Recherche de biomarqueurs en cancérologie : Recherche préclinique sur le rôle de la protéine DDB2 dans les cancers du pancréas, recherche translationnelle et clinique sur le déploiement des analyses de l'ADN tumoral circulant

Alexandre Harlé

▶ **To cite this version:**

Alexandre Harlé. Recherche de biomarqueurs en cancérologie : Recherche préclinique sur le rôle de la protéine DDB2 dans les cancers du pancréas, recherche translationnelle et clinique sur le déploiement des analyses de l'ADN tumoral circulant. Cancer. Université de Lorraine, 2021. tel-03111110

HAL Id: tel-03111110
https://hal.univ-lorraine.fr/tel-03111110

Submitted on 15 Jan 2021

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.

Ecole Doctorale BioSE (Biologie, Santé, Environnement)

Mémoire

Présenté pour la candidature au diplôme

D'HABILITATION À DIRIGER DES RECHERCHES

par

Alexandre HARLÉ

Recherche de biomarqueurs en cancérologie : Recherche préclinique sur le rôle de la protéine DDB2 dans les cancers du pancréas, recherche translationnelle et clinique sur le déploiement des analyses de l'ADN tumoral circulant

(Section CNU 82)

Parrain scientifique : Monsieur Jean-Louis MERLIN

PU-PH, Université de Lorraine, Faculté de Pharmacie, UMR 7039 CRAN, Institut de Cancérologie de Lorraine

---

UMR 7039 CRAN CNRS, Université de Lorraine, FST 1er Cycle Bât Henri Poincaré Campus Sciences BP 70239 54506 Vandoeuvre-lès-Nancy

Institut de Cancérologie de Lorraine, 6 Avenue de Bourgogne, 54519 Vandoeuvre-lès-Nancy

SOMMAIRE

1. Curriculum vitae 5

2. Activités d'enseignement 8

3. Activités de recherche 11

3.1. Travaux de recherche dans le cadre de mon diplôme de Doctorat d'Université 12

3.2. Activités de recherche depuis l'obtention du Doctorat d'Université 14

3.3. Activités de recherche au cours de mon année de mobilité internationale 15

3.4. Activités de recherche en cours 17

3.4.1. Rôle de la protéine DDB2 dans la cancérogenèse des cancers du pancréas exocrines 17

3.4.2. Activités de recherche translationnelle 22

4. Projet scientifique 2021-2025 30

4.1. Implication de la protéine DDB2 dans les cancers du pancréas 30

4.1.1. Exploration des voies de signalisation impliquées, voie du TGFβ et protéine DDB2.....30

4.1.2. Activation de DDB2 comme cible thérapeutique dans les cancers du pancréas.....31

4.1.3. Évaluation de DDB2 comme marqueur pronostique et diagnostique des cancers du pancréas 31

4.2. Hétérogénéité tumorale, évolution clonale et marqueurs prédictifs de réponse aux thérapeutiques chez les patients atteints d'un cancer du pancréas 32

4.2.1. Exploration de l'hétérogénéité tumorale et de l'évolution clonale par l'ADN tumoral circulant 32

4.2.2. Prédiction du retentissement antigénique des altérations somatiques et réponse à l'immunothérapie.....33

4.2.3. Détermination de la maladie résiduelle et prédiction précoce des rechutes chez les patients atteints de cancer 34

4.2.4. Evolution métastatique du cancer du pancréas 34

5. Financements 35

6. Liste des encadrements d'étudiants 36

7. Liste des publications et communications 42

8. Sélection de 5 publications 51

9. Références bibliographiques 111

Remerciements

Aux rapporteurs

À Madame le Professeur Christine Clavel,
Je vous remercie d'avoir accepté de rapporter mon mémoire, pour le temps que vous y avez consacré et d'apporter votre expertise en tant que biologiste. Je souhaite vous exprimer ici ma plus grande reconnaissance.

À Monsieur le Professeur Jean-Baptiste Bachet,
Merci d'avoir accepté de rapporter ce mémoire et d'apporter le point de vue du clinicien, notamment en tant qu'expert dans le traitement du cancer du pancréas. Veuillez-trouver ici ma plus sincère gratitude.

À Monsieur le Professeur Jean-Christophe Sabourin,
Après avoir été jury pour ma thèse d'Université il y a quelques années, je suis heureux que vous ayez accepté d'être rapporteur pour mon mémoire d'HDR. Je vous suis reconnaissant d'apporter votre expertise en tant que pathologiste dans ce rapport et nos discussions. Veuillez-trouver ici mes plus sincères remerciements.

Aux examinateurs de ce mémoire

À Madame le Professeur Karen Leroy,
Merci d'avoir accepté de faire partie des examinateurs lors de cette soutenance. Merci pour votre bienveillance et nos discussions scientifiques depuis toutes ces années. Veuillez-trouver ici l'expression de toute ma gratitude.

À Madame le Docteur Isabelle Soubeyran,
Je suis très heureux de vous compter parmi les examinateurs de ce travail. Merci pour votre soutien et pour nos discussions depuis toutes ces années. Veuillez-trouver ici tous mes remerciements.

À Monsieur le Professeur Thierry Conroy,
Je vous remercie chaleureusement d'avoir accepté de faire partie des examinateurs de ce mémoire, particulièrement en votre qualité d'expert reconnu mondialement du traitement des cancers du pancréas. Je sais compter sur votre soutien depuis toutes ces années et je vous en suis très reconnaissant. Je vous prie de trouver ici l'expression de toute ma gratitude.

À Monsieur le Professeur Pierre Laurent-Puig,
Je vous remercie d'avoir accepté de faire partie des examinateurs. Vous aviez également été membre de mon jury lors de ma thèse d'Université, et je vous remercie d'être à nouveau présent pour moi. Je vous prier de croire en mes sincères remerciements.

À Monsieur le Professeur Jean-Louis Merlin,
C'est du fond du cœur que je vous remercie pour toutes ces années de travail et de collaborations. J'étais bien loin de penser lors de mon premier stage d'internat en Novembre 2006 dans votre service que nous travaillerions toujours ensemble 14 ans plus tard. Vous avez toujours été présent pour moi lors de toutes ces longues étapes, dans les doutes et les réussites. Je vous suis très reconnaissant pour toutes ces longues années, pour votre bienveillance et votre soutien constant et indéfectible. Veuillez-trouver ici l'expression de ma plus sincère gratitude.

À Monsieur le Professeur Raphaäel Duval, Doyen de la Faculté de Pharmacie de Nancy,
Merci pour votre soutien et votre sincérité depuis toutes ces années, pour votre aide dans le gravissement de ces marches tout au long de ces années. Veuillez-trouver ici l'expression de toute ma reconnaissance.

À Monsieur le Docteur Julien Perrin, vice-Doyen de la Faculté de Pharmacie de Nancy,
Merci Julien pour toutes ces années passées ensemble comme collègues et je suis heureux de t'avoir aujourd'hui comme vice-Doyen de notre Faculté. Merci pour nos conversations et ton amitié.

À Monsieur le Professeur Patrick Menu, Directeur de l'École Doctorale BioSe,
Je vous remercie pour votre soutien et votre bienveillance, votre patience et votre accompagnement dans cette nouvelle étape. Veuillez-trouver ici l'expression de toute ma gratitude.

À Monsieur le Professeur Didier Wolf, Directeur de l'UMR 7039 CRAN,
Je vous remercie de votre soutien depuis toutes ces années, et je me réjouis de pouvoir continuer mes recherches dans notre laboratoire. Veuillez-trouver ici tous mes remerciements.

À tous mes collègues et anciens collègues de travail, à l'hôpital, à l'Université,
Je vous remercie sincèrement pour ces années passées ensemble et pour nos réussites collectives.

À mes parents, ma famille et mes amis,
Merci pour votre soutien depuis toutes ces années, vos preuves d'amour et d'amitié.

1. Curriculum vitae

Alexandre Harlé, né le 16/07/1980 à Paris 17ème

DIPLÔMES

2014 Diplôme de Docteur de l'Université de Lorraine, Mention Sciences de la vie et de la santé, Ecole Doctorale BioSe, Université de Lorraine, Nancy – « Marqueurs de réponse aux thérapies ciblées et personnalisation thérapeutique dans les cancers colorectaux métastatiques » sous la Direction du Pr. Jean-Louis Merlin.

2011 Diplôme d'études spécialisées de Pharmacie Industrielle et Biomédicale, Université Henri Poincaré, Nancy – « Mise au point du dosage de l'Imatinib, du Norimatinib, du Dasatinib et du Nilotinib par chromatographie liquide haute performance et détecteur à barrette de diodes » sous la Direction du Pr. MJ Royer-Morrot.
Diplôme d'Etat de Docteur en Pharmacie, Mention Très Bien, Université Henri Poincaré, Nancy.
M2 Ingénierie de la Santé, Science du Médicament, Mention Très Bien, Université Henri Poincaré, Nancy – « Mise au point de la recherche des mutations PIK3CA dans les cancers colorectaux » sous la Direction du Pr. Jean-Louis Merlin.

