 66+ surexprimant une forme recombinante d'ER $\alpha$ 36 (MCF-7/ER $\alpha$ 36) et dans une lignée ER $\alpha$ 66- qui exprime de façon endogène ce variant (HCC38) (Wang Z. et al., 2006 ; Lin S.L. et al., 2010 ; Chaudhri et al., 2012).

Il a également été montré dans une étude récente qu'ER $\alpha$ 36 via son résidu L297 est capable d'interagir directement avec la forme phosphorylée de ERK2 (P-ERK2). Cette interaction stabilise l'activité de P-ERK2 en retardant l'action de la MKP3 qui inactive ERK en déphosphorylant les thréonines et les tyrosines. De plus, la phosphorylation de ERK1/2 induite par ER $\alpha$ 36 pourrait dépendre de l'activation de Src ou de la PKC (Protéine Kinase C) (Omarjee et al., 2017).

L'ensemble de ces mécanismes renforce l'activité de cette voie qui joue un rôle important notamment dans le développement des mécanismes de résistance aux anti-estrogènes.

#### *iv) La voie dépendante de la PKC*

Les études de Chaudhri et al., (2014b) menées sur les cellules cancéreuses mammaires triples négatives HCC38 ont démontré l'importance de la voie dépendante de la PKC dans la signalisation impliquant ER $\alpha$ 36.

Dans cette voie, l'activation de la signalisation dépendante d'ER $\alpha$ 36 par l'E2 conduit à l'activation de la PLC (Phospholipase C) qui aboutit au clivage du PIP2 en IP3 et en DAG. L'IP3 permet le relargage du calcium contenu dans le réticulum endoplasmique lisse tandis que le DAG permet l'activation de la PKC. Comme nous l'avons vu précédemment, la PKC peut induire la phosphorylation de ERK1/2 et ainsi promouvoir la prolifération et le potentiel métastatique des cellules tumorales. De plus, en lien avec cette voie, Chaudhri et al.,(2014a) ont montré que le PIP2 peut, via l'activation d'AkT, inhiber JNK (c-Jun Kinase) qui est responsable de la cascade apoptotique. Ainsi, la signalisation dépendante d'ER $\alpha$ 36 est également associée au développement de mécanismes de résistance à l'apoptose.

#### *v) La voie PI3K/AkT*

L'activation de la voie PI3K/AkT en lien avec ER $\alpha$ 36 a été démontrée dans le contexte des cellules cancéreuses endométriales Hec1A. En effet, en réponse à l'E<sub>2</sub> ou au tamoxifène, on



```

-736  GGTACCCGCG CCCCGCCGC CCGTCGGGGT GGCCGCCGCG CCCGGCAGGA
      Sp1      Sp1      Sp1      Sp1      Sp1
-686  GGGAGGGGAGG GAGGGAGGGA GAAGGGAGAG CCTAGGGAGC TGCGGGAGCC
      Sp1      Sp1      AhR
-636  CGGGGACGCG CGACCCGAGG GTGCGCCAGG GGAGCCCGGG GCCGCGGCC
      Sp1      WT1
-586  CAGCCCGGGG GTTCTGCGTG CAGCCCGGCG TGCGTTCAGA GTCAAGTTCT
      Sp1      Sp1      AP1
-536  CTCGCCGGG AGCTGAAAA AACGTACTCT CCACCCACTT ACCGTCCGTG
      Erg-1      Sp1
-486  CGAGAGGCCAG ACCCGAAAGC CCGGGCTTCC TAACAAAACA CACGTTGGAA
      Sp1      NF-κB
-436  AACCAGACAA AGCAGCAGTT ATTTGTGGGG GAAAACACCT CCAGGCAAAT
      GATA-1      GR      NF-κB
-386  AAACACGGGG CGCTTTGAGT CAGTTGGGAA GGTCTCGCTC TTGGCATTTA
      Sp1      ER      AP1
-336  AAGTGGGGG TGTTTGAGT TAGCAGAGCT CAGCAGAGTT TTATTTATCC
      Sp1      AP1      TBP
-286  TTTTAATGTT TTTGTTTAAAT GTGCTCCCA AATTTCTTT CATCTAGACT
      PU.1
-236  ATTTGATTGG AAATATGTCA GCTATGATGA TGACTTTCTG GGAAGCGATT
      GATA-1      Elk-1
-186  CCTGTCACCC GCTTTCCCT CCTCCCACC CCACGTCTG GGGCTTTAGA
      GATA-1      Sp1      Erg-1
-136  GAGCGATTGG GAGTTGAATG GGTCTGATTT CGGAGTTAGC TGGCTGAGTC
      c-Jun
-86   CGCGCTGGAG CGGATTGCTG GCATGTGACTT CTGACAGCCG GAAATTTGT
      Sp1      AP-1
-36   AGGTGTCCCGC GAGTTAAAA CAAGCCATAT GGAAGCAAA GTGCTTAAA
      Sp1      GR      +1
+15   AA

```

**Figure 39 : Organisation de la région promotrice d'ERα36.**

La région promotrice d'ERα36 s'étend sur 750 pb et comporte une boîte TATA non canonique qui est encadrée sur la séquence. On trouve également différents sites de fixation pour les facteurs de transcriptions Sp1, AP1, AhR, WT1, Erg-1, NF-κB, GATA-1, GR, ER, Elk-1, c-Jun et PU.1. Ces différents sites sont soulignés. Enfin, le site d'initiation de la transcription d'ERα36 est représenté par une flèche et détermine la position +1 de la séquence.

**AhR** = Aryl Hydrocarbon Receptor, **AP1** = Activator Protein 1, **Elk-1** = ETS Like-1, **ER** = Estrogen Receptor, **Erg-1** = ETS-related gene, **NF-κB** = Nuclear Factor-kappa B, **GR** = Glucocorticoid Receptor, **PU.1** = PU-box binding factor 1, **SP1** = Specificity Protein 1, **WT1** = Wilm's Tumor suppressor 1.

(D'après Zou et al., 2009)

observe une phosphorylation, c'est-à-dire une activation rapide d'AkT qui peut conduire à une augmentation de la prolifération et, comme nous l'avons vu précédemment, à une inhibition de la cascade de signalisation apoptotique (Lin S.L. et al., 2010, Chaudhri et al., 2014a ; 2014b). De plus, il a été démontré qu'en présence d'E<sub>2</sub>, ER $\alpha$ 36 est capable d'interagir directement avec la PI3K dans le cytoplasme de cellules cancéreuses mammaires triples négatives exprimant ER $\alpha$ 36 de façon endogène (Lignée HBCc-12A dérivée de HBCx-12A PDX). Toutefois, les mécanismes à l'origine de cette interaction n'ont à l'heure actuelle pas été identifiés (Omarjee et al., 2017).

En conclusion, l'ensemble de ces résultats indique que le variant ER $\alpha$ 36 joue un rôle pivot dans différentes voies de signalisation, précédemment décrites pour leur implication dans la progression tumorale. Toutefois, il est important de préciser que l'ensemble de ces données ont été obtenu in vitro dans différents types de lignées cellulaires et principalement dans des modèles de surexpression. Il reste donc à démontrer leur validité in vivo dans un contexte cellulaire physiopathologique.

### C) Mécanismes de régulation de l'expression d'ER $\alpha$ 36

Récemment, Nagel et al., (2019) ont étudié le lien potentiel entre l'amplification du gène ESR1 et le niveau d'expression d'ER $\alpha$ 36. En effet, ils ont observé que la diminution de la DFS associée à l'expression d'ER $\alpha$ 36 dans leur cohorte est similaire à la diminution de la DFS associée à l'amplification du gène ESR1 détectée par qPCR. Toutefois, quelle que soit la technique utilisée, qPCR ou FISH, aucune corrélation n'a pu être établie entre le nombre de copies du gène ESR1 et le niveau d'expression d'ER $\alpha$ 36. Il est à noter qu'aucune des amorces utilisées ne cible l'exon 9 du gène ESR1 qui est pourtant spécifique d'ER $\alpha$ 36 (Markiewicz et al., 2013).

Concernant la régulation transcriptionnelle d'ER $\alpha$ 36, la région promotrice associée à l'expression de ce variant, est localisée dans le premier intron du gène ESR1 et a été clonée en 2009 par Zou et al., (Figure 39). Cette région, qui s'étend sur environ 750 pb, est riche en GC et contient une boîte TATA non canonique (-295 à -290) qui permet le recrutement de la machinerie transcriptionnelle via la protéine TBP. On trouve également des séquences consensus qui correspondent aux sites de fixation de différents facteurs de transcription représentés dans la Figure 39. En revanche, cette région promotrice ne contient pas la séquence consensus CCAAT, qui est impliquée dans le recrutement du facteur C/EBP.

La présence d'un demi-ERE imparfait (5'-AGTCA-3') dans le promoteur d'ER $\alpha$ 36 suggère que l'expression de ce variant pourrait être régulée directement par la forme canonique ER $\alpha$ 66. Cela a été confirmé *in vitro* en co-transfectant, dans les cellules HEK293 (ER $\alpha$ 66-), un vecteur d'expression d'ER $\alpha$ 66 et un vecteur contenant la région promotrice d'ER $\alpha$ 36 placée en amont de la séquence codant la luciférase. En présence d'ER $\alpha$ 66, l'activité du promoteur d'ER $\alpha$ 36 diminue de façon significative dans ce modèle. Cela indique que la forme canonique ER $\alpha$ 66 peut réguler négativement l'expression de son variant (Zou et al., 2009). De plus, cette relation inverse entre l'expression d'ER $\alpha$ 66 et celle d'ER $\alpha$ 36 est renforcée par l'action de WT1 (Wilm's Tumor 1). En effet, ce facteur régule à la hausse l'expression d'ER $\alpha$ 66 tandis qu'il régule à la baisse celle d'ER $\alpha$ 36 en se fixant au niveau de la séquence consensus 5' GGGGCGCGCG 3' localisée en amont de la boîte TATA (Kang et al., 2011b).

A l'heure actuelle, le lien direct entre l'expression d'ER $\alpha$ 36 et les autres facteurs de transcription possédant un site de liaison dans la région promotrice de ce variant n'est pas clairement décrit dans la littérature. Néanmoins, ces facteurs interviennent comme des médiateurs, parfois indirects, de la signalisation œstrogénique dépendante d'ER $\alpha$ 66. En effet, les données disponibles sur quelques-uns de ces facteurs dans le contexte des tumeurs mammaires indiquent que AhR est inversement corrélé à l'expression d'ER $\alpha$ 66 dans une cohorte de 439 patientes et que NF- $\kappa$ B présente une activité plus importante dans les tumeurs mammaires ER $\alpha$ 66- et HER2+ (Biswas et al., 2000 ; Vacher et al., 2018). De plus, une étude de Vivacqua et al., (2012) indique que l'expression d'Egr-1 est induite en réponse à l'E<sub>2</sub> et au tamoxifène dans les cellules cancéreuses mammaires SKBr3 (ER $\alpha$ 66-) de façon dépendante de l'activation de GPER. Enfin Kang et al., (2010) ont également mis en évidence que la signalisation dépendante de ce récepteur aux œstrogènes membranaire induit l'activité du promoteur d'ER $\alpha$ 36 via la signalisation Src/MEK1/2/AP-1.

D'autre part, Yin et al., (2014) ont montré que l'expression d'ER $\alpha$ 36, EGFR et HER2 est induite en réponse au tamoxifène dans les cellules cancéreuses mammaires MCF-7, T47D et H3396. Cela suggère l'existence d'une boucle de régulation positive entre ces 3 récepteurs. En effet, la signalisation induite par EGFR et HER2 est associée à une augmentation de l'expression d'ER $\alpha$ 36 de façon dépendante des sites AP-1 localisés dans la région promotrice de ce variant (Yin et al., 2014 ; Kang et al., 2011a). En retour, l'induction de l'expression d'ER $\alpha$ 36 est responsable d'une augmentation du taux de protéines HER2 et EGFR.

Ainsi, malgré ces quelques données éparses, peu d'informations sont disponibles concernant les mécanismes qui régulent l'expression d'ER $\alpha$ 36 dans le contexte des cellules cancéreuses

mammaires mais également dans un contexte physiologique. Pourtant, cela pourrait permettre de mieux appréhender l'implication éventuelle d'ER $\alpha$ 36 au moment de l'initiation et/ou de la progression tumorale mammaire, ainsi que lors de l'acquisition de résistances aux anti-œstrogènes.

#### D) Mécanismes de régulation de l'activité d'ER $\alpha$ 36

La majorité des mutations du gène ESR1 affectant le LBD et par conséquent l'activité d'ER $\alpha$ 66, sont localisées dans des régions qui sont absentes chez ER $\alpha$ 36. En effet, seuls les acides aminés 184 à 430 d'ER $\alpha$ 66 sont conservés dans la séquence d'ER $\alpha$ 36. Ainsi, bien que cela ne soit pas décrit dans la littérature, nous pouvons émettre l'hypothèse que les mutations K303R, E380Q et V392I pourraient avoir un impact sur l'activité d'ER $\alpha$ 36. De la même façon, les PTMs qui affectent la région 184 à 430 d'ER $\alpha$ 66, et notamment la phosphorylation des sérines S282 et S305, sont également susceptibles de moduler l'activité d'ER $\alpha$ 36.

D'autre part, l'activité du variant ER $\alpha$ 36 peut être régulée grâce à l'intervention de protéines chaperonnes comme HSP90 ou la synucléine gamma. De manière générale, les protéines chaperonnes permettent d'augmenter la demi-vie et de maintenir le récepteur dans une conformation active afin de faciliter l'interaction avec les ligands. Contrairement à HSP90 qui est également exprimée dans des conditions physiologiques, la synucléine gamma est exprimée uniquement dans le contexte des tumeurs mammaires. L'interaction physique entre la synucléine gamma et ER $\alpha$ 36, qui est indépendante de la présence d'HSP90, stimule l'activation des voies de signalisation non génomiques, faisant intervenir ERK1/2, notamment en réponse à l'E<sub>2</sub> et au tamoxifène (Shi Y.E. et al., 2010). Des études antérieures portant sur 3 modèles de cellules cancéreuses mammaires (T47D, MCF-7 et MDA-MB-435) ont également montré que l'expression de la synucléine gamma est associée à une augmentation de l'activité transcriptionnelle d'ER $\alpha$ 66. Cela suggère que cette protéine chaperonne pourrait également interagir avec ER $\alpha$ 66 (Jiang Y. et al., 2003 ; Liu et al., 2007). Ainsi, en stabilisant ER $\alpha$ 36 dans sa forme active, la synucléine gamma est associée à un mauvais pronostic car elle participe à l'induction de la progression tumorale (prolifération, migration) et au développement de la résistance au tamoxifène (Shi Y.E. et al., 2010 ; Panneerselvam et al., 2015).

#### E) Molécules régulant l'expression et l'activité d'ER $\alpha$ 36

Outre les œstrogènes, différentes molécules capables de réguler l'expression et l'activité d'ER $\alpha$ 36 ont été identifiées. Comme nous l'avons vu précédemment, l'agoniste de GPER, G1,

stimule de façon directe l'activité de ce variant. De plus, les anti-œstrogènes comme le tamoxifène et le fulvestrant sont associés à l'activation de voies de signalisation dans lesquelles ER $\alpha$ 36 est impliqué. Cela pourrait notamment s'expliquer par l'induction de l'expression d'ER $\alpha$ 36 en réponse à ces traitements. En effet, une exposition chronique au tamoxifène est associée à une augmentation de l'expression d'ER $\alpha$ 36 dans les cellules cancéreuses mammaires MCF-7 (Yin et al., 2014).

A l'inverse d'autres molécules comme le brousoflavonol B et l'icaritine qui sont des flavonoïdes purifiés respectivement à partir de l'écorce du murier et d'une herbe médicinale chinoise (Herba Epimedii), régulent négativement l'expression d'ER $\alpha$ 36. Ces 2 molécules sont associées à une inhibition de la croissance des cellules cancéreuses mammaires MDA-MB-231 (Guo M. et al., 2013 ; Wang X. et al., 2017).

Enfin, dans le contexte de l'activation de Src par ER $\alpha$ 36, un effet biphasique de l'E<sub>2</sub> a été mis en évidence. En effet, le traitement des cellules cancéreuses mammaires MDA-MB-231 et MDA-MB-436 avec 1 nM d'E<sub>2</sub> conduit à la phosphorylation activatrice de Src, au niveau de la Tyrosine Y416. A l'inverse, un traitement avec 5  $\mu$ M d'E<sub>2</sub> conduit à la phosphorylation inhibitrice de la Tyrosine Y527 de Src. Le même effet a été décrit en présence de 1nM ou de 5  $\mu$ M d'anti-œstrogènes (Zhang X. et al., 2012).

---

## *Contexte et objectifs de la thèse*

---

Dans un contexte où l'incidence du cancer du sein tend à nouveau à augmenter et constitue toujours la première cause de décès liée au cancer chez la femme, les équipes de recherche tentent d'apporter de nouveaux éléments pour permettre une prise en charge de plus en plus efficace et personnalisée des patientes.

En clinique, le choix de la stratégie thérapeutique à adopter est guidée principalement par la classification moléculaire qui repose essentiellement sur l'expression de 3 biomarqueurs (ER $\alpha$ 66, PR et HER2). A l'heure actuelle, malgré l'amélioration de la prise en charge d'un grand nombre de patientes, 1/3 des tumeurs mammaires ER $\alpha$ 66+ traitées par hormonothérapie développent des mécanismes de résistance qui se traduisent par la reprise de la progression tumorale en dépit du traitement (Piggott et al., 2018). Ce constat souligne la nécessité d'améliorer cette classification moléculaire grâce à l'identification de nouveaux biomarqueurs de la progression tumorale. Cela permettrait de prédire plus précisément la réponse des tumeurs à l'hormonothérapie mais aussi d'orienter l'utilisation des nouvelles thérapies ciblées.

Les travaux menés au sein du département BioSIS du laboratoire portent notamment sur la validation du variant des récepteurs alpha aux œstrogènes, ER $\alpha$ 36, comme nouveau biomarqueur prédictif de la progression tumorale mammaire. En effet, depuis sa découverte en 2005, de nombreuses études suggèrent que l'expression de ce variant dans les tumeurs mammaires ER $\alpha$ 66 positives ou négatives est associée à la progression des cancers du sein et à l'acquisition de résistances aux traitements.

Les premiers travaux du laboratoire en lien avec ER $\alpha$ 36 ont été menés dans le contexte du cancer testiculaire, qui est également un cancer hormono-sensible. Ces études ont montré que les alkylphénols, qui sont des perturbateurs endocriniens oestrogène-mimétiques, induisent la prolifération des cellules séminomateuses TCam-2 de façon dépendante d'ER $\alpha$ 36 (Ajj et al., 2013). La thématique de l'équipe a ensuite évolué pour se concentrer sur le rôle de ce variant dans la physiopathologie de la glande mammaire. Grâce au développement d'une méthode générique basée sur la modélisation et la comparaison de réseaux de gènes, Clémence Chamard-Jovenin a montré au cours de sa thèse qu'une faible expression d'ER $\alpha$ 36 dans les tumeurs mammaires est associée à une croissance tumorale dépendante des œstrogènes, tandis qu'une forte expression de ce variant est associée à la progression métastatique (Chamard-Jovenin et al., 2015).

Si l'ensemble des études disponibles indiquent que l'expression d'ER $\alpha$ 36 dans les cellules cancéreuses mammaires, mais également dans les cellules de cancer testiculaire d'origine

germinale, participe à la progression tumorale, peu de données sont disponibles dans la littérature concernant son rôle dans un contexte physiologique.

C'est pourquoi, la première partie de ce travail de thèse a été consacrée à la réalisation d'une étude *in vivo* permettant d'évaluer les conséquences de l'expression du variant ER $\alpha$ 36 sur le développement physiologique de la glande mammaire et de la lignée germinale mâle. Pour cela, nous avons établi 2 modèles de souris transgéniques, C57BL/6J MMTV-ER $\alpha$ 36 et C57BL/6J Oct4-ER $\alpha$ 36, exprimant respectivement ce variant typiquement humain dans la glande mammaire ou dans les cellules germinales murines. Afin d'approfondir l'étude des effets d'ER $\alpha$ 36 dans la glande mammaire saine nous avons également évalué, *in vitro* et *in silico*, l'impact de sa surexpression dans des cellules épithéliales mammaires saines immortalisées MCF-10A. Enfin, les conséquences de l'expression d'ER $\alpha$ 36 couplée à une exposition aux alkylphénols, ont été caractérisées car ces perturbateurs endocriniens induisent l'expression de ce variant dans les cellules séminomateuses TCam-2 et pourraient moduler son activité.

Dans la continuité, le second objectif de ce travail était d'étudier les modalités de régulation classiques et épigénétiques de l'expression de ce variant dans les cellules cancéreuses mammaires afin de mieux comprendre l'implication d'ER $\alpha$ 36 lors de l'initiation et/ou de la progression tumorale. Pour cela, nous avons étudié d'une part, l'impact de certains facteurs de transcription et du niveau de méthylation de la région promotrice d'ER $\alpha$ 36 sur son expression, et d'autre part, la régulation post-transcriptionnelle de l'expression de ce variant par les micro-ARNs. En parallèle, nous développons une approche basée sur l'utilisation d'outils bioinformatiques et statistiques afin de décrypter les réseaux de gènes associés à l'expression d'ER $\alpha$ 36 et d'identifier de nouveaux partenaires éventuels de ce variant.



---

# *Résultats*

---

Les travaux réalisés au cours de ma thèse avaient pour objectif de répondre à 2 problématiques :

**1) Quel est l'impact d'ER $\alpha$ 36 sur le développement physiologique de la lignée germinale mâle et de la glande mammaire ?**

Après avoir exposé des résultats non publiés concernant les conséquences de l'expression d'ER $\alpha$ 36 dans la lignée germinale mâle murine, je présenterai les principaux résultats issus des publications N°1 et N°2, qui concernent l'impact de l'expression de ce variant dans les cellules épithéliales mammaires saines.

**Publication N°1 :**

**C. Thiebaut**, C. Chamard-Jovenin, A. Chesnel, C. Morel, E.H. Djermoune, T. Boukhobza, H. Dumond. 2017. "Mammary epithelial cell phenotype disruption in vitro and in vivo through ER $\alpha$ 36 overexpression". PLoS ONE, 12, pp.e0173931.

**Publication N°2 :**

C. Chamard-Jovenin, **C. Thiebaut**, A. Chesnel, E. Bresso, C. Morel, M. Smaïl-Tabbone, M.D. Devignes, T. Boukhobza, H. Dumond. 2017. "Low-dose alkylphenol exposure promotes mammary epithelium alterations and transgenerational developmental defects, but does not enhance tumorigenic behaviour of breast cancer cells". Frontiers in Endocrinology, 8, pp.272.

**2) Quels sont les mécanismes qui régulent l'expression d'ER $\alpha$ 36 ?**

Afin de répondre à cette deuxième question, je présenterai successivement les différentes pistes explorées au cours de ma thèse en me focalisant plus particulièrement sur les mécanismes épigénétiques et post-transcriptionnels qui ont fait l'objet de la publication N°3. Enfin, je terminerai en présentant les étapes de développement et l'application d'une approche statistique permettant d'identifier de nouveaux facteurs impliqués dans la signalisation dépendante d'ER $\alpha$ 36. Cette dernière partie fait l'objet de la publication méthodologique N°4, portée par nos collaborateurs de l'IECL de Nancy.

**Publication N °3 :**

**C. Thiebaut**, A. Chesnel, J.L. Merlin, M. Chesnel, A. Leroux, A. Harlé, and H. Dumond. 2019. “Dual epigenetic regulation of ER $\alpha$ 36 expression in breast cancer cells”. International Journal of Molecular Sciences, 20(11), 2637.

**Publication N °4 :**

B. Bastien, T. Boukhobza, H. Dumond, **C. Thiebaut**, H. Chakir, A. Gegout-Petit, A. Muller-Gueudin, Y. Shi. “*A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology*”. Soumis le 04/07/2019 dans le journal Biostatistics.

Bien que certains éléments ponctuels de discussion soient apportés dans cette partie, les principaux résultats seront discutés de manière transversale et plus approfondie dans la partie « Discussion et Perspectives ».

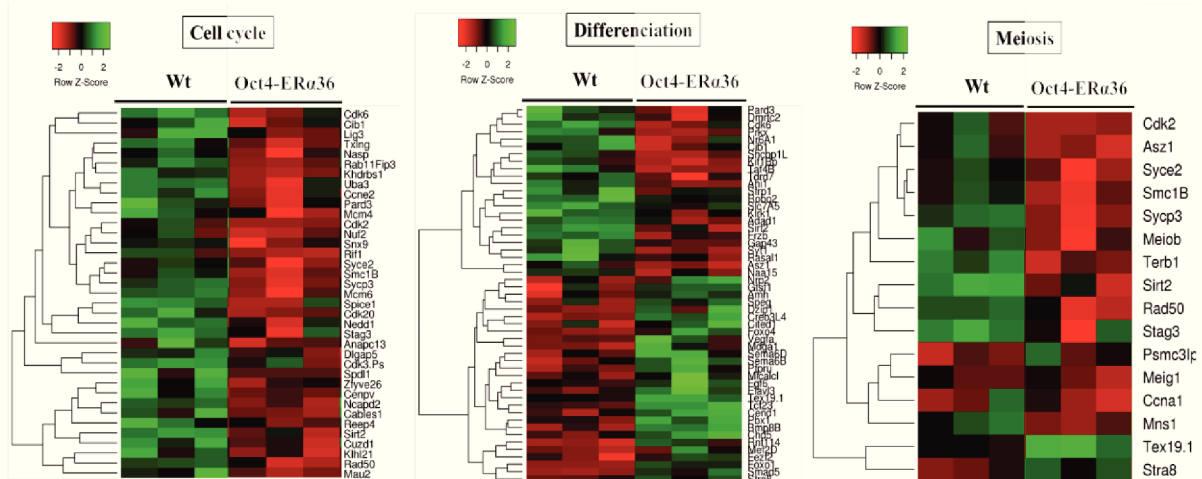
## I. Quel est l'impact d'ER $\alpha$ 36 sur le développement physiologique de la lignée germinale mâle et de la glande mammaire ?

Depuis son clonage en 2005, le variant *ER $\alpha$ 36* a été principalement décrit dans la littérature pour son rôle dans la progression de différents types de cancers. Au début de ma thèse peu d'informations étaient disponibles concernant son implication dans un contexte physiologique. Toutefois, il avait été décrit que l'ARN messenger codant *ER $\alpha$ 36* était exprimé dans différents tissus sains cibles des œstrogènes comme la glande mammaire, l'utérus, les tissus osseux et cartilagineux, les tubules rénaux, l'estomac... (Zheng et al., 2010 ; Xie H. et al., 2011 ; Wang J. et al., 2012 ; Sun et al., 2013 ; Wang Q. et al., 2015 ; Yan Y. et al., 2017). Cela suggèrerait donc que ce variant pourrait participer à la médiation des effets physiologiques des œstrogènes.

Les études antérieures menées au laboratoire avaient participé à la démonstration de l'implication d'ER $\alpha$ 36 dans la progression des tumeurs testiculaires d'origine germinale et des tumeurs mammaires. Nous avons choisi d'étudier au cours de ma thèse l'impact de l'expression de ce variant sur le développement physiologique de la lignée germinale mâle et de la glande mammaire murines. Etant donné que la séquence codant ER $\alpha$ 36 n'est pas présente dans le génome de la souris, nous avons développé en collaboration avec l'institut clinique de la souris (ICS, Illkirch-Graffenstaden) 2 lignées C57BL/6J transgéniques exprimant spécifiquement ER $\alpha$ 36 dans la lignée germinale ou dans la glande mammaire.

Plus précisément, afin de permettre l'expression d'ER $\alpha$ 36 dans la lignée germinale, la séquence codant ce variant a été placée sous le contrôle du promoteur d'Oct4 délété de son élément proximal (Oct4- $\Delta$ PE). En effet, contrairement à l'élément distal qui permet une expression restreinte à la lignée germinale, l'élément proximal de ce promoteur peut être activé dans d'autres types cellulaires dérivant de l'épiblaste. De plus, le promoteur Oct4- $\Delta$ PE permet l'expression du transgène dès le 8<sup>ème</sup> jpc (jour post-coïtum) dans les cellules germinales primordiales (Yeom et al., 1996 ; Yoshimizu et al., 1999). Concernant l'expression induite spécifiquement dans la glande mammaire, la séquence codant ER $\alpha$ 36 a été placée sous le contrôle du promoteur MMTV (Mouse Mammary Tumor Virus) qui présente une activité constitutive basale dans les cellules de l'épithélium mammaire des femelles vierges, contrairement aux promoteurs WAP (Whey acidic protein) et BLG ( $\beta$ -LactoGlobuline) qui sont activés spécifiquement au moment de la gestation et de la lactation (Menezes et al., 2014).

Afin de s'affranchir d'un potentiel effet de l'expression d'ER $\alpha$ 36 sur le comportement maternel, le transgène, Oct4-ER $\alpha$ 36 ou MMTV-ER $\alpha$ 36, a été transmis uniquement par des mâles



**Figure 40 : Impact de la surexpression d'ERα36 dans le testicule fœtal de souris sur les processus biologiques « Cell cycle », « Differentiation » et « Meiosis »**

Heatmaps représentant le niveau d'expression d'un certain nombre de gènes différentiellement régulés dans les cellules germinales fœtales mâles prélevées chez les animaux C57BL6/J sauvages (Wt), n'exprimant pas ERα36 ou chez les animaux C57BL6/J transgéniques (Oct4-ERα36) qui expriment ce variant dans la lignée germinale. Cell cycle : 37 gènes ; Differentiation : 51 gènes et Meiosis : 16 gènes. Rouge : faible expression vs Vert : forte expression.

Pour chaque condition, 3 réplicats biologiques correspondant à des lots de gonades fœtales sont représentés.

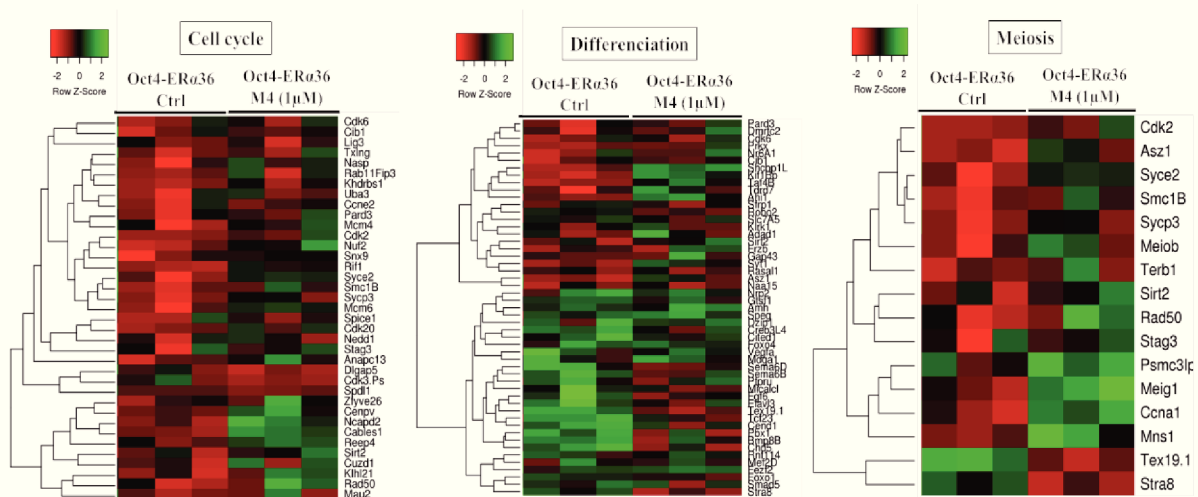
transgéniques hémizygotés, c'est-à-dire porteurs d'une seule copie du transgène, croisés avec des femelles sauvages de lignée consanguine C57BL/6J. (Voir la description détaillée du modèle MMTV-ER $\alpha$ 36 dans la partie « matériel et méthodes » de la publication N°1).

#### A) Impact d'ER $\alpha$ 36 sur le développement de la lignée germinale mâle murine

L'étude du modèle transgénique Oct4-ER $\alpha$ 36 a été menée en collaboration avec le laboratoire de développement des gonades (LDG UMR SGCSR, CEA, Fontenay aux Roses) dirigé par le professeur Gabriel Livera. La stratégie développée avait pour objectif de suivre les conséquences de l'expression du transgène sur la fertilité, l'histologie des gonades et la production de testostérone. Une analyse transcriptomique, par puce Affymetrix, sur des cellules germinales fœtales mâles collectées chez les animaux sauvages (wt) et transgéniques (Tg Oct4-ER $\alpha$ 36) au moment de l'entrée en quiescence, c'est-à-dire à environ 15,5 jpc, a également été réalisée.

Au total, Marie-Justine Guerquin a mis en évidence 1365 gènes dont l'expression varie de manière significative ( $p < 0,05$ ) entre le lot wt et le lot Tg Oct4-ER $\alpha$ 36. Afin de déterminer quelles sont les fonctions cellulaires impactées par l'expression d'ER $\alpha$ 36 dans les cellules germinales mâles la liste des 1365 gènes dérégulés a été soumise dans les bases de données « Panther (Protein ANalysis THrough Evolutionary Relationships) Classification System » (<http://pantherdb.org/>) et « DAVID (Database for Annotation, Visualization and Integrated Discovery) » (<https://david.ncifcrf.gov/>) qui permettent de déterminer les principaux processus biologiques (Termes GO BP (Gene Ontology, Biological Process)) qui sont enrichis dans les différentes listes de gènes.

Les résultats obtenus, qui n'ont pas encore fait l'objet d'une publication, indiquent que la signalisation dépendante d'ER $\alpha$ 36 module l'expression de gènes associés aux termes GO : « Cell cycle », « Meiosis » et « Differentiation » ( $p\text{-value} < 3 \times 10^{-2}$ ) (Figure 40). Cela suggère que l'expression de ce variant dans les cellules de la lignée germinale testiculaire est associée une modulation des processus de prolifération, de division méiotique et de différenciation qui pourrait conduire à une perte de l'identité germinale des cellules. Cependant, aucun défaut de différenciation, de qualité spermatique et de fertilité n'a été observé dans les testicules des animaux Tg Oct4-ER $\alpha$ 36 à l'âge adulte (non présenté ici). Etant donné que cette isoforme d'ER $\alpha$  n'est pas naturellement exprimée chez la souris, on peut émettre l'hypothèse que les effets de l'expression artificielle de ce variant sur la différenciation germinale pourraient être



**Figure 41 : Impact des alkylphénols associés à la surexpression d'ERα36 dans le testicule fœtal de souris sur les processus biologiques « Cell cycle », « Differentiation » et « Meiosis ».**

Heatmaps représentant le niveau d'expression des différents gènes, précédemment étudiés dans la figure 40, dans des cellules germinales fœtales mâles prélevées chez les animaux C57BL6/J transgéniques Oct-4-ERα36 contrôles, c'est-à-dire non exposés au M4, ou chez les animaux C57BL6/J transgéniques Oct4-ERα36 exposés au M4 (0,5 μg/kg/jour) durant la gestation.

Cell cyle : 37 gènes ; Differentiation : 51 gènes et Meiosis : 16 gènes. Rouge : faible expression vs Vert : forte expression.

Pour chaque condition, 3 répliquats biologiques correspondant à des lots de gonades fœtales sont représentés.

compensés par les voies de signalisation endogènes qui se mettent en place lors de la puberté pour réguler la spermatogénèse.

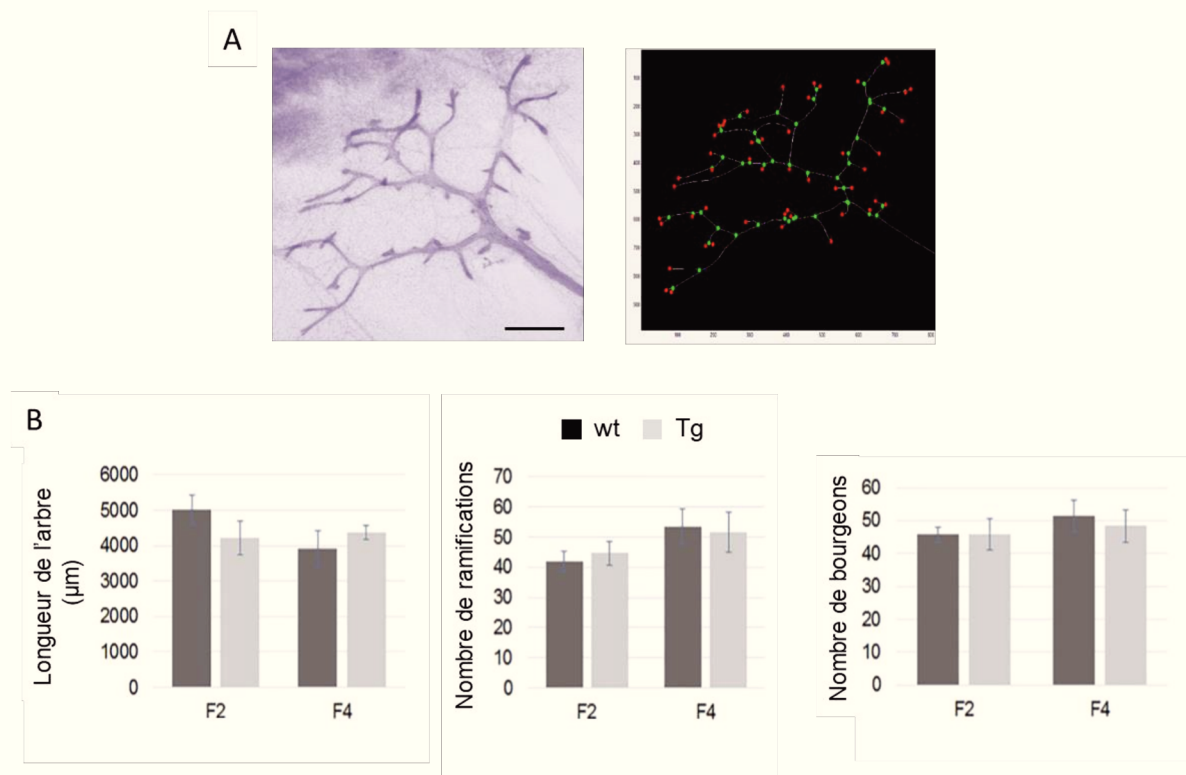
En parallèle, notre laboratoire avait précédemment montré *in vitro* qu'ER $\alpha$ 36 est un médiateur de la réponse des cellules de cancrs testiculaires d'origine germinale, TCam-2, aux alkylphénols à longues chaînes. En effet, le mélange d'alkylphénols M4 qui se compose de 4-nonylphénol et de 4-tert-octylphenol (en rapport 30:1), stimule la prolifération des cellules séminomateuses TCam-2 de façon dépendante d'une signalisation non génomique induite par ER $\alpha$ 36 en faisant intervenir la kinase PI3K et le facteur de transcription CREB (Ajj et al., 2013).

Afin d'évaluer si l'induction de la signalisation dépendante d'ER $\alpha$ 36 par les alkylphénols pourrait potentialiser l'impact de ce variant sur la différenciation des cellules germinales, nous avons étudié grâce au modèle de souris Tg Oct4-ER $\alpha$ 36, les effets combinés de l'expression de ce variant et d'une exposition *in utero* de 5 jours à une dose de M4, correspondant à l'exposition humaine (0,5  $\mu$ g/kg/jour).

L'analyse transcriptomique des cellules germinales fœtales mâles prélevées à 15,5 jpc chez les animaux Tg Oct4-ER $\alpha$ 36 exposés ou non au M4 *in utero* a permis de mettre en évidence une variation du niveau d'expression des gènes associés aux mêmes termes GO que précédemment à savoir : « Cell cycle », « Differentiation » et « Meiosis » (p-value < 1,5 x 10<sup>-2</sup>) (Figure 41). Toutefois, lorsque l'on compare les heatmaps des Figures 40 et 41, on constate que chez les animaux Tg Oct4-ER $\alpha$ 36 exposés au M4, le profil d'expression d'un certain nombre de gènes impliqués dans le contrôle de la prolifération et de la division méiotique tend à se rapprocher de celui observé chez les animaux wt. Cela suggère que le M4 perturbe les effets liés à l'expression d'ER $\alpha$ 36 dans la lignée germinale. Néanmoins, aucune modification de l'histologie testiculaire n'est observée à l'âge adulte chez les animaux Tg Oct4-ER $\alpha$ 36 qui avaient été exposés *in utero* pendant 5 jours au mélange M4.

Dans la continuité, afin de mieux comprendre le mécanisme moléculaire qui relie l'expression d'ER $\alpha$ 36 et la différenciation des cellules germinales mâles, il serait intéressant d'étudier la signalisation dépendante de la surexpression d'ER $\alpha$ 36 *in vitro* dans des lignées de cellules germinales testiculaires saines comme la lignée murine GC-1 spg. Etant donné qu'aucune lignée d'origine humaine n'a été établie, l'étude de l'impact d'une expression endogène d'ER $\alpha$ 36 n'est à l'heure actuelle pas réalisable. Enfin, il pourrait être intéressant de vérifier si le transgène semble avoir le même impact sur la différenciation des cellules germinales fœtales femelles.





**Figure 42 : Analyse quantitative de l'arbre mammaire des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 au sevrage.** (Figure 5 de la publication n°1)

Cette étude a été menée sur des arbres mammaires prélevés au moment du sevrage (21 jours) sur des souris sauvages (wt) ou transgéniques MMTV-ER $\alpha$ 36 (Tg) appartenant aux générations F2 ou F4. Cela permet d'évaluer si l'effet du transgène est amplifié ou au contraire réprimé au cours des générations.

(A) Après coloration de l'arbre mammaire avec du carmin et observation microscopique (à gauche), le logiciel à façon mis au point au laboratoire permet de générer un squelette (à droite), sur lequel sont représentés les terminaisons (points rouge) et les embranchements (points vert).  
Barre d'échelle : 50  $\mu$ m

(B) A partir de ce squelette, le logiciel détermine la longueur de l'arbre mammaire, le nombre de ramifications et le nombre de bourgeons.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. F2 wt : N = 4, F2 Tg : N = 6, F4 wt : N = 6, F4 Tg : N = 5.

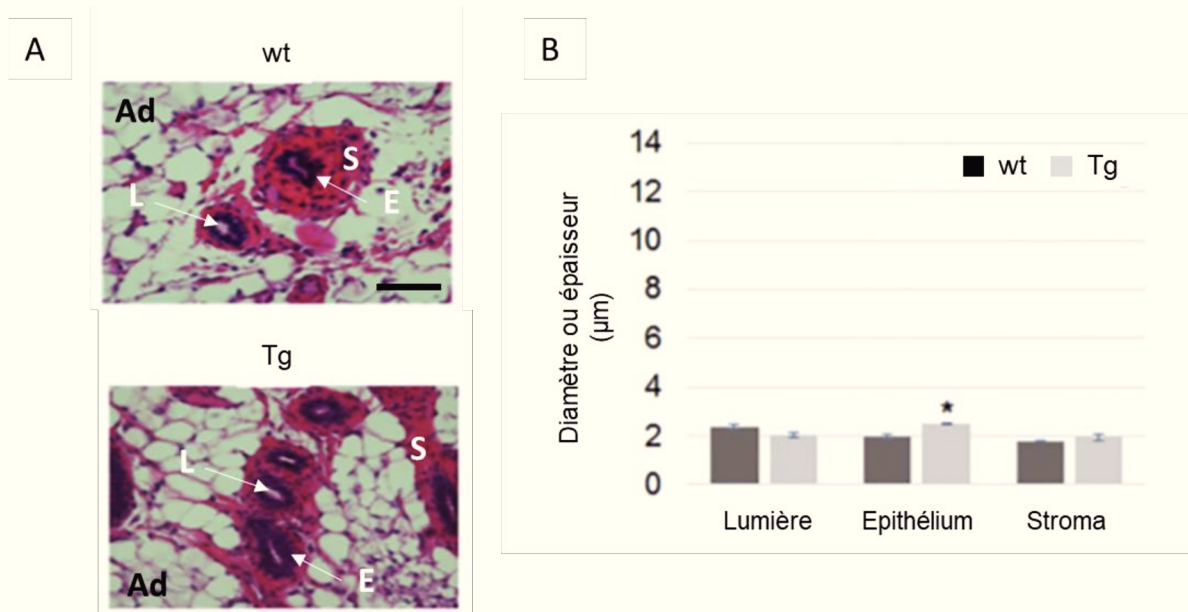
## B) Impact d'ER $\alpha$ 36 sur le développement de la glande mammaire murine

### i) Etudes *in vivo* sur le modèle de souris transgénique MMTV-ER $\alpha$ 36

Au cours du développement de la glande mammaire, les œstrogènes, jouent un rôle majeur durant la période post-pubertaire en permettant l'élongation et la ramification des canaux. En revanche, ils ne semblent pas intervenir dans le développement embryonnaire. En effet, différentes études de KO menées chez la souris ont montré que les récepteurs aux œstrogènes canoniques, ER $\alpha$ 66 et ER $\beta$ 1, ne sont pas nécessaires pour le développement pré-pubertaire de cet organe (Bocchinfuso et al., 2000). Etant donné qu'aucune information n'était disponible concernant l'implication éventuelle du variant ER $\alpha$ 36 dans les différentes étapes du développement physiologique de cette glande, nous avons étudié l'impact de l'expression de ce variant dans le tissu mammaire sain en utilisant le modèle de souris Tg MMTV-ER $\alpha$ 36 exprimant spécifiquement ce variant au niveau de la glande mammaire. L'ensemble des résultats obtenus sont décrits dans **la publication n°1**.

Lors de cette étude, nous avons dans un premier temps évalué l'impact de l'expression d'ER $\alpha$ 36 sur le développement embryonnaire et néonatal de la glande mammaire. Pour cela, une analyse quantitative de la structure de l'arbre mammaire et une étude de l'aspect histologique des canaux ont été menées sur les glandes mammaires des animaux wt et Tg MMTV-ER $\alpha$ 36 prélevées au moment du sevrage (21 jours), c'est-à-dire juste avant le début de la puberté. Plus précisément, la structure de l'arbre mammaire a été étudiée grâce à un logiciel à façon, mis au point au laboratoire par El-Hadi Djermoune (MCU, CRAN), qui permet de créer un squelette numérique à partir de l'arbre et de déterminer sa longueur, le nombre de ramifications et le nombre de bourgeons. Les résultats présentés dans les Figures 42 et 43 indiquent que la structure de l'arbre mammaire rudimentaire, qui se développe lors de la phase embryonnaire et néonatale, n'est pas affectée par l'expression d'ER $\alpha$ 36. Seule l'épaisseur de l'épithélium apparaît légèrement augmentée lorsque l'on s'intéresse à l'aspect histologique des canaux mammaires au sevrage.

A l'âge adulte (16 semaines), c'est-à-dire au terme du développement pubertaire de la glande mammaire, l'expression d'ER $\alpha$ 36 est associée à une modification de l'aspect histologique des canaux mammaires des souris vierges Tg MMTV-ER $\alpha$ 36 (Figure 44). En effet, on observe une augmentation significative du diamètre de la lumière et de l'épaisseur du stroma ainsi qu'une diminution significative de l'épaisseur de l'épithélium. De plus, nous avons mis en évidence

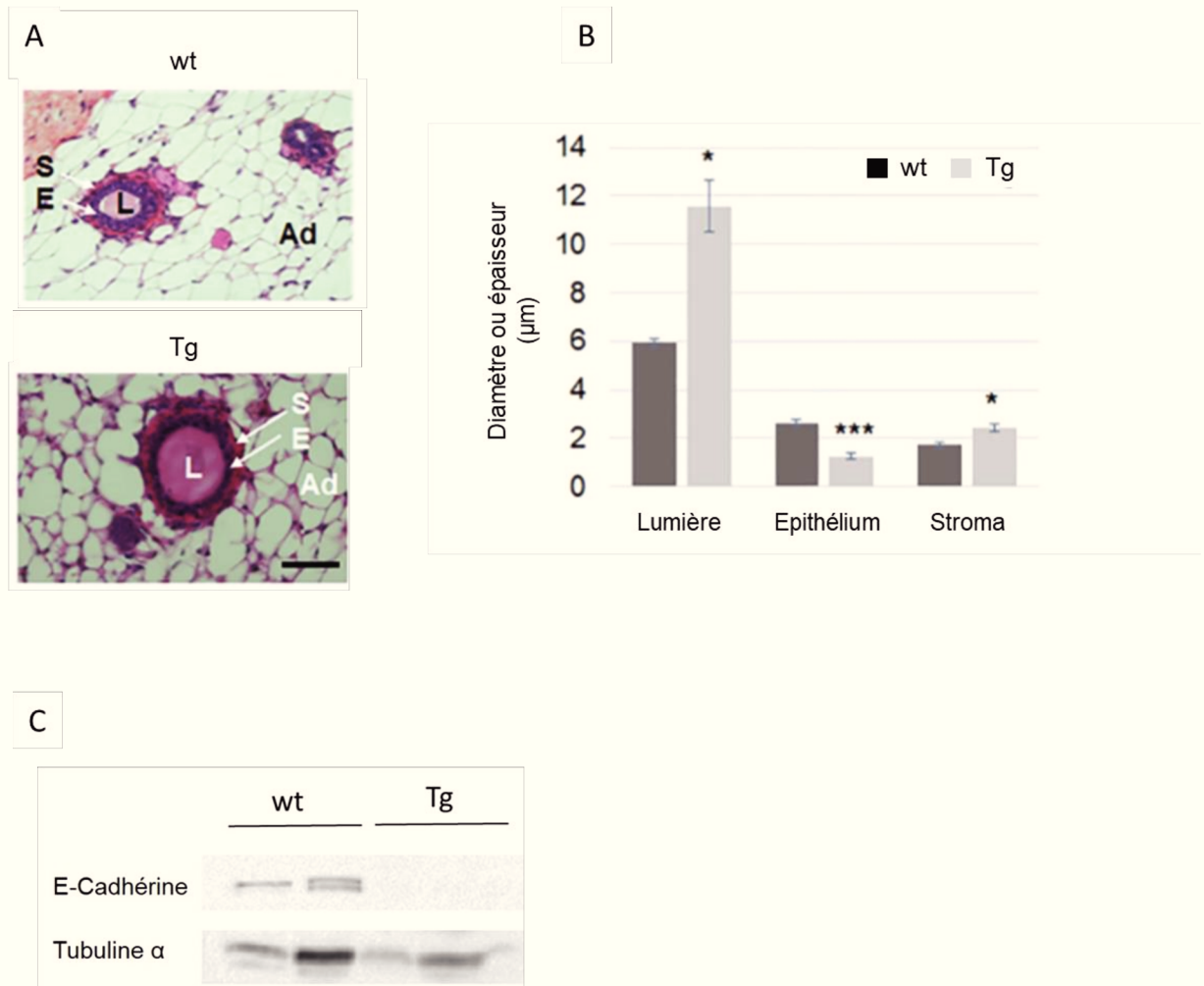


**Figure 43 : Etude de l'histologie des canaux mammaires des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 au sevrage.** (Figures 6A et 6B de la publication n°1)

(A) Coupes histologiques de glandes mammaires prélevées au moment du sevrage (21 jours) sur des souris sauvages (wt) ou transgéniques MMTV-ER $\alpha$ 36 (Tg) appartenant à la génération F4. Coloration Hématoxyline-éosine. S = Stroma ; E = Epithélium ; L = Lumière des canaux ; Ad = Adipocytes. Barre d'échelle : 15  $\mu$ m.

(B) Analyse quantitative du diamètre de la lumière, de l'épaisseur de l'épithélium et de l'épaisseur du stroma.

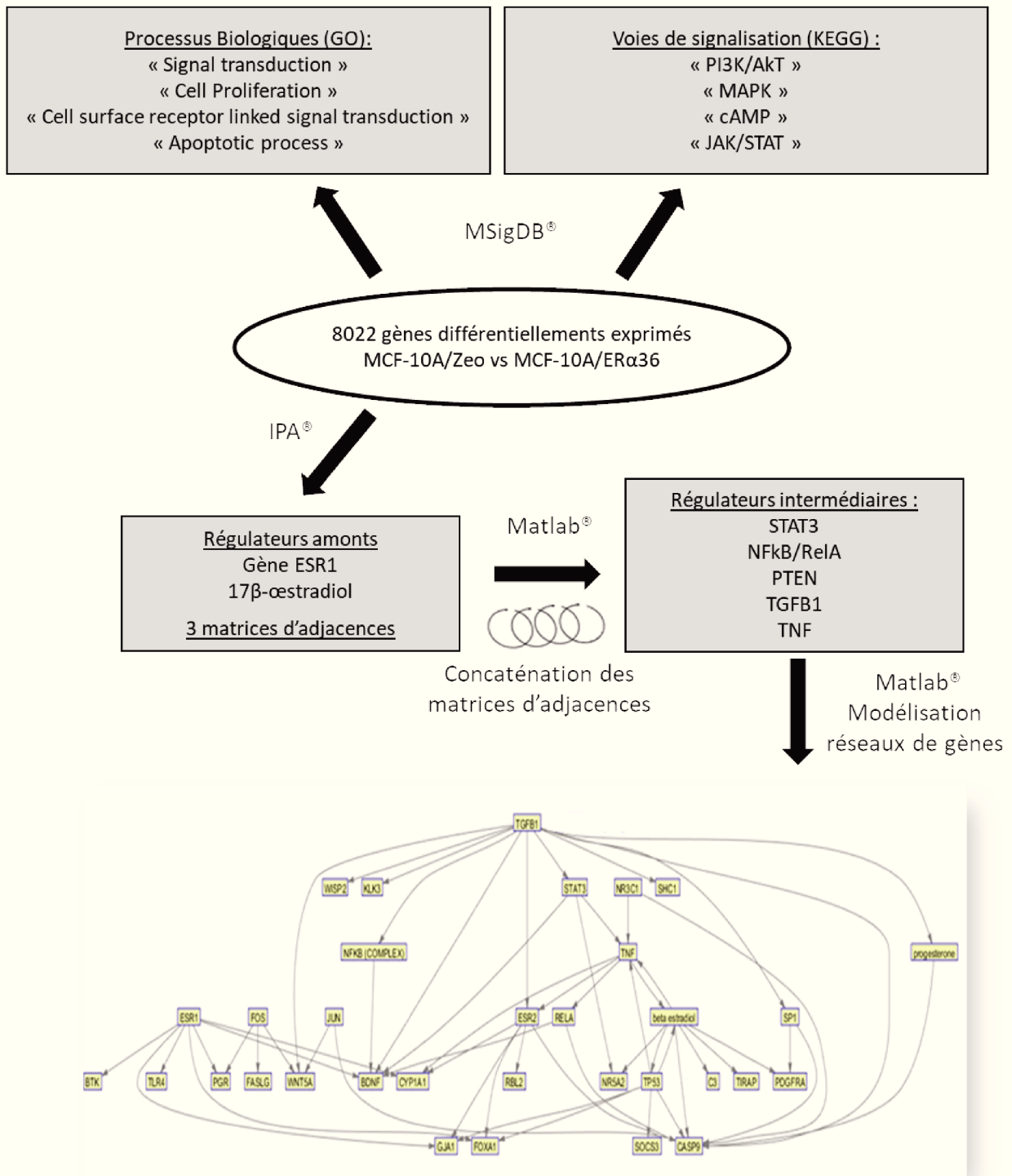
Les barres d'erreur correspondent à l'erreur standard à la moyenne. N>5 par groupe.\*: P < 0,05.



**Figure 44 : Etude de l'histologie des canaux mammaires des animaux sauvages et transgéniques MMTV/ERα36 à l'âge adulte.** (Figures 6A et 6C, Figure supplémentaire S6 de la publication n°1)

- (A) Coupes histologiques de glandes mammaires prélevées à l'âge adulte (16 semaines) sur des souris sauvages (wt) ou transgéniques MMTV-ERα36 (Tg) appartenant à la génération F4. Coloration Hématoxyline-éosine. S = Stroma ; E = Epithélium ; L = Lumière des canaux ; Ad = Adipocytes. Barre d'échelle : 15 µm.
- (B) Analyse quantitative du diamètre de la lumière, de l'épaisseur de l'épithélium et de l'épaisseur du stroma.
- (C) Western blot représentatif de l'expression de la E-cadhérine dans la glande mammaire des animaux wt et Tg. La tubuline α a été utilisée comme invariant.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. N>5 par groupe.\*: P < 0,05; \*\*\*: P < 0,001.



**Figure 45 : Stratégie d'enrichissement des données à partir du transcriptome des cellules MCF-10A/ERα36 et MCF-10A/Zeo.** (Figure supplémentaire S4 de la publication n°1)

L'ensemble de la stratégie est détaillée dans le paragraphe « Analyse du transcriptome des cellules MCF-10A/ERα36 et MCF-10A/Zeo »

une perte d'expression de la protéine de jonction E-cadhérine dans l'épithélium mammaire des animaux transgéniques. L'étude quantitative de l'arbre mammaire n'a pas pu être réalisée à l'âge adulte car à ce stade de développement l'arbre mammaire est très étendu et le logiciel utilisé ne permet pas de mesurer précisément les différents paramètres étudiés.

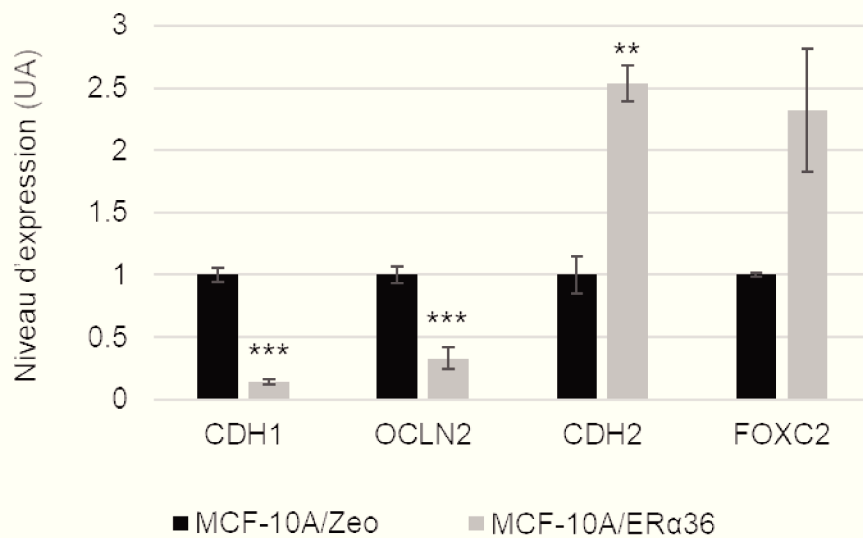
Ainsi, l'ensemble de ces résultats indiquent qu'en réponse aux œstrogènes endogènes produits à partir de la puberté, l'activation d'une signalisation œstrogénique dépendante d'ER $\alpha$ 36, conduit à une altération de l'aspect histologique post-pubertaire des canaux mammaires. Ces données rappellent ce qui est observé dans l'étude de Bocchinfuso et al., lors d'une hyperstimulation de la réponse œstrogénique mammaire chez des souris wt traitées à l'E<sub>2</sub>. Les modifications observées et plus particulièrement l'amincissement de l'épithélium et la perte d'expression de la E-cadhérine pourraient suggérer qu'ER $\alpha$ 36 joue un rôle dans le contrôle de la prolifération et des capacités migratoires des cellules épithéliales mammaires.

#### ii) Etudes *in vitro* et *in silico* sur les cellules épithéliales mammaires saines MCF-10A

Afin de décrypter les fonctions cellulaires et les voies de signalisation impactées par l'expression d'ER $\alpha$ 36 dans les cellules épithéliales mammaires, nous avons travaillé *in vitro* sur la lignée de cellules épithéliales mammaires saines immortalisées MCF-10A. Etant donné qu'aucune expression endogène de ce variant n'est détectée dans cette lignée, nous avons établi au laboratoire 2 lignées stables de cellules MCF-10A surexprimant ou non ER $\alpha$ 36 : MCF-10A/ER $\alpha$ 36 et MCF-10A/Zeo.

##### ➤ Analyse du transcriptome des cellules MCF-10A/ER $\alpha$ 36 et MCF-10A/Zeo

Dans un premier temps, nous avons réalisé une analyse transcriptomique, par puce Affymetrix GeneChip U133, sur les cellules MCF-10A/ER $\alpha$ 36 et MCF-10A/Zeo. Grâce à cette analyse, nous avons mis en évidence 8022 gènes dont l'expression varie significativement ( $p < 0,05$ ) entre nos 2 lignées cellulaires (Figure 45). Afin d'identifier les principaux processus biologiques et les principales voies de signalisation impactés par l'expression d'ER $\alpha$ 36, la liste des gènes dérégulés a été soumise respectivement dans les bases de données MSigDB (<http://software.broadinstitute.org/gsea/msigdb>) et KEGG (<https://www.genome.jp/kegg/>). Les résultats de cette analyse fonctionnelle indiquent que ces gènes dérégulés sont principalement associés aux termes GO : « Signal transduction », « cell proliferation », « cell surface receptor

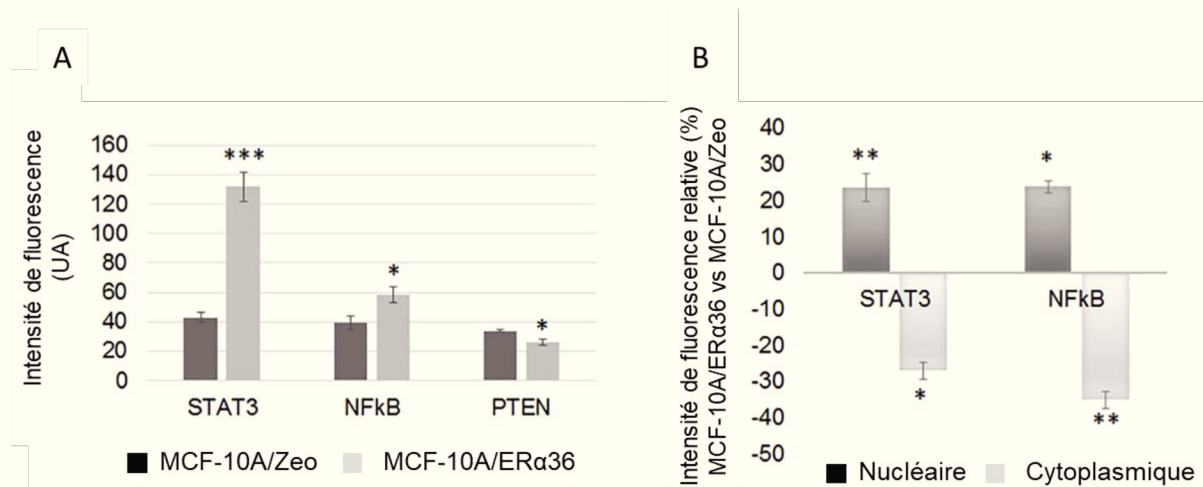


**Figure 46 : Impact d'ERα36 sur l'expression de marqueurs associés à la transition épithélio-mésenchymateuse.**

Représentation du niveau d'expression des gènes CDH1, OCLN2, CDH2 et FOXC2 dans le transcriptome des cellules MCF-10A/Zeo et des cellules MCF-10A/ERα36.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. N=3. \*\*: P < 0,01; \*\*\*: P < 0,001. UA = unité arbitraire.

**CDH1** = E-cadhérine, **CDH2** = N-cadhérine, **OCLN2** = Occludine 2, **FOXC2** = Forkhead box protein C2.



**Figure 47 : Implication des facteurs STAT3, NF-κB et PTEN dans la signalisation induite par la surexpression d'ERα36 dans les cellules MCF-10A.** (Figure 1 de la publication n°1)

(A) L'intensité de fluorescence, reflétant le niveau d'expression de STAT3, NF-κB et PTEN, a été mesurée par immunocytofluorescence dans les cellules MCF-10A/Zeo et MCF-10A/ERα36. On observe une augmentation significative de l'expression des facteurs STAT3 et NF-κB ainsi qu'une diminution significative de l'expression de PTEN dans les cellules MCF-10A/ERα36.

(B) Localisation intracellulaire (nucléaire ou cytoplasmique) des facteurs STAT3 et NF-κB étudiée par immunocytofluorescence dans les cellules MCF-10A/ERα36 par rapport aux cellules MCF-10A/Zeo. On observe une relocalisation nucléaire significative des facteurs STAT3 et NF-κB dans les cellules MCF-10A/ERα36.

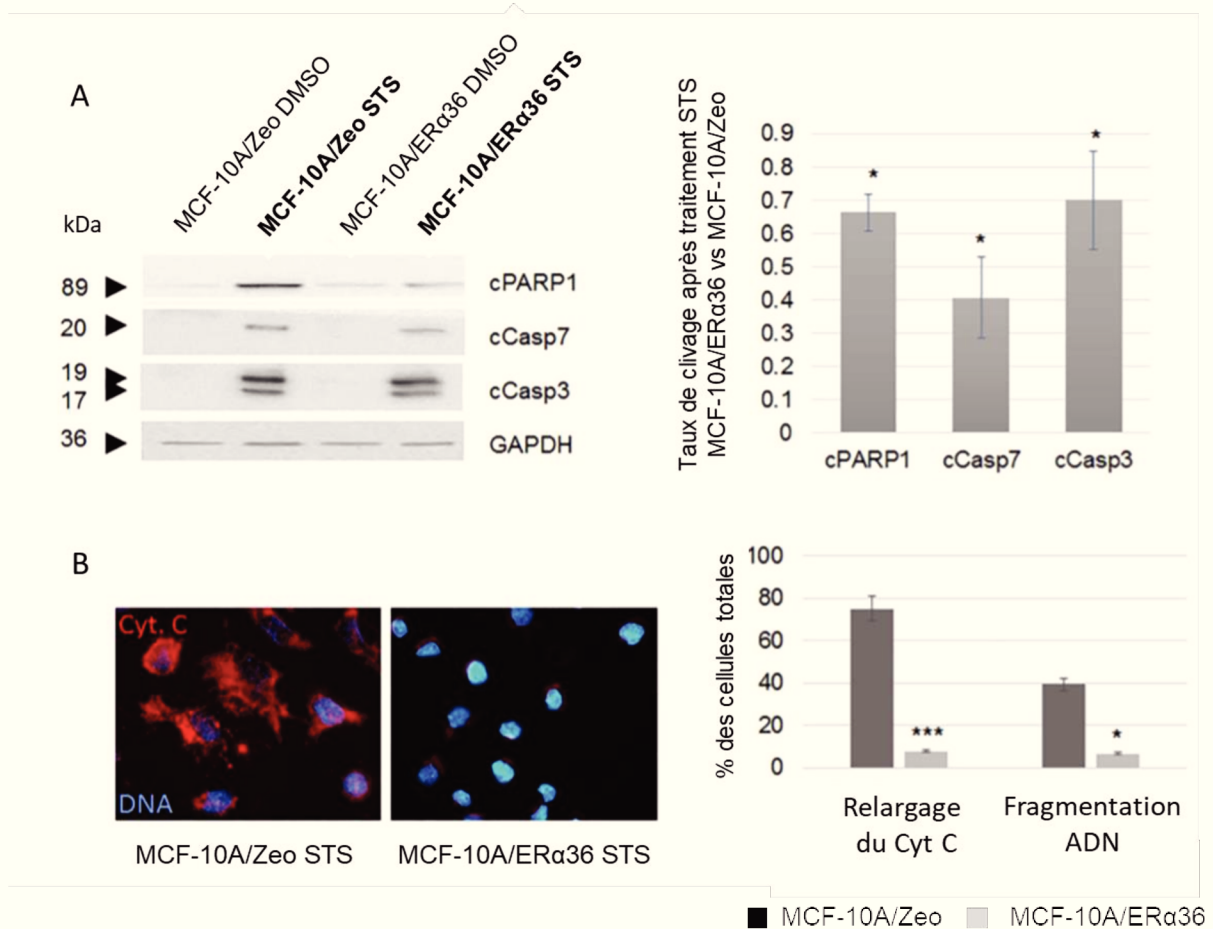
Les barres d'erreur correspondent à l'erreur standard à la moyenne. N=3. \*: P < 0,05; \*\*: P < 0,01; \*\*\*: P < 0,001. UA = unité arbitraire.

linked signal transduction » et « apoptotic process » ( $p$ -value  $< 1 \times 10^{-9}$ ). Concernant les voies de signalisation KEGG, l'expression d'ER $\alpha$ 36 semble avoir un impact sur les voies « PI3K/AkT Signaling Pathway », « MAPK Signaling Pathway », « cAMP Signaling Pathway » et « Jak/STAT Signaling Pathway » ( $p$ -value  $< 1 \times 10^{-3}$ ). De manière intéressante, ces processus biologiques et ces voies de signalisation sont fréquemment décrites dans la littérature pour leur implication dans la carcinogénèse mammaire (Velloso et al., 2017). De plus, parmi les gènes dérégulés on trouve quelques marqueurs associés à la transition épithélio-mésenchymateuse (Asiedu et al., 2011). En effet, dans les cellules MCF-10A/ER $\alpha$ 36, on observe une diminution significative des marqueurs épithéliaux CDH1 et Ocln qui s'accompagne d'une augmentation significative de l'expression du marqueur mésenchymateux CDH2. On observe également une tendance à l'augmentation du marqueur mésenchymateux Foxc2 (Figure 46).

Par la suite, nous avons souhaité déterminer plus précisément, quels sont les facteurs intermédiaires qui interviennent, de façon dépendante d'ER $\alpha$ 36, dans la régulation de l'expression des 8022 gènes différentiellement exprimés dans les lignées MCF-10A/ER $\alpha$ 36 et MCF-10A/Zeo.

Pour cela, nous avons utilisé successivement (i) le logiciel IPA® (Ingenuity Pathway Analysis) qui a permis d'obtenir des matrices d'adjacences décrivant les 3 types d'interactions fonctionnelles (modulation, activation, inhibition) pouvant exister entre les 8022 gènes et (ii) un programme à façon Matlab®, mis au point au laboratoire, qui a permis de modéliser des réseaux de gènes représentatifs des interdépendances fonctionnelles qui existent entre les gènes impliqués dans les 4 processus biologiques mis en évidence précédemment (Figure 45). Parmi les facteurs intermédiaires identifiés comme étant communs à l'ensemble des fonctions biologiques, on trouve STAT3, NF- $\kappa$ B (p65), PTEN, TGF $\beta$ 1 et TNF $\alpha$ . L'implication des facteurs STAT3, NF- $\kappa$ B (p65) et PTEN dans la signalisation induite par la surexpression d'ER $\alpha$ 36 a été validée *in vitro* en étudiant le niveau d'expression et la localisation intracellulaire (cytoplasme/noyau) de ces facteurs par immunocytochimie dans les cellules MCF-10A/Zeo et MCF-10A/ER $\alpha$ 36 (Figure 47). Ce résultat suggère, que les effets d'ER $\alpha$ 36 dans les cellules épithéliales mammaires saines impliquent préférentiellement les voies de signalisation NF- $\kappa$ B et JAK2/STAT3.





**Figure 48 : Impact de la surexpression d'ERα36 sur la réponse apoptotique des cellules MCF-10A.** (Figure 3 de la publication n°1)

La mort cellulaire par apoptose des cellules MCF-10A/Zeo et MCF-10A/ERα36 a été induite par un traitement de 6 heures avec de la Staurosporine (STS) à 0,25 μM.

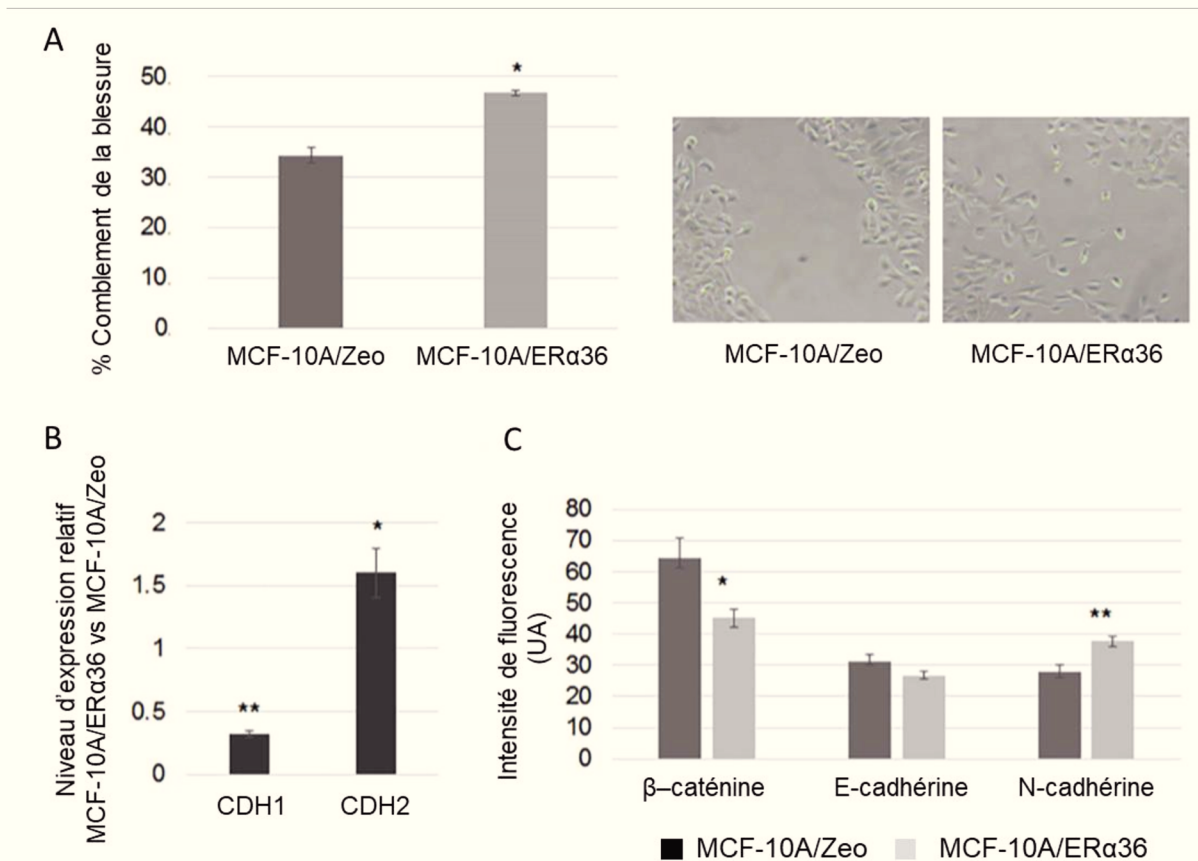
- (A) Clivage de PARP1 (cPARP1) et des caspases 3 (cCasp3) et 7 (cCasp7) après traitement à la STS, dans les cellules MCF-10A/Zeo et MCF-10A/ERα36. Une image représentative de western blot et une quantification du taux de clivage relatif de ces différents marqueurs apoptotiques dans les cellules MCF-10A/ERα36 par rapport aux cellules MCF-10A/Zeo sont présentées. La GAPDH a été utilisée comme invariant. On observe une diminution significative du taux de clivage de PARP1, Casp3 et Casp7 dans les cellules MCF-10A/ERα36 traitées à la STS par rapport aux cellules MCF-10A/Zeo.
- (B) Etude du relargage du cytochrome C (Cyt C) par immunocytofluorescence et de la fragmentation de l'ADN par la technique TUNEL, après traitement à la STS des cellules MCF-10A/Zeo et MCF-10A/ERα36. On observe une diminution significative du relargage du Cyt C et de la fragmentation de l'ADN dans les cellules MCF-10A/ERα36 traitées à la STS par rapport aux cellules MCF-10A/Zeo.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. (A) N=4, (B) N=3. \*: P < 0,05; \*\*\*: P < 0,001.

➤ Etude phénotypique des cellules MCF-10A/ER $\alpha$ 36 et MCF-10A/Zeo

Afin de valider et de préciser les données obtenues *in silico*, nous avons exploré les conséquences de la surexpression d'ER $\alpha$ 36 sur la prolifération, la migration et la réponse apoptotique des cellules épithéliales mammaires saines MCF-10A. Cette étude a permis de mettre en évidence que la signalisation dépendante d'ER $\alpha$ 36 est associée à la mise en place de mécanismes de résistances à l'apoptose induite par la staurosporine (Figure 48) et à une augmentation des capacités migratoires des cellules MCF-10A (Figure 49). De plus, nous avons montré que la surexpression d'ER $\alpha$ 36 conduit à la diminution du potentiel prolifératif des cellules MCF-10A de façon dépendante de STAT3 (Figure 50). En effet, lorsque l'on traite les cellules MCF-10A/ER $\alpha$ 36 avec un inhibiteur de STAT3, le 5,15 DPP, on restaure leur capacité à proliférer. De manière complémentaire, afin d'expliquer cet effet de STAT3, qui apparaît contradictoire étant donné que ce facteur est plutôt décrit comme un inducteur de la prolifération, il nous a paru intéressant d'explorer le rôle d'ER $\beta$  dans ce système (Banerjee et Resat, 2016). En effet, cette isoforme des récepteurs aux œstrogènes est principalement décrit dans la littérature pour son effet antiprolifératif sur les cellules cancéreuses mammaires (Haldosén et al., 2014) et une étude de Wang H.C. et al., (2011) a montré que son expression est induite par STAT3 dans le contexte des adénocarcinomes pulmonaires. Les résultats préliminaires obtenus, présentés dans la Figure 51, ont montré que l'expression d'ER $\alpha$ 36 dans les cellules MCF-10A semble conduire à une relocalisation nucléaire d'ER $\beta$ . De plus, ce phénomène apparaît atténué après un traitement avec l'inhibiteur de STAT3, le 5,15 DPP. Cela semble donc confirmer l'existence d'un lien entre ER $\alpha$ 36, STAT3 et ER $\beta$  qui conduirait à une diminution de la prolifération, déclenchée par ER $\alpha$ 36 mais dépendante d'ER $\beta$ .

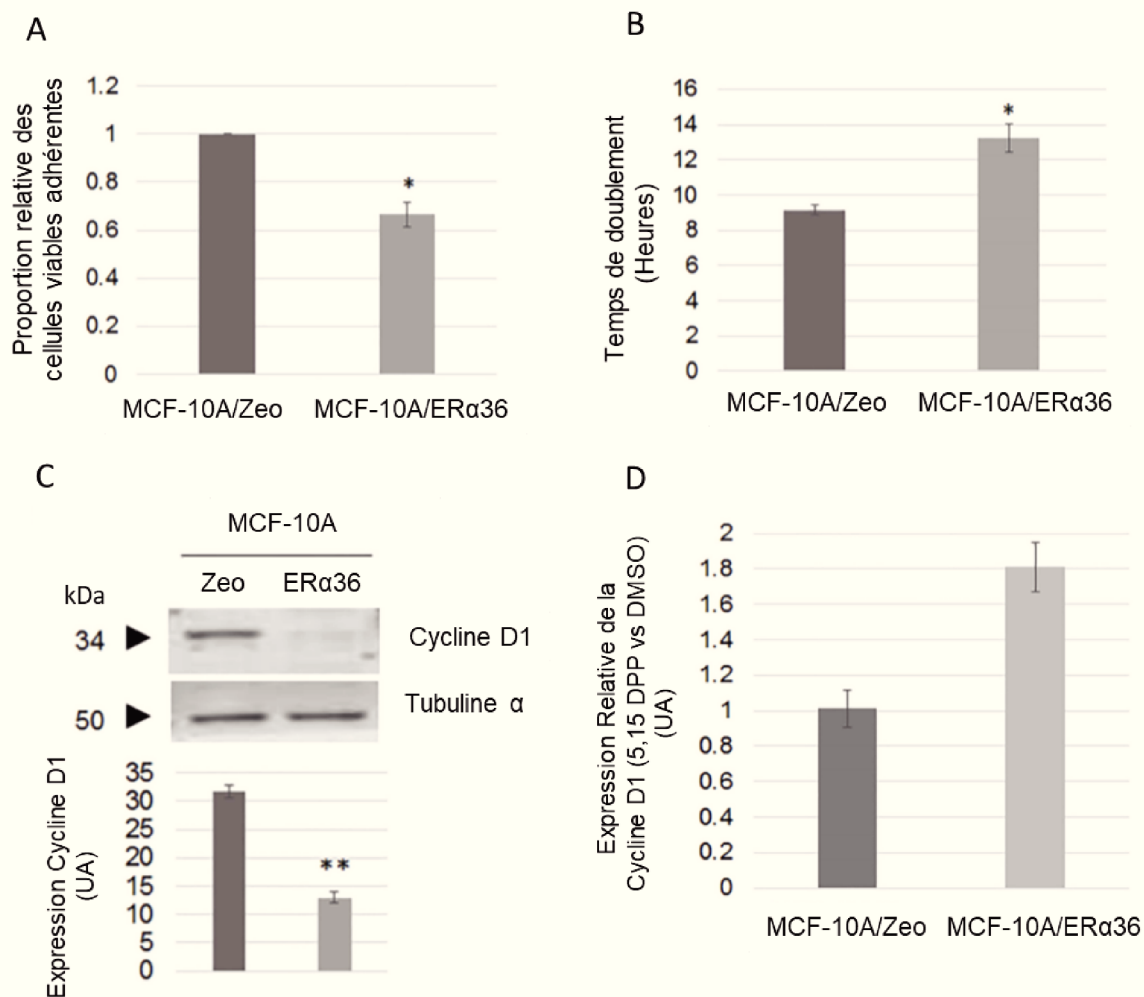
Ces résultats obtenus *in vitro* pourraient fournir des pistes pour expliquer l'aspect histologique de l'épithélium mammaire observé dans le modèle de souris Tg MMTV-ER $\alpha$ 36 *in vivo*. Tout d'abord, on peut émettre l'hypothèse que l'augmentation de l'épaisseur de l'épithélium, qui est observée au sevrage chez les animaux Tg MMTV-ER $\alpha$ 36, pourrait refléter la mise en place de mécanismes de résistance à l'apoptose dans les cellules épithéliales mammaires exprimant ER $\alpha$ 36. En effet, l'apoptose joue un rôle majeur au cours des processus de morphogénèse en permettant le remodelage des nouvelles structures tissulaires. Ainsi, la perturbation de ce processus pourrait avoir un impact sur la mise en place de la lumière des canaux de l'arbre mammaire primitif. D'autre part, la diminution de l'épaisseur de l'épithélium qui est observée à l'âge adulte pourrait plutôt s'expliquer par le ralentissement de la division cellulaire induit



**Figure 49 : Impact de la surexpression d'ER $\alpha$ 36 sur les capacités migratoires des cellules MCF-10A.** (Figure 4 de la publication n°1)

- (A) Pourcentage de comblement d'une blessure, réalisée sur un tapis de cellules MCF-10A/Zeo ou MCF-10A/ER $\alpha$ 36 à confluence. La taille de la blessure a été mesurée à t = 0 heure et à t = 6 heures. On observe une augmentation significative du comblement de la blessure par les cellules MCF-10A/ER $\alpha$ 36 qui, comme on le voit sur la photo de microscopie, ont migré à l'intérieur de la blessure.
- (B) Niveaux d'expression des marqueurs, CDH1 et CDH2, qui sont associés à la transition épithélio-mésenchymateuse, mesurés par RT-qPCR. On observe une augmentation significative du niveau d'expression des ARNm correspondant au gène CDH2 et une diminution significative de l'expression de CDH1 dans les cellules MCF-10A/ER $\alpha$ 36 par rapport aux cellules MCF-10A/Zeo.
- (C) Niveau d'expression protéique de la  $\beta$ -caténine, de la E-cadhérine (CDH1) et de la N-cadhérine (CDH2). On observe une augmentation significative du niveau d'expression de la N-cadhérine et une diminution significative de l'expression de la  $\beta$ -caténine dans les cellules MCF-10A/ER $\alpha$ 36 par rapport aux cellules MCF-10A/Zeo.

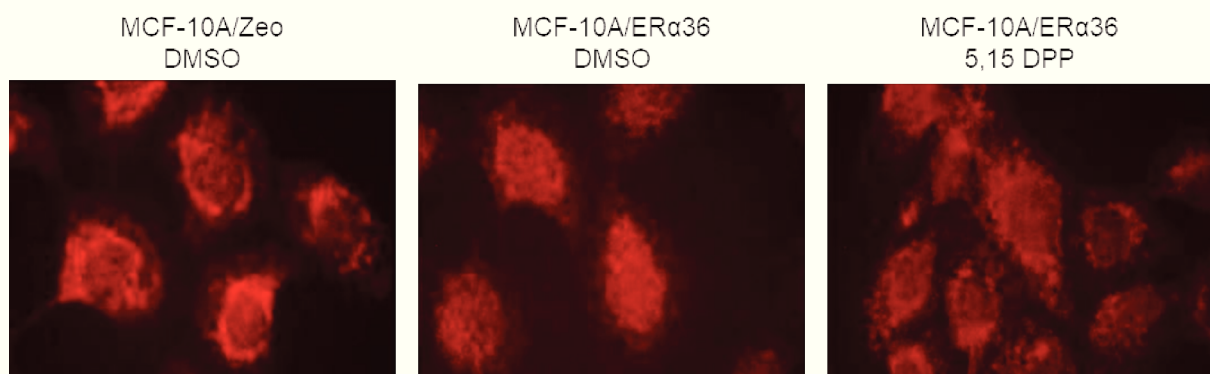
Les barres d'erreur correspondent à l'erreur standard à la moyenne. (A) N=5, (B-C) N=3.\*: P < 0,05; \*\*: P < 0,01. UA = unité arbitraire.



**Figure 50 : Impact de la surexpression d'ERα36 sur la prolifération des cellules MCF-10A.** (Figure 2 de la publication n°1)

- (A) Viabilité des cellules MCF-10A/Zeo et MCF-10A/ERα36 évaluée par la méthode du crystal violet. On observe une diminution significative du nombre de cellules MCF-10A/ERα36 viables adhérentes, par rapport aux cellules MCF-10A/Zeo. Cela pourrait refléter une diminution de la prolifération de ces cellules.
- (B) Temps de doublement des cellules MCF-10A/Zeo et MCF-10A/ERα36. On observe une augmentation significative du temps de doublement des cellules MCF-10A/ERα36 par rapport aux cellules MCF-10A/Zeo.
- (C) Niveau d'expression de la protéine Cycline D1 mesuré par western blot. La tubuline α a été utilisée comme invariant. On observe une diminution significative du niveau d'expression de la cycline D1 dans les cellules MCF-10A/ERα36 par rapport aux cellules MCF-10A/Zeo.
- (D) Impact de l'inhibiteur de STAT3 (5,15 DPP) sur le niveau d'expression de la Cycline D1. Dans les cellules MCF-10A/ERα36, l'inhibition de STAT3 est associée à une augmentation de l'expression de la cycline D1.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. (A,C,D) N=3, (B) N=4.\*: P < 0,05; \*\*: P < 0,01. UA = unité arbitraire.