2009 DU Pharmacie Clinique Oncologique – Université Claude Bernard Lyon 1
DU Antibiothérapie en milieu hospitalier – Université Henri Poincaré Nancy

2005 Concours national de l'internat – Pharmacie Industrielle et Biologie Médicale, zone Nord-est.

2005 Sixième année de Pharmacie, Université Joseph Fourier, Grenoble. UV Biologie médicale et UV Communication, Université Joseph Fourier, Grenoble. Certificat MSBM : Conception et structure des molécules d'intérêts thérapeutiques, Université Joseph Fourier, Grenoble. Stage d'initiation à la recherche : Laboratoire de Chimie organique, Université Joseph Fourier, Grenoble.

2000 Concours de première année de Pharmacie, Université Joseph Fourier, Grenoble.

1998 Baccalauréat série S option Physique-Chimie, Mention Assez-Bien, Lycée Berthollet, Annecy.

ACTIVITÉS PROFESSIONNELLES

Depuis Septembre 2015 Maitre de Conférences des Universités – Praticien Hospitalier, Université de Lorraine, CNRS UMR7039 CRAN, Institut de Cancérologie de Lorraine. 1ere classe depuis le 1/9/2019
Juillet 2018 à juillet 2019 Mobilité internationale (Visiting Scientist at The Francis Crick Institute, London, United-Kingdom, Research fellow at The Royal Marsden Hospital, London, United-Kingdom)
2011 à 2015 Assistant Hospitalo-Universitaire, Université de Lorraine, CNRS UMR7039 CRAN, Institut de Cancérologie de Lorraine.
2006 à 2011 Interne en Pharmacie PIBM, CHU de Nancy, Institut de Cancérologie de Lorraine

ACTIVITÉS D'ENSEIGNEMENT

Environ 100 HTD par an dans le domaine de la biologie cellulaire oncologique, de la théranostique et le diagnostic des cancers pour les étudiants en Faculté de Pharmacie, en LMD, en DU ou en formation continue.
Coordination de l'UEL DMC (Diagnostic Moléculaire en Cancérologie) en DFA-SP1, Faculté de Pharmacie depuis 2017.
Coordination de l'UEL MCTC (Médicament, Ciblage et thérapeutique de précision en cancérologie), en première année de Master parcours IS depuis 2018.

Coordination du DU « Pharmacie Clinique Oncologique à l'Officine » à la Faculté de Pharmacie de Nancy, ouverture prévue pour l'année Universitaire 2021-2022.

ACTIVITÉS D'INTERET COLLECTIF

Groupe de Travail validation de méthode NGS et base de variants somatiques : Membre du groupe de travail de l'Institut National du Cancer (INCa) pour la validation des méthodes de Next Generation Sequencing et la mise en place d'une base de données Nationale de variants somatiques rares.

Groupe de Travail validation des pipelines bioinformatiques pour l'analyse NGS du COFRAC : Membre du groupe de travail du Comité Français d'Accréditation.

Association Gen&Tiss (Assurance Qualité en génétique somatique) : Membre du comité de pilotage de Gen&Tiss pour la mise en place de programmes d'évaluation externe de la qualité en génétique somatique.

GFCO : Membre du bureau du Groupe Francophone de Cytogénomique Oncologique (GFCO).

DGOS : Expert génétique tumorale (commission n˚9) pour la sélection des Laboratoires de Biologie Médicale de Référence (LBMR) pour la Direction Générale de l'Offre des Soins.

UCGI : Responsable du bureau recherche translationnelle pour le groupe Unicancer des cancers gastrointestinaux.

Membre de la **Commission Recherche de la Faculté de Pharmacie de Nancy**.

Membre de l'American Society of Clinical Oncology (**ASCO**)
Membre de l'American Association for Cancer Research (**AACR**)
Membre de l'Association Française d'Assurance Qualité en Anatomie Cytologie Pathologique (**AFAQAP**)
Membre du Groupe Francophone de Cytogénomique Oncologique (**GFCO**)
Membre de l'European Liquid Biopsy Society (**ELBS**)

ACTIVITÉS D'EXPERTISE

Activité de reviewing pour des périodiques internationaux et nationaux : Archives of Medical Research, BMC Medical Notes, British Journal of Cancer, Bulletin du Cancer, Cancer Investigation, Expert Review of Anticancer Therapy, Internal Medicine: Open Access, Journal of Carcinogenesis & Mutagenesis, Journal of Clinical Pathology, Journal of Clinical Laboratory Analysis, Nature, PlosOne, The International Journal of Biological Markers, The Lancet Oncology, Theranostics, Tropical Medicine & Surgery, World Journal Of Surgical Oncology.

Membre du comité de rédaction de la Revue Scientifique d'Unicancer : Innovation & Thérapeutique en Oncologie

Evaluations scientifiques pour le Cancéropôle de l'Est, le GEFLUC Flandres-Artois et Normandie et Cancéropôle Nord-Ouest, l'AIFA (Agence Italienne de Pharmacie).

Expertise et séminaires pour l'industrie pharmaceutique.

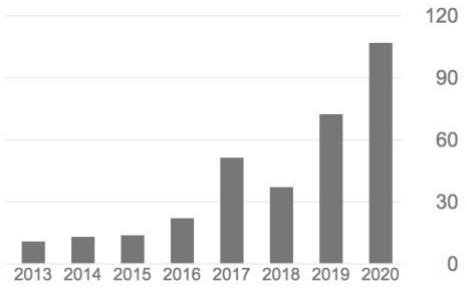
PUBLICATIONS ET COMMUNICATIONS

42 publications dont 35 dans des journaux internationaux à comité de lecture et 7 dans des journaux nationaux à comité de lecture.
14 publications en premier auteur, 5 en deuxième auteur, 2 en avant dernier auteur, 16 en dernier auteur et 5 en rang non-utile (*cf.* liste des publications)
Plus de 80 communications (51 communications orales dont 30 sur invitation, et 33 communications sous forme de poster).

Points SIGAPS : 529

Indices bibliométriques au 03/12/2020

	Toutes	Depuis 2015
Citations	335	305
indice h	10	10
indice i10	11	11



2. Activités d'enseignement

La partie Universitaire de mon travail consiste en l'enseignement de la biologie cellulaire oncologique et de ses applications en cancérologie. Une grande importance est donnée au diagnostic moléculaire et aux thérapies ciblées, dans un but de développement de la personnalisation thérapeutique des traitements anticancéreux. L'enseignement est ainsi adapté aux différentes disciplines pharmaceutiques : biologie, pharmacie hospitalière, pharmacie industrielle, officine. Je participe aux enseignements de la Faculté de Pharmacie de Nancy (Université de Lorraine), mais également aux enseignements de la Faculté des Sciences (Licence et Master), à certains Diplômes Universitaires ainsi qu'à des actions de formation continue.

Participation à la réalisation des sujets pour le Certificat de Synthèse Pharmaceutique (CSP) depuis 4 ans pour la partie biologie cellulaire et cancérologie.

Depuis l'année Universitaire 2017/2018, je suis responsable de l'UE libre DMC (Diagnostic Moléculaire en Cancérologie) et responsable de l'UE de M1 MCTC (Médicament, Ciblage et Thérapeutique en Cancérologie) depuis la rentrée 2018.

Depuis la rentrée 2020, je suis membre du jury de délibération de DFA-SP1.

Nombre d'heures ETD pour l'année universitaire 2017-2018 : 95,75
Score SIAPS (Période 2011 à 2019) : 685 (grille détaillée ci-après).
Pas d'activité d'enseignement sur l'année Universitaire 2019-2020 compte tenu de mon départ en mobilité au Crick Institute de Londres (équipe du Pr. Charles Swanton). Lauréat d'une bourse à la mobilité de la Fondation de France.
Nombre d'heures ETD pour l'année universitaire 2019-2020 : 76,5
Les enseignements auxquels je participe de manière régulière sont détaillés ci-dessous :

Préparation au Diplôme d'Etat de Docteur en Pharmacie, Faculté de Pharmacie de Nancy, Université de Lorraine
DFA SP1 : UEL Evaluation du Principe Actif (3h de TD et 6h de TP par an)
DFG SP2 : UEL Instrumentation et Techniques Médicales pour le Diagnostic et la Thérapie (1h de TP par an)
DFG SP3 : UEL Pharmacogénétique et Pharmacogénomique (2h de CM par an)
DFA SP1 : UEB Cancérologie (3h30 de CM par an)
DFA SP1 : UEB Connaissances Intégrées Pharmaceutiques (10h de TD par an)
DFA SP1 : UEL Diagnostic Moléculaire en Cancérologie (5h30 de CM, 6h de TD et 1h30 de TP par an)
DFA SP1 : UEF Développement clinique (3h30 de CM par an)
DFA SP2 : UEB Assurance Qualité à l'Officine (4,5h de TD par an)
DFA SP2 : UEH Fonctions hospitalières (2h de TD par an)
DFA SP3 : UEO Cancérologie à l'officine (2h de CM par an)
DES de Pharmacie Hospitalière (3h de CM par an)
DES de Biologie Médicale (1h30 de CM par an)

Participation au système LMD, Université de Lorraine
Licence Pro : Génie Moléculaire et Cellulaire (6h de CM par an)
M1 BioIngénierie et Médicaments (6h de CM par an)
M1 Biologie Cellulaire et Physiologie (2h de CM par an)
M2 BioSciences et Ingénierie de la Santé (10h30 de CM par an)

Participation à des Diplômes Universitaires
DU Pharmacie Clinique Oncologique – Faculté de Pharmacie – Université de Lorraine (6h30 de CM par an)

Participation à la formation continue
DU Pharmacie Clinique Oncologique – Faculté de Pharmacie – Université de Lorraine (6h30 de CM par an)

Projets pédagogiques

En accord avec le Doyen de la Faculté de Pharmacie, dans une volonté de développement de la formation continue, je proposerai en 2021, la création d'un Diplôme Universitaire de cancérologie destiné aux pharmaciens d'officine pour une ouverture à la rentrée universitaire 2021.

Les objectifs de ce DU est de maintenir les compétences de nos collègues officinaux de plus en plus amenés à prendre en charge à l'officine les patients atteints de cancer. Ce DU reprendra ainsi les bases de l'oncogenèse, des rappels concrets sur les grandes pathologies, et bien entendu des mises à jour pratique sur les nouveaux médicaments disponibles à l'officine, notamment les anticancéreux *per os*. Ce nouveau DU sera totalement d'actualité, compte tenu du 21ème avenant à la convention pharmaceutique qui ouvre un nouvel entretien pharmaceutique pour les patients sous anticancéreux oraux. Le suivi de ces médicaments en officine représente donc un enjeu important notamment en matière de formation.

Grille des points SIAPS

Nom	Harlé
Prénom	Alexandre
Service, Faculté	Faculté de Pharmacie, Université de Lorraine
Date de naissance	16/07/1980
Sous Section CNU	82
Candidature	MCUPH

1) Charges d'enseignement en présentiel.														
		2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019	Coefficients	Total
Nombre d'heures d'enseignement PACES	Cours				0	0	0	0	0	0	0	0	1.5	0
	ED				0	0	0	0	0	0	0	0	1	0
Nombre d'heures d'enseignement dans le Diplôme d'Etat de Docteur en Pharmacie (DFGSP hors PACES, DFASP, 6ème année).	Cours				2	4	11	8	14.5	11	17.5	22	1.5	135
	ED				0	0	42	34	26.5	36	29	25.5	1	193
	TP				0	0	16.5	16.5	13.5	7.5	6	9.5	1	69.5
Nombre d'heures d'enseignement niveau Licence (DEUST, Lic. Pro., IFSI, Para Med...)	Cours				2	4	4	4	4	4	6	6	1.5	51
	ED				0	0	0	0	0	0	0	0	1	0
	TP				0	0	0	0	0	0	0	0	1	0
Nombre d'heures d'enseignement MASTER	Cours				10	6	9	11	8	9.5	12.5	12.5	1.5	117.75
	ED				0	0	0	0	0	0	0	0	1	0
	TP				0	0	0	0	0	0	0	0	1	0
Nombre d'heures d'enseignement 3ème cycle (DES, Ecole Doctorale...)	Cours				6.5	6.5	5.5	5.5	4.5	2.5	0	0	1.5	46.5
	ED				0	0	0	0	0	0	0	0	1	0
	TP				0	0	0	0	0	0	0	0	1	0
Nombre d'heures d'enseignement Formation continue	Cours				0	0	0	4.5	0	4	6.5	6.5	1.5	32.25
	ED				0	0	0	0	0	0	0	0	1	0
	TP				0	0	0	0	0	0	0	0	1	0
Nombre d'heures d'enseignement Autres formations médicales (Chir. Dent, Med...)	Cours				0	0	0	0	0	0	0	0	1.5	0
	ED				0	0	0	0	0	0	0	0	1	0
	TP				0	0	0	0	0	0	0	0	1	0
Nombre de thèses d'exercice, DES (nombre en tant que directeur)					0	0	0	0	1	3	7	5	1	16
SCORES ANNUELS		0	0	0	30.75	30.75	102.75	100	87.5	93	105.75	110.5		
													24.5	661

2) Formation pédagogique														
		2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019	Coefficients	
Formation DU/DIU pédagogique (min. 80 heures) (1 = oui)					0	0	0	0	0	0	0	0	15	0
Nombre de 1/2 journées de formation en pédagogie suivies					0	0	0	0	0	0	0	0	2	0
Titulaire de la certification SIDES (Oui = 1)					0	0	0	0	0	0	0	0	3	0
SCORES ANNUELS		0	0	0	0	0	0	0	0	0	0	0		
													20	0

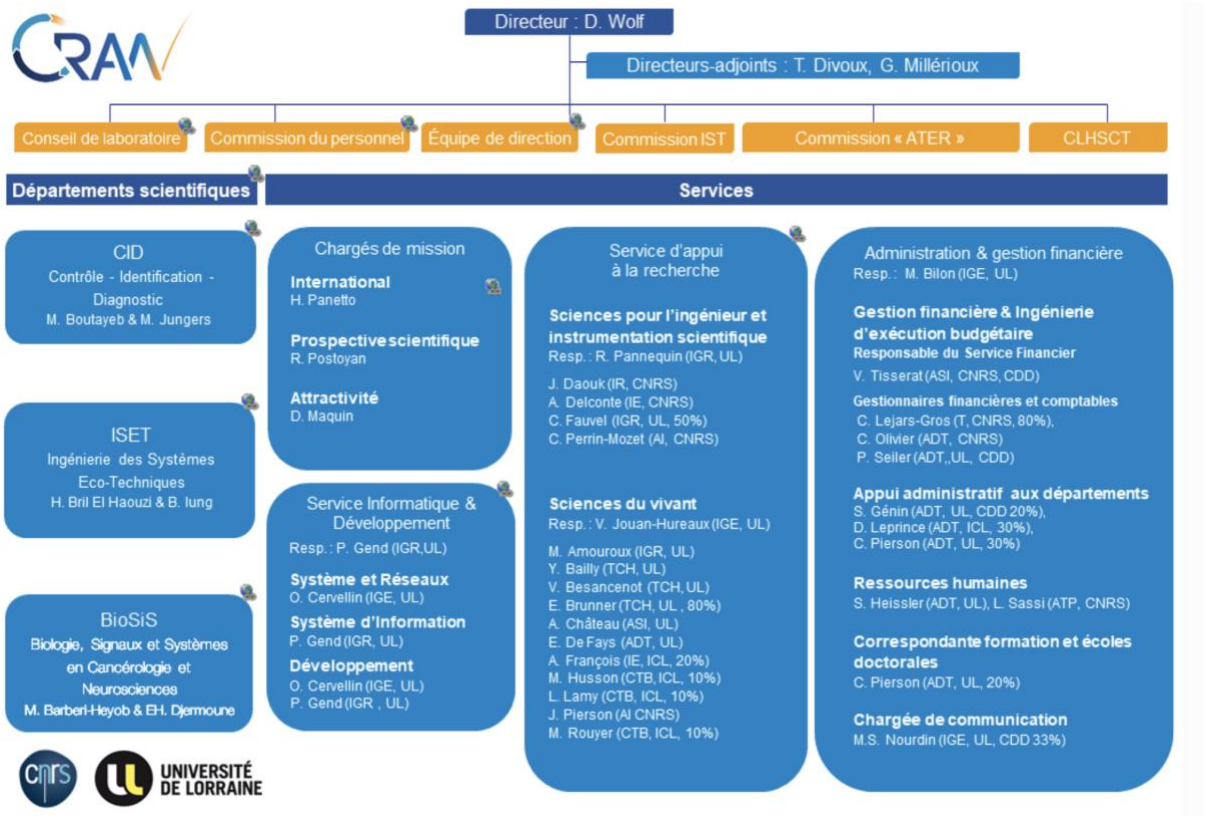
3) Activités (& responsabilités) pédagogiques et services rendus														
		2008/2009	2009/2010	2010/2011	2011/2012	2012/2013	2013/2014	2014/2015	2015/2016	2016/2017	2017/2018	2018/2019	Coefficients	
Assesseur pédagogique/directeur des études (1/année de mandat)					0	0	0	0	0	0	0	0	10	0
Responsabilité de diplômes : (MASTER 1Master 2, DEUST, LP...)					0	0	0	0	0	0	0	0	7	0
Responsabilité de diplômes formation continue : (DU, DIU...)					0	0	0	0	0	0	0	0	3	0
Responsable d'année ou de filière (1 par année et filières)					0	0	0	0	0	0	0	0	5	0
Coordonateur DES (1/an : pour une coordination locale , 2/an : régionale, 3/an : nationale)					0	0	0	0	0	0	0	0	7	0
Responsable d'UE (1/ECTS)					0	0	0	0	0	0	3	6	2	18
Production pédagogique numérique TICE (1h de cours = 3pts)					0	0	0	0	0	0	0	0	1	0
Mise à jour de ressources numériques (1h = 1pts)					0	0	0	0	0	0	0	0	0.5	0
Nombre d'heures d'innovations pédagogiques évaluées					0	0	0	0	0	0	0	0	5	0
Publication/présentation colloque en pédagogie ou didactique (1 par publication)					0	0	0	0	0	0	0	0	1	0
Participation à la rédaction d'un ouvrage pédagogique de référence (1 par ouvrage)					0	0	0	0	0	0	0	0	1	0
Responsabilité d'une commission pédagogique (1/commission/an)					0	0	0	0	0	0	0	0	2.5	0
Participation au CNCI en tant qu'expert (1/année de participation)					0	0	0	0	0	0	0	0	2.5	0
Encadrement tutorat 5AHU (1/année d'encadrement)					0	0	2	2	2	0	0	0	1	6
Membre CNU (1 / année de participation)					0	0	0	0	0	0	0	0	2	0
Membre commission pédagogique (1/commission /année de participation)					0	0	0	0	0	0	0	0	1	0
Responsable TP (1/TP/ année de responsabilité)					0	0	0	0	0	0	0	0	1	0
Responsabilité d'une structure pédagogique (officine virtuelle...) (1/ strature / année de responsabilité)					0	0	0	0	0	0	0	0	3	0
SCORES ANNUELS		0	0	0	0	0	2	2	2	0	6	12		24
SIAPS ANUUELS		0	0	0	30.75	30.75	104.75	102	89.5	93	111.75	122.5	**55.5**	685
													100	

SIAPS Total

3. Activités de recherche

Je développe mon activité de recherche dans l'Unité Mixte de Recherche 7039 CNRS CRAN Université de Lorraine, dirigée par le Professeur Didier Wolf.