**Figure 51 : Effet de la surexpression d'ER $\alpha$ 36 sur la localisation intracellulaire d'ER $\beta$ .**

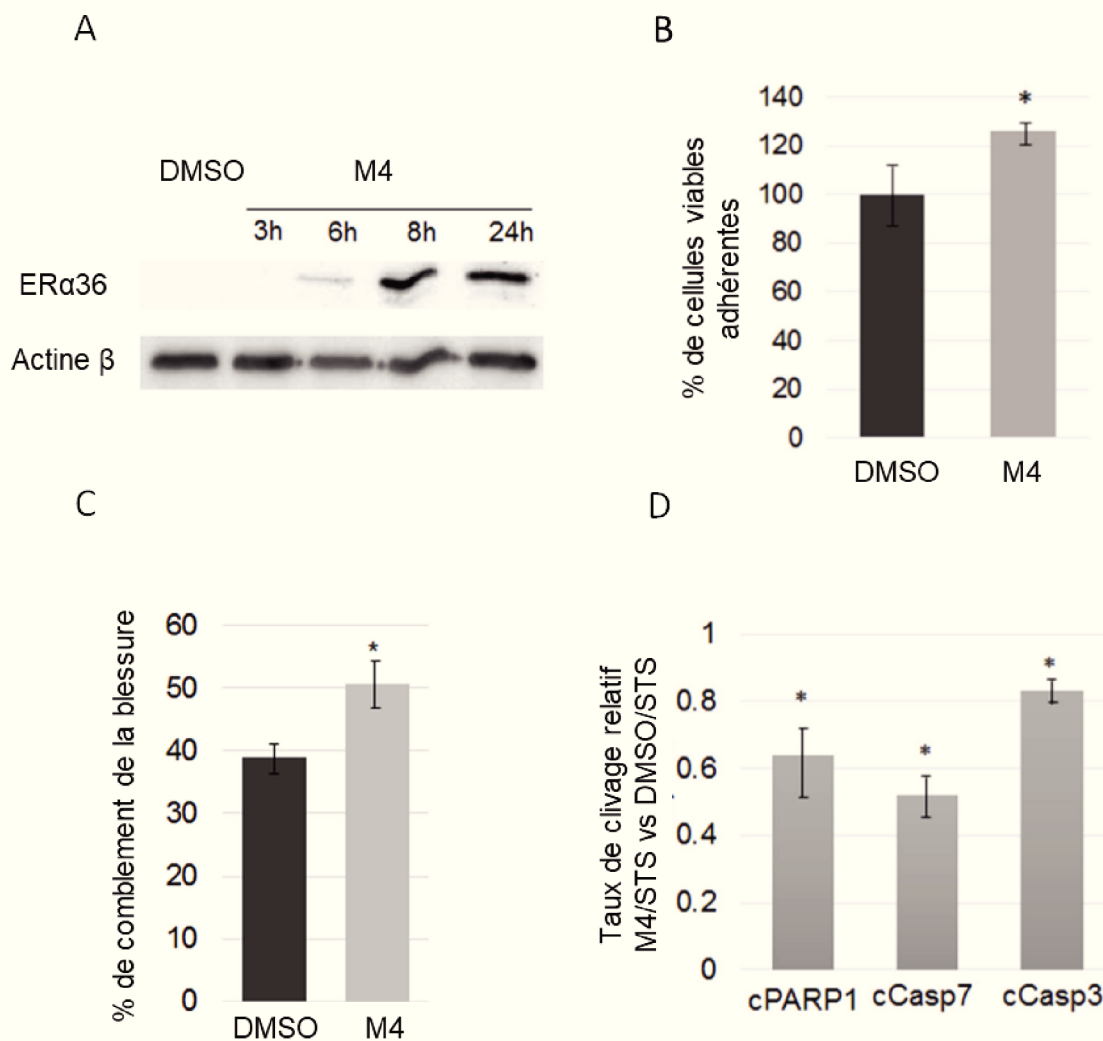
Etude par immunocytofluorescence de la localisation intracellulaire d'ER $\beta$  dans les cellules MCF-10A/Zeo exposées au DMSO 0,01% et dans les cellules MCF-10A/ER $\alpha$ 36 exposées au DMSO 0,01% ou à un inhibiteur de STAT3, le 5,15 DPP à 15  $\mu$ M, pendant 24h.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

On observe que la surexpression d'ER $\alpha$ 36 conduit à une relocalisation nucléaire d'ER $\beta$ . Toutefois cet effet semble atténué par le traitement avec du 5,15 DPP.

Les photos présentées sont représentatives de 3 expériences indépendantes.

Anticorps utilisé : anti-ER $\beta$  (Sc-8974, Santa Cruz)



**Figure 52 : Impact du mélange M4 sur la prolifération, la migration et la survie des cellules MCF-10A.** (Figure 1A, 2B, 3A, 4A de la publication n°2)

- (A) Western blot représentatif de l'induction de l'expression d'ER $\alpha$ 36 dans les cellules MCF-10A après des expositions de 3, 6, 8 ou 16 heures au mélange M4 1 nM. L'actine  $\beta$  a été utilisée comme invariant.
- (B) Viabilité des cellules MCF-10A, exposées au DMSO 0,01% ou au M4 1 nM pendant 24 heures, évaluée grâce à la méthode du crystal violet. On observe une augmentation significative du nombre de cellules MCF-10A viables adhérentes après traitement au M4. Cela pourrait refléter une augmentation de la prolifération de ces cellules.
- (C) Capacités migratoires des cellules MCF-10A traitées au DMSO 0,01% ou au M4 1 nM. Le pourcentage de comblement de la blessure a été calculé 6 heures après la réalisation d'une blessure sur un tapis de cellules MCF-10A à confluence. On observe que l'exposition au M4 favorise le comblement de la blessure par les cellules MCF-10A qui acquièrent des capacités migratoires.
- (D) Clivage de PARP1 (cPARP1) et des caspases 3 (cCasp3) et 7 (cCasp7), après un traitement de 6 heures à la staurosporine (STS) 0,25  $\mu$ M, des cellules MCF-10A prétraitées au DMSO 0,01% ou au M4 1 nM pendant 24 heures. On observe que l'exposition au M4 conduit au développement d'une résistance à l'apoptose induite par la STS dans les cellules MCF-10A.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. (B) N=3, (C) N=5, (D) N=4.\*: P < 0,05.

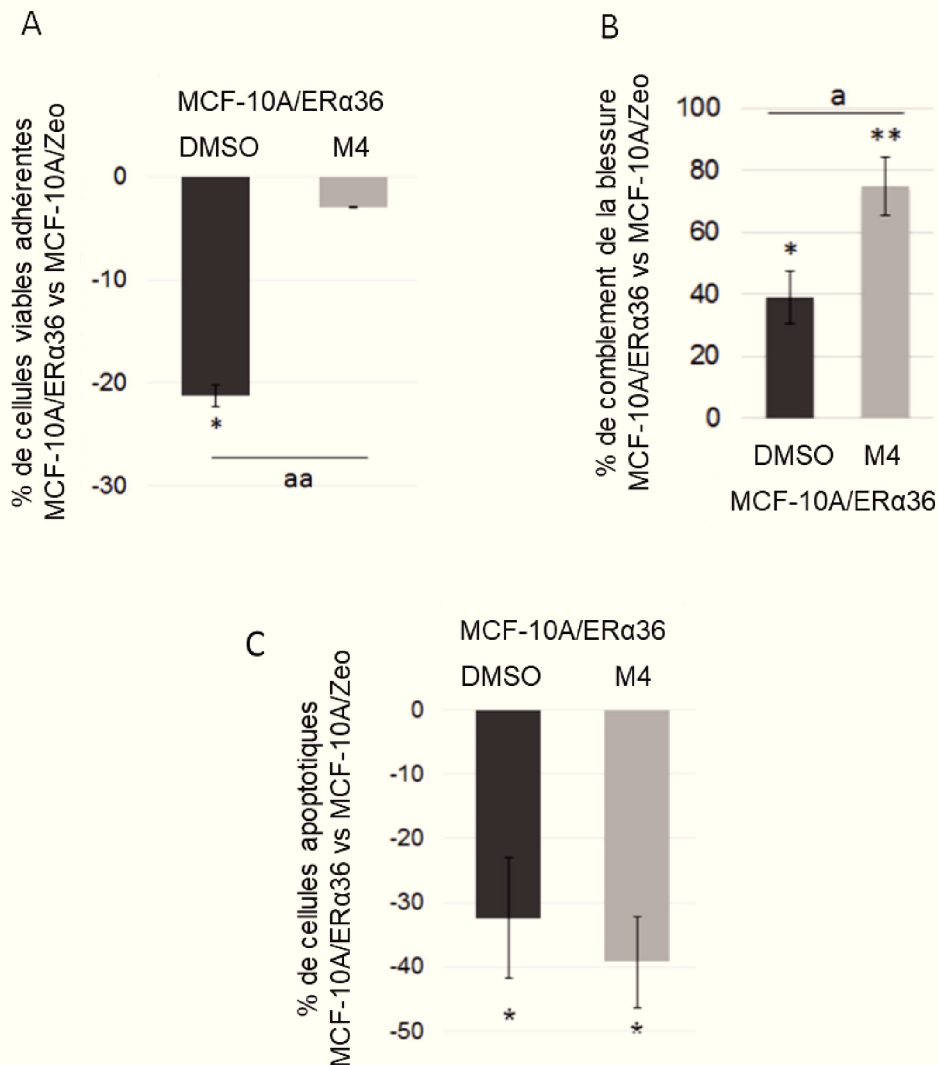
par l'expression d'ER $\alpha$ 36. En effet, à ce stade, les structures de l'arbre mammaire sont déjà mises en place et leur maintien dépend majoritairement de la capacité des cellules épithéliales à se renouveler grâce à un équilibre établi entre senescence et prolifération. Afin de valider ces hypothèses il serait intéressant de distinguer sur les coupes histologiques, les cellules apoptotiques et les cellules prolifératives en réalisant respectivement un marquage TUNEL et un marquage Ki67. Enfin, concernant la désorganisation de l'épithélium qui est observée à l'âge adulte, elle pourrait être liée à la perte de l'identité épithéliale des cellules qui bordent les canaux et à l'acquisition de propriétés migratoires.

Finalement, l'induction de la migration et de la survie cellulaire, qui sont 2 caractéristiques majeures des cellules cancéreuses, suggère que l'expression d'ER $\alpha$ 36 joue un rôle dans l'acquisition de propriétés qui rappellent une transformation néoplasique de l'épithélium mammaire. Dans la publication N°2, nous avons pourtant montré que lorsque que l'on greffe des cellules MCF-10A/ER $\alpha$ 36 ou MCF-10A/Zeo dans le coussin adipeux mammaire de souris nude, aucune tumeur ne se développe quelle que soit la lignée. Cela suggère que l'expression d'ER $\alpha$ 36 dans les cellules épithéliales mammaires n'est pas suffisante à elle seule pour induire la formation d'une tumeur.

➤ Impact des alkylphénols sur la signalisation dépendante d'ER $\alpha$ 36

D'autres études, menées en parallèle au laboratoire et décrites dans la première partie de la publication N°2, ont montré que les alkylphénols induisent l'expression d'ER $\alpha$ 36 et stimulent la prolifération, la survie et la migration des cellules épithéliales mammaires saines MCF-10A (Figure 52). Suite à ces observations, nous avons voulu aller plus loin en étudiant in vitro les effets du M4 sur la prolifération, la migration et la survie des cellules MCF-10A/ER $\alpha$ 36. En effet, ce mélange est suspecté de moduler la signalisation dépendante d'ER $\alpha$ 36 et notamment la voie JAK/STAT. L'ensemble des résultats obtenus sont décrits dans **la publication N°2**.

Les résultats de ces études montrent qu'une exposition au M4 conduit à la restauration du potentiel prolifératif des cellules MCF-10A/ER $\alpha$ 36, qui après traitement, présentent un taux de prolifération similaire aux cellules MCF-10A/Zeo (Figure 53A). Cela rappelle les résultats obtenus lors de l'étude transcriptomique de la lignée germinale testiculaire où nous avons observé que l'exposition des animaux Tg Oct4-ER $\alpha$ 36 au M4 semble atténuer les effets du transgène sur le processus biologique lié au cycle cellulaire. De plus, nous avons mis en évidence qu'après exposition au M4, le phénotype de résistance à l'apoptose des cellules MCF-



**Figure 53 : Impact du mélange M4 sur la prolifération, la migration et la survie des cellules MCF-10A/ERα36.** (Figure 4B, 4D, 4E de la publication n°2)

- (A) Viabilité des cellules MCF-10A/Zeo et MCF-10A/ERα36, exposées au DMSO 0,01% ou au M4 1 nM pendant 48 heures, évaluée grâce à la méthode du crystal violet. On observe une augmentation significative du nombre de cellules MCF-10A/ERα36 viables adhérentes après traitement au M4. Cela pourrait refléter que le M4 restaure le potentiel prolifératif des cellules MCF-10A qui était réprimé par la surexpression d'ERα36.
- (B) Capacités migratoires des cellules MCF-10A/Zeo et MCF-10A/ERα36 traitées avec du DMSO 0,01% ou du M4 1 nM. Le pourcentage de comblement de la blessure a été calculé 6 heures après la réalisation d'une blessure sur un tapis de cellules MCF-10A/Zeo ou MCF-10A/ERα36 à confluence. On observe que l'exposition au M4 potentialise le comblement de la blessure c'est-à-dire les capacités migratoires des cellules MCF-10A, qui étaient déjà induites par la surexpression d'ERα36.
- (C) Etude de la résistance à l'apoptose induite par la staurosporine (0,25 μM, 6 heures), dans les cellules MCF-10A/Zeo et MCF-10A/ERα36 prétraitées avec du DMSO 0,01% ou du M4 1 nM pendant 24h. Le pourcentage de cellules apoptotiques a été déterminé par la technique TUNEL. On observe que l'exposition au M4 n'impacte pas de manière significative la résistance à l'apoptose qui était déjà induite par la surexpression d'ERα36 dans les cellules MCF-10A.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

\* = MCF-10A/ERα36 vs MCF-10A ; a = MCF-10A/ERα36 M4 vs MCF-10A/ERα36 DMSO.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. (A, C) N = 5, (B) N=3.\* ou a: P < 0,05; \*\* ou aa : P < 0,01.



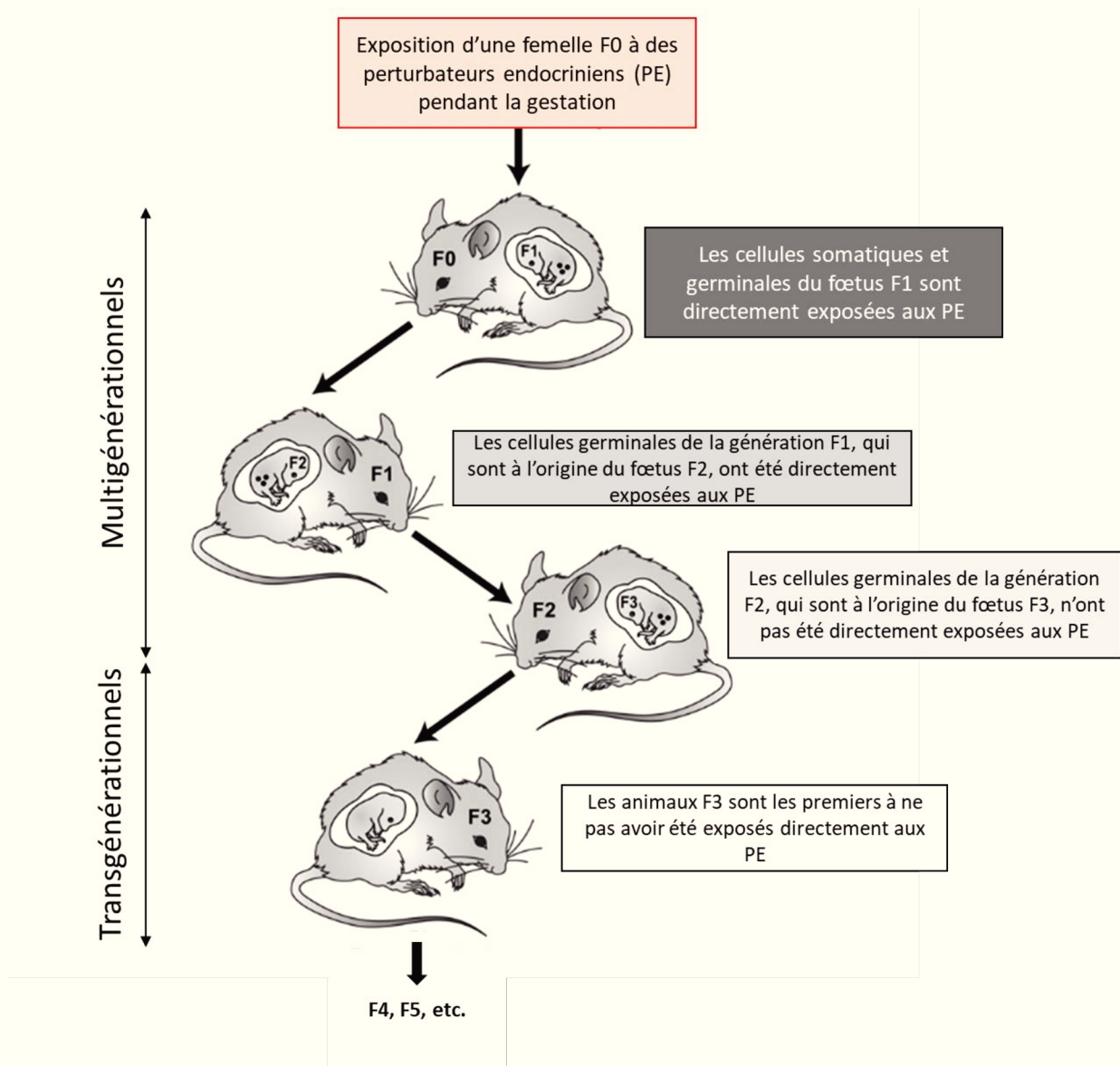
10A/ER $\alpha$ 36 est maintenu et les capacités migratoires de ces cellules sont augmentées. (Figures 53B et C). Autrement dit, les alkylphénols permettent de maintenir voire de renforcer les phénotypes pro-tumoraux (survie et migration) liés à l'expression d'ER $\alpha$ 36 tout en annulant l'effet antiprolifératif de l'expression de ce variant dans les cellules épithéliales mammaires saines MCF-10A.

En parallèle, un nombre croissant d'études indiquent qu'une exposition fœtale ou néonatale à des perturbateurs endocriniens œstrogène-mimétiques, comme le bisphénol A ou la vinclozoline, perturbe le développement physiologique de la glande mammaire et pourrait accroître le risque d'apparition d'un cancer du sein à l'âge adulte (Nilsson et al., 2008 ; Manikkam et al., 2012 ; Dhimolea et al., 2014). Cela suggère qu'une hyperstimulation ou un déclenchement précoce et aberrant de la signalisation œstrogénique par des perturbateurs endocriniens peut jouer un rôle dans la transformation néoplasique de l'épithélium mammaire. Il a également été clairement démontré que les perturbateurs endocriniens peuvent perturber le fonctionnement de la machinerie épigénétique et ainsi altérer la programmation épigénétique fœtale. Ce mécanisme d'action est particulièrement problématique, car si ces modifications de l'épigénome atteignent la lignée germinale, elles peuvent être associées à l'apparition d'effets transgénérationnels (Figure 54).

Etant donné que le mélange M4 est associé à une diminution de l'expression de l'effecteur épigénétique DNMT3L dans les cellules séminomateuses TCam-2 (Ajj et al., 2013), nous avons souhaité évaluer si une exposition fœtale à ce mélange pourrait avoir un impact transgénérationnel sur le développement de la glande mammaire et si cet effet pourrait être modulé par l'expression d'ER $\alpha$ 36.

Afin de répondre à cette problématique, nous avons étudié l'aspect histologique des canaux mammaires des animaux wt ou Tg MMTV-ER $\alpha$ 36, constituant la 3<sup>ème</sup> génération (F3) de descendants de femelles exposées quotidiennement au M4 (0,5 $\mu$ g/kg/jour) durant toute la durée de la gestation. En effet, ces animaux correspondent à la première génération n'ayant pas été directement exposée au mélange M4, mais leur développement pourrait être impacté par de potentiels effets transgénérationnels de ces perturbateurs endocriniens.

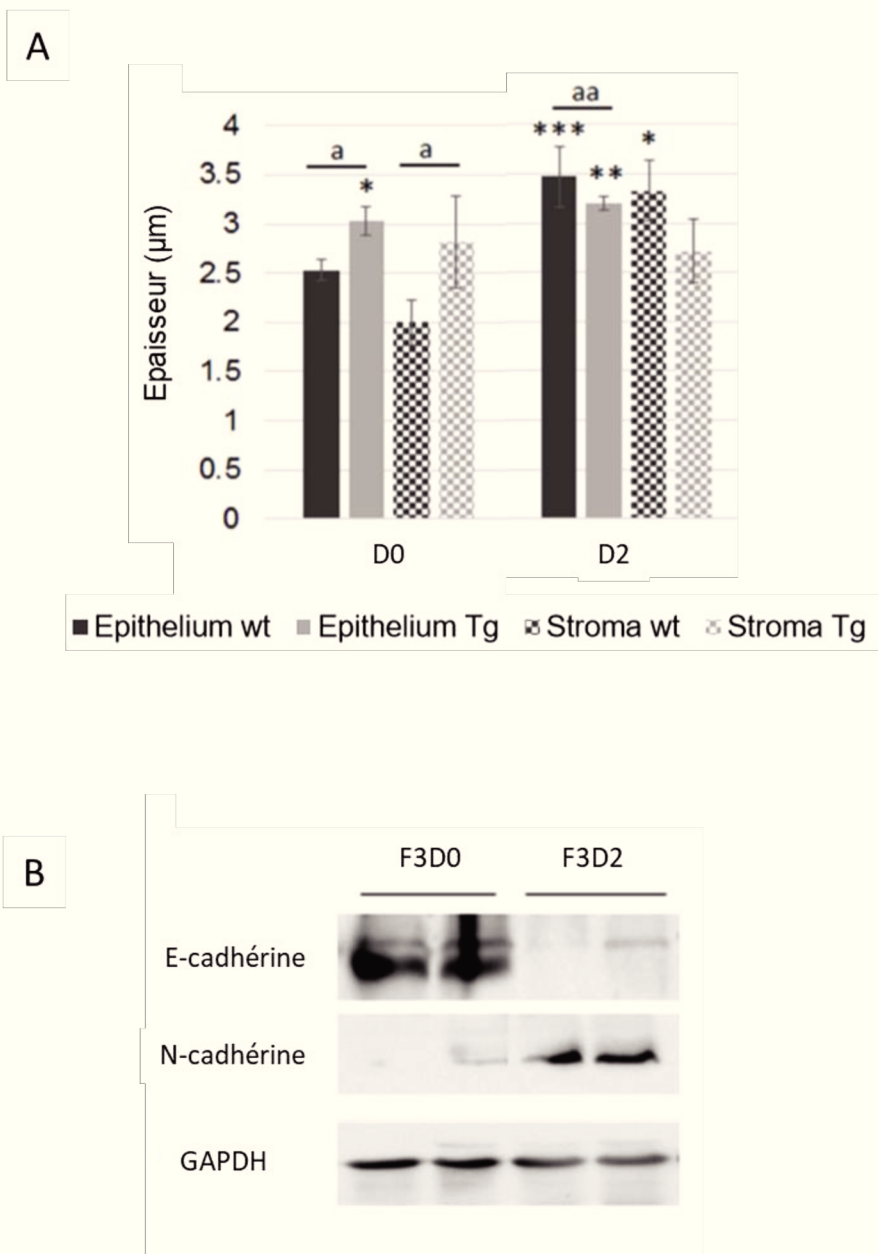
Les résultats obtenus au sevrage (PND21), indiquent que chez les animaux wt, l'exposition au mélange M4 (D2 = 0,5 $\mu$ g/kg/jour) conduit à l'apparition d'effets transgénérationnels avec une modification de l'aspect histologique des canaux mammaires (Figure 55). Chez ces animaux, on observe une augmentation de l'épaisseur du stroma et de l'épithélium qui s'accompagne



**Figure 54 : Etude des effets multigénérationnels et transgénérationnels des perturbateurs endocriniens.**

La femelle gestante F0 ainsi que sa descendance F1 et la lignée germinale de cette descendance, à l'origine de la génération F2, sont directement exposées aux perturbateurs endocriniens (PE). Dans ce cas, les effets des PE sur ces organismes sont dits « multigénérationnels ». La génération F3, formée à partir des cellules germinales de la F2, n'a pas été exposée directement aux PE. C'est également le cas des générations suivantes F4, F5, etc. Les effets sur cette descendance sont donc appelés « transgénérationnels ».

(D'après Schug et al., 2011)

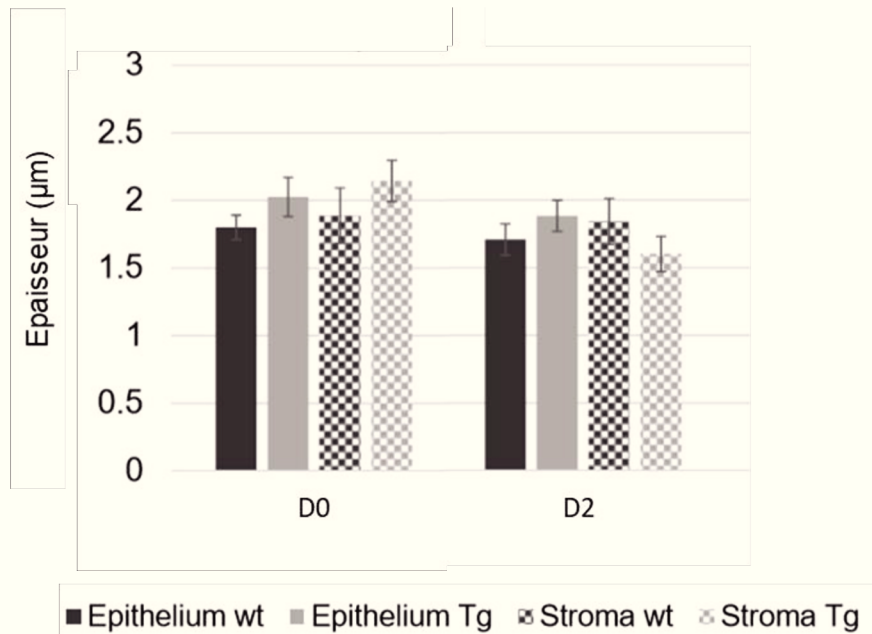


**Figure 55 : Etude des effets transgénérationnels du mélange M4 sur l'histologie des canaux mammaires des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 au sevrage.** (Figure 5A, 5C de la publication n°2)

(A) Variation de l'épaisseur du stroma et de l'épithélium, mesurée sur des coupes histologiques de glandes mammaires prélevées au sevrage (21 jours), sur des souris sauvages (wt) ou transgéniques MMTV-ER $\alpha$ 36 (Tg) appartenant à la génération F3. Ces animaux ont été exposés (D2) ou non (D0), de manière transgénérationnelle à une dose de 0,5  $\mu$ g/kg/jour de M4 (dose correspondant à l'exposition humaine).

(B) Western blot représentatif de l'expression de la E-cadhérine et de la N-cadhérine dans la glande mammaire des animaux wt, appartenant à la génération F3 qui a été exposée (D2) ou non (D0), de manière transgénérationnelle à une dose de 0,5  $\mu$ g/kg/jour de M4 . La GAPDH a été utilisée comme invariant.

\* = significativement différent des animaux D0 wt; a = différence significative entre les animaux wt et Tg exposés à la même dose. Les barres d'erreur correspondent à l'erreur standard à la moyenne. (D0 wt) N = 7, (D0 Tg) N = 5, (D2 wt) N = 4 et (D2 Tg) N = 8 . \* ou a: P < 0,05; \*\* ou aa : P < 0,01; \*\*\* : P < 0,001.



**Figure 56 : Etude des effets transgénérationnels du mélange M4 sur l’histologie des canaux mammaires des animaux sauvages et transgéniques MMTV/ER $\alpha$ 36 à l’âge adulte.** (Figure supplémentaire S7 de la publication n°2)

Variation de l’épaisseur du stroma et de l’épithélium, mesurée sur des coupes histologiques de glandes mammaires prélevées à l’âge adulte (16 semaines) sur des souris sauvages (wt) ou transgéniques MMTV-ER $\alpha$ 36 (Tg) appartenant à la génération F3. Ces animaux ont été exposés (D2) ou non (D0), de manière transgénérationnelle à une dose de 0,5  $\mu$ g/kg/jour de M4 (dose correspondant à l’exposition humaine).

Les barres d’erreur correspondent à l’erreur standard à la moyenne.

d'une perte d'expression de la E-cadhérine au profit d'une induction de l'expression de la N-cadhérine. Concernant l'implication d'ER $\alpha$ 36 dans ce mécanisme, la présence du transgène semble annuler les effets liés à l'impact transgénérationnel du M4. En effet, chez les animaux Tg MMTV-ER $\alpha$ 36, au sevrage, l'épaisseur du stroma et de l'épithélium n'est pas impactée par l'exposition transgénérationnelle au M4. Cela semble donc souligner à nouveau les effets antagonistes du transgène et du M4.

Concernant les effets du M4 à l'âge adulte, aucune modification de l'histologie des canaux mammaires n'est observée chez les animaux wt et Tg MMTV-ER $\alpha$ 36 (Figure 56). Cela peut suggérer qu'après la puberté des mécanismes compensatoires se mettent en place pour permettre un développement normal de la glande mammaire. Cependant, avant de conclure définitivement il serait intéressant d'analyser plus finement les coupes histologiques en regardant l'expression de marqueurs de prolifération et d'apoptose (Ki67 et TUNEL) ou en étudiant la composition de la matrice extracellulaire entourant les canaux. De plus, comme cela est décrit dans une étude de Ayyanan et al., (2011), la proportion de cellules épithéliales basales (CD24 low) et luminales (CD24 high) pourrait être évaluée par cytométrie en flux.

En conclusion, l'expression d'ER $\alpha$ 36 dans les cellules de la lignée germinale mâle murine et dans les cellules épithéliales mammaires saines est suffisante pour conduire à la modulation d'un certain nombre de gènes et de voies de signalisation impliqués dans le contrôle de processus biologiques, comme la prolifération, la différenciation, la migration et l'apoptose. In vivo, cela n'a pas de conséquences sur le développement et la fonction testiculaire des souris Tg Oct4-ER $\alpha$ 36, en revanche cela impacte l'histologie des canaux mammaires des souris Tg MMTV-ER $\alpha$ 36 adultes. Ces altérations structurales de la glande mammaire, qui apparaissent majoritairement après la puberté, semblent dépendre de l'induction de la signalisation dépendante d'ER $\alpha$ 36 par les hormones endogènes et font penser à des lésions pré-néoplasiques de l'épithélium.

Concernant les alkylphénols, in vitro, le mélange M4 induit la prolifération des cellules séminomateuses TCam-2 de façon dépendante d'ER $\alpha$ 36 et stimule la migration et la prolifération des cellules épithéliales mammaires saines MCF-10A surexprimant ce variant. Toutefois, in vivo, le traitement des animaux transgéniques avec ces perturbateurs endocriniens oestrogène-mimétiques n'est pas associé à une amplification des effets du transgène sur le développement testiculaire et mammaire.

# Publication N°1

“Mammary epithelial cell phenotype disruption *in vitro* and *in vivo* through ER $\alpha$ 36 overexpression”

C. Thiebaut\*, C. Chamard-Jovenin\*, A. Chesnel, C. Morel, E.H. Djermoune,  
T. Boukhobza, H. Dumond.

\* Contribution égale

Article publié dans PLoS ONE en Mars 2017.

Le récepteur aux œstrogènes alpha 36 (ER $\alpha$ 36) est un variant de la forme canonique du récepteur aux œstrogènes alpha 66 (ER $\alpha$ 66), exprimé dans de nombreuses lignées de cellules cancéreuses sensibles aux hormones, et dont l'expression est corrélée à un pronostic défavorable pour les patientes atteintes d'un cancer du sein. Alors que l'activité d'ER $\alpha$ 36 a été reliée à la progression du cancer du sein et à l'acquisition de résistances à l'hormono- et à la chimiothérapie, le niveau d'expression, la localisation d'ER $\alpha$ 36 et son rôle physiologique sont peu décrits dans la glande mammaire saine.

C'est pourquoi nous avons exploré les conséquences d'une surexpression d'ER $\alpha$ 36, *in vitro* dans une lignée de cellules épithéliales mammaires saines immortalisées, MCF-10A, et *in vivo* dans un modèle unique de souris transgéniques MMTV-ER $\alpha$ 36 dans lequel l'ARNm codant ER $\alpha$ 36 est spécifiquement exprimé dans la glande mammaire murine.

A l'aide d'une combinaison d'analyses bioinformatiques et d'études expérimentales du transcriptome de cellules MCF-10A surexprimant ou non ER $\alpha$ 36, nous avons identifié un réseau de gènes hiérarchisé, en aval d'ER $\alpha$ 36 et modulé par la voie JAK2/STAT3. De manière concomitante, nous avons montré que la surexpression d'ER $\alpha$ 36 induit, dans la lignée MCF-10A, une inhibition de la prolifération mais augmente le potentiel migratoire de ces cellules et leur résistance face à la mort cellulaire par apoptose induite par un traitement à la staurosporine.

*In vivo*, l'expression d'ER $\alpha$ 36 induit un amincissement et une dislocation de l'épithélium des canaux mammaires à l'âge adulte mais aucune altération majeure n'est observée au sevrage ou chez les femelles avant la puberté. Ces phénotypes sont corrélés à une perte de l'expression de la E-cadhérine, marqueur de jonction adhérente des cellules épithéliales.

En conclusion, nous avons montré qu'une surexpression d'ER $\alpha$ 36 est suffisante, à elle seule, pour induire des perturbations du phénotype épithélial du tissu mammaire sain, *in vivo* et *in vitro*, par l'intermédiaire d'un effet stimulateur dominant sur les voies de signalisation non génomiques des œstrogènes. Ces résultats suggèrent qu'en présence d'un niveau de stéroïdes endogènes suffisant chez l'adulte, la surexpression d'ER $\alpha$ 36 *in vivo* altère l'architecture de la glande mammaire. Cela pourrait conduire à l'apparition de lésions pré-néoplasiques et augmenter le risque de cancer du sein.

RESEARCH ARTICLE

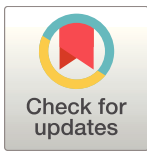
# Mammary epithelial cell phenotype disruption *in vitro* and *in vivo* through ERalpha36 overexpression

Charlène Thiebaut<sup>☯</sup>, Clémence Chamard-Jovenin<sup>☯</sup>, Amand Chesnel, Chloé Morel, El-Hadi Djermoune, Taha Boukhobza, Hélène Dumond\*

CNRS-Université de Lorraine, UMR 7039, Centre de Recherche en Automatique de Nancy, BP70239, Vandœuvre-lès-Nancy, France

☯ These authors contributed equally to this work.

\* [helene.dumond@univ-lorraine.fr](mailto:helene.dumond@univ-lorraine.fr)



**OPEN ACCESS**

**Citation:** Thiebaut C, Chamard-Jovenin C, Chesnel A, Morel C, Djermoune E-H, Boukhobza T, et al. (2017) Mammary epithelial cell phenotype disruption *in vitro* and *in vivo* through ERalpha36 overexpression. PLoS ONE 12(3): e0173931. <https://doi.org/10.1371/journal.pone.0173931>

**Editor:** Aamir Ahmad, University of South Alabama Mitchell Cancer Institute, UNITED STATES

**Received:** December 1, 2016

**Accepted:** February 28, 2017

**Published:** March 16, 2017

**Copyright:** © 2017 Thiebaut et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** The funders are all French: the Anses - Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail <https://www.anses.fr/> (fund 2012/2/014; recipient HD), the Institut national de la santé et de la recherche médicale [www.inserm.fr](http://www.inserm.fr) Plan Cancer/ environnement (fund ENV201304; recipient HD), La Ligue contre le cancer [www.ligue-cancer.net/](http://www.ligue-cancer.net/)

## Abstract

Estrogen receptor alpha 36 (ERα36) is a variant of the canonical estrogen receptor alpha (ERα66), widely expressed in hormone sensitive cancer cells and whose high expression level correlates with a poor survival prognosis for breast cancer patients. While ERα36 activity have been related to breast cancer progression or acquired resistance to treatment, expression level and location of ERα36 are poorly documented in the normal mammary gland. Therefore, we explored the consequences of a ERα36 overexpression *in vitro* in MCF-10A normal mammary epithelial cells and *in vivo* in a unique model of MMTV-ERα36 transgenic mouse strain wherein ERα36 mRNA was specifically expressed in the mammary gland. By a combination of bioinformatics and computational analyses of microarray data, we identified hierarchical gene networks, downstream of ERα36 and modulated by the JAK2/STAT3 signaling pathway. Concomitantly, ERα36 overexpression lowered proliferation rate but enhanced migration potential and resistance to staurosporin-induced apoptosis of the MCF-10A cell line. *In vivo*, ERα36 expression led to duct epithelium thinning and disruption in adult but not in prepubescent mouse mammary gland. These phenotypes correlated with a loss of E-cadherin expression. Here, we show that an enhanced expression of ERα36 is sufficient, by itself, to disrupt normal breast epithelial phenotype *in vivo* and *in vitro* through a dominant-positive effect on nongenomic estrogen signaling pathways. These results also suggest that, in the presence of adult endogenous steroid levels, ERα36 overexpression *in vivo* contributes to alter mammary gland architecture which may support pre-neoplastic lesion and augment breast cancer risk.

## Introduction

Twenty years ago, steroid hormones, especially sex hormones, were shown to stimulate gene transcription through binding a transcription factor in the cell nucleus called a nuclear steroid receptor. The so-called “genomic estrogen signaling” is mediated by direct actions of nuclear-localized estrogen receptors (ERs: ERalpha and ERbeta) as ligand-induced transcription



(fund 2016; recipient HD) and the Région Lorraine for a half PhD fellowship <http://www.lorraine.eu> (CC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

factors [1,2]. However, estrogen activities are also mediated through nongenomic signaling which involves extranuclear events such as activation of various protein kinases mediated by membrane associated ERs or the G protein-coupled estrogen receptor (GPER) [3].

In 2005, Wang and colleagues [4] identified and cloned a 36-kDa variant of ER $\alpha$ , ER $\alpha$ 36, which is mainly located on the plasma membrane and mediates nongenomic estrogenic signaling. ER $\alpha$ 36 differs from the canonical 66-kDa estrogen receptor alpha (ER $\alpha$ 66) by the lack of both AF-1 and AF-2 transcription activation domains, and a truncated ligand-binding domain deleted from helix9 to helix12. These structural characteristics are consistent with the fact that ER $\alpha$ 36 has no intrinsic transcriptional activity and suggest that it may have a spectrum of ligand selectivity different from the ER $\alpha$ 66 one. Although the ability of ER $\alpha$ 36 to interact directly with any ligand is still debated, tamoxifen has been demonstrated to bind ER $\alpha$ 36 as an agonist like 17 $\beta$ -estradiol, which triggers proliferation, migration and apoptosis escape of breast cancer cells [5–8]. Moreover, ER $\alpha$ 36 is generated from a promoter located in the first intron of the ESR1 gene, indicating that its expression can be regulated independently from ER $\alpha$ 66 and consistent with the findings that ER $\alpha$ 36 is expressed in ER-negative breast cancer cells that lack ER $\alpha$ 66 expression [4,9]. Even in the absence of known agonists, a high expression of ER $\alpha$ 36 is a marker of poor survival prognosis for breast cancer patients [10, 11].

Consistent with its subcellular localization, ER $\alpha$ 36 can mediate membrane-initiated signaling through physical interaction with the EGFR/Src/Shc complex, or functional coupling with GPER or HER2, depending of the tumor cell tested [12–15]. In various types of cancerous cell lines (endometrial, ER positive or negative breast cancer cells, seminoma cells. . .), ER $\alpha$ 36 may activate calcium release, the PKC, MAPK/ERK or the PI3K/AKT signaling pathways responsible for cell proliferation, migration and survival [16–18].

ER $\alpha$ 36 mRNAs have been found in normal ovary, uterus, breast and testis tissues as well as endothelial and vascular smooth muscle cells, kidney, cartilage, bone, lung and heart [18]. A sexual dimorphism of protein location was reported in the osteoarticular system where ER $\alpha$ 36 appears to associate with plasma membrane only in females. However, the ER $\alpha$ 36 physiological role(s), ligand(s) and associated transactivation mechanism(s) in those tissues remain to be determined [19, 20]. Despite ER $\alpha$ 36 activity have been related to breast cancer progression and acquired resistance to tamoxifen and chemotherapy breast tumor treatment, precise expression level and location of ER $\alpha$ 36 is still missing in the normal mammary gland [21].

The mammary gland is an epidermal appendage, originating from ectodermal and mesodermal elements, whose development is a step by step sequential process which begins in utero at 10.5 day post-coitum (dpc) in mouse and 4 to 6 weeks in human and ends at adulthood with lactation and involution [22, 23]. A rudimentary mammary tree is present in newborn females, whereas regression occurred under androgen production during male fetal development [24]. During female puberty, the epithelium forms into a branching, bilayered ductal structure, consisting of an outer basal myoepithelial layer of cells and an inner luminal cell layer [24]. Epithelial estrogen and progesterone were shown to be responsible for ductal elongation and side branching, respectively [25, 26]. Several studies based on KO mouse phenotype analyses demonstrated that the estrogen receptors ER $\alpha$  and ER $\beta$  are dispensable before puberty but required for a correct mammary gland development and function in adulthood [26, 27]. Indeed, ER $\alpha$  is expressed in epithelial and stromal cells and mediate duct elongation whereas ER $\beta$  is mostly involved in cell differentiation and function during gestation and lactation [27, 28]. Despite an abundant and detailed literature dealing with ER $\alpha$ 36 expression and function in breast cancer, no data has been reported to date about a putative role of ER $\alpha$ 36 in mammary gland normal development or the molecular and cellular consequences of its transactivation.

In this work, we explored the consequences of a potential ER $\alpha$ 36 overexpression in normal mammary epithelial cells in vitro and in a unique model of transgenic mouse strain, in which

ER $\alpha$ 36 is expressed specifically in the mammary gland. We identified proliferation, survival and migration as the main functions modulated by ER $\alpha$ 36 overexpression through STAT3 pathway, and assumed that this could lead to predispose normal mammary epithelial cells to neoplastic like transformation and augment breast cancer risk.

## Material and methods

### Animal husbandry

Mice were housed under constant conditions at a temperature of  $22\pm 1^\circ\text{C}$ , a humidity of approximately 40% and a 10h-light/14h-dark cycle. Food (SAFE, France) and water disposed in glass bottles were supplied *ad libitum*. All experimental procedures including the study of a total of 120 animals were approved by the French Minister of Research Committee for animal experiment (APAFIS#2168–2015110518268051 v5).

### ER $\alpha$ 36 transgenic mouse strain

ER $\alpha$ 36 transgenic (Tg) founders were obtained from the Mouse Clinical Institute, Illkirsch-Graffenstaden, France by injecting an MMTV/ER $\alpha$ 36 chimeric construction into one of the pronuclei of a hybrid B6SJL/F2 background (S1A Fig). The transgene is a transcriptional fusion composed of the MMTV promoter sequence from pGL4.36 plasmid (Promega) and the complete ER $\alpha$ 36 human cDNA sequence [4] as previously described for EGFR in Brandt [29]. The transgene copy number was determined by droplet digital PCR using specific primers, and a clone bearing a single copy of the MMTV-ER $\alpha$ 36 transgene was retained, in order to avoid transgene silencing and obtain 50% hemizygote transgenic offspring when crossed with wild-type (wt) C57BL/6J strain.

The presence or absence of the ER $\alpha$ 36 transgene was assessed at weaning (postnatal day 21, PND 21) by means of genomic DNA extraction from ear pinna samples and further real time PCR determination (forward 5' -GCTTCGATGATGGGCTTACT-3'; reverse 5' -CTAAACTGGAGGTACTAGTCC-3'). A sequence from chromosome 17 was used as a reference (forward 5' -AAGGAGCAAGGTGGCTTACA-3'; rev 5' -TGAGAAGGGTACCGTCACGG-3').

### ER $\alpha$ 36 transgene expression in transgenic strain

The male transmission of the transgene was preferred rather than the female one in order to avoid a maternal effect of ER $\alpha$ 36 expression, through potential physiological or behavioral alterations. Seven week old hemizygote ER $\alpha$ 36 transgenic males from F1 to F3 generations (6 males per generation) were crossed with seven week old C57BL/6J wt virgin females obtained from Charles River Laboratories, France. The mammary gland specific ER $\alpha$ 36 mRNA expression was determined in F1 to F4 litters. Mammary gland, uterus, ovary, testis, heart, liver, brain, kidney and salivary gland were then dissected from 21 day or 4 month old F2 to F4 progenies (at least 5 unrelated animals for each age and generation), RNA extracted and checked for ER $\alpha$ 36 expression by qRT-PCR (Forward 5' -ATGAATCTGCAGGGAGAGGA; reverse 5' -GGCTTTAGACACGAGGAAACC-3'). The RPLPO gene was used as a reference (forward 5' -GGCGACCGTGAAGTCCAAC-3'; reverse 5' -CCATCAGCACCACAGCCTTC-3'). In wt mice, we detected no ER $\alpha$ 36 expression, whatever the organ, the sex or the age of the animal tested. In transgenic mice, none of the tested organs except the mammary gland did express the transgene (not shown). The protein expression in adult mammary gland (4 month old females) was confirmed by western blot analysis (S1B Fig). The sex ratio and transmission rate of the transgene were determined by means of male/female and transgenic littermate number determination in each litter and generation in a total of 171 animals (S1C Fig).

## Mammary gland whole mounts

To prepare the whole mounts, the mammary glands were dissected from animals anesthetized with a lethal dose of pentobarbital, spread onto glass slides and stained as described by Vandenberg et al. [30]. A dedicated Matlab program, adapted from Tylcz et al. [31] was used to quantify mammary tree extension and branching. First, the ducts in all the images are enhanced using the multiscale filtering method of Frangi et al. [32] (S2A Fig). The objectives are (i) to increase the difference in pixel intensities between ducts and background tissues, and (ii) to prevent structures breaks. The ducts are then detected by segmentation (hard thresholding) and skeletonization. Finally, the mammary network is quantified in terms of tree extension, branching and amount of sprouts in F2 and F4 animals (F2wt: N = 4, F2 Tg: N = 6, F4wt: N = 6, F4Tg N = 5). (S2B Fig).

## Histology

To prepare paraffin sections, mammary glands were fixed with Davidson fixative [33] for 24h at room temperature, dehydrated and embedded with paraffin (VWR). Seven micrometer sections were cut on a Leitz rotary microtome (Leica) and mounted on Superfrost slides (Fisher Scientific); the entire mammary gland was sectioned at once. Sections were stained with hematoxylin/eosin/methyl green to determine the presence of epithelial cords. Sections with visible cords were used for further analysis. In order to determine mammary ducts parameters, only orthogonally sectioned ones were chosen. By using NIS-elements BR4.20.00 imaging software (Nikon), epithelium and stromal thickness as well as lumen width were measured along both the larger and smaller diameters of the ducts. At less 5 duct sections from 5 separate slices were measured in each mammary gland. A total of 27 animals were used for histology (at weaning wt: N = 6; Tg N = 8; adults wt: N = 6; Tg N = 7).

## Cell culture

MCF-10A cells were purchased in 2015 from ATCC<sup>®</sup> (CRL-10317) and maintained in DMEM/F12 (GIBCO) supplemented with 5% horse serum (HS), 1% Glutamine 1% Penicillin/Streptomycin and a mix of EGF (epidermal growth factor), Cholera toxin, hydrocortisone and insulin as described by Soule et al. [34].

## Transient transfection and stable cell line establishment

Stable MCF-10A cell lines transfected by pCDNA3.1-ER $\alpha$ 36 (pCDNA3.1vector containing the complete cDNA sequence of ER $\alpha$ 36) or the empty expression vector were obtained as previously described [17]. Corresponding cell lines were named MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo, respectively. The transfected cell lines were then subcloned and checked for ER $\alpha$ 36 expression by RT-PCR (S3A Fig), western blot (S3B Fig) or immunofluorescence analysis before each experiment (S3C Fig).

## Microarray experiment

Transcriptional profile analyses of MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo cells cultured for 2 days in standard medium were performed in triplicates on Affymetrix GeneChip U133 2.0 by the GenomEast Platform (IGBMC, Strasbourg, France). Biotinylated cRNA targets were prepared using the Ambion "MessageAmp" Premier RNA Amplification Kit" according to the Instruction Manual P/N 4386269 Revision D (Revision Date: May 16, 2008), starting from 200 ng of total RNA extracted with RNeasy Mini kit (Quiagen). Following fragmentation, 10  $\mu$ g of cRNAs were hybridized for 16 hours at 45°C, 60 rpm on Human GeneChip<sup>®</sup> HG-U133 plus 2.0 arrays (Affymetrix). The chips were washed and stained in the GeneChip<sup>®</sup> Fluidics

Station 450 (Affymetrix) using the FS450\_0004 script and scanned with the GeneChip<sup>®</sup> Scanner 3000 7G (Affymetrix) at a resolution of 1.56  $\mu$ m. Raw data (.CEL Intensity files) were extracted from the scanned images using the Affymetrix GeneChip<sup>®</sup> Command Console (AGCC) version 4.0. CEL files were further processed with Affymetrix Expression Console software version 1.3.1 to calculate probeset signal intensities, using Robust Multi-array Average (RMA) algorithms with default settings.

### Real-time PCR analysis

RT and real-time PCR analyses were performed as previously described [17]. The following primers were used for qRT-PCR: *RPLPO* forward (Fw) 5' -GGCGACCGTGAAGTCCAACCT-3', *RPLPO* reverse (Rev) 5' -CCATCAGCACCACAGCCTTC-3', *CDH1* forward (Fw) 5' -TGC CCAGAAAATGAAAAAGG-3', *CDH1* reverse (Rev) 5' -GTGTATGTGGCAATGCGTTC-3', *CDH2* forward (Fw) 5' -ACAGTGGCCACCTACAAAGG-3', *CDH2* reverse (Rev) 5' -CCGA GATGGGGTTGATAATG-3'. Assays were performed at least in triplicate, and the mean values were used to calculate expression levels, using the  $\Delta\Delta C(t)$  method referring to *RPLPO* house-keeping gene expression.

### Crystal violet assay

The crystal violet assay was performed in 24-well plates seeded with  $5 \times 10^2$  cells per well. After each well was washed with PBS, the cells attached to the bottom of the plate were fixed and stained with 0.4% crystal violet solution in 2% ethanol for 30 min. After the plate was washed with water and dried, crystal violet was solubilized in 10% acetic acid and the absorbance at 570 nm was measured by a microplate reader (Victor x3, Perkin-Elmer).

### Western immunoblotting

Western blots were performed as described previously [17]. The following primary antibodies were used: anti-Cyclin D1 (#2922, Cell Signaling), anti-ER $\alpha$ 36 (CY1109, Cell Applications), anti-PARP 1 cleaved (552596, BD Pharmingen) and anti-Caspase 7 cleaved (#9494, Cell Signaling), anti-Caspase 3 cleaved (#9664, Cell Signaling). The anti- $\beta$  Actin antibody (sc1615, Santa Cruz Biotechnology), anti-Glyceraldehyde-3-Phosphate-DesHydrogenase (GTX100118, Genetex) or anti- $\alpha$ -tubulin (GTX102079, Genetex) were used as a control. Protein expression profiles were revealed with Clarity Western ECL Substrate (Biorad) and banding quantification was performed using the Quantity One Chemidoc XRS software (Biorad).

### Immunofluorescence

Immunofluorescence was performed as described previously [17]. The following primary antibodies were used: anti-cytochrome c (sc13561, Santa Cruz Biotechnology), anti-NF $\kappa$ B p65 (GTX102090, GeneTex), anti- $\beta$ -catenin (E247, Epitomics #1247-s), anti-E-cadherin (GTX100443, GeneTex), anti-N-cadherin (TA326835, OriGene), anti-PTEN (#9188, Cell Signaling), anti-phospho-ERK1/2 (#4370, Cell Signaling) and anti-STAT3 (#12640, Cell Signaling). Goat anti-rabbit secondary antibody was coupled to AlexaFluor 555 (Invitrogen). Images were obtained with DS-Ri1 Nikon camera and Eclipse80i Nikon microscope and quantifications were performed using NIS-Elements BR 4.20.00 software (Nikon).

### Scratch-wound assay

Scratch assays were performed using the Ibidi-culture inserts (Ibidi<sup>®</sup>/Biovalley) following the manufacturer instructions. Cultures were then washed to remove detached cells and debris.

Quantification of wound mean width were performed at  $t = 0$  and  $t = 6$ h of culture by phase-contrast image analysis with NIS-elements BR 4.20.00 software (Nikon).

### TUNEL assay

TUNEL assay was performed using the Apo-BrdU-IHC in situ DNA fragmentation Assay kit (BioVision, USA) following the manufacturer instructions for the staining of cell preparations fixed on slides adapted for the use of AlexaFluor 555 (Invitrogen) goat anti-mouse secondary antibody. Nuclear DNA was stained with Hoechst (bisBenzimide H33342 Trihydrochloride, Sigma-Aldrich). Quantification of stained cell number was performed with NIS-elements BR 4.20.00 software (Nikon).

### Annexin V apoptosis detection assay

Cell apoptosis was assayed by using the BD Annexin V: FITC Apoptosis Detection system. Propidium iodide and FITC fluorescence of the stained cells were automatically acquired and analyzed with a BD FACSCalibur<sup>TM</sup> (BD BioSciences).

### Statistical analysis

All data are summarized as the mean  $\pm$  SEM and results considered statistically significant  $P < 0.05$ . Statistical analyses were performed with Matlab vR2014b software (MathWorks) by using Student t-test for unpaired samples with a significant p-value threshold below 5%.

## Results

### Microarray and bioinformatic analysis of differentially expressed genes in MCF-10A/ER $\alpha$ 36 versus MCF-10A/Zeo cells

In order to determine which cell function could be altered by ER $\alpha$ 36 overexpression, we analyzed and compared the transcriptional profiles of MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo cells. 8022 transcripts were significantly up- or down-regulated (absolute variation factor  $\geq 2.3$  in triplicate RNA samples, corrected p-value  $P < 0.05$  with Benjamini-Hochberg method) between the 2 cell lines. Corresponding genes were termed “differentially expressed genes” (DEGs). The overall strategy of DEG list analysis is described in [S4A Fig](#).

### Gene Ontology enrichment analysis of differentially expressed genes

The online tool MSigDBv5.0 (Molecular Signature Database <http://www.broadinstitute.org/gsea/msigdb/>) [35, 36] was used to achieve Gene Ontology (GO) enrichment analysis. The overlap between DEG list (see above) and GO gene sets derived from the Biological Process part of GO was computed. The four functions mostly affected by ER $\alpha$ 36 overexpression were signal transduction (GO:0007165;  $P = 1 \times 10^{-44}$ ), cell proliferation (GO:0008283;  $P = 1 \times 10^{-9}$ ), cell surface receptor linked signal transduction (GO:0007166;  $P = 1 \times 10^{-14}$ ) and apoptotic process (GO:0006915;  $P = 1 \times 10^{-9}$ ). Proportion of genes allocated to one GO gene set affected by the overexpression are 43.3%, 43.3%, 41.8% and 43.2%, respectively.

### Pathway enrichment of DEGs

Pathway enrichment analysis was also achieved with the online tool of the KEGG database (<http://www.genome.jp/kegg/>). The four best overlaps between the DEG list and KEGG database signaling pathways were identified for the “PI3K-AKT Signaling pathway” (hsa04151), “MAPK Signaling pathway” (hsa04010), “cAMP Signaling pathway” (hsa04024) and “Jak/STAT Signaling

pathway” (hsa04630) ( $P$ -value $<10^{-3}$ ). Proportion of genes allocated to one KEGG signaling pathway affected by the overexpression are 36.6%, 39.8%, 38.4% and 41.1%, respectively.

## Identification of key DEG regulators downstream of ER $\alpha$ 36

Thereafter, Ingenuity Pathway Analysis (IPA<sup>®</sup>) and Matlab<sup>®</sup> software were used to identify intermediate factors that may contribute to the target gene expression changes observed after ER $\alpha$ 36 overexpression.

The DEG list was submitted to the IPA<sup>®</sup> software which predicted ESR1 gene (encoding ER $\alpha$ 36) and 17 $\beta$ -estradiol (ER $\alpha$ 36 agonist) as key upstream regulators of the DEG list. The direct and indirect functional links (affected, activated and inhibited) between the DEGs and the predicted upstream regulators are described by the IPA<sup>®</sup> Software as 3 adjacency matrices in a digraph collecting the gene interactions.

In order to identify intermediate regulators acting downstream ER $\alpha$ 36 and involved in the control of each function previously identified by MSigDB<sup>®</sup> analysis, gene-regulator networks, corresponding to each of the four specific functions (see GO gene sets above), were built using the Matlab<sup>®</sup> software. By running a specific Matlab program on adjacency matrices, we identified among the upstream regulators determined via IPA (i) the first level of intermediate regulators that directly regulate DEGs, and (ii) the second level of intermediate regulators that control directly the intermediates from the first level and thus indirectly the DEGs. Iterative identification of intermediate regulator levels was performed until no more element could be added into the regulator set.

Then, a digraph was drawn for each function representing its regulator hierarchical network for the considered DEGs. In the networks, the DEGs, intermediate or upstream regulators are represented as nodes. The vertices indicate the regulation relationship between the nodes. An example of such a network is described in [S4B Fig](#).

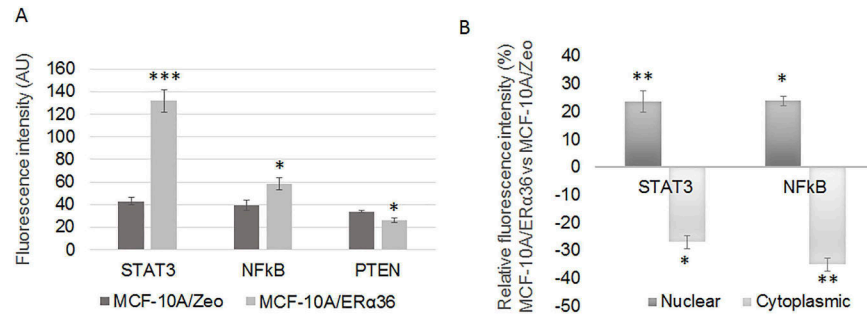
The common intermediate regulators (i.e. nodes) involved in the four functions were determined by comparing the four networks. For each node of any network, we associated an integer which indicated the number of networks in which it can be found. This led to the identification of 24 common intermediate factors that could mediate ER $\alpha$ 36 dependent target gene regulation. Among them, we selected for further experimental validation, the intermediates involved into the signaling pathways identified from KEGG database.

The expression and the localization of key effectors, associated with PI3K/AKT (PTEN, PI3K), MAPK/ERK1/2 (phosphorylated ERK1/2), JAK/STAT (STAT3) or NF $\kappa$ B (p65) signaling pathways, were studied *in vitro* by immunofluorescence. Fluorescence intensity quantification indicated that each protein was expressed in MCF-10A/Zeo cells at a similar level. Nevertheless, ER $\alpha$ 36 overexpression significantly induced STAT3 (206%) and NF $\kappa$ B (49%) expression but decreased PTEN (23%) expression ([Fig 1A](#) confirmed by western blot in [S5A Fig](#)). A nuclear translocation of NF $\kappa$ B (24%) and STAT3 (23%) was also observed in MCF-10A/ER $\alpha$ 36 cells compared to MCF-10A/Zeo cells ([Fig 1B](#), [S5B Fig](#)). This was consistent with an activation of the NF $\kappa$ B and JAK2/STAT3 signaling pathways.

In parallel, the accuracy of the *in silico* derived predictions was also verified by examining the MCF-10A/ER $\alpha$ 36 versus MCF-10A/Zeo cells phenotypes. First, we focused on proliferation and apoptosis that are 2 cellular functions previously highlighted by the GO enrichment analysis and then we studied the migration/invasion that could be affected by the TGF $\beta$ 1 intermediate factor.

## ER $\alpha$ 36 expression and cell division

Quantification of viable adherent cells by crystal violet staining indicated a 35% decrease of cell number in MCF-10A/ER $\alpha$ 36 compared to MCF-10A/Zeo cells ([Fig 2A](#)). Consistently,



**Fig 1. ER $\alpha$ 36 overexpression modulates MAPK/ERK1/2, NF $\kappa$ B and JAK2/STAT3 signaling pathways in MCF-10A cells.** Expression levels of PTEN, Phospho-ERK1/2 (P-Erk), NF $\kappa$ B and STAT3 were measured by immunofluorescence with specific antibodies: anti-PTEN, anti-Phospho-ERK1/2, anti-NF $\kappa$ B p65, anti-STAT3 in MCF-10A/Zeo and MCF-10A/ER $\alpha$ 36 cells. Total, nuclear and cytoplasmic average fluorescent signal intensities (AlexaFluor 555) are quantified from 5 cells in at least 5 separate fields per condition. Each bar represents mean  $\pm$  S.E.M. N = 3 independent experiments. \*:  $P < 0.05$ , \*\*\*:  $P < 0.001$ . A. Average fluorescence of PTEN, Phospho-ERK1/2, NF $\kappa$ B and STAT3 in whole MCF-10A/Zeo and MCF-10A/ER $\alpha$ 36 cells. B. Ratios of nuclear or cytoplasmic average fluorescence of STAT3 and NF $\kappa$ B intensities in MCF-10A/ER $\alpha$ 36 versus MCF-10A/Zeo cells indicating nuclear relocalization following ER $\alpha$ 36 overexpression.

<https://doi.org/10.1371/journal.pone.0173931.g001>

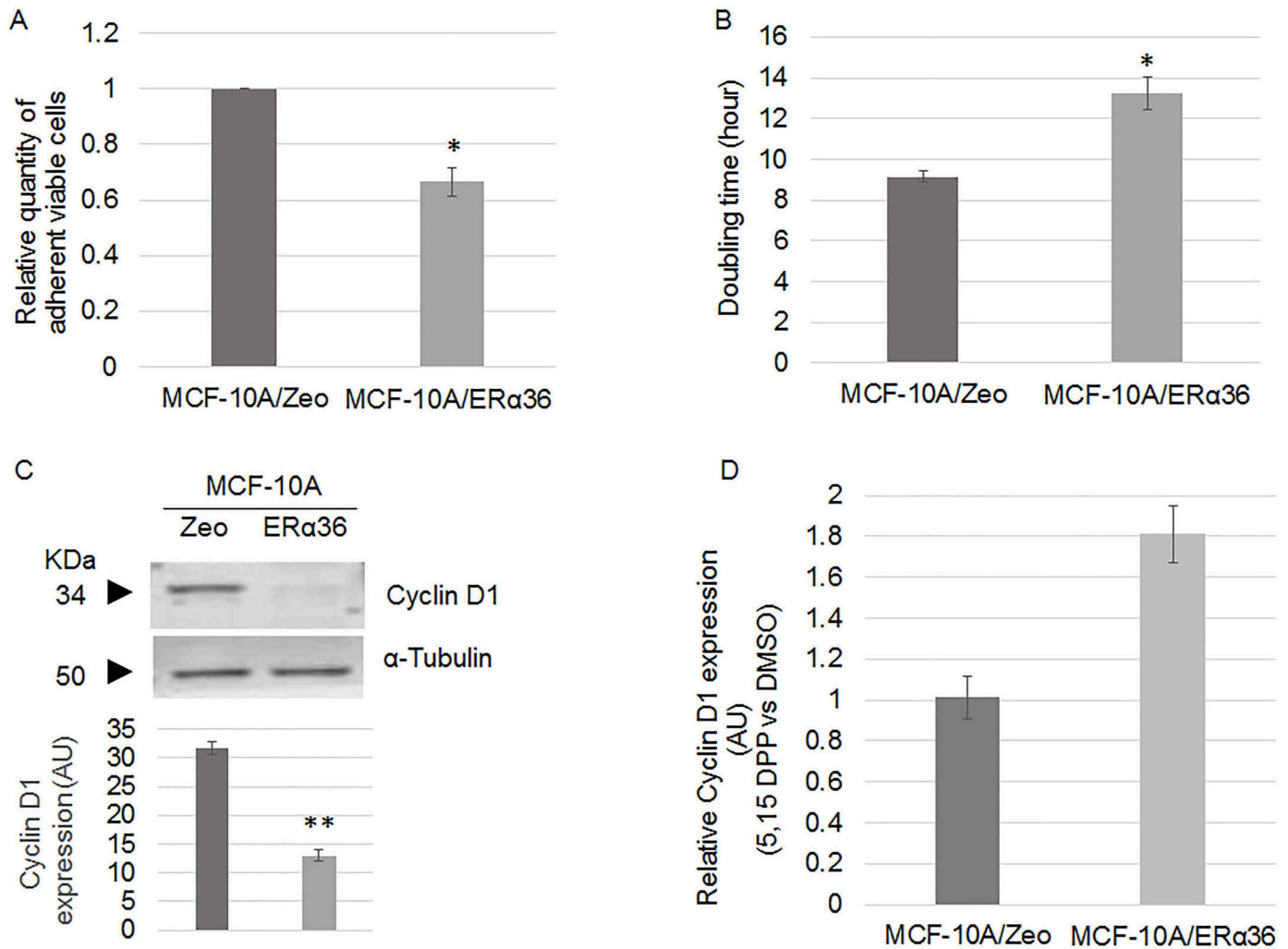
MCF-10A/ER $\alpha$ 36 doubling time was 40% higher than MCF-10A/Zeo one (Fig 2B), probably due to a 31% longer S phase, as measured by flow cytometry (not shown). Expression of cyclin D1, a key marker of cell cycle entry was also decreased by 60% in MCF-10A/ER $\alpha$ 36 compared to MCF-10A/Zeo cells (Fig 2C). Since this could appear to contradict to the previous observation that STAT3 mediates ER $\alpha$ 36 signaling, we measured cell proliferation and cyclin D1 expression after a 24 h treatment with 5,15 DPP, an inhibitor of STAT3 activity. 5, 15 DPP exposure had no effect on both parameters in the control MCF-10A cell line but triggered a 40% increase of cell number and a 81% increase of cyclin D1 expression in MCF-10A/ER $\alpha$ 36 cells. These results suggest that ER $\alpha$ 36 overexpression may slower cell division in a STAT3 dependent manner.

### ER $\alpha$ 36 expression and apoptosis escape

MCF-10A/Zeo and MCF-10A/ER $\alpha$ 36 cells were exposed for 6 h to DMSO (vehicle) or staurosporin (STS), a potent inducer of apoptosis. Fig 3A indicates that DMSO treated cells displayed no apoptosis. STS treatment triggered a significant expression of apoptotic markers in both cell lines with a respective 34%, 60% and 30% reduction of PARP 1, caspase 7 and caspase 3 cleavage in MCF-10A/ER $\alpha$ 36 compared to MCF-10A/Zeo cells. In order to confirm this result, we performed an FITC-Annexin V assay after both cell line STS treatment. After a 6-hour exposure, we detected less than 5% of stained cells in either control or ER $\alpha$ 36 overexpressing MCF-10A cells. An extended exposure time up to 24h led to a 10.85% reduction of apoptotic cell number in MCF-10A/ER $\alpha$ 36 compared to MCF-10A/Zeo cell line. No staining was detected in DMSO exposed control cells. We also performed a TUNEL assay after 24h vehicle or STS exposure: less than 0.3% vehicle treated cells were stained whereas 39.3% and 6.5% fragmented nuclei were detected in STS exposed MCF-10A/Zeo or MCF-10A/ER $\alpha$ 36, respectively (Fig 3B). Finally, a representative immunofluorescence image and the corresponding quantification of Fig 3B indicate that cytochrome c release by 75% in ER $\alpha$ 36 overexpressing MCF-10A cells. Taken together, these data strongly suggest that ER $\alpha$ 36 overexpression leads to a higher resistance to STS-induced apoptosis.

### ER $\alpha$ 36 expression and migration potential

Using the scratch-wound assay, we observed an accelerated closure of the wound created in a confluent monolayer culture of MCF-10A/ER $\alpha$ 36 compared to MCF-10A/Zeo cells (Fig 4A).



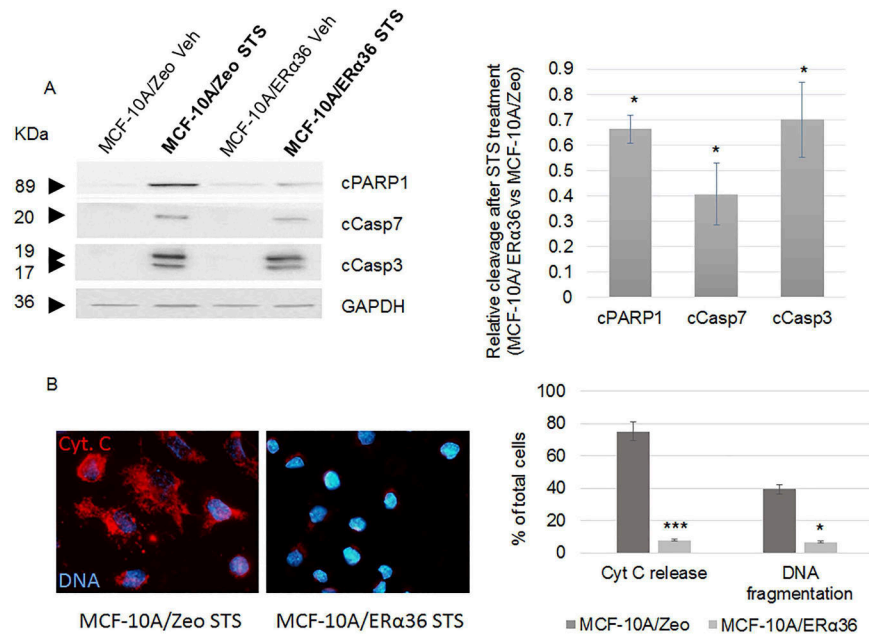
**Fig 2. ERα36 overexpression lowers MCF-10A cell proliferation rate.** A. Quantification of MCF-10A/Zeo and MCF-10A/ERα36 cell viability by crystal violet assay after 48h culture. ERα36 overexpression triggers a 35% decrease of cell proliferation. Each bar represents mean ± S.E.M. N = 3. \*:  $P < 0.05$ . B. Doubling time of MCF-10A/Zeo and MCF-10A/ERα36 cells was measured by crystal violet cell counting at  $t = 0$  (immediately after seeding),  $t = 24h$ ,  $48h$  and  $72h$  of culture. Given the measurements of living cells at  $t = 72h$  and  $t = 0$ , doubling time was calculated assuming a constant growth rate. Each bar represents mean ± S.E.M. N = 4. \*:  $P < 0.05$ . C. Representative western blot analysis of Cyclin D1 expression in MCF-10A/Zeo and MCF-10A/ERα36 cells (left panel), α-Tubulin was used as a loading control. Quantification of corresponding band intensity (right panel) indicate a 60% decrease of cyclin D1 expression in MCF-10A/ERα36 compared to MCF-10A/Zeo cells. Each bar represents mean ± S.E.M. N = 3. \*\*:  $P < 0.01$ . D. Relative Cyclin D1 expression in MCF-10A/Zeo and MCF-10A/ERα36 cells after 24h DMSO or  $10\mu M$  5,15 DPP exposure. Results depicted in the histogram are represented as 5,15 DPP versus DMSO ratio. In the presence of 5,15 DPP, cyclin D1 protein expression increases by 81% only in MCF-10A/ERα36 cells.

<https://doi.org/10.1371/journal.pone.0173931.g002>

Therefore, expression of epithelial-mesenchymal transition markers was analyzed by RT-PCR. CDH1 expression was reduced by 70% whereas CDH2 was induced 50% by ERα36 overexpression (Fig 4B).

Expression and localization of proteins known to be involved in cell-cell junctions were also examined: ERα36 overexpression triggered the decrease of E-cadherin expression (15%) and beta-catenin (30%) membrane immunofluorescence staining toward a dispersed cytoplasmic localization whereas N-cadherin staining was significantly augmented by 36% (Fig 4C).





**Fig 3. ER $\alpha$ 36 overexpression stimulates apoptosis resistance.** MCF-10A/Zeo and MCF-10A/ER $\alpha$ 36 cells were exposed to 0.25 $\mu$ M staurosporin (STS) or vehicle (Veh) for 6 hours. A. Cleavage of PARP1 (cPARP1), Caspase 7 (cCasp7) and Caspase 3 (cCasp3) were evaluated with specific antibodies (left panel). GAPDH was used as a loading control. Results depicted in the corresponding histogram are represented as STS versus Vehicle ratio (right panel). ER $\alpha$ 36 overexpression triggered a significant 34%, 60% and 30% decrease of PARP1, Caspase 7 and Caspase 3 cleavage, respectively. Each bar represents mean  $\pm$  S.E.M. N = 4. \*:  $P < 0.05$ . B. Cytochrome c (Cyt. C) release (red, AlexaFluor 555) and DNA fragmentation were respectively assessed by immunofluorescence and TUNEL assay after STS exposure in MCF-10A/Zeo and MCF-10A/ER $\alpha$ 36 cells (left panel), then quantified as shown in the corresponding histogram (right panel). No cytochrome c release or DNA fragmentation can be detected in untreated cells (not shown). Each bar represents mean  $\pm$  S.E.M. N = 3. \*:  $P < 0.05$ , \*\*\*:  $P < 0.001$ .

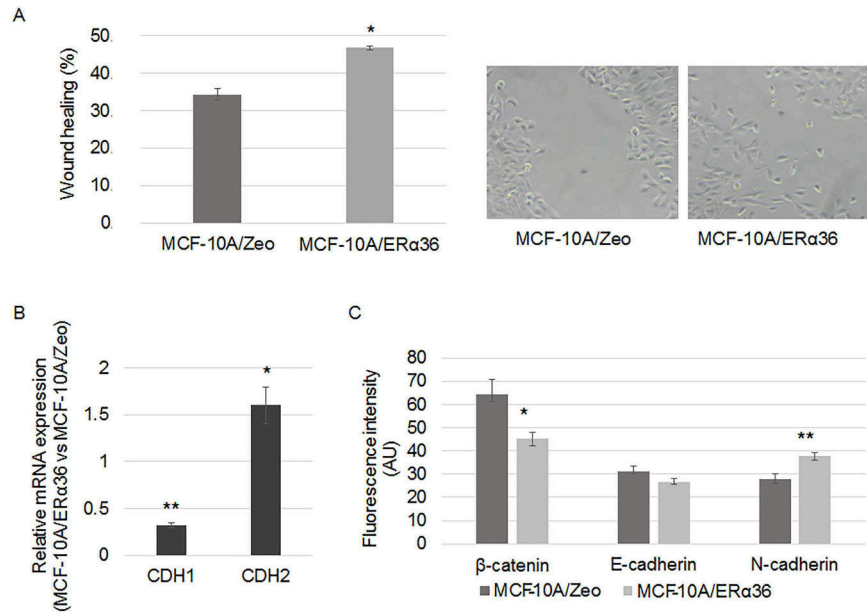
<https://doi.org/10.1371/journal.pone.0173931.g003>

### Consequences of ER $\alpha$ 36 expression in the mouse mammary gland

We hypothesized that ER $\alpha$ 36 overexpression could be sufficient to promote mammary epithelium alteration in vivo. Since ER $\alpha$ 36 sequence is absent in mouse, we produced ER $\alpha$ 36 transgenic mice in which ER $\alpha$ 36 is expressed under the control of MMTV (Mouse Mammary Tumor Virus) promoter, specifically in the mammary gland [29]. Mammary tree structure of wild-type (wt) and transgenic (Tg) mice was analyzed with a dedicated software designed to skeletonize mammary gland RGB images and compute mammary tree extension (Fig 5A), number of branching (Fig 5B) and end buds (Fig 5C). This study was completed by histological analyses of mammary gland slices and western blot analyses of protein expression.

First, we addressed the effects of an ER $\alpha$ 36 expression on fetal and neonatal mammary gland development by exploring the differences between wt and Tg mammary glands harvested on F2 or F4 mice at weaning and adulthood (Fig 6A). Neither mammary tree parameters nor measurement of epithelium or stroma thickness or lumen diameter varied at weaning (PND21) between wt and Tg animals (Fig 6B).

Conversely, at adulthood (16 week old), lumen diameter ( $P = 0.0036$ ) and stroma thickness ( $P = 0.0029$ ) were significantly augmented whereas epithelium thickness ( $P = 9 \times 10^{-7}$ ) was reduced in F2, F3 and F4 Tg female mice compared to wt ones (Fig 6C). This phenotype correlated with loss of E-cadherin expression in mammary glands from transgenic mice (S6 Fig).



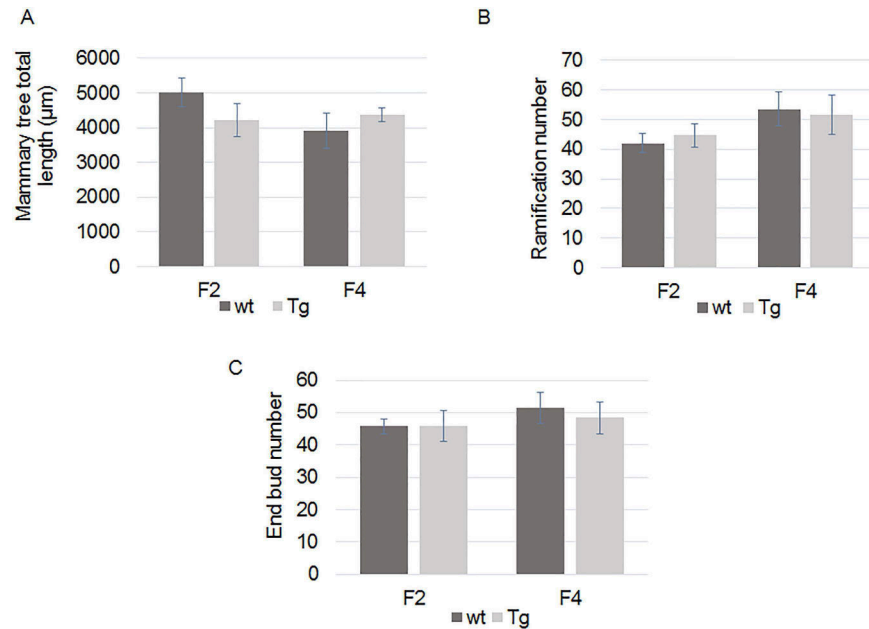
**Fig 4. ERα36 overexpression enhances migratory potential.** A. A wound was performed on a confluent monolayer culture of MCF-10A/Zeo and MCF-10A/ERα36 cells. Histogram depicts the wound healing when measured after a 6-hour culture (left panel). A representative picture of the migrating MCF-10A/ERα36 cells is presented in the right panel. Each bar represents mean ± S.E.M. N = 5. \*:  $P < 0.05$ . B-C. MCF-10A/Zeo and MCF-10A/ERα36 were cultured for 24 hours. B. CDH1 and CDH2 gene expression was measured by RT-PCR analysis. The housekeeping gene RPLPO was used as the reference gene. Each bar represents mean ± S.E.M. N = 3. \*:  $P < 0.05$ . \*\*:  $P < 0.01$ . C. Expression of cell-cell adhesion proteins was studied by immunofluorescence with specific antibodies: anti-β-catenin, anti-E-cadherin and anti-N-cadherin (AlexaFluor 555). Average fluorescent signal intensities are quantified from 5 cells from 5 fields per condition. Each bar represents mean ± S.E.M. N = 3. \*:  $P < 0.05$ , \*\*  $P < 0.01$ .

<https://doi.org/10.1371/journal.pone.0173931.g004>

These results suggest that ERα36 expression promotes mammary gland alteration after the puberty when potential estrogen receptor ligands such as steroid hormones are endogenously produced or reach the targeted mammary gland.

## Discussion

The data presented herein depict an integrated picture of normal mammary epithelial cells overexpressing ERα36. In the human MCF-10A cell line, ERα36 modulates the expression of genes that control proliferation, survival and migration. The pathways inferred from these data by computational modeling indicate that the regulation of those ERα36 target genes should involve preferentially the JAK2/STAT3 pathway as previously described in breast cancer cells in response to estrogens or anti-estrogens [13, 37]. These data are in line with the prior assumption of Björnström and Sjöberg [38] and Marino [39] who proposed that estrogenic signaling takes place through multiple pathways. Indeed, Zhang and Wang [13] and Fox [40] reported an ERα36 dependent control of proliferation and resistance to apoptosis through SRC/EGFR/STAT3/5 in breast cancer cells. Since ERα36 has been shown to collaborate with other estrogen or growth factor receptors in numerous cancer cells, our results also raise the possibility that an excess of ERα36 protein could be a dominant positive activator of either the G protein-coupled estrogen receptor GPER, or the EGFR downstream signaling. Indeed, GPER mRNA expression was stimulated by 2.09 fold in MCF-10A cells overexpressing ERα36. Moreover, ERα36 was shown to interact physically with GPER [41, 42] and/or to collaborate with GPER to trigger downstream signaling in female reproductive tract, seminoma or breast



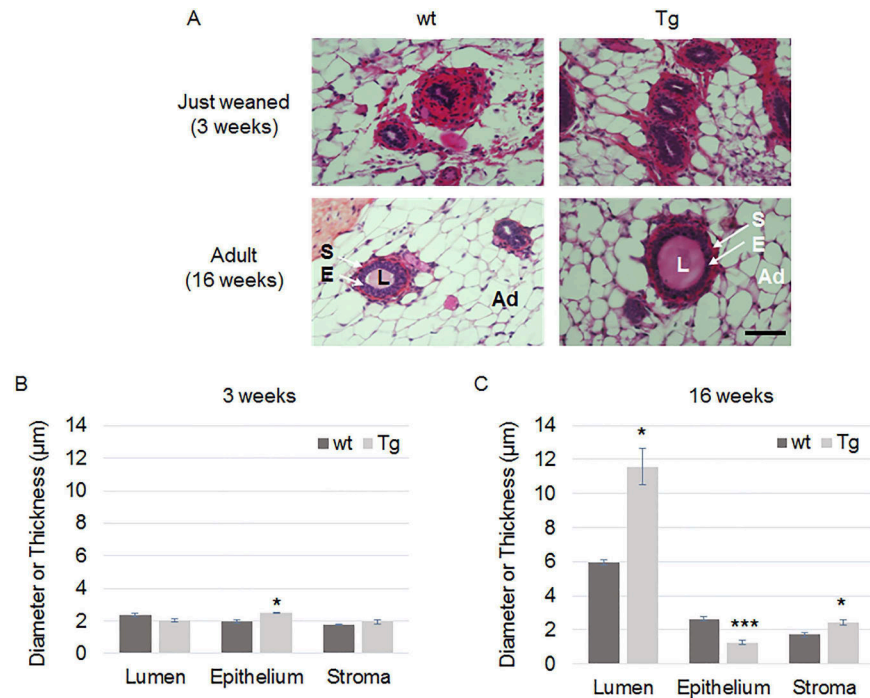
**Fig 5. Whole mount mammary tree parameters analyses in wt and Tg mice at weaning (PND21).** F2 or F4 mammary glands from wt or Tg mice were mounted. Computational analysis of mammary tree total extension (A), number of branching (B) or end buds (C) was performed using a dedicated software. No significant difference was observed between wt and Tg mice. F2wt: N = 4, F2 Tg: N = 6, F4wt: N = 6, F4Tg N = 5.

<https://doi.org/10.1371/journal.pone.0173931.g005>

cancer cells [5, 16, 17, 43, 44]. JAK2/STAT3 signaling was also reported to be involved in GPER signaling in the hypothalamus and SKBR-3 breast cancer cells [45, 46]. EGFR mRNA expression was not affected by ER $\alpha$ 36 overexpression in normal epithelial cells (data not shown) whereas a positive cross-activation of both gene expression has been reported in seminoma and breast cancer cells [16, 41, 42]. Since EGF is present in cell culture medium, ER $\alpha$ 36 protein could enhance a basal level of EGFR signaling through PI3K/AKT and STAT3/5. In vivo, a combination of EGF and progesterone treatment of ER $\alpha$ KO mice was shown to partially sustain mammary tree elongation and to modulate cytokine-cytokine receptor interaction (namely IL6/JAK2/STAT3 signaling), adhesion, TGF $\beta$  signaling and apoptosis, four functions identified in the functional analysis of DEGs [47].

The ER $\alpha$ 36 dependent stimulation of STAT3 expression and nuclear localization is of particular interest since constitutively activated STAT3 has been demonstrated to directly contribute to oncogenesis in various cancers by enhancing cell proliferation and migration but preventing cell apoptosis [48]. Indeed, STAT3 is supposed to be a promising therapeutic target of novel anticancer drugs like methyl ester derivative of synthetic triterpenoid which have been tested in a relevant model of ER negative breast cancers [49]. Together with NF- $\kappa$ B, which expression is also increased in ER $\alpha$ 36 overexpressing cells, STAT3 has been shown to play a pivotal role in the epigenetic switch linking inflammation to mammary gland tumorigenesis (i) in a model of MCF-10A cells containing and ER-src fusion exposed to tamoxifen as well as (ii) in the pre-tumorigenic state of mammary epithelium from mice constitutively overexpressing ERBB2 [50–52].

In this study, an in vivo mammary specific ER $\alpha$ 36 overexpression was performed through the production of a unique MMTV-ER $\alpha$ 36 transgenic mouse strain. At weaning, no difference could be observed at the microscopic level in the mammary gland between transgenic and wt mice. However, in ER $\alpha$ 36 overexpressing adult virgin females, the mammary gland displayed a



**Fig 6. Histological analysis of F4 wt and Tg mice at weaning and adulthood.** F<sub>4</sub> mammary gland slices from 3 week old (weaning) or 16 week old (adult) virgin female wt or Tg mice were stained with hematoxylin/eosin (A). S, stroma; E, epithelium; L, lumen; Ad, adipocyte. The epithelium thickness, the stromal thickness and lumen diameter of mammary ducts were measured in wt or Tg mice at weaning (B) and adulthood (C). Scale bar, 15  $\mu$ m. Each bar represents mean  $\pm$  S.E.M. N>5 per group. \*:  $P < 0.05$ , \*\*\*:  $P < 0.001$ .

<https://doi.org/10.1371/journal.pone.0173931.g006>

significant dilation of the ducts as well as stromal thickening and epithelium thinning and leakage. A similar but limited dilation was also reported by Bocchinfuso [53] after exogenous estradiol treatment of wild-type mice. These results suggest that the endogenous estrogens produced from puberty in ER $\alpha$ 36 transgenic mice could signal through the ER $\alpha$ 36 protein and mimic the overstimulated mammary estrogenic response observed in estradiol exposed wt mice. Since ER $\alpha$ 36 is able to down regulate ER $\alpha$ 66 expression or to behave as a dominant negative of ER $\alpha$ 66 activity [54], the abnormal epithelium and stromal phenotype could result from an alteration of the correct estrogen signaling pathway. In conclusion, molecular, cellular and tissue phenotype of ER $\alpha$ 36 overexpressing models evoke a constitutive estrogenic stimulation. Noteworthy, epithelium thinning and leakage are in line with lower proliferation rate and enhanced migration observed in vitro, respectively.

To date, commercially available antibodies gave poor results in our hands for the in situ detection of ER $\alpha$ 36 receptor in normal or cancerous mammary tissues. However, a basal expression of the ER $\alpha$ 36 mRNA (and sometimes of the protein) has been detected in many organs [18]. Since a high level of active ER $\alpha$ 36 is known to be one of the main actors of breast cancer progression [18], it would be noteworthy to investigate if ER $\alpha$ 36 function in cancer requires either an expression or/and an agonist concentration threshold. Indeed, our data from normal mammary epithelial cells suggest that, in the absence of ligand, ER $\alpha$ 36 overexpression is sufficient to trigger cell migration and partial resistance to apoptosis, two hallmarks of cancer cells. In the presence of a physiological adult level of estradiol, ER $\alpha$ 36 also mediates mammary epithelium disorganization in vivo, indicating a circuitous estrogen signaling in the presence of a high ER $\alpha$ 36 receptor level.

We previously demonstrated that estrogen mimicking compounds such as long chain alkyl-phenols stimulate GPER/ER $\alpha$ 36 signaling pathways in human seminoma cells [17]. Many papers, namely from Ana M Soto's laboratory also demonstrate that fetal or neonatal exposure to BPA or DES alter the estrogen signaling in the developing mouse or rat mammary gland of rodents and predispose animals to develop breast neoplasia later in adulthood [55]. Therefore, the unique model of MMTV-ER $\alpha$ 36 transgenic mice described in this study could serve to screen endogenous or xenobiotic molecules, acting as ER $\alpha$ 36 agonists or antagonists and to decipher their mechanisms of action. How and when ER $\alpha$ 36 expression is initiated and maintained in mammary epithelial cells remain to be investigated.

## Supporting information

**S1 Fig. Production of ER $\alpha$ 36 transgenic mice.** A. Structure of ER $\alpha$ 36 transgene. MMTV promoter was cloned from the pGL4.36[luc2P/MMTV/Hygro] reporter vector (Promega, France) and placed upstream ER $\alpha$ 36 complete cDNA sequence and a poly(A) stretch. Construction was then microinjected into B6SJLF2 fertilized eggs.

B. ER $\alpha$ 36 protein expression in transgenic adult mouse. Western blot analysis of ER $\alpha$ 36 protein expression in adult mammary gland (4 month old females). wt: wild-type; Tg: transgenic. Tubulin protein expression is used as loading control.

C. Sex ratio and transgene transmission rate in ER $\alpha$ 36 transgenic strain. The litters from 27 wild-type females mated with ER $\alpha$ 36 KI hemizygote males were analyzed for sex ratio and transgene transmission rate. Among 171 animals, 47.4% were males and 52.6% were females. A non-significant ( $P = 0.16$ ) lack of transgene transmission was observed, especially in females (40% measured *versus* 50% expected transgene transmission rate using Chi2 test). (TIF)

**S2 Fig. Mammary tree parameter analyses.** Mammary tree whole mounts images (A) from PND21 mice were skeletonized with a dedicated software (B). Total extension (white line), number of duct branching (green spots) and sprouts (red spots) were computed. (TIF)

**S3 Fig. Characterization of MCF-10A/ER $\alpha$ 36 cell line.** A. Real-time PCR analysis of ER $\alpha$ 36 expression in MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo cells. The ER $\alpha$ 36 mRNA expression level slightly detected by real-time PCR in MCF-10A/Zeo cells was set to 1. Several clones of MCF-10A/ER $\alpha$ 36 cells were tested for ER $\alpha$ 36 expression which varied from 2- to 42-fold the one of MCF-10A/Zeo cells. For further analyses, we selected the 36-4 clone in which ER $\alpha$ 36 expression was augmented by 13-fold compared to MCF-10A/Zeo cells. This overexpression level was in the range of the difference observed between MCF-10A cells and "naturally ER $\alpha$ 36 expressing" MDA-MB-231 breast cancer cells (Zou et al; 2009 [54]).