Les recherches développées par le laboratoire sont à caractère tant fondamental qu'appliqué, à finalité industrielle ou sociétale (santé et environnement notamment, *cf* Organigramme ci-dessous). Une originalité de l'unité est de développer également une recherche interdisciplinaire associant la biologie et les sciences de l'automatique au sens large appliquée à la santé (cancérologie et neurologie). Les thèmes de recherche couvrent les secteurs de l'automatique et de la productique, du traitement du signal et des images, de la biologie en cancérologie et en neurologie, du génie informatique et des réseaux de communication et de la sûreté de fonctionnement des systèmes.



Ma thématique principale est centrée sur la recherche de biomarqueurs prédictifs de réponse aux thérapies ciblées dans les cancers digestifs au sein du département BioSIS du CRAN (Biologie, Signaux et Systèmes en Cancérologie et Neurosciences) sur le site de l'Institut de Cancérologie de Lorraine au sein duquel je suis responsable de l'axe de recherche translationnelle « Diagnostic moléculaire approfondi, Biomarqueurs et réponse thérapeutique ». Il existe au sein du Département BioSis plusieurs équipes de recherche, elles aussi dédiées au cancer, avec notamment des travaux étudiant les cancers du sein ainsi que les tumeurs cérébrales.

Mon activité de recherche s'appuie sur différentes infrastructures et collaborations :

- L'Unité de Recherche Translationnelle de l'ICL pour les projets précliniques : avec deux enseignants chercheurs (Dr. Pauline Gilson et moi), 2 techniciens de laboratoire, 0,5 ETP d'assistant administratif, 2 étudiants en PhD (Andréa Witz et Julie Dardare), 1 étudiant en Master et un étudiant en Licence Professionnelle. Cette unité est intégrée au sein du CRAN. Viennent s'ajouter dans cette unité les Dr. Lina Bolotine Lina (MCU-PH), Dr. Henri-Pierre Lasalle (MCF) et le Dr. Ilya Yakavets (post-doc) travaillant sur une thématique « Nanovectorisation et évaluation préclinique de molécules à activité photodynamique ».
- La Direction de le recherche clinique de l'ICL pour la partie clinique, constituée d'une cellule de data biostatistiques (1 Biostatisticien), un centre de promotion des essais cliniques (3 chefs de projet, 2 ARC et 0,5 assistant administratif) et un centre de recherche clinique (1 cadre, 7 ARC, 2 IRC, 0,8 Assistant administratif).
- Le Réseau Biopsie liquide du Cancéropole de l'Est, coordonné par le Professeur Jean-Louis Merlin, regroupant les acteurs en biopsie liquide au sein du cancéropôle Est (régions Grand-Est et Bourgogne Franche Comté).
- Des Groupe Nationaux auxquels je participe, notamment l'UCGI (Unicancer gastrointestinal), l'ELBS (European Liquid Biopsy Society), le groupe PRODIGE de la FFCD (Fédération Française de Cancérologie Digestive) et le GFCO (Groupement Français de Cytogénomique en Oncologie).
- Des collaborations internationales, en particulier les équipes du Pr. Charles Swanton et du Dr. Samra Turajlic au Francis Crick Institute de Londres, l'équipe du Dr. Kevin Litchfield de l'UCL à Londres, et l'équipe du Pr. Nikolas von Bubnoff de l'Université de Freiburg en Allemagne.
- Des collaborations industrielles avec des fabricants d'instruments et/ou de réactifs, notamment Biocartis (Belgique), BioRad (USA), IT Solutions (France), Sophia Genetics (Suisse) et Nonacus (Royaume-Uni).

3.1. Travaux de recherche dans le cadre de mon diplôme de Doctorat d'Université

Récepteurs HER, voies de signalisation et prescription des thérapies ciblées chez les patients atteints d'un cancer colorectal métastatique

Le développement des thérapies personnalisées et plus récemment de la théranostique a complètement modifié l'approche biologique de la maladie cancéreuse. Ainsi, aux marqueurs sériques classiques et parfois peu spécifiques comme par exemple le CA-125, l'ACE, le CA 19-9 ou encore le CA 15-3 viennent aujourd'hui s'ajouter de nombreux marqueurs tissulaires tumoraux, plus représentatifs de la maladie et surtout prédictifs de la résistance ou de la réponse à une thérapie ciblée, associée ou non à une chimiothérapie traditionnelle.

Les récepteurs à tyrosine kinase de la famille de l'Epidermal Growth Factor (EGFR ou HER) constituent un domaine d'intérêt démontré en cancérologie. De nombreuses molécules de cette classe font l'objet d'une utilisation thérapeutique chez l'homme dans le cadre d'une AMM. C'est le cas de médicaments comme les anticorps monoclonaux anti-

EGFR (cetuximab, panitumumab) ou les inhibiteurs de kinase (regorafenib) utilisés dans le traitement des cancers colorectaux.

On trouve en aval de ces récepteurs des voies de signalisation majeures ayant démontré leur impact dans le développement de la maladie cancéreuse. C'est le cas par exemple de la voie PI3K/AKT/mTOR ou encore de la voie des MAP-kinases qui jouent un rôle important dans le contrôle de la prolifération et de la mort cellulaire. Ces voies de signalisation font intervenir des protéines effectrices, pouvant faire office de cibles thérapeutiques intéressantes, notamment en cas de présence d'une anomalie somatique ou fonctionnelle, ce qui est souvent le cas en cancérologie.

L'étude des biomarqueurs prédictifs de réponse aux thérapies ciblées dans les cancers colorectaux métastatiques a constitué le sujet de ma thèse d'Université, soutenue publiquement le 13 Novembre 2014.

Le cancer colorectal est l'un des cancers le plus diagnostiqué avec plus d'un million de nouveaux cas à travers le monde chaque année. En France, cette maladie concerne 35 000 nouveaux patients chaque année et est responsable de plus de 16 000 décès. On estime qu'environ 60% des patients ont une évolution métastatique de leur maladie.

Depuis le début des années 1980, on connait le rôle de l'oncogène RAS dans la voie de signalisation cellulaire des MAPKinases. Les mutations sur les codons 12 et 13 du gène *KRAS* ont été identifiées en 2008 comme des marqueurs prédictifs de résistance aux anticorps monoclonaux anti-EGFR associés à une chimiothérapie de type FOLFOX ou FOLFIRI chez les patients atteints d'un cancer colorectal métastatique. En 2013, des études cliniques ont démontré que la connaissance du statut mutationnel des codons 12 et 13 du gène *KRAS* ne suffisait pas à expliquer tous les échecs thérapeutiques aux anticorps anti-EGFR. Il s'avère que la présence d'une mutation sur les exons 2, 3 et 4 de *KRAS* ou sur les exons 2, 3 et 4 du gène *NRAS*, est également prédictive de la réponse aux traitements par anticorps anti-EGFR. Il est ainsi rendu obligatoire par l'EMA d'effectuer un génotypage des exons 2, 3 et 4 des gènes *KRAS* et *NRAS* avant toute prescription d'un anti-EGFR. Ces recherches de mutations peuvent s'effectuer de différentes façons, notamment en utilisant des techniques de PCR temps-réel.

Les mutations les plus fréquemment rencontrées dans les cancers colorectaux métastatiques se situent sur les codons 12, 13 (exon 2), 59, 61 (exon 3), 117 et 146 (exon 4) de *KRAS* et *NRAS* et sont présentes dans environ 50% des cas. Parmi les patients « sauvages », c'est à dire dont les tumeurs ne portent pas de mutations sur les gènes KRAS et NRAS, on rencontre dans environ 10% des cas, une mutation du gène *BRAF*. La présence d'une mutation (essentiellement sur le codon 600, exon 15 du gène) est synonyme de mauvais pronostic mais il n'existe pas de lien entre la réponse aux thérapies ciblées anti-EGFR et la mutation de *BRAF* dans les cancers colorectaux métastatiques. Au même titre que les mutations de *BRAF*, les mutations du gène *PIK3CA*, codant pour la sous-unité catalytique de PI3K (Phosphatidyl-inositol-3-kinase), enzyme intervenant directement en aval des récepteurs EGFR au niveau de la voie PI3K/AKT/mTOR, pourraient avoir un rôle pronostique, mais ne démontrent à priori pas d'intérêt dans la prédiction à la réponse aux thérapies ciblées actuellement sur le marché. Cependant, l'étude de ces mutations pourraient avoir un intérêt à court terme, notamment si l'utilisation d'inhibiteurs de PI3K venait à montrer un intérêt dans le traitement des cancers colorectaux métastatiques.