B. Western-blot analysis of ER $\alpha$ 36 in MCF-10A/Zeo and MCF-10A/ER $\alpha$ 36. ER $\alpha$ 36 protein is not detectable by western blotting in MCF10A/Zeo cells. However, ER $\alpha$ 36 expression is revealed by an anti-ER $\alpha$  antibody (G20) in MCF-10A/ER $\alpha$ 36 cells (clone 36-4 selected from panel A).

C. Immunofluorescence analysis of ER $\alpha$ 36 expression and localization in MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo cells. Merge images show nuclei stained blue with Hoechst and ER $\alpha$ 36 stained red by anti-ER $\alpha$ 36 rabbit polyclonal primary antibody and anti-rabbit-Alexafluor 555 secondary antibody (clone 36-4 selected from panel A). Scale bar = 50 $\mu$ m. (TIF)

**S4 Fig. Overall strategy of microarray data bioinformatic analysis.** A. Differentially expressed genes (DEGs) from MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo cells were submitted to

MSigDB<sup>®</sup> online tool to performed functional analyses. This led to the identification of the functions and signaling pathways mostly affected by ER $\alpha$ 36 overexpression. Ingenuity pathway analysis (IPA<sup>®</sup>) determined the DEG upstream regulators and a dedicated Matlab<sup>®</sup> software was designed to identify intermediate regulators acting downstream ER $\alpha$ 36 and upstream DEGs.

B. Example of one hierarchized gene network built by iterative computation of resulting adjacency matrices with a dedicated software from Matlab<sup>®</sup>.  
(TIF)

**S5 Fig. ER $\alpha$ 36 overexpression modulates MAPK/Erk1/2, NF $\kappa$ B and JAK2/STAT3 signaling pathways in MCF-10A cells.**

A. Representative western blot analysis of Phospho-Erk1/2 (P-Erk), total Erk1/2 (t-Erk) PTEN, NF $\kappa$ B and STAT3 expression in MCF-10A/Zeo and MCF-10A/ER $\alpha$ 36 cells.  $\beta$ -Actin was used as a loading control.

B. Localization of NF $\kappa$ B and STAT3 was studied by immunofluorescence with specific antibodies: anti-NF $\kappa$ B p65, anti-STAT3 (red, AlexaFluor 555). Hoechst was used to stain the nuclei (blue). A nuclear translocation of NF- $\kappa$ B and STAT3 was observed in MCF10A/ER $\alpha$ 36 cells compared to MCF10A/Zeo cells. Scale bar = 50 $\mu$ m.

(TIF)

**S6 Fig. ER $\alpha$ 36 overexpression triggers loss of E-cadherin expression in adult transgenic mice.**

Representative western blot analysis of E-cadherin expression in wt and Tg mammary glands.  $\alpha$ -tubulin was used as a loading control.

(TIF)

## Acknowledgments

We are grateful to Emmanuel Bresso, Malika Smail-Tabbone and Marie Dominique Desvignes from LORIA, UMR7503 CNRS INRIA for their help and advises concerning the gene enrichment analysis and data mining. We thank Dominique Meng and Martine Chillet for technical support and histology, Alexandra Kleinclauss for flow cytometry assays and Fabrice Augé from GenomEast Platform (IGBMC, Strasbourg, France) for microarray experiment.

## Author Contributions

**Conceptualization:** HD AC TB.

**Data curation:** EHD TB.

**Formal analysis:** AC HD EHD TB.

**Funding acquisition:** AC HD TB.

**Investigation:** CT CC CM.

**Methodology:** CT CC AC CM.

**Project administration:** AC TB HD.

**Resources:** AC CC HD.

**Software:** EHD TB.

**Supervision:** HD TB AC.

**Validation:** AC HD TB.

**Visualization:** CT CC AC HD.

**Writing – original draft:** HD.

**Writing – review & editing:** AC CC.

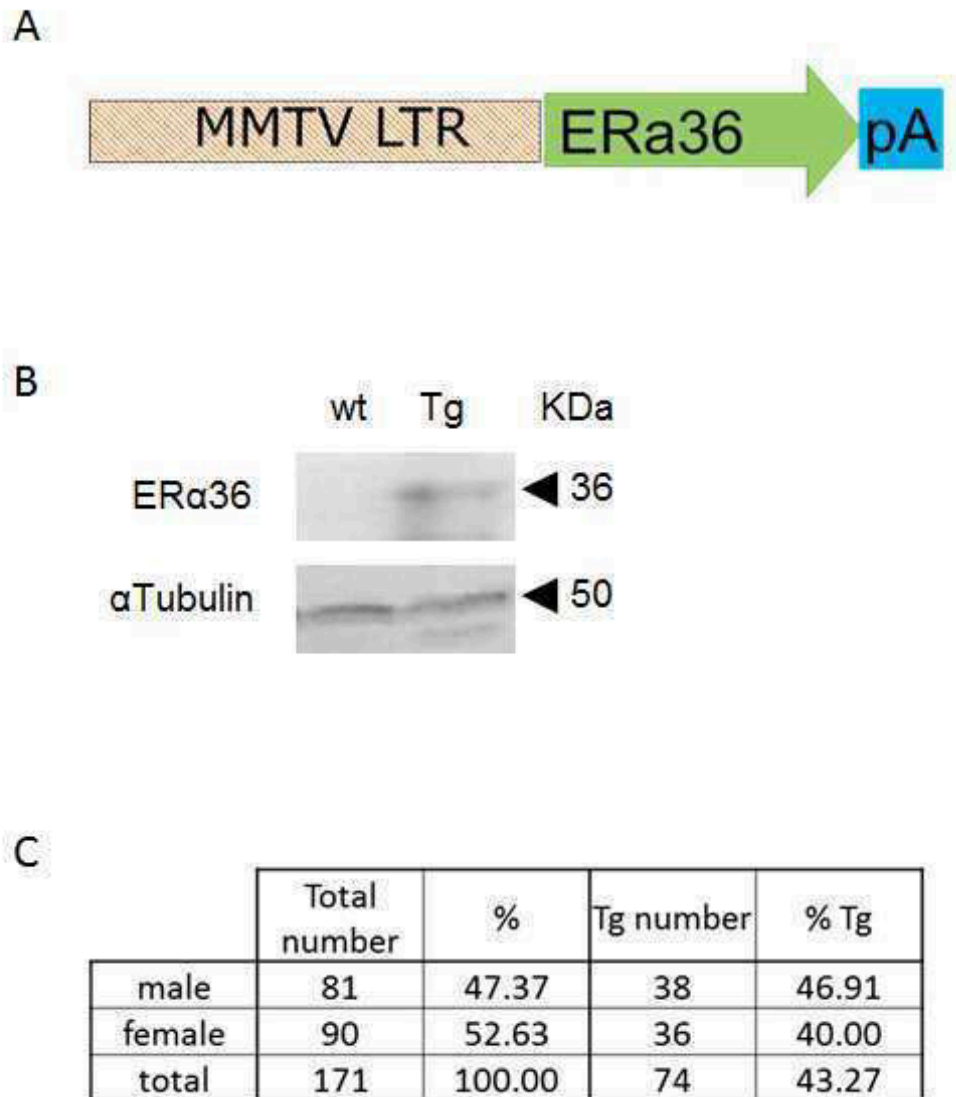
## References

1. Banerjee S, Chambliss KL, Mineo C, Shaul PW. Recent insights into non-nuclear actions of estrogen receptor alpha. *Steroids*. 2014; 81: 64–9. <https://doi.org/10.1016/j.steroids.2013.11.002> PMID: [24252382](https://pubmed.ncbi.nlm.nih.gov/24252382/)
2. Gibson D and Saunders P. Estrogen dependent signaling in reproductive tissues—A role for estrogen receptors and estrogen related receptors. *Mol Cell Endocrinol*. 2012; 348: 361–372. <https://doi.org/10.1016/j.mce.2011.09.026> PMID: [21964318](https://pubmed.ncbi.nlm.nih.gov/21964318/)
3. Qian H, Xuan J, Liu Y, Shi G. Function of G-Protein-Coupled Estrogen Receptor-1. In *Reproductive System Tumors*. *J Immunol Res*. 2016; 7128702.
4. Wang ZY, Zhang XT, Shen P, Loggie BW, Chang Y, Deuel TF. Identification, cloning, and expression of human estrogen receptor-alpha36, a novel variant of human estrogen receptor-alpha66. *Biochem Biophys Res Commun*. 2005; 336: 1023–1027. <https://doi.org/10.1016/j.bbrc.2005.08.226> PMID: [16165085](https://pubmed.ncbi.nlm.nih.gov/16165085/)
5. Kang LG, Zhang XT, Xie Y, Tu YP, Wang D, Liu ZM, et al. Involvement of estrogen receptor variant ER-alpha36, not GPR30, in nongenomic estrogen signaling. *Mol Endocrinol*. 2010; 24: 709–721. <https://doi.org/10.1210/me.2009-0317> PMID: [20197310](https://pubmed.ncbi.nlm.nih.gov/20197310/)
6. Zhang XT, Ding L, Kang LG, Wang Z-Y. Estrogen receptor alpha 36 mediates mitogenic antiestrogen signaling in ER-negative breast cancer cells. *PLoS ONE*. 2012; 7: e30174 <https://doi.org/10.1371/journal.pone.0030174> PMID: [22276155](https://pubmed.ncbi.nlm.nih.gov/22276155/)
7. Lin AH, Li RW, Ho EY, Leung GP, Leung SW, Vanhoutte PM, Man RY. Differential ligand binding affinities of human estrogen receptor- $\alpha$  isoforms. *PLoS ONE*. 2013; 30: e63199.
8. Panneerselvam M, Muthu K, Ramadas K. Structural insights into tumor-specific chaperoning activity of gamma synuclein in protecting estrogen receptor alpha 36 and its role in tamoxifen resistance in breast cancer. *Mol Biosyst*. 2015; 11:2998–3010. <https://doi.org/10.1039/c5mb00272a> PMID: [26299756](https://pubmed.ncbi.nlm.nih.gov/26299756/)
9. Chaudhri RA, Olivares-Navarrete R, Cuenca N, Hadadi A, Boyan BD, Schwartz Z. Membrane estrogen signaling enhances tumorigenesis and metastatic potential of breast cancer cells via estrogen receptor-alpha36 (ERalpha36). *J Biol Chem*. 2012; 287, 7169–7181 <https://doi.org/10.1074/jbc.M111.292946> PMID: [22247547](https://pubmed.ncbi.nlm.nih.gov/22247547/)
10. Shi L, Dong B, Li Z, Lu Y, Ouyang T, Li J, et al. Expression of ER- $\{\alpha\}$ 36, a novel variant of estrogen receptor  $\{\alpha\}$ , and resistance to tamoxifen treatment in breast cancer. *J Clin Oncol*. 2009; 27: 3423–3429. <https://doi.org/10.1200/JCO.2008.17.2254> PMID: [19487384](https://pubmed.ncbi.nlm.nih.gov/19487384/)
11. Chamard-Jovenin C, Jung AC, Chesnel A, Abecassis J, Flament S, Ledrappier S, Macabre C, Boukhobza T, Dumond H. From ER $\alpha$ 66 to ER $\alpha$ 36: a generic method for validating a prognosis marker of breast tumor progression. *BMC Syst Biol*. 2015; 9:28. <https://doi.org/10.1186/s12918-015-0178-7> PMID: [26080803](https://pubmed.ncbi.nlm.nih.gov/26080803/)
12. Kang L, Guo YM, Zhang XT, Wang ZY. A positive cross-regulation of HER2 and ER-alpha36 controls ALDH1 positive breast cancer cells. *J Steroid Biochem Mol Biol*. 2011; 127:262–268. <https://doi.org/10.1016/j.jsbmb.2011.08.011> PMID: [21907803](https://pubmed.ncbi.nlm.nih.gov/21907803/)
13. Zhang XT and Wang Z-Y. Estrogen receptor-  $\alpha$  variant, ER-  $\alpha$  36, is involved in tamoxifen resistance and estrogen hypersensitivity. *Endocrinology*. 2013; 154: 1990–1998. <https://doi.org/10.1210/en.2013-1116> PMID: [23546601](https://pubmed.ncbi.nlm.nih.gov/23546601/)
14. Zhang XT, Deng H, Wang ZY. Estrogen activation of the mitogen-activated protein kinase is mediated by ER-alpha36 in ER-positive breast cancer cells. *J Steroid Biochem Mol Biol*. 2014; 143: 434–443. <https://doi.org/10.1016/j.jsbmb.2014.06.009> PMID: [24973581](https://pubmed.ncbi.nlm.nih.gov/24973581/)
15. Yin L, Zhang XT, Bian XW, Guo YM, Wang ZY. Disruption of the ER-alpha36-EGFR/HER2 positive regulatory loops restores tamoxifen sensitivity in tamoxifen resistance breast cancer cells. *PLoS ONE*. 2014; 9 (9): e107369. <https://doi.org/10.1371/journal.pone.0107369> PMID: [25203051](https://pubmed.ncbi.nlm.nih.gov/25203051/)
16. Wallacides A, Chesnel A, Ajj H, Chillet M, Flament S, Dumond H. Estrogens promote proliferation of the seminoma-like Tcam-2 cell line through a GPER dependent ERalpha36 induction. *Mol Cell Endocrinol*. 2012; 350: 61–71. <https://doi.org/10.1016/j.mce.2011.11.021> PMID: [22138413](https://pubmed.ncbi.nlm.nih.gov/22138413/)
17. Ajj H, Chesnel A, Pinel S, Plenat F, Flament S, Dumond H. An alkylphenol mix promotes seminoma derived cell proliferation through an ERalpha36-mediated mechanism. *PLoS ONE*. 2013; 8(4):e61758. <https://doi.org/10.1371/journal.pone.0061758> PMID: [23626723](https://pubmed.ncbi.nlm.nih.gov/23626723/)

18. Sołtysik K and Czekaj P. ER $\alpha$ 36—Another piece of the estrogen puzzle. *Eur J Cell Biol.* 2015; 94(12):611–25. <https://doi.org/10.1016/j.ejcb.2015.10.001> PMID: 26522827
19. Elbaradie KB, Wang Y, Boyan BD, Schwartz Z. Sex-specific response of rat costochondral cartilage growth plate chondrocytes to 17 $\beta$ -estradiol involves differential regulation of plasma membrane associated estrogen receptors. *Biochim Biophys Acta.* 2013; 1833: 1165–1172. <https://doi.org/10.1016/j.bbamcr.2012.12.022> PMID: 23305904
20. Irsik DL, Carmines PK, Lane PH. Classical estrogen receptors and ER $\alpha$  splice variants in the mouse. *PLoS ONE.* 2013; 8:e70926. <https://doi.org/10.1371/journal.pone.0070926> PMID: 23940668
21. Yu L, Ke W, Wang Y, Ding W, Wang B, Huang S, et al. Predictive and prognostic value of ER- $\alpha$ 36 expression in breast cancer patients treated with chemotherapy. *Steroids.* 2014; 84:11–16. <https://doi.org/10.1016/j.steroids.2014.03.003> PMID: 24632025
22. Javed A and Lteif A. Development of the human breast. *Semin Plast Surg.* 2013; 27(1):5–12. <https://doi.org/10.1055/s-0033-1343989> PMID: 24872732
23. Inman JL, Robertson C, Mott JD, Bissell MJ. Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development.* 2015; 142: 1028–1042. <https://doi.org/10.1242/dev.087643> PMID: 25758218
24. Macias H, and Hinck L. Mammary gland development. *Wiley Interdiscip Rev Dev Biol.* 2012; 1(4):533–57.
25. Atwood CS, Hovey RC, Glover JP, Chepko G, Ginsburg E, Robison WG, Vonderhaar BK. Progesterone induces side-branching of the ductal epithelium in the mammary glands of peripubertal mice. *J Endocrinol.* 2000; 167(1):39–52. PMID: 11018751
26. Feng Y, Manka D, Wagner KU, Khan SA. Estrogen receptor- $\alpha$  expression in the mammary epithelium is required for ductal and alveolar morphogenesis in mice. *Proc Natl Acad Sci U S A.* 2007; 104(37):14718–23. <https://doi.org/10.1073/pnas.0706933104> PMID: 17785410
27. Förster C, Mäkela S, Wärrä A, Kietz S, Becker D, Hultenby K, Warner M, Gustafsson JA. Involvement of estrogen receptor beta in terminal differentiation of mammary gland epithelium. *Proc Natl Acad Sci U S A.* 2002; 99(24): 15578–83. <https://doi.org/10.1073/pnas.192561299> PMID: 12438700
28. Mallepell S, Krust A, Chambon P, Briskin C. Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. *Proc Natl Acad Sci U S A.* 2006; 103(7): 2196–201. <https://doi.org/10.1073/pnas.0510974103> PMID: 16452162
29. Brandt R, Eisenbrandt R, Leenders F, Zschesche W, Binas B, Juergensen C, Theuring F. Mammary gland specific hEGF receptor transgene expression induces neoplasia and inhibits differentiation. *Oncogene.* 2000; 19(17): 2129–37. <https://doi.org/10.1038/sj.onc.1203520> PMID: 10815804
30. Vandenberg LN, Maffini MV, Wadia PR, Sonnenschein C, Rubin BS, Soto AM. Exposure to environmentally relevant doses of the xenoestrogen bisphenol-A alters development of the fetal mouse mammary gland. *Endocrinology.* 2007; 148(1):116–27. <https://doi.org/10.1210/en.2006-0561> PMID: 17023525
31. Tylcz JB, El Alaoui-Lasmali K, Djermoune EH, Thomas N, Faivre B, Bastogne T. Data-driven modeling and characterization of anti-angiogenic molecule effects on tumoral vascular density. *Biomed Signal Process Control.* 2015; 20:52–60.
32. Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In: *Proceedings from MICCAI'98*, 11–13 October 1998, Cambridge, MA, Vol 1496 (Wells WM, Colchester A, Delp SL, Eds). Berlin, Germany: Computer Science Press. 130–137.
33. Howroyd P, Hoyle-Thacker R, Lyght O, Williams D, Kleymenova E. Morphology of the fetal rat testis preserved in different fixatives. *Toxicol Pathol.* 2005; 33(2):300–304. <https://doi.org/10.1080/01926230590896145> PMID: 15902974
34. Soule HD, Maloney TM, Wolman SR, Peterson WD Jr, Brenz R, McGrath CM, Russo J, Pauley RJ, Jones RF, Brooks SC. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res.* 1990; 50(18):6075–86. PMID: 1975513
35. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003; 34(3):267–273. <https://doi.org/10.1038/ng1180> PMID: 12808457
36. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005; 102(43):15545–15550. <https://doi.org/10.1073/pnas.0506580102> PMID: 16199517

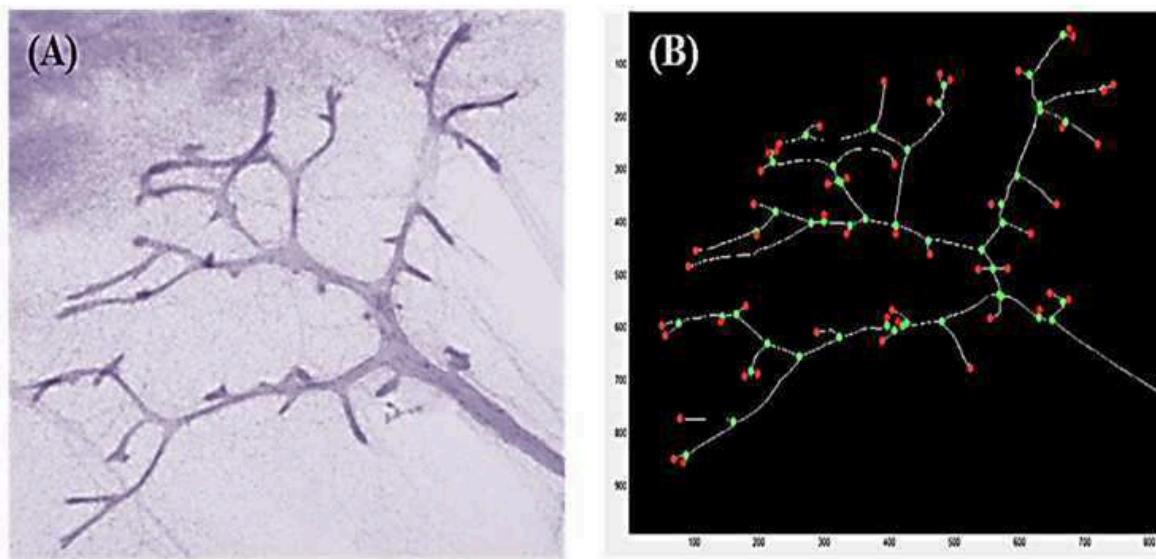


37. Chaudhri RA, Schwartz N, Elbradie K, Schwartz Z, Boyan BD. Role of ER $\alpha$ 36 in membrane associated signaling by estrogens. *Steroids*. 2014; 81:74–80. <https://doi.org/10.1016/j.steroids.2013.10.020> PMID: [24252378](https://pubmed.ncbi.nlm.nih.gov/24252378/)
38. Bjornstrom L, Sjoberg M. Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. *Mol Endocrinol*. 2005; 19:833–842. <https://doi.org/10.1210/me.2004-0486> PMID: [15695368](https://pubmed.ncbi.nlm.nih.gov/15695368/)
39. Marino M, Galluzzo P, Ascenzi P. Estrogen signaling multiple pathways to impact gene transcription. *Curr Genomics*. 2006; 7:497–508. PMID: [18369406](https://pubmed.ncbi.nlm.nih.gov/18369406/)
40. Fox EM, Bernaciak TM, Wen J, Weaver AM, Shupnik MA, Silva CM. Signal transducer and activator of transcription 5b, c-Src, and epidermal growth factor receptor signaling play integral roles in estrogen-stimulated proliferation of estrogen receptor-positive breast cancer cells. *Mol Endocrinol*. 2008; 22:1781–1796. <https://doi.org/10.1210/me.2007-0419> PMID: [18550772](https://pubmed.ncbi.nlm.nih.gov/18550772/)
41. Zhang XT, Kang LG, Ding L, Vranic S, Gatalica Z, Wang ZY. A positive feedback loop of ER- $\alpha$ 36/EGFR promotes malignant growth of ER-negative breast cancer cells. *Oncogene*. 2011; 30:770–780. <https://doi.org/10.1038/onc.2010.458> PMID: [20935677](https://pubmed.ncbi.nlm.nih.gov/20935677/)
42. Wang ZY, Yin L. Estrogen receptor  $\alpha$ -36 (ER- $\alpha$ 36): A new player in human breast cancer. *Mol Cell Endocrinol*. 2015; 3:193–206.
43. Gao F, Ma X, Ostmann AB, Das SK. GPR30 activation opposes estrogen dependent uterine growth via inhibition of stromal ERK1/2 and estrogen receptor  $\alpha$  (ER $\alpha$ ) phosphorylation signals. *Endocrinology*. 2011; 152:1434–1447. <https://doi.org/10.1210/en.2010-1368> PMID: [21303939](https://pubmed.ncbi.nlm.nih.gov/21303939/)
44. Albanito L, Madeo A, Lappano R, Vivacqua A, Rago V, Carpino A. G protein-coupled receptor 30 (GPR30) mediates gene expression changes and growth response to 17 $\beta$ -estradiol and selective GPR30 ligand G-1 in ovarian cancer cells. *Cancer Res*. 2007; 67:1859–1866. <https://doi.org/10.1158/0008-5472.CAN-06-2909> PMID: [17308128](https://pubmed.ncbi.nlm.nih.gov/17308128/)
45. Kwon O, Kang ES, Kim I, Shin S, Kim M, Kwon S, Oh SR, Ahn YS, Kim CH. GPR30 mediates anorectic estrogen-induced STAT3 signaling in the hypothalamus. *Metabolism*. 2014; 63(11):1455–1461. <https://doi.org/10.1016/j.metabol.2014.07.015> PMID: [25200186](https://pubmed.ncbi.nlm.nih.gov/25200186/)
46. Wang J, Xu J, An X, Lyu J. Estrogen activates GPER mediated IL-6 / STAT3 signaling pathway to enhance proliferation in breast cancer SKBR-3 cells. *Journal of Third Military Medical University*. 2015; 4:340–345.
47. Mehta RG, Hawthorne M, Mehta RR, Torres KE, Peng X, McCormick DL, Kopelovich L. Differential roles of ER $\alpha$  and ER $\beta$  in normal and neoplastic development in the mouse mammary gland. *PLoS ONE*. 2014; 9(11):e113175. <https://doi.org/10.1371/journal.pone.0113175> PMID: [25405629](https://pubmed.ncbi.nlm.nih.gov/25405629/)
48. Kamran MZ, Patil P, Gude RP. Role of STAT3 in cancer metastasis and translational advances. *Biomed Res Int*. 2013; 421821.
49. Tran K, Risingsong R, Royce D, Williams CR, Sporn MB, Liby K. The synthetic triterpenoid CDDO-methyl ester delays estrogen receptor-negative mammary carcinogenesis in polyoma middle T mice. *Cancer Prev Res (Phila)*. 2012; 5(5):726–34.
50. Iliopoulos D, Jaeger SA, Hirsch HA, Bulyk ML, Struhl K. STAT3 activation of miR-21 and miR-181b-1 via PTEN and CYLD are part of the epigenetic switch linking inflammation to cancer. *Mol Cell*. 2010; 39(4):493–506. <https://doi.org/10.1016/j.molcel.2010.07.023> PMID: [20797623](https://pubmed.ncbi.nlm.nih.gov/20797623/)
51. Iliopoulos Dimitrios, Hirsch Heather A., and Kevin Struhl<sup>\*</sup> An epigenetic switch involving NF- $\kappa$ B, Lin28, let-7 microRNA, and IL6 links inflammation to cell transformation. *Cell*. 2009; 139(4): 693–706. <https://doi.org/10.1016/j.cell.2009.10.014> PMID: [19878981](https://pubmed.ncbi.nlm.nih.gov/19878981/)
52. Peng M, Ball-Kell SM, Tyner AL. Protein tyrosine kinase 6 promotes ERBB2-induced mammary gland tumorigenesis in the mouse. *Cell Death Dis*. 2015; 6:e1848. <https://doi.org/10.1038/cddis.2015.210> PMID: [26247733](https://pubmed.ncbi.nlm.nih.gov/26247733/)
53. Bocchinfuso WP, Lindzey JK, Hewitt SC, Clark JA, Myers PH, Cooper R, Korach KS. Induction of mammary gland development in estrogen receptor- $\alpha$  knockout mice. *Endocrinology*. 2000; 141(8):2982–94.
54. Zou Y, Ding L, Coleman M, Wang ZY. Estrogen receptor- $\alpha$  (ER- $\alpha$ ) suppresses expression of its variant ER- $\alpha$  36. *FEBS Lett*. 2009; 583:1368–1374. <https://doi.org/10.1016/j.febslet.2009.03.047> PMID: [19328202](https://pubmed.ncbi.nlm.nih.gov/19328202/)
55. Paulose T, Speroni L, Sonnenschein C, Soto AM. Estrogens in the wrong place at the wrong time: Fetal BPA exposure and mammary cancer. *Reprod Toxicol*. 2015; 54:58–65. <https://doi.org/10.1016/j.reprotox.2014.09.012> PMID: [25277313](https://pubmed.ncbi.nlm.nih.gov/25277313/)



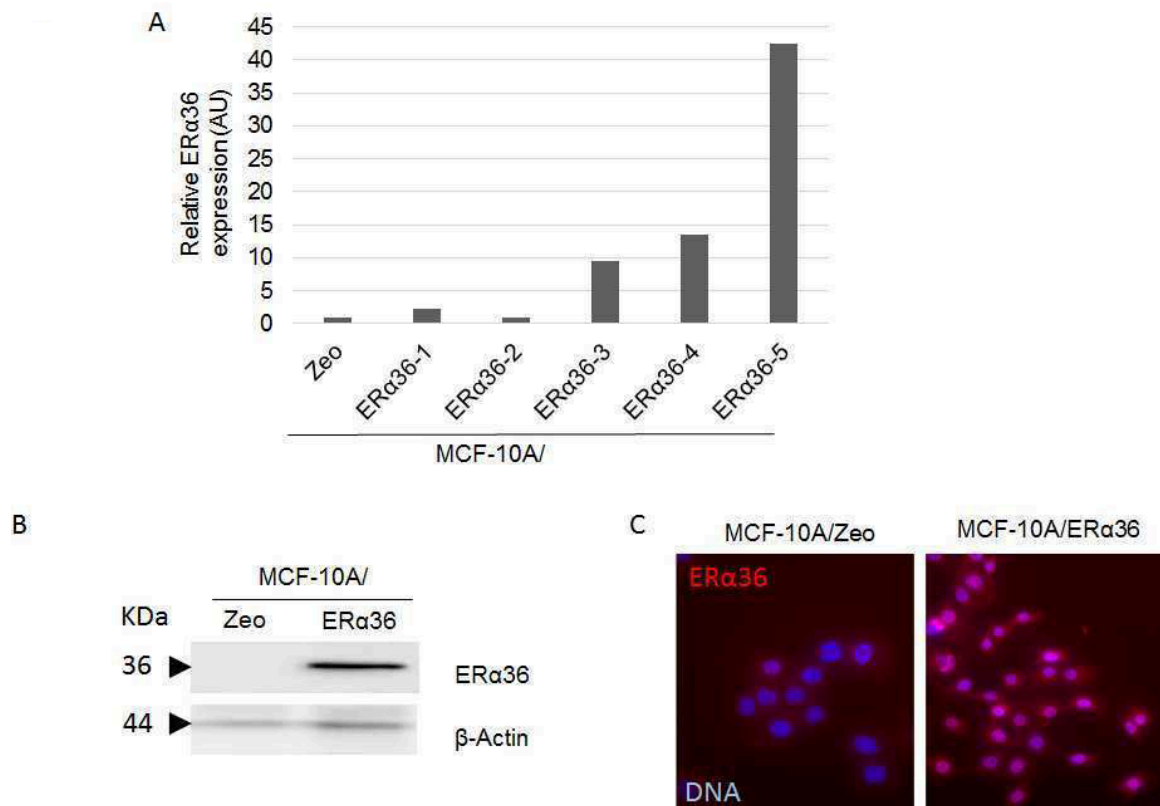
**Figure S1 : Production of ER $\alpha$ 36 transgenic mice.**

A. Structure of ER $\alpha$ 36 transgene. MMTV promoter was cloned from the pGL4.36[luc2P/MMTV/Hygro] reporter vector (Promega, France) and placed upstream ER $\alpha$ 36 complete cDNA sequence and a poly(A) stretch. Construction was then microinjected into B6SJLF2 fertilized eggs. B. ER $\alpha$ 36 protein expression in transgenic adult mouse. Western blot analysis of ER $\alpha$ 36 protein expression in adult mammary gland (4 month old females). wt: wild-type; Tg: transgenic. Tubulin protein expression is used as loading control. C. Sex ratio and transgene transmission rate in ER $\alpha$ 36 transgenic strain. The litters from 27 wild-type females mated with ER $\alpha$ 36 KI hemizygotic males were analyzed for sex ratio and transgene transmission rate. Among 171 animals, 47.4% were males and 52.6% were females. A non-significant ( $P=0.16$ ) lack of transgene transmission was observed, especially in females (40% measured versus 50% expected transgene transmission rate using Chi2 test).



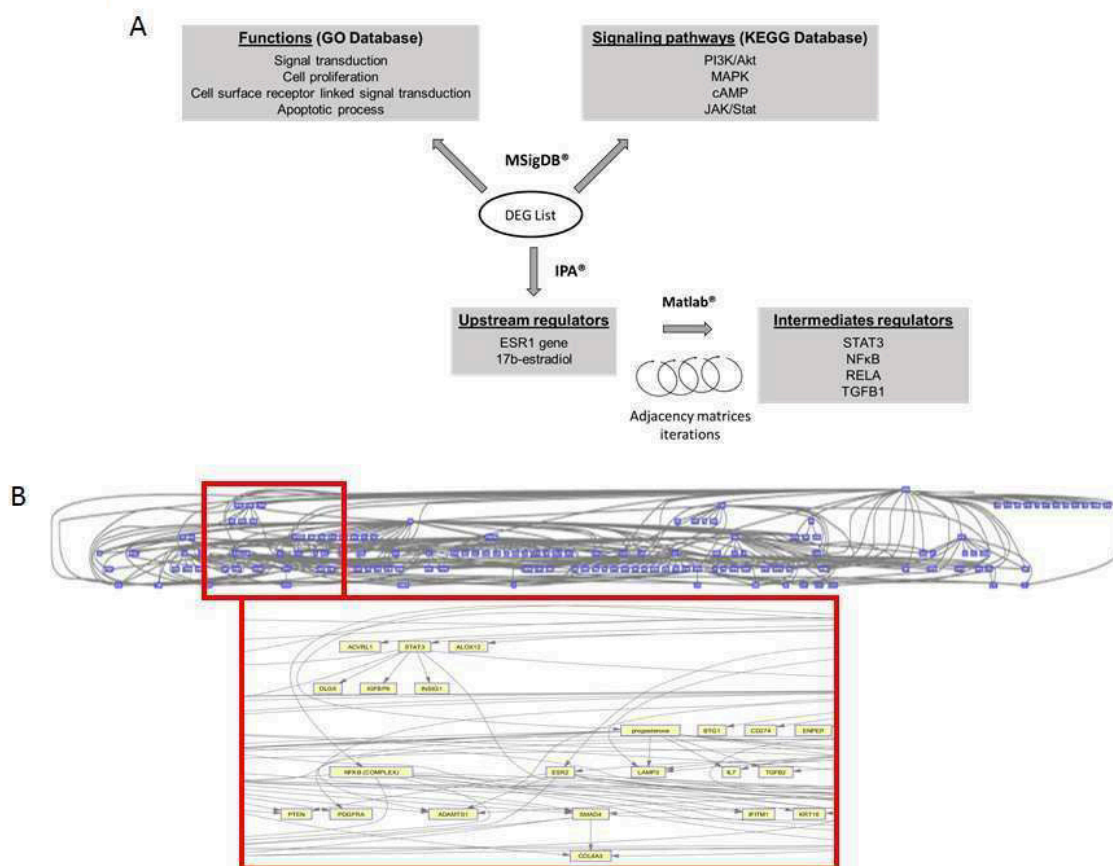
**Figure S2 : Mammary tree parameter analyses.**

Mammary tree whole mounts images (A) from PND21 mice were skeletonized with a dedicated software (B). Total extension (white line), number of duct branching (green spots) and sprouts (red spots) were computed.



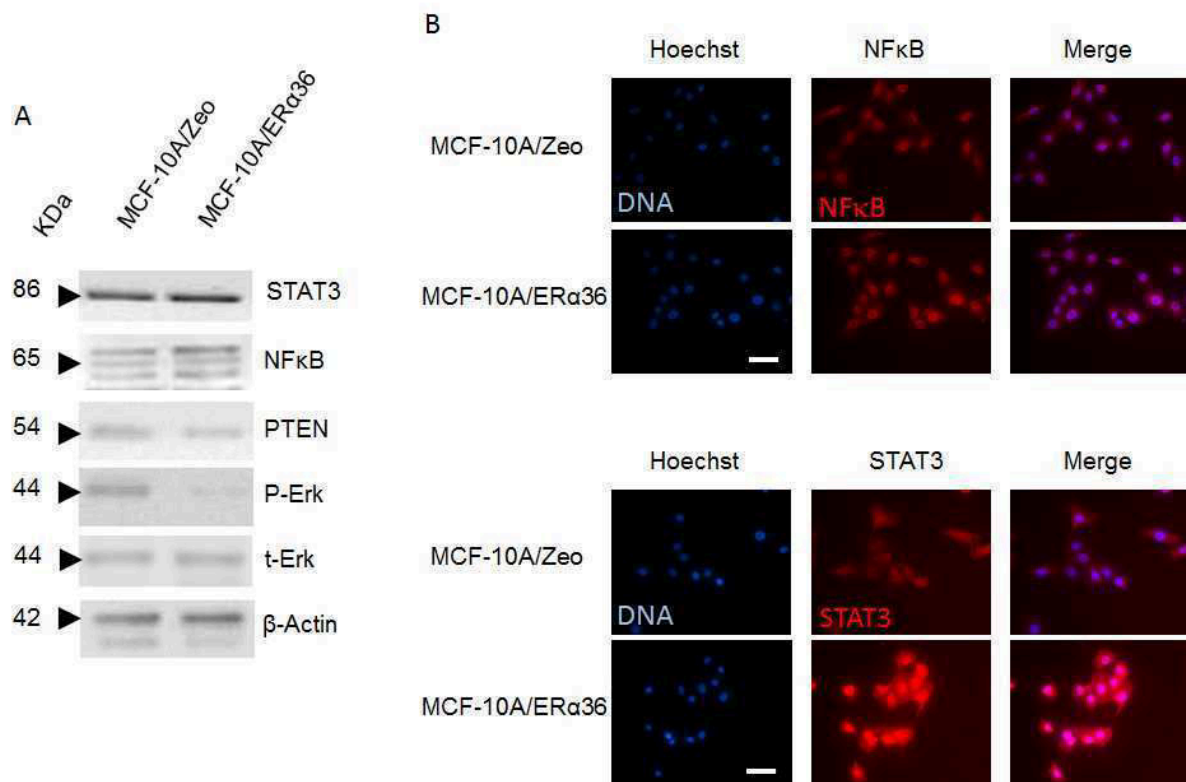
**Figure S3 : Characterization of MCF-10A/ERα36 cell line.**

A. Real-time PCR analysis of ERα36 expression in MCF-10A/ERα36 and MCF-10A/Zeo cells. The ERα36 mRNA expression level slightly detected by real-time PCR in MCF-10A/Zeo cells was set to 1. Several clones of MCF10A/ERα36 cells were tested for ERα36 expression which varied from 2- to 42-fold the one of MCF-10A/Zeo cells. For further analyses, we selected the 36–4 clone in which ERα36 expression was augmented by 13-fold compared to MCF-10A/Zeo cells. This overexpression level was in the range of the difference observed between MCF-10A cells and “naturally ERα36 expressing” MDA-MB-231 breast cancer cells (Zou et al; 2009 [54]). B. Western-blot analysis of ERα36 in MCF-10A/Zeo and MCF-10A/ERα36. ERα36 protein is not detectable by western blotting in MCF10A/Zeo cells. However, ERα36 expression is revealed by an anti-ERα antibody (G20) in MCF-10A/ERα36 cells (clone 36–4 selected from panel A). C. Immunofluorescence analysis of ERα36 expression and localization in MCF-10A/ERα36 and MCF-10A/Zeo cells. Merge images show nuclei stained blue with Hoechst and ERα36 stained red by anti-ERα36 rabbit polyclonal primary antibody and anti-rabbit-Alexafluor 555 secondary antibody (clone 36–4 selected from panel A). Scale bar = 50μm



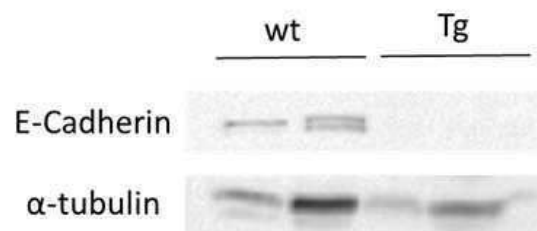
**Figure S4 : Overall strategy of microarray data bioinformatic analysis.**

A. Differentially expressed genes (DEGs) from MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo cells were submitted to MSigDB 1 online tool to performed functional analyses. This led to the identification of the functions and signaling pathways mostly affected by ER $\alpha$ 36 overexpression. Ingenuity pathway analysis (IPA 1) determined the DEG upstream regulators and a dedicated Matlab1 software was designed to identify intermediate regulators acting downstream ER $\alpha$ 36 and upstream DEGs. B. Example of one hierarchized gene network built by iterative computation of resulting adjacency matrices with a dedicated software from Matlab1.



**Figure S5 : ER $\alpha$ 36 overexpression modulates MAPK/Erk1/2, NF $\kappa$ B and JAK2/STAT3 signaling pathways in MCF-10A cells.**

A. Representative western blot analysis of Phospho-Erk1/2 (P-Erk), total Erk1/2 (t-Erk) PTEN, NF $\kappa$ B and STAT3 expression in MCF-10A/Zeo and MCF10A/ER $\alpha$ 36 cells.  $\beta$ -Actin was used as a loading control. B. Localization of NF $\kappa$ B and STAT3 was studied by immunofluorescence with specific antibodies: anti-NF $\kappa$ B p65, anti-STAT3 (red, AlexaFluor 555). Hoechst was used to stain the nuclei (blue). A nuclear translocation of NF- $\kappa$ B and STAT3 was observed in MCF10A/ER $\alpha$ 36 cells compared to MCF10A/Zeo cells. Scale bar = 50 $\mu$ m.



**Figure S6 : ER $\alpha$ 36 overexpression triggers loss of E-cadherin expression in adult transgenic mice.**

Representative western blot analysis of E-cadherin expression in wt and Tg mammary glands.  $\alpha$ -tubulin was used as a loading control

## *Publication N°2*

“Low-dose alkylphenol exposure promotes mammary epithelium alterations and transgenerational developmental defects, but does not enhance tumorigenic behaviour of breast cancer cells”

C. Chamard-Jovenin\*, C. Thiebaut\*, A. Chesnel, E. Bresso, C. Morel, M. Smaïl-Tabbone, M.D. Devignes, T. Boukhobza, H. Dumond.

\* Contribution égale

Article publié dans *Frontiers in Endocrinology* en Octobre 2017.

Une exposition fœtale ou néonatale aux alkylphénols à longue chaîne est suspectée d’altérer le développement physiologique de la glande mammaire, et d’augmenter ainsi le risque de cancer du sein. Néanmoins, les mécanismes de cette programmation fœtale des cancers diagnostiqués chez l’adulte restent à explorer.

Au cours de ce travail, des cellules épithéliales mammaires saines immortalisées (MCF-10A) ont été exposées *in vitro* à des faibles doses d’un mélange réaliste d’alkylphénols à longue chaîne (4-nonylphénol + 4-tert-octylphénol). Des analyses du transcriptome et du phénotype des cellules exposées en cinétique à ce mélange, combinées à une analyse fonctionnelle et à une modélisation des voies de signalisation impliquées, indiquent que les alkylphénols à longue chaîne induisent une augmentation des capacités prolifératives et migratrices, ainsi qu’une résistance accrue à l’apoptose. Nos résultats mettent également en lumière les mécanismes moléculaires sous-jacents faisant intervenir le variant du récepteur aux œstrogènes ER $\alpha$ 36.

En parallèle, un modèle transgénérationnel de souris exposées à trois doses réalistes d’alkylphénols a été mis en place pour déterminer si, et comment, cette exposition peut impacter l’architecture de la glande mammaire.



L'analyse des glandes mammaires de souris F3, obtenues après gavage buccal de souris gestantes C57BL/6J P0 pendant toute la durée de la gestation, montre effectivement une altération de l'histologie des canaux mammaires qui rappelle les phénotypes observés *in vitro* sur les cellules épithéliales mammaires saines.

Puisque les phénotypes *in vivo* et *in vitro* semblent similaires et font intervenir tous deux le variant humain ER $\alpha$ 36, les conséquences de l'exposition aux alkylphénols pourraient être transposées du modèle murin à l'Homme. Néanmoins, un traitement transitoire aux alkylphénols des cellules épithéliales mammaires saines MCF-10A, combiné à une surexpression d'ER $\alpha$ 36 n'est pas suffisante pour initier la tumorigenèse chez des souris nude xénotreffées.

En conclusion, il reste à déterminer si une exposition transgénérationnelle à de faibles doses d'alkylphénols et le développement anormal de la glande mammaire qui en découle, peuvent être à l'origine d'une plus grande susceptibilité au cancer du sein.



# Low-Dose Alkylphenol Exposure Promotes Mammary Epithelium Alterations and Transgenerational Developmental Defects, But Does Not Enhance Tumorigenic Behavior of Breast Cancer Cells

## OPEN ACCESS

### Edited by:

Vance Trudeau,  
University of Ottawa,  
Canada

### Reviewed by:

Antimo Migliaccio,  
Università degli Studi della  
Campania "L. Vanvitelli"  
Naples, Italy  
Giulia Ricci,  
Università degli Studi della  
Campania "Luigi Vanvitelli"  
Caserta, Italy

### \*Correspondence:

Hélène Dumond  
helene.dumond@univ-lorraine.fr

<sup>†</sup>These authors have equally  
contributed to the work and should  
be considered as co-first authors.

### Specialty section:

This article was submitted to  
Experimental Endocrinology,  
a section of the journal  
Frontiers in Endocrinology

**Received:** 16 June 2017

**Accepted:** 27 September 2017

**Published:** 23 October 2017

### Citation:

Chamard-Jovenin C, Thiebaut C,  
Chesnel A, Bresso E, Morel C,  
Small-Tabbone M, Devignes M-D,  
Boukhobza T and Dumond H (2017)  
Low-Dose Alkylphenol Exposure  
Promotes Mammary Epithelium  
Alterations and Transgenerational  
Developmental Defects, But Does  
Not Enhance Tumorigenic Behavior of  
Breast Cancer Cells.  
Front. Endocrinol. 8:272.  
doi: 10.3389/fendo.2017.00272

Clémence Chamard-Jovenin<sup>1†</sup>, Charlène Thiebaut<sup>1†</sup>, Amand Chesnel<sup>1</sup>,  
Emmanuel Bresso<sup>2,3,4</sup>, Chloé Morel<sup>1</sup>, Malika Smail-Tabbone<sup>2,3,4</sup>, Marie-Dominique  
Devignes<sup>2,3,4</sup>, Taha Boukhobza<sup>1</sup> and Hélène Dumond<sup>1\*</sup>

<sup>1</sup> CNRS-Université de Lorraine, UMR 7039, Centre de Recherche en Automatique de Nancy, BP70239,  
Vandoeuvre-lès-Nancy, France, <sup>2</sup> Université de Lorraine, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, France,  
<sup>3</sup> Inria, Villers-lès-Nancy, France, <sup>4</sup> CNRS, LORIA, UMR 7503, Vandoeuvre-lès-Nancy, France

Fetal and neonatal exposure to long-chain alkylphenols has been suspected to promote breast developmental disorders and consequently to increase breast cancer risk. However, disease predisposition from developmental exposures remains unclear. In this work, human MCF-10A mammary epithelial cells were exposed *in vitro* to a low dose of a realistic (4-nonylphenol + 4-tert-octylphenol) mixture. Transcriptome and cell-phenotype analyses combined to functional and signaling network modeling indicated that long-chain alkylphenols triggered enhanced proliferation, migration ability, and apoptosis resistance and shed light on the underlying molecular mechanisms which involved the human estrogen receptor alpha 36 (ER $\alpha$ 36) variant. A male mouse-inherited transgenerational model of exposure to three environmentally relevant doses of the alkylphenol mix was set up in order to determine whether and how it would impact on mammary gland architecture. Mammary glands from F3 progeny obtained after intrabuccal chronic exposure of C57BL/6J P0 pregnant mice followed by F1–F3 male inheritance displayed an altered histology which correlated with the phenotypes observed *in vitro* in human mammary epithelial cells. Since cellular phenotypes are similar *in vivo* and *in vitro* and involve the unique ER $\alpha$ 36 human variant, such consequences of alkylphenol exposure could be extrapolated from mouse model to human. However, transient alkylphenol treatments combined to ER $\alpha$ 36 overexpression in mammary epithelial cells were not sufficient to trigger tumorigenesis in xenografted Nude mice. Therefore, it remains to be determined if low-dose alkylphenol transgenerational exposure and subsequent abnormal mammary gland development could account for an increased breast cancer susceptibility.

**Keywords:** mammary gland, alkylphenol mix, cancer, development, estrogen receptor alpha 36

## INTRODUCTION

Mammary duct network development initiates during fetal life and continues until first lactation in adulthood. Cellular and molecular data support the parallel between mammary gland morphogenesis processes and breast cancer initiation and progression (1, 2). Recent works indicate that fetal and/or neonatal exposure to xenoestrogens—such as bisphenol A (BPA)—alters the mammary gland development and subsequently drives an increased long-term risk of breast cancer (3, 4). Like BPA, long-chain alkylphenols are known to exert estrogen-like activities through binding to classical nuclear estrogen receptor alpha (ER $\alpha$ ) and/or membrane-bound G protein-coupled estrogen receptor (GPER) (5). In 2004, the World Health Organization (WHO) published a nonylphenol NOAEL of 15 mg/kg/day for rat oral exposure and indicated that this value could be extrapolated to human (6). Indeed, previous studies by Chapin et al. (7) confirmed by Tyl et al. (8) demonstrated a lack of any transgenerational reproductive toxicity of dietary 4-nonylphenol (4NP) exposure during late gestation in Sprague–Dawley rats at doses ranging from 15 to 150 mg/kg/day. However, prenatal exposure to high doses of 4NP (25 mg/kg/day in MMTV-neu mice and 100 mg/kg/day in rats) causes altered development of the mammary gland, changes in steroid-receptor activation as well as increased synthesis of liver estriol (9, 10) in F1 pups. *In vitro*, 1  $\mu$ M 4NP can also stimulate the proliferation of MCF-7 estrogen-sensitive breast cancer cells (11). More recently, Raecker et al. (12) indicated that 4NP and in a less-extent 4-tert-octylphenol (4tOP) are ubiquitously found in commercially available food intended for babies and toddlers. Based on consumption studies, the authors concluded that the infant intake of this mix ranges from 0.23 to 0.65  $\mu$ g/kg/day, which is about 1,000 times less than the doses previously experimented before.

Recently, we have demonstrated that such a realistic dose of a 4NP:4tOP mixture can stimulate proliferation and modulate DNMT3L-dependent epigenetic status of testicular cancer germ cells (13). On the basis of this analysis, we assumed that a fetal exposure to low doses of alkylphenols could disturb male epigenetic imprinting establishment and drive male-inherited transgenerational alterations of reproductive organs. Indeed, such defects were described following fetal and/or neonatal exposure to environmental compounds, namely, BPA and vinclozolin (14). In the same study, we also demonstrated that the alkylphenol mix acts through a rapid, estrogen receptor alpha 36 (ER $\alpha$ 36) dependent non-genomic pathway (13).

Estrogen receptor alpha 36 is a variant of the canonical human ER $\alpha$  lacking the helix 9–12 domain, likely to bind a great diversity of compounds (15). Recently, we showed that expression of human ER $\alpha$ 36 transgene triggers neoplastic-like alteration of mice mammary gland (16). Since ER $\alpha$ 36 can enhance expression of migration/invasion markers in normal epithelial cells, we wondered if its expression or activity could be stimulated under alkylphenol exposure, alter mammary epithelial cell differentiation, and consequently lead to developmental, structural, or functional disruption of the mammary gland.

First, we explored the molecular and cellular impact of a 4NP:4tOP mix exposure in the human MCF-10A mammary

epithelial cell line by transcriptome and cell-phenotype analyses combined to functional and signaling network modeling. Our data point out ER $\alpha$ 36 variant as a key node of the gene networks modulated by alkylphenols, involved in proliferation, migration, and apoptosis escape control. In order to investigate low-dose alkylphenol effect on the fetal and neonatal mammary gland morphogenesis, we set up a transgenerational assay of intrabuccal chronic exposure of wild-type or ER $\alpha$ 36 transgenic (Tg) C57BL/6J P0 pregnant mice. Male-inherited exposure effects were assessed in F3 litters. The results strongly suggest that a low-dose alkylphenol exposure could promote a mammary gland abnormal development. Finally, we addressed if ER $\alpha$ 36 overexpression in xenografted MCF-10A cells combined to systemic alkylphenol exposure would be sufficient to trigger mammary tumor formation in adult Nude mice.

## MATERIALS AND METHODS

### Reagents

The test compounds, 4NP (CAS no. 84852-15-3) and 4tOP (CAS no. 140-66-9) were purchased from Sigma-Aldrich. 4tOP and 4NP were mixed based on their realistic concentration ratio (1:30) in infant food (12), thus forming the working mix called M4. In addition, 10-mM stock solutions (the concentration refers to 4NP in M4 mix) were prepared in either sesame oil (vehicle) for mice gavage or dimethylsulfoxide (DMSO) and further diluted in culture medium without phenol red for *in vitro* cell treatment. All working solutions were freshly made just before treatment and control cells were treated with DMSO, diluted with the same factor ( $1 \times 10^{-7}$ ) as M4 and indicated as “vehicle” in the figures.

### Animal Care and Use

#### Animals

Pathogen-free C57BL/6J mice (here considered as wt, Charles River Laboratories) or ER $\alpha$ 36 Tg C57BL/6J mice (16) were maintained in temperature-controlled and light-controlled (10 h light, 14 h dark cycle) conditions in the Nancy Faculty of Pharmacy animal facility. All experimental procedures were approved by the French Minister of Research Committee for animal experiment (Protocol no. APAFIS#2168-2015110518268051 v5) in accordance with the Guide for Care and Use of Laboratory Animals. Cages, food, and bedding displayed negligible estrogenicity as tested by the supplier (Tecniplast); water was supplied from glass bottles only. Food (SAFE, France) and water were supplied *ad libitum*.

Transgenic C57BL/6J mice expressing a single copy of a human ER $\alpha$ 36 transgene in the mammary gland under the control of the MMTV promoter were obtained as described in Thiebaut et al. (16). Tg males were mated with 7-week-old C57BL/6J wt females in order to obtain 50% hemizygotic Tg litters wherein the transgene is paternally inherited (Figure S1 in Supplementary Material). Paternal inheritance of the transgene was preferred to the maternal transmission in order to avoid a potential effect of the transgene expression during lactation. Before weaning, litters from wt females and Tg males were genotyped as previously described (16). In our study, wt litters

issued from wt males and wt litters issued from hemizygote Tg males were compared.

### Animal Treatment

Alkylphenol treatment composed of M4 dissolved in vehicle [0.05 µg/kg/day (D1), 0.5 µg/kg/day (D2), or 5 µg/kg/day (D3)] or either sesame oil (vehicle, D0) was administered to 7-week-old virgin C57BL/6J P0 females [5 P0 dams/dose 0 (D0) to dose 3 (D3)] from the first day of mating until delivery. Intrabuccal gavage was preferred to the stressful intragastric gavage which does not account for oral absorption (17), or to subcutaneously implanted minipumps which were reported to inappropriate for the delivery of estrogenic compounds to the mammary glands in rats (18, 19). These doses were thought to be environmentally relevant according to Raecker et al. (12). In order to avoid any hormonal interference with mammary gland development, we only used progeny from primipare C57BL/6J females, in spite of their tendency to cannibalism. In this work, the paternally inherited transgenerational effect of treatment can be measured in the F3 generation since F1 and F2 generations can be considered as directly exposed to the M4 mix in P0 mothers. At weaning (i.e., postnatal day 21, PND21), F1D0 to F1D3 wt or Tg litters as well as F2D0 to F2D3 or F3D0 to F3D3 paternally exposed wt or Tg litters were sacrificed by lethal intraperitoneal injection of 2 mg/kg pentobarbital (Sanofi/Aventis) and processed for further analyses.

### Cell Grafting in Nude Mice

Pathogen-free, 7-week-old athymic NMRI-nu (nu/nu) virgin female mice were purchased from Janvier Laboratories (Le-Genest-St-Isle, France). Animals were housed in the conditions described above, in the animal facility of the laboratory (CRAN, Nancy), with free access to water supplied from glass bottles and standard food *ad libitum*. Experimental procedures were approved by the French Minister of Research Committee for animal experiment (Protocol no. APAFIS#6158-2016072016446146 v5). As described above, sesame oil (vehicle) or 0.5 µg M4/kg/day treatment was administered by intrabuccal gavage, during 2 weeks before and 2 weeks after cell grafting. The dose of M4 was chosen based on significant results obtained in mammary ducts in C57BL/6J strain (see Results).

Orthotopic cell grafts were performed by a subcutaneous injection of a  $3 \times 10^6$  cells/50-µL phosphate-buffered saline (PBS) suspension, in the right mammary gland no. 4. The non-injected contralateral gland was used as a control. The following three mammary cell lines were used: ERα36 expressing MDA-MB-231 breast cancer line as a positive control for tumor formation, MCF-10A/Zeo as a non-cancerous negative control, or the previously described MCF-10A/ERα36 line overexpressing the human ERα36 receptor (16). Animals were sacrificed 4 weeks after grafting by lethal intraperitoneal injection of pentobarbital (see the above text).

### Mammary Gland Whole Mounts

The whole mounts were prepared by spreading the mammary glands onto glass slides and stained as described by Vandenberg et al. (20). A dedicated Matlab program for analysis of the

mammary network, adapted from Tylcz et al. (21, 22), was previously described (16). The mammary network was quantified in terms of tree extension, branching, and amount of sprouts.

### Histology

To prepare paraffin sections, mammary glands were fixed with Davidson (23) for 24 h at room temperature, dehydrated, and embedded with paraffin (VWR). Moreover, 7-µm sections were cut on a Leitz rotary microtome (Leica) and mounted on Superfrost slides (Fisher Scientific); the entire mammary gland was sectioned at once. Sections were stained with hematoxylin/eosin/methyl green to determine the presence of epithelial cords. Sections with visible cords were used for further analysis.

### Microarray Experiment

#### Sample Preparation and Data Normalization

Transcriptional profile analyses of MCF-10A cells exposed to 1 nM M4 for 1, 8, or 24 h were performed in triplicates on Affymetrix GeneChip U133 2.0 by the GenomEast Platform (IGBMC, Strasbourg, France). Biotinylated cRNA targets were prepared using the Ambion “MessageAmp™ Premier RNA Amplification Kit” according to the Instruction Manual P/N 4386269 Revision D (revision date: May 16, 2008), starting from 200 ng of total RNA extracted with RNeasy Mini Kit (Quiagen). Following fragmentation, 10 µg of cRNAs were hybridized for 16 h at 45°C, 60 rpm on Human GeneChip® HG-U133 plus 2.0 arrays (Affymetrix). The chips were washed and stained in the GeneChip® Fluidics Station 450 (Affymetrix) using the FS450\_0004 script and scanned with the GeneChip® Scanner 3000 7G (Affymetrix) at a resolution of 1.56 µm. Raw data (CEL Intensity files) were extracted from the scanned images using the Affymetrix GeneChip® Command Console (AGCC) version 4.0. CEL files were further processed with Affymetrix Expression Console software version 1.3.1 to calculate probe set signal intensities, using Robust Multi-array Average (RMA) algorithms with default settings. Overall, 3,362 transcripts are differentially expressed compared to  $t = 0$ . Distinct kinetic variation pattern and transcript number in each pattern are described in Figure S2A in Supplementary Material. A total of 278 transcripts, 76 transcripts, and 2,743 transcripts were significantly up- or downregulated (absolute variation factor  $\geq 2$  in triplicate RNA samples, corrected  $p$ -value ( $p < 0.05$ ) with Benjamini–Hochberg method) after 1-, 8-, or 24 h 1 nM M4 exposure, respectively. Differentially expressed genes (DEGs) were defined as genes whose expression is significantly affected by the M4 exposure, at any time of the kinetic (Figure S2B in Supplementary Material).

### Gene Ontology (GO) and Pathway Analyses

Gene ontology enrichment analyses were performed using the Panther online tool (Protein ANalysis THrough Evolutionary Relationships)<sup>1</sup> (24). MSigDBv5.0 (Molecular Signature

<sup>1</sup><http://pantherdb.org>.

Database)<sup>2</sup> (25, 26) was used to achieve the pathway enrichment analyses. The overlap between DEG list and (i) GO gene sets derived from the Biological Process part of GO (Table S1 in Supplementary Material) or (ii) Kyoto Encyclopedia Genes and Genomes (KEGG) pathways (Table S2 in Supplementary Material) were computed.

## Cell Culture

MCF-10A cells were purchased in 2015 from ATCC<sup>®</sup> (CRL-10317) and maintained as described by Soule et al. (27). Briefly, cells were plated at a density of  $2 \times 10^4$  cells per well in 6-well plates in 5% HS-supplemented DMEM/F12 (GIBCO) medium for 24 h, then starved for 48 h in 2.5% charcoal-stripped HS-containing medium without phenol red. Treatments were performed on 24 h 0% HS-cultured cells.

## Transient Transfection and Stable Cell Lines Establishment

Stable MCF-10A cell lines transfected by pCDNA3.1-ER $\alpha$ 36 or the empty expression vector were obtained as previously described (13) and named MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo, respectively. The transfected cell lines were then subcloned and checked for ER $\alpha$ 36 expression by western blot and reverse transcription polymerase chain reaction (RT-PCR) analyses before each experiment. MCF-10A/Zeo cells behaved similarly to the non-transfected MCF-10A cells in response to 1 nM M4 exposure whatever the assay performed (proliferation or migration or survival assay, data not shown).

## Real-time PCR Analysis

RT and real-time PCR analyses were performed as previously described (13). The following primers were used for qRT-PCR: *RPLPO* forward (Fw) 5'-GGCGACCGTGAAGTCCAAC-3', *RPLPO* reverse (Rev) 5'-CCATCAGCACACAGCCTTC-3', *VIM* forward (Fw) 5'-GGGACGCAGACATCGTCATC-3', *VIM* reverse (Rev) 5'-TCGTCATCGTCGAAATGGGC-3', *CDH2* forward (Fw) 5'-ACAGTGGCCACCTACAAAGG-3', *CDH2* reverse (Rev) 5'-CCGAGATGGGGTTGATAATG-3'. Assays were performed at least in triplicate, and the mean values were used to calculate expression levels, using the  $\Delta\Delta C(t)$  method referring to *RPLPO* housekeeping gene expression.

## Crystal Violet Assay

The crystal violet assay was performed in 24-well plates. After each well was washed with PBS, the cells attached to the bottom of the plate were fixed and stained with 0.4% crystal violet solution in 2% ethanol for 30 min. After the plate was washed with water and dried, crystal violet was solubilized in 10% acetic acid and the absorbance at 570 nm was measured by a microplate reader (Victor x3, Perkin-Elmer).

## Flow Cytometry

After a 24 h 1 nM M4 or vehicle exposure, medium and adherent cells were harvested. Cells were washed twice with cold

PBS, resuspended at  $1.10^6$  cells/mL in PBS containing 4  $\mu$ g/mL propidium iodide (PI) (P3566, Life Technologies), and 0.4  $\mu$ g/ $\mu$ L RNAase A (Sigma-Aldrich) for 20 min at room temperature. PI fluorescence level was determined using the argon laser of a FACScalibur flow cytometer (Becton-Dickinson). For each sample, fluorescence intensity of 10,000 cells was analyzed using CellQuest Pro Software.

## Western Immunoblotting

Western blots were performed as described previously (13). The following primary antibodies were used: anti-P-Src Tyr416 (#6943P, Epitomics), anti-Cyclin B1 (#1495-1, Epitomics), anti-Cyclin D1 (#2922, Cell Signaling), anti-Cyclin E1 (CPA1171, Clinisciences), anti-Phospho-Rb ser 807/811 (#8516, Cell Signaling), anti-ER $\alpha$ 36 (CY1109, Cell Applications), anti-P-Caspase 9 ser 196 (sc-11755, Santa Cruz Biotechnology), anti-P-Bad ser136 (sc-7999, Santa Cruz Biotechnology), anti-Bad (#9292, Cell Signaling), anti-PARP1 cleaved (552596, BD Pharmingen) and anti-Caspase 7 cleaved (#9494, Cell Signaling), and anti-Caspase 3 cleaved (#9664, Cell Signaling). The anti- $\beta$ -Actin antibody (sc1615, Santa Cruz Biotechnology) or anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH, GTX100118, Genetex) were used as a loading controls. Protein expression profiles were revealed with Clarity Western ECL Substrate (Biorad) and banding quantification was performed using the Quantity One Chemidoc XRS software (Biorad).

## Immunofluorescence

Immunofluorescence was performed as described previously (13). The following primary antibodies were used: anti-NF $\kappa$ B (GTX102090, GeneTex), anti- $\beta$ -catenin (E247, Epitomics #1247-s), anti-E-cadherin (GTX100443, GeneTex), and anti-N-cadherin (TA326835, OriGene). Goat anti-rabbit secondary antibody was coupled to AlexaFluor 555 (Invitrogen). Images were obtained with DS-Ri1 Nikon camera and Eclipse80i Nikon microscope and quantifications were performed using NIS-Elements BR 4.20.00 software (Nikon).

## TUNEL Assay

TUNEL assay was performed using the Apo-BrdU-IHC *in-situ* DNA fragmentation Assay kit (BioVision, USA) following the manufacturer instructions for the staining of cell preparations fixed on slides adapted for the use of AlexaFluor 555 (Invitrogen) goat anti-mouse secondary antibody. Nuclear DNA was stained with Hoechst (bisBenzimide H33342 Trihydrochloride, Sigma-Aldrich). Quantification of stained cell number was performed with NIS elements BR 4.20.00 software (Nikon).

## Scratch-Wound Assay

Scratch assays were performed using the Ibidi-culture inserts (Ibidi<sup>®</sup>/Biovalley) following the manufacturer instructions. Cultures were then washed to remove detached cells and debris and treated with vehicle or 1 nM M4. Quantification of wound mean diameter were performed at  $t = 0$  and after a 6 h exposure by Phase-contrast image analysis with NIS-elements BR 4.20.00 software (Nikon).

<sup>2</sup><http://www.broadinstitute.org/gsea/msigdb/>.

## Statistical Analysis

Statistical analyses were performed with Matlab vR2014b software (MathWorks) by using Student's *t*-test for unpaired samples and ANOVA followed by Bonferroni *post-hoc* test for multigroup comparison, with a significant *p*-value threshold below 5%. \**p* < 0.05; \*\**p* < 0.01; \*\*\**p* < 0.001. SDs or SEs were indicated on figures as advocated by Altman and Bland (28).

## RESULTS

### Microarray and Bioinformatic Analysis of MCF-10A DEGs after Kinetic Alkylphenol Exposure

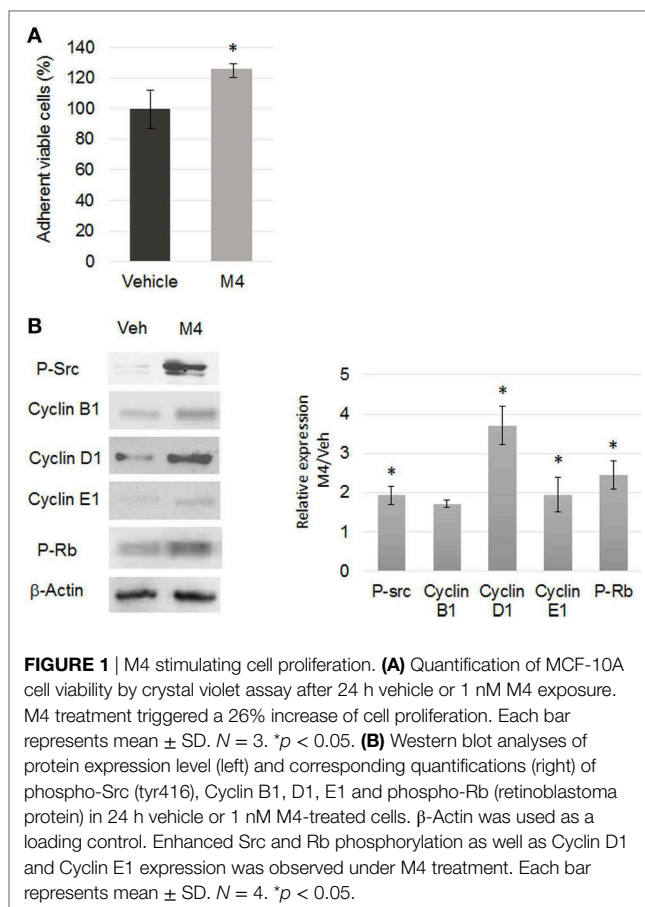
In order to determine which cell function could be altered by alkylphenol exposure, we analyzed the transcriptional profile of MCF-10A cells exposed to M4. A kinetic microarray analysis of gene expression was performed after a 1-, 8-, or 24 h 1 nM M4 exposure. GO enrichment analysis of DEGs (Table S1 in Supplementary Material) indicated low levels of enrichment after 1 h M4 exposure, which underscored cell response to stress and metabolic processes. No GO biological process was found significantly enriched after an 8 h exposure, probably due to the limited number of DEGs. After a 24 h M4 exposure, the DEG list was enriched with genes mainly assigned to GO terms related to DNA structure and stability, cell cycle and proliferation, and splicing and cell death (Table S1 in Supplementary Material).

At each time point of the kinetic (1, 8, and 24 h), pathway enrichment analysis identified the maximal overlap between DEG lists and KEGG pathways for the "Pathways in Cancer" pathway. DEGs related to this pathway were mainly involved in the three cellular functions previously highlighted by the GO enrichment analysis: proliferation, evading apoptosis, and migration/invasion (Table S2 in Supplementary Material). Moreover, this analysis suggested that the main intracellular signaling pathways involved in M4 response were the MAPK, JAK-STAT, and P53-dependent ones (Table S2 in Supplementary Material). The accuracy of the *in silico* derived predictions was investigated by examining the MCF-10A cell phenotypes after M4 exposure.

### Focus on Cell Cycle

Expression of key genes and corresponding proteins involved in cell cycle control and regulation was determined by western blot analysis after a 24 h M4 exposure. Quantification of cell viability by crystal violet staining indicated a  $23 \pm 0.5\%$  augmentation after M4 versus vehicle 24 h exposure (*p* < 0.05, **Figure 1A**). Flow cytometry analyses confirmed that a 24 h M4 exposure triggered cell cycle entry even if the data were not statistically significant ( $66 \pm 1.3\%$  G0/G1 phase in vehicle exposed cells versus  $60 \pm 0.8\%$  G0/G1 phase in M4 exposed cells; *N* = 3). Moreover, the expression of more than 25 genes involved in all cell cycle steps was modified by alkylphenol exposure (**Table 1**), all of them sustaining enhanced cell cycle entry or progression.

**Figure 1B** shows that Src phosphorylation, Cyclin D1 and Cyclin E1 expression as well as Retinoblastoma (Rb) protein phosphorylation were significantly induced (*p* < 0.05).



**FIGURE 1** | M4 stimulating cell proliferation. **(A)** Quantification of MCF-10A cell viability by crystal violet assay after 24 h vehicle or 1 nM M4 exposure. M4 treatment triggered a 26% increase of cell proliferation. Each bar represents mean  $\pm$  SD. *N* = 3. \**p* < 0.05. **(B)** Western blot analyses of protein expression level (left) and corresponding quantifications (right) of phospho-Src (tyr416), Cyclin B1, D1, E1 and phospho-Rb (retinoblastoma protein) in 24 h vehicle or 1 nM M4-treated cells.  $\beta$ -Actin was used as a loading control. Enhanced Src and Rb phosphorylation as well as Cyclin D1 and Cyclin E1 expression was observed under M4 treatment. Each bar represents mean  $\pm$  SD. *N* = 4. \**p* < 0.05.

According to the microarray analysis, almost all genes encoding the mini chromosome maintenance (MCM) protein family involved in proper positioning of DNA replication forks as well as DNA mismatch repair protein encoding genes, MSH2 and MSH6, were upregulated after M4 exposure (Figure S3A in Supplementary Material). Therefore, MCF-10A cell line exposed to M4 or not was also tested for random amplification of polymorphic DNA (RAPD): among the 10 GC-rich decamer primers randomly designed to fingerprint the genomic DNA [Figure S3B in Supplementary Material (29, 30)], 50% displayed an altered profile after M4 exposure, suggesting the outbreak of genomic instability (Figure S3C in Supplementary Material). Taken together, these results supported the GO enrichment prediction that M4 may stimulate cell cycle despite non-reliable DNA replication initiation, thus altering genomic stability.

### Focus on Apoptosis Escape

MCF-10A cells were exposed to DMSO or M4 for 24 h and then treated for 6 h by either vehicle or staurosporine (STS). M4 alone triggered Bad (ser136) and Caspase 9 (ser196) enhanced phosphorylation (*p* < 0.05; **Figure 2A**), suggesting that those proteins became protected against pro-apoptotic intracellular signals and NF- $\kappa$ B nuclear translocation (Figure S5A in Supplementary Material). Co-treatment with M4 and STS led

**TABLE 1** | List of genes associated with the KEGG pathway “Cell cycle” and corresponding fold expression variation after 24 h M4 exposure as measured in transcriptomic analyses.

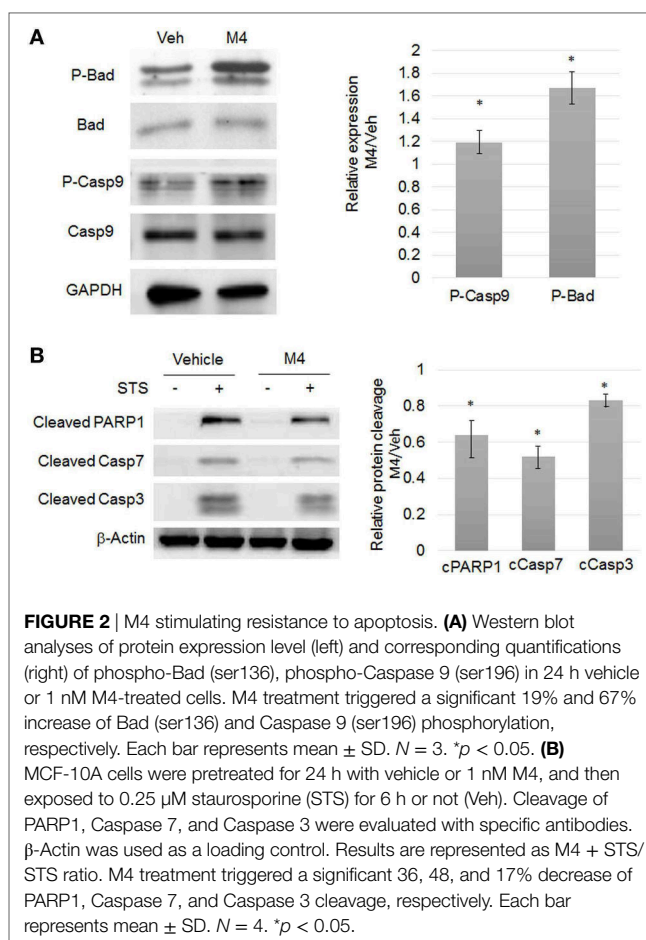
Gene symbol	Fold expression variation 24 h-M4/Veh
BUB1	10.85
BUB1B	16.43
CCNB1	11.11 <sup>a</sup>
CCNB2	2.33
CCND1	4.61 <sup>a</sup>
CDC20	5.44
CDC25A	6.85
CDC25C	5.47
CDC6	24.60
CDC7	5.20
CDK2	2.33
CDK4	2.05
CDKN2C	4.85
E2F2	2.98
MCM2	6.85 <sup>a</sup>
MCM3	8.77
MCM4	5.91
MCM5	5.19 <sup>a</sup>
MCM6	2.11
MCM8	2.51
MCM10	9.09 <sup>a</sup>
PCNA	3.58
PRKDC	2.12
RBL1	2.96
SKP1	2.18
TGFB2	2.18
TTK	17.24

<sup>a</sup>Fold expression variation that was validated by real-time PCR analyses (not shown). KEGG, Kyoto Encyclopedia Genes and Genomes.

to  $17 \pm 2\%$  reduction Caspase 3,  $48 \pm 6\%$  reduction of Caspase 7, and  $36 \pm 5\%$  of PARP1 cleavage compared to STS alone ( $p < 0.05$ ; **Figure 2B**). This apoptosis resistance triggered by M4 was confirmed by the observation of cytochrome c release and nuclear fragmentation measurement which was reduced by  $55 \pm 3\%$  ( $p < 0.05$ ; data not shown).

## Focus on EMT

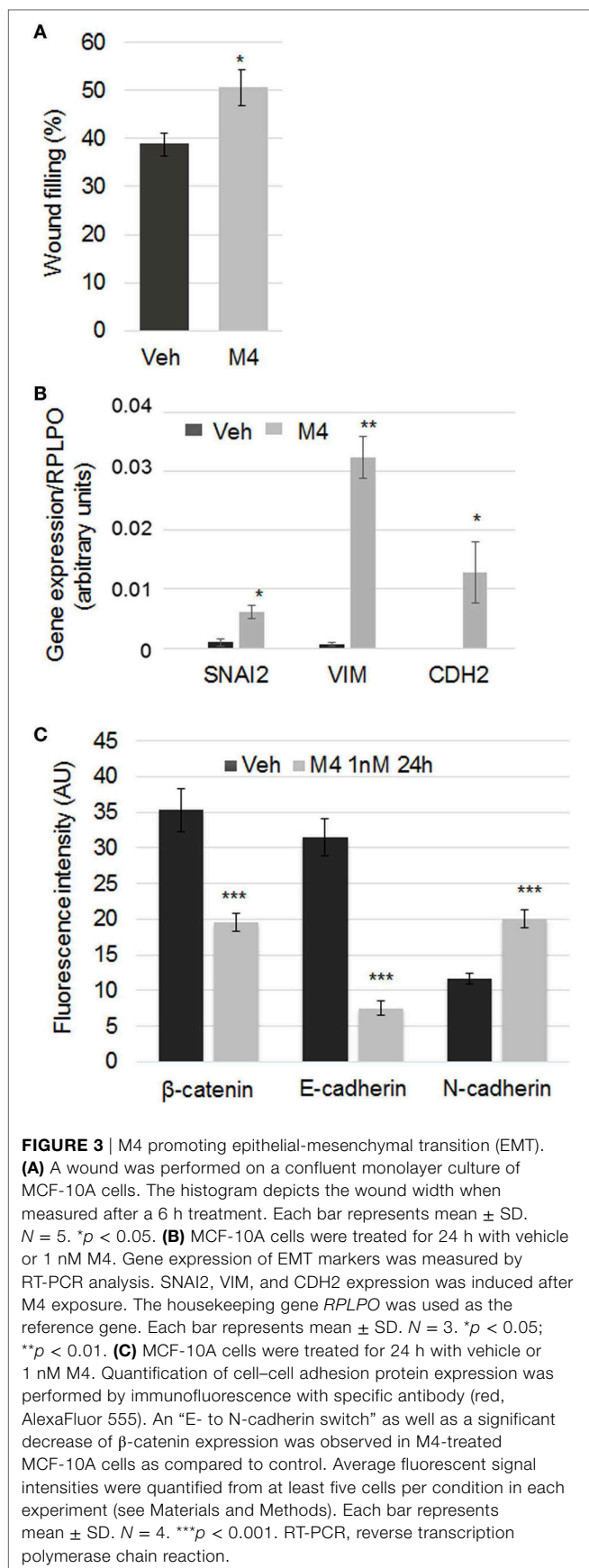
Using the scratch-wound assay, we observed that M4 exposure of MCF-10A cells triggered an accelerated closure of the wound created in a confluent monolayer culture ( $p < 0.05$ ; **Figure 3A**; Figure S4 in Supplementary Material). Therefore, expression of epithelial-mesenchymal transition (EMT) markers was analyzed by real-time PCR. SNAI2 ( $p < 0.05$ ), VIM ( $p < 0.01$ ), and CDH2 ( $p < 0.05$ ) expression were induced after M4 exposure (**Figure 3B**). Expression and localization of proteins known to be involved in cell-cell junctions were also examined: a list of genes associated with the KEGG pathway “Biological adhesion” and corresponding fold expression variation after 24 h M4 exposure was extracted from the transcriptomic data (**Table 2**). The expression of those 65 genes was repressed after M4 treatment. M4 treatment also triggered the decrease of beta-catenin (43%;  $p < 0.001$ ) and E-cadherin (78%;  $p < 0.001$ ) membrane immunofluorescence staining, whereas N-cadherin staining was augmented by 66% (**Figure 3C**; Figure S5B in Supplementary Material).



**FIGURE 2** | M4 stimulating resistance to apoptosis. **(A)** Western blot analyses of protein expression level (left) and corresponding quantifications (right) of phospho-Bad (ser136), phospho-Caspase 9 (ser196) in 24 h vehicle or 1 nM M4-treated cells. M4 treatment triggered a significant 19% and 67% increase of Bad (ser136) and Caspase 9 (ser196) phosphorylation, respectively. Each bar represents mean  $\pm$  SD.  $N = 3$ . \* $p < 0.05$ . **(B)** MCF-10A cells were pretreated for 24 h with vehicle or 1 nM M4, and then exposed to 0.25  $\mu$ M staurosporine (STS) for 6 h or not (Veh). Cleavage of PARP1, Caspase 7, and Caspase 3 were evaluated with specific antibodies.  $\beta$ -Actin was used as a loading control. Results are represented as M4 + STS/STS ratio. M4 treatment triggered a significant 36, 48, and 17% decrease of PARP1, Caspase 7, and Caspase 3 cleavage, respectively. Each bar represents mean  $\pm$  SD.  $N = 4$ . \* $p < 0.05$ .

## Involvement of ER $\alpha$ 36 in M4-Dependent Pathways in MCF-10A Cells

Microarray analysis indicated that none of the potential estrogen receptors (ER $\alpha$ , ER $\beta$ , GPER) expression was affected by M4 treatment. Since alkylphenol are described as estrogen mimicking compounds, we focused on the variant ER $\alpha$ 36 whose expression was stimulated 8-fold by 24 h M4 exposure (**Figure 4A**). The potential additive effects of ER $\alpha$ 36 overexpression and M4 treatment were assessed in the MCF-10A/ER $\alpha$ 36 cell line which stably overexpressed ER $\alpha$ 36 compared to MCF-10A/Zeo, both exposed to vehicle or M4. Quantification of cell viability by crystal violet staining indicated a marked reduction in MCF-10A/ER $\alpha$ 36 cell division rate compared to MCF-10A/Zeo (16). Nevertheless, MCF-10A/ER $\alpha$ 36 proliferation remained inducible by M4 exposure in the same range (18% versus 23%) as MCF-10A/Zeo ones ( $p < 0.01$ ; **Figure 4B**). MCF-10A/ER $\alpha$ 36 and MCF-10A/Zeo were exposed for 24 h to vehicle or M4 and then treated with STS as described above. The quantifications of Caspase 3, Caspase 7, and PARP1 cleavage in both cell lines indicated that ER $\alpha$ 36 overexpression led to a significant 34% reduction of PARP1 cleavage ( $p < 0.05$ ), whereas a non-significant trend toward cleavage moderation was also observed for Caspase 7 and Caspase 3 (**Figure 4C**). A TUNEL assay also indicated that ER $\alpha$ 36 overexpression was sufficient to reduce the number of



**TABLE 2** | List of genes associated with the GO term "Biological adhesion" and corresponding fold expression variation after 24 h M4 exposure as measured in transcriptomic analyses.

Gene symbol	Fold expression variation 24 h–M4/Veh	Gene symbol	Fold expression variation 24 h–M4/Veh
ADAM17	0.44	IL7R	0.17
ANTXR1	0.36	ITGA10 <sup>a,b,c</sup>	0.45
ANXA1	0.45	ITGA5 <sup>a,b,c</sup>	0.49
ARHGEF7 <sup>b</sup>	0.45	ITGAM <sup>b</sup>	0.28
ATP2A2	0.39	ITGB4 <sup>a,b,c</sup>	0.36
AZI2	0.41	ITGB8 <sup>a,b,c</sup>	0.46
B4GALT1	0.41	KIRREL	0.41
BCL2L11	0.31	LAMA3	0.38
CD164	0.50	LAMA5	0.41
CD36 <sup>a</sup>	0.09	LAMC1 <sup>a,c</sup>	0.37
CD44 <sup>a</sup>	0.41	LEPR	0.23
CD47 <sup>a</sup>	0.39	LMLN	0.40
CD99	0.42	LOXL2	0.16
CFDP1	0.46	LY6D	0.32
CLCA2	0.41	MICALL2	0.32
CLEC7A	0.17	MPZL2	0.32
COL12A1	0.48	MPZL3	0.33
COL7A1	0.29	NT5E	0.38
CTNNA1	0.47	OLR1	0.47
CTTN	0.31	PCDH7	0.48
CYR61	0.42	PCDH9	0.33
DLG1	0.47	PCDHGA4	0.31
DST	0.21	PDZD2	0.28
EGFR <sup>b,c</sup>	0.25	PKN2	0.40
ENG	0.48	PRKCE	0.40
FAT2	0.43	PTPRK	0.37
FN1 <sup>a,b,c</sup>	0.26	SCARB1	0.35
GNAS	0.37	SLC7A11	0.19
GPNMB	0.14	SRGAP2	0.30
HAS3	0.36	SRPX2	0.31
HPSE	0.29	WISP2	0.25
IGFBP7	0.40	WNT4	0.32

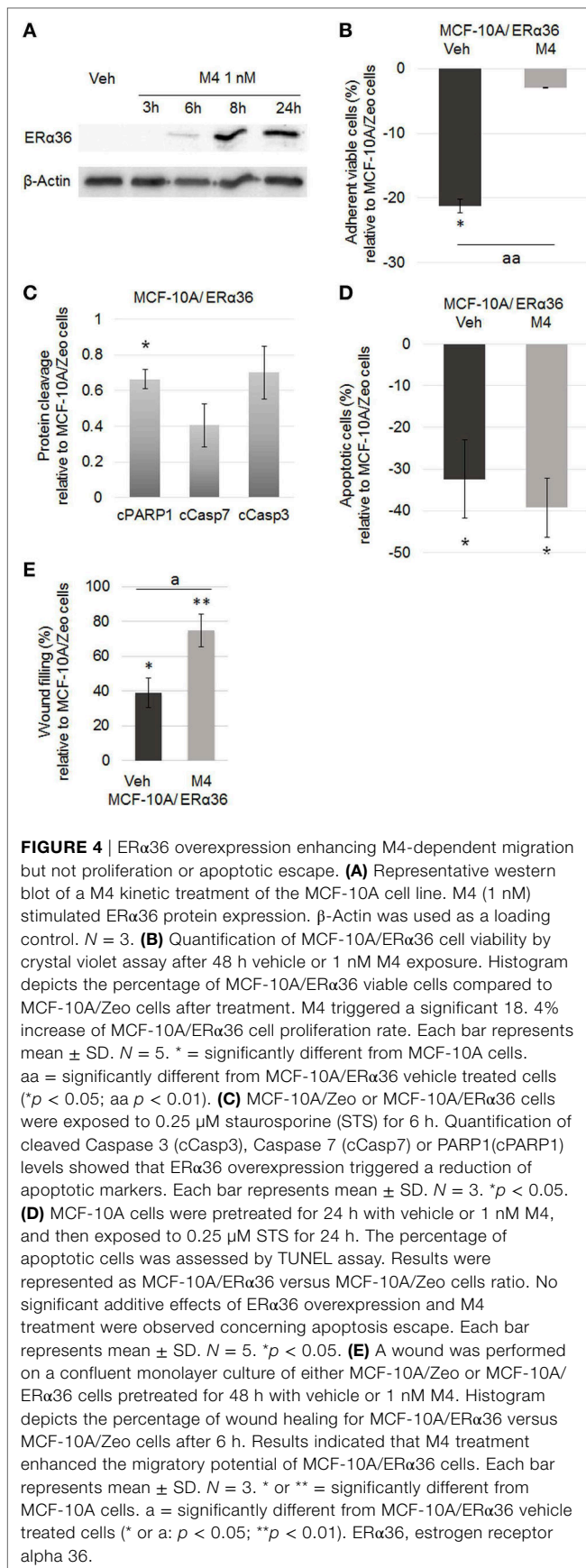
Genes that were also associated with KEGG pathways "ECM-receptor interaction" <sup>a</sup> and/or "Regulation of actin cytoskeleton" <sup>b</sup> and/or "Focal adhesion" <sup>c</sup> are indicated. GO, gene ontology; KEGG, Kyoto Encyclopedia Genes and Genomes.

MCF-10A/ER $\alpha$ 36 apoptotic cells compared to MCF-10A/Zeo ones after STS exposure ( $p < 0.05$ ; **Figure 4D**). Nevertheless, M4 pre-treatment of STS exposed MCF-10A/ER $\alpha$ 36 cells did not significantly potentialize resistance to apoptosis. Finally, scratch-wound assays performed on MCF-10A/Zeo or MCF-10A/ER $\alpha$ 36 indicated that ER $\alpha$ 36 overexpression enhanced M4-stimulated migration potential of mammary epithelial cells ( $p < 0.01$ ; **Figure 4E**; Figure S4 in Supplementary Material).