L'étude de l'expression des formes phosphorylées des protéines intervenant dans les voies de signalisation en aval des récepteurs à l'EGF peut également être d'intérêt dans la prédiction de la réponse aux thérapies ciblées anti-EGFR apportant une vision fonctionnelle aux données de génétique. Lors de mon travail de thèse, j'ai étudié l'impact de l'expression de certaines phosphoprotéines, impliquées dans les voies de signalisation RAS/RAF/MAPKinases et PI3K/AKT/mTOR sur des échantillons de tumeurs prélevées chez cent patients atteints d'un cancer colorectal métastatique et ayant reçu un traitement anti-EGFR en association à une chimiothérapie. Ce travail effectué en collaboration avec le Pr. Pierre Laurent-Puig (HEGP, INSERM, Université Paris Descartes) a permis de mettre en évidence, chez les patients dont les tumeurs étaient sauvages pour les exons 2, 3 et 4 de *KRAS* et *NRAS*, que la surexpression de pEGFR et de pAKT était prédictive d'une meilleure réponse aux anticorps monoclonaux ciblant l'EGFR au seuil fixé (Harlé *et al.*, Br J Cancer, 2015). Ainsi, les données recueillies lors de cette étude mettent en évidence l'intérêt potentiel d'une association anti-EGFR et anti-AKT chez les patients atteints d'un cancer colorectal métastatique et dont les tumeurs sont RAS sauvage. Ces données suggèrent également que l'étude de l'expression des phosphoprotéines en aval des voies de signalisation RAS/RAF/MAPKinases et PI3K/AKT/mTOR, permettrait de mieux sélectionner les patients pouvant bénéficier des anticorps monoclonaux anti-EGFR, puisque d'après les études actuelles, on estime que 40 à 60% des patients avec des tumeurs RAS sauvage ne sont pas ou peu répondeurs à ces thérapies (*cf.* Publication 3).

3.2. Activités de recherche depuis l'obtention du Doctorat d'Université

Depuis l'obtention de mon Doctorat d'Université en 2014, nous avons poursuivi au laboratoire nos travaux dans le cancer colorectal, mais également diversifié nos travaux de recherche en les consacrant au développement du diagnostic moléculaire dans différents types de cancers, que ce soit sur tissu ou en biopsie liquide.

Cancer colorectal

Il a été précédemment décrit qu'un certain nombre de mutations dites « hotspot » des gènes *KRAS* et *NRAS* étaient responsables de résistances au traitement par anticorps monoclonaux anti-EGFR chez les patients atteints d'un cancer colorectal métastatique. Nous avons alors voulu nous poser la question de l'intérêt des autres mutations de *KRAS*, *NRAS* mais également du gène *BRAF* chez ces patients. Nous avons ainsi pu démontrer que chez ces patients, des mutations du gènes *KRAS* plus rares que les mutations hotspots, pouvaient potentiellement induire une résistance aux anti-EGFR (Harlé *et al.*, Target Oncol, 2016). Grâce à des modèles prédictifs *in silico*, nous avons également pu démontrer l'intérêt de l'étude des mutations des gènes *KRAS*, *NRAS* et *BRAF* et leur implication dans la résistance aux anti-EGFR (Franczak *et al.*, Sci Rep, 2019).

Développement du diagnostic moléculaire

Compte tenu de l'importance grandissante du diagnostic moléculaire en cancérologie, nous avons également voulu ces dernières années évaluer les nouvelles technologies disponibles. Nous avons ainsi pu évaluer et comparer un certain nombre de

nouvelles techniques de biologie moléculaire, PCR ou NGS, notamment pour la détermination fiable et rapide, sur tissu FFPE, des mutations du gène *BRAF* (Harlé *et al.*, PLoS One, 2016, Franczak *et al.*, Mol Diagn Ther, 2017, Serre *et al.*, Oncotarget, 2018), *PIK3CA* (Lambert *et al.*, Pathol Oncol Res, 2019) ou *KRAS*, *NRAS* (Chen *et al.*, Oncol Lett, 2016, Franczak *et al.*, PLoS One, 2019). Des études concernant l'ADN tumoral circulant ont été également effectuées pour la détermination fiable des mutations des gènes KRAS, NRAS et BRAF en biopsie liquide (Franczak *et al.*, PLoS One, 2020).

3.3. Activités de recherche au cours de mon année de mobilité internationale

Pendant mon année de mobilité internationale au Francis Crick Institute, sous la Direction du Professeur Charles Swanton et du Docteur Samra Turajlic, j'ai pu participer aux recherches déjà en cours dans le domaine de l'hétérogénéité tumorale et de l'évolution clonale des tumeurs. J'ai ainsi participé au programme TRACERx renal et TRACERx melanoma. Ces travaux sont toujours en cours. Nous avons cependant pu mettre en évidence l'intérêt d'une nouvelle approche, le Representative Sequencing (RepSeq).

Dans l'étude TRACERx, le recours au multi-échantillonnage des lésions primitives ou tumorales a pu démontrer l'importance de l'évolution clonale des tumeurs au cours de la maladie. Le multi-échantillonnage, c'est-à-dire parfois plus d'une centaine de biopsies tissulaires par patient, n'est pas réalisable en routine clinique, pour des raisons de temps, mais également financières. Il a ainsi été développé une approche innovante consistant à substituer le multi-échantillonnage par le recours à un broyat des tissus tumoraux (Figure 1). Cette approche permet ainsi de ne séquencer qu'un seul échantillon par patient au lieu de plusieurs dizaines, afin de reconstituer l'arbre phylogénétique de la tumeur.

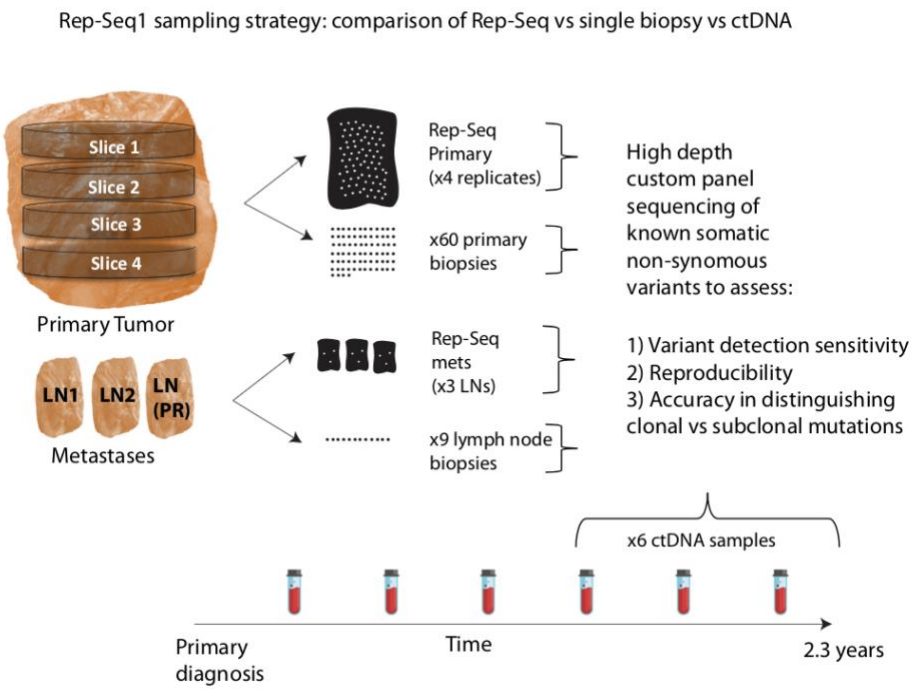


Figure 1 : Principe du RepSeq. La technique a été appliquée aussi bien sur la lésion primitive que sur la biopsie, et a été comparée à une approche de multi-échantillonnage de 60 prélèvements pour le primitif et de 9 pour les lésions métastatiques (d'après Litchfield *et al.*).

L'étude publiée dans Cell Repors en 2020 démontre ainsi tout l'intérêt de cette technique (Litchfield *et al.*, Cell Rep, 2020). En effet, la quasi-totalité des mutations retrouvées en multi-échantillonnage sont retrouvée par l'approche RepSeq. Il a également été possible d'identifier les mutations clonales des mutations sous-clonales, ainsi que les variations du nombre de copies somatiques des différents gènes (Figure 2). Dans cette publication, nous avons pu également mettre en évidence que l'analyse de l'ADN tumoral circulant extrait du plasma de ces patients permettait de retrouver plus de 80% des mutations, qu'elles soient clonales ou sous clonales (*cf.* Publication sélectionnée n˚ 1).

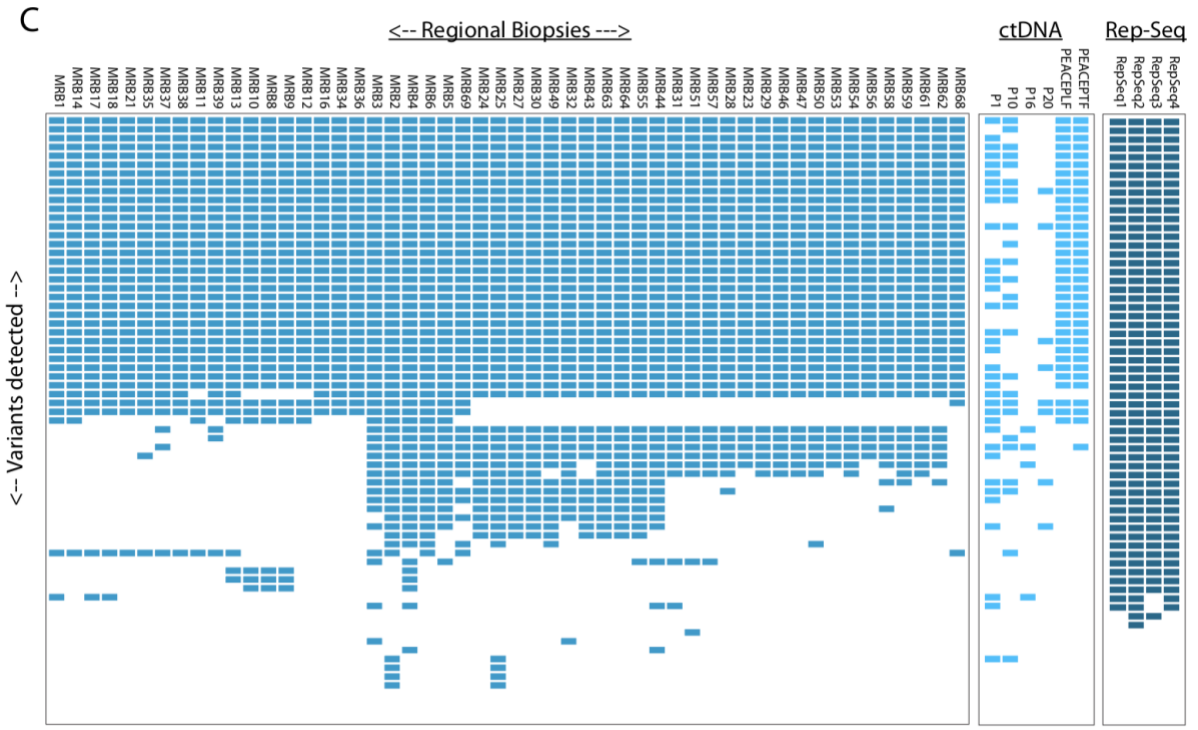


Figure 2 : Résultats obtenus pour l'étude d'un patient en RepSeq. Les quatre analyses en RepSeq donnent des résultats équivalents entre eux, et permettent également d'identifier la grande majorité des mutations identifiées en technique de multi-échantillonnage. Les données obtenues en biopsie liquide permettent également de retrouver une majorité des anomalies identifiées en RepSeq et en multi-échantillonnage (d'après Litchfield *et al.*).

J'ai pu également apprendre beaucoup dans le domaine des signatures génomiques et mettre ces nouvelles connaissances à profit pour l'identification de l'origine pulmonaire des tumeurs lorsque l'analyse anatomopathologique ne permettait pas leur identification de manière directe. Dans le travail publié dans Scientific Reports en 2019 (Busca *et al.*, Sci Rep, 2019), nous montrons, à l'aide d'une signature génomique simple et peu coûteuse, que l'origine pulmonaire sur d'autres origines, est possible avec une sensibilité de 83%. Pour cela, nous avons utilisé les données publiques du TCGA, puis testé de manière indépendante et prospective cette signature sur un set de données provenant du Centre Georges François Leclerc de Dijon (Figure 3, *cf.* publication sélectionnée n˚2).

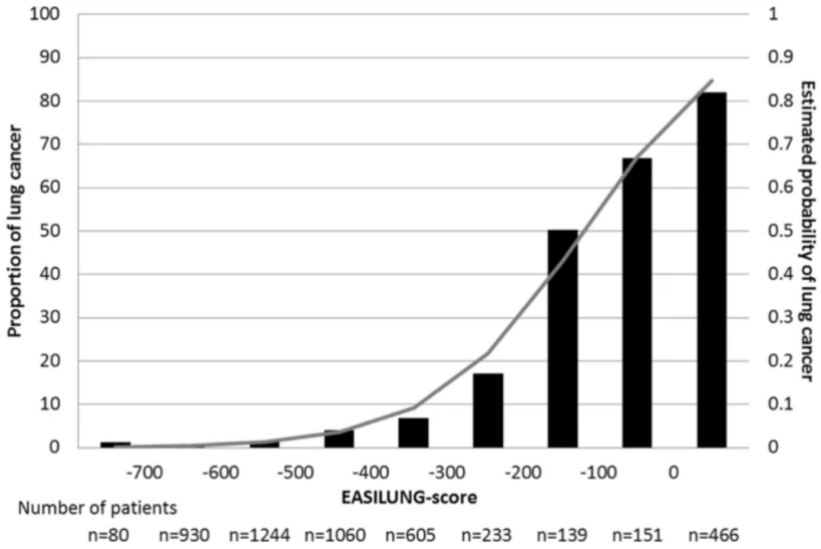


Figure 3 : Description du score EASILUNG sur le set de données de développement. Cette analyse inclus 4908 tumeurs avec une proportion de mutations somatiques supérieure à 1 substitution par mégabase. Ce graphique montre la probabilité que le cancer soit d'origine pulmonaire en fonction du score calculé (d'après Buscat *et al.*).

3.4. Activités de recherche en cours

Mes activités de recherche en cancérologie s'inscrivent dans un contexte hospitalo-universitaire et me permettent de développer une thématique préclinique, notamment sur le rôle de la protéine DDB2 dans la cancérogénèse des cancers du pancréas, et une thématique translationnelle centrée sur l'apport de l'étude de l'ADN tumoral circulant dans l'évaluation de l'hétérogénéité tumorale et l'évolution clonale.

3.4.1. Rôle de la protéine DDB2 dans la cancérogenèse des cancers du pancréas exocrines

Contexte

Les adénocarcinomes canalaires pancréatiques (PDAC) représentent un problème majeur de santé publique en raison de leur taux de survie à 5 ans inférieur à 6% (Conroy *et al.*, Eur J Cancer, 2016) et de l'inefficacité des thérapies ciblées actuelles pour ce type de cancer (Moore *et al.*, Cells, 2020). D'ici 2030, les PDACs pourraient représenter une des causes majeures de décès liées au cancer (Rahib *et al.*, Cancer Res, 2014). Chez les patients non opérables d'emblée, les seules options thérapeutiques à disposition des cliniciens sont les chimiothérapies de type FOLFIRINOX ou gemcitabine (Conroy *et al.*, N Engl J Med, 2018) et à ce jour aucun biomarqueur pronostique et prédictif de réponse au traitement n'a encore été identifié. Dans ce contexte, les PDACs représentent aujourd'hui un des principaux axes de recherche développés à l'Institut de Cancérologie de Lorraine (Pr. T. Conroy) et par le Cancéropôle Est.

La protéine DDB2 (DNA-Damaged Binding Protein 2) est une protéine nucléaire de 48 kDa, historiquement connue pour son implication dans l'initiation de la réparation de l'ADN

endommagé via le système NER de réparation par excision de nucléotides (Luijsterburg *et al.*, J Cell Biol, 2012). La protéine DDB2 intervient dans ce système de réparation en interagissant avec la protéine DDB1 pour former le complexe UV-DDB (Scrima *et al.*, Cell, 2008). DDB2 est également impliqué dans d'autres processus cellulaires essentiels tels que le remodelage de la chromatine (Kattan *et al.*, PLoS One, 2008, Zhao *et al.*, J Biol Chem, 2009), la transcription de certains gènes (Nag *et al.*, J Virol, 2001), la régulation du cycle cellulaire et la dégradation de certaines protéines (Yan *et al.*, Cell Cycle, 2011).

Nos collègues de l'UMR CNRS 7039 CRAN (équipe du Pr Becuwe) et d'autres chercheurs ont mis en évidence l'intérêt de la protéine DDB2 dans le développement et la progression de divers cancers (tels que les cancers du sein, de l'ovaire, colorectal, de la prostate, de la peau, de la tête et du cou) (Gilson *et al.*, Int J Mol Sci, 2019) (Figure 4). Des études sur différentes modèles de cancers humains ont permis de mieux définir l'activité biologique de DDB2 dans les cellules tumorales et de mettre en évidence sa dualité de rôle dans ces cancers, ayant parfois des propriétés de suppresseur de tumeurs et fonctionnant parfois comme un oncogène (Gilson *et al.*, Int J Mol Sci, 2019). DDB2 exerce une régulation à plusieurs niveaux, modulant la prolifération, la survie, la migration et l'invasion des cellules cancéreuses, l'angiogenèse et la formation de cellules souches cancéreuses ainsi que la sensibilité aux chimiothérapies et aux thérapies ciblées.