## In Vivo Assessment of Developmental and Transgenerational Impact of Alkylphenol Exposure

We first addressed the effects of maternal exposure to dietary relevant levels of long-chain alkylphenol M4 mix on fetal mammary gland development (see Materials and Methods). Whole mount analyses of mammary gland from F1 litters sacrificed at weaning (PND21) showed no significant mammary tree total extension or duct branching number in M4 maternally exposed F1 pups when compared to vehicle exposed ones. However, premature





lumen opening ( $p < 0.01$ ) and epithelium ( $p < 0.05$ ) thickening ( $p < 0.05$ ) were observed (Figure S6 in Supplementary Material).

The second part of the *in vivo* study used a transgenerational model of M4 paternally inherited exposure (Figure S1 in Supplementary Material). F3 mammary gland were harvested at weaning and submitted to histological analysis. A non-monotonic dose response to M4 exposure was observed in terms of epithelium and stroma thickening. The maximal impact was measured for the dose 2 (0.5  $\mu$ g/kg bw/day) which augmented epithelium and stroma thickness by 38% ( $p < 0.01$ ) and 60% ( $p < 0.05$ ), respectively (Figure 5A). Moreover, more than 80% of F3D2 mammary ducts displayed non-cohesive epithelium, cellular material into the lumen or cribriform patterns (Figure 5B). At the molecular level, we observed a loss of E-cadherin along with the induction of N-cadherin expression (Figure 5C). Finally, mammary tree whole mount computational analyses indicated a trend toward an increase of mammary tree total extension (F3D0:  $4,515 \pm 108 \mu$ m, F3D2:  $5,272 \pm 83 \mu$ m;  $p = 0.073$ ;  $N = 9$ ; Amand Chesnel, unpublished data). A significant 21% reduced number of end buds ( $p < 0.01$ ), whereas no branching point number variation in F3D2 pups were observed (Figure 5D).

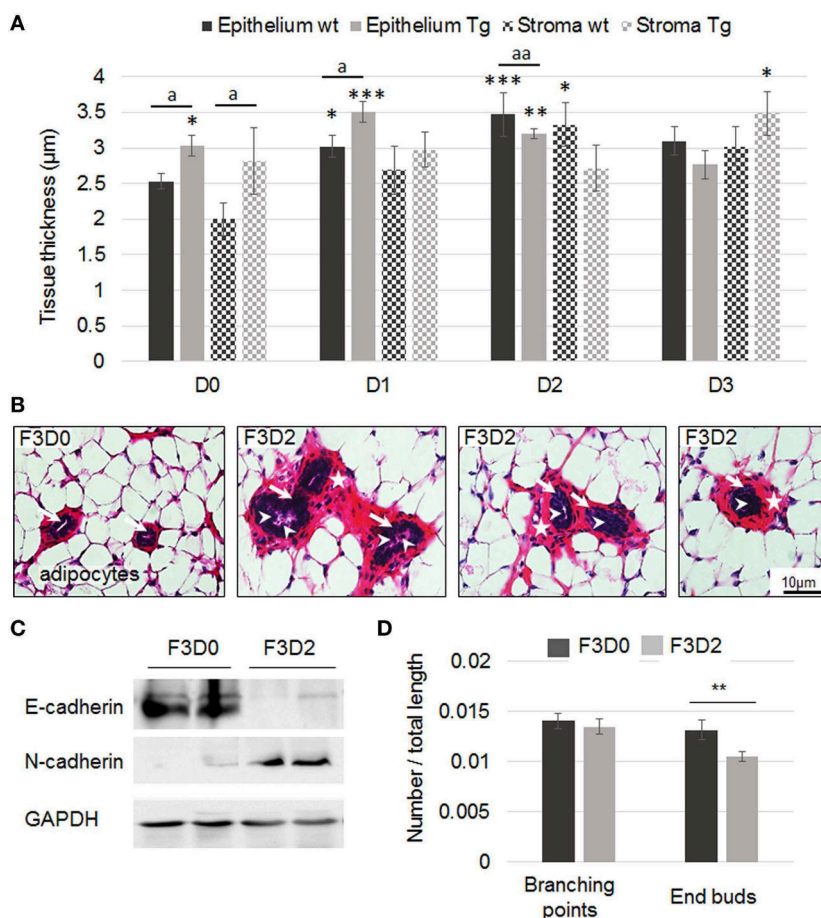
## M4 Exposure of Transgenic Mice Expressing a Human ER $\alpha$ 36 Transgene in the Mammary Gland

Then, we addressed the combined effects of M4 exposure and human ER $\alpha$ 36 expression in the mammary gland of mice transgenerationally exposed to M4, as described above for the wild-type mice. Epithelium thickness was affected by both the transgene expression and the treatment as a non-monotonic dose-response relationship: compared to wt animals, the transgene expression seemed to enhance the treatment consequences for D1 ( $p < 0.05$ ) but conversely to lower this effect for D2 ( $p < 0.01$ ) and D3 (Figure 5A). The stroma displayed a similar but statistically non-significant profile (Figure 5A).

At adulthood, we observed a non-significant trend to epithelium thickening in F3 control, D1- or D2-treated Tg mice but a particular phenotype resembling the one depicted at weaning for D3-treated animals (Figure S7 in Supplementary Material). No combined effect of human ER $\alpha$ 36 expression and M4 exposure was detected for either the epithelium or the stroma.

## ER $\alpha$ 36 Overexpression in M4 Exposed Xenografted Nude Mice

In order to address the ability of M4 treatment combined to ER $\alpha$ 36 overexpression to trigger carcinogenesis of the adult mammary gland, we orthotopically grafted the MCF-10A cell line, either expressing or not the ER $\alpha$ 36 receptor. No tumor was observed in MCF-10A/Zeo or MCF-10A/ER $\alpha$ 36 grafted mice whatever the treatment. Conversely, 80% of MDA-MB-231 grafted mice developed a tumor 5 days after cell injection. Once gavage was stopped, tumor volume was slightly increased in MDA-MB-231 grafted mice previously exposed to M4 compared to vehicle treated ones. At end point, tumor volume and histology were similar in M4 or vehicle-exposed animals (Figure S8 in Supplementary Material).



**FIGURE 5** | Mammary gland histology of F3 wild-type or transgenic (Tg) mice exposed to M4. **(A)** Quantification of epithelium and stroma thickness in F3 mammary gland following exposure to sesame oil (D0) or M4 [0.05 µg (D1), 0.5 µg (D2), or 5 µg M4/kg/day (D3)]. Measurements were performed on PND21. Each bar represents mean  $\pm$  SEM. Epithelium or stroma thickness was measured on at least five independent slices from each animal included in the experiment. N: number of animals/dose/genotype. D0wt:  $N = 7$ ; D0Tg:  $N = 5$ ; D1wt:  $N = 6$ ; D1Tg:  $N = 4$ ; D2wt:  $N = 4$ ; D2Tg:  $N = 8$ ; D3wt:  $N = 3$ ; D3Tg:  $N = 4$ . \*, \*\*, or \*\*\* = significantly different from D0 wild-type (wt) mice. a or aa = Tg mice significantly different from wild-type littermates exposed to the same dose (\* or  $a p < 0.05$ ; \*\*; aa  $p < 0.01$ ; \*\*\* $p < 0.001$ ). **(B)** F3 mammary gland exposed to M4 (D2 = 0.5 µg/kg/day). The slices show mammary ducts surrounded by stroma (stars) and adipocytes. Disorganization of the epithelium (arrows) was observed in D2 exposed F3 mammary glands and cellular material (arrowheads) was present into the lumen of some F3D2 female ducts. Scale bar, 10 µm. **(C)** Representative western blot analyses of E-cadherin, N-cadherin, and GAPDH (loading control) protein expression levels in total mammary gland harvested from F3D0 or F3D2 pups. **(D)** Mammary trees from F3D0 ( $N = 4$ ) or F3D2 ( $N = 9$ ) exposed mice were whole mounted. Computational analysis of the mammary tree total extension, number of branching or end buds was performed using dedicated software. Results are represented as a ratio of number of branching or end buds/mammary tree total length. The number of end buds was significantly reduced in F3D2 pups. Each bar represents mean  $\pm$  SEM. \*\* $p < 0.01$ . GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

## DISCUSSION

The data presented herein support for the first time a deleterious impact of exposure to low doses of the M4 alkylphenol realistic mixture through P0 maternal exposure followed by a second-generation male transmission. Instead, most studies available depict the direct consequences of several milligrams of either 4NP or 4tOP, separately. The study by Moon et al. (10) indicated that Long Evans rats exposed to 100 mg/kg during gestation days 15–19 displayed advanced lobular development of their mammary gland on PND21, while the glands from the “low” dose (10 mg/kg) were unaffected. The conversion of immature to mature structures was also reported in Noble rats in response to 7.1 mg/24 h or 0.01 mg/24 h nonylphenol exposure (31). In our

study, PND21 F1 pups, fully exposed to a maximal dose of 5 µg/kg M4 through gestation exhibited mammary gland architecture alterations but no precocious differentiation or extension of the mammary tree. Nevertheless, Colerangle and Roy (31) showed a significant stimulation of cell cycle and proliferation of mammary epithelial cells (24 and 38% for the low and high exposure, respectively) which correlates to epithelium thickening observed in our F1 mammary gland samples and the 23% accelerated proliferation combined to cyclin up-regulation observed in MCF-10A cells.

While testing the combined effects of M4 exposure and ER $\alpha$ 36 expression on tumor initiation, we performed a positive control group in which the Nude mice, xenografted with ER $\alpha$ 36 expressing MDA-MB-231 breast tumor cells, were gavigated with

sesame oil or M4 (dose 2). Mice developed a tumor without any difference between vehicle and M4 treated animals in terms of tumor latency, multiplicity of histology. Nevertheless, we observed a trend to tumor growth rate enhancement in mice exposed to alkylphenols, which suggest that such an exposure could be considered as a tumor promoting environment (= boost tumor growth). This model is quite different from MMTV-erbB2 Tg mice used by Jenkins et al (32) or the carcinogen DMBA (dimethylbenz[a]anthracene) induced tumorigenesis described by Betancourt et al. (33) in which both teams demonstrated a pro-tumoral effect of BPA exposure. Opposite results were obtained by Peng et al. (34), who reported a protective effect of a 40 mg/kg octylphenol exposure in DMBA-treated Sprague–Dawley rats. Nevertheless, these data suggest that a transient direct exposure to alkylphenols at adulthood is not sufficient by itself to trigger mammary carcinogenesis. Indeed, Sprague et al. (35) found no relationship between breast density of postmenopausal women and blood concentration of either nonyl- or octylphenol.

Since lifetime exposure to alkylphenols is the most common feature for human populations in western countries, we addressed the transgenerational male transmission of mammary gland abnormalities, especially for a dose of 0.5 µg/kg/day, representative of human exposure (12). F3 pups from P0-treated dams, which were not directly exposed to the M4 mixture, displayed paternally inherited elevated lumen diameter, epithelium, and stroma thickness as well as end bud number when compared to F3 originating from untreated dams. The main phenotype was the presence of numerous stoppered mammary ducts which display either cellular material scattered into the lumen or several septa and epithelium misorganization. These phenotypes could be interpreted as a partial disruption of stroma-epithelium interactions (36) as confirmed by severe reduction of E-cadherin and stimulation of N-cadherin expression in F3 pups originating from animals exposed to M4 *in utero*. Such nonylphenol-dependent disruption of cell–cell adherence and epithelial structure was also observed in rat seminiferous tubules after high-dose exposures (37). The mixed phenotypes observed in mammary glands from this study resemble ductal atypia rather than intraductal hyperplasia described by Vandenberg et al. (38) in CD-1 mice exposed to BPA. In a review, Paulose et al. (36) also listed systemic, tissular, cellular, and molecular alterations generated in adult mice by BPA perinatal exposure. Namely, the authors describe a partial and genome-wide modification of DNA methylation (36). Even if we did not observe any histological alteration in adult mice, we detected *in vitro* significant M4-dependent changes in ERα36 promoter methylation status (data not shown), suspected to be linked to the protein expression level. Taken together, our data suggest that, like BPA, M4 exposure could have multiscale consequences on mammary gland development. Therefore, it would be of great interest to follow aging exposed mice compared to control ones to address the hypothesis of any alkylphenol-dependent fetal programming of adult disease.

To address the molecular pathways that could be involved in alkylphenol-dependent mammary epithelial cell alteration in human, we performed a comprehensive analysis of gene expression under alkylphenol *in vitro* exposure of the human

MCF-10A cell line. GO and KEGG pathway enrichment analysis confirmed by *in vitro* experiments indicated that the main functions affected by the treatment—i.e., proliferation, survival, and migration—could ultimately result in cell transformation. Indeed, perturbation of cell cycle and enhanced proliferation concomitant with genetic instability are considered as a risk factor for the development of cancer. Nonylphenol-dependent stimulation of clastogenic and mutagenic mechanisms as well as enhanced cell cycle and proliferation has been reported many times in different *in vitro* and *in vivo* models (31, 39–41).

Our data reporting an enhanced survival to STS-induced apoptosis in MCF-10A are in line with inhibition of induced cell death by BPA or nonylphenol reported in rat PC12, hippocampal, and neuronal cells (42, 43). Another consequence of alkylphenol exposure was the stimulation of migratory potential as supported by enhanced wound filling and EMT marker expression. These *in vitro* data point in the same direction as mammary epithelium morphological changes observed *in vivo*. Similar morphological and transcriptomic results were obtained in low-dose BPA-exposed mice which show ductal alterations correlated to induction of anti-apoptotic and reduction of focal adhesion-related genes (44).

Alkylphenols are known to bind estrogen receptors with a relative binding affinity to 17β-estradiol in the range of 0.1% for ERα and 50% for GPER (5, 45). In order to understand whether the molecular and global phenotypes observed after alkylphenol exposure were due to their estrogenicity and could be extrapolated to human, we turned back to the microarray analysis of human normal epithelial MCF-10A cells but observed that none of the probe covering known nuclear or membrane estrogen receptor genes displayed differential expression at any time. Since KEGG pathway enrichment pointed to JAK-STAT and MAPK signaling pathways, we assumed that the ERα36 variant could be involved in M4 response. Indeed, ERα36 was previously demonstrated to trigger (i) STAT3 signaling in mammary epithelial cells, (ii) MAPK signaling in breast cancer cells, and (iii) alkylphenol-dependent response in seminoma cells (13, 16, 46, 47). Recently, we also demonstrated that ERα36 overexpression in MCF-10A cells or mammary gland is sufficient, by itself, to alter normal breast epithelial phenotype *in vitro* and *in vivo* (16). Therefore, we showed that ERα36 expression was induced by M4 exposure in mammary epithelial cells and addressed the consequences of an alkylphenol treatment in epithelial cells displaying a forced ERα36 expression. An additive effect of both conditions was only observed for the enhanced migration phenotype of MCF-10A cells. This suggests that alkylphenols modulate (i) proliferation and survival in an ERα36 dose-independent manner and (ii) migration in an ERα36 dose-dependent way, probably through a combination of cofactor association and transduction pathway activation. Therefore, ERα36 expression level could be a promising marker of breast epithelium leakage in the context of endocrine disruptor exposure. Similar results were obtained from breast cancer patient samples, in which we and others demonstrated that ERα36 expression level is a relevant predictor of metastatic progression (17, 46, 48) but to our knowledge, this is the first description of a functional role of ERα36 in normal epithelial mammary cells exposed to xenoestrogens. Therefore,

ER $\alpha$ 36-dependent modifications of cell phenotype that could be related to cell transformation or cancer progression phenotype seem to be a general feature of mammary cell exposed to estrogens or xenoestrogens.

## ETHICS STATEMENT

All experimental procedures were approved by the French Minister of Research Committee for animal experiment in accordance with the Guide for Care and Use of Laboratory Animals. The transgenerational model of C57BL/6J exposed to alkylphenols was carried out in accordance with the recommendations of French Minister of Research Committee for animal experiment (Protocol no. APAFIS#2168-2015110518268051 v5). Cell grafting in Nude mice exposed to alkylphenols was carried out in accordance with the recommendations of French Minister of Research Committee for animal experiment (Protocol no. APAFIS#6158-2016072016446146 v5).

## AUTHOR CONTRIBUTIONS

Conceptualization: HD, AC, and TB. Data curation: EB and TB. Formal analysis: AC, HD, M-DD, and TB. Funding acquisition: HD and TB. Investigation: CT, CC-J, and CM. Methodology: CT, CC-J, AC, and MS-T. Project administration: AC, TB, and HD.

## REFERENCES

- Soto AM, Sonnenschein C. Environmental causes of cancer: endocrine disruptors as carcinogens. *Nat Rev Endocrinol* (2010) 6(7):363–70. doi:10.1038/nrendo.2010.87
- Trichopoulos D. Is breast cancer initiated in utero? *Epidemiology* (1990) 1(2):95–6. doi:10.1097/00001648-199003000-00001
- Murray TJ, Maffini MV, Ucci AA, Sonnenschein C, Soto AM. Induction of mammary gland ductal hyperplasias and carcinoma in situ following fetal BPA exposure. *Reprod Toxicol* (2007) 23(3):383–90. doi:10.1016/j.reprotox.2006.10.002
- Yang O, Kim HL, Weon JI, Seo YR. Endocrine-disrupting chemicals: review of toxicological mechanisms using molecular pathway analysis. *J Cancer Prev* (2015) 20(1):12–24. doi:10.15430/JCP.2015.20.1.12
- Thomas P, Dong J. Binding and activation of the seven-transmembrane estrogen receptor GPR30 by environmental estrogens: a potential novel mechanism of endocrine disruption. *J Steroid Biochem Mol Biol* (2006) 102(1–5):175–9. doi:10.1016/j.jsbmb.2006.09.017
- WHO/UNEP/ILO. *Integrated Risk Assessment: Nonylphenol Case Study. International Program on Chemical Safety 2004*. Geneva: International Programme on Chemical Safety, WHO (2004).
- Chapin RE, Delaney J, Wang Y, Lanning L, Davis B, Collins B, et al. The effects of 4-nonylphenol in rats: a multigeneration reproduction study. *Toxicol Sci* (1999) 52(1):80–91. doi:10.1093/toxsci/52.1.80
- Tyl RW, Myers CB, Marr MC, Castillo NP, Seely JC, Sloan CS, et al. Three-generation evaluation of dietary para-nonylphenol in CD-(Sprague-Dawley) rats. *Toxicol Sci* (2006) 92(1):295–310. doi:10.1093/toxsci/kfj203
- Acevedo R, Parnell PG, Villanueva H, Chapman LM, Gimenez T, Gray SL, et al. The contribution of hepatic steroid metabolism to serum estradiol and estrion concentrations in nonylphenol-treated MMTV<sup>neu</sup> mice and its potential effects on breast cancer incidence and latency. *J Appl Toxicol* (2005) 25(5):339–53. doi:10.1002/jat.1078
- Moon HJ, Han SY, Shin JH, Kang IH, Kim TS, Hong JH, et al. Gestational exposure to nonylphenol causes precocious mammary gland development in female rat offspring. *J Reprod Dev* (2007) 53(2):333–44. doi:10.1262/jrd.18055

Resources: AC, CC-J, and HD. Software: EB, MS-T, M-DD, and TB. Supervision: HD, TB, and AC. Validation: AC, HD, and TB.

## ACKNOWLEDGMENTS

We thank greatly Dominique Meng for histology and Nude mice rearing, Maelle Chesnel for technical assistance, and Christine Capdeville-Atkinson and Patrick Liminana for their help in using the animal facility of the Faculté de Pharmacie de l'Université de Lorraine.

## FUNDING

This work was supported by the “Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail” (ANSES; grant number PNR EST 2012-2-014), “INSERM Cancer-Environnement” (grant number ENV201304), and “La Ligue contre le Cancer (2016).” CCJ was the recipient of a “Région Lorraine” PhD fellowship.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fendo.2017.00272/full#supplementary-material>.

- Soto AM, Justicia H, Wray JW, Sonnenschein C. p-Nonyl-phenol: an estrogenic xenobiotic released from “modified” polystyrene. *Environ Health Perspect* (1991) 92:167–73. doi:10.1289/ehp.9192167
- Raecker T, Thiele B, Boehme RM, Guenther K. Endocrine disrupting nonyl- and octylphenol in infant food in Germany: considerable daily intake of nonylphenol for babies. *Chemosphere* (2011) 82(11):1533–40. doi:10.1016/j.chemosphere.2010.11.065
- Ajj H, Chesnel A, Pinel S, Plenat F, Flament S, Dumond H. An alkylphenol mix promotes seminoma derived cell proliferation through an ER $\alpha$ 36-mediated mechanism. *PLoS One* (2013) 8(4):e61758. doi:10.1371/journal.pone.0061758
- Manikkam M, Guerrero-Bosagna C, Tracey R, Haque MM, Skinner MK. Transgenerational actions of environmental compounds on reproductive disease and identification of epigenetic biomarkers of ancestral exposures. *PLoS One* (2012) 7(2):e31901. doi:10.1371/journal.pone.0031901
- Wang Z, Zhang X, Shen P, Loggie BW, Chang Y, Deuel TF. Identification, cloning, and expression of human ER $\alpha$ 36, a novel variant of human ER $\alpha$ 66. *Biochem Biophys Res Commun* (2005) 336(4):1023–7. doi:10.1016/j.bbrc.2005.08.226
- Thiebaut C, Chamard-Jovenin C, Chesnel A, Morel C, Djermoune EH, Boukhobza T, et al. Mammary epithelial cell phenotype disruption in vitro and in vivo through ER $\alpha$ 36 overexpression. *PLoS One* (2017) 12(3):e0173931. doi:10.1371/journal.pone.0173931
- Gayraud V, Lacroix MZ, Collet SH, Viguie C, Bousquet-Melou A, Toutain PL, et al. High bioavailability of bisphenol A from sublingual exposure. *Environ Health Perspect* (2013) 121:951–6. doi:10.1289/ehp.1206339
- Odum J, Pyrah IT, Soames AR, Foster JR, Van Miller JB, Joiner RL, et al. Effects of p-nonylphenol (NP) and diethylstilboestrol (DES) on the Alderley Park (Alpk) rat: comparison of mammary gland and uterus sensitivity following oral gavage or implanted mini-pumps. *J Appl Toxicol* (1999) 19(5):367–78. doi:10.1002/(SICI)1099-1263(199909/10)19:5<367::AID-JAT593>3.0.CO;2-C
- Odum J, Pyrah IT, Foster JR, Van Miller JB, Joiner RL, Ashby J. Comparative activities of p-nonylphenol and diethylstilbestrol in noble rat mammary gland and uterotrophic assays. *Regul Toxicol Pharmacol* (1999) 29(2):184–95. doi:10.1006/rtph.1999.1286

20. Vandenberg LN, Maffini MV, Wadia PR, Sonnenschein C, Rubin BS, Soto AM. Exposure to environmentally relevant doses of the xenoestrogen bisphenol-A alters development of the fetal mouse mammary gland. *Endocrinology* (2007) 148(1):116–27. doi:10.1210/en.2006-0561
21. Tylcz JB, El Alaoui-Lasmali K, Djermoune EH, Thomas N, Faivre B, Bastogne T. Data-driven modeling and characterization of anti-angiogenic molecule effects on tumoral vascular density. *Biomed Signal Process Control* (2015) 20:52–60. doi:10.1016/j.bspc.2015.04.008
22. Frangi AF, Niessen WJ, Vincken KL, Vieregger MA. Multiscale vessel enhancement filtering. In: Wells WM, Colchester A, Delp SL, editors. *Proceedings from MICCAI'98, Cambridge, MA*. (Vol. 1496), Berlin, Germany: Computer Science Press (1998). p. 130–7.
23. Howroyd P, Hoyle-Thacker R, Lyght O, Williams D, Kleymenova E. Morphology of the fetal rat testis preserved in different fixatives. *Toxicol Pathol* (2005) 33(2):300–4. doi:10.1080/01926230590896145
24. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* (2016) 45(Database issue):D331–8. doi:10.1093/nar/gkw1138
25. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* (2003) 34(3):267–73. doi:10.1038/ng1180
26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* (2005) 102(43):15545–50. doi:10.1073/pnas.0506580102
27. Soule HD, Maloney TM, Wolman SR, Peterson WD, Brenz R, McGrath CM, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. *Cancer Res* (1990) 50(18):6075–86.
28. Altman DG, Bland JM. Standard deviations and standard errors. *BMJ* (2005) 331:903. doi:10.1136/bmj.331.7521.903
29. Ong TM, Song B, Qian HW, Wu ZL, Whong WZ. Detection of genomic instability in lung cancer tissues by random amplified polymorphic DNA analysis. *Carcinogenesis* (1998) 19(1):233–5. doi:10.1093/carcin/19.1.233
30. Papadopoulos S, Benter T, Anastassiou G, Pape M, Gerhard S, Bornfeld N, et al. Assessment of genomic instability in breast cancer and uveal melanoma by random amplified polymorphic DNA analysis. *Int J Cancer* (2002) 99(2):193–200. doi:10.1002/ijc.10297
31. Colerangle JB, Roy D. Exposure of environmental estrogenic compound nonylphenol to noble rats alters cell-cycle kinetics in the mammary gland. *Endocrine* (1996) 4(2):115–22. doi:10.1007/BF02782756
32. Jenkins S, Wang J, Eltoum I, Desmond R, Lamartiniere CA. Chronic oral exposure to bisphenol A results in a nonmonotonic dose response in mammary carcinogenesis and metastasis in MMTV-erbB2 mice. *Environ Health Perspect* (2011) 119:1604–9. doi:10.1289/ehp.1103850
33. Betancourt AM, Wang J, Jenkins S, Mobley J, Russo J, Lamartiniere CA. Altered carcinogenesis and proteome in mammary glands of rats after prepubertal exposures to the hormonally active chemicals bisphenol A and genistein. *J Nutr* (2012) 142(7):1382S–8S. doi:10.3945/jn.111.152058
34. Peng JH, Zhang F, Zhang HX, Fan HY. Prepubertal octylphenol exposure up-regulate BRCA1 expression, down-regulate ER  $\alpha$  expression and reduce rat mammary tumorigenesis. *Cancer Epidemiol* (2009) 33(1):51–5. doi:10.1016/j.canep.2009.04.017
35. Sprague BL, Trentham-Dietz A, Hedman CJ, Wang J, Hemming JD, Hampton JM, et al. Circulating serum xenoestrogens and mammographic breast density. *Breast Cancer Res* (2013) 15(3):R45. doi:10.1186/bcr3432
36. Paulose T, Speroni L, Sonnenschein C, Soto AM. Estrogens in the wrong place at the wrong time: fetal BPA exposure and mammary cancer. *Reprod Toxicol* (2015) 54:58–65. doi:10.1016/j.reprotox.2014.09.012
37. Duan P, Hu C, Butler HJ, Quan C, Chen W, Huang W, et al. 4-Nonylphenol induces disruption of spermatogenesis associated with oxidative stress-related apoptosis by targeting p53-Bcl-2/Bax-Fas/FasL signaling. *Environ Toxicol* (2016) 32(3):739–53. doi:10.1002/tox.22274
38. Vandenberg LN, Maffini MV, Schaeberle CM, Ucci AA, Sonnenschein C, Rubin BS, et al. Perinatal exposure to the xenoestrogen bisphenol-A induces mammary intraductal hyperplasias in adult CD-1 mice. *Reprod Toxicol* (2008) 26(3–4):210–9. doi:10.1016/j.reprotox.2008.09.015
39. Dobrzyńska MM. Male-mediated F1 effects in mice exposed to nonylphenol or to a combination of X-rays and nonylphenol. *Drug Chem Toxicol* (2012) 35(1):36–42. doi:10.3109/01480545.2011.586036
40. Frassinetti S, Barberio C, Caltavuturo L, Fava F, Di Gioia D. Genotoxicity of 4-nonylphenol and nonylphenol ethoxylate mixtures by the use of *Saccharomyces cerevisiae* D7 mutation assay and use of this text to evaluate the efficiency of biodegradation treatments. *Ecotoxicol Environ Saf* (2011) 74(3):253–8. doi:10.1016/j.ecoenv.2010.10.039
41. Forte M, Di Lorenzo M, Carrizzo A, Valiante S, Vecchione C, Laforgia V, et al. Nonylphenol effects on human prostate non tumorigenic cells. *Toxicology* (2016) 357–358:21–32. doi:10.1016/j.tox.2016.05.024
42. Negishi T, Ishii Y, Kyuwa S, Kuroda Y, Yoshikawa Y. Inhibition of staurosporine-induced neuronal cell death by BPA and nonylphenol in primary cultured rat hippocampal and cortical neurons. *Neurosci Lett* (2003) 353(2):99–102. doi:10.1016/j.neulet.2003.09.031
43. Liu C, Sun Y, Song Y, Saito T, Kurasaki M. Nonylphenol diethoxylate inhibits apoptosis induced in PC12 cells. *Environ Toxicol* (2016) 31(11):1389–98. doi:10.1002/tox.22144
44. Wadia PR, Cabaton NJ, Borrero MD, Rubin BS, Sonnenschein C, Shioda T, et al. Low-dose BPA exposure alters the mesenchymal and epithelial transcriptomes of the mouse fetal mammary gland. *PLoS One* (2013) 8(5):e63902. doi:10.1371/journal.pone.0063902
45. Li Z, Zhang H, Gibson M, Li J. An evaluation on combination effects of phenolic endocrine disruptors by ER binding assay. *Toxicol In Vitro* (2012) 26(6):769–74. doi:10.1016/j.tiv.2012.05.017
46. Wang ZY, Yin L. ER $\alpha$ 36: a new player in human breast cancer. *Mol Cell Endocrinol* (2015) 418(3):193–206. doi:10.1016/j.mce.2015.04.017
47. Omarjee S, Jacquemetton J, Poulard C, Rochel N, Dejaegere A, Chebaro Y, et al. The molecular mechanisms underlying the ER $\alpha$ -36-mediated signaling in breast cancer. *Oncogene* (2017) 36(18):2503–14. doi:10.1038/onc.2016.415
48. Chamard-Jovenin C, Jung AC, Chesnel A, Abecassis J, Flament S, Ledrappier S, et al. From ER $\alpha$ 66 to ER $\alpha$ 36: a generic method for validating a prognosis marker of breast tumor progression. *BMC Syst Biol* (2015) 9:28. doi:10.1186/s12918-015-0178-7

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Chamard-Jovenin, Thiebaut, Chesnel, Bresso, Morel, Smail-Tabbone, Devignes, Boukhobza and Dumond. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A	GO biological process		GO biological process	
		Fold Enrichment t = 1 h		Fold Enrichment t = 1h
	response to stress	1,83	regulation of primary metabolic process	1,51
	cellular nitrogen compound metabolic process	1,63	regulation of cellular metabolic process	1,5
	organic cyclic compound metabolic process	1,57	primary metabolic process	1,37

B	GO biological process		GO biological process	
		Fold Enrichment t = 24 h		Fold Enrichment t = 24 h
	DNA strand elongation involved in DNA replication	8,09	response to ionizing radiation	2,54
	kinetochore organization	7,85	nucleocytoplasmic transport	2,51
	CENP-A containing nucleosome assembly	7,30	positive regulation of DNA metabolic process	2,44
	nuclear DNA replication	6,92	nucleic acid phosphodiester bond hydrolysis	2,43
	telomere maintenance via recombination	6,73	mRNA splicing, via spliceosome	2,28
	DNA replication initiation	6,68	negative regulation of cell cycle phase transition	2,23
	spindle checkpoint	5,88	negative regulation of organelle organization	2,08
	mitotic metaphase plate congression	5,74	cell proliferation	1,97
	DNA damage response, detection of DNA damage	5,56	regulation of cellular response to stress	1,91
	regulation of transcription involved in G1/S transition of mitotic cell cycle	5,46	regulation of cellular amide metabolic process	1,82
	sister chromatid cohesion	5,37	posttranscriptional regulation of gene expression	1,8
	DNA-dependent DNA replication maintenance of fidelity	4,94	regulation of cellular protein localization	1,76
	chromosome condensation	4,94	protein complex assembly	1,62
	regulation of mitotic metaphase/anaphase transition	4,81	cellular macromolecule catabolic process	1,57
	regulation of DNA-dependent DNA replication	4,76	organonitrogen compound catabolic process	1,53
	mitotic spindle assembly	4,71	cell death	1,51
	microtubule polymerization or depolymerization	4,41	small molecule metabolic process	1,44
	regulation of cholesterol biosynthetic process	4,20	regulation of cell proliferation	1,43
	DNA synthesis involved in DNA repair	3,77	response to oxygen-containing compound	1,42
	positive regulation of cell cycle phase transition	3,60	negative regulation of cellular biosynthetic process	1,41
	regulation of DNA recombination	3,53	regulation of multicellular organismal development	1,39
	cell division	3,52	negative regulation of macromolecule metabolic process	1,39
	DNA geometric change	3,32	negative regulation of nitrogen compound metabolic process	1,38
	double-strand break repair via homologous recombination	3,32	regulation of cellular protein metabolic process	1,36
	G2/M transition of mitotic cell cycle	2,98	multi-organism process	1,35
	positive regulation of mitotic cell cycle	2,92	regulation of catalytic activity	1,31
	mitotic DNA damage checkpoint	2,91	animal organ development	1,3
	meiotic nuclear division	2,76	G-protein coupled receptor signaling pathway	0,39
	regulation of cell division	2,74	sensory perception of smell	<0,2
	regulation of microtubule cytoskeleton organization	2,74	detection of chemical stimulus involved in sensory perception	<0,2
	regulation of signal transduction by p53 class mediator	2,63		

**Supplementary Table 1. Gene Ontology enrichment analysis of differentially expressed genes after a 1 h or 24 h 1nM M4 exposure.**

Biological process GO terms significantly (p-value <0.05) enriched after 1h (A) or 24h (B) alkylphenol exposure and respective Fold enrichment were obtained using the PANTHER database (Protein Analysis Through Evolutionary Relationships, <http://pantherdb.org>).

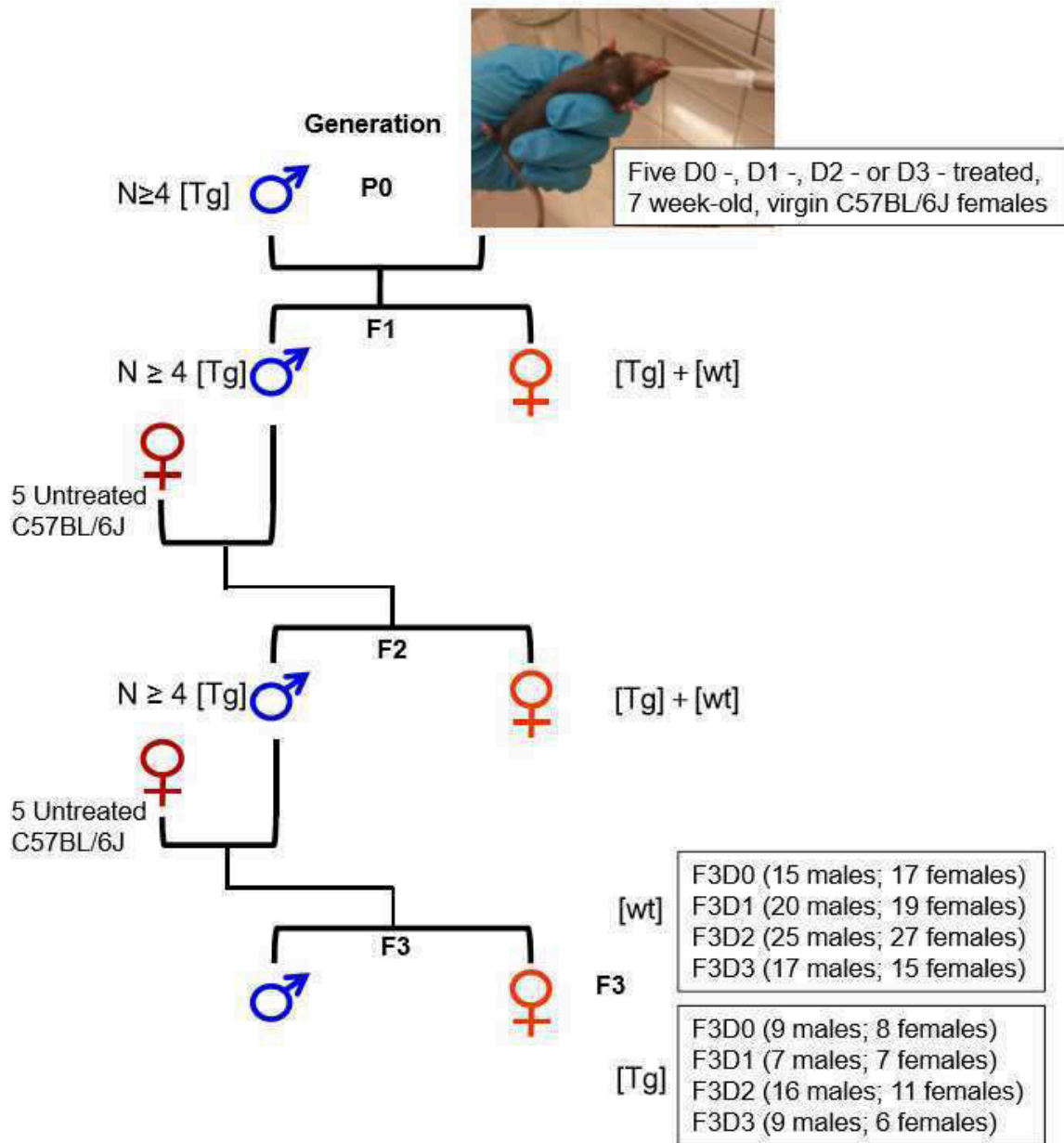
The overlap between the 76 DEGs obtained after 8 h M4 exposure and GO gene sets derived from the Biological Process didn't highlight GO terms significantly enriched at this point of the kinetic.

1 hour		8 hours		24 hours	
KEGG	Genes in overlap	KEGG	Genes in overlap	KEGG	Genes in overlap
KEGG_PATHWAYS_IN_CANCER	7	KEGG_PATHWAYS_IN_CANCER	4	KEGG_PATHWAYS_IN_CANCER	41
KEGG_FOCAL_ADHESION	6	KEGG_CELL_CYCLE	4	KEGG_CELL_CYCLE	37
KEGG_REGULATION_OF_ACTIN_CYTOSKELETON	6	KEGG_DNA_REPLICATION	3	KEGG_SPLICEOSOME	31
KEGG_MAPK_SIGNALING_PATHWAY	6	KEGG_P53_SIGNALING_PATHWAY	3	KEGG_PYRIMIDINE_METABOLISM	26
KEGG_JAK_STAT_SIGNALING_PATHWAY	5	KEGG_JAK_STAT_SIGNALING_PATHWAY	3	KEGG_DNA_REPLICATION	25

**Supplementary Table 2. Pathway enrichment analysis of differentially expressed genes after a 1 h, 8 h or 24 h 1nM M4 exposure.**

KEGG pathways significantly ( $p$ -value  $<0.05$ ) enriched after 1h, 8 h or 24h alkylphenol were obtained using the MSigDBv5.0 database (Molecular Signature Database <http://www.broadinstitute.org/gsea/msigdb/>).

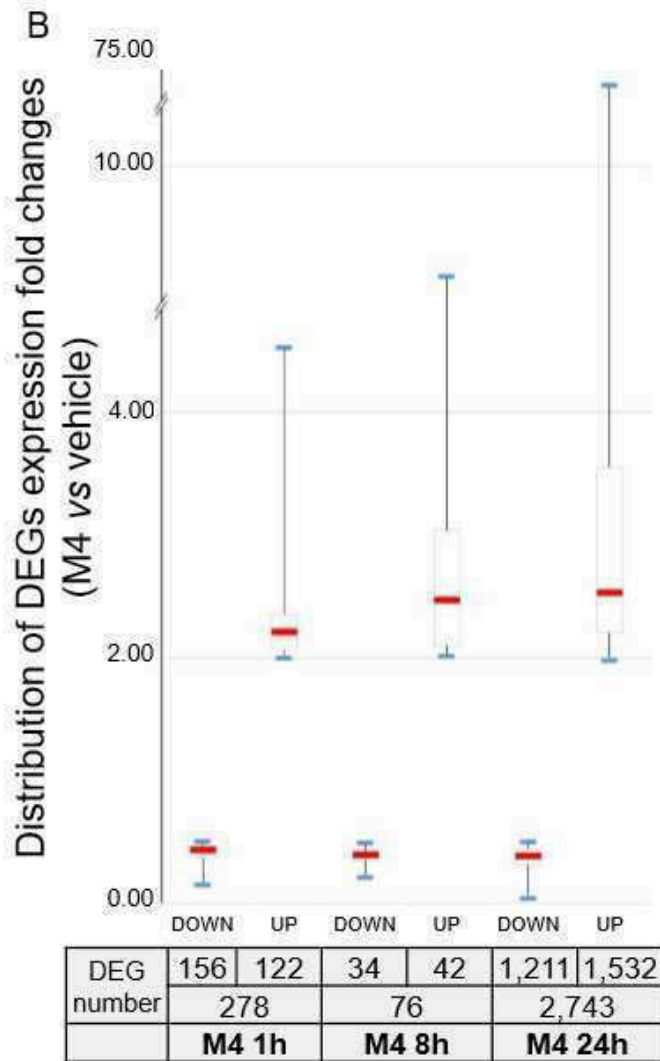
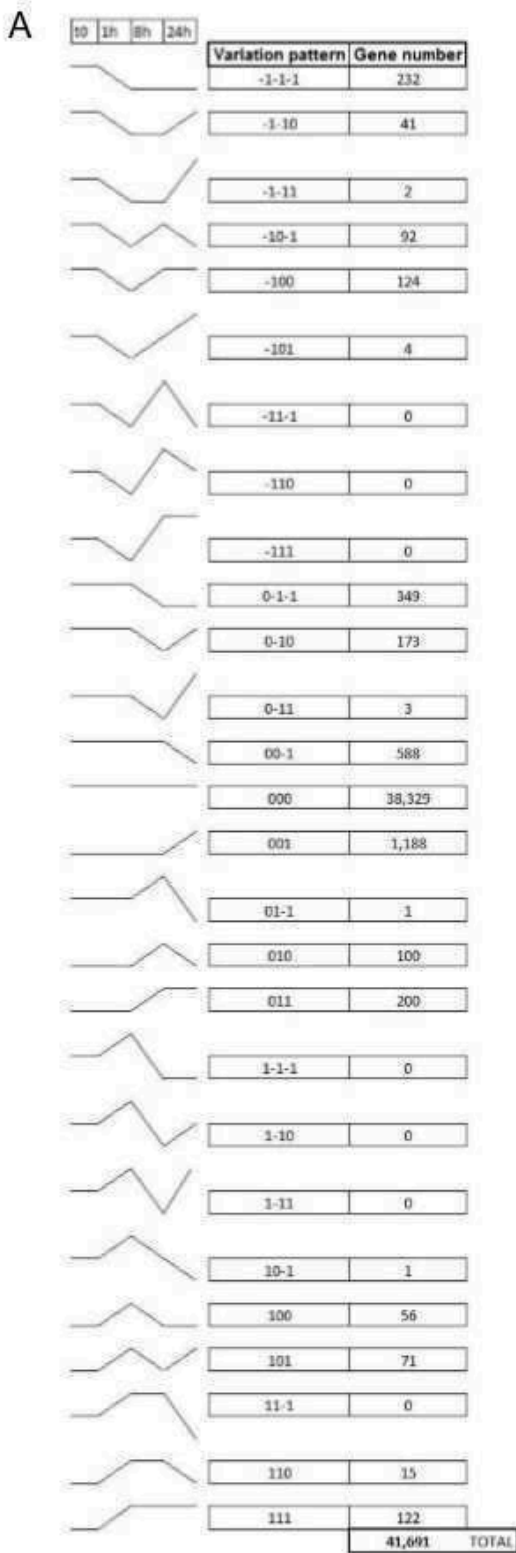
At each point of the kinetics, the maximal number of genes in overlap is obtained for the KEGG Pathway “Pathways in cancer”.



**Supplementary Figure 1: Male inherited transgenerational model of exposure in ERα36 transgenic mouse.**

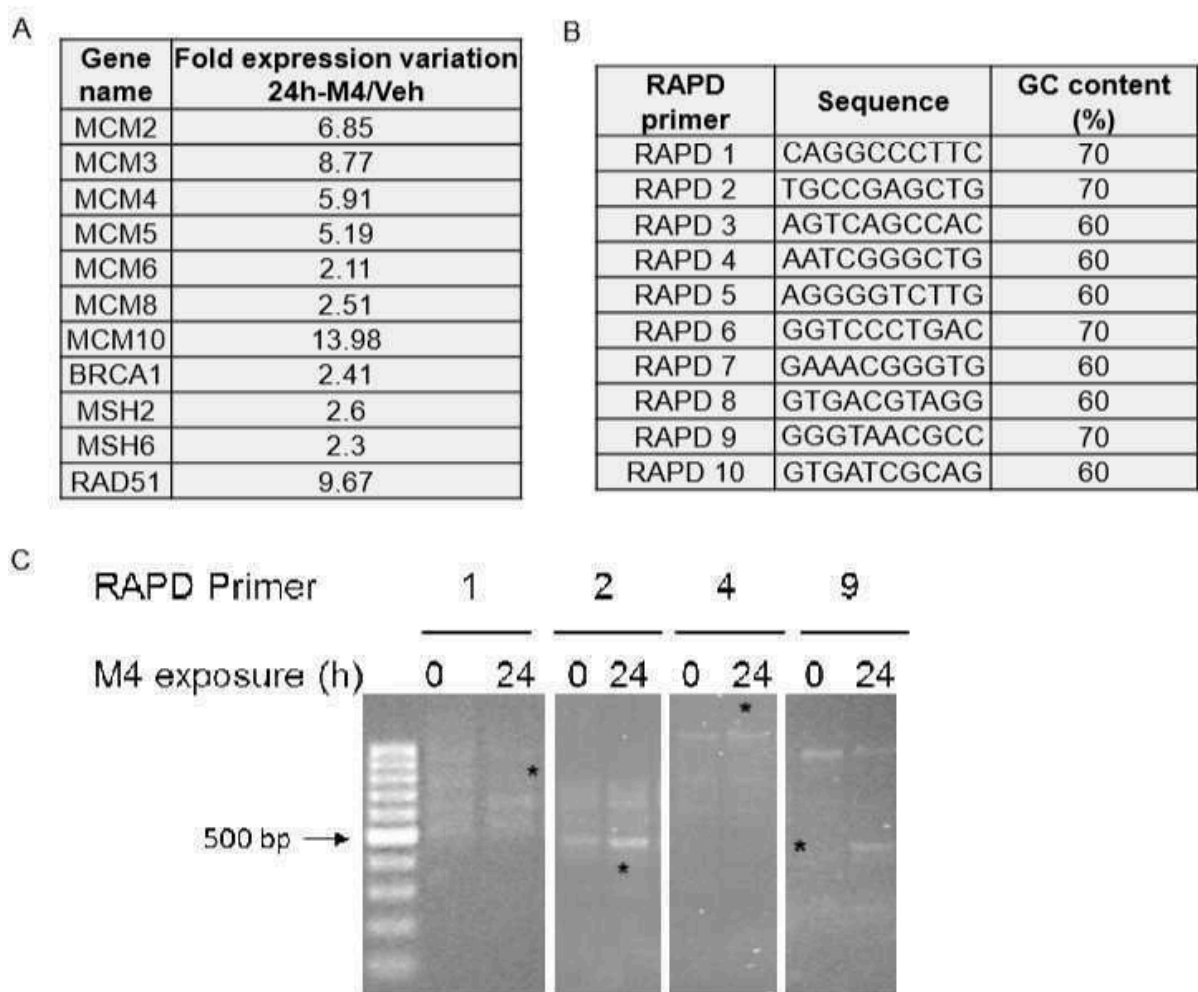
Previously described (16) ERα36 hemizygotic transgenic (Tg) males were mated with P0 C57BL/6J (wt) virgin adult females treated with M4 dissolved in vehicle (0.05μg/kg/day (D1), 0.5μg/kg/day (D2) or 5μg/kg/day (D3) or either sesame oil (vehicle, D0). See materials and methods section for details. Hemizygotic F1 transgenic males were selected as previously described (16) and at least 4 of them, issued from D0 to D3 maternal treatment, were mated with 5 or more wt adult virgin females. The same experiment was repeated with F2 transgenic males. The male inherited transgenerational effect of the 3 different doses of alkylphenol treatment was observed in the mammary gland from F3 progenies. Neither the transgene nor the treatments affected significantly the sex ratio whereas we observed a statistically significant deficiency of transgenic animals (-9% and -8% transgenic males and females respectively;  $p = 0.009$ ) under alkylphenol exposure whatever the dose of M4 given to P0 dams.





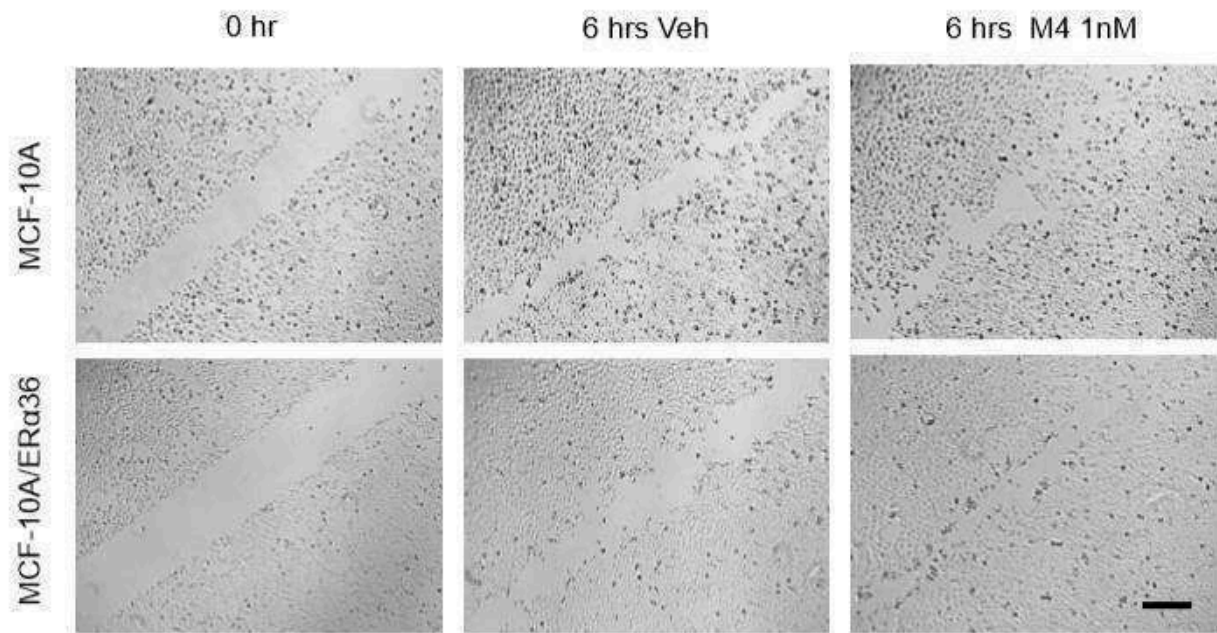
**Supplementary Figure 2. DEG transcriptional profiles and distribution after a kinetic of MCF-10A cell exposure to alkylphenol mixture.**

A) Kinetic variation pattern of MCF10A genes after alkylphenol exposure. The variation of expression is evaluated at 1 hour, 8 hours and 24 hours of treatment versus control. If genes expression variation is under 0.5, index of variation is set at -1 (down-regulated genes). If genes expression variation is over 2, index variation is set at 1 (up-regulated genes). Between 0.5 and 2, gene expression variation is considered non-significantly changed, index variation is set at 0. Gene number in each variation pattern is indicated in the right column. B) Boxplots indicating the distribution of up- or down-regulated DEG expression fold changes (M4 1nM vs vehicle) for each kinetic time point.



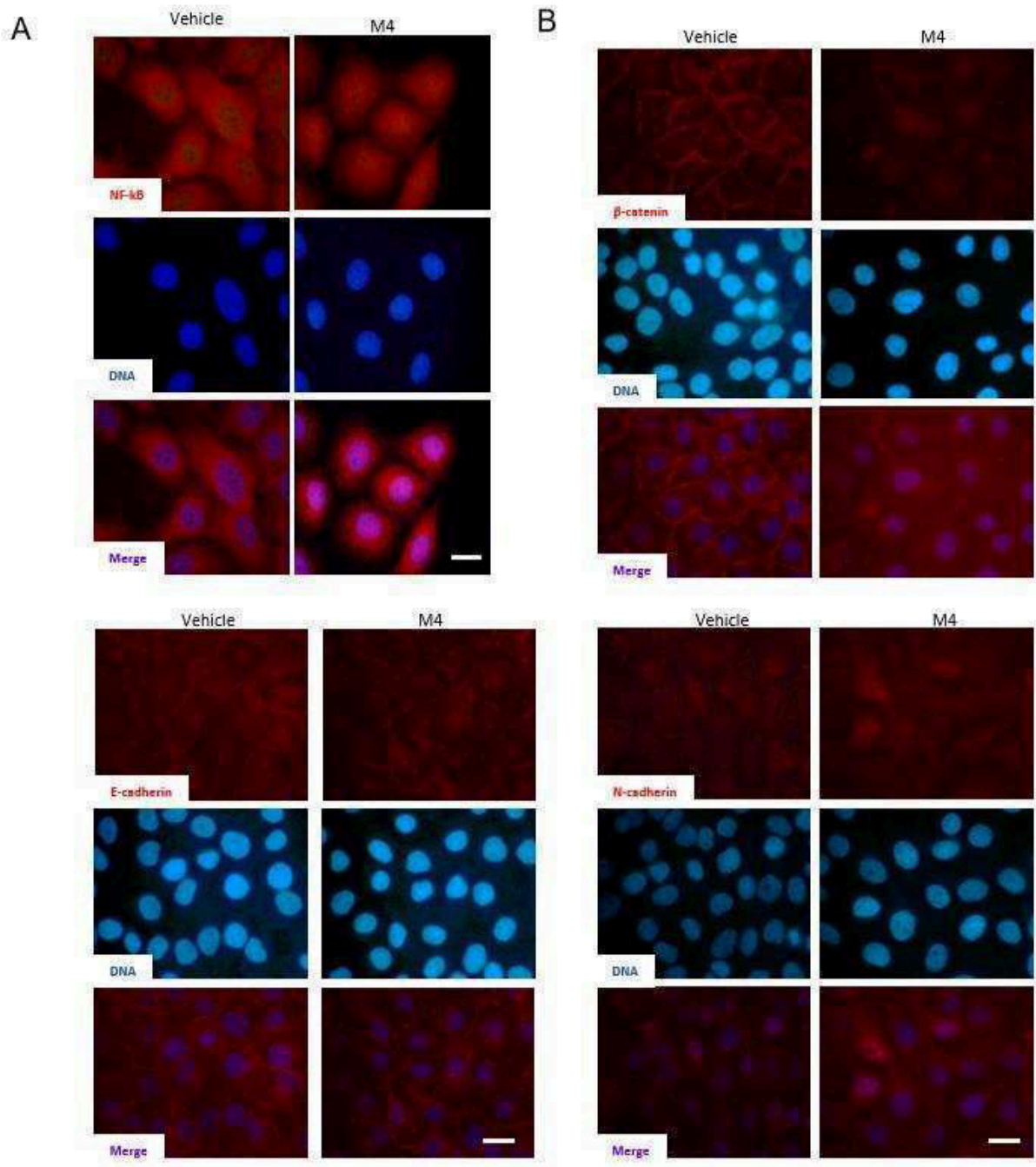
### Supplementary Figure 3. M4 alters genomic stability of MCF-10A cells.

A) Fold expression variation after 24h 1nM M4 exposure of DEGs assigned to “Failed repair of genes”. B) RAPD primer sequences and GC base contents. C) RAPD profiles (26, 27). Ten random GC-rich decanucleotide primers were used to score genomic alterations of MCF-10A cells exposed 24h to 1nM M4 or vehicle. Each genomic DNA pool was amplified with each individual primer. The random amplified polymorphic DNA (RAPD) reaction was performed in a volume of 15 $\mu$ L containing 7.5 $\mu$ L iQ SYBR Green Supermix (Biorad), 0.3 $\mu$ M primer and 20 ng genomic DNA. Cycling parameters were as follows: 95°C for 5 minutes, 40 cycles (95°C for 30 seconds, 37°C for 30 seconds, and 72°C for 1 minute). Polymerase chain reaction (PCR) products were resolved by electrophoresis on a 2% tris-acetateEDTA-agarose gel and visualized by ethidium bromide staining (Geldoc, Biorad). Each experiment was repeated three times. Samples that display loss, addition or significant fluorescence intensity variation of the banding profile are marked (\*) and termed as « genomic unstable ». The frequency of altered RAPD profile is about 50%.



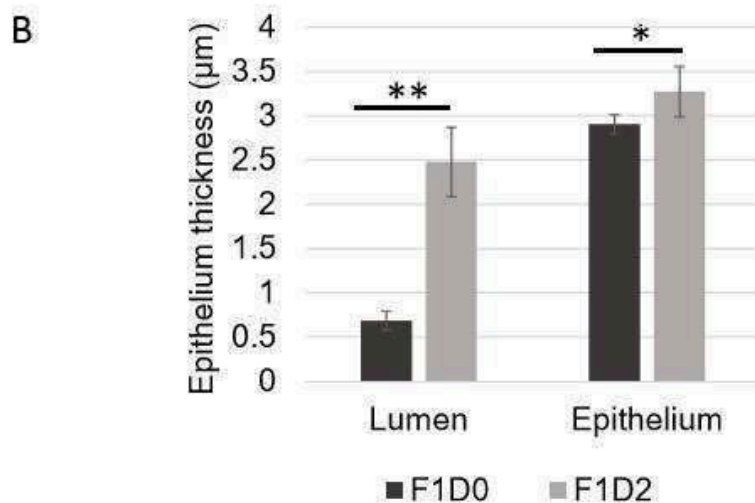
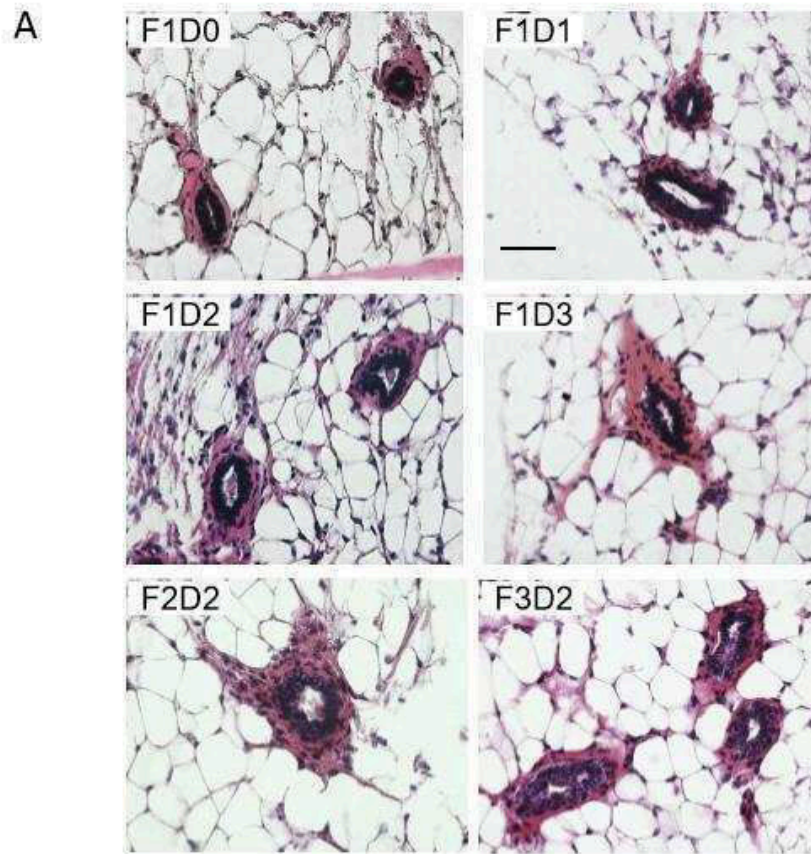
**Supplementary Figure 4. Representative images of scratch wound assays.**

Scratch assays were performed on MCF-10A and MCF-10A/ER $\alpha$ 36 cultures using the Ibidi-culture inserts (Ibidi®/Biovalley) following the manufacturer instructions. Cultures were then washed to remove detached cells and debris and treated with vehicle or M4. Quantification of wound mean diameter were performed at t=0 and after a 6 h exposure by Phase-contrast image analysis with NIS-elements BR 4.20.00 software (Nikon). Scale bar: 1mm.



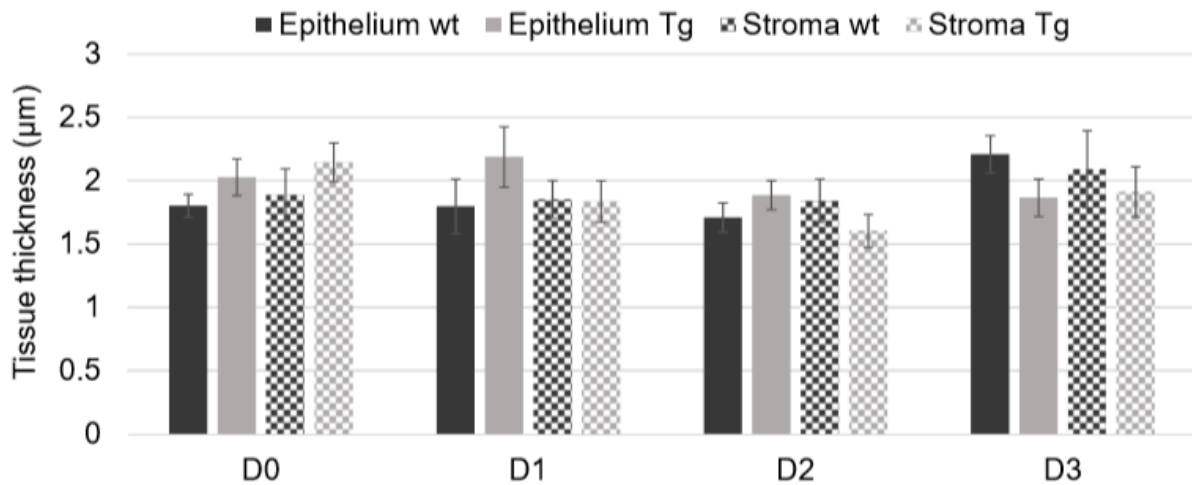
**Supplementary Figure 5. Cellular relocation of NF- $\kappa$ B,  $\beta$ -catenin, E-cadherin, N-cadherin and ZO-1 after M4 treatment.**

MCF-10A cells were treated for 24 h with vehicle or 1nM M4. Localization of NF- $\kappa$ B and cell-cell adhesion proteins were performed by immunofluorescence with specific antibodies (red, AlexaFluor 555). Hoechst was used to stain the cell nuclei (blue). A) NF- $\kappa$ B nuclear translocation is observed in MCF-10A cells exposed to M4. B) A decrease of beta-catenin membrane immunofluorescence staining toward a dispersed cytoplasmic localization and an “E to N-cadherin switch” are observed in M4-treated cells compared to control. Scale bar: 5 $\mu$ m.



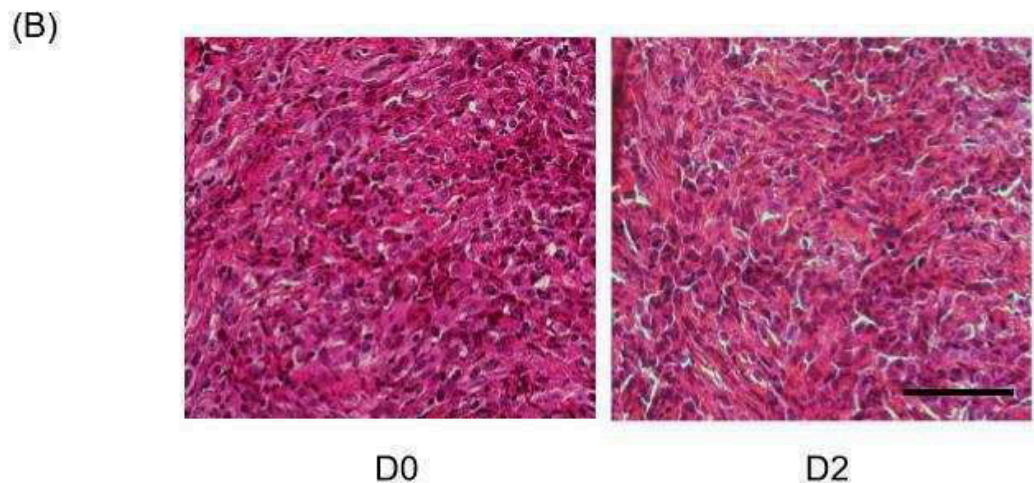
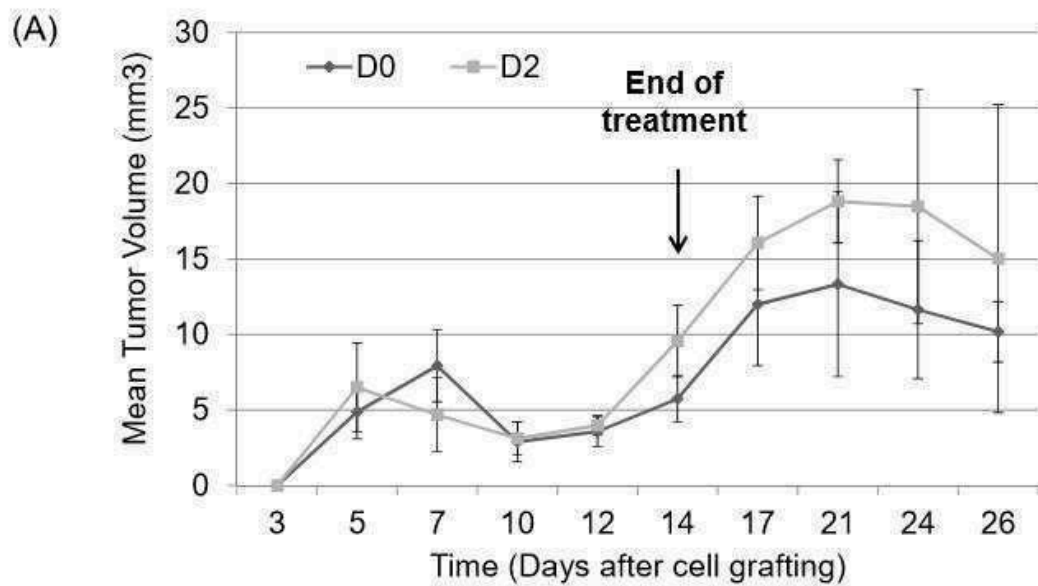
**Supplementary Figure 6. Hematoxylin/Eosin staining of mammary gland at weaning (PND 21).**

A) Representative histology of F1 mammary gland exposed to vehicle (D0) or M4 (D1= 0.05µg/kg/day, D2=0.5µg/kg/day, D3=5µg/kg/day) as well as F2D2 and F3D2 mammary glands. B) Corresponding quantification of lumen diameter and epithelium thickness for F1D0 and F1D2 mammary glands. A significant increase in all measured parameters is observed in D2 exposed animals (N=5) compared to D0 ones (N=8). Scale bar, 10 µm. Each bar represents mean ± S.E.M. All parameters were measured on at least 5 independent slices from each animal included in the experiment. \*: p<0.05; \*\*: p< 0.01.



**Supplementary Figure 7. Quantification of epithelium and stroma thickness in F3 mammary gland following exposure to sesame oil (D0) or M4 (0.05µg (D1), 0.5µg (D2) or 5µg M4/kg/day (D3)).**

Measurements were performed at adulthood (16 weeks). Each bar represents mean  $\pm$  S.E.M. Epithelium or stroma thickness was measured on at least 5 independent slices from each animal included in the experiment.  $N \geq 3$ .



**Supplementary Figure 8 : Effects of M4 exposure on tumor growth in nude mice xenografted with MDA-MB-231 breast cancer cells.**

As described in the Materials and Methods section, D0 and D2 treatments were administered by intrabuccal gavage, during two weeks before and two weeks after cell grafting in the right mammary gland no.4 of nude mice. (A) Tumor growth was measured during four weeks after MDA-MB-231 cells injection. In each group (D0 and D2), one mouse did not develop a tumor. For each point of the kinetics the mean tumor volume (mm<sup>3</sup>)  $\pm$  S.E.M was indicated. N=4 per dose. (B) Hematoxylin/Eosin staining of D0 and D2 mammary tumors collected four weeks after cell grafting. Scale bar, 50  $\mu$ m. No significant differences were observed concerning tumor growth kinetics and tumor histology between the 2 groups.

## II. Quels sont les mécanismes qui régulent l'expression d'ER $\alpha$ 36 ?

*Comme nous l'avons vu en introduction, le variant ER $\alpha$ 36 est décrit dans la littérature pour son rôle majeur dans la progression tumorale et dans le développement de résistances aux traitements anti-tumoraux. La première partie des études menées au cours de ma thèse a montré que l'expression d'ER $\alpha$ 36 dans la glande mammaire saine est associée à l'acquisition de propriétés cellulaires (migration et survie) rappelant une transformation néoplasique de l'épithélium mammaire. Ainsi, l'expression de ce variant pourrait être impliquée dans l'initiation du cancer du sein. Dans ce contexte, l'identification et la compréhension des mécanismes qui déclenchent/régulent l'expression d'ER $\alpha$ 36 pourrait permettre de développer des stratégies thérapeutiques pour prévenir les effets délétères de l'expression d'ER $\alpha$ 36 dans les tumeurs.*

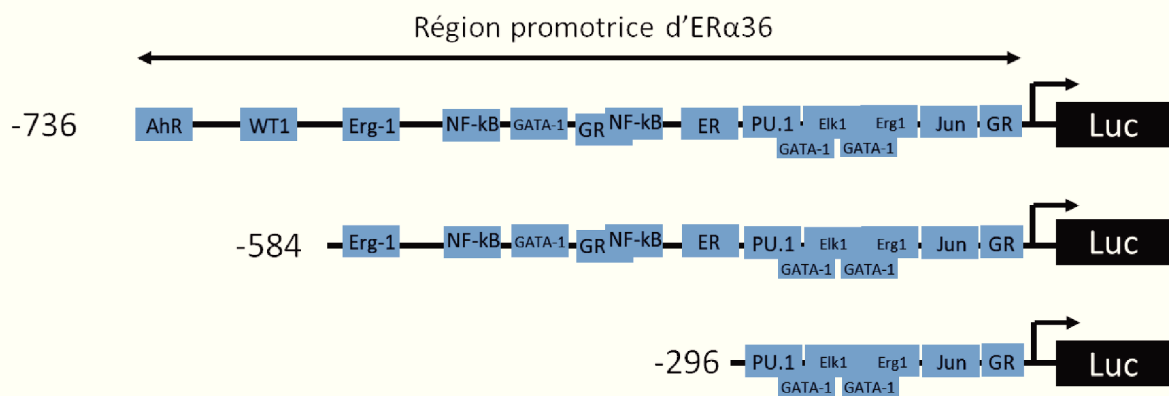
### A) Régulation de l'expression d'ER $\alpha$ 36 par des facteurs trans-régulateurs

*Les résultats obtenus concernant la régulation de l'expression d'ER $\alpha$ 36 par des facteurs trans-régulateurs n'ont pas encore fait l'objet d'une publication mais ils sont présentés comme résultats complémentaires.*

Avant ma thèse, seuls quelques mécanismes classiques de régulation transcriptionnelle de l'expression d'ER $\alpha$ 36 avaient été décrits dans la littérature. En effet, il avait été montré que la forme canonique ER $\alpha$ 66 et le facteur de transcription WT1 agissaient comme des régulateurs négatifs de l'expression de ce variant tandis que la signalisation induite par EGFR et/ou HER2 conduisait à l'induction de l'expression d'ER $\alpha$ 36 par l'intermédiaire des facteurs de transcription de la famille AP-1 (Zou et al., 2009, Kang et al., 2011b, Yin et al., 2014).

Afin d'apporter de nouveaux éléments concernant la régulation de l'expression d'ER $\alpha$ 36 par des facteurs trans-régulateurs, nous avons examiné la séquence de la région promotrice de ce variant décrite par Zou et al., (2009). Nous avons choisi de nous focaliser sur AhR, un récepteur décrit comme un médiateur des effets des xénobiotiques, qui joue un rôle majeur dans l'initiation et la progression des tumeurs mammaires (Goode et al., 2013). En effet, on peut émettre l'hypothèse que ce récepteur, qui possède un élément de réponse dans la région promotrice d'ER $\alpha$ 36, pourrait médier l'induction de l'expression de ce variant en réponse à des xénobiotiques œstrogène-mimétiques comme les alkylphénols. De plus, il a été démontré qu'en présence de dioxine (TCDD), AhR est capable d'inhiber l'activité de la forme canonique des récepteurs aux œstrogènes, ER $\alpha$ 66 selon plusieurs mécanismes. AhR peut notamment





**Figure 57 : Système rapporteur de l'activité du promoteur d'ERα36.**

Ce système rapporteur a été décrit par Kang et al., (2010).

Le gène codant la luciférase (Luc) est placé sous le contrôle de la région promotrice d'ERα36 sous sa forme longue (-736 pb) ou tronquée (-584 pb et -296 pb). Les différents éléments de réponse sont représentés par des boîtes bleues et le nom des facteurs de transcription associés est indiqué.

s'associer avec ER $\alpha$ 66 pour l'empêcher d'interagir avec les ERE ou induire sa dégradation par le protéasome (Matthews et Gustafsson, 2006). Ainsi, AhR pourrait réguler de manière indirecte l'expression d'ER $\alpha$ 36 en neutralisant l'effet inhibiteur d'ER $\alpha$ 66 sur l'expression de son variant. C'est pourquoi nous avons mis en œuvre plusieurs stratégies expérimentales dans la lignée MCF-7 (ER $\alpha$ 66+) afin d'évaluer l'impact d'AhR sur la régulation de l'expression d'ER $\alpha$ 36.

Après avoir stimulé l'activité d'AhR en traitant les cellules cancéreuses mammaires MCF-7 avec du BaP à 25  $\mu$ M pendant 24h, nous avons analysé :

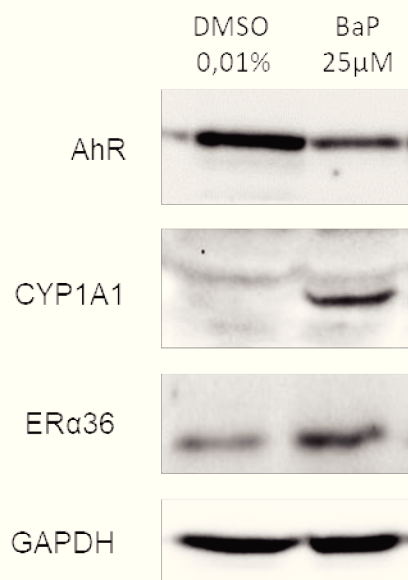
- (i) La variation du niveau d'expression d'ER $\alpha$ 36 par western blot.
- (ii) L'activité de la région promotrice d'ER $\alpha$ 36 grâce à un système rapporteur.

Plus précisément, nous avons voulu vérifier si l'effet d'AhR dépendait de son élément de réponse spécifique localisé dans la région promotrice d'ER $\alpha$ 36, en position -642 à -651 pb. Pour cela, nous avons reproduit au laboratoire le système rapporteur décrit par Kang et al., (2010) qui correspond au gène codant la luciférase sécrétée placé sous le contrôle de tout ou partie de la région promotrice d'ER $\alpha$ 36 décrite par Zou et al., (2009) (Figure 57).

- (iii) L'interaction entre les protéines AhR et ER $\alpha$ 36 par la technique PLA (Proximity Ligation Assay) qui permet de visualiser des interactions entre 2 protéines intracellulaires d'intérêt.

Etant donné que AhR est capable d'interagir avec ER $\alpha$ 66 pour neutraliser son activité, nous avons voulu déterminer si ce type d'interaction peut également être détecté entre AhR et ER $\alpha$ 36. Cette expérimentation, a été menée au sein de l'équipe « Signalisation des stéroïdes et cancer du sein » dirigée par Muriel Le Romancer (DR INSERM, CRCL, Centre Léon Bérard, Lyon). Le kit « Duolink<sup>®</sup> PLA Technology » (Sigma-Aldrich<sup>®</sup>, St. Quentin Fallavier, France) a été utilisé en suivant les instructions du fournisseur.

Quelle que soit la technique utilisée, les résultats obtenus n'ont pas permis de mettre en évidence de variation de l'activité du promoteur d'ER $\alpha$ 36, ni d'interaction entre AhR et ce variant en présence de BaP. Plusieurs hypothèses pourraient permettre d'expliquer ce constat. Tout d'abord, il est possible que notre système rapporteur ne permette pas de détecter de manière significative des variations de l'activité du promoteur d'ER $\alpha$ 36. En effet, lorsque l'on module l'activité d'ER $\alpha$ 66 avec un inhibiteur comme le fulvestrant, aucune modification de l'activité luciférase, c'est-à-dire de l'activité du promoteur complet (-736) d'ER $\alpha$ 36, n'est



**Figure 58 : Impact de l'activation d'AhR par son ligand, le BaP, sur l'expression d'ERα36.**

Cette étude a été réalisée sur des cellules MDA-MB-231 traitées pendant 24h avec du DMSO 0,01% ou du BaP (ligand d'AhR) à 25 µM. L'expression d'AhR, de CYP1A1 (gène cible d'AhR) et d'ERα36 a été détectée par Western Blot. La GAPDH a été utilisée comme invariant. L'image présentée est représentative de 3 réplicats réalisés de manière indépendante.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

Anticorps utilisés : anti-AhR (17840, Proteintech), anti-CYP1A1 (13241, Proteintech), anti-ERα36 (G20, sc-544, Santa Cruz), anti-GAPDH (sc-2055, Santa Cruz).

**BaP** = Benzo[a]Pyrène, **CYP1A1** = Cytochrome P450 1A1

observée alors que l'effet inhibiteur d'ER $\alpha$ 66 sur l'expression de son variant a déjà été démontrée. De plus, nous avons observé un faible niveau d'expression d'AhR et d'ER $\alpha$ 36 dans les cellules cancéreuses mammaires MCF-7, qui pourrait être insuffisant pour déclencher une variation d'expression ou détecter une interaction.

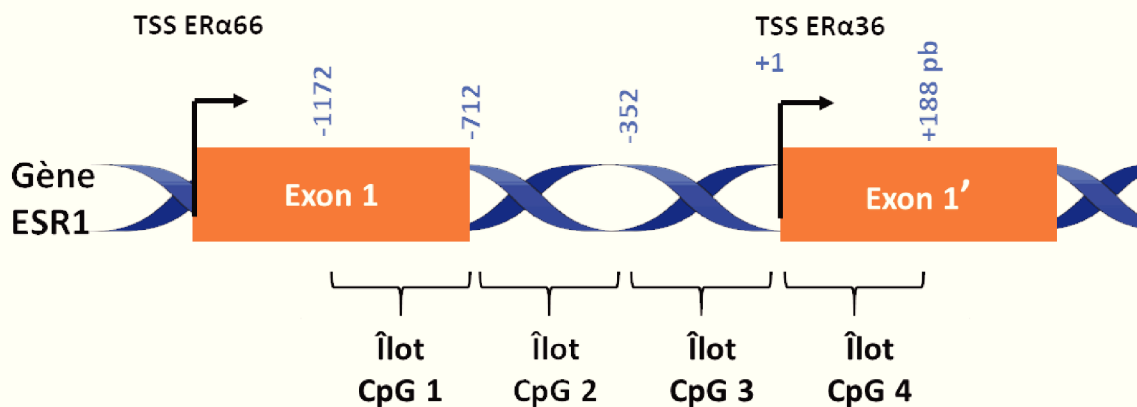
Par conséquent, nous avons débuté une nouvelle série d'expérimentations sur les cellules cancéreuses mammaires triples négatives MDA-MB-231 qui contrairement aux cellules cancéreuses MCF-7 présentent un fort niveau d'expression endogène d'AhR.

Les résultats préliminaires obtenus indiquent que lorsque l'on stimule l'activité d'AhR en traitant ces cellules pendant 24h avec du BaP (Benzo-[a]-pyrène) à 25 $\mu$ M, on observe en western blot une augmentation de l'expression de l'enzyme du métabolisme des xénobiotiques, CYP1A1, qui témoigne de l'activation d'AhR. On observe également une augmentation de l'expression d'ER $\alpha$ 36 qui semble valider le fait qu'AhR pourrait réguler l'expression de ce variant dans cette lignée cellulaire (Figure 58).

Dans la continuité, il serait donc intéressant d'étudier l'activité du promoteur d'ER $\alpha$ 36 et l'interaction entre ce variant et AhR dans le contexte des cellules MDA-MB-231. Toutefois, dans cette lignée ER $\alpha$ 66-, nous ne pourrions pas conclure sur l'implication éventuelle d'ER $\alpha$ 66 dans la régulation d'ER $\alpha$ 36 par AhR. Pour pallier à ce problème, une des solutions serait de déterminer le niveau d'expression d'AhR et d'ER $\alpha$ 36 dans différents types de cellules cancéreuses mammaires appartenant au sous-type « Luminal A » et de répéter cette étude dans celles qui expriment le plus ces 2 facteurs.

Enfin, l'interaction directe entre AhR et la région promotrice d'ER $\alpha$ 36 sera étudiée grâce à la technique d'immunoprécipitation de la chromatine (ChIP) qui est actuellement en cours de mise au point au laboratoire.

En parallèle des modalités de régulations classiques de l'expression d'ER $\alpha$ 36 par des facteurs trans-régulateurs, nous avons souhaité concentré la majeure partie de nos études sur les mécanismes de régulation épigénétique et post-transcriptionnelle (micro-ARNs) de l'expression de ce variant. En effet, un nombre croissant d'études démontrent que ces mécanismes jouent un rôle central dans les processus de carcinogénèse et de progression tumorale (Chatterjee et al., 2018, Tan et al., 2018).



**Figure 59 : Représentation schématique de la localisation des 4 îlots CpGs susceptibles de réguler l'expression d'ERα36.**

Les exons 1 et 1' du gène ESR1 sont représentés par des rectangles oranges. Les sites d'initiation de la transcription (TSS) d'ERα66 et d'ERα36 sont représentés par des flèches.

Les positions de chaque îlots CpG, relatives au site d'initiation de la transcription d'ERα36 (+1), sont indiquées en bleu et en paires de bases (pb). Sur cette représentation schématique, les échelles ne sont pas respectées.

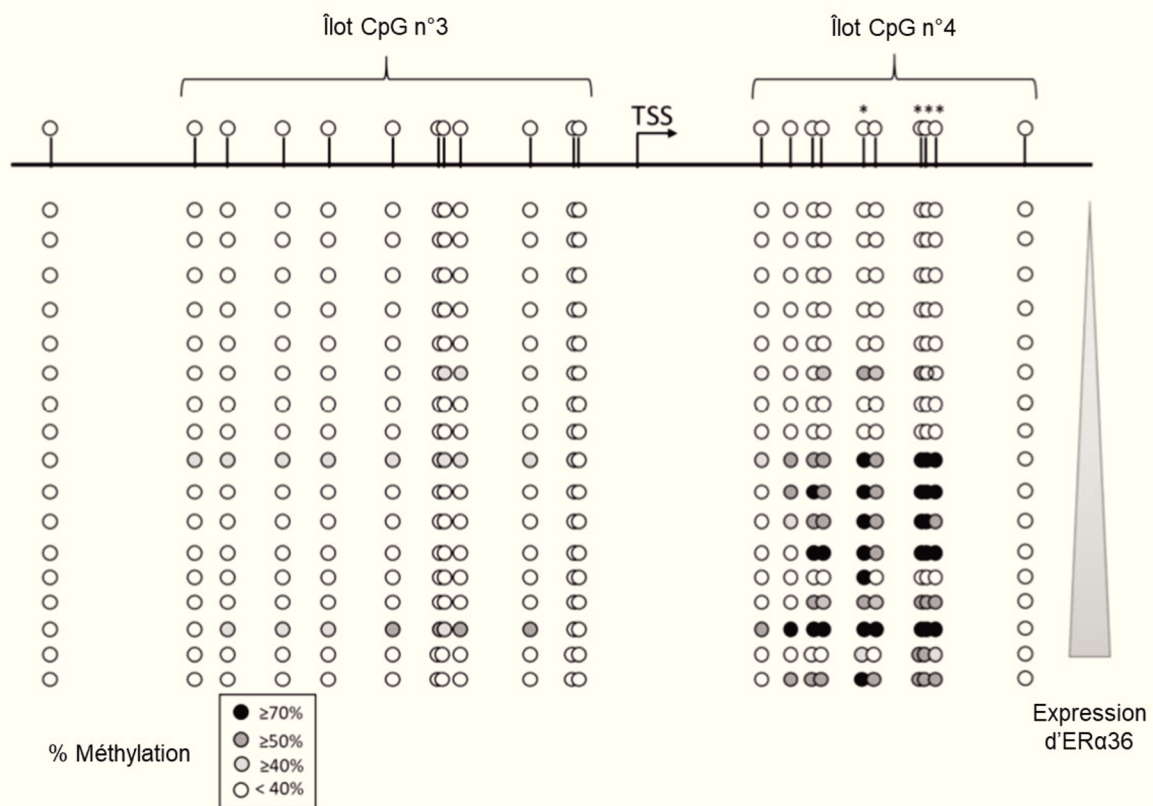
### B) Régulation de l'expression d'ER $\alpha$ 36 par le statut de méthylation de son promoteur

Les études précédentes menées au laboratoire ont montré que l'expression d'ER $\alpha$ 36 est induite dans les cellules séminomateuses TCam-2 et dans les cellules épithéliales mammaires saines MCF-10A en réponse aux alkylphénols, qui sont suspectés de moduler le statut épigénétique des cellules. Sur la base de ce constat nous avons voulu déterminer si l'activité du promoteur d'ER $\alpha$ 36 pourrait être régulée par son statut de méthylation. L'ensemble des résultats obtenus sont décrits dans **la publication N°3**.

Grâce à l'utilisation de la base de données «EMBOSS CpG Plot» ([https://www.ebi.ac.uk/Tools/seqstats/emboss\\_cpgplot/](https://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/)), 4 îlots CpG localisés dans la région promotrice d'ER $\alpha$ 36 ont été identifiés (Figure 59). Le statut de méthylation de chacun de ces îlots a été déterminé par séquençage au bisulfite dans 17 biopsies de tumeurs mammaires présentant des niveaux d'expression variables d'ER $\alpha$ 36. Le calcul de la corrélation entre le niveau de méthylation de chacun des sites CpG localisés dans ces îlots et le niveau d'expression d'ER $\alpha$ 36, grâce à un programme « Matlab », indique que seule la méthylation de 4 sites CpG localisés au niveau de l'îlot 4 a un impact significatif sur l'expression de ce variant dans les tumeurs mammaires (Figure 60). En effet, il apparaît que plus ces sites CpG sont méthylés, plus le niveau d'expression d'ER $\alpha$ 36 augmente. Bien que cela semble contraire à ce qui est classiquement décrit, cette corrélation entre une hyperméthylation et l'induction de l'expression d'un gène est fréquemment observée lorsque les sites CpG impliqués dans la régulation sont localisés en aval du TSS (Ball et al., 2009 ; Ehrlich et Lacey, 2013).

Ce lien direct entre le niveau de méthylation du promoteur d'ER $\alpha$ 36 et l'expression de ce variant, a également été confirmé in vitro en exposant les cellules cancéreuses mammaires MCF-7 à un agent déméthylant, la Décitabine (DAC). En effet, ce traitement qui induit une déméthylation des 4 sites CpG précédemment identifiés, conduit à une diminution du niveau d'expression d'ER $\alpha$ 36 sans impacter la viabilité cellulaire (Figure 61).