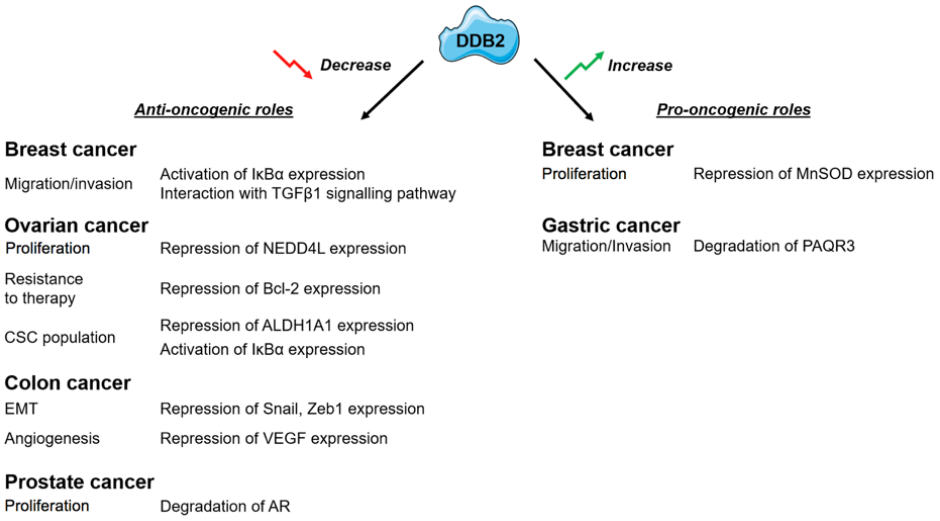


Figure 4 : Rôle de la protéine DDB2 sur différents cancers solides et ses partenaires identifiés (Gilson *et al.*, 2019). Abréviations: AR (Androgen Receptor), ALDH1A1 (Aldehyde dehydrogenase 1 family, member A1), Bcl-2 (B-cell lymphoma 2), c-Flip (Cellular FLICE-like inhibitory protein, EMT (Epithelial–mesenchymal transition), HEKs (Human epidermal keratinocytes), I_B_ (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha), MEFs (Mouse embryonic fibroblasts), MnSOD (Manganese superoxide dismutase), NEDD4L (Neural precursor cell expressed developmentally downregulated gene 4-like), PAQR3 (Progestin and adipoQ receptor family member 3), TGF-_1 (Transforming growth factor beta 1), VEGF (vascular endothelial growth factor), Zeb1 (Zinc finger E-box-binding homeobox 1). (Gilson *et al.*, Int J Mol Sci, 2019)

À ce jour, aucune donnée de la littérature ne mentionne le rôle de DDB2 dans la cancérogenèse pancréatique. Or DDB2 a été décrite comme capable de réguler la voie du TGF-β (Zhao *et al.*, Int J Clin Exp Pathol, 2015) et l'analyse transcriptomique du TCGA a mis en évidence que cette voie de signalisation est activée dans près de la moitié des adénocarcinomes pancréatiques (Bailey *et al.*, Nature, 2016). Notre objectif est donc

d'étudier cette hypothèse et d'étendre la connaissance de la protéine DDB2 aux cancers du pancréas (Dardare *et al.*, Int J Mol Sci, 2020).

Résultats préliminaires et recherche en cours

Dans le cadre de cette étude, 3 modèles cellulaires de PDAC (BxPC-3, Capan-2 et T3M4) sont utilisés. Ces lignées présentent des profils mutationnels et un potentiel de prolifération, de différenciation, de migration et d'invasion variés (d'après les données de la littérature et confirmation par données préliminaires réalisées au laboratoire, tableau 1). Les cellules Capan2, bien différenciées, *KRAS* et *CDKN2A* mutées, présentent une prolifération lente avec un temps de doublement cellulaire de 75,3h ainsi qu'un potentiel métastatique important. Les cellules BxPc3 et T3M4 ont un état de différenciation modéré à faible et un temps de prolifération plus faible (57,7h et 36,3h respectivement). Les cellules BxPC3 présentent une mutation homozygote du gène *TP53* ainsi que des délétions des gènes *SMAD4* et *CDKN2A* et ont des capacités d'invasion importantes tandis que les cellules T3M4, mutées au niveau des gènes *TP53, KRAS* et *SMAD4,* sont peu invasives.

Tableau 1. Caractéristiques des lignées cellulaires de PDAC Capan-2, BxPc3 et T3M4 : profil mutationnels, potentiel de prolifération, de différenciation, de migration et d'invasion

lignée cellulaire	origine	état de différenciation	prolifération (temps de doublement en heures)	potentiel migratoire et invasif	profil mutationnel
BxPC3	tumeur primaire de PDAC	modéré à faible	57,7h	++	TP53 c.659A>G p.(Tyr220Cys) (homozygote) CDKN2A c.-10_458del p.(Met1_*157del) (homozygote) SMAD4 c.905-10_1448del p.(Trp302_*553del) (homozygote)
Capan-2	tumeur primaire de PDAC	bien différencié	75,3h	++	KRAS c.35G>T p.(Gly12Val) (hétérozygote) CDKN2A c.52_57dupACGGCC p.(Thr18_Ala19dup) (homozygote)
T3M-4	tumeur primaire de PDAC	faible	36,3h	-	SMAD4 c.1139+2T>A p.? (homozygote) TP53 c.659A>G p.(Tyr220Cys) (homozygote) KRAS c.183A>C p.(Gln61His) (hétérozygote)

Le niveau d'expression protéique et transcriptionnel de DDB2 a été déterminé dans ces différentes lignées par western blot et RT-PCR respectivement. Les données préliminaires montrent une faible expression de DDB2 dans la lignée Capan2 et une forte expression de DDB2 dans les lignées BxPC3 et T3M4. Ces résultats suggèrent une relation entre niveau d'expression de DDB2 et potentiel prolifératif (figure 5).

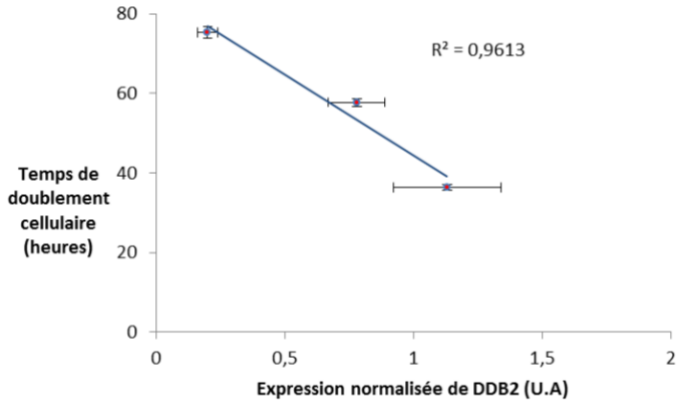


Figure 5. Temps de doublement des cellules de PDAC (Capan-2, BxPc3, T3M4) en fonction du niveau d'expression protéique de DDB2 dans ces cellules.

Afin de confirmer cette hypothèse et établir le lien entre expression de DDB2 et propriétés des lignées de PDAC, nous avons réalisé le knock-out de DDB2 (KO DDB2) sur les cellules exprimant fortement cette protéine (BxPc3 et T3M4) par techniques de shRNA et CRISPR-Cas9 (Figure 6).

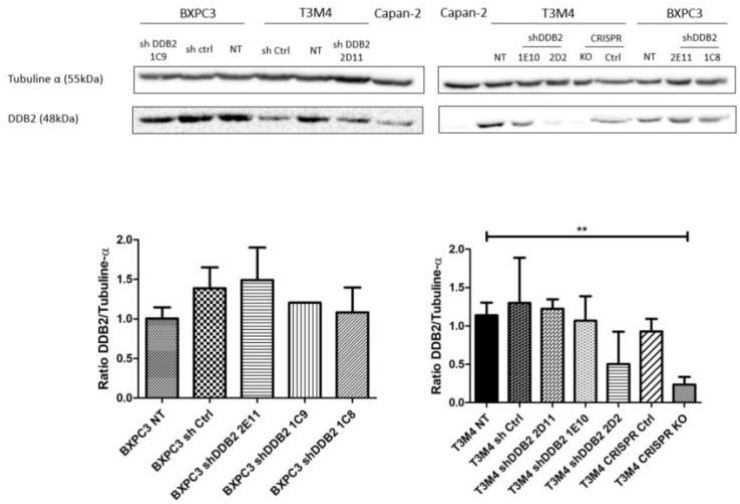


Figure 6. Semi-quantification de la protéine DDB2 dans les lignées BxPC3 et T3M4 transfectées et non transfectées (**p=0,0048 ; n=3).

Nous avons ensuite comparé les capacités prolifératives, migratoires et invasives des cellules KO DDB2 comparativement à des cellules contrôles par tests de croissance, tests de blessure et tests d'invasion (transwell) respectivement. L'expression de DDB2 était synonyme d'un caractère moins prolifératif et moins invasif.

L'étude morphologique des cellules KO DDB2 par microscopie n'a pas montré de différences significatives.