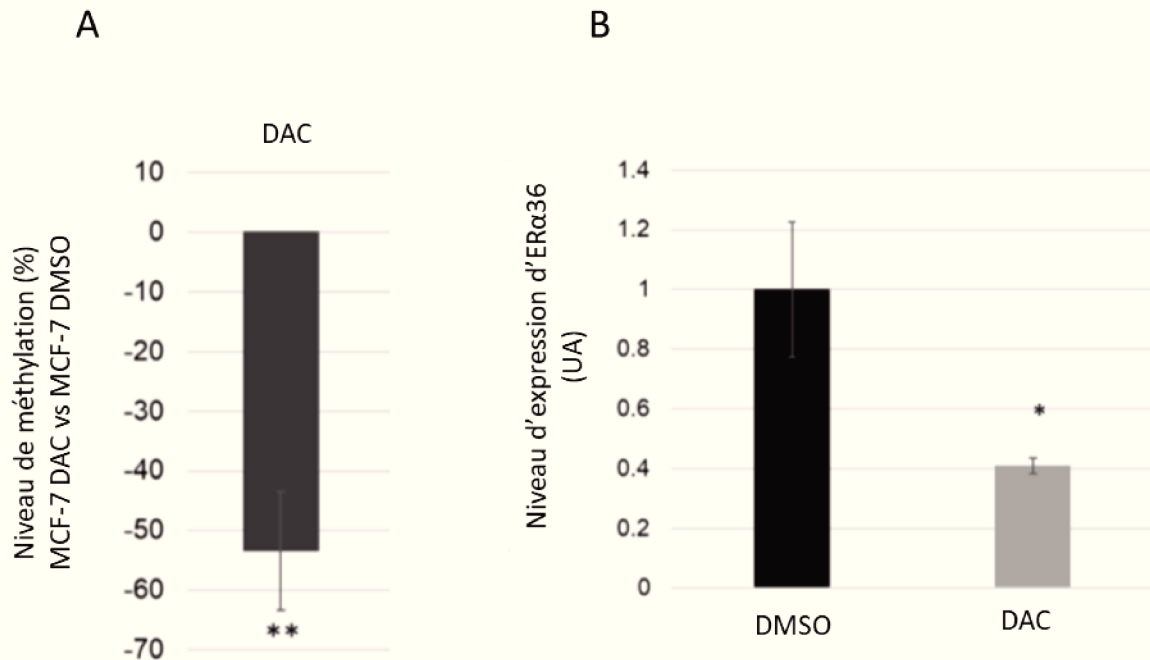
De manière intéressante, la DAC a été récemment décrite dans la littérature pour son rôle dans la réinduction de l'expression d'ER $\beta$ 1 dans les tumeurs mammaires et dans la réinduction de l'expression d'ER $\alpha$ 66 dans le contexte des ostéosarcomes (Al-nakhle et al., 2013 ; Lillo Osuna et al., 2019). Sachant, que cette forme canonique régule négativement l'expression de son variant, nous avons étudié l'éventuelle implication d'ER $\alpha$ 66 dans la modulation du statut de méthylation du promoteur d'ER $\alpha$ 36. En effet, il a été récemment décrit qu'ER $\alpha$ 66 est capable de réprimer l'expression de ces gènes cibles en recrutant différents facteurs impliqués dans la



**Figure 60 : Etude de la corrélation entre le statut de méthylation du promoteur d'ER $\alpha$ 36 et le niveau d'expression de ce variant dans des biopsies de tumeurs mammaires.** (Figure 5 de la publication n°3)

Le niveau de méthylation des sites CpG, localisés au niveau des îlots 3 et 4 de la région promotrice d'ER $\alpha$ 36, a été déterminé par séquençage au bisulfite dans 17 biopsies de tumeurs mammaires extraites de la biobanque de l'Institut de Cancérologie de Lorraine (ICL). En parallèle, le niveau d'expression d'ER $\alpha$ 36 dans chacune des tumeurs a été mesuré par RT-qPCR. La corrélation entre le statut de méthylation de ces différents sites CpG et le niveau d'expression d'ER $\alpha$ 36 a été calculée grâce au test de corrélation de Pearson.

Les 4 sites CpG dont le niveau de méthylation est significativement corrélé à l'expression d'ER $\alpha$ 36 dans les biopsies de tumeurs mammaires sont indiqués par une étoile (\*).



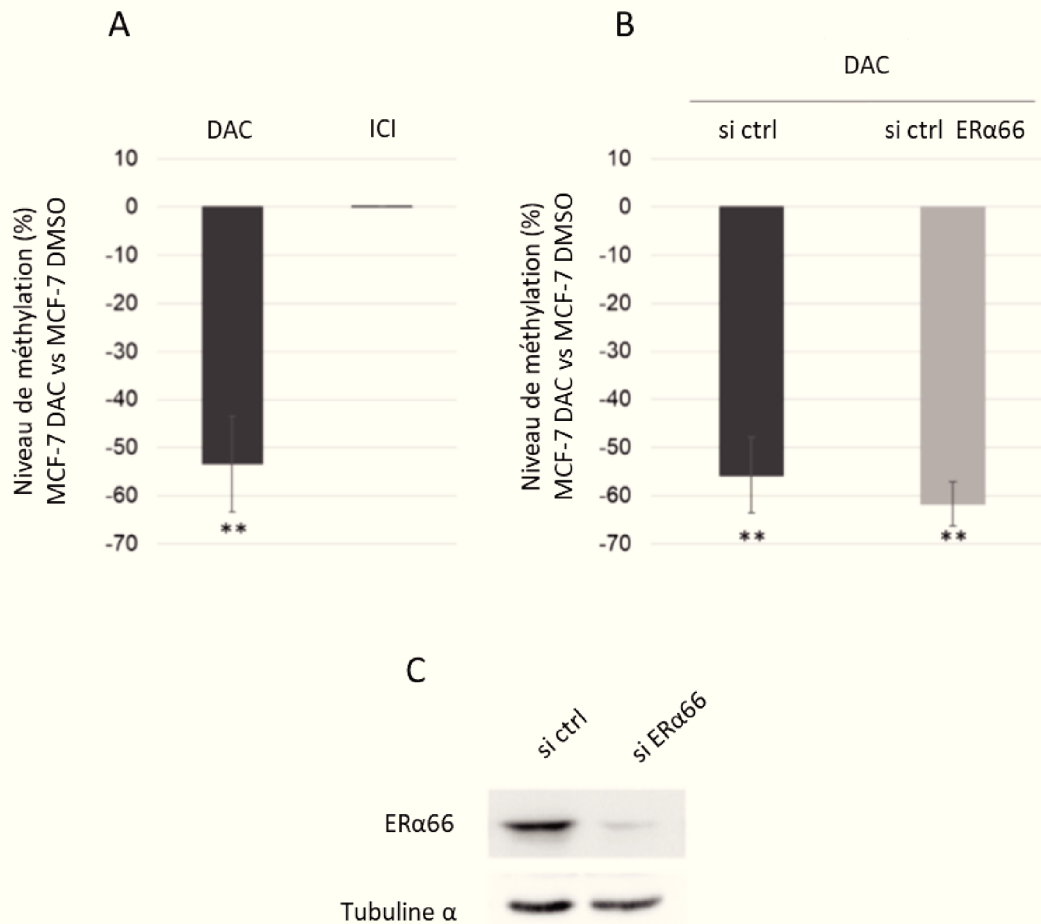
**Figure 61 : Impact de la décitabine sur la méthylation du promoteur d'ER $\alpha$ 36 et l'expression de ce variant.** (Figures 3B et 6A de la publication n°3)

- (A) Impact de la décitabine (DAC) sur le statut de méthylation des 4 sites CpG corrélés à l'expression d'ER $\alpha$ 36. Le niveau de méthylation de ces 4 sites a été déterminé par PCR en temps réel après conversion au bisulfite de l'ADN des cellules MCF-7, exposées pendant 48 heures au DMSO 0,01% ou à la DAC 5  $\mu$ M. On observe une diminution significative du niveau de méthylation de ces 4 sites CpG après traitement à la DAC.
- (B) Impact de la DAC sur l'expression d'ER $\alpha$ 36. Le niveau d'expression d'ER $\alpha$ 36 a été déterminé par RT-qPCR à partir des ARNs totaux des cellules MCF-7 exposées pendant 48 heures au DMSO 0,01% ou à la DAC 5  $\mu$ M. On observe une diminution significative du niveau d'expression d'ER $\alpha$ 36 après traitement à la DAC.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. (A) N=4, (B) N=3.\*: P < 0,05; \*\*: P < 0,01. UA = Unité arbitraire





**Figure 62 : Implication d'ERα66 dans la déméthylation de la région promotrice d'ERα36 induite par la décitabine.** (Figure 6 de la publication n°3)

- (A) Impact du Fulvestrant (ICI), un inhibiteur d'ERα66, sur le statut de méthylation des 4 sites CpG, associés à l'expression d'ERα36. Le niveau de méthylation de ces 4 sites a été déterminé par PCR en temps réel après conversion au bisulfite de l'ADN des cellules MCF-7 exposées pendant 48 heures au DMSO 0,01% ou à la DAC 5 μM ou à l'ICI 1 μM. On observe que, contrairement à la DAC, qui déméthyle les 4 sites CpG associés à l'expression d'ERα36, l'ICI n'a pas d'impact sur leur statut de méthylation.
- (B) Niveau de méthylation des 4 sites CpG associés à l'expression d'ERα36, dans des cellules MCF-7, (i) transfectées avec un siRNA contrôle (si ctrl) ou un siRNA ciblant les ARNm codant ERα66 (si ERα66) et (ii) exposées pendant 48 heures au DMSO 0,01% ou à la DAC 5 μM. On observe que l'effet de la DAC, sur la déméthylation des sites CpG étudiés, n'est pas impacté par l'absence d'ERα66.
- (C) La diminution de l'expression d'ERα66 en présence du siRNA ciblant l'ARNm d'ERα66 a été confirmée par western blot. La tubuline α a été utilisée comme invariant.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. (A) N=4, (B) si ctrl : N=3; si ERα66 : N=4. \*\* : P < 0,01.

régulation du statut épigénétique comme DNMT3B, EZH2 (HMT) ou des HDACs. Ainsi ER $\alpha$ 66 peut agir aussi bien sur les modifications des histones que sur le statut de méthylation des îlots CpG (Ariazi et al., 2017).

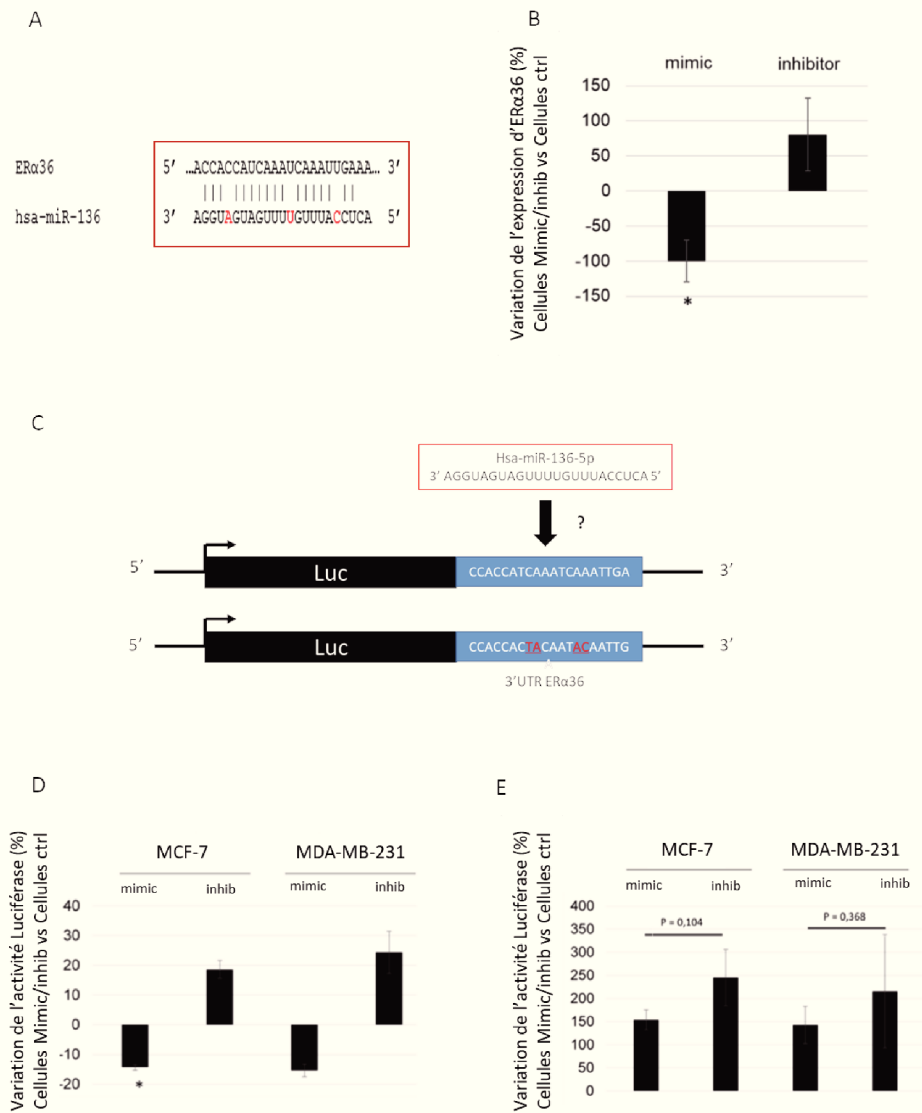
Les résultats présentés dans la Figure 62 indiquent que l'inhibition d'ER $\alpha$ 66 par le fulvestrant n'impacte pas le niveau de méthylation des 4 sites CpG corrélés à l'expression d'ER $\alpha$ 36. De plus, la DAC déméthyle ces 4 sites de manière similaire en présence d'un siRNA contrôle ou d'un siRNA ciblant ER $\alpha$ 66. Ainsi, il semble que la forme canonique n'impacte pas le statut de méthylation du promoteur d'ER $\alpha$ 36.

### C) Régulation post-transcriptionnelle de l'expression d'ER $\alpha$ 36 par des micro-ARNs

Etant donné qu'un nombre croissant d'études mettent en avant l'implication des micro-ARNs dans la progression des tumeurs de diverses origines et notamment des tumeurs mammaires, nous avons émis l'hypothèse que l'expression d'ER $\alpha$ 36 pourrait être régulée au niveau post-transcriptionnel par des micro-ARNs. L'ensemble des résultats obtenus sont décrits dans **la publication N°3**.

Afin d'identifier les micro-ARNs susceptibles de réguler négativement l'expression d'ER $\alpha$ 36 sans impacter celle d'ER $\alpha$ 66, nous avons soumis la séquence des ARNm codant chacun de ces 2 isoformes d'ER $\alpha$  dans la base de données « miRBase » (<http://www.mirbase.org/>). Cette analyse a permis d'identifier 12 micro-ARNs susceptibles de cibler la région 3'UTR de l'ARNm codant ER $\alpha$ 36 mais pas celle de l'ARNm codant ER $\alpha$ 66. Parmi ces micro-ARNs, nous avons choisi de nous intéresser plus particulièrement à hsa-miR-136-5p qui d'après la base de données « miRmine » (Panwar et al., 2017) est 2 fois moins exprimé dans le sérum des patientes atteintes d'un cancer du sein que dans le sérum des patientes non atteintes.

De plus, les informations disponibles dans la littérature indiquent que l'expression de ce micro-ARNs est inversement corrélée avec le stade/grade et le potentiel métastatique des tumeurs et des lignées cellulaires mammaires. En effet, hsa-miR-136-5p est : (i) fortement exprimé dans le tissu sain et la lignée de cellules épithéliales mammaires saines MCF-10A qui exprime faiblement ER $\alpha$ 36, (ii) moyennement exprimé dans les tumeurs de stades I et II équivalentes à la lignée de cellules cancéreuses mammaires luminales MCF-7 et (iii) faiblement exprimé dans les tumeurs de haut grade et les lignées de cellules cancéreuses mammaires triples négatives comme la lignée MDA-MB-231 qui exprime fortement ER $\alpha$ 36 (Yan M. et al., 2016). De la même façon, hsa-miR-136-5p est décrit pour son rôle suppresseur de tumeur dans le contexte



**Figure 63 : Etude de l'interaction entre hsa-miR-136-5p et la région 3'UTR de l'ARNm d'ERα36.** (Figure 2 de la publication n°3)

- (A) Représentation de l'appariement imparfait, prédit par la base de données « miRBase », entre hsa-miR-136-5p et la région 3'UTR de l'ARNm d'ERα36.
- (B) Impact de la variation du niveau d'hsa-miR-136-5p intracellulaire sur l'expression d'ERα36. Les cellules MCF-7 ont été transfectées avec un oligonucléotide « hsa-miR-136-5p mimic » ou « hsa-miR-136-5p inhibiteur » afin d'augmenter ou de diminuer artificiellement la quantité d'hsa-miR-136-5p intracellulaire. 24 heures après la transfection, le niveau d'expression d'ERα36 a été mesuré par RT-qPCR. Le pourcentage de variation, de l'expression d'ERα36 dans les cellules transfectées (mimic ou inhib) par rapport aux cellules contrôles (ctrl = non transfectées), est représenté sur le graphique.
- (C) Représentation schématique des 2 constructions utilisées pour valider l'interaction directe entre hsa-miR-136-5p et la région 3'UTR de l'ARNm d'ERα36. Ce système rapporteur correspond au gène codant la luciférase à l'extrémité duquel a été cloné une partie de la séquence de la région 3'UTR de l'ARNm d'ERα36, sous sa forme sauvage (wt) ou mutante (mut). Les mutations correspondent à des inversions de nucléotides et sont représentées en rouge sur le schéma.
- (D) Etude de l'interaction directe entre hsa-miR-136-5p et la région 3'UTR de l'ARNm d'ERα36. Les cellules MCF-7 et MDA-MB-231 ont été co-transfectées avec les oligonucléotides « hsa-miR-136-5p mimic » ou « hsa-miR-136-5p inhibiteur » et avec le système rapporteur wt. L'activité luciférase a été mesurée 24 heures après la transfection.
- (E) Afin de valider la spécificité de l'interaction entre hsa-miR-136-5p et la région 3'UTR de l'ARNm d'ERα36, la même expérimentation a été réalisée avec le système rapporteur mut.

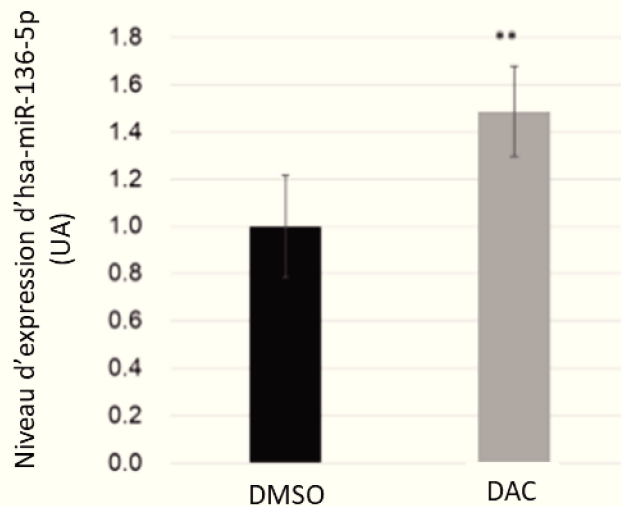
Les barres d'erreur correspondent à l'erreur standard à la moyenne. (B, E) N=5, (D) N=3. \* : P < 0,05.

du cancer de la prostate, des carcinomes rénaux, des adénocarcinomes pulmonaires ou encore des ostéosarcomes (Chen P. et al., 2018 ; Guo et Pan, 2018 ; Li T.T. et al., 2018; Xie Z.C. et al., 2018 ; Zhu Y. et al., 2018).

Dans un premier temps, l'impact de l'expression d'hsa-miR-136-5p sur l'expression d'ER $\alpha$ 36 a été évalué *in vitro* dans les cellules MCF-7 grâce à l'utilisation d'un oligonucléotide « hsa-miR-136-5p mimic » permettant d'augmenter artificiellement le niveau d'expression de ce micro-ARN ou d'un oligonucléotide « hsa-miR-136-5p inhibiteur » permettant de neutraliser son expression endogène. Les résultats obtenus, présentés dans les Figures 63A et B, démontrent qu'il existe bien une relation inverse entre l'expression de ce micro-ARN et celle d'ER $\alpha$ 36. De plus, grâce à un système rapporteur, correspondant au gène codant la luciférase à l'extrémité duquel nous avons cloné la séquence de la région 3'UTR de l'ARNm codant ER $\alpha$ 36 sous une forme sauvage ou mutante, nous avons montré que hsa-miR-136-5p est bien capable de réguler l'expression de ce variant en interagissant directement et spécifiquement avec la région 3'UTR de son ARNm (Figure 63 C, D et E).

Sachant que hsa-miR-136-5p semble être important pour la régulation de l'expression d'ER $\alpha$ 36 dans le contexte des tumeurs mammaires, nous avons ensuite voulu déterminer quels sont les mécanismes qui pourraient moduler l'expression de ce micro-ARN.

Le gène codant hsa-miR-136-5p est localisé au niveau du locus DLK1-DIO3 du chromosome 14q32 qui comprend l'un des plus importants cluster d'ARNs non codants du génome. Dans cette région, on trouve également le gène codant RTL1 qui est transcrit en orientation antisens par rapport à hsa-miR-136-5p (Benetatos et al., 2013). Kagami et al., (2010), ont montré que l'expression de ce locus, qui est soumis à l'empreinte parentale, est régulée par le statut de méthylation de 2 DMRs (Differentially Methylated Region) : IG-DMR et MEG3-DMR. Cependant, la région qui régule l'expression de hsa-miR-136-5p n'a pas encore été précisément identifiée. Dans le but d'obtenir des informations sur cette région, nous avons extrait de la base de données « TCGA » (The Cancer Genome Atlas), le niveau d'expression de hsa-miR-136-5p et le profil de méthylation global de 167 tumeurs mammaires invasives. Ces données ont été analysées par Pierre Filhine-Trésarrieu (Bioinformaticien) grâce au package « Limma » de R, qui est classiquement utilisé pour l'analyse de données de microarray mais qui peut également permettre d'identifier des régions différenciellement méthylées. Les résultats obtenus n'ont pas permis d'identifier de région dont le statut de méthylation est significativement corrélé à l'expression d'hsa-miR-136-5p. Malgré tout, nous avons émis l'hypothèse que le niveau d'expression de ce micro-ARN pourrait être impacté par un traitement à la DAC et ainsi



**Figure 64 : Effet de la décitabine sur l'expression d'hsa-miR-136-5p.** (Figure 3A de la publication n°3)

Les cellules MCF-7 ont été exposées pendant 48h au DMSO 0,01% ou à la décitabine (DAC) 5  $\mu$ M. Le niveau d'expression d'hsa-miR-136-5p a ensuite été mesuré grâce à la technique « Taqman™ microRNA Assay ». On observe que la DAC induit une augmentation significative de l'expression d'hsa-miR-136-5p.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. N=6. \*\* : P < 0,01. UA = Unité arbitraire.

moduler l'expression d'ER $\alpha$ 36 en réponse à ce traitement. Cette hypothèse a été vérifiée sur les cellules cancéreuses mammaires MCF-7 qui, après traitement à la DAC, présentent une augmentation de l'expression d'hsa-miR-136-5p, associée à une diminution de l'expression d'ER $\alpha$ 36 (Figure 64).

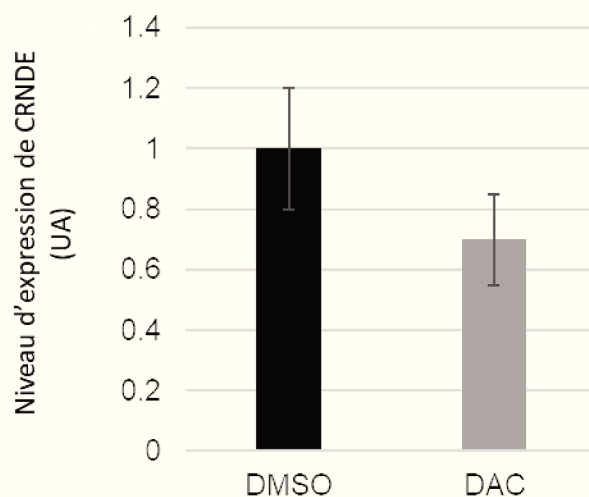
En parallèle, il a été décrit, notamment dans le contexte du cancer colorectal, des gliomes et du cancer du sein, que l'expression du lncRNA CRNDE (ColoRectal Neoplasia Differentially Expressed) est corrélée à la progression de ces tumeurs via la séquestration et donc l'inactivation du suppresseur de tumeur hsa-miR-136-5p (Gao et al., 2017 ; Huan et al., 2017 ; Li D.X. et al., 2017). Ainsi, nous avons voulu vérifier si la DAC pourrait également avoir un effet sur l'expression de CRNDE. Toutefois, cette étude n'a pas permis de mettre en évidence de variation significative de l'expression de ce lncRNA après traitement à la DAC malgré une tendance à la diminution (Figure 65).

#### D) Impact de la régulation de l'expression d'ER $\alpha$ 36 par la DAC sur la résistance au tamoxifène

Le développement de la résistance au tamoxifène dans les cellules cancéreuses mammaires MCF-7 dépend notamment de l'induction de l'expression d'ER $\alpha$ 36 en réponse à ce traitement. Etant donné que la DAC régule négativement l'expression de ce variant dans ces mêmes cellules, nous avons voulu étudier l'impact d'une combinaison de traitement 4-OHT/DAC sur l'expression d'ER $\alpha$ 36 (4-OHT = métabolite actif du tamoxifène). Les résultats obtenus, présentés dans la Figure 66, indiquent que l'addition de DAC bloque l'induction de l'expression d'ER $\alpha$ 36 par le 4-OHT. Cela semble dépendre de l'induction de l'expression d'hsa-miR-136-5p par la DAC. Comme précédemment, lorsque nous avons étudié le rôle potentiel de CRNDE dans ce système, aucune variation significative de l'expression de ce lncRNA n'a été mise en évidence en réponse à le 4-OHT et à la combinaison 4-OHT/DAC.

En conclusion, l'ensemble des résultats obtenus au cours de cette étude a permis de mettre en évidence 2 nouveaux mécanismes de régulation de l'expression d'ER $\alpha$ 36. Nous avons montré que l'expression de ce variant dépend (i) du statut de méthylation de son promoteur, plus précisément de 4 sites CpG localisés en aval du TSS mais aussi (ii) du micro-ARN hsa-miR-136-5p qui régule l'expression d'ER $\alpha$ 36 au niveau post-transcriptionnel.

De plus, ce travail a permis de démontrer un effet direct et indirect de la DAC sur l'expression d'ER $\alpha$ 36. En effet, ce traitement régule négativement l'expression de ce variant, non seulement en déméthylant son promoteur mais aussi en induisant l'expression du micro-ARN hsa-miR-

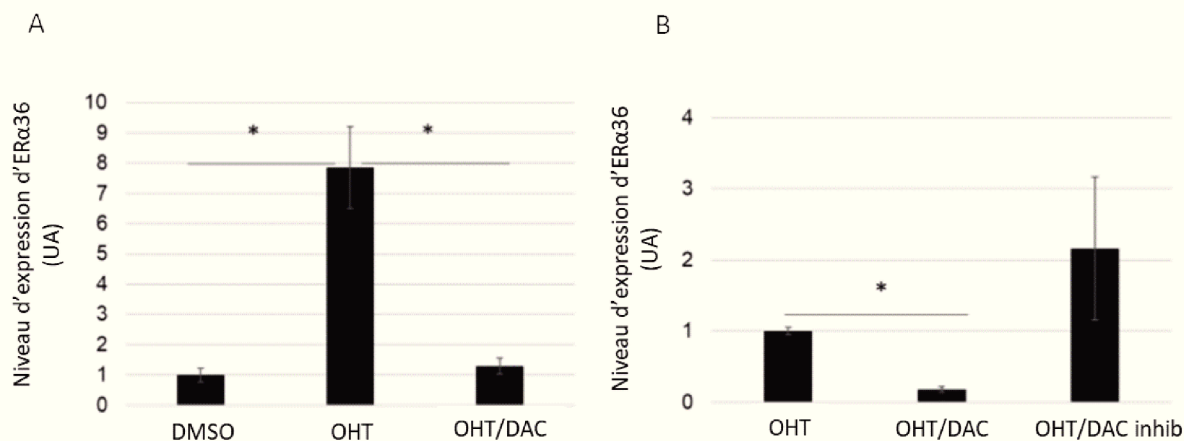


**Figure 65 : Effet de la décitabine sur l'expression du long ARN non codant CRNDE.**  
(Figure S2 de la publication n°3)

Les cellules MCF-7 ont été exposées pendant 48h au DMSO 0,01% ou à la décitabine (DAC) 5  $\mu$ M. Le niveau d'expression de CRNDE a ensuite été mesuré par RT-qPCR. On observe que la DAC n'impacte pas de manière significative l'expression de CRNDE.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. N=3. UA = Unité arbitraire.



**Figure 66 : Impact de la décitabine sur la résistance au tamoxifène dépendante de l'expression d'ERα36.** (Figure 4 de la publication n°3)

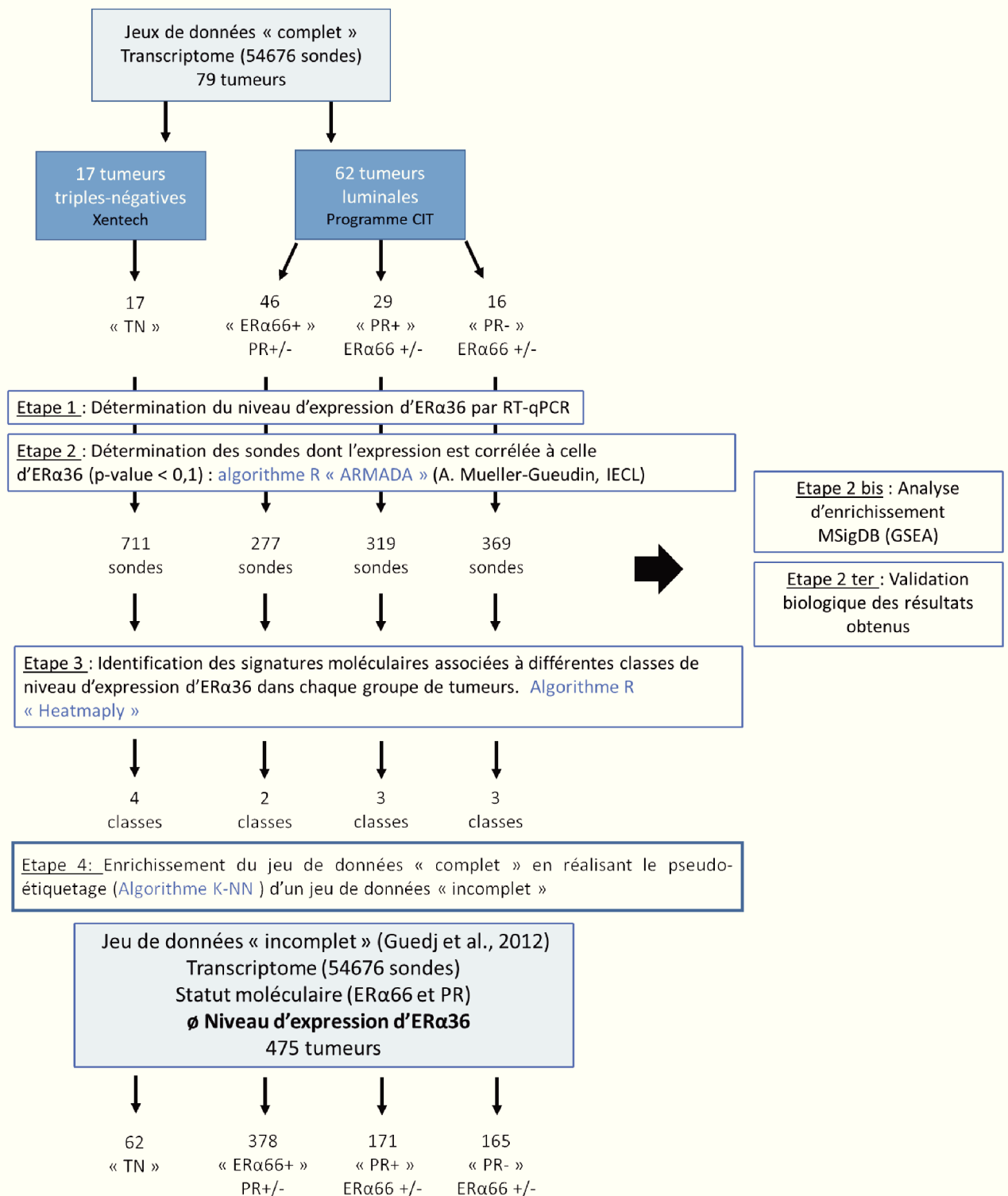
(A) Impact de la décitabine (DAC) sur l'induction de l'expression d'ERα36 par le 4- hydroxy-tamoxifène (OHT). Les cellules MCF-7 ont été exposées, pendant 48 heures, au DMSO 0,01%, à l'OHT 1 μM ou à la combinaison OHT 1 μM/ DAC 5 μM. Le niveau d'expression d'ERα36 a été mesuré par RT-qPCR. On observe que la décitabine réprime l'induction de l'expression d'ERα36 induite par l'OHT.

(B) Afin de vérifier si cet effet de la décitabine dépend de hsa-miR-136-5p, le co-traitement, OHT/DAC, a été réalisé sur des cellules MCF-7 transfectées avec l'oligonucléotide « hsa-miR-136-5p inhibiteur ». On observe que lorsque l'on bloque l'induction de l'expression d'hsa-miR-136-5p, dépendante de la DAC, on restaure l'effet inducteur de l'OHT sur l'expression d'ERα36.

Le DMSO 0,01% qui correspond au solvant, sert de contrôle.

Les barres d'erreur correspondent à l'erreur standard à la moyenne. N=3. \* : P < 0,05. UA = Unité arbitraire.





**Figure 67 : Schéma récapitulatif de la stratégie développée pour identifier de nouveaux facteurs impliqués dans la signalisation dépendante d'ERα36.**

136-5p. Cette molécule, qui a déjà été approuvée par la FDA pour le traitement des syndromes myélodysplasiques, pourrait être utilisée en combinaison avec le 4-OHT pour aider à prévenir l'apparition de résistances à ce traitement.

E) Développement d'une approche bioinformatique permettant d'identifier de nouveaux facteurs impliqués dans la signalisation dépendante d'ER $\alpha$ 36

En parallèle, nous avons développé une approche non supervisée basée sur l'utilisation d'outils bioinformatiques et statistiques afin d'identifier de nouvelles variables explicatives, c'est-à-dire de nouveaux régulateurs de la signalisation dépendante d'ER $\alpha$ 36. Nous avons également voulu vérifier grâce à cette étude si la signalisation dépendante d'ER $\alpha$ 36 pouvait être fonction de celle modulée par les récepteurs ER $\alpha$ 66 et PR, qui sont utilisés pour la classification moléculaire des tumeurs mammaires. L'enchaînement des différentes étapes de cette approche est représenté dans la Figure 67.

Au départ, nous avons accès à 2 jeux de données correspondant au transcriptome (i) de 17 tumeurs mammaires PDX triples négatives (TN) issues de la base de données Xentech (Reyal et al., 2012), et (ii) de 62 tumeurs de types luminales fournies par Aurélien De Reynies et Jacqueline Métral qui sont les coordinateurs du programme « Carte d'identité des tumeurs » (CIT) mené par la ligue contre le cancer. Afin d'affiner notre analyse, le second jeu de données correspondant aux tumeurs du sous-type « Luminal » a été subdivisé en 3 groupes : « **ER $\alpha$ 66+** ; PR+/- » (46 tumeurs), « ER $\alpha$ 66 +/- ; **PR+** » (29 tumeurs) et « ER $\alpha$ 66 +/- ; **PR-** » (16 tumeurs). Au total, nous avons donc analysé et comparé 4 jeux de données : « **TN** », « **ER $\alpha$ 66+** », « **PR+** », « **PR-** ».

➤ Etape 1 :

Etant donné qu'aucune sonde présente sur les puces à ADN ne permet de détecter spécifiquement ER $\alpha$ 36, nous avons mesuré au laboratoire le niveau d'expression de ce variant en réalisant des RT-qPCR sur les 17 échantillons de tumeurs PDX et sur les ADNc correspondants aux 62 tumeurs de type « Luminal », fournis avec les données transcriptomiques (Tableau IX).

➤ Etape 2 :

A partir de ces données, notre challenge d'un point de vue statistique a été de pouvoir expliquer les variations d'expression d'ER $\alpha$ 36 obtenues dans un faible nombre de tumeurs mais à partir

**Tableau IX : Caractéristiques des 62 tumeurs du jeu de données « Luminal ».**

N° identification de la tumeur	Statut ER $\alpha$ 66 et PR	Age de la patiente	Expression d'ER $\alpha$ 36 (UA)
STAM_RL_2	ER+ / PR+	65	1.70
STAM_RL_3	ER+ / PR+	60	0.02
STAM_RL_4	ER+ / PR?	65	1.92
STAM_RL_5	ER+ / PR-	76	11.71
STAM_RL_6	ER+ / PR+	78	0.04
STAM_RL_7	ER+ / PR+	67	2.46
STAM_RL_9	ER+ / PR+	72	0.81
STAM_RL_10	ER+ / PR-	73	15.78
STAM_RL_12	ER+ / PR-	57	0.80
STAM_RL_13	ER+ / PR+	78	16.22
STAM_RL_14	ER+ / PR?	69	1.40
STAM_RL_15	ER+ / PR+	75	10.48
STAM_RL_17	ER+ / PR?	76	16.22
STAM_RL_20	ER? / PR+	58	7.83
STAM_RL_21	ER+ / PR+	72	0.28
STAM_RL_24	ER+ / PR-	55	1.41
STAM_RL_26	ER+ / PR+	68	20.39
STAM_RL_27	ER? / PR-	69	12.73
STAM_RL_28	ER+ / PR+	70	4.17
STAM_RL_29	ER+ / PR+	62	3.86
STAM_RL_31	ER? / PR-	64	6.32
STAM_RL_32	ER+ / PR-	53	6.19
STAM_RL_33	ER+ / PR-	64	2.04
STAM_RL_34	ER+ / PR+	74	20.39
STAM_RL_36	ER+ / PR-	53	43.40
STAM_RL_37	ER+ / PR-	75	8.57
STAM_RL_39	ER? / PR+	57	15.67
STAM_RL_40	ER+ / PR-	70	20.11
STAM_RL_41	ER+ / PR+	78	23.58
STAM_RL_42	ER+ / PR-	79	0.41
STAM_RL_45	ER+ / PR+	61	2.35
STAM_RL_46	ER+ / PR+	71	3.68
STAM_RL_47	ER? / PR+	70	9.25
STAM_RL_49	ER+ / PR?	65	0.53
STAM_RL_50	ER? / PR-	70	19.69
STAM_RL_51	ER+ / PR?	71	1.04
STAM_RL_52	ER+ / PR+	75	25.99
STAM_RL_53	ER+ / PR+	72	7.78
STAM_RL_56	ER+ / PR-	53	1.88
STAM_RL_57	ER+ / PR+	46	6.06
STAM_RL_58	ER? / PR+	68	4.56
STAM_RL_59	ER+ / PR+	73	6.63
STAM_RL_60	ER+ / PR?	58	15.03
STAM_RL_61	ER+ / PR-	61	18.89
STAM_RL_62	ER+ / PR?	70	0.14
STAM_RL_64	ER+ / PR?	53	2.25
STAM_RL_66	ER+ / PR+	62	17.38
STAM_RL_69	ER+ / PR+	72	3.73
STAM_RL_70	ER+ / PR-	61	3.27
STAM_RL_71	ER+ / PR+	73	11.39
STAM_RL_72	ER+ / PR+	70	0.35
STAM_RL_73	ER+ / PR+	65	8.57
STAM_RL_74	ER+ / PR+	60	0.71

Pour chacune des tumeurs, désignées par un identifiant de type « STAM-RL-X », le statut ER $\alpha$ 66 (ER) et PR, ainsi que l'âge de la patiente sont indiqués. Concernant le niveau d'expression d'ER $\alpha$ 36, il a été mesuré au laboratoire par RT-qPCR.

Le transcriptome, les données cliniques et les ADNc correspondants aux tumeurs du jeu de données « Luminal » ont été fournis par Aurélien De Reynies et Jacqueline Métral qui sont les coordinateurs du programme « Carte d'identité des tumeurs » (CIT).

Les informations concernant les 17 tumeurs du jeu de données « Triple négatif », ont été fournies par la société Xentech. Cependant, elles ne sont pas présentées dans ce manuscrit car elle sont soumises à un accord de transfert de matériel (MTA CXT-295). UA = Unité arbitraire

**Tableau X : Résultats de l'analyse d'enrichissement des données.**

Jeux de données	TN	ER $\alpha$ 66 +
Facteurs de transcription	SP1, LEF1, E12, NFAT, FOXO4, MAZ, AP4, PAX4	LEF1, VDR, SP1, FOXO4, NFAT, NFY, ELK1, AREB6
Micro-ARNs	hsa-miR-19A et B hsa-miR-20A et B hsa-miR-21 hsa-miR-29A,B et C hsa-miR-106A et B hsa-miR-124A hsa-miR-130A et B hsa-miR-181A,B,C et D hsa-miR-519A, B et C	hsa-miR-331
Processus Biologiques	Regulation_of_hydrolase_activity Positive_regulation_of_catalytic_activity Positive_regulation_of_biosynthetic_process Positive_regulation_of_gene_expression	Phosphate_containing_compound_metabolic_process Organonitrogen_compound_metabolic_process Golgi_apparatus Establishment_of_localization_in_cell

Jeux de données	PR+	PR-
Facteurs de transcription	MAZ, LEF1, E12, SP1, ERR1, FREAC2, ETS2, NFAT	LEF-1, NFAT, FOXO4, MYOD, SP1, SOX5, PAX4, ELK1
Micro-ARNs	hsa-miR-15A et B hsa-miR-16 hsa-miR-18A et B hsa-miR-34B hsa-miR-124A hsa-miR-132 hsa-miR-153 hsa-miR-195 hsa-miR-205 hsa-miR-212	hsa-miR-34A, B et C hsa-miR-103 hsa-miR-135A et B hsa-miR-141 hsa-miR-181A, B, C et D hsa-miR-186 hsa-miR-200A, B et C hsa-miR-493 hsa-miR-527
Processus Biologiques	Cytosolic_ribosome Structural_constituent_of_ribosome Translational_initiation Establishment_of_protein_localization_to_endoplasmic_reticulum	Regulation_of_response_to_stress Cell_projection Cytoskeleton Anatomical_structure_formation_involved_in_morphogenesis

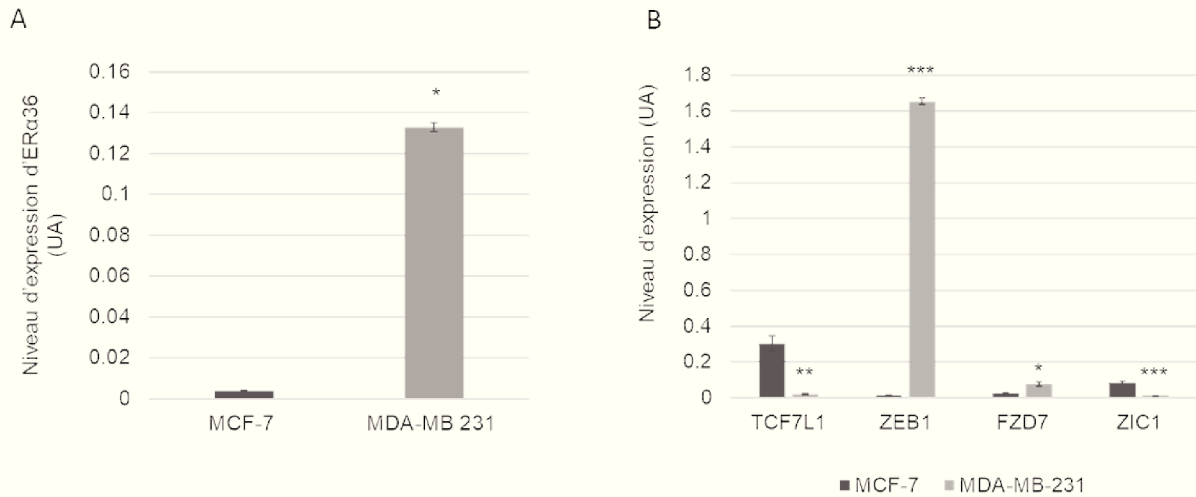
Ce tableau regroupe l'ensemble des facteurs de transcription et des micro-ARNs identifiés, grâce à la base de donnée MSigDB, comme étant les principaux régulateurs potentiels des listes de gènes, dont l'expression est corrélée à celle d'ER $\alpha$ 36, dans les jeux de données : « TN », « ER $\alpha$ 66+ », « PR+ » et « PR- ». Les processus biologiques associés à chacune de ces listes de gènes sont également répertoriés.

**TN** = Triple-négatif, **PR** = Récepteur à la progestérone.

d'un grand nombre de variables explicatives potentielles correspondant aux 54 676 sondes transcriptomiques. Pour répondre à cette problématique, nous avons adapté un algorithme R appelé « ARMADA », développé par Aurélie Mueller-Gueudin (IECL : Institut Élie Cartan de Lorraine), qui correspond à une association de plusieurs tests statistiques permettant de sélectionner, parmi les 54 676 sondes initiales, quelques centaines de sondes dont l'expression est supposée être corrélée à celle d'ER $\alpha$ 36 dans nos différents jeux de données. Plus précisément, un pré filtrage est réalisé en calculant des coefficients de corrélations de Pearson afin de sélectionner les variables corrélées à ER $\alpha$ 36 avec une p-value < 0,1. Ces corrélations sont ensuite affinées en réalisant 8 tests statistiques successifs. Ainsi, nous avons obtenus une liste de 711 sondes corrélées à l'expression d'ER $\alpha$ 36 dans le jeu « TN », une liste de 277 sondes dans le jeu « ER $\alpha$ 66+ », une liste de 319 sondes dans le jeu « PR+ » et une liste de 369 sondes dans le jeu « PR- ».

➤ Etape 2 bis :

A partir des listes de gènes correspondant à ces différentes sondes transcriptomiques nous avons réalisé une analyse d'enrichissement grâce à la base de données MSigDB (GSEA). Plus précisément, nous avons recherché les facteurs de transcription et les micro-ARNs qui régulent la majorité des gènes contenus dans les différentes listes (TN, ER $\alpha$ 66+, PR+ et PR-) (Tableau X). Les résultats préliminaires obtenus indiquent que 4 facteurs de transcription : NFAT1, FOXO4, SP1 et LEF-1, sont des régulateurs communs de l'ensemble des listes et pourraient donc être des médiateurs de l'effet d'ER $\alpha$ 36 dans les tumeurs mammaires. De manière intéressante, une étude de Antanavičiūtė et al., (2017) a montré que les facteurs de transcription FOXO4, SP1 et LEF-1 sont des hallmarks transcriptionnels caractéristiques des cellules cancéreuses qui régulent positivement l'expression de gènes dont la combinaison induit la transition épithélio-mésenchymateuse et l'acquisition de propriétés de cellules souches, de façon dépendante de l'activation de la voie Wnt/ $\beta$ -caténine. Bien que ces facteurs soient communs à l'ensemble des groupes de tumeurs, lorsque l'on analyse les termes GO BP, on observe que ces facteurs semblent impliqués dans la régulation de processus biologiques différents en fonction des groupes. En effet, pour les tumeurs « TN » et « ER $\alpha$ 66+ », les gènes corrélés à l'expression d'ER $\alpha$ 36 semblent principalement associés à des processus métaboliques tandis que dans les tumeurs « PR+ » et « PR- » on observe plutôt un lien avec le fonctionnement de machinerie traductionnelle et la morphologie cellulaire, respectivement. Ces termes restent très génériques et il serait nécessaire d'augmenter le nombre de données



**Figure 68 : Validation biologique des résultats obtenus grâce l'algorithme R « ARMADA » .**

- (A) Le niveau d'expression d'ERα36 a été mesuré par RT-qPCR dans les cellules MCF-7 et MDA-MB-231.
- (B) Le niveau d'expression des gènes TCF7L1, ZEB1, FZD7 et ZIC1 a été mesuré par RT-qPCR dans les cellules MCF-7 et MDA-MB-231. Ces gènes ont été étudiés car leur expression apparait corrélée à celle d'ERα36, dans tous les types de tumeur, après application de l'algorithme « ARMADA ».

Les barres d'erreur correspondent à l'erreur standard à la moyenne. N=3. \* : P < 0,05; \*\* : P < 0,01; \*\*\* : P < 0,001. UA = unité arbitraire

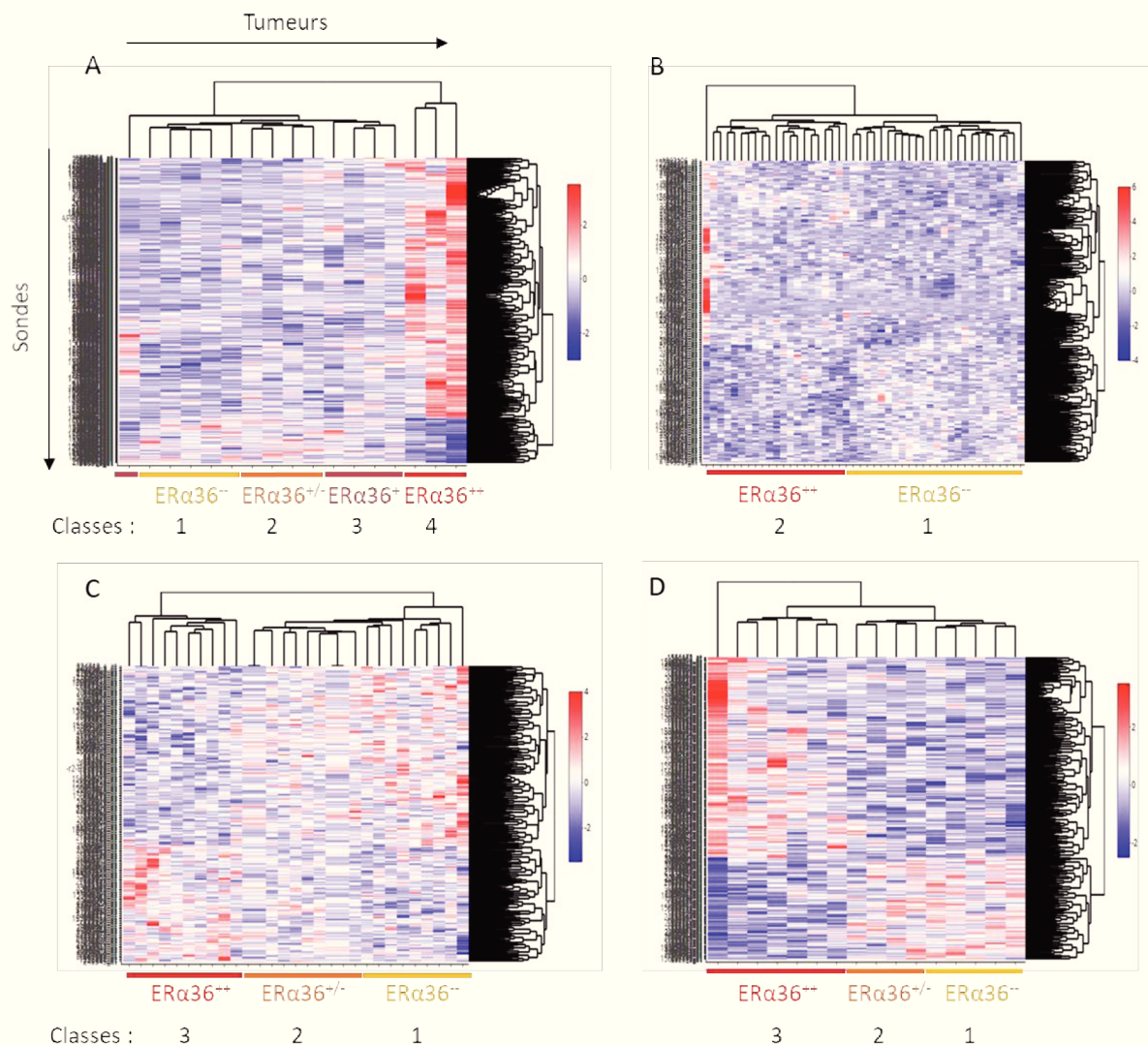
transcriptomiques disponibles afin de préciser le lien entre ER $\alpha$ 36 et ces différents processus biologiques.

Concernant l'analyse des micro-ARNs, les résultats obtenus et présentés dans le Tableau X indiquent que la majorité des gènes dont l'expression est corrélée à ER $\alpha$ 36 dans le jeu « **TN** » sont notamment régulés par les micro-ARNs : hsa-miR-106B, hsa-miR-21 et hsa-miR-29A. Ceux-ci sont répertoriés dans la revue de Le Quesne et Caldas, (2010), comme des micro-ARNs oncogéniques impliqués dans les processus métastatiques, la survie et l'autosuffisance en facteurs de croissance des tumeurs mammaires. De manière intéressante, ces données rappellent les résultats d'une étude antérieure menée au laboratoire par Clémence Chamard-Jovenin, qui a montré qu'une forte expression d'ER $\alpha$ 36 dans les tumeurs mammaires est associée à une augmentation du potentiel métastatique et à une croissance tumorale indépendante des œstrogènes (Chamard-Jovenin et al., 2015). A l'inverse, dans les tumeurs du jeu « **PR -** », les gènes corrélés à l'expression d'ER $\alpha$ 36 sont plutôt associés à des micro-ARNs répertoriés comme suppresseurs de tumeurs dans les tumeurs mammaires : hsa-miR-34 et hsa-miR-200. Ces résultats indiquent que le pronostic associé à l'expression d'ER $\alpha$ 36 pourrait varier en fonction des sous-types moléculaires et notamment en fonction du statut ER $\alpha$ 66 et PR de la tumeur. De plus, il semble qu'il existe une composante post transcriptionnelle majeure dans les réseaux biologiques impliquant ER $\alpha$ 36.

➤ Etape 2 ter :

Par la suite, nous avons voulu apporter une première confirmation de la validité biologique des résultats obtenus. Pour cela, la corrélation entre l'expression d'ER $\alpha$ 36 et celle des gènes ZEB1, FZD7, ZIC1 et TCF7LD, qui a été prédite par l'algorithme dans toutes les catégories de tumeurs, a été vérifiée *in vitro* par RT-qPCR dans les lignées de cellules cancéreuses mammaires MCF-7 (ER $\alpha$ 66+ ; PR+/-) et MDA-MB-231 (TN). Les résultats présentés dans la Figure 68 confirment l'existence d'une corrélation (positive ou négative) entre l'expression d'ER $\alpha$ 36 et celle des gènes identifiés grâce à l'algorithme R « ARMADA » dans les 2 lignées.

Nous avons également mesuré l'expression des 4 régulateurs amonts de ces gènes, à savoir NFAT1, FOXO4, SP1 et LEF-1, par RT-qPCR et western blot. Nous avons observé que le niveau d'expression de NFAT1, SP1 et LEF-1 varie peu entre les 2 lignées qui expriment différemment ER $\alpha$ 36. Concernant FOXO4 il n'est pas détecté quelle que soit la technique utilisée. De plus, les lignées MCF-7 et MDA-MB-231 ont été transfectées de façon transitoire avec un vecteur de surexpression d'ER $\alpha$ 36 (pcDNA3.1-ER36, voir publications N°1 et N°2).



**Figure 69 : Détermination des différentes classes d'expression d'ERα36 dans chaque jeu de données.**

Ces heatmaps représentent le niveau d'expression des différentes sondes, corrélées à ERα36, dans chacune des tumeurs des 4 jeux de données. Les différentes classes d'expression d'ERα36 sont déterminées à partir du dendrogramme situé au dessus des heatmaps. Pour chacune de ces classes, numérotées de 1 à 4, une signature moléculaire a été déterminée.

- (A) 17 tumeurs « Triples négatives », 711 sondes : 4 signatures moléculaires corrélées à ERα36
- (B) 46 tumeurs « ERα66+, PR+/- », 277 sondes : 2 signatures moléculaires corrélées à ERα36
- (C) 29 tumeurs « PR+, ERα66 +/- », 319 sondes : 3 signatures moléculaires corrélées à ERα36
- (D) 16 tumeurs « PR-, ERα66 +/- », 369 sondes : 3 signatures moléculaires corrélées à ERα36

Rouge : forte expression vs Bleu : faible expression



Cela a conduit à augmenter artificiellement l'expression de ce variant d'un facteur 6 et 16, respectivement. Cependant, après surexpression d'ER $\alpha$ 36, nous n'avons toujours pas mis en évidence de variation significative de l'expression des gènes NFAT1, FOXO4, SP1 et LEF-1. Cela indique qu'ils ne constituent pas des cibles directes d'ER $\alpha$ 36. Dans la continuité, nous pourrions invalider ou surexprimer ces différents facteurs pour savoir si, en retour, ER $\alpha$ 36 pourrait constituer une de leur cible. De plus, étant donnée qu'une interaction physique entre la forme canonique ER $\alpha$ 66 et le facteur Sp1 a déjà été décrite dans le contexte des voies génomiques ERE-indépendantes (O'Lone et al., 2004), il serait intéressant d'étudier l'interaction potentielle d'ER $\alpha$ 36 avec ces facteurs de transcription en réalisant une co-immunoprécipitation. Cela permettrait de voir si ER $\alpha$ 36 joue un rôle de corégulateur des fonctions cellulaires associées à ces facteurs de transcription.

➤ Etape 3 :

Dans la continuité, afin de renforcer la puissance statistique de notre étude, il serait intéressant de pouvoir augmenter le nombre de données transcriptomiques disponibles en enrichissant des jeux de données dit « incomplets » pour lesquels nous connaissons le statut moléculaire (ER $\alpha$ 66+/- ; PR+/- ; TN), le niveau d'expression des 54 676 sondes du transcriptome mais pas le niveau d'expression d'ER $\alpha$ 36. En effet, de nombreuses données transcriptomiques sont disponibles dans les bases de données en ligne mais dans la majorité des cas il est impossible d'avoir accès aux échantillons biologiques correspondants pour mesurer l'expression d'ER $\alpha$ 36.

Afin de pouvoir compiler et exploiter ces données nous avons commencé à développer en collaboration avec Marianne Clausel (IECL), une méthode d'apprentissage semi-supervisé pour la complétion de données biologiques. La première étape consiste à définir, à partir de nos jeux de données « TN », « ER $\alpha$ 66+ », « PR+ » et « PR- », des signatures moléculaires caractéristiques du niveau d'expression d'ER $\alpha$ 36 dans ces différents types de tumeurs. Ces signatures ont été identifiées grâce à l'algorithme « Heatmaply » de R qui permet d'obtenir pour chacun des 4 jeux de données, un heatmap représentatif du niveau d'expression des gènes corrélés à ER $\alpha$ 36 dans chacune des tumeurs. Grâce au dendrogramme associé, différentes classes de tumeurs ont été définies et sont chacune caractérisées par un niveau d'expression d'ER $\alpha$ 36 et une signature moléculaire (Figure 69, Tableau XI).

**Tableau XI : Répartition des tumeurs du jeu de données « Luminal » dans les différentes classes d'expression d'ER $\alpha$ 36.**

Identification Tumeur	Statut ER $\alpha$ 66 et PR	Age de la patiente	Expression d'ER $\alpha$ 36 (UA)	Classes d'expression d'ER $\alpha$ 36		Classes d'expression d'ER $\alpha$ 36 Gjeu ER $\alpha$ 66+
				Jeu PR+	Jeu PR-	
STAM_RL_2	ER+ / PR+	65	1.70	1		1
STAM_RL_3	ER+ / PR+	60	0.02	1		1
STAM_RL_4	ER+ / PR?	65	1.92			1
STAM_RL_5	ER+ / PR-	76	11.71		3	2
STAM_RL_6	ER+ / PR+	78	0.04	2		2
STAM_RL_7	ER+ / PR+	67	2.46	1		1
STAM_RL_9	ER+ / PR+	72	0.81	2		1
STAM_RL_10	ER+ / PR-	73	15.78		3	2
STAM_RL_12	ER+ / PR-	57	0.80		1	1
STAM_RL_13	ER+ / PR+	78	16.22	3		2
STAM_RL_14	ER+ / PR?	69	1.40			2
STAM_RL_15	ER+ / PR+	75	10.48	3		1
STAM_RL_17	ER+ / PR?	76	16.22			2
STAM_RL_20	ER? / PR+	58	7.83	3		
STAM_RL_21	ER+ / PR+	72	0.28	1		1
STAM_RL_24	ER+ / PR-	55	1.41		1	1
STAM_RL_26	ER+ / PR+	68	20.39	3		2
STAM_RL_27	ER? / PR-	69	12.73		3	
STAM_RL_28	ER+ / PR+	70	4.17	2		2
STAM_RL_29	ER+ / PR+	62	3.86	1		1
STAM_RL_31	ER? / PR-	64	6.32		2	
STAM_RL_32	ER+ / PR-	53	6.19		2	1
STAM_RL_33	ER+ / PR-	64	2.04		1	1
STAM_RL_34	ER+ / PR+	74	20.39	3		2
STAM_RL_36	ER+ / PR-	53	43.40		3	2
STAM_RL_37	ER+ / PR-	75	8.57		2	2
STAM_RL_39	ER? / PR+	57	15.67	3		
STAM_RL_40	ER+ / PR-	70	20.11		3	2
STAM_RL_41	ER+ / PR+	78	23.58	3		2
STAM_RL_42	ER+ / PR-	79	0.41		2	1
STAM_RL_45	ER+ / PR+	61	2.35	1		1
STAM_RL_46	ER+ / PR+	71	3.68	2		1
STAM_RL_47	ER? / PR+	70	9.25	2		
STAM_RL_49	ER+ / PR?	65	0.53			1
STAM_RL_50	ER? / PR-	70	19.69		3	
STAM_RL_51	ER+ / PR?	71	1.04			1
STAM_RL_52	ER+ / PR+	75	25.99	3		2
STAM_RL_53	ER+ / PR+	72	7.78	2		1
STAM_RL_56	ER+ / PR-	53	1.88		1	1
STAM_RL_57	ER+ / PR+	46	6.06	2		2
STAM_RL_58	ER? / PR+	68	4.56	2		
STAM_RL_59	ER+ / PR+	73	6.63	2		2
STAM_RL_60	ER+ / PR?	58	15.03			2
STAM_RL_61	ER+ / PR-	61	18.89		3	2
STAM_RL_62	ER+ / PR?	70	0.14			2
STAM_RL_64	ER+ / PR?	53	2.25			1
STAM_RL_66	ER+ / PR+	62	17.38	3		1
STAM_RL_69	ER+ / PR+	72	3.73	2		2
STAM_RL_70	ER+ / PR-	61	3.27		1	1
STAM_RL_71	ER+ / PR+	73	11.39	3		2
STAM_RL_72	ER+ / PR+	70	0.35	1		1
STAM_RL_73	ER+ / PR+	65	8.57	1		1
STAM_RL_74	ER+ / PR+	60	0.71	1		1

A partir des heatmaps présentés dans la Figure 71, différentes classes d'expression d'ER $\alpha$ 36 ont été définies pour chacun des jeux de données : « TN » (4 classes), « ER $\alpha$ 66+ » (2 classes), « PR+ » (3 classes) et « PR- » (3 classes). La répartition des 62 tumeurs du jeu de données « Luminal » dans ces différentes classes est indiquée dans le tableau. Les cases grises indiquent que la tumeur n'entre pas dans la classification car elle ne fait pas partie du jeu de données associé à la colonne (« ER $\alpha$ 66+ », « PR+ » ou « PR- »).

Les informations concernant le jeu de données « TN » étant soumises à un accord de transfert de matériel (MTA CXT-295), elles ne sont pas présentées ici.

## ➤ Etape 4 :

Grâce à l'utilisation d'un algorithme K-NN (Méthode des plus proches voisins) nous pouvons ensuite réaliser un pseudo-étiquetage, c'est-à-dire attribué une valeur de classe d'expression d'ER $\alpha$ 36 à chacune des tumeurs d'un jeu de données « incomplet ». Plus précisément, cet algorithme permet de comparer le profil d'expression génique de chacune des tumeurs à étiqueter aux signatures moléculaires associées à l'expression d'ER $\alpha$ 36 précédemment identifiées grâce au jeu de données « complet ».

Afin de tester la pertinence de cette méthode, nous l'avons appliqué à un jeu de données « incomplet » de 475 tumeurs, provenant de l'étude de Guedj et al., (2012), pour lesquelles nous connaissions le transcriptome, le statut moléculaire (ER $\alpha$ 66 et PR+) mais pas le niveau d'expression d'ER $\alpha$ 36. Après avoir attribué une valeur de classe d'expression d'ER $\alpha$ 36 à chacune des tumeurs utilisées pour cette analyse transcriptomique, nous disposons d'un jeu de données complété de 62 tumeurs « **TN** », 378 « **ER $\alpha$ 66+** », 171 « **PR+** » et 165 « **PR -** ».

## ➤ Etape 5 et perspectives :

Le jeu de données complet ayant été enrichi, la suite du travail consiste à :

- Renouveler l'identification de signatures moléculaires associées au niveau d'expression d'ER $\alpha$ 36 dans chaque groupe de tumeurs afin d'obtenir des profils statistiquement plus robustes.
- Déterminer quels sont les gènes clés au sein de ces signatures afin de raffiner la classification actuelle des sous types moléculaires de tumeurs mammaires en y incluant une composante associée à l'expression d'ER $\alpha$ 36.
- Etablir des corrélations potentielles entre ces signatures et les paramètres cliniques associés aux différentes classes de tumeurs afin de proposer le niveau d'expression d'ER $\alpha$ 36 comme marqueur prédictif de la survie, de la progression tumorale, de la réponse au traitement...

Du point de vue fondamental, nous aurons également progresser dans la compréhension, aux niveaux moléculaires et cellulaires, du rôle d'ER $\alpha$ 36 dans la physiopathologie de la glande mammaire.

## *Publication N°3*

“Dual epigenetic regulation of ER $\alpha$ 36 expression in breast cancer cells”

**C. Thiebaut**, A. Chesnel, J.L. Merlin, M. Chesnel, A. Leroux, A. Harlé, H. Dumond.

Article publié dans International Journal of Molecular Sciences en Mai 2019.

Le cancer du sein constitue la principale cause de mortalité par cancer chez les femmes. Parmi les différents sous-types moléculaires décrits, les tumeurs lumineales sont considérées comme de bon pronostic et sont susceptibles d’être traitées par hormonothérapie. Cependant, dans environ 30% des cas, des résistances intrinsèques ou acquises à ce type de traitement sont observées. Récemment, il a été démontré que le variant ER $\alpha$ 36 pourrait jouer un rôle crucial dans ce processus. En effet, une forte expression de cette isoforme d’ER $\alpha$  est associée à l’apparition d’un phénotype pré-néoplasique dans les cellules épithéliales mammaires saines, ainsi qu’à la progression de la maladie et à une résistance accrue aux agents thérapeutiques dans les tumeurs mammaires. Ainsi, la compréhension des mécanismes de régulation classiques et épigénétiques de l’expression d’ER $\alpha$ 36 pourrait permettre de mieux appréhender l’implication de ce variant lors de l’initiation et/ou de la progression des tumeurs.

Dans cette étude nous avons identifié 2 mécanismes qui pourraient contribuer à la régulation de l’expression d’ER $\alpha$ 36. Nous nous sommes tout d’abord concentré sur hsa-miR-136-5p, un micro-ARN ciblant directement la région 3’UTR de l’ARNm d’ER $\alpha$ 36, dont l’expression est inversement corrélée à celle de ce variant dans les cellules cancéreuses mammaires. Nous avons montré que l’augmentation artificielle du taux d’hsa-miR-136-5p dans les cellules cancéreuses mammaires MCF-7, grâce à la transfection d’un oligonucléotide « hsa-miR-136-5p mimic », conduit à une diminution de l’expression d’ER $\alpha$ 36.

En parallèle, nous avons également observé que le statut de méthylation de la région promotrice d'ER $\alpha$ 36 influence directement son niveau d'expression. En effet, la déméthylation de 4 sites CpG localisés en aval du site d'initiation de la transcription d'ER $\alpha$ 36, est corrélée à une diminution significative de l'expression de ce variant.

De manière intéressante nous avons mis en évidence qu'un traitement avec de la décitabine, qui est un agent déméthylant utilisé dans le traitement des syndromes myélodysplasiques, conduit à une diminution de l'expression d'ER $\alpha$ 36. En effet, ce traitement régule négativement l'expression de ce variant en déméthylant sa région promotrice mais aussi en induisant l'expression du micro-ARN hsa-miR-136-5p. Cette molécule, pourrait donc être utilisée en combinaison avec le 4-OHT pour aider à prévenir l'apparition des résistances à ce traitement qui semblent dépendre de l'induction de l'expression d'ER $\alpha$ 36 dans les cellules cancéreuses mammaires.



Article

# Dual Epigenetic Regulation of ER $\alpha$ 36 Expression in Breast Cancer Cells

Charlène Thiebaut<sup>1</sup>, Amand Chesnel<sup>1</sup>, Jean-Louis Merlin<sup>2</sup>, Maelle Chesnel<sup>1</sup>, Agnès Leroux<sup>3</sup>, Alexandre Harlé<sup>2</sup>  and Hélène Dumond<sup>1,\*</sup> 

<sup>1</sup> Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France; charlene.thiebaut@univ-lorraine.fr (C.T.); amand.chesnel@univ-lorraine.fr (A.C.); maelle.chesnel22@gmail.com (M.C.)

<sup>2</sup> Université de Lorraine, CNRS, CRAN, Institut de Cancérologie de Lorraine, F-54000 Nancy, France; jl.merlin@nancy.unicancer.fr (J.-L.M.); alexandre.harle@crick.ac.uk (A.H.)

<sup>3</sup> Institut de Cancérologie de Lorraine, F-54000 Nancy, France; a.leroux@nancy.unicancer.fr

\* Correspondence: helene.dumond@univ-lorraine.fr

Received: 21 March 2019; Accepted: 27 May 2019; Published: 29 May 2019



**Abstract:** Breast cancer remains the major cause of cancer-induced morbidity and mortality in women. Among the different molecular subtypes, luminal tumors yet considered of good prognosis often develop acquired resistance to endocrine therapy. Recently, misregulation of ER $\alpha$ 36 was reported to play a crucial role in this process. High expression of this ER $\alpha$  isoform was associated to preneoplastic phenotype in mammary epithelial cells, disease progression, and enhanced resistance to therapeutic agents in breast tumors. In this study, we identified two mechanisms that could together contribute to ER $\alpha$ 36 expression regulation. We first focused on hsa-miR-136-5p, an ER $\alpha$ 36 3'UTR-targeting microRNA, the expression of which inversely correlated to the ER $\alpha$ 36 one in breast cancer cells. Transfection of hsa-miR136-5p mimic in MCF-7 cells resulted in downregulation of ER $\alpha$ 36. Moreover, the demethylating agent decitabine was able to stimulate hsa-miR-136-5p endogenous expression, thus indirectly decreasing ER $\alpha$ 36 expression and counteracting tamoxifen-dependent stimulation. The methylation status of ER $\alpha$ 36 promoter also directly modulated its expression level, as demonstrated after decitabine treatment of breast cancer cell and confirmed in a set of tumor samples. Taken together, these results open the way to a direct and an indirect ER $\alpha$ 36 epigenetic modulation by decitabine as a promising clinical strategy to counteract acquired resistance to treatment and prevent relapse.

**Keywords:** breast cancer; ER $\alpha$ 36; microRNA; methylation; endocrine therapy resistance

## 1. Introduction

Despite considerable improvement of therapeutic strategies, breast cancer remains a leading cause of cancer mortality in women, mainly due to acquired resistance to treatment. Indeed, 30–50% of luminal tumors, yet considered as highly curable, develop endocrine therapy resistance through unclear and diverse mechanisms, including alterations of estrogen receptor (ER) gene sequence or epigenetic marks, alternative signaling pathway usage or variation in the pattern of ER isoform [1]. Recently, Wang and colleagues argued that one of the key actors of such acquired resistance could be the ER variant ER $\alpha$ 36, the expression and/or activity of which become stimulated under estrogen deprivation or low dose tamoxifen exposure [2–5].

ER $\alpha$ 36 is a 36kDa receptor encoded by the ESR1 locus and transcribed from an alternative promoter located into the first intron [6]. Compared to the canonical ER $\alpha$ 66 protein, alternative splicing results in a shortened receptor lacking both AF-1 and AF-2 transactivation domains as well as the C-terminal part of the ligand-binding domain. ER $\alpha$ 36 retains dimerization and DNA-binding domains and contains a unique 27 amino acid C-terminal sequence involved in cofactors interactions [7,8]. 17 $\beta$ -estradiol (E2)

and hydroxy-tamoxifen (OHT) are able to activate ER $\alpha$ 36, which is primarily located below the cell surface and triggers nongenomic pathway stimulation (JAK/STAT, PI3K/AKT, MEK/ERK) [9]. ER $\alpha$ 36 overexpression in mouse normal mammary epithelium is related to several phenotypic hallmarks of breast cancer cells, such as loss of adhesion and enhanced migration potential, resistance to apoptosis, and genetic instability [10]. Forced or endogenously high expression of ER $\alpha$ 36 in breast cancer cell lines or tumor samples is also known to stimulate proliferation, migration, stemness, and resistance to therapeutic agents, independently of the ER $\alpha$ 66 protein level [2], [11–13].

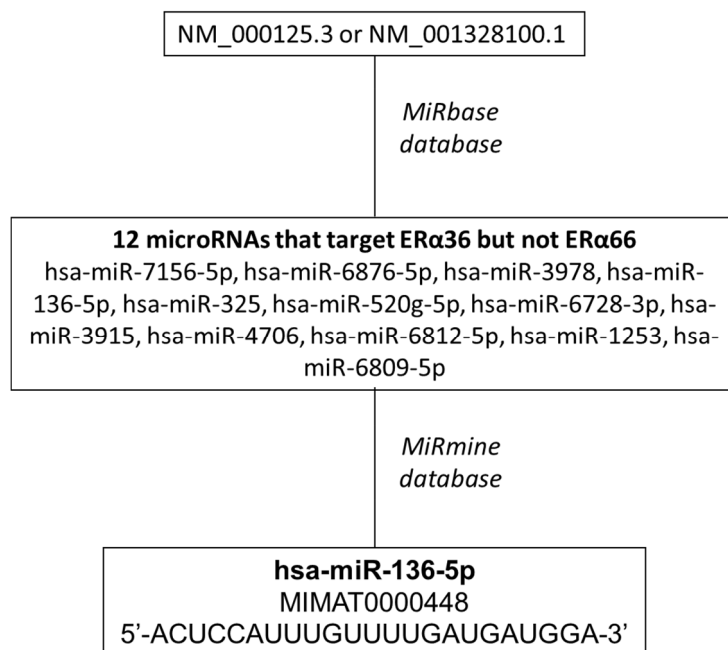
Classic regulation of ER $\alpha$ 36 expression by transacting factors has previously been investigated in breast cancer cells: ER $\alpha$ 66 was described as a negative regulator [14], whereas AP-1 family proteins, EGFR and HER2 as positive ones [15]. An epigenetic regulation of ER $\alpha$ 36 promoter activity was also suggested after mammary epithelial cell exposure to a long chain alkylphenol mixture [16]. In addition, accumulative data point out the role of microRNAs in the regulation of oncogenes and tumor suppressors [17]. Among the myriad of microRNAs identified in regulating oncogenesis, hsa-miR-136-5p could be a good candidate for ER $\alpha$ 36 regulation. Indeed, miR-136 expression negatively correlates to the tumor stage and metastatic potential in breast cell lines and tumor biopsies: hsa-miR-136-5p expression is high in normal tissue and MCF-10A cells where ER $\alpha$ 36 is absent, moderate in WHO stage I and II samples and luminal MCF-7 cells where ER $\alpha$ 36 expression level is low, and low in high grade tumors and triple-negative breast cancer cell (TNBC) lines such as MDA-MB-231 in which ER $\alpha$ 36 level is high [3,18,19]. Additionally, hsa-miR-136-5p is able to target RASAL2 in TNBC and subsequently suppresses migration and invasion, two processes which may involve ER $\alpha$ 36 functionality [18]. More generally, hsa-miR-136-5p is described as a tumor suppressor in many solid tumor models such as breast, prostate, ovarian, cervical, colon, glioma, osteosarcoma, and lung squamous cell cancer. Depending on the organ, it targets key actors of kinase-dependent transduction pathways (RASAL2, TCFPC2, MAP2K4, PTEN, wnt2) or downstream effectors (Bcl2, MMP3, MIEN1, metadherin, etc.) and thus participates to the control of proliferation, migration, invasion, and apoptosis [18], [20–24].

In an attempt to identify modulators of ER $\alpha$ 36 expression, we first screened microRNAs specifically targeting ER $\alpha$ 36. We selected 12 microRNAs and focused on hsa-miR-136-5p because (i) its expression in breast cancer cells appeared inversely correlated to ER $\alpha$ 36, (ii) both ER $\alpha$ 36 and hsa-miR-136-5p were involved in the same cancer cell features, and (iii) hsa-miR-136-5p was predicted to bind ER $\alpha$ 36 3'UTR. Moreover, we showed that the demethylating agent decitabine (DAC) modulated ER $\alpha$ 36 expression, directly by targeting the ER $\alpha$ 36 promoter region and indirectly by stimulating hsa-miR-136-5p expression, thus counteracting tamoxifen-induced ER $\alpha$ 36 overexpression.

## 2. Results

### 2.1. Screening of microRNAs that Target ER $\alpha$ 36 but not ER $\alpha$ 66

In a first step, we screened MiRbase database for microRNAs that target Homo sapiens estrogen receptor 1 (ESR1), transcript variant 7, 3'UTR (Acc number: NM\_001328100.1) encoding ER $\alpha$ 36 protein but not Homo sapiens estrogen receptor 1 (ESR1), transcript variant 1, 3'UTR (Acc number: NM\_000125.3) encoding ER $\alpha$ 66 (Figure 1) [25–28]. Twelve microRNAs predicted to potentially hybridize specifically to the 3'UTR of ER $\alpha$ 36 encoding transcript: hsa-miR-7156-5p, hsa-miR-6876-5p, hsa-miR-3978, hsa-miR-136-5p, hsa-miR-325, hsa-miR-520g-5p, hsa-miR-6728-3p, hsa-miR-3915, hsa-miR-4706, hsa-miR-6812-5p, hsa-miR-1253, hsa-miR-6809-5p. Each microRNA was submitted to the miRmine database to check its expression in normal/tumor breast tissue or cell lines. hsa-miR-136-5p (mature\_acc = MIMAT0000448; 5'-ACUCCAUUUGUUUGAUGAUGGA-3') displayed the highest expression level among the 12 candidates in MCF-7 breast cancer cells and was found two times less in sera from breast tumor patients than from normal ones [29].



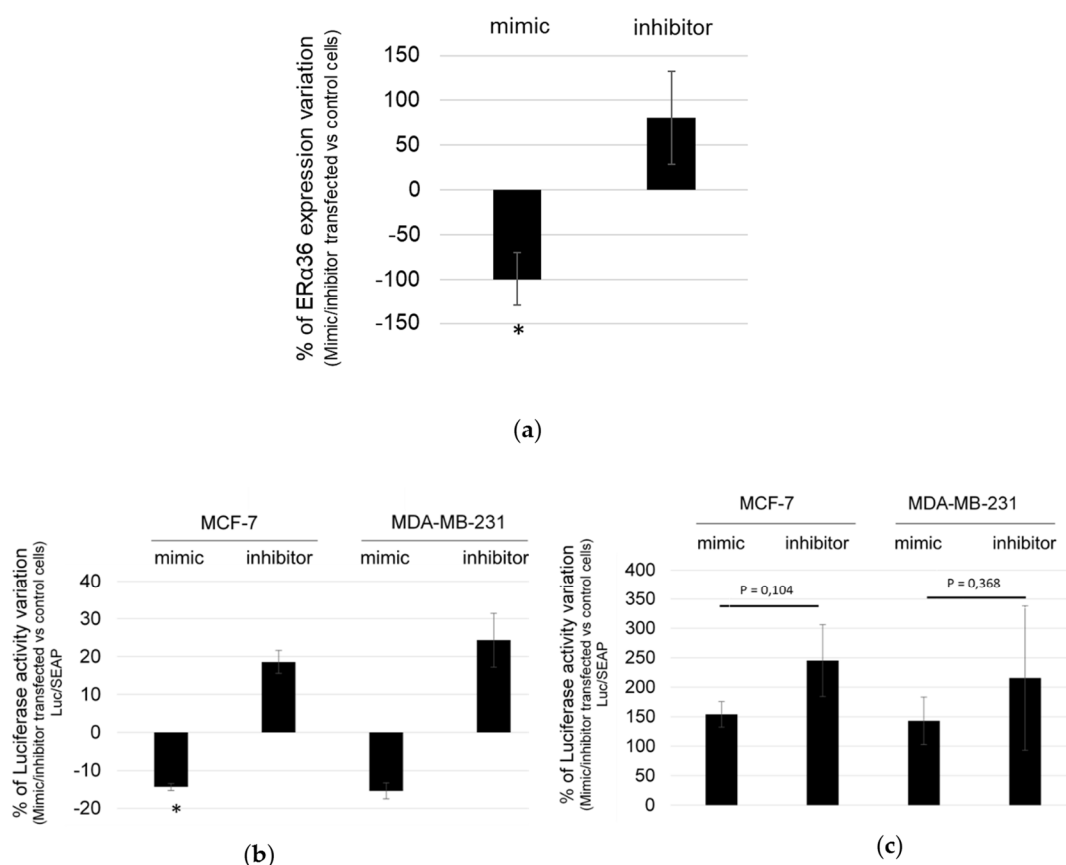
**Figure 1.** Strategy for the screening of microRNAs that target ER $\alpha$ 36 but not ER $\alpha$ 66.

## 2.2. hsa-miR-136-5p Under/over Expression Modulates ER $\alpha$ 36 mRNA Expression Level

To confirm that hsa-miR-136-5p modulate ER $\alpha$ 36 expression, MCF-7 cells were transfected with either hsa-miR-136-5p mimic or inhibitor. ER $\alpha$ 36 mRNA and hsa-miR-136-5p expression levels were measured by RT-qPCR analysis. The ratio of ER $\alpha$ 36 expression in hsa-miR-136-5p mimic- or inhibitor-transfected versus control cells indicated that ER $\alpha$ 36 mRNA expression level significantly decreased by 100% in the case of hsa-miR-136-5p mimic transfection compared to the hsa-miR-136-5p inhibitor one (Figure 2A). Transfection with hsa-miR-136-5p inhibitor led to an enhanced ER $\alpha$ 36 expression (Figure 2A), but this tendency remained not statistically significant, probably due to experimental variation in inhibitor transfection efficacy (data not shown).

The specific interaction between hsa-miR-136-5p and the 3'UTR region of ER $\alpha$ 36 mRNA was then assessed in both MCF-7 and MDA-MB-231 cells, which both express ER $\alpha$ 36 but belong either to luminal A or to triple-negative molecular subtype, respectively (Figure S1). The plasmid pMIR-Report-Luc-ER $\alpha$ 36wt (wt) contained the wild-type hsa-miR-136-5p target sequence of ER $\alpha$ 36 3'UTR and pMIR-Report-Luc-ER $\alpha$ 36mut (mut), a corresponding mutated sequence. Both the wt and mut sequence were cloned downstream the luciferase gene, thus driving luciferase mRNA stability in the presence of hsa-miR-136-5p. The cells were co-transfected with either the wt or the mut plasmid and hsa-miR-136-5p mimic or inhibitor. Luciferase activity was measured 24 h after transfection and normalized for transfection efficacy using the alkaline phosphatase activity from the co-transfected pEZX-PG04-Secreted alkaline phosphatase plasmid. Transfection of hsa-miR-136-5p mimic or inhibitor in MCF-7 or MDA-MB-231 cells resulted in a 120% to 150% decrease and a 180% to 240% increase of luciferase activity compared to control cells, respectively (Figure 2B). Both cell lines transfected by the mut plasmid displayed a higher basal level of luciferase activity than the wt one. There was no statistical variation between the hsa-miR-136-5p mimic or inhibitor transfected cells (Figure 2C). Therefore, we concluded that hsa-miR-136-5p directly and specifically targeted ER $\alpha$ 36 mRNA, thus modulating its expression level in breast cancer cells.

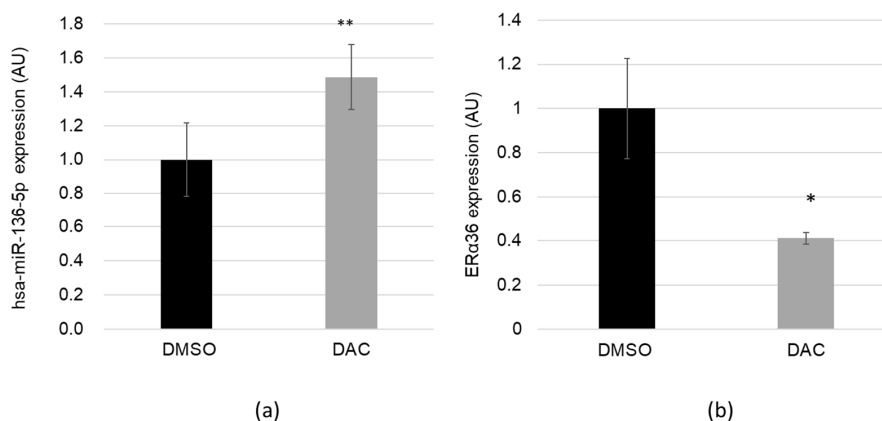




**Figure 2.** hsa-miR-136-5p mimic/inhibitor directly modulates ERα36 expression. (a) MCF-7 cells were transfected with either hsa-miR-136-5p mimic or inhibitor. ERα36 expression was measured by RT-qPCR analysis 24 h after transfection, and the percentage of ERα36 expression variation in hsa-miR-136-5p mimic or inhibitor transfected cells versus control ones was calculated. MCF-7 or MDA-MB-231 cells were co-transfected with hsa-miR-136-5p mimic or inhibitor and (b) the pMIR-Report-Luc-ERα36wt plasmid or (c) the pMIR-Report-Luc-ERα36mut one. Luciferase activity was measured 24 h after transfection using the Bright-Glo™ luciferase assay system. To evaluate the transfection efficacy and normalize the results, cells were also transfected at the same time with the pEZX-PG04-Secreted alkaline phosphatase plasmid. The alkaline phosphatase activity was measured 24 h after transfection in the culture medium with the Secrete-Pair™ Gaussia luciferase dual and single luminescence assay Kit. Each bar represents mean ± S.E.M. (a)  $N = 5$  (b)  $N = 3$  and (c)  $N = 5$ . \*  $P < 0.05$ .

### 2.3. The Demethylating Agent DAC Stimulates hsa-miR-136-5p and Represses ERα36 Expression

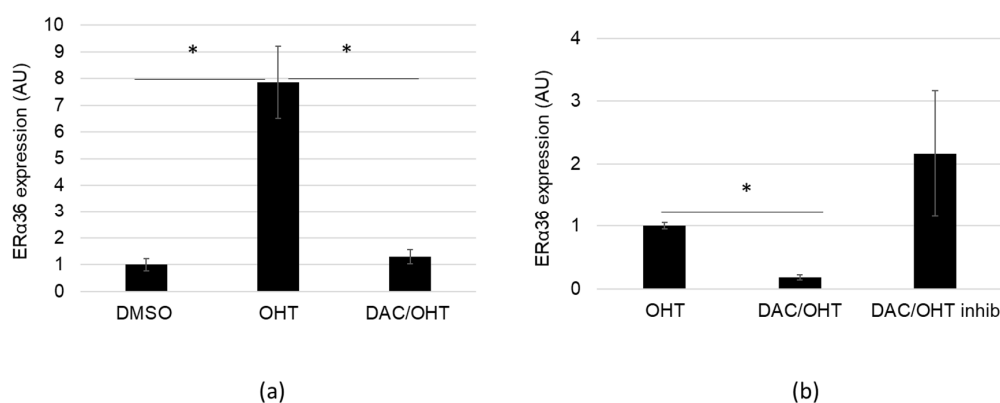
Many microRNAs are epigenetically regulated through the methylation status of their promoter [30]. hsa-miR-136-5p is located in the DLK1-DIO3 imprinted locus of chromosome 14q32, which contains a cluster of noncoding RNAs and the antisense gene RTL1, all differentially regulated by the methylation status of the IG-DMR. Methylation changes in this locus, which may affect gene or microRNA expression associated with Temple or Kagami–Ogata syndrome and oligometastatic phenotype [31,32]. Therefore, we hypothesized that expression of hsa-miR-136-5p could be upregulated after cell exposure to the demethylating agent DAC and thus downregulate ERα36. The MCF-7 cells were treated for 48 h with DAC or DMSO as control. DAC exposure did not alter cell viability (data not shown). MicroRNAs and long mRNAs were harvested from the same cells. hsa-miR-136-5p expression level (Figure 3A) and ERα36 expression (Figure 3B) were measured using RT-qPCR. DAC treatment induced a significant 48% increase of hsa-miR-136-5p expression as well as a 59% decrease of ERα36 expression. The upregulation was independent of the hsa-miR-136-5p sponge CRNDE (colorectal neoplasia differentially expressed gene) (Figure S2).



**Figure 3.** The demethylating agent decitabine (DAC) stimulates hsa-miR-136-5p expression and represses ERα36 one. MCF-7 cells were treated for 48 h with DMSO 0.01% (as control) or the demethylating agent DAC 5 μM. (a) hsa-miR-136-5p expression level was assessed using TaqMan™ microRNA assay and (b) ERα36 expression was measured by RT-qPCR analysis. Each bar represents mean ± S.E.M. (a)  $N = 6$  (b)  $N \geq 3$ . \*  $P < 0.05$ ; \*\*  $P < 0.01$ .

#### 2.4. DAC Counteracts OHT-dependent Stimulation of ERα36 Expression

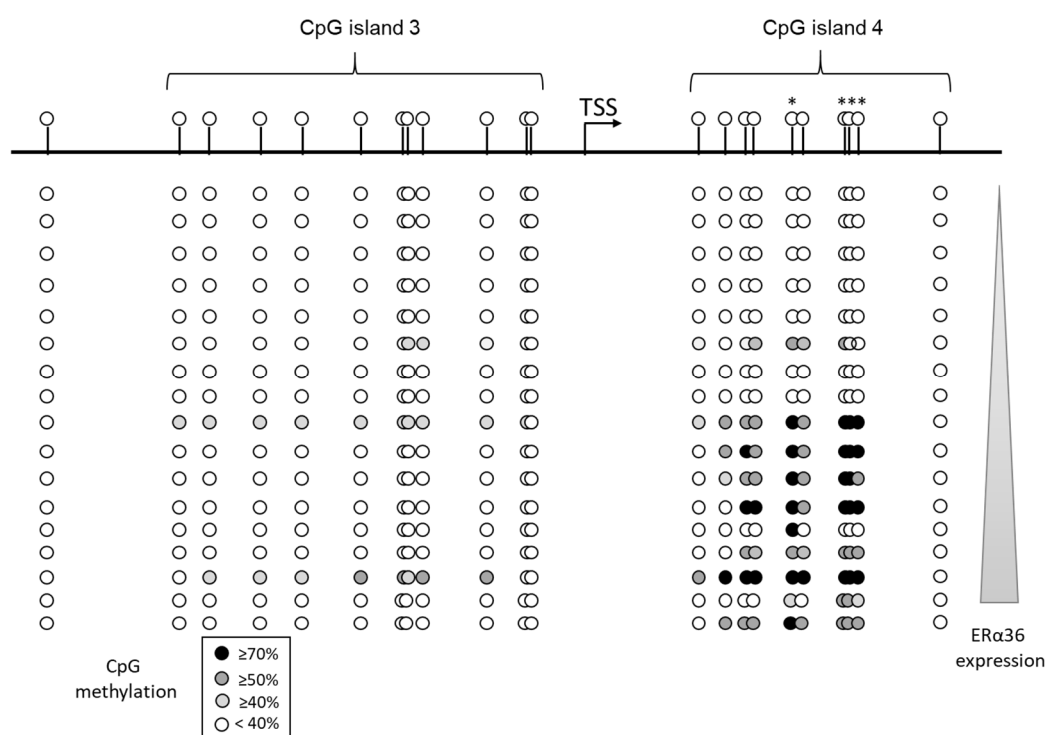
OHT was previously reported to either stimulate or repress ERα36 expression in MCF-7 cells exposed to 1 μM or 5 μM, respectively [5,33]. Since DAC (5 μM) and OHT (1 μM) displayed inverse effects on ERα36 expression, MCF-7 cells were exposed either to DAC or to OHT (1 μM), or co-treated with both compounds for 48 h. Exposure to 1 μM OHT-induced ERα36 expression by 8-fold although hsa-miR-136-5p expression was not affected by the treatment (data not shown). Moreover, co-treatment with DAC and OHT was able to counteract OHT-dependent stimulation of ERα36 expression (Figure 4A). Transfection of DAC/OHT co-treated cells with hsa-miR-136-5p inhibitor seemed to cancel the DAC-dependent impairment of OHT effect (Figure 4B), indicating that hsa-miR-136-5p was indeed a key element in this process. The variability of ERα36 expression observed in the [DAC/OHT + hsa-miR-136-5p inhibitor] condition was most likely due to uncontrolled experimental variations of the hsa-miR-136-5p inhibitor amount present into the cells. Taken together, these data suggested that DAC could be used to chemically repress ERα36 expression and prevent its induction by OHT through the stimulation of hsa-miR-136-5p.



**Figure 4.** Decitabine (DAC) treatment counteracts 4-hydroxy-tamoxifen (OHT) induced ERα36 expression. (a) MCF-7 cells were treated for 48 h with DMSO (as control), OHT 1 μM or by a combination OHT 1 μM/DAC 5 μM. ERα36 expression was measured by RT-qPCR analysis. (b) MCF-7 cells were transfected or not with hsa-miR-136-5p inhibitor. Both control and hsa-miR-136-5p inhibitor transfected cells were treated for 48 h by a combination of OHT 1 μM/DAC 5 μM. ERα36 expression was measured by RT-qPCR analysis. OHT treatment was used as reference. Each bar represents mean ± S.E.M.  $N \geq 3$ . \*  $P < 0.05$ .

### 2.5. ER $\alpha$ 36 Expression May be Modulated by the Methylation Status of Its Promoter, In Vivo and In Vitro

Previous studies from our laboratory [16,34] suggested that ER $\alpha$ 36 expression may be modulated by the methylation status of its promoter sequence, as defined by Wang's laboratory [14]. We searched for CpG islands located in the first intron of ESR1 using the EMBOSS CpG plot website ([http://www.ebi.ac.uk/Tools/seqstats/emboss\\_cpgplot/](http://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/)) and identified four clusters (Figure S3). The methylation status of corresponding sequences was determined after bisulfite conversion in a subset of 17 breast primary tumors collected from the biobank of the Cancerology Institute of Lorraine (Vandoeuvre-lès-Nancy, France). Total RNAs were harvested from the same samples in order to measure ER $\alpha$ 36 expression level by RT-qPCR. The correlation between each CpG methylation level and ER $\alpha$ 36 expression level was addressed via a Pearson rank statistical test (Figure 5). The results clearly suggested that the methylation of 4 CpGs located on the island 4 was correlated to a high ER $\alpha$ 36 expression.

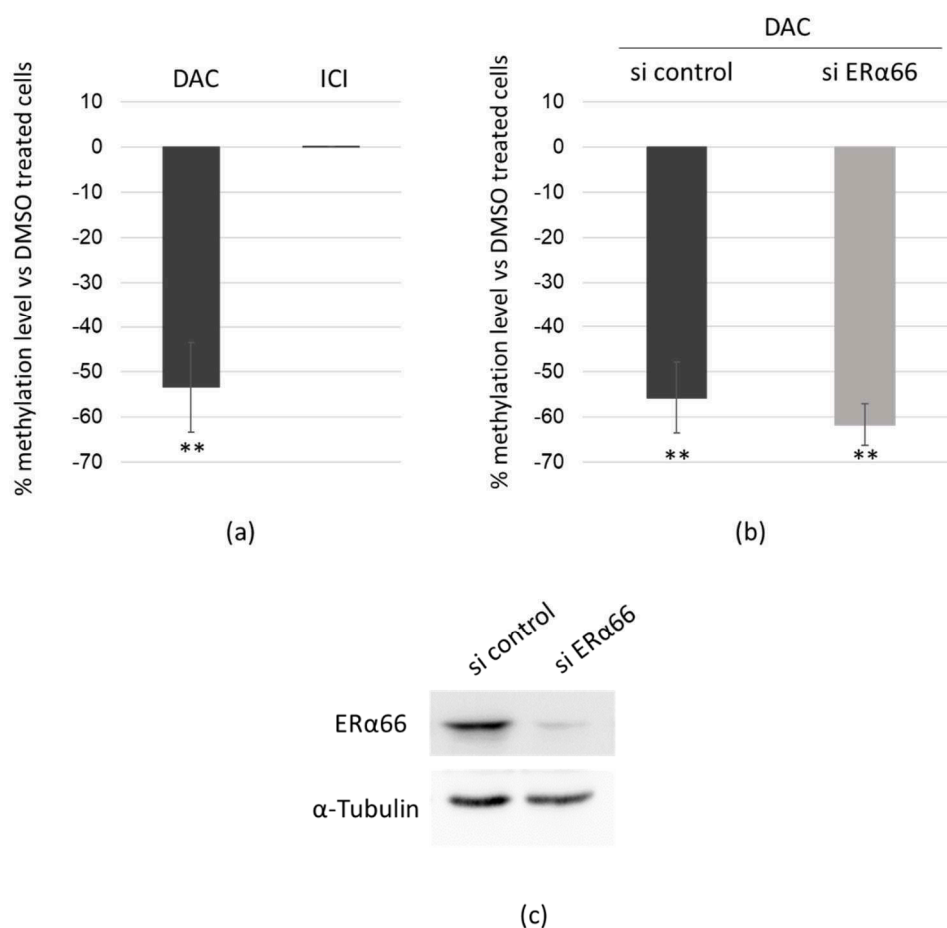


**Figure 5.** ER $\alpha$ 36 expression is related to the methylation status of its promoter in breast tumor samples. Seventeen breast tumor samples were collected from the tumor biobank of the Cancerology Institute of Lorraine. ER $\alpha$ 36 promoter genomic DNA was submitted to bisulfite sequencing and ER $\alpha$ 36 expression level was performed by real-time PCR analyses in triplicates for each tumor sample. The correlation between each CpG methylation and ER $\alpha$ 36 expression was established by a MATLAB program (Pearson rank correlation test). \* CpG which methylation level is significantly ( $P < 0.05$ ) correlated to ER $\alpha$ 36 expression level.

### 2.6. DAC Directly Targets ER $\alpha$ 36 Genomic Sequence

Since the methylation status of ER $\alpha$ 36 promoter seemed to be linked to its expression in tumor biopsies, we addressed a direct effect of the demethylating agent DAC on ER $\alpha$ 36 expression through demethylation of its promoter. MCF-7 cells were exposed to DAC for 48 h, and the methylation status of the four CpGs significantly related to ER $\alpha$ 36 expression was assessed after bisulfite conversion by real-time PCR. ER $\alpha$ 66 is known to repress ER $\alpha$ 36 expression and was reported to regulate its target genes through epigenetic modulator recruitment [35]. Thus, we compared the effect of the exposure to DAC or to the ER $\alpha$ 66 inhibitor fulvestrant (Figure 6A). DAC triggered a 53% decrease in the methylation level of the ER $\alpha$ 36 promoter sequence where ICI treatment had no effect. The same experiment was

performed in MCF-7 cells transfected by an ER $\alpha$ 66 siRNA to knockdown ER $\alpha$ 66 and exposed to DAC for 48 h (Figure 6B). The ER $\alpha$ 66 knockdown was confirmed by Western blot (Figure 6C). DAC significantly and similarly demethylated ER $\alpha$ 36 promoter in both si control and siER $\alpha$ 66 transfected cells. This confirmed that DAC directly acted on ER $\alpha$ 36 promoter sequence independently of ER $\alpha$ 66 and further suggested that the DAC-dependent repression of ER $\alpha$ 36 expression could account for (i) an indirect effect through the hsa-miR-136-5p targeting of ER $\alpha$ 36 3'UTR and (ii) a direct effect through demethylation of its promoter.



**Figure 6.** Decitabine directly targets ER $\alpha$ 36 genomic sequence. (a) MCF-7 cells were treated for 48 h with DMSO (as control), DAC or ICI 182,780 ( $N = 4$ ); (b) MCF-7 cells transfected with either scrambled siRNA (si control,  $N = 4$ ) or siRNA targeting the exon 1 of ESR1 (siER $\alpha$ 66,  $N = 3$ ) were treated for 48 h with DMSO (as control) or DAC. The methylation level of the ER $\alpha$ 36 promoter fourth CpG island was assessed after bisulfite conversion. A percentage of methylation level variation was calculated for DAC or ICI treated cells versus DMSO ones. (c) Representative Western blot analysis of ER $\alpha$ 66 expression in MCF-7 cells transfected with either si control or si ER $\alpha$ 66.  $\alpha$ -Tubulin was used as a loading control. Each bar represents mean  $\pm$  S.E.M. \*\*  $P < 0.01$ .

### 3. Discussion

This study depicts, for the first time, two epigenetic mechanisms that regulate ER $\alpha$ 36 expression in breast cancer cells. In line with the study of Li and colleagues [5], we showed that ER $\alpha$ 36 expression increased significantly after a low dose tamoxifen exposure in MCF-7 cell line. This upregulation was previously associated with a decreased sensitivity to tamoxifen and enhanced proliferative, migratory, and invasive abilities of breast cancer cells [5,15]. Mechanistically, computational modeling and docking analysis revealed that tamoxifen is able to bind directly ER $\alpha$ 36 through its ligand-binding domain [2]. This interaction activated various estrogen nongenomic pathways such as MAPK/Erk

and PI3K/Akt as well as the ER $\alpha$ 36/EGFR/HER2 regulatory loop responsible for tamoxifen therapy escape [9,15,36]. These experimental results provided support and explanations to clinical studies wherein a high ER $\alpha$ 36 expression in tumors treated with tamoxifen was significantly associated with a poor prognosis and an increased rate of metastases [2,37]. In this context, the understanding of mechanisms that regulate ER $\alpha$ 36 expression was important to try preventing the deleterious effect of ER $\alpha$ 36-triggered response to tamoxifen.

First, we showed in vitro and in breast tumor samples that a low methylation level of 4 CpGs localized in the promoter of ER $\alpha$ 36 correlated with a low ER $\alpha$ 36 mRNA expression. A demethylation of this CpG rich region was namely obtained in vitro after a 48 h DAC exposure which led to ER $\alpha$ 36 downregulation. This supported the causal relationship between the methylation status of the promoter, and the expression level of ER $\alpha$ 36. ER $\alpha$ 66 has been described as the main direct transcriptional regulator of ER $\alpha$ 36 expression and is able to regulate his target genes through the methylation level of their promoter [14,35,38,39]. When we silenced ER $\alpha$ 66 expression or activity by either siRNAs or ICI 280 182, we showed that ER $\alpha$ 66 did not drive the methylation-dependent repression of ER $\alpha$ 36.

DAC is an FDA-approved hypomethylating agent, currently used to treat blood cancer. Recently, a study conducted by researchers from Florida on animal models demonstrated that a low dose of DAC prevented breast cancer cells from spreading and could also be used to treat aggressive chemoresistant triple-negative breast cancer [40,41]. As previously described in the case of ER $\beta$  silencing [42], treatment with DAC may promote ER $\alpha$ 36 promoter demethylation and thus override the molecular mechanisms underlying ER $\alpha$ 36-dependent tamoxifen resistance.

Second, our data indicated that DAC also acted indirectly on ER $\alpha$ 36 expression since it upregulated hsa-miR-136-5p expression which, in turn, targeted ER $\alpha$ 36 mRNA. hsa-miR-136-5p was described as a tumor suppressor microRNA, associated with impaired tumorigenesis and metastasis in prostate cancer by targeting MAP2K4 [43], renal cell carcinoma [23], lung squamous cell carcinoma by controlling metabolic processes [22], lung adenocarcinoma by suppressing cell adhesion gene expression [24], and glioma and cutaneous squamous cell carcinoma by targeting WNT signaling [44,45]. The downregulation of hsa-miR-136-5p has been described in a chemo-radioresistant group of patients with advanced stage cervical squamous cell carcinoma. Conversely, a high ER $\alpha$ 36 expression was associated to estrogen nongenomic signaling driving metastatic progression, acquired resistance to hormone treatment, and poor outcome [11,13,46].

Our in vitro data demonstrated that hsa-miR-136-5p directly targeted ER $\alpha$ 36 3'UTR and led to mRNA expression downregulation. These data suggested that the upregulation of ER $\alpha$ 36 expression in response to OHT could be impaired either by the direct mimicking or by the DAC-dependent indirect stimulation of hsa-miR-136-5p that may consequently override ER $\alpha$ 36-dependent resistance to OHT. Further in vivo experiments are needed to explore whether treatment with DAC, prior or in combination with OHT, would be a relevant therapeutic strategy.

In the mammary gland, hsa-miR-136-5p expression was higher in normal breast tissue than in breast carcinoma and in luminal vs. triple-negative breast cancer cell lines [18,19,47]. In line, Pascek and coworkers described a significant higher expression in luminal tumors than in TNBCs [48]. However, other studies from tumor biopsies did not report hsa-miR-136-5p expression as a reliable marker for predicting either breast cancer molecular type [49], pathological stage [50] or patient outcome [51]. Little is known about hsa-miR-136-5p regulation by DNA methylation status. The promoter sequence of hsa-miR-136-5p has not been clearly defined. A 135 bp sequence is located upstream hsa-miR-136-5p and separates hsa-miR-437-5p from hsa-miR-136-5p, but the promoter status of this sequence or the relationship between its methylation level and hsa-miR-136-5p expression are not described. Moreover, hsa-miR-136-5p is located in the 14q32.2 Dlk1-Dio3 chromosomal region, which has been described to be submitted to global and complex regulation by parental imprinting through two DMR [52]. Therefore, it would be interesting to determine precisely the regulatory elements of hsa-miR-136-5p and to set up a prospective study in order to investigate both its expression and the methylation status of its putative promoter in breast tumors.

Expression of hsa-miR-136-5p was also tightly regulated by a combination of sponge lncRNAs, such as CRNDE, HOX transcript antisense RNA (HOTAIR), lncRNA-small nucleolar RNA host gene 14 (SNHG14) or the circular RNA hsa\_circ\_0008309 [19,53–55]. Like in many solid tumors, the main factor to which expression could be relevant to predict patient outcome is CRNDE, the hsa-miR-136-5p sponge [19]. CRNDE was of peculiar interest in breast cancer since a recent meta-analysis confirmed this lncRNA could act as an oncogene which upregulated expression correlated with poor prognosis (late Tumor Node Metastasis grade) and tumor progression (high tumor size and lymph node metastasis) [56,57]. In vitro, CRNDE controls hsa-miR-136-5p level which, in turn, is able to target RASAL2 and Wnt/beta-catenin signaling in triple-negative breast cancer cells [18,19]. In the present study, we showed that hsa-miR-136-5p could modulate ER $\alpha$ 36 expression independently of CRNDE. Thus, we added ER $\alpha$ 36 as another piece to the puzzle of the complex epigenetic regulation of breast cancer progression.

## 4. Materials and Methods

### 4.1. Reagents

4-hydroxytamoxifen (OHT), fulvestrant (ICI 182, 780), and 5-Aza-2'-deoxycytidine (decitabine, DAC) were purchased from Sigma-Aldrich, France. Stock solutions (10 mM) of each compound were prepared in dimethyl sulfoxide (DMSO) and further diluted in culture medium for in vitro cell treatment. All working solutions were freshly made just before treatment, and control cells were treated with DMSO, diluted with the same factor.

### 4.2. Plasmids

Reporter plasmids containing the miR-136-5p target sequence from ER $\alpha$ 36 3'UTR downstream the luciferase open reading frame were constructed as follows: pMIR-Report-Luc-KLF4-FL vector was purchased from the nonprofit Addgene repository (Addgene plasmid #34597) as a gift from Michael Ruppert [58]. The KLF4 sequence was removed and replaced by either the wild-type ER $\alpha$ 36 3'UTR sequence (5'-CCACCATCAAATCAAATTGA-3') or the mutated one (5'-CCACCACTACAATACAATTGA-3'). The corresponding plasmids were renamed pMIR-Report-Luc-ER $\alpha$ 36wt and pMIR-Report-Luc-ER $\alpha$ 36mut, respectively. pEZX-PG04-Secreted Alkaline Phosphatase plasmid was purchased from GeneCoppoeia™ (Tebu-bio, Le Perray-en-Yvelines, France).

### 4.3. Cell Culture

MCF-7 and MDA-MB-231 cells were purchased in 2015 from ATCC® (HTB-22™ and HTB-26™, respectively; Molsheim, France) and maintained in 10% fetal bovine serum (FBS) 1% Glutamine supplemented Dulbecco/Vogt modified Eagle's minimal essential medium (DMEM) (GIBCO) or Roswell Park Memorial Institute medium (RPMI) (GIBCO) medium, respectively.

### 4.4. Transfection

Transient transfections of MCF-7 or MDA-MB-231 cells were performed using JetPRIME transfection reagent (Ozyme, Saint-Cyr-l'École, France) in accordance with the manufacturer's instructions. Briefly, cells were plated at a density of  $2.5 \times 10^5$  cells per well in 6-well plates. After 24 h, transient transfection was performed with either 1  $\mu$ g plasmid and/or 30 pmol (15 nM final) miRVana hsa-miR-136-5p mimic (5'-CAUCAUCAAAAACAAAUGGAGUTT-3') or inhibitor (5'-UCCAUCAUCAAAAACAAAUGGAGU-3') (Thermo Fischer, Illkirch, France) in serum-containing medium. The growth medium was renewed 4 hours after the transfection, and cells were treated with selected agents for 24 h.

#### 4.5. Luciferase/Alkaline Phosphatase Assays

The luciferase activity associated with the plasmid pMIR-Report-Luc-ER $\alpha$ 36wt and the plasmid pMIR-Report-Luc-ER $\alpha$ 36mut was assessed using the Bright-Glo™ luciferase assay system (Promega Corporation, Madison, WI, USA) in accordance with the manufacturer's instructions. To normalize the results, pEZX-PG04-Secreted alkaline phosphatase plasmid was used, and the secreted alkaline phosphatase activity was detected on cell culture supernatant using the Secrete-Pair™ Gaussia luciferase dual and single luminescence assay kit (GeneCoppoeia™ Tebu-bio, Le Perray-en-Yvelines, France) in accordance with the manufacturer's instructions. Luminescence was measured by a microplate reader (Victor x3, Perkin-Elmer, Villebon-sur-Yvette, France).

#### 4.6. MicroRNA Extraction and Reverse Transcription

Micro and long RNAs were isolated using the mirVana™ miRNA Isolation Kit (Ambion, Austin, TX, USA) in accordance with the manufacturer's instructions. The purity and quantity of RNA were assessed using the NanoDrop ND-2000C spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The samples were used immediately or stored at  $-80^{\circ}\text{C}$  for future use. Reverse transcription (RT) and real-time PCR (qPCR) were carried out using TaqMan™ microRNA assay (Applied Biosystems, Foster City, CA, USA). RT reactions were performed in a volume of 15  $\mu\text{L}$ , and each reaction contained 60 ng of microRNA. RT reactions were performed with a Mastercycler® Gradient (Eppendorf, Montesson, France) with the following conditions:  $16^{\circ}\text{C}$  for 30 min,  $42^{\circ}\text{C}$  for 30 min,  $85^{\circ}\text{C}$  for 5 min, and  $4^{\circ}\text{C}$  on hold. Reactions without addition of reverse transcriptase (RT (-) controls) were performed alongside with cDNA synthesis of each sample and used in subsequent procedures to control the potential genomic DNA contamination. A total of 1.4  $\mu\text{L}$  of RT reaction product was added in the 20  $\mu\text{L}$  qPCR reaction mix containing TaqMan™ Universal PCR Master Mix II without UNG (Applied Biosystems). All TaqMan assays were run in triplicate on a CFX96 Touch™ real-time PCR detection system (Biorad, Marnes-la-Coquette, France). Real-time PCR cycling conditions consisted of  $95^{\circ}\text{C}$  for 10 min, followed by 50 cycles of  $95^{\circ}\text{C}$  for 15 s and  $60^{\circ}\text{C}$  for 1 min. TaqMan Probe (hsa-miR-136-5p: 5'-UGAGCCCUCGGAGGACUCCAUUUGUUUUGAUGAUGGAUUCUUAUGCUCUCAUCAUCG UCUCAAAUGAGUCUUCUGUGGGUUCU-3'; hsa-miR-423-5p: 5'-UGAGGGGCAGAGAGCGAGAC UUUUCUAUUUCCAAAAGCUCGGUCUGAGGCCCCUCAGU-3'). Assays were performed at least as independent experimental triplicates, and the mean values were used to calculate expression levels, using the  $\Delta\Delta\text{C}(t)$  method referring to hsa-miR-423 housekeeping miR expression (as recommended by the manufacturer, Applied Biosystems).

#### 4.7. Long RNAs RT and Real-time PCR Analysis

RT and real-time PCR analyses were performed as previously described [16]. The following primers were used for qRT-PCR: GAPDH forward (Fw) 5'-TGC-ACC-ACC-AAC-TGC-TTA-GC-3', GAPDH reverse (Rev) 5'-GGC-ATG-GAC-TGT-GGT-CAT-GAG-3', ER $\alpha$ 36 forward (Fw) 5'-ATG-AAT-CTG-CAG-GGA-GAG-GA-3', ER $\alpha$ 36 reverse (Rev) 5'-GGC-TTT-AGA-CAC-GAG-GAA-ACC-3', CRNDE forward (Fw) 5'-ATA-TTC-AGC-CGT- TGG-TCT-TTG-A-3', and CRNDE reverse (Rev) 5'-ATA-TTC-AGC-CGT-TGG-TCT-TTG-A-3'. Assays were performed at least in triplicate, and the mean values were used to calculate expression levels, using the  $\Delta\Delta\text{C}(t)$  method referring to GAPDH housekeeping gene expression.

#### 4.8. Western Immunoblotting

Western blots were performed as described previously [10]. We used the primary antibody anti-ER $\alpha$ 66 (F10, Santa Cruz Biotechnology, Heidelberg, Germany). The anti- $\alpha$ -Tubulin (GTX102079, Euromedex, Souffelweyersheim, France) was used as a loading control. Protein expression profiles were revealed with Clarity Western ECL Substrate (Biorad), and banding quantification was performed using the Image Lab 6.0.0 software (Biorad).

#### 4.9. Patients and Samples

Cryopreserved samples (Biobank registration number AC-2008-174) from 17 patients treated at Cancerology Institute of Lorraine (Nancy, France) between 2014 and 2015 with histology proven T1 to T4 invasive breast cancer were used for this study. Among the 17 patients, 14 (82%) had an invasive ductal carcinoma and 3 (18%) had a nonspecific invasive carcinoma. Breast tumors were examined by hematoxylin and eosin staining and histopathology defined in accordance with the sixth edition of TNM [59]. Hormonal receptors, HER2 and Ki-67 expressions were assessed using immunohistochemistry (Dako, Les Ulis, France). Patients' characteristics are detailed in Table S1. The investigations were carried out following the rules of the Declaration of Helsinki of 1975 (<https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/>), revised in 2013. All patients gave their informed consent for the use of their samples, and all data were anonymized prior to analysis to protect patient confidentiality. This study was performed within the framework of the tumor banking program of the Cancerology Institute of Lorraine including ethical considerations (approval AC-2013-1916 French Ministry of Research).

#### 4.10. Bisulfite Sequencing

Bisulfite sequencing data of tumor samples were obtained from Active Motif's Targeted Next-Gen Bisulfite Sequencing Service (Active Motif Europe, La Hulpe, Belgium).

#### 4.11. DNA Methylation Analysis of ER $\alpha$ 36 Promoter Sequence

Genomic DNA was harvested from MCF-10A, MCF-7 or MDA-MB-231, and bisulfite conversion of DNA was performed with the bisulfite conversion kit (Active Motif Europe, La Hulpe, Belgium) in accordance with the manufacturer's instructions. Real-Time PCR analysis of methylated/unmethylated DNA was performed after conversion using specific primers (forward 5'-GAG-TTT-AAA-ATA-AGT-TAT-ATG-GAA-GTA-TAA-GTG-3'; Meth reverse 5'-CGA-CCG-CGC-TCC-TTC-CAC-AAT-AAC-TAC-GA-3'; Unmeth reverse 5'-CAA-CCA-CAC-TCC-TTC-CAC-AAT-AAC-TAC-AA-3').

#### 4.12. Statistical Analysis

Statistical analyses were performed with MATLAB R2018a software (MathWorks) using Student's *t*-test for unpaired samples and ANOVA followed by Bonferroni post-hoc test for multigroup comparison, with a significant *p*-value threshold below 5%. \* *p* < 0.05; \*\* *p* < 0.01; \*\*\* *p* < 0.001. Standard deviations or standard errors were indicated on figures as advocated by Altman and Bland [60].

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/11/2637/s1>.

**Author Contributions:** Conceptualization, H.D.; Methodology, H.D., A.C. and A.L.; Validation, H.D., A.H., A.C. and J.L.M.; Investigation, C.T. and M.C.; Resources, A.L.; Writing—Original Draft Preparation, C.T. and A.H.; Writing—Review and Editing, H.D. and A.H.; Supervision, A.H. and H.D.; Project Administration, H.D.; Funding Acquisition, H.D. and J.L.M.

**Funding:** This research was funded by the Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES; PNR EST 2012-2-014), INSERM Cancer-Environnement (ENV201304) and La Ligue contre le Cancer (2016).

**Acknowledgments:** We thank Audrey Taverniers for technical assistance.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Szostakowska, M.; Trębińska-Stryjewska, A.; Grzybowska, E.A.; Fabisiwicz, A. Resistance to endocrine therapy in breast cancer: Molecular mechanisms and future goals. *Breast Cancer Res. Treat.* **2019**, *173*, 489–497. [[CrossRef](#)]



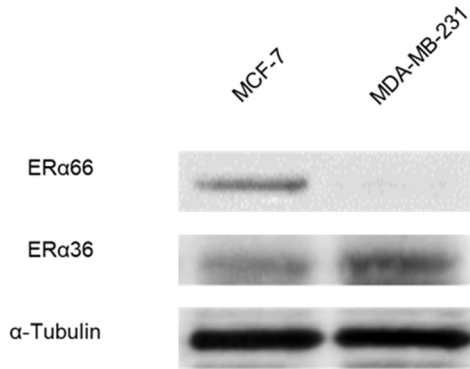
2. Wang, Q.; Jiang, J.; Ying, G.; Xie, X.-Q.; Zhang, X.; Xu, W.; Zhang, X.; Song, E.; Bu, H.; Ping, Y.-F.; et al. Tamoxifen enhances stemness and promotes metastasis of ER $\alpha$ 36+ breast cancer by upregulating ALDH1A1 in cancer cells. *Cell Res.* **2018**, *28*, 336–358. [[CrossRef](#)]
3. Wang, Z.; Zhang, X.; Shen, P.; Loggie, B.W.; Chang, Y.; Deuel, T.F. A variant of estrogen receptor- $\alpha$ , hER- $\alpha$ 36: Transduction of estrogen- and antiestrogen-dependent membrane-initiated mitogenic signaling. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9063–9068. [[CrossRef](#)]
4. Zhang, X.; Wang, Z.-Y. Estrogen receptor- $\alpha$  variant, ER- $\alpha$ 36, is involved in tamoxifen resistance and estrogen hypersensitivity. *Endocrinology* **2013**, *154*, 1990–1998. [[CrossRef](#)]
5. Li, G.; Zhang, J.; Jin, K.; He, K.; Zheng, Y.; Xu, X.; Wang, H.; Wang, H.; Li, Z.; Yu, X.; et al. Estrogen receptor- $\alpha$ 36 is involved in development of acquired tamoxifen resistance via regulating the growth status switch in breast cancer cells. *Mol. Oncol.* **2013**, *7*, 611–624. [[CrossRef](#)] [[PubMed](#)]
6. Wang, Z.; Zhang, X.; Shen, P.; Loggie, B.W.; Chang, Y.; Deuel, T.F. Identification, cloning, and expression of human estrogen receptor-alpha36, a novel variant of human estrogen receptor-alpha66. *Biochem. Biophys. Res. Commun.* **2005**, *336*, 1023–1027. [[CrossRef](#)]
7. Chaudhri, R.A.; Hadadi, A.; Lobachev, K.S.; Schwartz, Z.; Boyan, B.D. Estrogen receptor-alpha 36 mediates the anti-apoptotic effect of estradiol in triple negative breast cancer cells via a membrane-associated mechanism. *Biochim. Biophys. Acta* **2014**, *1843*, 2796–2806. [[CrossRef](#)]
8. Omarjee, S.; Jacquemetton, J.; Poulard, C.; Rochel, N.; Dejaegere, A.; Chebaro, Y.; Treilleux, I.; Marangoni, E.; Corbo, L.; Romancer, M.L. The molecular mechanisms underlying the ER $\alpha$ -36-mediated signaling in breast cancer. *Oncogene* **2017**, *36*, 2503–2514. [[CrossRef](#)]
9. Chaudhri, R.A.; Schwartz, N.; Elbaradie, K.; Schwartz, Z.; Boyan, B.D. Role of ER $\alpha$ 36 in membrane-associated signaling by estrogen. *Steroids* **2014**, *81*, 74–80. [[CrossRef](#)] [[PubMed](#)]
10. Thiebaut, C.; Chamard-Jovenin, C.; Chesnel, A.; Morel, C.; Djermoune, E.-H.; Boukhobza, T.; Dumond, H. Mammary epithelial cell phenotype disruption in vitro and in vivo through ERalpha36 overexpression. *PLoS ONE* **2017**, *12*, e0173931. [[CrossRef](#)] [[PubMed](#)]
11. Wang, Z.-Y.; Yin, L. Estrogen receptor alpha-36 (ER- $\alpha$ 36): A new player in human breast cancer. *Mol. Cell. Endocrinol.* **2015**, *418*, 193–206. [[CrossRef](#)]
12. Sołtysik, K.; Czekaj, P. ER $\alpha$ 36—Another piece of the estrogen puzzle. *Eur. J. Cell Biol.* **2015**, *94*, 611–625. [[CrossRef](#)]
13. Chamard-Jovenin, C.; Jung, A.C.; Chesnel, A.; Abecassis, J.; Flament, S.; Ledrappier, S.; Macabre, C.; Boukhobza, T.; Dumond, H. From ER $\alpha$ 66 to ER $\alpha$ 36: A generic method for validating a prognosis marker of breast tumor progression. *BMC Syst. Biol.* **2015**, *9*, 28. [[CrossRef](#)]
14. Zou, Y.; Ding, L.; Coleman, M.; Wang, Z. Estrogen receptor-alpha (ER- $\alpha$ ) suppresses expression of its variant ER- $\alpha$ 36. *FEBS Lett.* **2009**, *583*, 1368–1374. [[CrossRef](#)]
15. Yin, L.; Wang, Z.-Y. Roles of the ER- $\alpha$ 36-EGFR/HER2 positive regulatory loops in tamoxifen resistance. *Steroids* **2016**, *111*, 95–99. [[CrossRef](#)]
16. Chamard-Jovenin, C.; Thiebaut, C.; Chesnel, A.; Bresso, E.; Morel, C.; Smail-Tabbone, M.; Devignes, M.-D.; Boukhobza, T.; Dumond, H. Low-Dose Alkylphenol Exposure Promotes Mammary Epithelium Alterations and Transgenerational Developmental Defects, But Does Not Enhance Tumorigenic Behavior of Breast Cancer Cells. *Front. Endocrinol.* **2017**, *8*, 272. [[CrossRef](#)]
17. Piletič, K.; Kunej, T. MicroRNA epigenetic signatures in human disease. *Arch. Toxicol.* **2016**, *90*, 2405–2419. [[CrossRef](#)]
18. Yan, M.; Li, X.; Tong, D.; Han, C.; Zhao, R.; He, Y.; Jin, X. miR-136 suppresses tumor invasion and metastasis by targeting RASAL2 in triple-negative breast cancer. *Oncol. Rep.* **2016**, *36*, 65–71. [[CrossRef](#)]
19. Huan, J.; Xing, L.; Lin, Q.; Xui, H.; Qin, X. Long noncoding RNA CRNDE activates Wnt/ $\beta$ -catenin signaling pathway through acting as a molecular sponge of microRNA-136 in human breast cancer. *Am. J. Transl. Res.* **2017**, *9*, 1977–1989.
20. Guo, T.; Pan, G. MicroRNA-136 functions as a tumor suppressor in osteosarcoma via regulating metadherin. *Cancer Biomark. Sect. Dis. Markers* **2018**, *22*, 79–87. [[CrossRef](#)]
21. Deng, G.; Gao, Y.; Cen, Z.; He, J.; Cao, B.; Zeng, G.; Zong, S. miR-136-5p Regulates the Inflammatory Response by Targeting the IKK $\beta$ /NF- $\kappa$ B/A20 Pathway After Spinal Cord Injury. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* **2018**, *50*, 512–524. [[CrossRef](#)]

22. Xie, Z.-C.; Li, T.-T.; Gan, B.-L.; Gao, X.; Gao, L.; Chen, G.; Hu, X.-H. Investigation of miR-136-5p key target genes and pathways in lung squamous cell cancer based on TCGA database and bioinformatics analysis. *Pathol. Res. Pract.* **2018**, *214*, 644–654. [[CrossRef](#)]
23. Chen, P.; Zhao, L.; Pan, X.; Jin, L.; Lin, C.; Xu, W.; Xu, J.; Guan, X.; Wu, X.; Wang, Y.; et al. Tumor suppressor microRNA-136-5p regulates the cellular function of renal cell carcinoma. *Oncol. Lett.* **2018**, *15*, 5995–6002. [[CrossRef](#)]
24. Li, T.-T.; Gao, X.; Gao, L.; Gan, B.-L.; Xie, Z.-C.; Zeng, J.-J.; Chen, G. Role of upregulated miR-136-5p in lung adenocarcinoma: A study of 1242 samples utilizing bioinformatics analysis. *Pathol. Res. Pract.* **2018**, *214*, 750–766. [[CrossRef](#)]
25. Kozomara, A.; Griffiths-Jones, S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **2014**, *42*, D68–D73. [[CrossRef](#)]
26. Kozomara, A.; Griffiths-Jones, S. miRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **2011**, *39*, D152–D157. [[CrossRef](#)]
27. Griffiths-Jones, S.; Grocock, R.J.; van Dongen, S.; Bateman, A.; Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* **2006**, *34*, D140–D144. [[CrossRef](#)]
28. Griffiths-Jones, S.; Saini, H.K.; van Dongen, S.; Enright, A.J. miRBase: Tools for microRNA genomics. *Nucleic Acids Res.* **2008**, *36*, D154–D158. [[CrossRef](#)]
29. Panwar, B.; Omenn, G.S.; Guan, Y. miRmine: A database of human miRNA expression profiles. *Bioinform. Oxf. Engl.* **2017**, *33*, 1554–1560. [[CrossRef](#)]
30. Gulyaeva, L.F.; Kushlinskiy, N.E. Regulatory mechanisms of microRNA expression. *J. Transl. Med.* **2016**, *14*, 143. [[CrossRef](#)]
31. Eggermann, T.; Perez de Nanclares, G.; Maher, E.R.; Temple, I.K.; Tümer, Z.; Monk, D.; Mackay, D.J.G.; Grønskov, K.; Riccio, A.; Linglart, A.; et al. Imprinting disorders: A group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. *Clin. Epigenetics* **2015**, *7*, 123. [[CrossRef](#)]
32. Uppal, A.; Wightman, S.C.; Mallon, S.; Oshima, G.; Pitroda, S.P.; Zhang, Q.; Huang, X.; Darga, T.E.; Huang, L.; Andrade, J.; et al. 14q32-encoded microRNAs mediate an oligometastatic phenotype. *Oncotarget* **2015**, *6*, 3540–3552. [[CrossRef](#)]
33. Xu, Z.; Huang, B.; Liu, J.; Wu, X.; Luo, N.; Wang, X.; Zheng, X.; Pan, X. Combinatorial anti-proliferative effects of tamoxifen and naringenin: The role of four estrogen receptor subtypes. *Toxicology* **2018**, *410*, 231–246. [[CrossRef](#)]
34. Ajj, H.; Chesnel, A.; Pinel, S.; Plénat, F.; Flament, S.; Dumond, H. An alkylphenol mix promotes seminoma derived cell proliferation through an ER $\alpha$ 36-mediated mechanism. *PLoS ONE* **2013**, *8*, e61758. [[CrossRef](#)]
35. Ariazi, E.A.; Taylor, J.C.; Black, M.A.; Nicolas, E.; Slifker, M.J.; Azzam, D.J.; Boyd, J. A New Role for ER $\alpha$ : Silencing via DNA Methylation of Basal, Stem Cell, and EMT Genes. *Mol. Cancer Res.* **2017**, *15*, 152–164. [[CrossRef](#)]
36. Chaudhri, R.A.; Olivares-Navarrete, R.; Cuenca, N.; Hadadi, A.; Boyan, B.D.; Schwartz, Z. Membrane estrogen signaling enhances tumorigenesis and metastatic potential of breast cancer cells via estrogen receptor- $\alpha$ 36 (ER $\alpha$ 36). *J. Biol. Chem.* **2012**, *287*, 7169–7181. [[CrossRef](#)]
37. Shi, L.; Dong, B.; Li, Z.; Lu, Y.; Ouyang, T.; Li, J.; Wang, T.; Fan, Z.; Fan, T.; Lin, B.; et al. Expression of ER- $\alpha$ 36, a novel variant of estrogen receptor  $\alpha$ , and resistance to tamoxifen treatment in breast cancer. *J. Clin. Oncol.* **2009**, *27*, 3423–3429. [[CrossRef](#)]
38. Fleischer, T.; Tekpli, X.; Mathelier, A.; Wang, S.; Nebdal, D.; Dhakal, H.P.; Sahlberg, K.K.; Schlichting, E.; Børresen-Dale, A.-L.; Borgen, E.; et al. DNA methylation at enhancers identifies distinct breast cancer lineages. *Nat. Commun.* **2017**, *8*, 1379. [[CrossRef](#)]
39. Xue, X.; Yang, Y.A.; Zhang, A.; Fong, K.; Kim, J.; Song, B.; Li, S.; Zhao, J.C.; Yu, J. LncRNA HOTAIR enhances ER signaling and confers tamoxifen resistance in breast cancer. *Oncogene* **2016**, *35*, 2746–2755. [[CrossRef](#)]
40. Borges, S.; Döppler, H.; Perez, E.A.; Andorfer, C.A.; Sun, Z.; Anastasiadis, P.Z.; Thompson, E.A.; Geiger, X.J.; Storz, P. Pharmacologic reversion of epigenetic silencing of the PRKD1 promoter blocks breast tumor cell invasion and metastasis. *Breast Cancer Res.* **2013**, *15*, R66. [[CrossRef](#)]
41. Yu, J.; Qin, B.; Moyer, A.M.; Nowsheen, S.; Liu, T.; Qin, S.; Zhuang, Y.; Liu, D.; Lu, S.W.; Kalari, K.R.; et al. DNA methyltransferase expression in triple-negative breast cancer predicts sensitivity to decitabine. *J. Clin. Invest.* **2018**, *128*, 2376–2388. [[CrossRef](#)]

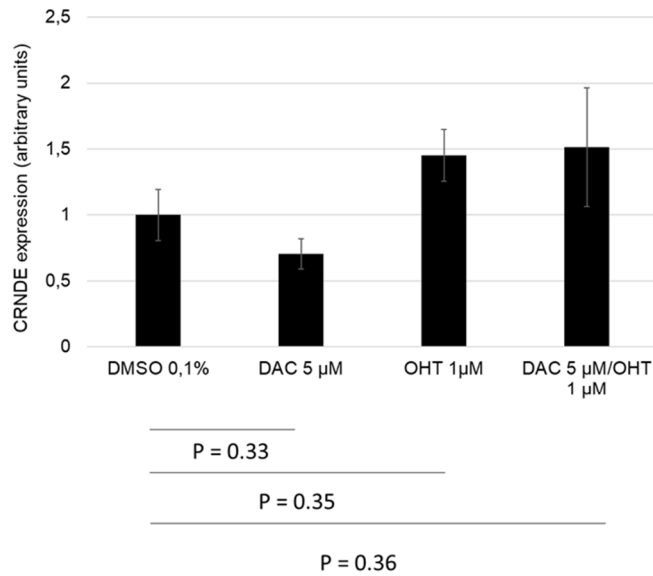
42. Pitta, C.A.; Papageorgis, P.; Charalambous, C.; Constantinou, A.I. Reversal of ER- $\beta$  silencing by chromatin modifying agents overrides acquired tamoxifen resistance. *Cancer Lett.* **2013**, *337*, 167–176. [[CrossRef](#)]
43. Zhu, Y.; Shao, S.; Pan, H.; Cheng, Z.; Rui, X. MicroRNA-136 inhibits prostate cancer cell proliferation and invasion by directly targeting mitogen-activated protein kinase 4. *Mol. Med. Rep.* **2018**, *17*, 4803–4810. [[CrossRef](#)]
44. Li, D.-X.; Fei, X.-R.; Dong, Y.-F.; Cheng, C.-D.; Yang, Y.; Deng, X.-F.; Huang, H.-L.; Niu, W.-X.; Zhou, C.-X.; Xia, C.-Y.; et al. The long non-coding RNA CRNDE acts as a ceRNA and promotes glioma malignancy by preventing miR-136-5p-mediated downregulation of Bcl-2 and Wnt2. *Oncotarget* **2017**, *8*, 88163–88178. [[CrossRef](#)]
45. Singh, A.; Willems, E.; Singh, A.; Ong, I.M.; Verma, A.K. Ultraviolet radiation-induced differential microRNA expression in the skin of hairless SKH1 mice, a widely used mouse model for dermatology research. *Oncotarget* **2016**, *7*, 84924–84937. [[CrossRef](#)]
46. Wallacides, A.; Chesnel, A.; Ajj, H.; Chillet, M.; Flament, S.; Dumond, H. Estrogens promote proliferation of the seminoma-like TCam-2 cell line through a GPER-dependent ER $\alpha$ 36 induction. *Mol. Cell. Endocrinol.* **2012**, *350*, 61–71. [[CrossRef](#)]
47. Iorio, M.V.; Ferracin, M.; Liu, C.-G.; Veronese, A.; Spizzo, R.; Sabbioni, S.; Magri, E.; Pedriali, M.; Fabbri, M.; Campiglio, M.; et al. MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* **2005**, *65*, 7065–7070. [[CrossRef](#)]
48. Paszek, S.; Gabło, N.; Barnaś, E.; Szybka, M.; Morawiec, J.; Kołacińska, A.; Zawlik, I. Dysregulation of microRNAs in triple-negative breast cancer. *Ginekol. Pol.* **2017**, *88*, 530–536. [[CrossRef](#)]
49. Braicu, C.; Raduly, L.; Morar-Bolba, G.; Cojocneanu, R.; Jurj, A.; Pop, L.-A.; Pileczki, V.; Ciocan, C.; Moldovan, A.; Irimie, A.; et al. Aberrant miRNAs expressed in HER-2 negative breast cancers patient. *J. Exp. Clin. Cancer Res.* **2018**, *37*, 257. [[CrossRef](#)]
50. Yerukala Sathipati, S.; Ho, S.-Y. Identifying a miRNA signature for predicting the stage of breast cancer. *Sci. Rep.* **2018**, *8*, 16138. [[CrossRef](#)]
51. Cheng, D.; He, H.; Liang, B. A three-microRNA signature predicts clinical outcome in breast cancer patients. *Eur. Rev. Med. Pharmacol. Sci.* **2018**, *22*, 6386–6395.
52. Kagami, M.; O'Sullivan, M.J.; Green, A.J.; Watabe, Y.; Arisaka, O.; Masawa, N.; Matsuoka, K.; Fukami, M.; Matsubara, K.; Kato, F.; Ferguson-Smith, A.C.; Ogata, T. The IG-DMR and the MEG3-DMR at human chromosome 14q32.2: Hierarchical interaction and distinct functional properties as imprinting control centers. *PLoS Genet.* **2010**, *6*, e1000992. [[CrossRef](#)]
53. Zhu, L.; Liu, Y.; Chen, Q.; Yu, G.; Chen, J.; Chen, K.; Yang, N.; Zeng, T.; Yan, S.; Huang, A.; et al. Long-Noncoding RNA Colorectal Neoplasia Differentially Expressed Gene as a Potential Target to Upregulate the Expression of IRX5 by miR-136-5P to Promote Oncogenic Properties in Hepatocellular Carcinoma. *Cell. Physiol. Biochem.* **2018**, *50*, 2229–2248. [[CrossRef](#)]
54. Li, B.; Wang, F.; Li, X.; Sun, S.; Shen, Y.; Yang, H. Hsa\_circ\_0008309 May Be a Potential Biomarker for Oral Squamous Cell Carcinoma. *Dis. Markers* **2018**, *2018*, 7496890. [[CrossRef](#)]
55. Duan, C.; Liu, Y.; Li, Y.; Chen, H.; Liu, X.; Chen, X.; Yue, J.; Zhou, X.; Yang, J. Sulfasalazine alters microglia phenotype by competing endogenous RNA effect of miR-136-5p and long non-coding RNA HOTAIR in cuprizone-induced demyelination. *Biochem. Pharmacol.* **2018**, *155*, 110–123. [[CrossRef](#)]
56. Dai, M.; Li, S.; Qin, X. Colorectal neoplasia differentially expressed: A long noncoding RNA with an imperative role in cancer. *OncoTargets Ther.* **2018**, *11*, 3755–3763. [[CrossRef](#)]
57. Wang, W.; Yuan, F.; Xu, J. The prognostic role of long noncoding RNA CRNDE in cancer patients: A systematic review and meta-analysis. *Neoplasma* **2019**, *66*, 73–82. [[CrossRef](#)]
58. Lin, C.-C.; Liu, L.-Z.; Addison, J.B.; Wonderlin, W.F.; Ivanov, A.V.; Ruppert, J.M. A KLF4-miRNA-206 autoregulatory feedback loop can promote or inhibit protein translation depending upon cell context. *Mol. Cell. Biol.* **2011**, *31*, 2513–2527. [[CrossRef](#)]
59. Sobin, L.H.; Wittekind, C.H. *TNM classification of malignant tumors, 6th ed*; John Wiley & Sons: Hoboken, NJ, USA, 2002.
60. Altman, D.G.; Bland, J.M. Standard deviations and standard errors. *BMJ* **2005**, *331*, 903. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



**Figure S1: Representative image of MCF-7 or MDA-MB-231 cells submitted to western blotting as described previously [10].** We used the primary antibodies anti-ERα66 (anti-ERα-F10, Santa Cruz Biotechnology) and anti-ERα36 (anti-ERα-G20, Santa Cruz Biotechnology). The anti-α-Tubulin (GTX102079, Genetex) was used as a loading control. Protein expression profiles were revealed with Clarity Western ECL Substrate (Biorad).



**Figure S2: Expression of CRNDE after 48h treatment with DAC, OHT or DAC+OHT as measured by RT-qPCR.**

MCF-7 cells were treated for 48h with DMSO 0.1% (as control), DAC 5 μM, OHT 1μM or by a combination OHT 1μM/DAC 5μM. CRNDE expression was measured by RT-qPCR analysis. Each bar represents mean ± S.E.M. N ≥ 3. \*: P <0.05



**Figure S3. Exon1/intron1 part of ESR1 genomic sequence.**

(>gi|224589818|ref|NC\_000006.11|:152129448-152130362 Homo sapiens chromosome 6, GRCh37.p10 Primary Assembly). Exon1 sequence is italicized and refers to the canonical ER $\alpha$ 66 transcribed sequence while exon 1' refers to the variant ER $\alpha$ 36. Nucleotides are numbered from ER $\alpha$ 36 transcription start site (TSS, arrow). Exon1/intron1 and exon1'/intron1 splice sites are annotated and CpG dinucleotides are in bold. The use of EMBOSS CpG plot ([http://www.ebi.ac.uk/Tools/seqstats/emboss\\_cpgplot/](http://www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot/)) allowed to design four ESR1 genomic regions as CpG islands on both sides of (i) the ESR-1 exon1/intron1 boundary (CpG islands No.1 and 2, respectively) and (ii) the ER $\alpha$ 36 transcription start site (CpG islands No.3 and 4). PCR primers used for CpG island amplification after bisulfite conversion are underlined.

## *Publication N°4*

*“A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology”*

B. Bastien, T. Boukhobza, H. Dumond, **C. Thiebaut**, H. Chakir, A. Gegout-Petit, A. Muller-Gueudin, Y. Shi.

Article soumis dans Biostatistics le 4 Juillet 2019.

Cette publication soumise à la revue Biostatistics et portée par nos collaboratrices de l'IECL de Nancy, propose une nouvelle méthodologie permettant de sélectionner et classer des variables de grande dimension, explicatives d'un phénomène biologique représenté par un petit nombre d'expériences. Cette méthodologie, qui correspond à l'algorithme "ARMADA" prévu pour fonctionner avec le logiciel R, fait intervenir successivement différentes approches mathématiques et statistiques : le regroupement des covariables, la décorrélation des covariables en utilisant l'analyse factorielle latente, la sélection par agrégation de méthodes adaptées et finalement des méthodes de classification.

Dans un premier temps, la méthode a été appliquée et testée sur des données transcriptomiques issues de 37 patients atteints d'un cancer du poumon non à petites cellules et ayant reçu un traitement de chimiothérapie. L'objectif est, dans ce cas, de sélectionner les ARNms covariants qui expliquent le taux de survie des patients en fonction du traitement reçu.

Dans un second temps, cette méthode a été validée sur des données transcriptomiques de 79 patientes atteintes d'un cancer du sein afin de définir des profils transcriptomiques et des réseaux de gènes, associés à de nouveaux biomarqueurs dans le but de mieux personnaliser les traitements.

C'est plus spécifiquement dans cette seconde application que s'inscrit mon travail de thèse car, grâce à "ARMADA", nous avons recherché des variables explicatives des variations d'expression d'ER $\alpha$ 36. Pour cela, nous avons déterminé par RT-qPCR l'expression de ce variant dans chacune des 79 tumeurs à notre disposition puis nous avons identifié les gènes dont les variations d'expression sont corrélées à celles d'ER $\alpha$ 36. Nous avons ainsi pu montrer que ces gènes sont majoritairement eux-mêmes régulés par 4 facteurs de transcription, impliqués dans la voie de signalisation Wnt/ $\beta$ -caténine, et par certains micro-ARNs décrits comme pro-tumoraux. Au final, l'algorithme "ARMADA" a permis de regrouper des patientes dont la tumeur mammaire exprime fortement ER $\alpha$ 36 et les gènes associés. Ainsi, ces patientes pourraient être traitées par des inhibiteurs de la signalisation Wnt ou par des modulateurs de micro-ARNs spécifiques et donc bénéficier de stratégies thérapeutiques personnalisées.



# A statistical methodology to select covariates in high-dimensional data under dependence. Application to the classification of genetic profiles in oncology.

BERANGERE BASTIEN

*Transgene S.A., Boulevard Gonthier d'Andernach, Parc d'innovation, CS80166, F-67405  
Illkirch-Graffenstaden Cedex - France*

TAHA BOUKHOBZA, HELENE DUMOND, CHARLENE THIEBAUT

*Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France*

HAFID CHAKIR, ANNE GEGOUT-PETIT, AURELIE MULLER-GUEUDIN\*, YAOJIE SHI

*Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France*

## SUMMARY

We propose a new methodology for selecting and ranking covariates associated with a variable of interest in a context of high-dimensional data under dependence but few observations. The methodology successively intertwines the clustering of covariates, decorrelation of covariates using Factor Latent Analysis, selection using aggregation of adapted methods and finally ranking. Simulations study shows the interest of the decorrelation inside the different clusters of covariates. We first apply our method to transcriptomic data of  $n = 37$  patients with advanced non-small-cell

\*aurelie.gueudin@univ-lorraine.fr

lung cancer who have received chemotherapy, to select the transcriptomic covariates that explain the survival outcome of the treatment. Secondly, we apply our method to  $n = 79$  breast tumor samples to define patient profiles for a new metastatic biomarker and associated gene network in order to personalize the treatments.

*Key words:* Aggregated methods, Correlated covariates selection, Genetic profiles, High dimension, Multiple testing procedures, Personalized medicine, Ranking, Variable selection.

## 1. INTRODUCTION

The purpose of personalized medicine is to select appropriate and optimal therapies based on the context of a patient's genetic content or other molecular or cellular analysis. One of the main challenges faced by biologist and mathematician consortium for the construction of explanatory models of multivariable biological processes, is the relatively low amount of experimental data available compared to the huge number of variables. The point is of great concern when the biological question deals with transcriptomic data in order to build gene networks and decipher the role of a rare isoform, for which no specific probe is currently available. In that context, the purpose of this paper is to propose a method to select the covariates that are linked with the outcome of a given therapy or a biological marker, among a set of more than tens of thousands covariates. For instance, if we consider a transcriptomic dataset of patients, we want to select the covariates associated to a variable of interest that could be the outcome of a given treatment or a promising biomarker of the treatment response. The variable of interest being known (treatment outcome or biological marker), the question is to find its link with the transcriptomic profile of the patients. We propose a methodology, that, firstly selects and ranks the transcriptomic covariates that are the most linked with the outcome treatment, and secondly, that visualises the profiles of the selected transcriptomic covariates, for all the patients of the study.

More generally, the problem to detect association between a variable of interest and many covariates has been tackled by many biologists and statisticians (Dudoit *and others* (2002); Aubert *and others* (2004); Bar-Hen *and others* (2005); Lê Cao *and others* (2011); Günther *and others* (2014)). A common example, coming from biology, is testing which of  $p$  genes' expression levels given in a dataset  $\mathbf{X}$  is linked significantly with a variable  $Y$ , which we will call the variable of interest. The variable of interest may be a binary variable like an outcome of treatment or it may be a quantitative variable such as a phenotype or physiological parameter. Sometimes, the aim of the biologist is not necessarily to detect exhaustively all the genes involved in his problem but to have a list of the most important of them in order to study their biological functions. For this purpose, it is interesting to rank the genes according to the strength of their link with the variable of interest. We will use the gene expression example for concreteness, but our aim is to propose a general methodology in a context of high dimensional data (the number  $p$  of covariates is in the order of thousands) while the total number  $n$  of samples could be small (for instance between 25 and 100).

In the context of transcriptomic data, the covariates are high dimensional and correlated. This correlation between covariates, in a high-dimensional context, has to be taken into account in the statistical analysis. Moreover, we are in a context of small sample size ( $n \ll p$ ). Thus, robustness of the statistical analysis has to be quantified.

We cite here some statistical methods that have been developed to select covariates in high-dimensional contexts. The state of the art about the control of false discoveries in multiple testing procedures is very extensive. The famous correction proposed by Bonferroni (1936) to control the Family Wise Error Rate (FWER) has been emulated and we can find a review about these methods in Dudoit *and others* (2004). Alternative methods focused on the control of the False Discovery Rate (FDR) (Benjamini and Hochberg (1995); Benjamini and Yekutieli (2001)), or of the local FDR (Efron (2005)) or the q-value (Storey (2002); Storey and Tibshirani

(2003); Storey (2003)). For a review (in french) of the methods, see Bar-Hen *and others* (2005). Regarding regression in the framework of high dimensional data ( $n \ll p$ ), many methods are available. For example, the PLS approach of Tenenhaus *and others* (2005) is a kind of principal component regression. The Lasso regression proposed by Tibshirani (1996) performs both variable selection and regularization in penalizing the sums of squares by the  $L_1$ -norm of the coefficients. This method has been derived for many kinds of problems like logistic-regression in the case of binary data (Meier *and others* (2008)), or network inference (Meinshausen and Bühlmann (2006)). Another versatile tool to select covariates in different non parametric contexts is given by the random forests, with the concept of importance of covariates (see for instance Genuer *and others* (2010)).

Another important characteristic of the data that has to be taken into account is the structure of covariance of the covariates. Most of the multiple testing corrections make the assumption of the independence between the tests. However it is well-known that omics data for instance are correlated by blocks. In the context of multiple testing, covariance between the covariates could bias the uniform repartition of the p-values under the null hypothesis and also inflates the variance of the estimation of the FDR (Friguet *and others* (2009); Friguet and Causeur (2011)). In Genuer *and others* (2010) it is also shown that despite the robustness of the random forests, importance of covariates calculated by random forests is perturbed by adding other correlated covariates. One of the ways to deal with dependence is to model it by latent factors; it is a way to reduce the information in supposing that the common information of the  $p$  covariates is given by  $q \ll p$  latent factors as Friguet *and others* (2009) and Friguet and Causeur (2011). More precisely, they propose a way to correct the data according to a regression link with the variable of interest  $Y$  in such a way that covariates are independent conditionally to  $Y$  (leading to the independence of the tests). After this correction, they propose a multiple testing procedure based on the method of Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001). This method of correction

will be called FAMT correction (for Factor Analysis for Multiple Testing) in the sequel.

However, the framework of FAMT is to consider the data  $\mathbf{X}$  as an only one block of correlated covariates and has to be adapted if  $\mathbf{X}$  is structured in several independent clusters of correlated covariates. As we will see in Section 3, the FAMT does not give good results if it is applied directly on the whole set of data  $\mathbf{X}$ , without taking into account its decomposition in clusters with strong within correlation. Then, we propose to identify the clusters of correlated covariates before performing FAMT correction on each of the clusters. The clustering of covariates as proposed by [Chavent \*and others\* \(2012\)](#) is a good way to arrange covariates into homogeneous clusters, i.e., groups inside of which covariates are strongly related to each other.

Our purpose in this paper is to propose a method adapted to the selection (and ranking) of correlated quantitative covariates associated with a variable of interest. For this, we propose a methodology that takes into account (1) the structure of correlation by clusters of covariates; (2) the correlation inside each cluster of correlated covariates.

Our methodology is divided in two steps: a pretreatment of the covariates (step 1) and a procedure of selection of the pretreated covariates (step 2). The pretreatment consists of (step 1.1) detecting the clusters of covariates by using the clustering of covariates proposed by [Chavent \*and others\* \(2012\)](#), and (step 1.2) applying a "decorrelation" between the covariates inside each cluster using the factor analysis proposed by [Friguet \*and others\* \(2009\)](#); [Friguet and Causeur \(2011\)](#); [Causeur \*and others\* \(2011\)](#). Their method performs a decorrelation of the covariates and calculates the corrected covariates suitable for statistical testing and/or regression.

After that pretreatment, we propose a procedure to select and rank the covariates, by combining different selection methods that take into account the nature of the outcome  $Y$  (qualitative or quantitative) and the high dimensional context (multiple testing procedures for the tests, penalised regression, ...). We define a score for each covariate, which is defined by the number of selections among all the selection methods involved in this step. This score can be used to classify

the covariates like in [Su and others \(2003\)](#).

The paper is organized as follows. In [Section 2](#), we detail the model and explain the principle of the main steps of our methodology: the pretreatment of the covariates and the construction of the covariates scores of selection. [Section 3](#) is dedicated to simulations studies in order to assess the interest of the proposed pretreatment on one hand and the good working of the whole selection strategy on the other hand. The simulations are performed in two different designs in the case where the variable of interest is binary. [Section 4](#) is dedicated to two real data analysis: the purpose of the first analysis is to select covariates that are linked with the outcome of a lung cancer treatment, whereas the second analysis selects covariates linked with a breast cancer biomarker. In both analysis, the selected covariates are used to define genetic profiles of patients. [Section 5](#) gives some conclusions and perspectives. The Supplementary Materials give a simulation study in the case where the variable of interest is a continuous quantitative variable ([Section 1](#) of Supplementary Materials). Technical details on the two real data applications are also given ([Sections 2 and 3](#) of the Supplementary Materials).

## 2. METHODOLOGY

### 2.1 Framework and model

We suppose that we have  $n$  i.i.d replications of  $(Y, \mathbf{X})$  where  $Y$  is the variable of interest, and  $\mathbf{X} = (X_1, X_2, \dots, X_p)$  is the vector of covariates, taking its values in  $\mathbb{R}^p$ . We make the assumption that the covariates are decomposed into  $K$  independent clusters:

$$\mathbf{X} = (\underbrace{X_1^{(1)}, \dots, X_{p_1}^{(1)}}_{\mathbf{X}^{(1)}}, \dots, \underbrace{X_i^{(k)}, \dots}_{\mathbf{X}^{(k)}}, \dots, \underbrace{X_{p_K}^{(K)}}_{\mathbf{X}^{(K)}}) = (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(K)}),$$

where  $p_1 + \dots + p_K = p$ .

On one hand, we model the dependence in the  $K$  clusters of covariates as in the framework of [Friguet and others \(2009\)](#): inside each cluster  $\mathbf{X}^{(k)}$ , the common information between the  $p_k$

covariates is modeled by regression on  $q_k$  latent factors:

$$X_i^{(k)} = \delta_i^{(k)}(Y) + \mathbf{b}_i^{(k)} \mathbf{Z}^{(k)} + \varepsilon_i^{(k)}, \quad \text{for } i = 1, \dots, p_k, \quad (2.1)$$

where  $\delta_i^{(k)}(Y)$  is a function of  $Y$ ,  $\mathbf{Z}^{(k)}$  is a random centered  $q_k$ -vector such that  $\mathbb{E}(\mathbf{Z}^{(k)} \mathbf{Z}^{(k)'}) = \mathbf{I}_{q_k}$ ,  $\mathbf{b}_i^{(k)}$  is a  $q_k$ -vector, and  $\varepsilon^{(k)} = (\varepsilon_1^{(k)}, \dots, \varepsilon_{p_k}^{(k)})$  is a random centered  $p_k$ -vector with independent components, and independent of  $\mathbf{Z}^{(k)}$ . The common information contained in  $\mathbf{X}^{(k)}$  is then concentrated in a small dimension space by  $q_k$  latent factors  $\mathbf{Z}^{(k)}$ . Under the model (2.1), we have:

$$\boldsymbol{\Sigma}^{(k)} := \mathbb{V}(\mathbf{X}^{(k)} | Y) = \mathbf{B}^{(k)} (\mathbf{B}^{(k)})' + \boldsymbol{\Psi}^{(k)} \quad (2.2)$$

$$\mathbb{V}(\mathbf{X}^{(k)} | Y, \mathbf{Z}^{(k)}) = \boldsymbol{\Psi}^{(k)} \quad (2.3)$$

$$\text{Cov}(\varepsilon_i^{(k)}, Z_j^{(k)}) = 0, \quad \forall i, j, k \quad (2.4)$$

where  $\boldsymbol{\Psi}^{(k)}$  is a diagonal  $p_k \times p_k$  matrix (the covariance matrix of  $\varepsilon^{(k)}$ ) and  $\mathbf{B}^{(k)}$  is a  $p_k \times q_k$  matrix of factor loadings  $\mathbf{b}_i^{(k)}$  (Equation (2.1), the  $\mathbf{b}_i^{(k)}$  being the  $i$ th row of  $\mathbf{B}^{(k)}$ ). In the decomposition given in Equation 2.2, the diagonal element  $\Psi_i^{(k)}$  is the specific variance of the response  $X_i^{(k)}$  while  $\mathbf{B}^{(k)} (\mathbf{B}^{(k)})'$  appears as the shared variance in the common factor structure. [Friguet and others \(2009\)](#) define the common variance by

$$\text{ComVar}^{(k)} = \frac{\text{trace}(\mathbf{B}^{(k)} (\mathbf{B}^{(k)})')}{\text{trace}(\boldsymbol{\Sigma}^{(k)})}. \quad (2.5)$$

On the other hand, we suppose that the specific informations at each cluster (that is vectors  $(\mathbf{Z}^{(k)}, \varepsilon^{(k)})_{1 \leq k \leq K}$ ) are independent, then, given  $Y$ , the covariance matrix of the whole vector of covariates has the form given by the Figure 1.

## 2.2 Main procedure

The procedure is decomposed in two steps: a pretreatment of the covariates (step 1) and a selection method of the covariates (step 2).

### 2.2.1 Step 1: pretreatment of data (clustering of covariates and decorrelation inside clusters).

The aim of this pretreatment is to perform a decorrelation of the covariates, to obtain corrected covariates that are suitable for testing and/or regression. Indeed, the correlation between covariates has an impact on all the classical selection procedures: the conventional methods, namely the multiple testing procedures (the p-value adjustment methods proposed by [Bonferroni \(1936\)](#), [Benjamini and Hochberg \(1995\)](#); [Benjamini and Yekutieli \(2001\)](#), or the q-value proposed by [Storey \(2002\)](#); [Storey and Tibshirani \(2003\)](#); [Storey \(2003\)](#), or the local FDR presented in [Aubert and others \(2004\)](#), [Bar-Hen and others \(2005\)](#)) are all built on the assumption that tests are independent. As a results, they are no longer promising if the independence is not verified. A very detailed discussion can be found in the Friguet's thesis ([Friguet \(2012\)](#)).

In estimating together the latent factors  $\mathbf{Z}^{(k)}$  and the coefficients of regressions ( $\mathbf{B}^{(k)}, \Psi^{(k)}$ ) by an E.M. algorithm in model (2.1), the FAMT procedure of [Friguet and others \(2009\)](#) can correct the covariates such that they are almost independent and as a result, suitable for multiple testing procedures or selection by regression or random forests. More precisely, the corrected data, noted  $X_i^{(k)\star} = X_i^{(k)} - \mathbf{b}_i^{(k)} \mathbf{Z}^{(k)} = \delta_i^{(k)}(Y) + \varepsilon_i^{(k)}$ ,  $i = 1, \dots, p_k$ , lead to a standard multiple regression problem where the errors  $\varepsilon_i^{(k)}$  are independent. Note that this correction of the data  $\mathbf{X}$  is done conditionally on the variable of interest  $Y$  ([Friguet and others \(2009\)](#)).

Of course, the whole vector  $\mathbf{X}$  satisfies assumption of Equation (2.1), and [Friguet and others \(2009\)](#) apply this decorrelation procedure on the whole set of covariates  $\mathbf{X}$ . But instead of applying Friguet's procedure on the whole set of covariates  $\mathbf{X}$ , we propose to first detect the different clusters  $(\mathbf{X}^{(k)})_{1 \leq k \leq K}$  and then to apply the decorrelation method on each cluster. Indeed, [Shi \(2016\)](#) has shown with some simulation studies that the decorrelation was degraded by the dimension of the vector of covariates, whereas it was better after the detection of the independent clusters. By this way, the covariates selection procedure can be highly improved by clustering of covariates (step 1.1) before applying factor analysis to correct the correlation within each cluster



(step 1.2), as it is shown in Section 3.

*Step 1.1: clustering of covariates.* We apply a clustering of covariates in the purpose to find clusters of correlated variables as we assumed in Section 2.1. We propose to use the algorithm of [Chavent and others \(2012\)](#) to cluster covariates into homogeneous clusters and thus to reveal structures. This algorithm maximizes an homogeneity criterion, where the homogeneity of a cluster is defined by the sum of squared Pearson correlations between the covariates present in the cluster and the first principal component of this cluster. This algorithm is expected to roughly find the highly correlated clusters of covariates as we assumed in the Section 2.1. The procedure proposes also a method (based on bootstrap resampling) to find the number  $K$  of clusters if it is unknown.

*Step 1.2: Factor analysis to correct dependency structure in each cluster.* As already explained in the beginning of this section, clustering is followed by decorrelation inside each cluster using the Friguet's procedure.

At the end of this pretreatment procedure, we obtain corrected data, noted  $\mathbf{X}_Y^*$  in the sequel. Note that  $\mathbf{X}_Y^*$  depends on  $Y$ . To simplify the notations,  $\mathbf{X}_Y^*$  will be noted  $\mathbf{X}^*$ .

*2.2.2 Step 2: Aggregation of statistical methods applied on the resulting dataset.* The statistical methods proposed in this part are not fixed and can be adapted by the practitioner according to its preferred selection methods and the characteristics of the data (nature of variable of interest  $Y$ , samples' sizes and so on).

The idea is the following: we choose several methods to select the pretreated covariates  $\mathbf{X}^*$ . We perform  $L$  methods, then each covariate  $X_i^*$  obtains a score  $S_i \in \{0, 1, \dots, L\}$  that is the number of selections among the  $L$  methods. By this way, the covariates can be ranked according to their link with the outcome  $Y$ .

For instance, in the examples developed in our simulation studies and in real data,  $Y$  is binary, the size of the samples are low and we choose eight different methods of selection: five different multiple testing procedures applied to the Wilcoxon test (Bonferroni, Benjamin-Hochberg,  $q$ -values, local FDR, FAMT), logistic regression penalised by Lasso, and two selections by random forests (threshold step and interpret step, see [Genuer and others \(2010\)](#)). The outcomes of this procedure are the scores  $S_i, i = 1, \dots, p$  which are integers included in  $[0, 8]$ . For example, if  $S_i = 8$ , then the corresponding variable has been selected by all the eight methods, whereas if  $S_i = 0$ , the corresponding variable has been selected by none of them. The scores can be used to rank the covariates according to the strength of their link with the variable of interest.

### 2.3 R Package *armada*

In the sequel, we call our procedure ARMADA for AggRegated Methods for covAriates selection under Dependence. Our procedure has been implemented in an R package, called `armada`, available on the CRAN (see [Muller-Gueudin and Gégout-Petit \(2019\)](#)). The package proposes also a graphical representation of the selected covariates, through heatmaps, as presented in Figures 7 and 10.

## 3. SIMULATIONS

We first explain the two different simulation designs in Section 3.1. We then describe the effect of the pretreatment in Section 3.2 and finally, we study the selection procedure in Section 3.3.

### 3.1 Simulation designs

We propose a simulation study with  $p = 1600$  covariates and sample size  $n = 60$ . We first describe how to create dependence in the covariates  $\mathbf{X}$ , then we present two simulation designs in a classification study. One design in a regression case is given in Section 1 of the Supplementary

Materials.

The covariates  $\mathbf{X} = (\mathbf{X}^{(k)})_{k=1,\dots,4}$  are clustered into four independent clusters, each of them containing  $p_k = 400$  covariates. For this, before to model the dependence with the outcome  $Y$ , we generate for each cluster  $k$ , a preliminary vector  $\tilde{\mathbf{X}}^{(k)}$  that is a gaussian 400-vector, with mean 0 and non-diagonal variance-covariance matrix  $\Sigma^{(k)}$ . The correlation between the covariates of  $\tilde{\mathbf{X}}^{(k)}$  inside the cluster  $k$  is designed by a factor analysis model described in Equation (2.2). More precisions on the simulation procedure of data with covariance design defined by (2.2) can be found in Friguet (2012). We simulate data with common variances  $\text{ComVar}^{(k)}$  equal to 0.8 in each cluster (recall that the common variance is defined in Equation (2.5)). Moreover, the numbers of latent factors in each cluster are  $(q^{(1)}, \dots, q^{(4)}) = (4, 6, 8, 10)$ .

Now, we create the dependence with outcome  $Y$  in perturbing some component of  $\tilde{\mathbf{X}}$ . The two following simulation designs consider an equiprobable two-class problem,  $Y \in \{0, 1\}$  (i.e.  $Y = 1$  for  $\frac{n}{2}$  subjects, and  $Y = 0$  for  $\frac{n}{2}$  subjects). In the two designs, there are either 160 (for design 1) or 240 (for design 2) influential covariates, whose links with the response variable  $Y$  have different intensities. More precisely, in both cases, the 10 first covariates of each cluster are the most strongly linked with the response variable  $Y$  and the strength of the link is decreasing in the successive groups of 10 influential covariates. The link between the influential covariates and  $Y$  is described for each design in the two following sections.

**3.1.1 Design 1.** This simulation design is inspired from the simulation design of Friguet *and others* (2009).  $Y$  is linked with 160 influential covariates in  $\mathbf{X}$ , the others being noise covariates. More precisely, in each cluster  $k = 1, \dots, 4$ , and for  $i = 1, \dots, 400$ ,

$$X_i^{(k)} = \tilde{X}_i^{(k)} + \delta_i \mathbf{1}_{Y=0}$$

with  $\delta = (\delta_i)_{i=1,\dots,400} = (\underbrace{1.5, \dots, 1.5}_{i=1,\dots,10}, \underbrace{1, \dots, 1}_{i=11,\dots,20}, \underbrace{0.75, \dots, 0.75}_{i=21,\dots,30}, \underbrace{0.5, \dots, 0.5}_{i=31,\dots,40}, \underbrace{0, \dots, 0}_{i=41,\dots,400})$ . In other words,  $Y$  is linked with the  $m_1 = 40$  first covariates of each cluster, and the  $m_0 = 360$  remaining

covariates of each cluster are independent of  $Y$ .

3.1.2 *Design 2.* This simulation design is inspired from the toys-data of [Genuer and others \(2010\)](#).  $Y$  is linked with 240 influential covariates in  $\mathbf{X}$ , the others being noise covariates. Let us define the simulation model by giving the conditional distribution of  $X_i$  given the value  $y$  of  $Y$ : in each cluster  $k = 1, \dots, 4$ , and for  $i = 1, \dots, 400$ ,

$$X_i^{(k)} = \tilde{X}_i^{(k)} + \delta_i^{(k)}(y)$$

where  $\delta_i(y)$  is a random variable.

- The relevant covariates are the  $m_1 = 60$  first covariates of each cluster. The distribution of the  $\delta_i^{(k)}(y)$  leading to the links between the relevant covariates and  $Y$  is given in Table 1.
- The  $m_0 = 340$  remaining covariates of each cluster are independent of  $Y$ :  $\delta_i^{(k)}(y) = 0$  whatever  $y$  for  $i = 61, \dots, 400$ .

We can remark that these two designs respect the covariance matrix given in Figure 1. The design 1 is exactly in the scope of our model given by Equation (2.1). Design 2 differs a little bit from the model of Equation (2.1) because  $\delta_i^{(k)}(Y)$  is a random function of  $Y$ . Note that in real data analysis, we don't know the model from which they are generated. It is why it is interesting to analyse the performance of our method on different kinds of simulated data.

### 3.2 Interest of our data pretreatment

In order to emphasize the interest of our data pretreatment, we compare the results of a Wilcoxon test after three different data pretreatments:

Procedure 1: nothing is done on the dataset  $\mathbf{X}$ .

Procedure 2: the covariates  $\mathbf{X}$  are decorrelated, taking  $Y$  into account, with the factor analysis procedure

of Friguet *and others* (2009); Causeur *and others* (2011), implemented in the R package FAMD. This gives a new dataset  $\mathbf{X}_Y^\dagger$ .

Procedure 3: the 4 clusters are estimated with the procedure of Chavent *and others* (2012), implemented in the R package `ClustOfVar`; then the covariates are decorrelated in each cluster, taking  $Y$  into account, with the factor analysis procedure of Friguet *and others* (2009); Causeur *and others* (2011), implemented in the R package FAMD. This gives a new dataset  $\mathbf{X}^*_Y$  obtained by the concatenation of the decorrelated clusters.

**Remark:** *our data pretreatment is the Procedure 3. We have supposed that the number of clusters is known. If that is not the case, the user can choose its own number of clusters by using the graphical tools of the `ClustOfVar` procedure (plots of the dendrogram).*

Our objective is to find out the differently expressed covariates in the two groups (groups  $Y = 0$  and  $Y = 1$ ) with sample sizes  $\frac{n}{2} = 30$ . For this, we perform Wilcoxon tests on each of the  $p$  pretreated covariates of the dataset (that is  $\mathbf{X}$  for Procedure 1,  $\mathbf{X}_Y^\dagger$  for Procedure 2,  $\mathbf{X}^*_Y$  for Procedure 3), given a three sets of  $p$  p-values. For each of these procedures, the selected covariates are those with p-values lower than 0.05. We compare these procedures on  $N = 100$  runs of  $(\mathbf{X}, Y)$ . For the comparison, we count the number of influential covariates that are correctly detected (this number is noted TP, for True Positive), this indicator gives an idea of the sensibility of the test after the procedure. To assess the specificity, we count the number of non-influential detected covariates (this number is noted FP, for False Positive). Note that the perfect method would detect all the influential covariates (that is 160 in design 1 and 240 in design 2) and no False Positive. However, according to the detection threshold chosen for the p-value, the expected number of FP is  $72 = 5\% \times (1600 - 160)$  for design 1 and  $68 = 5\% \times (1600 - 240)$  for design 2. The results are shown in Figure 2 for the design 1 and in Figure 3 for the design 2.

If we analyse the results given by Figures 2 and 3, we can see that Procedure 1 is in fact the one that has the lowest rate of FP but its power is also the poorest whatever the design. Our

Procedure reduces the mean and the variability of the distributions of the false positive rates, in comparison to the Procedure 2 (i.e. the FAMT procedure). The power of our Procedure is comparable with Procedure 2. This results show the interest of our proposed pretreatment before performing selection.

### 3.3 Results of the whole method (pretreatment and selection)

In order to describe the performances of our method, we show in Figures 4 and 5 the mean ARMADA obtained on the  $N = 100$  runs of  $(\mathbf{X}, Y)$  for each design. The scores are given for all the covariates individually, and also by group of influential and noise covariates (the groups of influential covariates are noted by "1.5", "1", "0.75", "0.5" in the design 1, and by "(0.7,3)", "(0.7,2)", "(0.7,1)", etc. in the design 2 (see Section 3.1); the group of noise covariates is noted by "-").

We can see on the Figures 4 and 5 that the scores give a clear ranking of the covariates, according to the strength of their link with the response variable  $Y$ . The highest scores are obtained by the covariates which are the most strongly linked with the response variable  $Y$ . The ARMADA method is particularly performant in the design 1, the mean score clearly distinguishes the five groups of covariates according to their link with  $Y$ . The distribution of the individual scores inside each group is given by the boxplots. For design 1, scores clearly separate the influential covariates from the others; and inside the influential covariates the two first groups are clearly separated of the last one. The method is not so performant for the design 2 probably because we are not exactly in the model of the study (Equation (2.1)) but also because the strength of the link with  $Y$  is low excepted for the two first groups of covariates that have scores which are well separated from the others by the selection method. Whatever the design, we can precise that around 95% of the noise covariates obtained an ARMADA score that was exactly 0.

### 3.4 Comparison with other selection methods

We propose the following selection criterion in our procedure: the selected covariates are those with scores greater or equal to 1.

We compare this selection procedure with two other selection methods:

- the Wilcoxon test: the selected covariates are those with raw-p-values (i.e. p-values without any correction) lower than 0.05,
- the FAMT procedure (*Causeur and others (2011)*): the selected covariates are those with adjusted p-values lower than 0.05.

To compare the three selection methods, Tables 2 and 3 give the rates of selection for each group of influential covariates, and for the group of noise covariates.

- We can see that our method respect the expected rate of false positives that is not the case for the FAMT method which exhibits a greater rate of 10 %.
- Moreover, our method gives the best results in the design 1. The rate of selection of the influential covariates is very good compared with the other methods even if the strength of the link is poor.
- In the design 2, our method is competitive with the FAMT procedure for the detection of influential covariates, but again FAMT procedure has more false positives than ours.

Finally, we can conclude with the ROC curves given in Figure 6 that our method outperforms the two others selection methods (the ordinates of the points of the ARMADA ROC curve are all higher than the ordinates of the points of the two other ROC curves). Note that the ROC curves of the design 2 give the impression that our method is not competitive with the two others, but this is only caused by the fact that we have traced a solid line between the points  $(1-\text{specificity}, \text{sensibility})_{\text{ARMADA score}=0}$  and  $(1-\text{specificity}, \text{sensibility})_{\text{ARMADA score}=1}$ . The ROC

curves have been obtained by the mean of the  $N = 100$  ROC curves obtained in the  $N = 100$  runs of  $(\mathbf{X}, Y)$ .

#### 4. APPLICATION TO REAL DATA

In this section we will show two applications of our method, both in oncology. The first one concerns the selection of transcriptomic covariates linked with the outcome of a chemotherapy for lung cancer. The second one concerns the selection of covariates linked with a quantitative biomarker ER $\alpha$ 36 in breast cancer.

##### 4.1 *Outcome of chemotherapy for advanced non-small-cell lung cancer*

We apply our method on transcriptomic data of  $n = 37$  patients with advanced non-small-cell lung cancer, who have received chemotherapy. Even if we are aware of the fact that chemotherapy is not a target therapy, the problematic is really to select suitable transcriptomic covariates in the purpose to detect profiles associated with the effect of a treatment. For each patient, we have 51336 transcriptomic covariates, and its survival status: 24 patients whose death occurred before 12 months and 13 patients whose death occurred after 12 months. This criteria of death before one year is very common in clinical trials. We applied a first filtering of the covariates, where we decided to ignore the covariates for which the Wilcoxon test does not detect a difference between the 24 patients whose survival time is lower than 12 months and the 13 other ones (we eliminate covariates with Wilcoxon-pvalue greater than 0.05). After this filtering we obtained a dataset with  $n = 37$  patients and  $p = 6810$  covariates. In the pretreatment step, we found that the  $p = 6810$  covariates are decomposed in 3 independent group of covariates.

4.1.1 *Classification study.* In a first time, the biological question was to find the genes which can explain a survival time greater or lower than 12 months. We then consider a binary response



variable  $Y$ :  $Y = 1$  for the 24 patients whose survival time is lower than 12 months and  $Y = 0$  for the 13 patients whose survival time is greater than 12 months. The results are shown in Table 4. We see that 10 covariates are particularly important, with a score equal to 7, whereas 2827 covariates have a score equal to 0, and 3983 covariates have a score greater or equal to 1.

It is clear that, the biologist will not focus on the 3983 covariates with a positive score. But the method clearly gives a hierarchy between the genes and it is sure that the function of the 10 genes with a score at 7 has to be studied to understand its link with the "success" of the treatment.

*4.1.2 Regression study.* As the survival time was known for all the 37 patients without any censoring, we also apply our method on the same dataset (6810 covariates) but here,  $Y$  is the survival time. We then have a regression problem. We have used eight selection methods in Step 2 of our method: five different multiple testing procedures applied to the Pearson correlation test (Bonferroni, Benjamin-Hochberg, q-values, local FDR, FAMT), regression penalised by Lasso, and two selections by random forests (threshold step and interpret step, see [Genuer and others \(2010\)](#)). The results are given in Table 5.

*4.1.3 Joint analysis of the regression and classification studies.* Joint results of the classification and regression studies are given in Table 6. This table is a little disappointing, because regression and classification do not select the same covariates. Whatever, among the covariates with a C-score (score in the classification case) equal to 7, there is only one with a R-score (score in the regression case) lower than 4 (equal to 0!). But these two analyses are not looking for the same kind of link with the covariates. Moreover, these two approaches give two tools to detect influential covariates. We can combine these two approaches and consider the covariates that are selected by at least one approach, or consider the covariates that are selected by both of them. In the Figure 7, we show the heatmap of the selected covariates which have a classification score

**and** a regression score greater than five. For the visualisation of the results, we then build an heatmap obtained thanks to the R package `heatmaply` after co-clustering of the survival times (on the  $x$ -axis) and of the covariates (on the  $y$ -axis) with the function `hclust` (Figure 7).

The visualisation of the co-clustering of the selected genes and the survival leads to the distinction of three different groups of patients (noted  $P_1$ ,  $P_2$ ,  $P_3$  in Figure 7) of respective sizes 7, 8, 22 from the left to the right of the  $x$ -axis. The co-clustering identifies also two clusters of genes (noted  $G_1$  and  $G_2$  for simplicity). All the people except 2 of the two first group  $P_1$  and  $P_2$  have a life status  $Y = 1$  (among the two exceptions, one is at the threshold with a survival of 11.5 months), all of the people of the third group  $P_3$  have a life status  $Y = 0$ . The selected covariates clearly discriminates groups  $P_1$  and  $P_3$ . Indeed, the patients of the group  $P_1$  have a low expression of the covariates in  $G_1$  and a high expression of the covariates in  $G_2$  and the inverse for group  $P_3$ . Patients of group  $P_2$  have intermediate expressions according the two others groups.

As the number of patients  $n = 37$  is small compared to the number of covariates even after filtering ( $p = 6810$ ), we have checked our results with a bootstrap study. The results (reported in the Section 2 of the Supplementary Materials) show that our method is robust: the distributions of the bootstrapped scores faithfully reproduce the original scores.

#### 4.2 *Biological network involving ER $\alpha$ 36 in breast cancer*

ER $\alpha$ 36 is a variant of the oestrogen receptor  $\alpha$  encoded by the ESR1 locus and expressed only in humans (Wang *and others* (2005)). ER $\alpha$ 36 expression and activity have been mainly studied in vitro and in vivo in the context of breast cancer. However, due to the lack of comprehensive transcriptomic data that include ER $\alpha$ 36, only sparse information is available on factors that could act up- and downstream ER $\alpha$ 36 in biological networks. Our challenge from a statistical point of view was to explain the ER $\alpha$ 36 expression variation obtained in a small number of breast tumors from a large number of potential explanatory variables that correspond to the 54676

transcriptomic probes. For this, we analysed the biological network involving ER $\alpha$ 36 through the use of 4 sets of Affymetrix transcriptomic data obtained from breast tumors of different molecular subtypes: the triple negative (noted TN), ERa66+, PR+ and PR- datasets (details are provided in Section 3 of the Supplementary Materials).

The analysis was performed in three steps (Figure 8).

**Step 1, clinical data completion.** These 4 transcriptomic datasets were completed by the measurement of ER $\alpha$ 36 expression level in each tumor. Biological details are given in Section 3 of the Supplementary Materials.

**Step 2a, statistical analysis.** To explain the ER $\alpha$ 36 expression variation obtained in a small number of tumors from a large number of potential explanatory variables, we used the R package *armada*, that allowed to select, among the 54676 initial probes, a few hundreds of genes whose expression is supposed to be correlated to that of ER $\alpha$ 36 (ARMADA score  $\geq 1$ ). We obtained four lists of respectively 711, 277, 319 and 369 probes correlated to the expression of ER $\alpha$ 36 in the TN, ERa66+, PR+ and PR- groups.

**Step 2b, functional analysis.** From these four lists of transcriptomic probes, we carried out a functional analysis using the MSigDB database (GSEA). In particular, we looked for transcription factors and microRNAs involved in the regulation of the majority of genes from the different lists (TN, ERa66+, PR+ and PR-). The results indicated that four transcription factors: NFAT, FOXO4, SP1 and LEF-1, were common regulators and could therefore be mediators of the ER $\alpha$ 36 effect in all breast tumor subtypes. Interestingly, a study by [Antanavičiūtė and others \(2017\)](#) has shown that these transcription factors FOXO4, SP1 and LEF-1 are transcriptional hallmarks characteristic of cancer cells and associated to the Wnt signaling pathway (involved in metastasis and maintenance of cancer stem cells). Regarding the analysis of microRNAs, the results indicated that the majority of the genes whose expression correlated to ER $\alpha$ 36 one in the TN set were regulated by the microRNAs: hsa-miR-106B, hsa-miR21 and hsa-miR-29A, listed as oncogenic

microRNAs involved in metastatic processes, survival and self-sufficiency in growth factors of mammary tumors (Le Quesne and Caldas (2010)). These results recalled those of a previous study of Chamard-Jovenin *and others* (2015), which showed that a high ER $\alpha$ 36 expression in mammary tumors is associated with an increased metastatic potential and an estrogen-independent tumor growth.

**Step 2c, experimental validation.** Subsequently, we provided an experimental confirmation of the biological reliability of the results: the correlation between the expression of ER $\alpha$ 36 and that of ZEB1, FZD7, ZIC1 and TCF7LD genes, identified by `armada` as correlated to ER $\alpha$ 36 in all tumor sets, was verified in vitro by RT-qPCR in two breast cancer cell lines (MCF-7 (ER $\alpha$ 66+, PR+, PR-) and MDA-MB-231 (TN)). The results of Figure 9 confirmed the correlation (positive or negative) between the expression of ER $\alpha$ 36 and that of the genes identified by `armada` in the both cell lines.

**Step 3, tumor classification according to ER $\alpha$ 36 expression level.** The final goal of our study was to identify the molecular signatures accounting for the ER $\alpha$ 36 expression level in the four different sets of tumors. These signatures were identified thanks to the R package `heatmaply` after co-clustering of the ER $\alpha$ 36 expression (on the  $x$ -axis) and of the covariates (on the  $y$ -axis) with the function `hclust`. For each of the four tumor datasets, a heatmap was built which accounted for the expression level of the genes correlated to ER $\alpha$ 36. Thanks to the associated dendrogram, different classes of tumors were defined and characterized by both the level of ER $\alpha$ 36 expression and an associated molecular signature. The Figure 10 illustrates the results for the study on the dataset ER $\alpha$ 66+: two classes of tumors were identified, called ER $\alpha$ 36<sup>++</sup> and ER $\alpha$ 36<sup>-</sup>. Taken together, the `armada` package helped to cluster patients which breast tumor highly express ER $\alpha$ 36 and associated genes. These patients could be treated by Wnt signaling inhibitors or specific microARN modulators and therefore benefit such promising new personalized medicine.

## 5. CONCLUSION AND PERSPECTIVES

We have proposed a new methodology which is able to select the covariates (here the genes) that are linked with a variable of interest (here the treatment of an outcome or a biological marker). The method is of particular interest in the high dimensional case and when the covariates are correlated. The algorithms corresponding to this method are available through the R package *armada*. After this selection obtained with our method, it is then easy to visualise the selected genes (or probes) for all the patients, and to classify the genetic profiles of patients with respect to their treatment outcome or biological marker. In the study of the treatment by chemotherapy in the advanced non-small-cell lung cancer, we have identify three types of genetic profiles defined with two clusters of genes. In the study of the mammary tumors, the covariates selection allows the biologist to study the functional role of selected probes and also to classify tumors and associated transcriptomic signatures. This kind of results is very promising for the identification of new therapeutic targets and the development of more efficient and personalized anti-cancer treatment.

## 6. SUPPLEMENTARY MATERIALS

The online Supplementary Materials give additional simulations on a regression model where  $Y$  is a continuous quantitative variable (Section 1 of Supplementary Materials), some further analysis of the lung cancer data presented in Section 4.1 (Section 2 of Supplementary Materials), and technical informations on the biological material used in Section 4.2 (Section 3 of Supplementary Materials).

## ACKNOWLEDGMENTS

The first case study of Section 4.1 comes from Transgene team thanks to Bérangère Bastien. Taha Boukhobza, Hélène Dumond and Charlène Thiebaut conducted the biological study of Section 4.2

and performed the functional analysis that follows the covariates selection with **armada**. Stefano Cairo and Olivier Déas ; XenTech, Genopole - Campus 3 ; 4 rue Pierre Fontaine - 91000 Evry (France) provided the TN dataset; Aurélien De Reynies, Ivan Bieche and Jacqueline Métral from the Carte d'Identité des Tumeurs program provided the access to transcriptomic raw data and biological samples (ERa66+, PR+, PR- datasets). The statistical methodology has been developed by Anne Gégout-Petit and Aurélie Muller-Gueudin, and thanks to Yaojie Shi and Hafid Chakir during their Master internships.

## REFERENCES

- ANTANAVIČIŪTĖ, IEVA, MIKALAYEVA, VALERYIA, CESLEVIČIENĖ, IEVA, MILAŠIŪTĖ, GINTARĖ, SKEBERDIS, VYTENIS ARVYDAS AND BORDEL, SERGIO. (2017). Transcriptional hallmarks of cancer cell lines reveal an emerging role of branched chain amino acid catabolism. *Scientific reports* **7**(1), 7820.
- AUBERT, JULIE, BAR-HEN, AVNER, DAUDIN, JEAN-JACQUES AND ROBIN, STÉPHANE. (2004). Determination of the differentially expressed genes in microarray experiments using local fdr. *BMC bioinformatics* **5**(1), 1.
- BAR-HEN, AVNER, DAUDIN, JEAN-JACQUES AND ROBIN, STÉPHANE. (2005). Comparaisons multiples pour les microarrays. *Journal de la Société Française de Statistique* **146**(1-2), 45–62.
- BENJAMINI, YOAV AND HOCHBERG, YOSEF. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.
- BENJAMINI, YOAV AND YEKUTIELI, DANIEL. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 1165–1188.

- BONFERRONI, CARLO EMILIO. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **8**, 3–62.
- CAUSEUR, DAVID, FRIGUET, CHLOÉ, HOUEE-BIGOT, MAGALI AND KLOAREG, MAELA. (2011). Factor analysis for multiple testing (famt): an R package for large-scale significance testing under dependence. *Journal of Statistical Software*. **40**(14), 19.
- CHAMARD-JOVENIN, CLÉMENCE, JUNG, ALAIN C, CHESNEL, AMAND, ABECASSIS, JOSEPH, FLAMENT, STÉPHANE, LEDRAPPIER, SONIA, MACABRE, CHRISTINE, BOUKHOBZA, TAHA AND DUMOND, HÉLÈNE. (2015). From  $ER_{\alpha 66}$  to  $ER_{\alpha 36}$ : a generic method for validating a prognosis marker of breast tumor progression. *BMC systems biology* **9**(1), 28.
- CHAVENT, MARIE, KUENTZ, VANESSA, LIQUET, BENOIT AND SARACCO, L. (2012). Clustofvar: an R package for the clustering of variables. *Journal of Statistical Software* **50**, 91–116.
- DUDOIT, SANDRINE, VAN DER LAAN, MARK J AND POLLARD, KATHERINE S. (2004). Multiple testing. part I. single-step procedures for control of general type I error rates. *Statistical Applications in Genetics and Molecular Biology* **3**(1), 1–69.
- DUDOIT, SANDRINE, YANG, YEE HWA, CALLOW, MATTHEW J AND SPEED, TERENCE P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica sinica*, 111–139.
- EFRON, BRADLEY. (2005). *Local false discovery rates*. Division of Biostatistics, Stanford University.
- FRIGUET, CHLOE. (2012). Impact of dependence in large-scale multiple testing [Ph.D. Thesis]. Université de Bretagne-Sud.
- FRIGUET, CHLOÉ AND CAUSEUR, DAVID. (2011). Estimation of the proportion of true null hy-

- potheses in high-dimensional data under dependence. *Computational Statistics & Data Analysis* **55**(9), 2665–2676.
- FRIGUET, CHLOÉ, KLOAREG, MAELA AND CAUSEUR, DAVID. (2009). A factor model approach to multiple testing under dependence. *Journal of the American Statistical Association* **104**(488), 1406–1415.
- GENUER, ROBIN, POGGI, JEAN-MICHEL AND TULEAU-MALOT, CHRISTINE. (2010). Variable selection using random forests. *Pattern Recognition Letters* **31**(14), 2225–2236.
- GÜNTHER, OLIVER P, SHIN, HEESUN, NG, RAYMOND T, McMASTER, W ROBERT, McMANUS, BRUCE M, KEOWN, PAUL A, TEBBUTT, SCOTT J AND LÊ CAO, KIM-ANH. (2014). Novel multivariate methods for integration of genomics and proteomics data: applications in a kidney transplant rejection study. *Omics: a journal of integrative biology* **18**(11), 682–695.
- LÊ CAO, KIM-ANH, BOITARD, SIMON AND BESSE, PHILIPPE. (2011). Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC bioinformatics* **12**(1), 1.
- LE QUESNE, JOHN AND CALDAS, CARLOS. (2010). Micro-rnas and breast cancer. *Molecular oncology* **4**(3), 230–241.
- MEIER, LUKAS, VAN DE GEER, SARA AND BÜHLMANN, PETER. (2008). The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**(1), 53–71.
- MEINSHAUSEN, NICOLAI AND BÜHLMANN, PETER. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 1436–1462.
- MULLER-GUEUDIN, AURELIE AND GÉGOUT-PETIT, ANNE. (2019). *armada: A Statistical*



- Methodology to Select Covariates in High-Dimensional Data under Dependence*. R package version 0.1.0.
- SHI, YAOJIE. (2016). Microarray data analysis : feature selection, clustering and prediction. *Master Internship report*, 1–40.
- STOREY, JOHN D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(3), 479–498.
- STOREY, JOHN D. (2003). The positive false discovery rate: a bayesian interpretation and the q-value. *The Annals of Statistics*, 2013–2035.
- STOREY, JOHN D AND TIBSHIRANI, ROBERT. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**(16), 9440–9445.
- SU, YANG, MURALI, TM, PAVLOVIC, VLADIMIR, SCHAFFER, MICHAEL AND KASIF, SIMON. (2003). Rankgene: identification of diagnostic genes based on expression data. *Bioinformatics* **19**(12), 1578–1579.
- TENENHAUS, MICHEL, VINZI, VINCENZO ESPOSITO, CHATELIN, YVES-MARIE AND LAURO, CARLO. (2005). PLS path modeling. *Computational Statistics & Data Analysis* **48**(1), 159–205.
- TIBSHIRANI, ROBERT. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.
- WANG, ZHAOYI, ZHANG, XINTIAN, SHEN, PENG, LOGGIE, BRIAN W, CHANG, YUNCHAO AND DEUEL, THOMAS F. (2005). Identification, cloning, and expression of human estrogen receptor- $\alpha$ 36, a novel variant of human estrogen receptor- $\alpha$ 66. *Biochemical and biophysical research communications* **336**(4), 1023–1027.

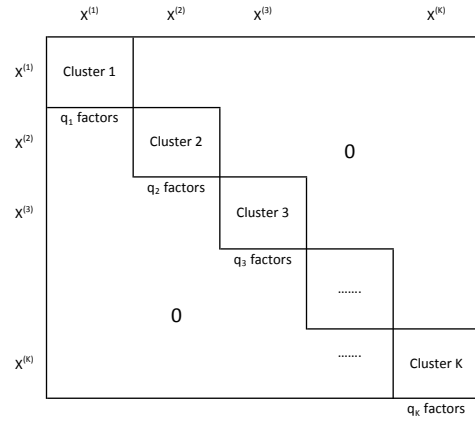


Fig. 1. Covariance structure of covariates given  $Y$

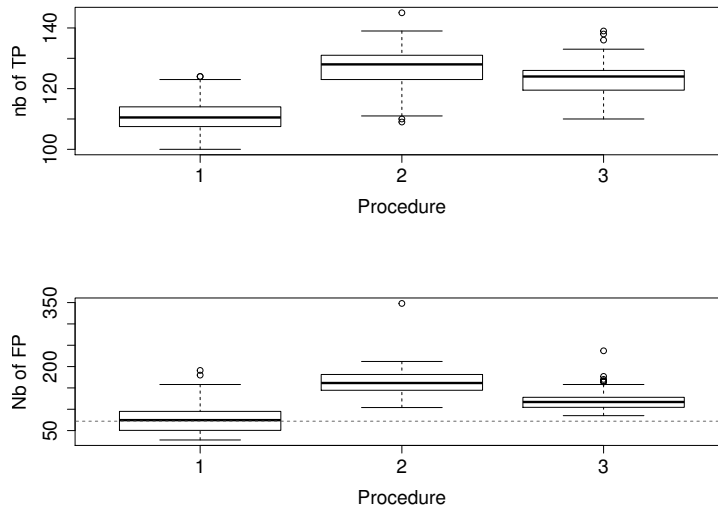


Fig. 2. Number of true positive tests (top), false positive tests (bottom) in the design 1 according to the different pretreatment procedures (1: Nothing, 2: FAMT, 3: clustering followed by FAMT in each cluster). Dotted lines: expected number of FP. Boxplots are calculated on  $N = 100$  runs.

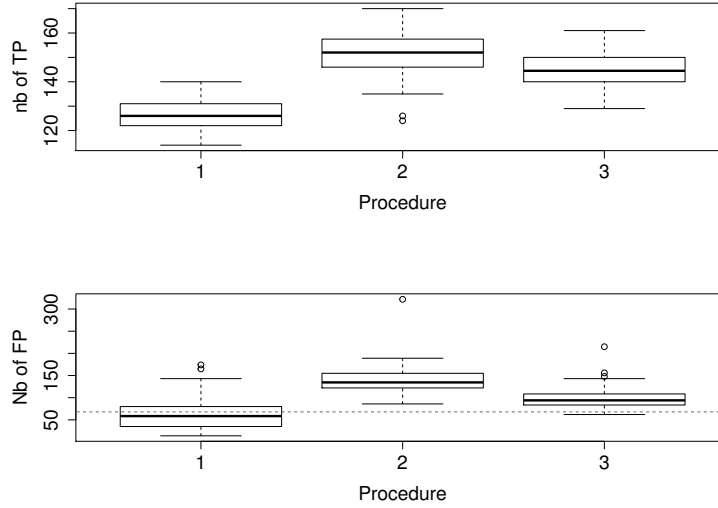


Fig. 3. Number of true positive tests (top), false positive tests (bottom) in the design 2 according to the different pretreatment procedures (1: Nothing, 2: FAMT, 3: clustering followed by FAMT in each cluster). Dotted lines: expected number of FP. Boxplots are calculated on  $N = 100$  runs.

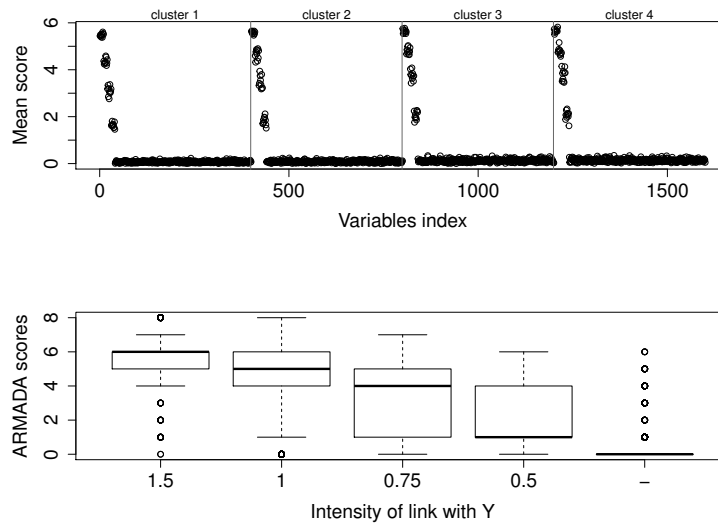


Fig. 4. Top: mean of the ARMADA scores obtained by all the covariates. Bottom: boxplot of the scores of the covariates, ranked by levels of link with  $Y$ . Means and boxplots are calculated on  $N = 100$  runs. Simulation in the design 1.

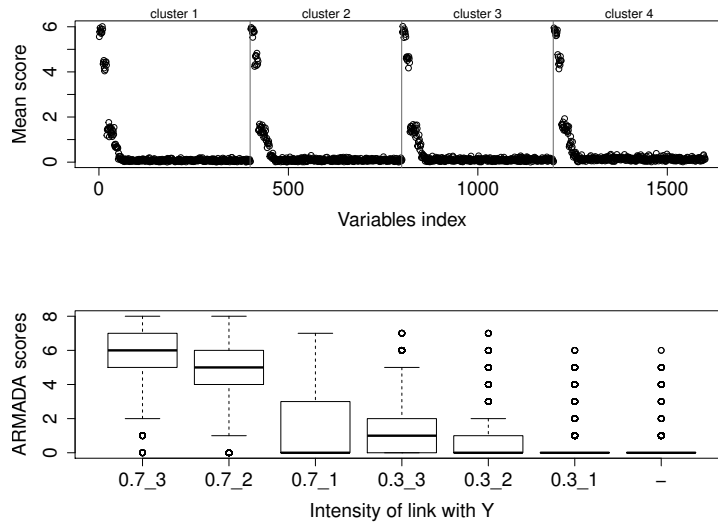


Fig. 5. Top: mean of the ARMADA scores obtained by all the covariates. Bottom: boxplot of the scores of the covariates, ranked by levels of link with  $Y$ . Means and boxplots are calculated on  $N = 100$  runs. Simulation in the design 2.

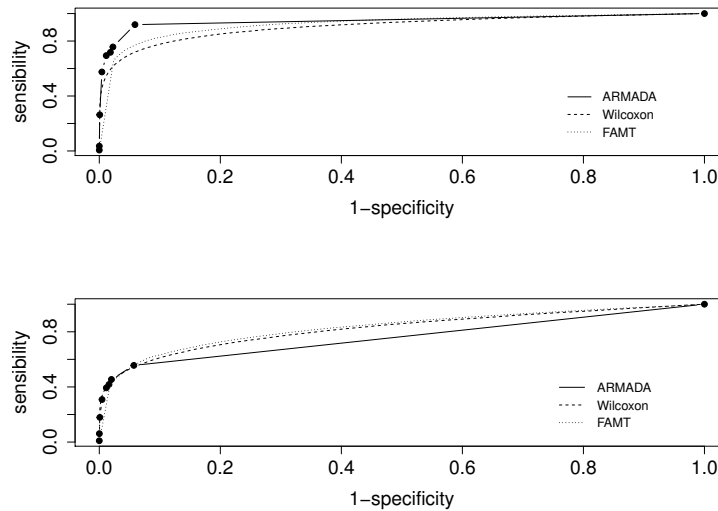


Fig. 6. ROC curves for the three selection methods, in the case of design 1 (top), design 2 (bottom).

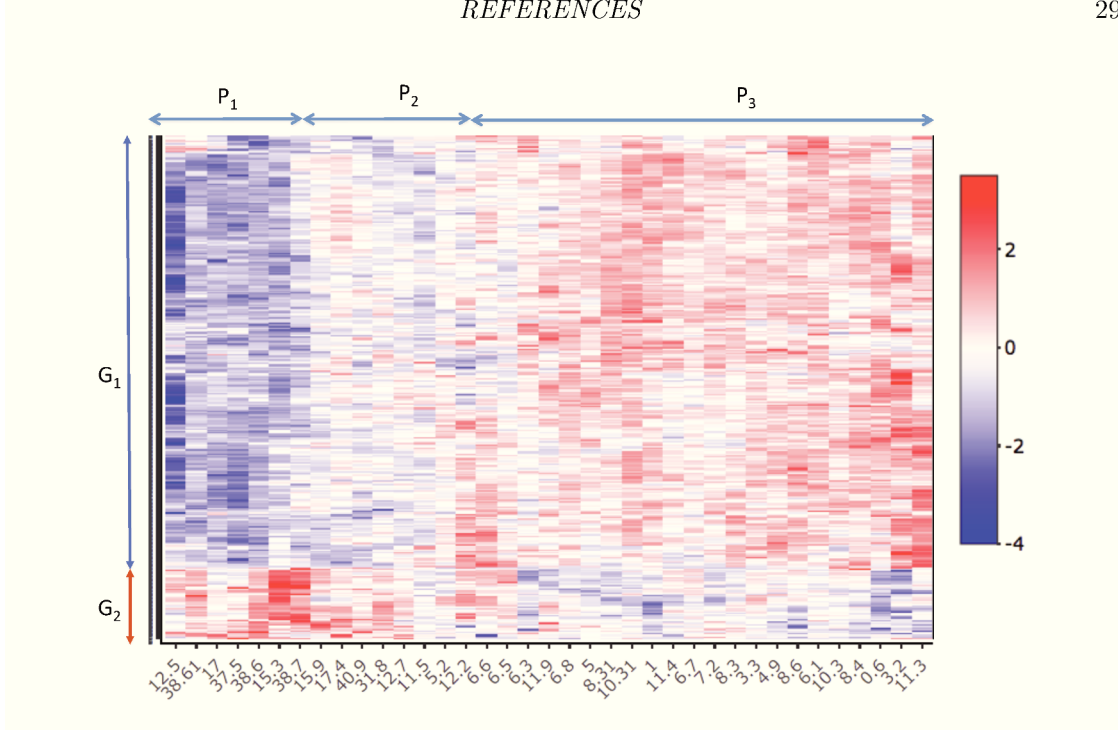


Fig. 7. Heatmap of the 342 covariates which have ARMADA scores greater or equal to 5 in **both** classification **and** regression studies. Each column corresponds to one patient. The  $x$ -axis represents the patients (marked with their survival time) and the  $y$ -axis the covariates. Color gives the normalised expression of the covariates.

Table 1. Links between the relevant covariates and  $Y$  in the simulation design 2. The notation  $\delta_i^{(k)} \sim 0.7\mathcal{N}(3y, 1) + 0.3\mathcal{N}(0, 1)$  means that, with probability 0.7,  $\delta_i^{(k)} \sim \mathcal{N}(3y, 1)$ , and with probability 0.3,  $\delta_i^{(k)} \sim \mathcal{N}(0, 1)$ .

$i$	model for $\delta_i^{(k)}$
for $i = 1, \dots, 10$	$\delta_i^{(k)} \sim 0.7\mathcal{N}(3y, 1) + 0.3\mathcal{N}(0, 1)$
for $i = 11, \dots, 20$	$\delta_i^{(k)} \sim 0.7\mathcal{N}(2y, 1) + 0.3\mathcal{N}(0, 1)$
for $i = 21, \dots, 30$	$\delta_i^{(k)} \sim 0.7\mathcal{N}(y, 1) + 0.3\mathcal{N}(0, 1)$
for $i = 31, \dots, 40$	$\delta_i^{(k)} \sim 0.3\mathcal{N}(3y, 1) + 0.7\mathcal{N}(0, 1)$
for $i = 41, \dots, 50$	$\delta_i^{(k)} \sim 0.3\mathcal{N}(2y, 1) + 0.7\mathcal{N}(0, 1)$
for $i = 51, \dots, 60$	$\delta_i^{(k)} \sim 0.3\mathcal{N}(y, 1) + 0.7\mathcal{N}(0, 1)$

Table 2. Results of the  $N = 100$  runs in the design 1: rates of selection of the different groups of influential and noise covariates by the ARMADA method, the Wilcoxon test and the FAMT procedure. The corresponding standard deviations are given in brackets.

	ARMADA	Wilcoxon	FAMT
1.5	0.99 (0.04)	0.99 (0.07)	0.99 (0.02)
1	0.97 (0.15)	0.85 (0.35)	0.95 (0.20)
0.75	0.91 (0.27)	0.62 (0.48)	0.82 (0.38)
0.5	0.79 (0.40)	0.33 (0.47)	0.52 (0.49)
-	0.05 (0.23)	0.05 (0.22)	0.10 (0.30)

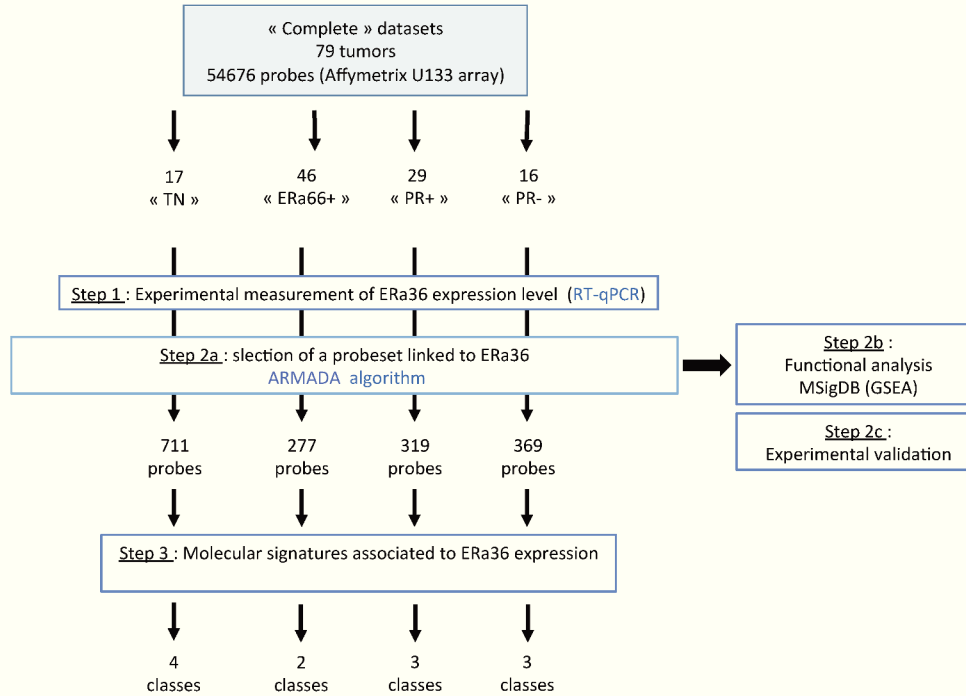


Fig. 8. Workflow describing the 3 steps of four breast tumor transcriptomic datasets.

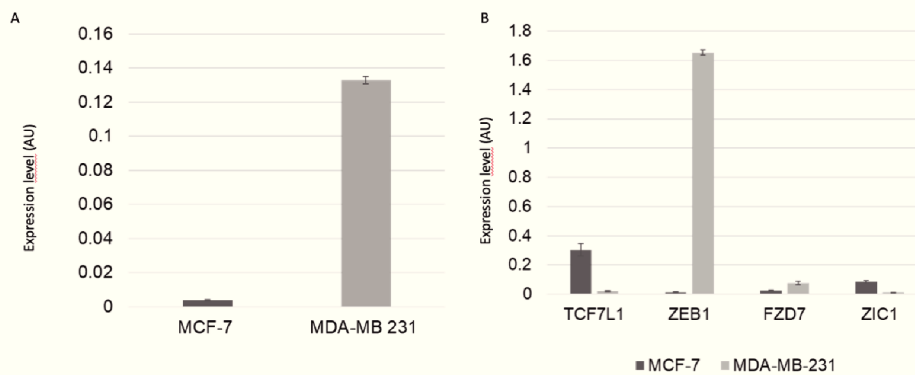


Fig. 9. Experimental validation of biological data inferred from *armada*. (A) ERα36 expression level as measured by RT-qPCR in MCF-7 and MDA-MB-231 breast cancer cells. (B) TCF7L1, ZEB1, FZD7 and ZIC1 expression level as measured by RT-qPCR in MCF-7 and MDA-MB-231 breast cancer cells.

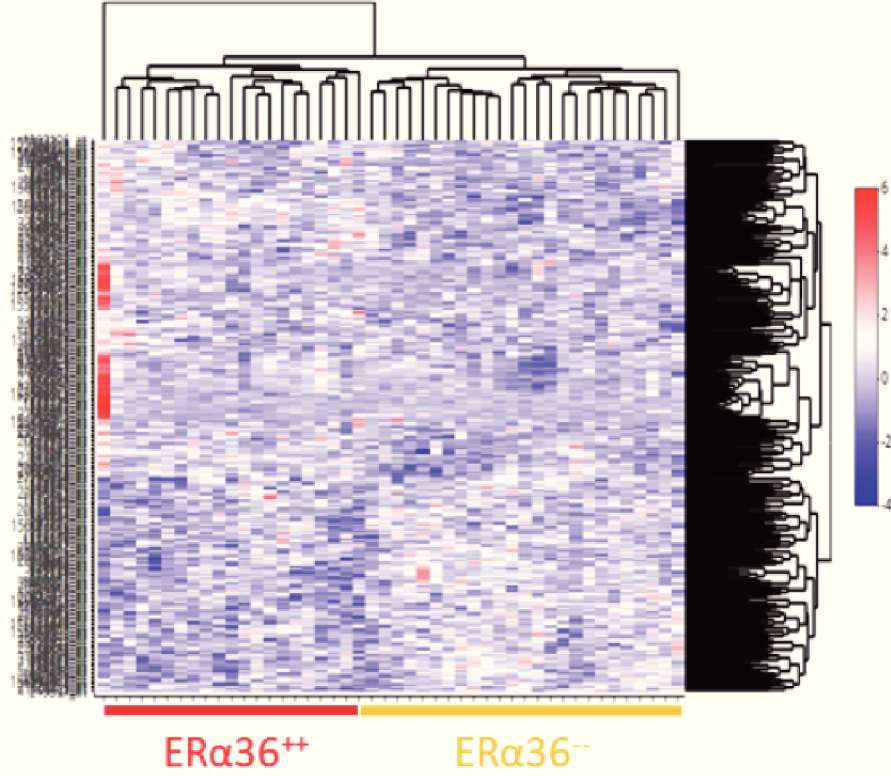


Fig. 10. Heatmap of the 711 covariates which have ARMADA scores greater or equal to 1 in the study of the dataset ERa66+. Each column corresponds to one ERa66+ tumor. The  $x$ -axis represents the tumors (marked with their ER $\alpha$ 36 expression level) and the  $y$ -axis represents the selected probes. Color gives the normalised expression of the covariates.

Table 3. Results of the  $N = 100$  runs in the design 2: rates of selection of the different groups of influential and noise covariates by the ARMADA method, the Wilcoxon test and the FAMT procedure. The corresponding standard deviations are given in brackets.

	ARMADA	Wilcoxon	FAMT
(0.7-3)	0.99 (0.08)	0.99 (0.07)	0.99 (0.04)
(0.7-2)	0.92 (0.27)	0.92 (0.26)	0.96 (0.17)
(0.7-1)	0.44 (0.49)	0.43 (0.49)	0.58 (0.49)
(0.3-3)	0.54 (0.49)	0.41 (0.49)	0.61 (0.48)
(0.3-2)	0.32 (0.46)	0.28 (0.45)	0.41 (0.49)
(0.3-1)	0.12 (0.32)	0.12 (0.33)	0.19 (0.39)
-	0.05 (0.23)	0.05 (0.22)	0.09 (0.29)

Table 4. Distribution of the covariates scores in the transcriptomic dataset: classification study.

Score	0	1	2	3	4	5	6	7	8
Number of covariates	2827	553	460	596	1170	888	306	10	0

Table 5. *Distribution of the covariates scores in the transcriptomic dataset: regression study.*

Score	0	1	2	3	4	5	6	7	8
Number of covariates	3988	89	456	509	984	692	86	5	1

Table 6. *Repartition of the covariates scores in the transcriptomic dataset. The R-scores are given in the 9 rows, the C-scores are given in the 8 columns. For instance, 41 covariates have a R-score equal to 1, and a C-score equal to 0.*

Regression score	Classification score							
	0	1	2	3	4	5	6	7
0	2227	328	273	337	531	257	34	1
1	41	7	3	9	17	10	2	0
2	131	35	39	52	119	71	9	0
3	119	48	44	50	117	114	17	0
4	174	65	56	86	256	241	102	4
5	119	64	40	57	116	176	116	4
6	15	4	4	5	12	19	26	1
7	1	2	1	0	1	0	0	0
8	0	0	0	0	1	0	0	0



**A statistical methodology to select covariates in  
high-dimensional data under dependence.  
Application to the classification of genetic  
profiles in oncology.  
Supplementary materials**

BERANGERE BASTIEN

*Transgene S.A., Boulevard Gonthier d'Andernach, Parc d'innovation, CS80166, F-67405*

*Illkirch-Graffenstaden Cedex - France*

TAHA BOUKHOBZA, HELENE DUMOND, CHARLENE THIEBAUT

*Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France*

HAFID CHAKIR, ANNE GEGOUT-PETIT, AURELIE MULLER-GUEUDIN\*, YAOJIE SHI

*Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France*

1. DESIGN 3: REGRESSION

In this section, we give results of simulations to study the behavior of our algorithm to select covariates linked with a continuous variable of interest (like survival time here). We simulate  $\tilde{\mathbf{X}} = (\tilde{\mathbf{X}}^{(k)})_{k=1,\dots,4}$  as in Section 3.1 of the main article, and  $Y$  as a standard gaussian variable. Now, we create the dependence with outcome  $Y$  in perturbing some component of  $\tilde{\mathbf{X}}$ : in all cluster

\*aurelie.gueudin@univ-lorraine.fr

$k = 1, \dots, 4$ , and for all  $i = 1, \dots, 400$ :

$$X_i^{(k)} = \tilde{X}_i^{(k)} + \delta_i Y \quad (1.1)$$

where  $\delta = (\delta_j)_{j=1, \dots, 400} = (1, 0.8, 0.6, 0.4, 0.2, 0, 0, \dots, 0)$ . Only the first 5 covariates of each cluster are linked with  $Y$ .

As in Section 3.2 of the main article, we show the interest of our pretreatment, comparing the three procedures detailed in Section 3.2 of the main article. As  $Y$  is a gaussian variable, we use the Pearson correlation test (instead of the Wilcoxon test used in Section 3.2 of the main article). We produce  $N = 100$  runs of  $(\mathbf{X}, Y)$  and count the number of false and true positive, and the ARMADA scores (shown in Figures 1 and 2).

Similarly to the classification study presented in Section 3 of the main article, our Procedure reduces the mean and the variability of the distributions of the false positive rates, in comparison to the Procedure 2 (i.e. the FAMTprocedure), and the power of our Procedure is comparable with Procedure 2.

The Figure 2 shows the ARMADA scores obtained on these  $N = 100$  runs of  $(\mathbf{X}, Y)$ . Again, similarly to the Section 3 of the main article, the scores give a ranking of the covariates, according to the intensity of their link with respect to the response variable  $Y$ . The true covariates are clearly separated of the noise covariates. We can also precise that 96% of the noise covariates obtained a score that was 0.

As in Section 3.4 of the main article, the Table 1 and the ROC curve in Figure 3 allow us to compare our method with the Pearson test and the FAMT procedure. Our method seems to be a good compromise to have quite good detection rates for the true covariates, but small detection rates for the noise covariates. Even though true covariates are not always enough detected, compared to the FAMT procedure, detection rate of noisy covariates is lower than FAMT. The Pearson test has the lowest levels of detection rates, and the true covariates with a small link with  $Y$  are not well detected. On the whole, our method seems to be appropriate for

sparse models particularly when the goal is to avoid false positive detections.

## 2. LUNG CANCER REAL DATASET: BOOTSTRAP ANALYSIS

As the number of patients  $n = 37$  is small compared to the number of covariates even after filtering ( $p = 6810$ ), we have checked our results with a bootstrap study. We have calculated the C-scores and R-scores of each covariates on  $B = 100$  bootstrap samples and the mean of the  $B$  results. We give the distribution of the bootstrapped means according to the original scores for the original dataset (Figure 4). We can see that the distributions of the bootstrapped means of the scores have a quite small dispersion and faithfully reproduce the original scores. The same conclusion holds for the bootstrapped median scores (shown in Tables 2 and 3).

Moreover, we can emphasize that our method is robust to detect the most important covariates (for instance, the 10 covariates that have a C-score equal to 7, or the 6 covariates that have an R-score greater than 7): their corresponding bootstrapped means of scores are also high, and their corresponding bootstrapped median scores are greater than 5.

## 3. BIOLOGICAL MATERIAL FOR THE STUDY OF ER $\alpha$ 36 IN BREAST CANCER

We analysed the biological network involving ER $\alpha$ 36 through the use of 4 sets of Affymetrix transcriptomic data obtained from breast tumors of different molecular subtypes: the triple negative (noted TN), ER66+, PR+ and PR- datasets:

- the TN dataset corresponding to Affymetrix transcriptomic comprehensive data from 17 patients derived xenografts (PDX) breast tumors was extracted from the Xentech<sup>TM</sup> database with the permission of Olivier Déas and Stefano Cairo (MTA CXT-295 Xentech SAS/University of Lorraine ; *Reyal and others (2012)*).
- the 3 other datasets (46 tumors ER66+, 29 tumors PR+, 16 tumors PR-) were part of

those from the Carte d'Identité des Tumeurs Program (CIT) from the Ligue Nationale Contre le Cancer described in [Guedj \*and others\* \(2012\)](#). Transcriptomic raw data were kindly provided by Aurélien De Reynies and Jacqueline Métral. One microgram of cDNAs from each tumor sample gathered at the Oncogenetics laboratory, INSERM U735, Institut Curie-Hôpital-Centre René Huguenin, St Cloud, France was also kindly provided by Ivan Bieche to measure ER $\alpha$ 36 expression.

The measurement of ER $\alpha$ 36 expression in each tumor (Step 1: clinical data completion) has been done as described in [Thiebaut \*and others\* \(2017\)](#). Total RNA extraction of PDX samples and qPCR analyses were performed. The following primers were used for qRT-PCR : GAPDH forward (Fw) 5'-TGC-ACC-ACC-AAC-TGC-TTA-GC -3', GAPDH reverse (Rev) 5'-GGC-ATG-GAC-TGT-GGT-CAT-GAG -3', ER $\alpha$ 36 forward (Fw) 5'- ATG-AAT-CTG-CAG-GGA-GAG-GA-3', ER $\alpha$ 36 reverse (Rev) 5'- GGC-TTT-AGA-CAC-GAG-GAA-ACC-3'. Assays were performed at least in triplicate, and the mean values were used to calculate expression levels, using the  $\Delta\Delta C(t)$  method referring to GAPDH housekeeping gene expression.

## REFERENCES

GUEDJ, MICHAEL, MARISA, LAËTITIA, DE REYNIES, AURÉLIEN, ORSETTI, BÉATRICE, SCHIAPPA, RENAUD, BIBEAU, FRÉDÉRIC, MACGROGAN, GAËTAN, LEREBOURS, FLORENCE, FINETTI, PASCAL, LONGY, MICHEL *and others*. (2012). A refined molecular taxonomy of breast cancer. *Oncogene* **31**(9), 1196.

REYAL, FABIEN, GUYADER, CHARLOTTE, DECRAENE, CHARLES, LUCCHESI, CARLO, AUGER, NATHALIE, ASSAYAG, FRANCK, DE PLATER, LUDMILLA, GENTIEN, DAVID, POUPON, MARIE-FRANCE, COTTU, PAUL *and others*. (2012). Molecular profiling of patient-derived breast cancer xenografts. *Breast cancer research* **14**(1), R11.

THIEBAUT, CHARLÈNE, CHAMARD-JOVENIN, CLÉMENCE, CHESNEL, AMAND, MOREL, CHLOÉ, DJERMOUNE, EL-HADI, BOUKHOBZA, TAHA AND DUMOND, HÉLÈNE. (2017). Mammary epithelial cell phenotype disruption in vitro and in vivo through ER $\alpha$ 36 overexpression. *PLoS one* **12**(3), e0173931.

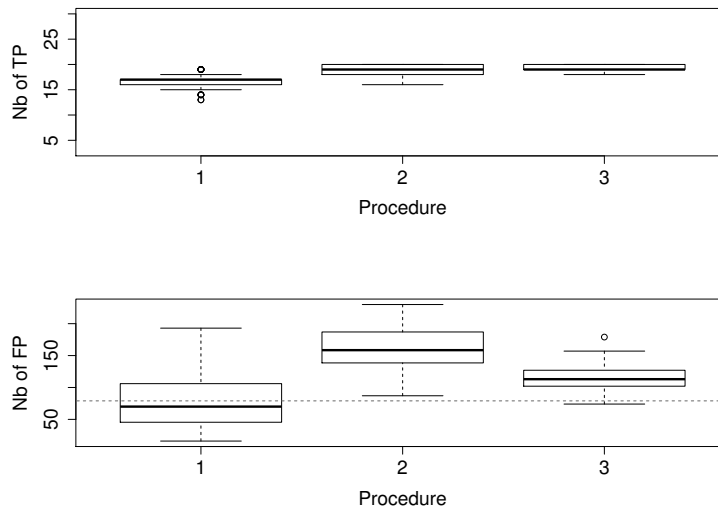


Fig. 1. Number of: true positive tests (top), false positive tests (bottom) in the design 3. Boxplots are calculated on  $N = 100$  runs.

Table 1. Results of the  $N = 100$  runs in the design 3: rates of selection of the different groups of influential and noise covariates by the ARMADA method, the Pearson correlation test and the FAMT procedure. The corresponding standard deviations are given in brackets.

	ARMADA	Pearson	FAMT
1	1 (0)	1 (0)	1 (0)
0.8	1 (0)	1 (0)	1 (0)
0.6	0.99 (0.08)	0.99 (0.08)	1 (0)
0.4	0.97 (0.18)	0.82 (0.38)	0.98 (0.13)
0.2	0.67 (0.47)	0.33 (0.47)	0.76 (0.43)
-	0.07 (0.26)	0.05 (0.22)	0.10 (0.30)

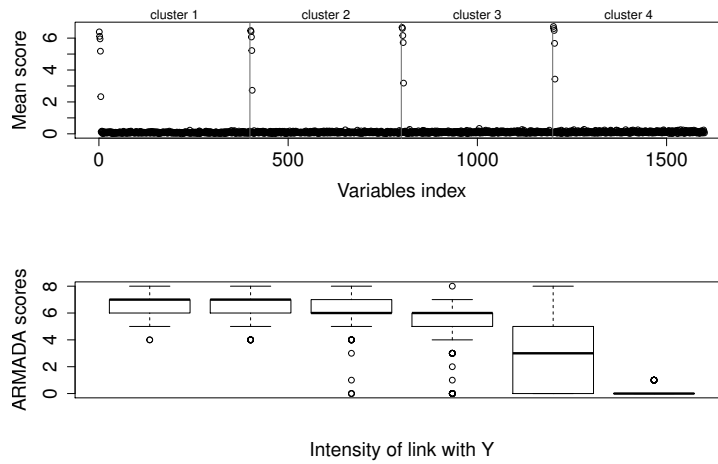


Fig. 2. Top: mean of the ARMADA scores obtained by all the covariates. Bottom: boxplot of the scores of the covariates, ranked by levels of link with  $Y$ . Means and boxplots are calculated on  $N = 100$  runs. Simulation in the design 3.

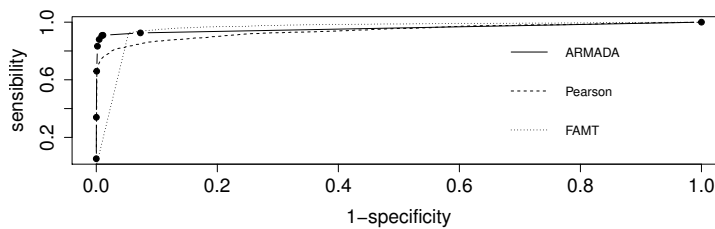


Fig. 3. ROC curves for the three selection methods, in the case of design 3. The ROC curves have been obtained by the mean of the  $N = 100$  ROC curves obtained in the  $N = 100$  runs of  $(\mathbf{X}, Y)$ .

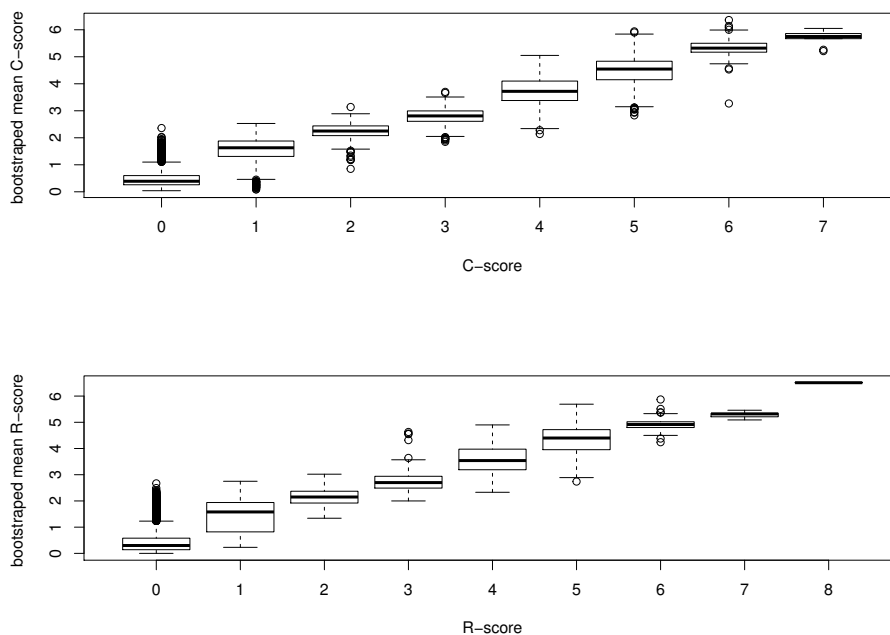


Fig. 4. Distribution of the bootstrapped mean of C-scores (resp. R-scores), i.e. means of C-(or R-)scores obtained on  $B = 100$  bootstrap samples), according to the corresponding C-scores (resp. R-scores) in the original dataset for all the  $p = 6810$  covariates.

Table 2. *Distribution of the bootstrapped median C-scores of the  $p = 6810$  covariates, obtained on  $B = 100$  bootstrap samples, versus the corresponding C-scores.*

Bootstrapped median C-score	ARMADA C-score							
	0	1	2	3	4	5	6	7
0	2698	53	0	0	0	0	0	0
0.5	11	1	0	0	0	0	0	0
1	108	315	29	1	0	0	0	0
1.5	0	19	5	0	0	0	0	0
2	9	162	308	76	5	0	0	0
2.5	0	2	28	14	0	0	0	0
3	1	1	87	321	55	2	0	0
3.5	0	0	1	29	12	1	0	0
4	0	0	2	155	922	218	1	0
4.5	0	0	0	0	19	6	0	0
5	0	0	0	0	157	644	221	2
5.5	0	0	0	0	0	0	6	0
6	0	0	0	0	0	17	78	8

Table 3. *Distribution of the bootstrapped median R-scores of the  $p = 6810$  covariates, obtained on  $B = 100$  bootstrap samples, versus the corresponding R-scores.*

Bootstrapped median R-score	ARMADA C-score								
	0	1	2	3	4	5	6	7	8
0	3773	29	11	0	0	0	0	0	0
0.5	20	2	5	0	0	0	0	0	0
1	67	22	17	0	0	0	0	0	0
1.5	8	1	9	0	0	0	0	0	0
2	109	32	243	40	1	0	0	0	0
2.5	4	0	22	8	2	0	0	0	0
3	7	3	147	295	80	2	0	0	0
3.5	0	0	0	14	13	2	0	0	0
4	0	0	2	149	788	210	0	0	0
4.5	0	0	0	0	10	14	0	0	0
5	0	0	0	3	90	462	85	5	0
5.5	0	0	0	0	0	1	0	0	0
6	0	0	0	0	0	1	1	0	1



---

## *Discussion et Perspectives*

---

Les objectifs de ce travail de thèse étaient d'évaluer l'impact de l'expression d'ER $\alpha$ 36 sur le développement physiologique de la glande mammaire et de préciser les modalités de régulation de l'expression de ce variant, au moment de l'initiation et/ou de la progression des tumeurs mammaires.

Cette problématique s'inscrit dans un contexte où les méthodes actuelles de classification des tumeurs mammaires commencent à montrer leurs limites notamment en ce qui concerne leurs capacités à orienter et à prédire l'efficacité des différentes stratégies thérapeutiques. En effet, à l'heure actuelle, l'un des grands enjeux de la recherche en cancérologie concerne la mise en évidence et la validation de nouveaux biomarqueurs prédictifs de la progression et de la réponse thérapeutique des tumeurs. La mise en perspective de nos résultats avec ceux disponibles dans la littérature nous permet d'apporter de nouveaux éléments et de proposer ER $\alpha$ 36 comme une piste prometteuse pour comprendre et moduler l'initiation et la résistance aux traitements des tumeurs mammaires.

Dans ce contexte, la discussion générale s'articulera autour de 3 questions :

- L'expression d'ER $\alpha$ 36 dans les cellules épithéliales mammaires saines est-elle nécessaire et/ou suffisante pour induire l'apparition d'un phénotype néoplasique ?
- Quels sont les signaux et les mécanismes responsables de la régulation de l'expression d'ER $\alpha$ 36 ?
- Quelle est la part de l'implication d'ER $\alpha$ 36 dans l'apparition de résistances intrinsèques ou acquises aux thérapies anti-cancéreuses ?

I. L'expression d'ER $\alpha$ 36 dans les cellules épithéliales mammaires saines est-elle nécessaire et/ou suffisante pour induire l'apparition d'un phénotype néoplasique ?

Depuis son clonage en 2005, ER $\alpha$ 36 a été principalement décrit dans la littérature pour son rôle dans la progression de différents types de cancers comme le cancer de l'endomètre, le cancer gastrique et le cancer du sein. Cet effet délétère associé à l'expression d'ER $\alpha$ 36 résulte de l'activation des voies de signalisation non génomiques qui interviennent dans la modulation de fonctions fréquemment dérégulées dans les cellules cancéreuses comme la prolifération, la migration et la survie (Wang Z. et al., 2005 ; Chaudhri et al., 2014b).

Bien que l'ARNm codant ce variant soit détecté dans plusieurs tissus sains, le rôle et les conséquences moléculaires et cellulaires de son activation sur le développement et la fonctionnalité de ces tissus restent à explorer (Zheng *et al.*, 2010 ; Xie H. *et al.*, 2011 ; Wang J. *et al.*, 2012 ; Sun *et al.*, 2013 ; Wang Q. *et al.*, 2015 ; Yan Y. *et al.*, 2017).

Grâce à une approche pluridisciplinaire incluant des études *in vitro* et *in silico* sur des cellules épithéliales mammaires saines immortalisées MCF-10A, nous avons montré que l'expression d'ER $\alpha$ 36 est suffisante pour perturber le phénotype de ces cellules de façon dépendante de NF- $\kappa$ B et STAT3. Dans la littérature, ces 2 facteurs sont décrits pour leur rôle physiologique au moment de l'étape d'involution de la glande mammaire. Plus précisément, l'activation de NF- $\kappa$ B et STAT3 participe à l'induction de l'apoptose des cellules épithéliales mammaires alvéolaires sécrétoires, qui est nécessaire pour permettre la régression du tissu alvéolaire mammaire au terme de la lactation (Connelly et al., 2010 ; Hughes et Watson, 2018). Cependant, cet effet pro-apoptotique ne coïncide pas avec les résultats obtenus dans notre étude puisque nous avons observé que l'expression d'ER $\alpha$ 36 induit plutôt une résistance à l'apoptose dans les cellules MCF-10A. Cet effet contradictoire peut s'expliquer par le fait que ces 2 facteurs sont également décrits dans la littérature pour leur implication dans la tumorigenèse mammaire (Zhang et Lai, 2014 ; Hughes et Watson, 2018). D'ailleurs, il a été démontré que STAT3 et NF- $\kappa$ B peuvent interagir pour promouvoir le maintien des cellules souches cancéreuses mammaires (Chung et al., 2013).

Si l'on s'intéresse plus particulièrement au facteur STAT3, qui est constitutivement actif dans 40% des tumeurs mammaires, il participe à la progression tumorale en favorisant notamment la survie et la transition épithélio-mésenchymateuse qui sont 2 processus qui semblent induits par l'expression d'ER $\alpha$ 36 dans les cellules épithéliales mammaires saines (Banerjee et Resat,

2016). De plus, il a été démontré que la voie JAK2/STAT3 est préférentiellement activée dans les cellules souches cancéreuses mammaires CD44<sup>+</sup>/CD24<sup>-low</sup>/ALDH1<sup>+</sup> (Marotta et al., 2011). Cela fait écho à l'étude de Kang et al., (2011a), qui a mis en évidence le rôle d'ER $\alpha$ 36 dans le maintien de cette population de cellules souches. Dans le même ordre d'idée, une étude de Asiedu et al., (2011), indique qu'un co-traitement de 40 jours des cellules MCF-10A avec du TGF $\beta$  et du TNF $\alpha$ , qui sont 2 facteurs intermédiaires liés à la signalisation d'ER $\alpha$ 36 dans notre étude, induit la transition épithélio-mésenchymateuses de ces cellules qui acquièrent des caractéristiques similaires à celles des cellules souches cancéreuses mammaires.

Concernant la modulation de la prolifération il semble que l'impact de STAT3 soit dépendant du contexte cellulaire. En effet, une étude de Yang S.F. et al., (2016), a montré que ce facteur, qui stimule la prolifération des cellules cancéreuses mammaires triples négatives MDA-MB-231 (haut grade), a un effet antiprolifératif sur les cellules cancéreuses luminales MCF-7 et T47D (bas grade). Cela coïncide avec 2 autres études qui ont montré que dans les tumeurs mammaires de bas grade, l'activation de STAT3 est associée à un bon pronostic (Dolled-Filhart et al., 2003 ; Sato et al., 2011). Ainsi, l'effet antiprolifératif observé dans les cellules immortalisées MCF-10A/ER $\alpha$ 36 de notre étude, est similaire à celui décrit dans les cellules cancéreuses MCF-7 et T47D qui correspondent à des tumeurs de bas grade. Dans la continuité, afin d'expliquer cet effet de STAT3 nous avons commencé à explorer l'implication d'ER $\beta$  et nous avons montré que cette isoforme des récepteurs aux œstrogènes, qui est principalement décrite dans la littérature pour son effet antiprolifératif dans les cellules cancéreuses mammaires (Haldosén et al., 2014), est activée de manière dépendante de STAT3 dans les cellules MCF-10A surexprimant ER $\alpha$ 36.

En parallèle, grâce à l'approche bioinformatique que nous avons développée au cours de ce travail de thèse, nous avons mis en évidence 4 facteurs de transcription, NFAT1, FOXO4, SP1 et LEF-1, qui semblent être des médiateurs de l'effet d'ER $\alpha$ 36 dans le contexte des tumeurs mammaires. Selon le contexte cellulaire, ces facteurs sont tous susceptibles d'agir comme des effecteurs transcriptionnels de la voie de signalisation Wnt/ $\beta$ -caténine qui intervient dans la régulation du développement physiologique embryonnaire et post-natal de la glande mammaire mais aussi dans l'autorenouvellement des cellules souches mammaires saines (Yu Q.C. et al., 2016). De plus, cette voie de signalisation est interconnectée à d'autres voies impliquant les facteurs STAT3 et NF- $\kappa$ B dans le contexte des cellules cancéreuses mammaires (Armanious et al., 2010 ; Ma et Hottiger, 2016).

In vivo, dans le modèle de souris Tg MMTV-ER $\alpha$ 36 exprimant spécifiquement ce variant dans la glande mammaire, nous avons observé que l'histologie des canaux est principalement perturbée lorsque la signalisation dépendante d'ER $\alpha$ 36 est activée probablement par les hormones endogènes à partir de la puberté. Ces altérations histologiques des canaux mammaires rappellent celles observées après une hyperstimulation de la signalisation œstrogénique et pourraient conduire à une transformation néoplasique de l'épithélium (Bocchinfuso et al., 2000). De plus, étant donné que les facteurs STAT3 et NF- $\kappa$ B, qui participent au réseau de signalisation dépendant d'ER $\alpha$ 36, sont impliqués dans le développement physiologique de la glande mammaire au moment de l'involution, il serait intéressant d'observer les effets de ce variant à cette étape du développement. Cela permettrait également d'évaluer l'éventuel impact d'ER $\alpha$ 36 sur la différenciation lactogénique au moment de la gestation.

Ainsi, les phénotypes cellulaires et tissulaires observés au cours de ces études indiquent que l'expression d'ER $\alpha$ 36 dans la glande mammaire saine participe à l'acquisition de propriétés associées à la transformation néoplasique. Cependant, cela n'est pas suffisant pour permettre le développement d'une tumeur in vivo. En effet, nous n'avons pas observé de développement de tumeurs chez les souris Tg MMTV-ER $\alpha$ 36 à l'âge adulte et l'expression de ce variant dans les cellules MCF-10A n'est pas suffisante pour leur conférer des propriétés tumorigéniques lorsqu'elles sont xénotreffées chez des souris Nude. Cela suggère donc que l'expression d'ER $\alpha$ 36 dans le tissu sain constitue un marqueur de sensibilité des cellules épithéliales mammaires qui seront plus réceptives à la dérégulation d'autres voies de signalisation oncogéniques auxquelles le variant ER $\alpha$ 36 participe indirectement. En d'autres termes, il est possible que l'induction de l'expression d'ER $\alpha$ 36 corresponde à l'une des étapes du processus d'initiation tumorale.

De manière complémentaire, on peut également se demander si une exposition continue, précoce (fœtal/néonatale) ou tardive (post-pubertaire), d'un individu à des molécules susceptibles de moduler l'expression et l'activité d'ER $\alpha$ 36 pourrait accroître le risque d'apparition d'une tumeur mammaire. En effet, dans notre étude, nous avons montré que les alkylphénols induisent l'expression d'ER $\alpha$ 36 dans les cellules épithéliales mammaires saines MCF-10A et contribuent in vitro à induire la prolifération, la migration et la survie de ces cellules.

## II. Quels sont les signaux et les mécanismes responsables de la régulation de l'expression d'ER $\alpha$ 36 ?

Dans ce travail de thèse, nous avons montré que l'expression d'ER $\alpha$ 36 est induite dans les cellules cancéreuses mammaires MDA-MB-231 après un traitement au BaP, qui est un agoniste du récepteur aux hydrocarbures aromatiques, AhR. Bien que l'interaction directe entre ce récepteur et son élément de réponse localisé dans la région promotrice d'ER $\alpha$ 36 n'ait pas encore été démontrée, plusieurs données laissent à penser que le rôle d'AhR dans l'initiation et la progression des tumeurs mammaires pourrait dépendre de sa capacité à induire l'expression d'ER $\alpha$ 36.

Tout d'abord, le niveau d'expression d'AhR et d'ER $\alpha$ 36 dans différentes lignées de cellules cancéreuses mammaires semble corrélé à l'agressivité de ces cellules. En effet, ces 2 facteurs sont plus exprimés dans les cellules cancéreuses mammaires triples négatives MDA-MB-231 (haut grade), que dans les cellules du sous-type « Luminal » MCF-7 (bas grade) (Zou et al., 2009 ; Powell et al., 2013). De plus, d'un point de vue clinique, dans une cohorte de 439 patientes l'expression d'AhR apparaît corrélée à des paramètres qui sont également associés à l'expression d'ER $\alpha$ 36 dans l'étude de Shi L. et al., (2009), à savoir l'âge des patientes ainsi que le statut HER2 des tumeurs mammaires (Vacher et al., 2018). Il a également été démontré qu'une forte expression d'AhR est corrélée à l'induction de l'expression d'un certain nombre de gènes associés à l'inflammation comme IL-6, TNF $\alpha$  et TGF $\beta$ 1 dans les tumeurs mammaires ER $\alpha$ 66+ et ER $\alpha$ 66- (Vacher et al., 2018). Cela est particulièrement intéressant car l'IL-6, qui apparaît également dérégulé dans notre étude transcriptomique sur les cellules MCF-10A/ER $\alpha$ 36, est un inducteur de la voie de signalisation JAK2/STAT3 dont l'activation semble liée à ER $\alpha$ 36 (Banerjee et Resat, 2016).

Parmi les ligands d'AhR décrits on trouve des membres de la famille des flavonoïdes qui agissent comme des agonistes ou antagonistes de ce récepteur (Zhang S. et al., 2003). De manière intéressante, certains flavonoïdes, comme le brousoflavonol B ou la kuromanine, sont également décrits pour leur impact sur la signalisation dépendante d'ER $\alpha$ 36 et pourraient être utilisés dans des études ultérieures afin de préciser les interconnexions entre l'activation d'AhR et l'induction de l'expression d'ER $\alpha$ 36 dans le contexte des tumeurs mammaires (Guo M. et al., 2013 ; Wang L. et al., 2016 ; Wang X. et al., 2017).

En parallèle, nos études ont montré que les alkylphénols sont capables d'induire l'expression d'ER $\alpha$ 36 dans les cellules séminomateuses Tcam-2 mais également dans les cellules MCF-

10A (Ajj et al., 2013). Etant donné que de nombreuses études portant sur le mode d'action des perturbateurs endocriniens ont montré que l'exposition à ces molécules pouvait perturber le fonctionnement de la machinerie épigénétique, nous avons émis l'hypothèse que l'expression d'ER $\alpha$ 36 pourrait être régulée par des mécanismes épigénétiques (Alavian-Ghavanini et Rüegg, 2018).

Dans une étude antérieure menée au laboratoire sur les cellules séminomateuses TCam-2, il a été démontré que l'expression de DNMT3L, une enzyme impliquée dans la méthylation de l'ADN, est réprimée en réponse aux alkylphénols (Ajj et al., 2013). Cela suggère donc que l'exposition à ce type de perturbateurs endocriniens est plutôt associée à un état hypométhylé de l'ADN. Cependant, dans notre étude menée sur les tumeurs mammaires nous avons observé qu'une hypométhylation de la région promotrice d'ER $\alpha$ 36, et plus précisément de 4 sites CpG localisés en aval du TSS, n'est pas associée à une augmentation mais plutôt à une diminution de l'expression de ce variant. Ce résultat qui apparaît contradictoire pourrait notamment s'expliquer par le contexte cellulaire des études.

En effet, la diminution de l'expression de DNMT3L en réponse aux alkylphénols a été mise en évidence dans le contexte des cellules séminomateuses TCam-2 qui ont conservé une identité de cellules germinales. Cela coïncide avec les données disponibles dans la littérature qui montrent qu'en réponse à d'autres perturbateurs endocriniens, comme la vinclozoline ou le benzoate d'œstradiol, l'expression des DNMTs 1 et 3 impliquées respectivement dans la maintenance et la mise en place du statut de méthylation de l'ADN, est réprimée dans les cellules germinales mâles de rats adultes. Cette modification de l'expression des DNMTs qui est elle-même soumise à une régulation épigénétique et post-transcriptionnelle, est associée à l'apparition d'effets transgénérationnels (Denis et al., 2011 ; Meunier et al., 2012).

A l'inverse, d'autres études portant sur les cellules somatiques ont montré que dans ce contexte cellulaire l'exposition au 17  $\alpha$ -éthynyl- œstradiol (EE2), qui est un composé synthétique classé parmi les perturbateurs endocriniens, conduit à une augmentation de l'expression de l'enzyme de maintenance DNMT1 dans la glande mammaire chez le rat (de Assis et al., 2012 ; Hilakivi-Clarke, 2014). Afin de préciser le lien entre l'exposition aux alkylphénols et l'induction de l'expression d'ER $\alpha$ 36 dans le contexte de la glande mammaire, il apparaît donc nécessaire de déterminer le niveau d'expression des DNMTs et le statut de méthylation du promoteur d'ER $\alpha$ 36 dans les cellules MCF-10A exposées ou non au M4. De plus, l'étude de l'impact des alkylphénols sur le niveau d'expression d'enzymes impliquées dans les modifications post

traductionnelles des histones, comme les HDACs ou les HMTs, permettrait de prendre en compte cette dimension des mécanismes épigénétiques.

En parallèle, étant donné que différentes études menées par l'équipe d'Ana M. Soto ont mis en évidence que l'exposition fœtale ou néonatale au BPA perturbe la signalisation œstrogénique et conduit à l'apparition d'altérations structurales de la glande mammaire (Muñoz-de-Toro et al., 2005 ; Ayyanan et al., 2011 ; Dhimolea et al., 2014), il serait intéressant d'évaluer l'impact du BPA et d'autres perturbateurs endocriniens sur l'expression d'ER $\alpha$ 36. Cela permettra de déterminer si l'expression de ce variant pourrait être considérée plus généralement comme un marqueur d'exposition à certains perturbateurs endocriniens.

Pour finir, nous avons exploré au cours de ces travaux les mécanismes de régulation post-transcriptionnelle de l'expression d'ER $\alpha$ 36 par des micro-ARNs. Au total, nous avons mis en évidence 12 micro-ARNs susceptibles d'interagir avec la région 3'UTR de l'ARNm de ce variant et nous avons validé l'implication d'hsa-miR-136-5p dans la régulation négative de l'expression d'ER $\alpha$ 36. A l'heure actuelle, les études disponibles dans la littérature indiquent que le niveau d'expression de hsa-miR-136-5p ne constitue pas un marqueur pronostic suffisant pour prédire le devenir des patientes. Toutefois, une étude de Huan et al., (2017) a montré que l'expression de son éponge, le lncRNA CRNDE, est significativement augmentée dans les biopsies de tumeurs mammaires et corrélée positivement à la taille et au stade/grade de la tumeur. Plus généralement, il a été démontré grâce à une analyse de type microarray que l'expression de milliers d'ARNs non codants est dérégulée dans le contexte des tumeurs mammaires (Chen B. et al., 2018). Ainsi, l'analyse de ces ARNs qui se comportent comme des éponges de micro-ARNs présente un intérêt majeur et demeure une piste ouverte pour caractériser l'ensemble des réseaux de régulation d'ER $\alpha$ 36.

Dans les perspectives il serait intéressant d'explorer le lien entre l'expression d'ER $\alpha$ 36 et les 11 autres micro-ARNs mis en évidence par le crible réalisé au début de notre étude. Parmi les pistes les plus prometteuses on peut citer le micro-ARN hsa-miR-6809-5p et son éponge l'ARN circulaire circEPSTI1 (epithelial stromal interaction 1). En effet, la neutralisation, c'est-à-dire l'inactivation, d'hsa-miR-6809-5p par son éponge est associée à une diminution de la survie de patientes atteintes d'un cancer du sein triple négatif (Chen B. et al., 2018). Cela pourrait notamment dépendre d'une augmentation de l'expression d'ER $\alpha$ 36.

De manière intéressante, plusieurs études ont également montré que l'exposition à certains perturbateurs endocriniens est associée à une modulation de l'expression des micro-ARNs



(Derghal et al., 2016). Par exemple, le répertoire en micro-ARNs dans des testicules de rats adultes est fortement perturbé après une exposition chronique (du jour de la conception à l'âge adulte) à de faibles doses d'un mélange d'alkylphénols et de phtalates (Buñay et al., 2017). Dans le même ordre d'idée, Tilghman et al., (2012), ont montré que le BPA module l'expression de plusieurs micro-ARNs, et notamment du micro-ARN oncogénique hsa-miR-21, dans les cellules cancéreuses mammaires MCF-7. Ainsi, l'impact des alkylphénols sur le niveau d'expression d'hsa-miR-136-5p est une autre piste intéressante à explorer pour déterminer l'impact de l'environnement sur l'expression d'ER $\alpha$ 36.

### III. Quelle est la part de l'implication d'ER $\alpha$ 36 dans l'apparition de résistances intrinsèques ou acquises aux thérapies anti-cancéreuses ?

Dans environ 60 à 70 % des cas les tumeurs mammaires sont dites hormono-dépendantes, c'est-à-dire que leur croissance et leur progression est contrôlée par les œstrogènes et/ou la progestérone. Dans le contexte des tumeurs ER $\alpha$ 66+ l'hormonothérapie est souvent incluse dans le protocole thérapeutique, en complément des stratégies thérapeutiques « classiques » basées sur la chirurgie et la chimiothérapie. Cependant, des résistances intrinsèques ou adaptatives à ce type de traitement sont observées dans environ 30% des cas et cela souligne l'importance des études visant à identifier les mécanismes moléculaires pouvant expliquer la survenue de ces résistances.

Dans le cadre de notre étude nous nous sommes principalement intéressés aux mécanismes de résistances au tamoxifène, qui est un anti-œstrogène principalement utilisé pour la prise en charge des patientes non ménopausées atteintes d'une tumeur ER $\alpha$ 66+. Il peut également être prescrit comme traitement adjuvant pour prévenir d'éventuelles rechutes. A l'heure actuelle, différents mécanismes basés sur des perturbations de la signalisation œstrogénique ont été proposés pour tenter d'expliquer le développement des résistances acquises au tamoxifène. Une étude récente de Poulard et al., (2019), a par exemple montré que cette résistance pourrait être liée à l'activation d'une signalisation dépendante du complexe ER $\alpha$ 66/Src/PI3K qui est préférentiellement induite dans différents modèles de résistance au tamoxifène. De plus, la forme canonique ER $\alpha$ 66 peut être activée de façon ligand-indépendante via l'induction de la voie PI3K/AkT qui aboutit à la phosphorylation de son résidu S167 (Campbell et al., 2001). Enfin, Hsu et al., (2019), ont mis en évidence que l'interconnexion de la signalisation dépendante du récepteur membranaire aux œstrogènes, GPER, et de celle dépendante d'EGFR pourrait également être impliquée dans l'acquisition de résistances au tamoxifène.

En parallèle, bien que cela soit encore débattu il semble que l'expression d'ER $\beta$ 1 dans les tumeurs soit associée à un bon pronostic et à une augmentation de la sensibilité au tamoxifène dans plusieurs cohortes de patientes atteintes d'un cancer du sein (Hopp et al., 2004 ; Honma et al., 2008 ; Omoto et Iwase, 2015). Cela pourrait être lié aux effets anti-prolifératifs et anti-invasifs d'ER $\beta$ 1, décrits dans le contexte des tumeurs mammaires (Haldosén et al., 2014).

D'autre part, les études de cohortes menées par Shi L. et al., (2009) et Wang Q. et al., (2018), ont montré que l'expression d'ER $\alpha$ 36 dans les tumeurs mammaires « ER $\alpha$ 66+ » traitées au tamoxifène est de mauvais pronostic. Cet effet délétère semble en grande partie lié aux effets

agonistes du tamoxifène qui dépendent d'ER $\alpha$ 36 et qui conduisent à l'activation de différentes voies de signalisation impliquées notamment dans l'acquisition de propriétés de cellules souches et de propriétés métastatiques (Wang Q. et al., 2018). En effet, il a été observé qu'après traitement des cellules cancéreuses mammaires MDA-MB-436 avec du tamoxifène, le variant ER $\alpha$ 36 phosphorylé est relocalisé au niveau nucléaire où il est associé à l'induction de l'expression du marqueur des cellules souches cancéreuses mammaires ALDH1 (Burness et Wicha, 2018). Toutefois, le mécanisme à l'origine de cette régulation transcriptionnelle dépendante d'ER $\alpha$ 36 reste à déterminer. Enfin, il semble que la boucle de signalisation impliquant ER $\alpha$ 36/EGFR/HER2 joue un rôle important car Yin et al., (2014) ont montré que l'inhibition d'EGFR et d'HER2 avec un inhibiteur de tyrosine kinase, le lapatinib, restaure la sensibilité des cellules MCF-7 qui étaient résistantes au tamoxifène (Yin et al., 2014 ; Teymourzadeh et al., 2017). Ce constat est particulièrement intéressant car il a été montré que le brousoflavonol B, la Cyanidine-3-o-glucoside (Kuromanine) et l'Icaritine, qui sont des flavonoïdes connus pour leurs propriétés anti-oxydantes, ont des effets antagonistes sur la signalisation dépendante d'ER $\alpha$ 36 et plus particulièrement sur la boucle ER $\alpha$ 36/EGFR/HER2. En effet, le brousoflavonol B est associé à une diminution de l'expression d'HER2 et d'EGFR dans les cellules cancéreuses mammaires MCF-7 résistantes au tamoxifène et dans les cellules MDA-MB-231 qui présentent un haut niveau d'expression d'ER $\alpha$ 36 (Guo M. et al., 2013 ; Yin et al., 2014). Quant à la Kuromanine et l'Icaritine, ils sont associés à une inhibition de la signalisation dépendante d'EGFR et ils induisent l'apoptose des cellules cancéreuses mammaires triples négatives MDA-MB-231 (Wang L. et al., 2016 ; Wang X. et al., 2017). Ainsi, leur utilisation pourrait constituer une piste pour tenter de contrer les effets dépendants de la boucle ER $\alpha$ 36/EGFR/HER2 dans la progression des tumeurs mammaires triples négatives et la résistance au tamoxifène des tumeurs « ER $\alpha$ 66+ ».

Ces mécanismes dépendants d'ER $\alpha$ 36 s'inscrivent dans le contexte des résistances acquises au tamoxifène. En effet, il a été démontré que l'exposition à cet anti-œstrogène est associée à une induction de l'expression d'ER $\alpha$ 36 dans les cellules cancéreuses mammaires MCF-7 (Yin et al., 2014). De la même façon, Wang Q. et al., (2018), ont observé que l'expression d'ER $\alpha$ 36 est significativement plus importante dans les biopsies de tumeurs mammaires résistantes au tamoxifène que dans les biopsies de tumeurs n'ayant pas été exposées au traitement. En partant de ce constat, nous avons montré au cours de ce travail de thèse qu'un co-traitement 4-OHT/DAC permet de contrer l'induction de l'expression d'ER $\alpha$ 36 induite par le tamoxifène dans les cellules MCF-7. Cette combinaison de traitement pourrait donc permettre de restaurer

la sensibilité des cellules cancéreuses mammaires au tamoxifène. Dans la continuité, il serait intéressant d'évaluer les effets combinés de la DAC et de l'4-OHT sur la survie des cellules cancéreuses mammaires *in vitro* mais également *in vivo* sur des modèles de tumeurs PDX résistantes ou non au tamoxifène (Poulard et al., 2019).

D'autre part, différentes études ont permis de mettre en évidence des signatures d'expression génique associées à un mauvais pronostic chez les patientes traitées au tamoxifène. De manière intéressante certaines de ces signatures semblent pouvoir prédire l'existence d'une résistance intrinsèque à cet anti-oestrogène chez des patientes qui n'ont pas encore été traitées (Musgrove et Sutherland, 2009). Etant donné qu'ER $\alpha$ 36 est impliqué dans le développement de la résistance au tamoxifène, nous pouvons émettre l'hypothèse que certaines de ces signatures pourraient être liées à l'expression de ce variant. Ainsi, il serait intéressant de comparer ces signatures moléculaires à celles obtenues grâce à notre analyse bioinformatique afin d'établir un profil plus précis des gènes dont l'expression serait corrélée à celle d'ER $\alpha$ 36 et qui pourraient être impliqués dans les résistances intrinsèques mais aussi acquises au tamoxifène.

A l'heure actuelle, le lien entre l'expression d'ER $\alpha$ 36 et la réponse des tumeurs mammaires à la chimiothérapie n'est pas clairement établi. En effet, les études de Shi L. et al., (2009) et Wang Q. et al., (2018) ont montré que l'expression de ce variant n'est pas corrélée à la survie des patientes présentant une tumeur ER $\alpha$ 66+ traitée par chimiothérapie (Cyclophosphamide ou Méthotrexate). Toutefois, une étude de Zhu L. et al., (2018), a démontré que le variant ER $\alpha$ 36 est impliqué dans le développement d'une résistance au Cisplatine dans les cellules MCF-7 et MDA-MB-231, de façon dépendante de la boucle de signalisation ER $\alpha$ 36/EGFR/HER2. L'implication d'ER $\alpha$ 36 dans le développement d'une résistance au Paclitaxel des cellules MDA-MB-231 a également été démontrée (Zhang J. et al., 2012).

De manière intéressante, nous avons mis en évidence que l'ARNm codant ER $\alpha$ 36 est une cible potentielle du micro-ARN hsa-miR-1253, dont l'expression est réprimée dans les lignées de cellules cancéreuses mammaires Bads-200 et Bats-72 résistantes au Paclitaxel (Lou et al., 2018). Ainsi, il serait intéressant d'étudier le lien entre l'expression de ce micro-ARN et l'expression d'ER $\alpha$ 36 dans le contexte du développement de la résistance au Paclitaxel. Enfin, il a également été démontré que la voie JAK/STAT3/MAPK/AkT, qui est modulée par l'expression d'ER $\alpha$ 36 dans notre étude, est impliquée dans la résistance des cellules cancéreuses mammaires triples négatives MDA-MB-231 et MDA-MB-468 au Paclitaxel (Wang S. et al., 2018).

En conclusion, ce travail de thèse a permis de montrer que l'expression d'ER $\alpha$ 36 joue un rôle important au moment de la transformation néoplasique des cellules épithéliales mammaires saines et de mettre en évidence des mécanismes de régulation classiques et épigénétiques de l'expression de ce variant. D'un point de vue clinique, cela permet de proposer des pistes de réflexion pour :

(i) Améliorer les méthodes de classification des tumeurs mammaires qui existent actuellement en ajoutant une composante moléculaire liée à l'expression d'ER $\alpha$ 36. Pour cela l'utilisation des approches « -omiques » couplées à la bioinformatique ouvre de grandes perspectives.

(ii) Développer de nouvelles stratégies thérapeutiques ciblant des partenaires d'ER $\alpha$ 36 dans les tumeurs mammaires.

(iii) Identifier de nouveaux biomarqueurs circulants prédictifs de l'expression d'ER $\alpha$ 36 qui pourraient participer à l'amélioration du diagnostic et du suivi des patientes. En effet, il serait intéressant d'évaluer si la détection de l'ARNm codant ER $\alpha$ 36 ou d'hsa-miR-136-5p et de CRNDE dans le plasma de patientes atteintes d'un cancer du sein pourrait permettre de prédire l'évolution de la tumeur et le devenir des patientes.

---

# *Bibliographie*

---

## A

- Acconcia, F., Ascenzi, P., Fabozzi, G., Visca, P., and Marino, M. (2004). S-palmitoylation modulates human estrogen receptor- $\alpha$  functions. *Biochem. Biophys. Res. Commun.* 316, 878–883.
- Adams, B.D., Furneaux, H., and White, B.A. (2007). The micro-ribonucleic acid (miRNA) miR-206 targets the human estrogen receptor- $\alpha$  (ER $\alpha$ ) and represses ER $\alpha$  messenger RNA and protein expression in breast cancer cell lines. *Mol. Endocrinol. Baltim. Md* 21, 1132–1147.
- Adomas, A.B., Grimm, S.A., Malone, C., Takaku, M., Sims, J.K., and Wade, P.A. (2014). Breast tumor specific mutation in GATA3 affects physiological mechanisms regulating transcription factor turnover. *BMC Cancer* 14, 278.
- Ajj, H., Chesnel, A., Pinel, S., Plénat, F., Flament, S., and Dumond, H. (2013). An alkylphenol mix promotes seminoma derived cell proliferation through an ER $\alpha$ 36-mediated mechanism. *PLoS ONE* 8, e61758.
- Alavian-Ghavanini, A., and Rüegg, J. (2018). Understanding Epigenetic Effects of Endocrine Disrupting Chemicals: From Mechanisms to Novel Test Methods. *Basic Clin. Pharmacol. Toxicol.* 122, 38–45.
- Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3983–3988.
- Al-nakhle, H., Smith, L., Bell, S., A Burns, P., Cummings, M., Hanby, A., Lane, S., D Parker, M., Hughes, T., and Speirs, V. (2013). Regulation of estrogen receptor  $\beta$ 1 expression in breast cancer by epigenetic modification of the 5' regulatory region. *Int. J. Oncol.* 43.
- Aloraifi, F., McDevitt, T., Martiniano, R., McGreevy, J., McLaughlin, R., Egan, C.M., Cody, N., Meany, M., Kenny, E., Green, A.J., et al. (2015). Detection of novel germline mutations for breast cancer in non-BRCA1/2 families. *FEBS J.* 282, 3424–3437.
- Amaral, S., Schroth, W., Kugler, S., Fritz, P., Simon, W., and Brauch, H. (2009). The promoter C specific ER $\alpha$  isoform is associated with tamoxifen outcome in breast cancer. *Breast Cancer Res. Treat.* 118, 323–331.
- Antanavičiūtė, I., Mikalayeva, V., Ceslevičienė, I., Milašiūtė, G., Skeberdis, V.A., and Bordel, S. (2017). Transcriptional hallmarks of cancer cell lines reveal an emerging role of branched chain amino acid catabolism. *Sci. Rep.* 7.
- Aranda, A., and Pascual, A. (2001). Nuclear hormone receptors and gene expression. *Physiol. Rev.* 81, 1269–1304.
- Ariazi, E.A., Taylor, J.C., Black, M.A., Nicolas, E., Slifker, M.J., Azzam, D.J., and Boyd, J. (2017). A New Role for ER $\alpha$ : Silencing via DNA Methylation of Basal, Stem Cell, and EMT Genes. *Mol. Cancer Res.* 15, 152–164.
- Armanious, H., Gelebart, P., Mackey, J., Ma, Y., and Lai, R. (2010). STAT3 upregulates the protein expression and transcriptional activity of  $\beta$ -catenin in breast cancer. *Int. J. Clin. Exp. Pathol.* 3, 654–664.
- Aroeira, L.S., Aguilera, A., Sánchez-Tomero, J.A., Bajo, M.A., Peso, G. del, Jiménez-Heffernan, J.A., Selgas, R., and López-Cabrera, M. (2007). Epithelial to Mesenchymal Transition and Peritoneal

Membrane Failure in Peritoneal Dialysis Patients: Pathologic Significance and Potential Therapeutic Interventions. *J. Am. Soc. Nephrol.* 18, 2004–2013.

Asiedu, M.K., Ingle, J.N., Behrens, M.D., Radisky, D.C., and Knutson, K.L. (2011). TGF $\beta$ /TNF $\alpha$ -Mediated Epithelial-Mesenchymal Transition Generates Breast Cancer Stem Cells with a Claudin-Low Phenotype. *Cancer Res.* 71, 4707–4719.

Asselin-Labat, M.-L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat. Cell Biol.* 9, 201–209.

de Assis, S., Warri, A., Cruz, M.I., Laja, O., Tian, Y., Zhang, B., Wang, Y., Huang, T.H.-M., and Hilakivi-Clarke, L. (2012). High-fat or ethinyl-oestradiol intake during pregnancy increases mammary cancer risk in several generations of offspring. *Nat. Commun.* 3, 1053.

Augereau, P., Miralles, F., Cavaillès, V., Gaudalet, C., Parker, M., and Rochefort, H. (1994). Characterization of the proximal estrogen-responsive element of human cathepsin D gene. *Mol. Endocrinol. Baltim. Md* 8, 693–703.

Ayyanan, A., Laribi, O., Schuepbach-Mallepell, S., Schrick, C., Gutierrez, M., Tanos, T., Lefebvre, G., Rougemont, J., Yalcin-Ozuysal, Ö., and Brisken, C. (2011). Perinatal Exposure to Bisphenol A Increases Adult Mammary Gland Progesterone Response and Cell Number. *Mol. Endocrinol.* 25, 1915–1923.

## B

Balasenthil, S., and Vadlamudi, R.K. (2003). Functional Interactions between the Estrogen Receptor Coactivator PELP1/MNAR and Retinoblastoma Protein. *J. Biol. Chem.* 278, 22119–22127.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.-H., LeProust, E., Park, I.-H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale methylomics reveals gene body signatures in human cell lines. *Nat. Biotechnol.* 27, 361–368.

Banerjee, K., and Resat, H. (2016). Constitutive activation of STAT3 in breast cancer cells: A review. *Int. J. Cancer* 138, 2570–2578.

Baulande, S., Criqui, A., and Duthieuw, M. (2014). Les microARN circulants, une nouvelle classe de biomarqueurs pour la médecine. *médecine/sciences* 30, 289–296.

Benetatos, L., Hatzimichael, E., Londin, E., Vartholomatos, G., Loher, P., Rigoutsos, I., and Briasoulis, E. (2013). The microRNAs within the DLK1-DIO3 genomic region: involvement in disease pathogenesis. *Cell. Mol. Life Sci. CMLS* 70, 795–814.

Bennesch, M.A., and Picard, D. (2015). Minireview: Tipping the Balance: Ligand-Independent Activation of Steroid Receptors. *Mol. Endocrinol.* 29, 349–363.

Biswas, D.K., Cruz, A.P., Gansberger, E., and Pardee, A.B. (2000). Epidermal growth factor-induced nuclear factor kappa B activation: A major pathway of cell-cycle progression in estrogen-receptor negative breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 97, 8542–8547.

Bocchinfuso, W.P., Lindzey, J.K., Hewitt, S.C., Clark, J.A., Myers, P.H., Cooper, R., and Korach, K.S. (2000). Induction of mammary gland development in estrogen receptor-alpha knockout mice. *Endocrinology* 141, 2982–2994.



- Bourguet, W., Germain, P., and Gronemeyer, H. (2000). Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. *Trends Pharmacol. Sci.* 21, 381–388.
- Brann, D.W., Zhang, Q.-G., Wang, R.-M., Mahesh, V.B., and Vadlamudi, R.K. (2008). PELP1 – A Novel Estrogen Receptor-Interacting Protein. *Mol. Cell. Endocrinol.* 290, 2–7.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* 68, 394–424.
- Brisken, C., and O'Malley, B. (2010). Hormone action in the mammary gland. *Cold Spring Harb. Perspect. Biol.* 2, a003178.
- Brooks, W.H., and Renaudineau, Y. (2015). Epigenetics and autoimmune diseases: the X chromosome-nucleolus nexus. *Front. Genet.* 6.
- Bulun, S.E., Chen, D., Moy, I., Brooks, D.C., and Zhao, H. (2012). Aromatase, breast cancer and obesity: a complex interaction. *Trends Endocrinol. Metab. TEM* 23, 83–89.
- Buñay, J., Larriba, E., Moreno, R.D., and del Mazo, J. (2017). Chronic low-dose exposure to a mixture of environmental endocrine disruptors induces microRNAs/isomiRs deregulation in mouse concomitant with intratesticular estradiol reduction. *Sci. Rep.* 7, 3373.
- Burness, M.L., and Wicha, M.S. (2018). Tamoxifen and ER $\alpha$ 36: Fertilizing the seeds of breast cancer metastasis. *Cell Res.* 28, 391–392.

## C

- Caizzi, L., Ferrero, G., Cutrupi, S., Cordero, F., Ballaré, C., Miano, V., Reineri, S., Ricci, L., Friard, O., Testori, A. (2014). Genome-wide activity of unliganded estrogen receptor- $\alpha$  in breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4892–4897.
- Campbell, R.A., Bhat-Nakshatri, P., Patel, N.M., Constantinidou, D., Ali, S., and Nakshatri, H. (2001). Phosphatidylinositol 3-Kinase/AKT-mediated Activation of Estrogen Receptor  $\alpha$ : A NEW MODEL FOR ANTI-ESTROGEN RESISTANCE. *J. Biol. Chem.* 276, 9817–9824.
- Carmeci, C., Thompson, D.A., Ring, H.Z., Francke, U., and Weigel, R.J. (1997). Identification of a gene (GPR30) with homology to the G-protein-coupled receptor superfamily associated with estrogen receptor expression in breast cancer. *Genomics* 45, 607–617.
- Carreau, S., Bois, C., Zanatta, L., Silva, F.R.M.B., Bouraima-Lelong, H., and Delalande, C. (2011). Estrogen signaling in testicular cells. *Life Sci.* 89, 584–587.
- Carroll, J.S. (2016). Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer. *Eur. J. Endocrinol.* 175, R41–R49.
- Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122, 33–43.
- Carroll, J.S., Hickey, T.E., Tarulli, G.A., Williams, M., and Tilley, W.D. (2017). Deciphering the divergent roles of progestogens in breast cancer. *Nat. Rev. Cancer* 17, 54–64.

- Castellano, L., Giamas, G., Jacob, J., Coombes, R.C., Lucchesi, W., Thiruchelvam, P., Barton, G., Jiao, L.R., Wait, R., Waxman, J. (2009). The estrogen receptor- $\alpha$ -induced microRNA signature regulates itself and its transcriptional response. *Proc. Natl. Acad. Sci. U. S. A.* 106, 15732–15737.
- Castet, A., Boulahtouf, A., Versini, G., Bonnet, S., Augereau, P., Vignon, F., Khochbin, S., Jalaguier, S., and Cavaillès, V. (2004). Multiple domains of the Receptor-Interacting Protein 140 contribute to transcription inhibition. *Nucleic Acids Res.* 32, 1957–1966.
- Castet, A., Carascossa, S., Duong, V., Augereau, P., Jalaguier, S., and Cavaillès, V. (2005). Rôle de RIP140 dans la signalisation hormonale. *médecine/sciences* 21, 273–278.
- Catalano, S., Barone, I., Giordano, C., Rizza, P., Qi, H., Gu, G., Malivindi, R., Bonofiglio, D., and Andò, S. (2009). Rapid Estradiol/ER $\alpha$  Signaling Enhances Aromatase Enzymatic Activity in Breast Cancer Cells. *Mol. Endocrinol.* 23, 1634–1645.
- Chakraborty, S., Jaiswal, N., and Nag, A. (2014). Biology of FOXM1 and its emerging role in cancer therapy. *J. Proteins Proteomics* 5.
- Chamard-Jovenin, C., Jung, A.C., Chesnel, A., Abecassis, J., Flament, S., Ledrappier, S., Macabre, C., Boukhobza, T., and Dumond, H. (2015). From ER $\alpha$ 66 to ER $\alpha$ 36: a generic method for validating a prognosis marker of breast tumor progression. *BMC Syst. Biol.* 9, 28.
- Chambliss, K.L., Yuhanna, I.S., Anderson, R.G.W., Mendelsohn, M.E., and Shaul, P.W. (2002). ERbeta has nongenomic action in caveolae. *Mol. Endocrinol. Baltim. Md* 16, 938–946.
- Chandarlapaty, S., Chen, D., He, W., Sung, P., Samoila, A., You, D., Bhatt, T., Patel, P., Voi, M., Gnant, M. (2016). Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial. *JAMA Oncol.* 2, 1310–1315.
- Chang, D.H., Cattoretti, G., and Calame, K.L. (2002). The dynamic expression pattern of B lymphocyte induced maturation protein-1 (Blimp-1) during mouse embryonic development. *Mech. Dev.* 117, 305–309.
- Charles, A.K., and Darbre, P.D. (2013). Combinations of parabens at concentrations measured in human breast tissue can increase proliferation of MCF-7 human breast cancer cells. *J. Appl. Toxicol.* 33, 390–398.
- Chatterjee, A., Rodger, E.J., and Eccles, M.R. (2018). Epigenetic drivers of tumourigenesis and cancer metastasis. *Semin. Cancer Biol.* 51, 149–159.
- Chaudhri, R.A., Olivares-Navarrete, R., Cuenca, N., Hadadi, A., Boyan, B.D., and Schwartz, Z. (2012). Membrane estrogen signaling enhances tumorigenesis and metastatic potential of breast cancer cells via estrogen receptor- $\alpha$ 36 (ER $\alpha$ 36). *J. Biol. Chem.* 287, 7169–7181.
- Chaudhri, R.A., Hadadi, A., Lobachev, K.S., Schwartz, Z., and Boyan, B.D. (2014a). Estrogen receptor-alpha 36 mediates the anti-apoptotic effect of estradiol in triple negative breast cancer cells via a membrane-associated mechanism. *Biochim. Biophys. Acta* 1843, 2796–2806.
- Chaudhri, R.A., Schwartz, N., Elbaradie, K., Schwartz, Z., and Boyan, B.D. (2014b). Role of ER $\alpha$ 36 in membrane-associated signaling by estrogen. *Steroids* 81, 74–80.
- Chen, B., Tang, H., Liu, X., Liu, P., Yang, L., Xie, X., Ye, F., Song, C., Xie, X., and Wei, W. (2015). miR-22 as a prognostic factor targets glucose transporter protein type 1 in breast cancer. *Cancer Lett.* 356, 410–417.

Chen, B., Wei, W., Huang, X., Xie, X., Kong, Y., Dai, D., Yang, L., Wang, J., Tang, H., and Xie, X. (2018). circEPSTII as a Prognostic Marker and Mediator of Triple-Negative Breast Cancer Progression. *Theranostics* 8, 4003–4015.

Chen, P., Chen, P., Zhao, L., Zhao, L., Pan, X., Pan, X., Jin, L., Jin, L., Lin, C., Lin, C., et al. (2018). Tumor suppressor microRNA-136-5p regulates the cellular function of renal cell carcinoma. *Oncol. Lett.* 15, 5995–6002.

Chetrite, G.S., Cortes-Prieto, J., Philippe, J.C., Wright, F., and Pasqualini, J.R. (2000). Comparison of estrogen concentrations, estrone sulfatase and aromatase activities in normal, and in cancerous, human breast tissues. *J. Steroid Biochem. Mol. Biol.* 72, 23–27.

Chiche, A. (2012). Study of mammary stem and progenitor cells and of their contribution in tumorigenesis : role of transcriptional factors Myc and p53. Thèse de l'Université de Paris 11.

Chu, I., Blackwell, K., Chen, S., and Slingerland, J. (2005). The dual ErbB1/ErbB2 inhibitor, lapatinib (GW572016), cooperates with tamoxifen to inhibit both cell proliferation- and estrogen-dependent gene expression in antiestrogen-resistant breast cancer. *Cancer Res.* 65, 18–25.

Chung, S.S., Aroh, C., and Vadgama, J.V. (2013). Constitutive Activation of STAT3 Signaling Regulates hTERT and Promotes Stem Cell-Like Traits in Human Breast Cancer Cells. *PLOS ONE* 8, e83971.

Clarke, C.H., Norfleet, A.M., Clarke, M.S., Watson, C.S., Cunningham, K.A., and Thomas, M.L. (2000). Perimembrane localization of the estrogen receptor alpha protein in neuronal processes of cultured hippocampal neurons. *Neuroendocrinology* 71, 34–42.

Clarke, M., Collins, R., Darby, S., Davies, C., Elphinstone, P., Evans, V., Godwin, J., Gray, R., Hicks, C., James, S. (2005). Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet Lond. Engl.* 366, 2087–2106.

Coelingh Bennink, H.J.T., Holinka, C.F., and Diczfalusy, E. (2008). Estetrol review: profile and potential clinical applications. *Climacteric J. Int. Menopause Soc.* 11 Suppl 1, 47–58.

Coin, F., and Egly, J. (2000). Formation du complexe d'initiation de la transcription : des facteurs généraux aux complexes qui déstabilisent la chromatine. *médecine/sciences* 16, 593.

Connelly, L., Barham, W., Pigg, R., Saint-Jean, L., Sherrill, T., Cheng, D.-S., Chodosh, L.A., Blackwell, T.S., and Yull, F.E. (2010). Activation of nuclear factor-kappa B in mammary epithelium promotes milk loss during mammary development and infection. *J. Cell. Physiol.* 222, 73–81.

Cvoro, A., Tzagarakis-Foster, C., Tatomer, D., Paruthiyil, S., Fox, M.S., and Leitman, D.C. (2006). Distinct roles of unliganded and liganded estrogen receptors in transcriptional repression. *Mol. Cell* 21, 555–564.

## D

D'Amato, N.C., Gordon, M.A., Babbs, B., Spoelstra, N.S., Carson Butterfield, K.T., Torkko, K.C., Phan, V.T., Barton, V.N., Rogers, T.J., Sartorius, C.A. (2016). Cooperative Dynamics of AR and ER Activity in Breast Cancer. *Mol. Cancer Res. MCR* 14, 1054–1067.

Davies, C., Pan, H., Godwin, J., Gray, R., Arriagada, R., Raina, V., Abraham, M., Medeiros Alencar, V.H., Badran, A., Bonfill, X. (2013). Long-term effects of continuing adjuvant tamoxifen to 10 years

versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet Lond. Engl.* 381, 805–816.

deConinck, E.C., McPherson, L.A., and Weigel, R.J. (1995). Transcriptional regulation of estrogen receptor in breast carcinomas. *Mol. Cell. Biol.* 15, 2191–2196.

DeNardo, D.G., Kim, H.-T., Hilsenbeck, S., Cuba, V., Tsimelzon, A., and Brown, P.H. (2005). Global Gene Expression Analysis of Estrogen Receptor Transcription Factor Cross Talk in Breast Cancer: Identification of Estrogen-Induced/Activator Protein-1-Dependent Genes. *Mol. Endocrinol.* 19, 362–378.

DeNardo, D.G., Cuba, V.L., Kim, H., Wu, K., Lee, A.V., and Brown, P.H. (2007). Estrogen receptor DNA binding is not required for estrogen-induced breast cell growth. *Mol. Cell. Endocrinol.* 277, 13–25.

Deng, H., Huang, X., Fan, J., Wang, L., Xia, Q., Yang, X., Wang, Z., and Liu, L. (2010). A variant of estrogen receptor- $\alpha$ , ER- $\alpha$ 36 is expressed in human gastric cancer and is highly correlated with lymph node metastasis. *Oncol. Rep.* 24, 171–176.

Denis, H., Ndlovu, 'Matladi N, and Fuks, F. (2011). Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Rep.* 12, 647–656.

Derghal, A., Djelloul, M., Trouslard, J., and Mounien, L. (2016). An Emerging Role of micro-RNA in the Effect of the Endocrine Disruptors. *Front. Neurosci.* 10.

Dhamad, A.E., Zhou, Z., Zhou, J., and Du, Y. (2016). Systematic Proteomic Identification of the Heat Shock Proteins (Hsp) that Interact with Estrogen Receptor Alpha (ER $\alpha$ ) and Biochemical Characterization of the ER $\alpha$ -Hsp70 Interaction. *PLoS One* 11, e0160312.

Dhimolea, E., Wadia, P.R., Murray, T.J., Settles, M.L., Treitman, J.D., Sonnenschein, C., Shioda, T., and Soto, A.M. (2014). Prenatal Exposure to BPA Alters the Epigenome of the Rat Mammary Gland and Increases the Propensity to Neoplastic Development. *PLoS ONE* 9.

Ding, W., Ren, J., Ren, H., and Wang, D. (2017). Long Noncoding RNA HOTAIR Modulates MiR-206-mediated Bcl-w Signaling to Facilitate Cell Proliferation in Breast Cancer. *Sci. Rep.* 7, 17261.

Dittmer, J. (2018). Breast cancer stem cells: Features, key drivers and treatment options. *Semin. Cancer Biol.* 53, 59–74.

Dolled-Filhart, M., Camp, R.L., Kowalski, D.P., Smith, B.L., and Rimm, D.L. (2003). Tissue microarray analysis of signal transducers and activators of transcription 3 (Stat3) and phospho-Stat3 (Tyr705) in node-negative breast cancer shows nuclear localization is associated with a better prognosis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 9, 594–600.

Driscoll, M.D., Sathya, G., Muyan, M., Klinge, C.M., Hilf, R., and Bambara, R.A. (1998). Sequence requirements for estrogen receptor binding to estrogen response elements. *J. Biol. Chem.* 273, 29321–29330.

Duffy, M.J., Harbeck, N., Nap, M., Molina, R., Nicolini, A., Senkus, E., and Cardoso, F. (2017). Clinical use of biomarkers in breast cancer: Updated guidelines from the European Group on Tumor Markers (EGTM). *Eur. J. Cancer Oxf. Engl.* 1990 75, 284–298.

## E

Eckert, D., Buhl, S., Weber, S., Jäger, R., and Schorle, H. (2005). The AP-2 family of transcription factors. *Genome Biol.* 6, 246.

Eeckhoute, J., Keeton, E.K., Lupien, M., Krum, S.A., Carroll, J.S., and Brown, M. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Res.* 67, 6477–6483.

Ehrlich, M., and Lacey, M. (2013). DNA methylation and differentiation: silencing, upregulation and modulation of gene expression. *Epigenomics* 5, 553–568.

Eroles, P., Bosch, A., Alejandro Pérez-Fidalgo, J., and Lluch, A. (2012). Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. *Cancer Treat. Rev.* 38, 698–707.

## F

Fallahpour, S., Navaneelan, T., De, P., and Borgo, A. (2017). Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data. *CMAJ Open* 5, E734–E739.

Farooq, A. (2015). Structural and Functional Diversity of Estrogen Receptor Ligands. *Curr. Top. Med. Chem.* 15, 1372–1384.

Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., Ji, X., Liu, W., Huang, B., Luo, W. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. *Genes Dis.* 5, 77–106.

Filardo, E.J., and Thomas, P. (2012). Minireview: G protein-coupled estrogen receptor-1, GPER-1: its mechanism of action and role in female reproductive cancer, renal and vascular physiology. *Endocrinology* 153, 2953–2962.

Flouriot, G., Brand, H., Denger, S., Metivier, R., Kos, M., Reid, G., Sonntag-Buck, V., and Gannon, F. (2000). Identification of a new isoform of the human estrogen receptor-alpha (hER- $\alpha$ ) that is encoded by distinct transcripts and that is able to repress hER- $\alpha$  activation function 1. *EMBO J.* 19, 4688–4700.

Fox, E.M., Davis, R.J., and Shupnik, M.A. (2008). ER $\beta$  in Breast Cancer – Onlooker, Passive Player, or Active Protector? *Steroids* 73, 1039–1051.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.

Friedrichs, N., Jäger, R., Paggen, E., Rudlowski, C., Merkelbach-Bruse, S., Schorle, H., and Buettner, R. (2005). Distinct spatial expression patterns of AP-2alpha and AP-2gamma in non-neoplastic human breast and breast cancer. *Mod. Pathol. Off. J. U. S. Can. Acad. Pathol. Inc* 18, 431–438.

## G

Gangloff, Y.G., Romier, C., Thuault, S., Werten, S., and Davidson, I. (2001). The histone fold is a key structural motif of transcription factor TFIID. *Trends Biochem. Sci.* 26, 250–257.

Gao, H., Song, X., Kang, T., Yan, B., Feng, L., Gao, L., Ai, L., Liu, X., Yu, J., and Li, H. (2017). Long noncoding RNA CRNDE functions as a competing endogenous RNA to promote metastasis and oxaliplatin resistance by sponging miR-136 in colorectal cancer. *OncoTargets Ther.* 10, 205–216.

- Gibson, D.A., and Saunders, P.T.K. (2012). Estrogen dependent signaling in reproductive tissues - a role for estrogen receptors and estrogen related receptors. *Mol. Cell. Endocrinol.* 348, 361–372.
- Giguère, V. (2002). To ERR in the estrogen pathway. *Trends Endocrinol. Metab. TEM* 13, 220–225.
- Giguère, V. (2008). Transcriptional control of energy homeostasis by the estrogen-related receptors. *Endocr. Rev.* 29, 677–696.
- Giguère, V., Yang, N., Segui, P., and Evans, R.M. (1988). Identification of a new class of steroid hormone receptors. *Nature* 331, 91–94.
- Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* 1, 555–567.
- Ginestier, C., Korkaya, H., Dontu, G., Birnbaum, D., Wicha, M., and Charafe-Jauffret, E. (2008). The cancer stem cell: The breast cancer driver. *médecine/sciences* 23, 1133-9.
- Giordano, C., Cui, Y., Barone, I., Ando, S., Mancini, M.A., Berno, V., and Fuqua, S.A.W. (2010). Growth factor-induced resistance to tamoxifen is associated with a mutation of estrogen receptor alpha and its phosphorylation at serine 305. *Breast Cancer Res. Treat.* 119, 71–85.
- Gong, W., Yu, J., Wang, Q., Li, S., Song, J., Jia, Z., Huang, S., and Zhang, A. (2016). Estrogen-related receptor (ERR)  $\gamma$  protects against puromycin aminonucleoside-induced podocyte apoptosis by targeting PI3K/Akt signaling. *Int. J. Biochem. Cell Biol.* 78, 75–86.
- Gonzalez de Valdivia, E., Broselid, S., Kahn, R., Olde, B., and Leeb-Lundberg, L.M.F. (2017). G protein-coupled estrogen receptor 1 (GPER1)/GPR30 increases ERK1/2 activity through PDZ motif-dependent and -independent mechanisms. *J. Biol. Chem.* 292, 9932–9943.
- Goode, G., Ballard, B.R., Manning, H.C., Freeman, M.L., Kang, Y., and Eltom, S.E. (2013). Knockdown of aberrantly upregulated aryl hydrocarbon receptor reduces tumor growth and metastasis of MDA-MB-231 human breast cancer cell line. *Int. J. Cancer J. Int. Cancer* 133, 2769–2780.
- Goss, P.E., Ingle, J.N., Pritchard, K.I., Robert, N.J., Muss, H., Gralow, J., Gelmon, K., Whelan, T., Strasser-Weippl, K., Rubin, S. (2016). Extending Aromatase-Inhibitor Adjuvant Therapy to 10 Years. *N. Engl. J. Med.* 375, 209–219.
- Grandien, K.F., Berkenstam, A., Nilsson, S., and Gustafsson, J.A. (1993). Localization of DNase I hypersensitive sites in the human oestrogen receptor gene correlates with the transcriptional activity of two differentially used promoters. *J. Mol. Endocrinol.* 10, 269–277.
- Greene, G.L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., and Shine, J. (1986). Sequence and expression of human estrogen receptor complementary DNA. *Science* 231, 1150–1154.
- Gruber, C.J., Tschugguel, W., Schneeberger, C., and Huber, J.C. (2002). Production and Actions of Estrogens. *N. Engl. J. Med.* 346, 340–352.
- Gudjonsson, T., Adriance, M.C., Sternlicht, M.D., Petersen, O.W., and Bissell, M.J. (2005). Myoepithelial Cells: Their Origin and Function in Breast Morphogenesis and Neoplasia. *J. Mammary Gland Biol. Neoplasia* 10, 261–272.
- Guedj, M., Marisa, L., de Reynies, A., Orsetti, B., Schiappa, R., Bibeau, F., MacGrogan, G., Lerebours, F., Finetti, P., Longy, M. (2012). A refined molecular taxonomy of breast cancer. *Oncogene* 31, 1196–1206.

Guo, S., and Sonenshein, G.E. (2004). Forkhead box transcription factor FOXO3a regulates estrogen receptor alpha expression and is repressed by the Her-2/neu/phosphatidylinositol 3-kinase/Akt signaling pathway. *Mol. Cell. Biol.* 24, 8681–8690.

Guo, T., and Pan, G. (2018). MicroRNA-136 functions as a tumor suppressor in osteosarcoma via regulating metadherin. *Cancer Biomark. Sect. Dis. Markers* 22, 79–87.

Guo, M., Wang, M., Deng, H., Zhang, X., and Wang, Z.-Y. (2013). A novel anticancer agent Brousoflavonol B downregulates estrogen receptor (ER)- $\alpha$ 36 expression and inhibits growth of ER-negative breast cancer MDA-MB-231 cells. *Eur. J. Pharmacol.* 714, 56–64.

Guo, Y., Yu, P., Liu, Z., Maimaiti, Y., Chen, C., Zhang, Y., Yin, X., Wang, S., Liu, C., and Huang, T. (2017). Prognostic and clinicopathological value of GATA binding protein 3 in breast cancer: A systematic review and meta-analysis. *PLOS ONE* 12, e0174843.

Gusterson, B.A., and Stein, T. (2012). Human breast development. *Semin. Cell Dev. Biol.* 23, 567–573.

## H

Häggsström, M., and Richfield, D. (2014). Diagram of the pathways of human steroidogenesis. *WikiJournal Med.* 1, 5.

Haldosén, L.-A., Zhao, C., and Dahlman-Wright, K. (2014). Estrogen receptor beta in breast cancer. *Mol. Cell. Endocrinol.* 382, 665–672.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. *Cell* 144, 646–674.

Hao, H., d'Alincourt-Salazar, M., Kelley, K.M.M., Shatnawi, A., Mukherjee, S., Shah, Y.M., and Ratnam, M. (2007). Estrogen-induced and TAFII30-mediated gene repression by direct recruitment of the estrogen receptor and co-repressors to the core promoter and its reversal by tamoxifen. *Oncogene* 26, 7872–7884.

Hawkins, S.M., and Matzuk, M.M. (2008). Menstrual Cycle: Basic Biology. *Ann. N. Y. Acad. Sci.* 1135, 10–18.

He, T.-T., Zuo, A.-J., Wang, J.-G., and Zhao, P. (2017). Organochlorine pesticides accumulation and breast cancer: A hospital-based case-control study. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* 39, 1010428317699114.

Heckler, M.M., Thakor, H., Schafer, C.C., and Riggins, R.B. (2014). ERK/MAPK regulates ERR $\gamma$  expression, transcriptional activity and receptor-mediated tamoxifen resistance in ER+ breast cancer. *FEBS J.* 281, 2431–2442.

Heery, D.M., Kalkhoven, E., Hoare, S., and Parker, M.G. (1997). A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature* 387, 733–736.

Heldring, N., Pawson, T., McDonnell, D., Treuter, E., Gustafsson, J.-Å., and Pike, A.C.W. (2007). Structural Insights into Corepressor Recognition by Antagonist-bound Estrogen Receptors. *J. Biol. Chem.* 282, 10449–10455.

Hennequin, C., and Fourquet, A. (2014). Controverse sur l'irradiation de la chaîne mammaire interne dans le cancer du sein. *Cancer/radiothérapie* 18 : 351–355.

Hens, J.R., and Wysolmerski, J.J. (2005). Key stages of mammary gland development: Molecular mechanisms involved in the formation of the embryonic mammary gland. *Breast Cancer Res.* 7, 220–224.

Hilakivi-Clarke, L. (2014). Maternal exposure to diethylstilbestrol during pregnancy and increased breast cancer risk in daughters. *Breast Cancer Res. BCR* 16, 208.

Holinka, C.F., Diczfalusy, E., and Coelingh Bennink, H.J.T. (2008). Estetrol: a unique steroid in human pregnancy. *J. Steroid Biochem. Mol. Biol.* 110, 138–143.

Holst, F. (2016). Estrogen receptor alpha gene amplification in breast cancer: 25 years of debate. *World J. Clin. Oncol.* 7, 160–173.

Hong, H., Yang, L., and Stallcup, M.R. (1999). Hormone-independent transcriptional activation and coactivator binding by novel orphan nuclear receptor ERR3. *J. Biol. Chem.* 274, 22618–22626.

Honma, N., Horii, R., Iwase, T., Saji, S., Younes, M., Takubo, K., Matsuura, M., Ito, Y., Akiyama, F., and Sakamoto, G. (2008). Clinical importance of estrogen receptor-beta evaluation in breast cancer patients treated with adjuvant tamoxifen therapy. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* 26, 3727–3734.

Hopp, T.A., Weiss, H.L., Parra, I.S., Cui, Y., Osborne, C.K., and Fuqua, S.A.W. (2004). Low levels of estrogen receptor beta protein predict resistance to tamoxifen therapy in breast cancer. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 10, 7490–7499.

Horard, B., and Vanacker, J.-M. (2003). Estrogen receptor-related receptors: orphan receptors desperately seeking a ligand. *J. Mol. Endocrinol.* 31, 349–357.

Hsu, L.-H., Chu, N.-M., Lin, Y.-F., and Kao, S.-H. (2019). G-Protein Coupled Estrogen Receptor in Breast Cancer. *Int. J. Mol. Sci.* 20, 306.

Hua, H., Zhang, H., Kong, Q., and Jiang, Y. (2018). Mechanisms for estrogen receptor expression in human cancer. *Exp. Hematol. Oncol.* 7.

Huan, J., Xing, L., Lin, Q., Xui, H., and Qin, X. (2017). Long noncoding RNA CRNDE activates Wnt/ $\beta$ -catenin signaling pathway through acting as a molecular sponge of microRNA-136 in human breast cancer. *Am. J. Transl. Res.* 9, 1977–1989.

Hughes, K., and Watson, C.J. (2018). The Multifaceted Role of STAT3 in Mammary Gland Involution and Breast Cancer. *Int. J. Mol. Sci.* 19.

Hurtado, A., Holmes, K.A., Geistlinger, T.R., Hutcheson, I.R., Nicholson, R.I., Brown, M., Jiang, J., Howat, W.J., Ali, S., and Carroll, J.S. (2008). Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. *Nature* 456, 663–666.

## I

Ing, N.H., Massuto, D.A., and Jaeger, L.A. (2008). Estradiol up-regulates AUF1p45 binding to stabilizing regions within the 3'-untranslated region of estrogen receptor alpha mRNA. *J. Biol. Chem.* 283, 1764–1772.

Ingman, W.V., and Robertson, S.A. (2008). Mammary gland development in transforming growth factor beta1 null mutant mice: systemic and epithelial effects. *Biol. Reprod.* 79, 711–717.



Inman, J.L., Robertson, C., Mott, J.D., and Bissell, M.J. (2015). Mammary gland development: cell fate specification, stem cells and the microenvironment. *Dev. Camb. Engl.* 142, 1028–1042.

## J

Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994). Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* 79, 107–117.

Javed, A., and Lteif, A. (2013). Development of the Human Breast. *Semin. Plast. Surg.* 27, 5–12.

Jéhannin-Ligier, K. (2017). Projection de l'incidence et de la mortalité par cancer en France métropolitaine en 2017. Rapport technique INCa. 80.

Jensen, E.V. (1962). On the mechanism of estrogen action. *Perspect. Biol. Med.* 6, 47–59.

Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M., and Schiff, R. (2015). ESR1 mutations as a mechanism for acquired endocrine resistance in breast cancer. *Nat. Rev. Clin. Oncol.* 12, 573–583.

Jeselsohn, R., Bergholz, J.S., Pun, M., Cornwell, M., Liu, W., Nardone, A., Xiao, T., Li, W., Qiu, X., Buchwalter, G. (2018). Allele-Specific Chromatin Recruitment and Therapeutic Vulnerabilities of ESR1 Activating Mutations. *Cancer Cell* 33, 173-186.e5.

Jia, M., Dahlman-Wright, K., and Gustafsson, J.-Å. (2015). Estrogen receptor alpha and beta in health and disease. *Best Pract. Res. Clin. Endocrinol. Metab.* 29, 557–568.

Jiang, H., Teng, R., Wang, Q., Zhang, X., Wang, H., Wang, Z., Cao, J., and Teng, L. (2008). Transcriptional analysis of estrogen receptor alpha variant mRNAs in colorectal cancers and their matched normal colorectal tissues. *J. Steroid Biochem. Mol. Biol.* 112, 20–24.

Jiang, Y., Liu, Y.E., Lu, A., Gupta, A., Goldberg, I.D., Liu, J., and Shi, Y.E. (2003). Stimulation of estrogen receptor signaling by gamma synuclein. *Cancer Res.* 63, 3899–3903.

Josso, N., and Rey, R. (1995). La cellule de Sertoli, une cellule endocrine. *médecine/sciences* 11, 537.

## K

Kagami, M., O'Sullivan, M.J., Green, A.J., Watabe, Y., Arisaka, O., Masawa, N., Matsuoka, K., Fukami, M., Matsubara, K., Kato, F., et al. (2010). The IG-DMR and the MEG3-DMR at Human Chromosome 14q32.2: Hierarchical Interaction and Distinct Functional Properties as Imprinting Control Centers. *PLoS Genet.* 6.

Kalaitzidis, D., and Gilmore, T.D. (2005). Transcription factor cross-talk: the estrogen receptor and NF- $\kappa$ B. *Trends Endocrinol. Metab.* 16, 46–52.

Kalet, B.T., Anglin, S.R., Handschy, A., O'Donoghue, L.E., Halsey, C., Chubb, L., Korch, C., and Duval, D.L. (2013). Transcription factor Ets1 cooperates with estrogen receptor  $\alpha$  to stimulate estradiol-dependent growth in breast cancer cells and tumors. *PLoS One* 8, e68815.

Kang, L., Zhang, X., Xie, Y., Tu, Y., Wang, D., Liu, Z., and Wang, Z.-Y. (2010). Involvement of Estrogen Receptor Variant ER- $\alpha$ 36, Not GPR30, in Nongenomic Estrogen Signaling. *Mol. Endocrinol.* 24, 709–721.

Kang, L., Guo, Y., Zhang, X., Meng, J., and Wang, Z.-Y. (2011a). A positive cross-regulation of HER2 and ER- $\alpha$ 36 controls ALDH1 positive breast cancer cells. *J. Steroid Biochem. Mol. Biol.* 127, 262–268.

- Kang, L., Wang, L., and Wang, Z.-Y. (2011b). Opposite regulation of estrogen receptor- $\alpha$  and its variant ER- $\alpha$ 36 by the Wilms' tumor suppressor WT1. *Oncol. Lett.* 2, 337–341.
- Kastrati, I., Semina, S., Gordon, B., and Smart, E. (2019). Insights into how phosphorylation of estrogen receptor at serine 305 modulates tamoxifen activity in breast cancer. *Mol. Cell. Endocrinol.* 483, 97–101.
- Keaveney, M., Klug, J., Dawson, M.T., Nestor, P.V., Neilan, J.G., Forde, R.C., and Gannon, F. (1991). Evidence for a previously unidentified upstream exon in the human oestrogen receptor gene. *J. Mol. Endocrinol.* 6, 111–115.
- Khan, D., and Ansar Ahmed, S. (2016). The Immune System Is a Natural Target for Estrogen Action: Opposing Effects of Estrogen in Two Prototypical Autoimmune Diseases. *Front. Immunol.* 6.
- Khosla, S., Oursler, M.J., and Monroe, D.G. (2012). Estrogen and the skeleton. *Trends Endocrinol. Metab.* TEM 23, 576–581.
- Kim, J., Petz, L.N., Ziegler, Y.S., Wood, J.R., Potthoff, S.J., and Nardulli, A.M. (2000). Regulation of the estrogen-responsive pS2 gene in MCF-7 human breast cancer cells. *J. Steroid Biochem. Mol. Biol.* 74, 157–168.
- Kirn, V., Strake, L., Thangarajah, F., Richters, L., Eischeid, H., Koitzsch, U., Odenthal, M., and Fries, J. (2018). ESR1-promoter-methylation status in primary breast cancer and its corresponding metastases. *Clin. Exp. Metastasis* 35, 707–712.
- Klein-Hitpass, L., Kaling, M., and Ryffel, G.U. (1988). Synergism of closely adjacent estrogen-responsive elements increases their regulatory potential. *J. Mol. Biol.* 201, 537–544.
- Klinge, C.M. (2001). Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res.* 29, 2905–2919.
- Kok, M., Holm-Wigerup, C., Hauptmann, M., Michalides, R., Stål, O., Linn, S., and Landberg, G. (2009). Estrogen receptor- $\alpha$  phosphorylation at serine-118 and tamoxifen response in breast cancer. *J. Natl. Cancer Inst.* 101, 1725–1729.
- Kong, S.L., Li, G., Loh, S.L., Sung, W.-K., and Liu, E.T. (2011). Cellular reprogramming by the conjoint action of ER $\alpha$ , FOXA1, and GATA3 to a ligand-inducible growth state. *Mol. Syst. Biol.* 7, 526.
- Kos, M., Reid, G., Denger, S., and Gannon, F. (2001). Minireview: genomic organization of the human ER $\alpha$  gene promoter region. *Mol. Endocrinol. Baltim. Md* 15, 2057–2063.
- Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D., and Werb, Z. (2006). GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* 127, 1041–1055.
- Kovats, S. (2015). Estrogen receptors regulate innate immune cells and signaling pathways. *Cell. Immunol.* 294, 63–69.
- Krishnan, V., Porter, W., Santostefano, M., Wang, X., and Safe, S. (1995). Molecular mechanism of inhibition of estrogen-induced cathepsin D gene expression by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in MCF-7 cells. *Mol. Cell. Biol.* 15, 6710–6719.
- Kuiper, G.G., Enmark, E., Peltö-Huikko, M., Nilsson, S., and Gustafsson, J.A. (1996). Cloning of a novel receptor expressed in rat prostate and ovary. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5925–5930.
- Kumar, R., and Thompson, E.B. (1999). The structure of the nuclear hormone receptors. *Steroids* 64, 310–319.

Kumar, R., Zakharov, M.N., Khan, S.H., Miki, R., Jang, H., Toraldo, G., Singh, R., Bhasin, S., and Jasuja, R. (2011). The Dynamic Structure of the Estrogen Receptor. *J. Amino Acids* 2011.

## L

Lakhani, N.J., Sarkar, M.A., Venitz, J., and Figg, W.D. (2003). 2-Methoxyestradiol, a promising anticancer agent. *Pharmacotherapy* 23, 165–172.

Laudet, V. (1997). Evolution of the nuclear receptor superfamily: early diversification from an ancestral orphan receptor. *J. Mol. Endocrinol.* 19, 207–226.

Le Quesne, J., and Caldas, C. (2010). Micro-RNAs and breast cancer. *Mol. Oncol.* 4, 230–241.

Le Romancer, M., Poulard, C., Cohen, P., Sentis, S., Renoir, J.-M., and Corbo, L. (2011). Cracking the estrogen receptor's posttranslational code in breast tumors. *Endocr. Rev.* 32, 597–622.

Lee, E.Y.H.P., and Muller, W.J. (2010). *Oncogenes and Tumor Suppressor Genes.* Cold Spring Harb. Perspect. Biol. 2.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. *Annu. Rev. Genet.* 34, 77–137.

Li, D.-X., Fei, X.-R., Dong, Y.-F., Cheng, C.-D., Yang, Y., Deng, X.-F., Huang, H.-L., Niu, W.-X., Zhou, C.-X., Xia, C.-Y. (2017). The long non-coding RNA CRNDE acts as a ceRNA and promotes glioma malignancy by preventing miR-136-5p-mediated downregulation of Bcl-2 and Wnt2. *Oncotarget* 8.

Li, L., Haynes, M.P., and Bender, J.R. (2003). Plasma membrane localization and function of the estrogen receptor alpha variant (ER46) in human endothelial cells. *Proc. Natl. Acad. Sci. U. S. A.* 100, 4807–4812.

Li, T.-T., Gao, X., Gao, L., Gan, B.-L., Xie, Z.-C., Zeng, J.-J., and Chen, G. (2018). Role of upregulated miR-136-5p in lung adenocarcinoma: A study of 1242 samples utilizing bioinformatics analysis. *Pathol. Res. Pract.* 214, 750–766.

Lillo Osuna, M.A., Garcia-Lopez, J., El Ayachi, I., Fatima, I., Khalid, A.B., Kumpati, J., Slayden, A.V., Seagroves, T.N., Miranda-Carboni, G.A., and Krum, S.A. (2019). Activation of Estrogen Receptor Alpha by Decitabine Inhibits Osteosarcoma Growth and Metastasis. *Cancer Res.* 79, 1054–1068.

Lin, A.H.Y., Li, R.W.S., Ho, E.Y.W., Leung, G.P.H., Leung, S.W.S., Vanhoutte, P.M., and Man, R.Y.K. (2013). Differential Ligand Binding Affinities of Human Estrogen Receptor- $\alpha$  Isoforms. *PLoS ONE* 8.

Lin, H., Dai, T., Xiong, H., Zhao, X., Chen, X., Yu, C., Li, J., Wang, X., and Song, L. (2010). Unregulated miR-96 Induces Cell Proliferation in Human Breast Cancer by Downregulating Transcriptional Factor FOXO3a. *PLoS ONE* 5.

Lin, S.-L., Yan, L.-Y., Zhang, X.-T., Yuan, J., Li, M., Qiao, J., Wang, Z.-Y., and Sun, Q.-Y. (2010). ER- $\alpha$ 36, a Variant of ER- $\alpha$ , Promotes Tamoxifen Agonist Action in Endometrial Cancer Cells via the MAPK/ERK and PI3K/Akt Pathways. *PLoS ONE* 5.

Lin, Z., Yin, P., Reierstad, S., O'Halloran, M., Coon, V.J.S., Pearson, E.K., Mutlu, G.M., and Bulun, S.E. (2010). Adenosine A1 receptor, a target and regulator of estrogen receptoralpha action, mediates the proliferative effects of estradiol in breast cancer. *Oncogene* 29, 1114–1122.

Lipovka, Y., and Konhilas, J.P. (2016). The complex nature of oestrogen signalling in breast cancer: enemy or ally? *Biosci. Rep.* 36.

Liu, Y.E., Pu, W., Jiang, Y., Shi, D., Dackour, R., and Shi, Y.E. (2007). Chaperoning of estrogen receptor and induction of mammary gland proliferation by neuronal protein synuclein gamma. *Oncogene* 26, 2115–2125.

Lou, W., Liu, J., Ding, B., Xu, L., and Fan, W. (2018). Identification of chemoresistance-associated miRNAs in breast cancer. *Cancer Manag. Res.* 10, 4747–4757.

Low, S., Zembutsu, H., and Nakamura, Y. (2018). Breast cancer: The translation of big genomic data to cancer precision medicine. *Cancer Sci.* 109, 497–506.

Lu, Q., Pallas, D.C., Surks, H.K., Baur, W.E., Mendelsohn, M.E., and Karas, R.H. (2004). Striatin assembles a membrane signaling complex necessary for rapid, nongenomic activation of endothelial NO synthase by estrogen receptor alpha. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17126–17131.

## M

Ma, B., and Hottiger, M.O. (2016). Crosstalk between Wnt/ $\beta$ -Catenin and NF- $\kappa$ B Signaling Pathway during Inflammation. *Front. Immunol.* 7.

Macias, H., and Hinck, L. (2012). Mammary Gland Development. *Wiley Interdiscip. Rev. Dev. Biol.* 1, 533–557.

Madureira, P.A., Varshochi, R., Constantinidou, D., Francis, R.E., Coombes, R.C., Yao, K.-M., and Lam, E.W.-F. (2006). The Forkhead box M1 protein regulates the transcription of the estrogen receptor alpha in breast cancer cells. *J. Biol. Chem.* 281, 25167–25176.

Maekawa, R., Sato, S., Okada, M., Lee, L., Tamura, I., Jozaki, K., Kajimura, T., Asada, H., Yamagata, Y., Tamura, H. (2016). Tissue-Specific Expression of Estrogen Receptor 1 Is Regulated by DNA Methylation in a T-DMR. *Mol. Endocrinol. Baltim. Md* 30, 335–347.

Magnani, L., Ballantyne, E.B., Zhang, X., and Lupien, M. (2011). PBX1 Genomic Pioneer Function Drives ER $\alpha$  Signaling Underlying Progression in Breast Cancer. *PLOS Genet.* 7, e1002368.

Malhotra, G.K., Zhao, X., Band, H., and Band, V. (2010). Histological, molecular and functional subtypes of breast cancers. *Cancer Biol. Ther.* 10, 955–960.

Maller, O., Martinson, H., and Schedin, P. (2010). Extracellular matrix composition reveals complex and dynamic stromal-epithelial interactions in the mammary gland. *J. Mammary Gland Biol. Neoplasia* 15, 301–318.

Manavathi, B., Dey, O., Gajulapalli, V.N.R., Bhatia, R.S., Bugide, S., and Kumar, R. (2013). Derailed Estrogen Signaling and Breast Cancer: An Authentic Couple. *Endocr. Rev.* 34, 1–32.

Manikkam, M., Guerrero-Bosagna, C., Tracey, R., Haque, M.M., and Skinner, M.K. (2012). Transgenerational actions of environmental compounds on reproductive disease and identification of epigenetic biomarkers of ancestral exposures. *PloS One* 7, e31901.

Markiewicz, A., Wełnicka-Jaśkiewicz, M., Skokowski, J., Jaśkiewicz, J., Szade, J., Jassem, J., and Zaczek, A.J. (2013). Prognostic significance of ESR1 amplification and ESR1 PvuII, CYP2C19\*2, UGT2B15\*2 polymorphisms in breast cancer patients. *PloS One* 8, e72219.

- Marotta, L.L.C., Almendro, V., Marusyk, A., Shipitsin, M., Schemme, J., Walker, S.R., Bloushtain-Qimron, N., Kim, J.J., Choudhury, S.A., Maruyama, R., et al. (2011). The JAK2/STAT3 signaling pathway is required for growth of CD44+CD24- stem cell-like breast cancer cells in human tumors. *J. Clin. Invest.* 121, 2723–2735.
- Marshall, W.A., and Tanner, J.M. (1969). Variations in pattern of pubertal changes in girls. *Arch. Dis. Child.* 44, 291–303.
- Mathews, J., and Gustafsson, J.-Å. (2006). Estrogen receptor and aryl hydrocarbon receptor signaling pathways. *Nucl. Recept. Signal.* 4.
- Mathews, J.R., and Clarke, A.R. (2005). p53 mediates a default programme of mammary gland involution in the absence of STAT3. *Oncogene* 24, 3083–3090.
- McBryan, J., Howlin, J., Napoletano, S., and Martin, F. (2008). Amphiregulin: role in mammary gland development and breast cancer. *J. Mammary Gland Biol. Neoplasia* 13, 159–169.
- McEwan, I.J. (2016). The Nuclear Receptor Superfamily at Thirty. *Methods Mol. Biol. Clifton NJ* 1443, 3–9.
- McPherson, L.A., and Weigel, R.J. (1999). AP2alpha and AP2gamma: a comparison of binding site specificity and trans-activation of the estrogen receptor promoter and single site promoter constructs. *Nucleic Acids Res.* 27, 4040–4049.
- McPherson, L.A., Baichwal, V.R., and Weigel, R.J. (1997). Identification of ERF-1 as a member of the AP2 transcription factor family. *Proc. Natl. Acad. Sci. U. S. A.* 94, 4342–4347.
- Mehra, R., Varambally, S., Ding, L., Shen, R., Sabel, M.S., Ghosh, D., Chinnaiyan, A.M., and Kleer, C.G. (2005). Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res.* 65, 11259–11264.
- Menasce, L.P., White, G.R., Harrison, C.J., and Boyle, J.M. (1993). Localization of the estrogen receptor locus (ESR) to chromosome 6q25.1 by FISH and a simple post-FISH banding technique. *Genomics* 17, 263–265.
- Menezes, M.E., Das, S.K., Emdad, L., Windle, J.J., Wang, X.Y., Sarkar, D., and Fisher, P.B. (2014). Genetically engineered mice as experimental tools to dissect the critical events in breast cancer. *Adv. Cancer Res.* 121, 331–382.
- Menon, K.M.J., and Goldstrohm, A. (2015). *Post-transcriptional Mechanisms in Endocrine Regulation* (Springer).
- Merlin, J.L. (2008). Les inhibiteurs de tyrosine kinase en oncologie – Tyrosine kinase inhibitors in oncology. *La lettre du pharmacologue* 22(2), 51–62.
- Meunier, L., Siddeek, B., Vega, A., Lakhdari, N., Inoubli, L., Bellon, R.P., Lemaire, G., Mauduit, C., and Benahmed, M. (2012). Perinatal programming of adult rat germ cell death after exposure to xenoestrogens: role of microRNA miR-29 family in the down-regulation of DNA methyltransferases and Mcl-1. *Endocrinology* 153, 1936–1947.
- Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.-S., Tam, W.-L., Thomson, A.M., Lim, B., and Rigoutsos, I. (2006). A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* 126, 1203–1217.
- Misawa, A., and Inoue, S. (2015). Estrogen-Related Receptors in Breast Cancer and Prostate Cancer. *Front. Endocrinol.* 6, 83.

Mohammed, H., Russell, I.A., Stark, R., Rueda, O.M., Hickey, T.E., Tarulli, G.A., Serandour, A.A., Serandour, A.A.A., Birrell, S.N., Bruna, A., et al. (2015). Progesterone receptor modulates ER $\alpha$  action in breast cancer. *Nature* 523, 313–317.

Mohandass, J., Ravichandran, S., Srilakshmi, K., Rajadurai, C.P., Sanmugasamy, S., and Kumar, G.R. (2010). BCDB - A database for breast cancer research and information. *Bioinformatics* 5, 1–3.

Morales, S., Monzo, M., and Navarro, A. (2017). Epigenetic regulation mechanisms of microRNA expression. *Biomol. Concepts* 8.

Moriarty, K., Kim, K.H., and Bender, J.R. (2006). Estrogen Receptor-Mediated Rapid Signaling. *Endocrinology* 147, 5557–5563.

Mosselman, S., Polman, J., and Dijkema, R. (1996). ER beta: identification and characterization of a novel human estrogen receptor. *FEBS Lett.* 392, 49–53.

Moumen, M., Chiche, A., Cagnet, S., Petit, V., Raymond, K., Faraldo, M.M., Deugnier, M.-A., and Glukhova, M.A. (2011). The mammary myoepithelial cell. *Int. J. Dev. Biol.* 55, 763–771.

Mueller, S.O., Clark, J.A., Myers, P.H., and Korach, K.S. (2002). Mammary gland development in adult mice requires epithelial and stromal estrogen receptor alpha. *Endocrinology* 143, 2357–2365.

Muñoz-de-Toro, M., Markey, C.M., Wadia, P.R., Luque, E.H., Rubin, B.S., Sonnenschein, C., and Soto, A.M. (2005). Perinatal Exposure to Bisphenol-A Alters Peripubertal Mammary Gland Development in Mice. *Endocrinology* 146, 4138–4147.

Musgrove, E.A., and Sutherland, R.L. (2009). Biological determinants of endocrine resistance in breast cancer. *Nat. Rev. Cancer* 9, 631–643.

## N

Nagel, A., Szade, J., Iliszko, M., Elzanowska, J., Welnicka-Jaskiewicz, M., Skokowski, J., Stasiłojc, G., Bigda, J., Sadej, R., Zaczek, A. (2019). Clinical and Biological Significance of ESR1 Gene Alteration and Estrogen Receptors Isoforms Expression in Breast Cancer Patients. *Int. J. Mol. Sci.* 20.

Naimi, A., Rajabi, P., Ghasemibasir, H., Eftekhari, A., and Mohammadizadeh, F. (2009). Expression of basal and luminal cytokeratins in breast cancer and their correlation with clinicopathological prognostic variables. *Indian J. Med. Sci.* 63, 152.

Najim, O., Huizing, M., Papadimitriou, K., Trinh, X.B., Pauwels, P., Goethals, S., Zwaenepoel, K., Peterson, K., Weyler, J., Altintas, S. (2019). The prevalence of estrogen receptor-1 mutation in advanced breast cancer: The estrogen receptor one study (EROS1). *Cancer Treat. Res. Commun.* 19, 100123.

Nebbioso, A., Tambaro, F.P., Dell'Aversana, C., and Altucci, L. (2018). Cancer epigenetics: Moving forward. *PLoS Genet.* 14, e1007362.

Need, E.F., Atashgaran, V., Ingman, W.V., and Dasari, P. (2014). Hormonal regulation of the immune microenvironment in the mammary gland. *J. Mammary Gland Biol. Neoplasia* 19, 229–239.

Nettles, K.W., Gil, G., Nowak, J., Métivier, R., Sharma, V.B., and Greene, G.L. (2008). CBP Is a dosage-dependent regulator of nuclear factor-kappaB suppression by the estrogen receptor. *Mol. Endocrinol. Baltim. Md* 22, 263–272.

Neville, M.C., McFadden, T.B., and Forsyth, I. (2002). Hormonal Regulation of Mammary Differentiation and Milk Secretion. *J. Mammary Gland Biol. Neoplasia* 19.

Nilsson, E.E., Anway, M.D., Stanfield, J., and Skinner, M.K. (2008). Transgenerational epigenetic effects of the endocrine disruptor vinclozolin on pregnancies and female adult onset disease. *Reprod. Camb. Engl.* 135, 713–721.

Norfleet, A.M., Thomas, M.L., Gametchu, B., and Watson, C.S. (1999). Estrogen receptor-alpha detected on the plasma membrane of aldehyde-fixed GH3/B6/F10 rat pituitary tumor cells by enzyme-linked immunocytochemistry. *Endocrinology* 140, 3805–3814.

## O

Obsil, T., and Obsilova, V. (2008). Structure/function relationships underlying regulation of FOXO transcription factors. *Oncogene* 27, 2263–2275.

Ogawa, S., Inoue, S., Watanabe, T., Orimo, A., Hosoi, T., Ouchi, Y., and Muramatsu, M. (1998). Molecular cloning and characterization of human estrogen receptor betacx: a potential inhibitor of estrogen action in human. *Nucleic Acids Res.* 26, 3505–3512.

O’Lone, R., Frith, M.C., Karlsson, E.K., and Hansen, U. (2004). Genomic Targets of Nuclear Estrogen Receptors. *Mol. Endocrinol.* 18, 1859–1875.

Omarjee, S., Jacquemetton, J., Poulard, C., Rochel, N., Dejaegere, A., Chebaro, Y., Treilleux, I., Marangoni, E., Corbo, L., and Romancer, M.L. (2017). The molecular mechanisms underlying the ER $\alpha$ -36-mediated signaling in breast cancer. *Oncogene* 36, 2503–2514.

Omoto, Y., and Iwase, H. (2015). Clinical significance of estrogen receptor  $\beta$  in breast and prostate cancer from biological aspects. *Cancer Sci.* 106, 337–343.

## P

Panda, A.C. (2018). Circular RNAs Act as miRNA Sponges. *Adv. Exp. Med. Biol.* 1087, 67–79.

Pandey, D.P., and Picard, D. (2009). miR-22 inhibits estrogen signaling by directly targeting the estrogen receptor alpha mRNA. *Mol. Cell. Biol.* 29, 3783–3790.

Panneerselvam, M., Muthu, K., and Ramadas, K. (2015). Structural insights into tumor-specific chaperoning activity of gamma synuclein in protecting estrogen receptor alpha 36 and its role in tamoxifen resistance in breast cancer. *Mol. Biosyst.* 11, 2998–3010.

Panwar, B., Omenn, G.S., and Guan, Y. (2017). miRmine: a database of human miRNA expression profiles. *Bioinforma. Oxf. Engl.* 33, 1554–1560.

Pavlopoulou, A., Spandidos, D.A., and Michalopoulos, I. (2015). Human cancer databases (Review). *Oncol. Rep.* 33, 3–18.

Pellikainen, J., Naukkarinen, A., Ropponen, K., Rummukainen, J., Kataja, V., Kellokoski, J., Eskelinen, M., and Kosma, V.-M. (2004). Expression of HER2 and its association with AP-2 in breast cancer. *Eur. J. Cancer* 40, 1485–1495.

Penolazzi, L., Lambertini, E., Giordano, S., Sollazzo, V., Traina, G., del Senno, L., and Piva, R. (2004). Methylation analysis of the promoter F of estrogen receptor alpha gene: effects on the level of transcription on human osteoblastic cells. *J. Steroid Biochem. Mol. Biol.* 91, 1–9.

Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslén, L.A. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747–752.

Peters, A.A., Buchanan, G., Ricciardelli, C., Bianco-Miotto, T., Centenera, M.M., Harris, J.M., Jindal, S., Segara, D., Jia, L., Moore, N.L. (2009). Androgen receptor inhibits estrogen receptor- $\alpha$  activity and is prognostic in breast cancer. *Cancer Res.* 69, 6131–6140.

Pietras, R.J., and Szego, C.M. (1975). Endometrial cell calcium and oestrogen action. *Nature* 253, 357.

Piggott, L., Silva, A., Robinson, T., Santiago-Gómez, A., Simões, B.M., Becker, M., Fichtner, I., Andera, L., Young, P., Morris, C. (2018). Acquired Resistance of ER-Positive Breast Cancer to Endocrine Treatment Confers an Adaptive Sensitivity to TRAIL through Posttranslational Downregulation of c-FLIP. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 24, 2452–2463.

Ponglikitmongkol, M., Green, S., and Chambon, P. (1988). Genomic organization of the human oestrogen receptor gene. *EMBO J.* 7, 3385–3388.

Poulard, C., Jacquemetton, J., Trédan, O., Cohen, P.A., Vendrell, J., Ghayad, S.E., Treilleux, I., Marangoni, E., and Le Romancer, M. (2019). Oestrogen Non-Genomic Signalling is Activated in Tamoxifen-Resistant Breast Cancer. *Int. J. Mol. Sci.* 20.

Powell, J.B., Goode, G.D., and Eltom, S.E. (2013). The Aryl Hydrocarbon Receptor: A Target for Breast Cancer Therapy. *J. Cancer Ther.* 4, 1177–1186.

Pratt, W.B., and Toft, D.O. (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp. Biol. Med.* Maywood NJ 228, 111–133.

Pryzbylkowski, P., Obajimi, O., and Keen, J.C. (2008). Trichostatin A and 5 Aza-2' deoxycytidine decrease estrogen receptor mRNA stability in ER positive MCF7 cells through modulation of HuR. *Breast Cancer Res. Treat.* 111, 15–25.

Pupo, M., Vivacqua, A., Perrotta, I., Pisano, A., Aquila, S., Abonante, S., Gasperi-Campani, A., Pezzi, V., and Maggiolini, M. (2013). The nuclear localization signal is required for nuclear GPER translocation and function in breast Cancer-Associated Fibroblasts (CAFs). *Mol. Cell. Endocrinol.* 376, 23–32.

Pupo, M., Maggiolini, M., and Musti, A.M. (2016). GPER Mediates Non-Genomic Effects of Estrogen. *Methods Mol. Biol.* Clifton NJ 1366, 471–488.

Pupo, M., Bodmer, A., Berto, M., Maggiolini, M., Dietrich, P.-Y., and Picard, D. (2017). A genetic polymorphism repurposes the G-protein coupled and membrane-associated estrogen receptor GPER to a transcription factor-like molecule promoting paracrine signaling between stroma and breast carcinoma cells. *Oncotarget* 8, 46728–46744.

Puzianowska-Kuznicka, M., Pawlik-Pachucka, E., Owczarż, M., Budzińska, M., and Polosak, J. (2013). Small-Molecule Hormones: Molecular Mechanisms of Action. *Int. J. Endocrinol.* 2013.

## Q

Qian, H., Xuan, J., Liu, Y., and Shi, G. (2016). Function of G-Protein-Coupled Estrogen Receptor-1 in Reproductive System Tumors. *J. Immunol. Res.* 2016.

## R

Rakesh, K., and W, O.B. (2008). *Nuclear Receptors Coregulators And Human Diseases* (World Scientific).



Rao, J., Jiang, X., Wang, Y., and Chen, B. (2011). Advances in the understanding of the structure and function of ER- $\alpha$ 36, a novel variant of human estrogen receptor- $\alpha$ . *J. Steroid Biochem. Mol. Biol.* 127, 231–237.

Razandi, M., Pedram, A., Rosen, E.M., and Levin, E.R. (2004). BRCA1 Inhibits Membrane Estrogen and Growth Factor Receptor Signaling to Cell Proliferation in Breast Cancer. *Mol. Cell. Biol.* 24, 5900–5913.

Rebeck, T.R., Kauff, N.D., and Domchek, S.M. (2009). Meta-analysis of risk reduction estimates associated with risk-reducing salpingo-oophorectomy in BRCA1 or BRCA2 mutation carriers. *J. Natl. Cancer Inst.* 101, 80–87.

Reid, G., Hübner, M.R., Métivier, R., Brand, H., Denger, S., Manu, D., Beaudouin, J., Ellenberg, J., and Gannon, F. (2003). Cyclic, proteasome-mediated turnover of unliganded and liganded ER $\alpha$  on responsive promoters is an integral feature of estrogen signaling. *Mol. Cell* 11, 695–707.

Revankar, C.M., Cimino, D.F., Sklar, L.A., Arterburn, J.B., and Prossnitz, E.R. (2005). A Transmembrane Intracellular Estrogen Receptor Mediates Rapid Cell Signaling. 307, 6.

Reyal, F., Guyader, C., Decraene, C., Lucchesi, C., Auger, N., Assayag, F., De Plater, L., Gentien, D., Poupon, M.-F., Cottu, P. (2012). Molecular profiling of patient-derived breast cancer xenografts. *Breast Cancer Res.* 14, R11.

Rezaei, R., Wu, Z., Hou, Y., Bazer, F.W., and Wu, G. (2016). Amino acids and mammary gland development: nutritional implications for milk production and neonatal growth. *J. Anim. Sci. Biotechnol.* 7.

Reznikov, A. (2015). Hormonal impact on tumor growth and progression. *Exp. Oncol.* 37, 162–172.

Robinson, G.W. (2007). Cooperation of signalling pathways in embryonic mammary gland development. *Nat. Rev. Genet.* 8, 963–972.

Robinson, D.R., Wu, Y.-M., Vats, P., Su, F., Lonigro, R.J., Cao, X., Kalyana-Sundaram, S., Wang, R., Ning, Y., Hodges, L. (2013). Activating ESR1 mutations in hormone-resistant metastatic breast cancer. *Nat. Genet.* 45, 1446–1451.

Rochira, V., and Carani, C. (2009). Aromatase deficiency in men: a clinical perspective. *Nat. Rev. Endocrinol.* 5, 559–568.

Rodgers, K.M., Udesky, J.O., Rudel, R.A., and Brody, J.G. (2018). Environmental chemicals and breast cancer: An updated review of epidemiological literature informed by biological mechanisms. *Environ. Res.* 160, 152–182.

## S

Sabbah, M., Kang, K.I., Tora, L., and Redeuilh, G. (1998). Oestrogen receptor facilitates the formation of preinitiation complex assembly: involvement of the general transcription factor TFIIB. *Biochem. J.* 336, 639–646.

Safe, S., and Kim, K. (2008). Nonclassical genomic ER/Sp and ER/AP-1 signaling pathways. *J. Mol. Endocrinol.* 41, 263–275.

Sakakura, T., Suzuki, Y., and Shiurba, R. (2013). Mammary stroma in development and carcinogenesis. *J. Mammary Gland Biol. Neoplasia* 18, 189–197.

- Samavat, H., and Kurzer, M.S. (2015). Estrogen metabolism and breast cancer. *Cancer Lett.* 356, 231–243.
- Sasano, H., Nagura, H., Harada, N., Goukon, Y., and Kimura, M. (1994). Immunolocalization of aromatase and other steroidogenic enzymes in human breast disorders. *Hum. Pathol.* 25, 530–535.
- Sato, T., Neilson, L.M., Peck, A.R., Liu, C., Tran, T.H., Witkiewicz, A., Hyslop, T., Nevalainen, M.T., Sauter, G., and Rui, H. (2011). Signal transducer and activator of transcription-3 and breast cancer prognosis. *Am. J. Cancer Res.* 1, 347–355.
- Savkur, R.S., and Burris, T.P. (2004). The coactivator LXXLL nuclear receptor recognition motif. *J. Pept. Res.* 63, 207–212.
- Scarabin-Carré, V., Canonico, M., Brailly-Tabard, S., Trabado, S., Ducimetière, P., Giroud, M., Ryan, J., Helmer, C., Plu-Bureau, G., Guiochon-Mantel, A. (2012). High Level of Plasma Estradiol as a New Predictor of Ischemic Arterial Disease in Older Postmenopausal Women: The Three-City Cohort Study. *J. Am. Heart Assoc.* 1.
- Schug, T.T., Janesick, A., Blumberg, B., and Heindel, J.J. (2011). Endocrine Disrupting Chemicals and Disease Susceptibility. *J. Steroid Biochem. Mol. Biol.* 127, 204–215.
- Schulster, M., Bernie, A.M., and Ramasamy, R. (2016). The role of estradiol in male reproductive function. *Asian J. Androl.* 18, 435–440.
- Schuur, E.R., McPherson, L.A., Yang, G.P., and Weigel, R.J. (2001). Genomic structure of the promoters of the human estrogen receptor-alpha gene demonstrate changes in chromatin structure induced by AP2gamma. *J. Biol. Chem.* 276, 15519–15526.
- Seachrist, D.D., Bonk, K.W., Ho, S.-M., Prins, G.S., Soto, A.M., and Keri, R.A. (2016). A Review of the Carcinogenic Potential of Bisphenol A. *Reprod. Toxicol. Elmsford N* 59, 167–182.
- Selye, H. (1942). Correlations between the chemical structure and the pharmacological actions of the steroids. *Endocrinology* 30, 437–453.
- Sempere, L.F. (2019). Celebrating 25 Years of MicroRNA Research: From Discovery to Clinical Application. *Int. J. Mol. Sci.* 20, 1987.
- Sharma, G.N., Dave, R., Sanadya, J., Sharma, P., and Sharma, K.K. (2010). Various types and management of breast cancer: an overview. *J. Adv. Pharm. Technol. Res.* 1, 109–126.
- Shehu, A., Albarracin, C., Devi, Y.S., Luther, K., Halperin, J., Le, J., Mao, J., Duan, R.W., Frasor, J., and Gibori, G. (2011). The stimulation of HSD17B7 expression by estradiol provides a powerful feed-forward mechanism for estradiol biosynthesis in breast cancer cells. *Mol. Endocrinol. Baltim. Md* 25, 754–766.
- Shi, L., Dong, B., Li, Z., Lu, Y., Ouyang, T., Li, J., Wang, T., Fan, Z., Fan, T., Lin, B. (2009). Expression of ER- $\alpha$ 36, a Novel Variant of Estrogen Receptor  $\alpha$ , and Resistance to Tamoxifen Treatment in Breast Cancer. *J. Clin. Oncol.* 27, 3423–3429.
- Shi, Y.E., Chen, Y., Dackour, R., Potters, L., Wang, S., Ding, Q., Wang, Z., and Liu, Y.E. (2010). Synuclein gamma stimulates membrane-initiated estrogen signaling by chaperoning estrogen receptor (ER)-alpha36, a variant of ER-alpha. *Am. J. Pathol.* 177, 964–973.
- Siersbæk, R., Kumar, S., and Carroll, J.S. (2018). Signaling pathways and steroid receptors modulating estrogen receptor  $\alpha$  function in breast cancer. *Genes Dev.* 32, 1141–1154.

Sołtysik, K., and Czekał, P. (2015). ER $\alpha$ 36--Another piece of the estrogen puzzle. *Eur. J. Cell Biol.* 94, 611–625.

Stresemann, C., and Lyko, F. (2008). Modes of action of the DNA methyltransferase inhibitors azacytidine and decitabine. *Int. J. Cancer* 123, 8–13.

Suetsugi, M., Su, L., Karlsberg, K., Yuan, Y.-C., and Chen, S. (2003). Flavone and isoflavone phytoestrogens are agonists of estrogen-related receptors. *Mol. Cancer Res. MCR* 1, 981–991.

Sun, L., Wang, J., Zhang, L., Li, X., and Shen, D. (2013). Expression of ER- $\alpha$ 36, a novel variant of estrogen receptor in endometrial carcinoma and its clinical significance. *Gynecol. Obstet. Invest.* 75, 68–72.

Szego, C.M., and Davis, J.S. (1967). Adenosine 3',5'-monophosphate in rat uterus: acute elevation by estrogen. *Proc. Natl. Acad. Sci. U. S. A.* 58, 1711–1718.

## T

Tan, W., Liu, B., Qu, S., Liang, G., Luo, W., and Gong, C. (2018). MicroRNAs and cancer: Key paradigms in molecular therapy. *Oncol. Lett.* 15, 2735–2742.

Tanimoto, K., Eguchi, H., Yoshida, T., Hajiro-Nakanishi, K., and Hayashi, S. (1999). Regulation of estrogen receptor alpha gene mediated by promoter B responsible for its enhanced expression in human breast cancer. *Nucleic Acids Res.* 27, 903–909.

Tanos, T., and Brisken, C. (2008). What Signals Operate in the Mammary Niche? *Breast Dis.* 29, 69–82.

Tata, J.R. (2005). One hundred years of hormones. *EMBO Rep.* 6, 490–496.

Tay, Y., Zhang, J., Thomson, A.M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455, 1124–1128.

Teymourzadeh, A., Mansouri, S., Farahmand, L., Hosseinzade, A., and Majidzadeh-A, K. (2017). ER- $\alpha$ 36 Interactions With Cytosolic Molecular Network in Acquired Tamoxifen Resistance. *Clin. Breast Cancer* 17, 403–407.

Thomas, P., and Dong, J. (2006). Binding and activation of the seven-transmembrane estrogen receptor GPR30 by environmental estrogens: a potential novel mechanism of endocrine disruption. *J. Steroid Biochem. Mol. Biol.* 102, 175–179.

Thomas, P., Pang, Y., Filardo, E.J., and Dong, J. (2005). Identity of an estrogen membrane receptor coupled to a G protein in human breast cancer cells. *Endocrinology* 146, 624–632.

Thornton, J.W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5671–5676.

Tilghman, S.L., Bratton, M.R., Segar, H.C., Martin, E.C., Rhodes, L.V., Li, M., McLachlan, J.A., Wiese, T.E., Nephew, K.P., and Burow, M.E. (2012). Endocrine disruptor regulation of microRNA expression in breast carcinoma cells. *PloS One* 7, e32754.

To, S.Q., Cheung, V., Lazarus, K.A., Knowler, K.C., and Clyne, C.D. (2014). Estradiol regulates Tumor Necrosis Factor- $\alpha$  expression and secretion in Estrogen Receptor positive breast cancer cells. *Mol. Cell. Endocrinol.* 394, 21–28.

Tremblay, G.B., Bergeron, D., and Giguere, V. (2001). 4-Hydroxytamoxifen is an isoform-specific inhibitor of orphan estrogen-receptor-related (ERR) nuclear receptors beta and gamma. *Endocrinology* 142, 4572–4575.

Tsuboi, K., Nagatomo, T., Gohno, T., Higuchi, T., Sasaki, S., Fujiki, N., Kurosumi, M., Takei, H., Yamaguchi, Y., Niwa, T. (2017). Single CpG site methylation controls estrogen receptor gene transcription and correlates with hormone therapy resistance. *J. Steroid Biochem. Mol. Biol.* 171, 209–217.

Turashvili, G., and Brogi, E. (2017). Tumor Heterogeneity in Breast Cancer. *Front. Med.* 4.

Turner, B.C., Zhang, J., Gumbs, A.A., Maher, M.G., Kaplan, L., Carter, D., Glazer, P.M., Hurst, H.C., Haffty, B.G., and Williams, T. (1998). Expression of AP-2 transcription factors in human breast cancer correlates with the regulation of multiple growth factor signalling pathways. *Cancer Res.* 58, 5466–5472.

## V

Vacher, S., Castagnet, P., Chemlali, W., Lallemand, F., Meseure, D., Pocard, M., Bieche, I., and Perrot-Appinat, M. (2018). High AHR expression in breast tumors correlates with expression of genes from several signaling pathways namely inflammation and endogenous tryptophan metabolism. *PLoS ONE* 13.

Velloso, F., Bianco, A., Farias, J.O., Torres, N.E., Ferruzo, P.Y., Anschau, V., Jesus-Ferreira, H.C., Chang, T.H.-T., Sogayar, M., Zerbini, L.F. (2017). The crossroads of breast cancer progression: insights into the modulation of major signaling pathways. *Onco Targets Ther.* 10, 5491–5524

Veltmaat, J.M., Mailleux, A.A., Thiery, J.P., and Bellusci, S. (2003). Mouse embryonic mammaryogenesis as a model for the molecular regulation of pattern formation. *Differ. Res. Biol. Divers.* 71, 1–17.

Visvader, J.E. (2009). Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev.* 23, 2563–2577.

Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev.* 28, 1143–1158.

Vivacqua, A., Romeo, E., De Marco, P., De Francesco, E.M., Abonante, S., and Maggiolini, M. (2012). GPER mediates the Egr-1 expression induced by 17 $\beta$ -estradiol and 4-hydroxytamoxifen in breast and endometrial cancer cells. *Breast Cancer Res. Treat.* 133, 1025–1035.

Von Roemeling, C.A., Mathias, Adam, Dawson, L., Harrison, Steadman, Marlow, L., Miller, J., Hollister, Beth, Radisky, D., and Copland, J. (2014). Estrogen receptor mRNA directed therapy for triple negative breast cancer. Poster.

Vrtačnik, P., Ostanek, B., Mencej-Bedrač, S., and Marc, J. (2014). The many faces of estrogen signaling. *Biochem. Medica* 24, 329–342.

## W

Wagner, K.-U., Krempler, A., Triplett, A.A., Qi, Y., George, N.M., Zhu, J., and Rui, H. (2004). Impaired alveologenesis and maintenance of secretory mammary epithelial cells in Jak2 conditional knockout mice. *Mol. Cell. Biol.* 24, 5510–5520.

- Wajapeyee, N., and Somasundaram, K. (2003). Cell cycle arrest and apoptosis induction by activator protein 2alpha (AP-2alpha) and the role of p53 and p21WAF1/CIP1 in AP-2alpha-mediated growth inhibition. *J. Biol. Chem.* 278, 52093–52101.
- Wang, H.-C., Yeh, H.-H., Huang, W.-L., Lin, C.-C., Su, W.-P., Chen, H.H.W., Lai, W.-W., and Su, W.-C. (2011). Activation of the signal transducer and activator of transcription 3 pathway up-regulates estrogen receptor-beta expression in lung adenocarcinoma cells. *Mol. Endocrinol. Baltim. Md* 25, 1145–1158.
- Wang, J., Li, J., Fang, R., Xie, S., Wang, L., and Xu, C. (2012). Expression of ER $\alpha$ 36 in gastric cancer samples and their matched normal tissues. *Oncol. Lett.* 3, 172–175.
- Wang, L., Li, H., Yang, S., Ma, W., Liu, M., Guo, S., Zhan, J., Zhang, H., Tsang, S.Y., Zhang, Z. (2016). Cyanidin-3-o-glucoside directly binds to ER $\alpha$ 36 and inhibits EGFR-positive triple-negative breast cancer. *Oncotarget* 7, 68864–68882.
- Wang, Q., Zhang, W., Yang, J., Liu, Y.-L., Yan, Z.-X., Guo, Z.-J., Li, Y.-J., and Bian, X.-W. (2015). High ER $\alpha$ 36 Expression Level and Membrane Location Predict Poor Prognosis in Renal Cell Carcinoma: *Medicine (Baltimore)* 94, e1048.
- Wang, Q., Jiang, J., Ying, G., Xie, X.-Q., Zhang, X., Xu, W., Zhang, X., Song, E., Bu, H., Ping, Y.-F. (2018). Tamoxifen enhances stemness and promotes metastasis of ER $\alpha$ 36+ breast cancer by upregulating ALDH1A1 in cancer cells. *Cell Res.* 28, 336–358.
- Wang, S., Yao, Y., Yao, M., Fu, P., and Wang, W. (2018). Interleukin-22 promotes triple negative breast cancer cells migration and paclitaxel resistance through JAK-STAT3/MAPKs/AKT signaling pathways. *Biochem. Biophys. Res. Commun.* 503, 1605–1609.
- Wang, X., Belguise, K., O'Neill, C.F., Sánchez-Morgan, N., Romagnoli, M., Eddy, S.F., Mineva, N.D., Yu, Z., Min, C., Trinkaus-Randall, V. (2009). RelB NF- $\kappa$ B Represses Estrogen Receptor  $\alpha$  Expression via Induction of the Zinc Finger Protein Blimp1. *Mol. Cell. Biol.* 29, 3832–3844.
- Wang, X., Zheng, N., Dong, J., Wang, X., Liu, L., and Huang, J. (2017). Estrogen receptor- $\alpha$ 36 is involved in icaritin induced growth inhibition of triple-negative breast cancer cells. *J. Steroid Biochem. Mol. Biol.* 171, 318–327.
- Wang, Z., Zhang, X., Shen, P., Loggie, B.W., Chang, Y., and Deuel, T.F. (2005). Identification, cloning, and expression of human estrogen receptor-alpha36, a novel variant of human estrogen receptor-alpha66. *Biochem. Biophys. Res. Commun.* 336, 1023–1027.
- Wang, Z., Zhang, X., Shen, P., Loggie, B.W., Chang, Y., and Deuel, T.F. (2006). A variant of estrogen receptor- $\alpha$ , hER- $\alpha$ 36: Transduction of estrogen- and antiestrogen-dependent membrane-initiated mitogenic signaling. *Proc. Natl. Acad. Sci. U. S. A.* 103, 9063–9068.
- Watson, C.J. (2006). Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. *Breast Cancer Res. BCR* 8, 203.
- Watson, C.J., and Neoh, K. (2008). The Stat family of transcription factors have diverse roles in mammary gland development. *Semin. Cell Dev. Biol.* 19, 401–406.
- Webb, P., Nguyen, P., Valentine, C., Lopez, G.N., Kwok, G.R., McInerney, E., Katzenellenbogen, B.S., Enmark, E., Gustafsson, J.-Å., Nilsson, S. (1999). The Estrogen Receptor Enhances AP-1 Activity by Two Distinct Mechanisms with Different Requirements for Receptor Transactivation Functions. *Mol. Endocrinol.* 13, 1672–1685.
- Welch, D.R. (2016). Tumor heterogeneity 2016. *Cancer Res.* 76, 4–6.

White, R.E. (2002). Estrogen and vascular function. *Vascul. Pharmacol.* 38, 73–80.

Wiseman, B.S., Sternlicht, M.D., Lund, L.R., Alexander, C.M., Mott, J., Bissell, M.J., Soloway, P., Itohara, S., and Werb, Z. (2003). Site-specific inductive and inhibitory activities of MMP-2 and MMP-3 orchestrate mammary gland branching morphogenesis. *J. Cell Biol.* 162, 1123–1133.

## X

Xie, H., Sun, M., Liao, X.-B., Yuan, L.-Q., Sheng, Z.-F., Meng, J.-C., Wang, D., Yu, Z.-Y., Zhang, L.-Y., Zhou, H.-D. (2011). Estrogen receptor  $\alpha 36$  mediates a bone-sparing effect of 17 $\beta$ -estradiol in postmenopausal women. *J. Bone Miner. Res.* 26, 156–168.

Xie, Z.-C., Li, T.-T., Gan, B.-L., Gao, X., Gao, L., Chen, G., and Hu, X.-H. (2018). Investigation of miR-136-5p key target genes and pathways in lung squamous cell cancer based on TCGA database and bioinformatics analysis. *Pathol. Res. Pract.* 214, 644–654.

## Y

Yamashita, H., Nishio, M., Kobayashi, S., Ando, Y., Sugiura, H., Zhang, Z., Hamaguchi, M., Mita, K., Fujii, Y., and Iwase, H. (2005). Phosphorylation of estrogen receptor alpha serine 167 is predictive of response to endocrine therapy and increases postrelapse survival in metastatic breast cancer. *Breast Cancer Res. BCR* 7, R753-764.

Yan, M., Li, X., Tong, D., Han, C., Zhao, R., He, Y., and Jin, X. (2016). miR-136 suppresses tumor invasion and metastasis by targeting RASAL2 in triple-negative breast cancer. *Oncol. Rep.* 36, 65–71.

Yan, Y., Yu, L., Castro, L., and Dixon, D. (2017). ER $\alpha 36$ , a variant of estrogen receptor  $\alpha$ , is predominantly localized in mitochondria of human uterine smooth muscle and leiomyoma cells. *PLoS ONE* 12.

Yang, S.-F., Hou, M.-F., Chen, F.-M., Ou-Yang, F., Wu, Y.-C., Chai, C.-Y., and Yeh, Y.-T. (2016). Prognostic value of protein inhibitor of activated STAT3 in breast cancer patients receiving hormone therapy. *BMC Cancer* 16, 20.

Yang, X., Wang, H., and Jiao, B. (2016). Mammary gland stem cells and their application in breast cancer. *Oncotarget* 8, 10675–10691.

Yaşar, P., Ayaz, G., User, S.D., Güpür, G., and Muyan, M. (2017). Molecular mechanism of estrogen-estrogen receptor signaling. *Reprod. Med. Biol.* 16, 4–20.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M., Hübner, K., and Schöler, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. 14.

Yi, P., Driscoll, M.D., Huang, J., Bhagat, S., Hilf, R., Bambara, R.A., and Muyan, M. (2002). The effects of estrogen-responsive element- and ligand-induced structural changes on the recruitment of cofactors and transcriptional responses by ER alpha and ER beta. *Mol. Endocrinol. Baltim. Md* 16, 674–693.

Yin, L., Zhang, X.-T., Bian, X.-W., Guo, Y.-M., and Wang, Z.-Y. (2014). Disruption of the ER- $\alpha 36$ -EGFR/HER2 positive regulatory loops restores tamoxifen sensitivity in tamoxifen resistance breast cancer cells. *PloS One* 9, e107369.

York, B., and O'Malley, B.W. (2010). Steroid Receptor Coactivator (SRC) Family: Masters of Systems Biology. *J. Biol. Chem.* 285, 38743–38750.

Yoshida, T., Eguchi, H., Nakachi, K., Tanimoto, K., Higashi, Y., Suemasu, K., Iino, Y., Morishita, Y., and Hayashi, S. (2000). Distinct mechanisms of loss of estrogen receptor alpha gene expression in human breast cancer: methylation of the gene and alteration of trans-acting factors. *Carcinogenesis* 21, 2193–2201.

Yoshimizu, T., Sugiyama, N., Felice, M.D., Yeom, Y.I., Ohbo, K., Masuko, K., Obinata, M., Abe, K., Schöler, H.R., and Matsui, Y. (1999). Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev. Growth Differ.* 41, 675–684.

Yu, J., Angelin-Duclos, C., Greenwood, J., Liao, J., and Calame, K. (2000). Transcriptional Repression by Blimp-1 (PRDI-BF1) Involves Recruitment of Histone Deacetylase. *Mol. Cell. Biol.* 20, 2592–2603.

Yu, Q.C., Verheyen, E.M., and Zeng, Y.A. (2016). Mammary Development and Breast Cancer: A Wnt Perspective. *Cancers* 8.

Yuan, F., and Wang, W. (2015). MicroRNA-802 suppresses breast cancer proliferation through downregulation of FoxM1. *Mol. Med. Rep.* 12, 4647–4651.

## Z

Zeng, X., Liu, L., Lü, L., and Zou, Q. (2018). Prediction of potential disease-associated microRNAs using structural perturbation method. *Bioinforma. Oxf. Engl.* 34, 2425–2432.

Zhang, H.-F., and Lai, R. (2014). STAT3 in Cancer—Friend or Foe? *Cancers* 6, 1408–1440.

Zhang, C., Luo, X., Liu, L., Guo, S., Zhao, W., Mu, A., Liu, Z., Wang, N., Zhou, H., and Zhang, T. (2013). Myocardin-related transcription factor A is up-regulated by 17 $\beta$ -estradiol and promotes migration of MCF-7 breast cancer cells via transactivation of MYL9 and CYR61. *Acta Biochim. Biophys. Sin.* 45, 921–927.

Zhang, J., Brewer, S., Huang, J., and Williams, T. (2003). Overexpression of transcription factor AP-2alpha suppresses mammary gland growth and morphogenesis. *Dev. Biol.* 256, 127–145.

Zhang, J., Li, G., Li, Z., Yu, X., Zheng, Y., Jin, K., Wang, H., Gong, Y., Sun, X., Teng, X., et al. (2012). Estrogen-independent effects of ER- $\alpha$ 36 in ER-negative breast cancer. *Steroids* 77, 666–673.

Zhang, J., Zhou, C., Jiang, H., Liang, L., Shi, W., Zhang, Q., Sun, P., Xiang, R., Wang, Y., and Yang, S. (2017). ZEB1 induces ER- $\alpha$  promoter hypermethylation and confers antiestrogen resistance in breast cancer. *Cell Death Dis.* 8, e2732.

Zhang, S., Qin, C., and Safe, S.H. (2003). Flavonoids as aryl hydrocarbon receptor agonists/antagonists: effects of structure and cell context. *Environ. Health Perspect.* 111, 1877–1882.

Zhang, S., Qiu, C., Wang, L., Liu, Q., and Du, J. (2014). The elevated level of ER $\alpha$ 36 is correlated with nodal metastasis and poor prognosis in lung adenocarcinoma. *Steroids* 87, 39–45.

Zhang, X., Tang, N., Hadden, T.J., and Rishi, A.K. (2011a). Akt, FoxO and regulation of apoptosis. *Biochim. Biophys. Acta* 1813, 1978–1986.

Zhang, X., Kang, L., Ding, L., Vranic, S., Gatalica, Z., and Wang, Z.-Y. (2011b). A Positive Feedback Loop of ER- $\alpha$ 36/EGFR Promotes Malignant Growth of ER-negative Breast Cancer Cells. *Oncogene* 30, 770–780.

Zhang, X., Li, Y., Wang, D., and Wei, X. (2017). miR-22 suppresses tumorigenesis and improves radiosensitivity of breast cancer cells by targeting Sirt1. *Biol. Res.* 50, 27.

- Zhang, X.-T., Ding, L., Kang, L.-G., and Wang, Z.-Y. (2012). Involvement of ER- $\alpha$ 36, Src, EGFR and STAT5 in the biphasic estrogen signaling of ER-negative breast cancer cells. *Oncol. Rep.* 27, 2057–2065.
- Zhao, C., Yasui, K., Lee, C.J., Kurioka, H., Hosokawa, Y., Oka, T., and Inazawa, J. (2003). Elevated expression levels of NCOA3, TOP1, and TFAP2C in breast tumors as predictors of poor prognosis. *Cancer* 98, 18–23.
- Zhao, C., Matthews, J., Tujague, M., Wan, J., Ström, A., Toresson, G., Lam, E.W.-F., Cheng, G., Gustafsson, J.-A., and Dahlman-Wright, K. (2007). Estrogen receptor beta2 negatively regulates the transactivation of estrogen receptor alpha in human breast cancer cells. *Cancer Res.* 67, 3955–3962.
- Zhao, C., Dahlman-Wright, K., and Gustafsson, J.-A. (2008). Estrogen receptor beta: an overview and update. *Nucl. Recept. Signal.* 6, e003.
- Zheng, Y., Zhang, J., Xu, Z., Sheng, J., Zhang, X., Wang, H., Teng, X., Liu, X., Cao, J., and Teng, L. (2010). Quantitative profiles of the mRNAs of ER- $\alpha$  and its novel variant ER- $\alpha$ 36 in breast cancers and matched normal tissues\*. *J. Zhejiang Univ. Sci. B* 11, 144–150.
- Zhu, H., Huang, Y., Su, H., Ma, Y., Tao, Y., Liao, D.J., Liu, Y., and Feng, Z. (2018). Identification of a novel human estrogen receptor- $\alpha$  splice variant able to enhance malignant biological behaviors of breast cancer cells. *Oncol. Lett.* 15, 5339–5344.
- Zhu, L., Zou, J., Zhao, Y., Jiang, X., Wang, Y., Wang, X., and Chen, B. (2018). ER- $\alpha$ 36 mediates cisplatin resistance in breast cancer cells through EGFR/HER-2/ERK signaling pathway. *J. Exp. Clin. Cancer Res.* CR 37, 123.
- Zhu, Y., Shao, S., Pan, H., Cheng, Z., and Rui, X. (2018). MicroRNA-136 inhibits prostate cancer cell proliferation and invasion by directly targeting mitogen-activated protein kinase kinase 4. *Mol. Med. Rep.* 17, 4803–4810.
- Zimmerman, S. (2019). Get to Know Your Sex Hormones.
- Zou, Y., Ding, L., Coleman, M., and Wang, Z. (2009). Estrogen receptor-alpha (ER- $\alpha$ ) suppresses expression of its variant ER- $\alpha$ 36. *FEBS Lett.* 583, 1368–1374.
- Zwart, W., Griekspoor, A., Berno, V., Lakeman, K., Jalink, K., Mancini, M., Neeffjes, J., and Michalides, R. (2007). PKA-induced resistance to tamoxifen is associated with an altered orientation of ERalpha towards co-activator SRC-1. *EMBO J.* 26, 3534–3544.



## Résumé :

Les récepteurs nucléaires aux œstrogènes, représentés par les formes canoniques ER $\alpha$ 66 et ER $\beta$ 1, sont les principaux médiateurs des effets des œstrogènes dans l'organisme. Ces hormones qui régulent le développement physiologique de la glande mammaire participent également à l'initiation et à la progression du cancer sein. L'expression d'ER $\alpha$ 66 est d'ailleurs utilisée dans la classification moléculaire des tumeurs mammaires afin d'orienter la stratégie thérapeutique vers l'hormonothérapie. Cependant, dans environ 30% des cas, des résistances thérapeutiques sont observées et cela souligne la nécessité d'identifier de nouveaux biomarqueurs. Depuis son clonage en 2005, le variant des récepteurs alpha aux œstrogènes, ER $\alpha$ 36, a été principalement décrit dans la littérature pour son rôle majeur dans la progression des tumeurs mammaires et dans l'acquisition de résistances aux anti-œstrogènes comme le Tamoxifène. Si une forte expression d'ER $\alpha$ 36 dans les cellules cancéreuses mammaires apparaît nettement comme un facteur de mauvais pronostic, peu de données sont disponibles concernant son rôle dans le développement de la glande mammaire saine. C'est pourquoi le premier objectif de ce travail était de déterminer le rôle d'ER $\alpha$ 36 dans le développement physiologique de cette glande. Grâce à une approche pluridisciplinaire, incluant des études *in vivo* sur un modèle de souris transgéniques MMTV-ER $\alpha$ 36 et des études *in vitro* et *in silico* sur des cellules épithéliales mammaires saines immortalisées (MCF-10A), nous avons montré que l'expression d'ER $\alpha$ 36 est suffisante pour perturber le phénotype des cellules épithéliales mammaires et conduire à l'apparition d'altérations structurales des canaux mammaires à l'âge adulte. De plus, nous avons montré qu'une exposition aux alkylphénols, qui sont des perturbateurs endocriniens œstrogène-mimétiques, stimule l'expression endogène de ce variant dans les cellules MCF-10A et augmente leurs capacités migratoires sans pour autant amplifier les effets d'ER $\alpha$ 36 sur l'histologie des canaux mammaires. Dans un second temps, afin de mieux comprendre l'implication d'ER $\alpha$ 36 au moment de l'initiation et/ou de la progression tumorale, nous avons étudié les modalités de régulation classique et épigénétique de l'expression de ce variant dans les cellules cancéreuses mammaires. Les résultats obtenus indiquent que l'expression d'ER $\alpha$ 36 est positivement corrélée au statut de méthylation de sa région promotrice et que l'ARNm codant ce variant est la cible du micro-ARN, hsa-miR-136-5p. Enfin, le dernier objectif de ce travail était de développer une approche bio-informatique afin d'identifier *in silico* de nouveaux partenaires d'ER $\alpha$ 36. Finalement, l'ensemble de ce travail s'inscrit dans une démarche de raffinement de la classification moléculaire actuelle des tumeurs mammaires en y ajoutant une composante associée à l'expression d'ER $\alpha$ 36.

**Mots-clés :** Cancer du sein, ER $\alpha$ 36, physiopathologie, épigénétique, micro-ARN, bio-informatique.

## Abstract :

The estrogen nuclear receptors, represented by the canonical forms ER $\alpha$ 66 and ER $\beta$ 1, are the main mediators of the estrogenic effects in mammals. These hormones, which regulate the physiological development of the mammary gland, participate in the initiation and progression of breast cancer. In fact, ER $\alpha$ 66 expression is a key molecular classifier of breast tumors used in order to guide the therapeutic strategies toward hormone therapy. However, in 30% of cases, therapeutic failures are observed, which highlights the importance of identifying new biomarkers. The estrogen receptor variant, ER $\alpha$ 36, has been cloned in 2005 and mainly described in the literature to be involved in the progression of mammary tumors and in the acquired resistance to anti-estrogen drugs, such as Tamoxifen. Even if a high expression of ER $\alpha$ 36 in breast cancer cells appears to be associated with a poor prognosis, few data are available concerning its role in the normal development of the mammary gland. Therefore, the aim of this work was to determine the role of ER $\alpha$ 36 in the physiological development of the mammary gland. Thanks to a multidisciplinary approach, that combines *in vivo* studies on MMTV-ER $\alpha$ 36 transgenic mice, and *in vitro* and *in silico* studies on immortalized normal epithelial mammary cells (MCF-10A), we showed that ER $\alpha$ 36 expression is sufficient to disturb the mammary epithelial cells phenotype, leading to the emergence of structural alterations of mammary ducts at adulthood. Moreover, we showed that exposure to the estrogen mimicking compounds alkylphenols stimulates the endogenous expression of this variant in MCF-10A cells, and increases their migratory ability. Then, in order to get a better understanding of ER $\alpha$ 36 contribution to tumor initiation and/or progression, we studied classical and epigenetic regulation of this variant expression in breast cancer cells. Our results show that ER $\alpha$ 36 expression is positively correlated with the methylation status of its promoter region, and that the ER $\alpha$ 36 mRNA is the target of the microRNA, has-miR-136-5p. Finally, the last aim of this work was to develop a bioinformatic approach in order to study the ER $\alpha$ 36 partners. To summarize, all of this work falls within a need of the current breast tumor molecular classification refinement by adding a component related with ER $\alpha$ 36 expression.

**Keywords :** Breast cancer, ER $\alpha$ 36, pathophysiology, epigenetics, micro-RNA, bioinformatics.