Le niveau d'expression des facteurs impliqués dans la TEM a été étudié par Western Blot. Une différence significative de l'expression de ces marqueurs en fonction de l'expression de DDB2 a été retrouvée, montrant une potentielle implication de la protéine DDB2 dans le phénomène de TEM.

Nous avons également évalué par des tests de cytotoxicité la sensibilité des cellules KO DDB2 aux différentes chimiothérapies utilisées pour le traitement des PDAC, notamment le protocole FOLFIRINOX, la gemcitabine et l'olaparib (exemples avec les figures 7 et 8).

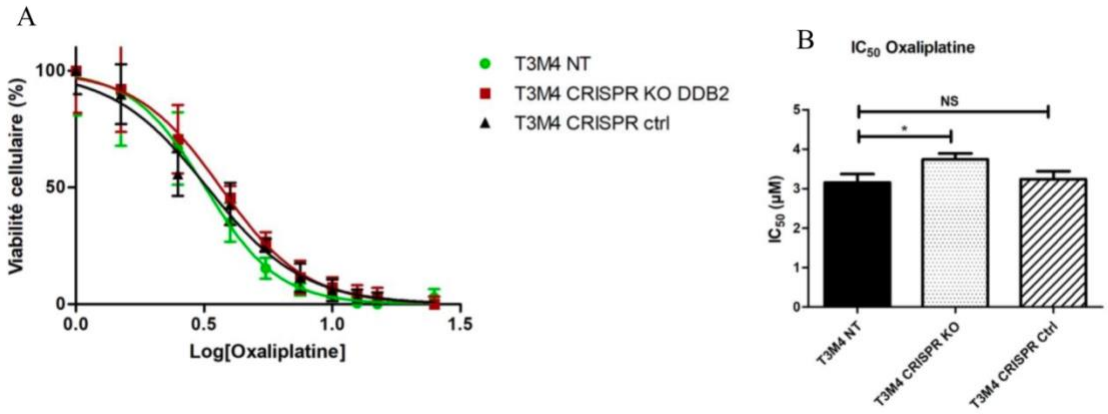


Figure 7. A) Courbe de viabilité cellulaire des lignées T3M4 NT, T3M4 CRISPR KO, T3M4 CRISPR Ctrl sous l'effet de doses croissantes d'oxaliplatine [0-25µM] (n=3) B) Concentrations inhibitrices 50 d'oxaliplatine obtenues pour les lignées T3M4 (n=3) ; * p=0,0433 ; NS : non significatif.

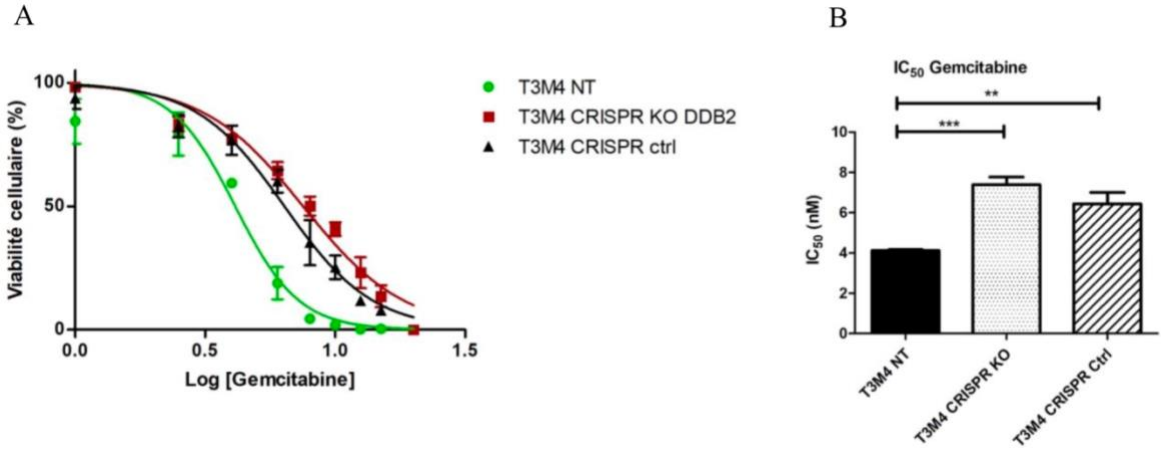


Figure 8. A) Courbe de viabilité cellulaire des lignées T3M4 NT, T3M4 CRISPR KO, T3M4 CRISPR Ctrl sous l'effet de doses croissantes de gemcitabine [0-20nM] (n=3) B). Concentrations inhibitrices 50 de gemcitabine obtenues pour les lignées T3M4 (n=3) ; ***p=0,0005 ; **p=0,0076

Nos résultats préliminaires sont donc en faveur d'une plus grande chimiosensibilité lorsque les cellules expriment DDB2.

L'activation transcriptionnelle de DDB2 dans les cellules exprimant faiblement DDB2 (Capan2) a été réalisée par technique CRISPRa et le modèle cellulaire maintenant disponible au laboratoire. Les conséquences de cette surexpression de DDB2 sont en cours d'évaluation selon les mêmes modalités que celles décrites précédemment.

Nous estimons que les résultats finaux de ces travaux pourront être soumis à publication d'ici fin 2020 (*cf.* Publications sélectionnées 4 et 5)

Collaborateurs : Dr. A. Harlé, Dr. P. Gilson, Pr JL. Merlin, A. Witz (PhD débuté en Novembre 2019), J. Dardare (PhD débuté en Novembre 2020) et un étudiant en M2R
Financement : Actions incitatives du CRAN et Institut de Cancérologie de Lorraine.

3.4.2. Activités de recherche translationnelle

Une part importante de mon activité de recherche consiste en l'élaboration et à la coordination biologique d'essais thérapeutiques. Les différents essais en cours sont ont en commun la détection d'ADN tumoral circulant, l'exploration génomique des tumeurs et de l'hétérogénéité tumorale.

Nous étudions également, dans une moindre mesure, les cancers HPV-induits et l'apport de la biopsie liquide pour ces patients.

ADN tumoral circulant et hétérogénéité tumorale

La présence d'acides nucléiques dans le sang est décrite depuis 1948 (Mandel *et al.*, C R Seances Soc Biol Fil, 1948), mais c'est seulement depuis quelques années, que l'on sait que les tumeurs sont capables de diffuser activement ou passivement leur ADN dans les fluides biologiques (Franczak *et al.*, Arch Med Res, 2018). Les progrès technologiques, notamment dans le domaine de la biologie moléculaire ont permis leur détection en vue d'applications dans le domaine de l'oncologie médicale. La détection de l'ADN tumoral circulant, matériel génétique tumoral relargué par la tumeur lors de la nécrose, de l'apoptose ou encore de phénomènes cellulaires actifs, est ainsi envisagée pour des applications de détection précoce, de théranostique, de suivi des patients, d'évaluation de la qualité d'un acte chirurgical ou encore de détection précoce de la rechute des patients atteints de cancer (Franczak *et al.*, Arch Med Res, 2018).
Nous avons développé depuis 2015 au laboratoire une expertise dans la détection de l'ADN tumoral circulant en utilisant à la fois des techniques de PCR (temps-réel, digital droplet, BEAMing) et de séquençage haut-débit (NGS).
Différents objectifs sont affichés pour l'activité biopsie liquide du laboratoire : L'amélioration des techniques de détection afin d'en améliorer leur sensibilité au travers de programmes techniques et leur application dans le domaine de la théranostique et du suivi des patients atteints de cancer au travers d'essais cliniques. Ces programmes de recherche ont pour finalité de déterminer la place de la détection de l'ADN tumoral circulant comme marqueur substitutif, exhaustif, dynamique et moins invasif à la biopsie tissulaire dans la prise en charge des patients atteints de cancer.

Nous travaillons également plus récemment sur des projets de recherche visant à étudier dans quelle mesure l'ADN tumoral circulant pourrait être un bon support pour prédire l'évolution clonale et sous clonale de la maladie. En effet, il est décrit depuis quelques années que le cancer est une maladie clonale avec une évolution de type Darwinien (Greaves *et al.*, Nature, 2012) et des programmes ambitieux, comme les programmes TRACERx sur lesquels j'ai eu l'opportunité de travailler sous la direction du Pr C Swanton

pendant mon année de mobilité au Francis Crick Institute de Londres, ont permis de mieux décrire et comprendre les aspects évolutifs du cancer.
Le programme biopsie liquide implique une partie du personnel du service de biologie moléculaire des tumeurs de l'ICL ainsi que des médecins oncologues ou spécialistes d'organes. Des sujets relatifs à la détection de l'ADN tumoral circulant sont aussi régulièrement proposés aux étudiants souhaitant valider un Master2 Recherche. Les collaborations sont ensuite détaillées par projet. Tous les projets sont systématiquement discutés, élaborés et suivis par la cellule méthodologie data-biostatistiques et le service de promotion des essais cliniques de l'ICL en respect des bonnes pratiques cliniques et de la réglementation en vigueur (loi Jardé) dans un environnement certifié vis à vis de la norme ISO9001:2015.

TransPANDAS (2016-en cours)
L'étude TransPANDAS est une étude translationnelle à l'étude PANDAS – PRODIGE 44, étude nationale (PI Pr T Conroy) supportée par un PHRC et la Ligue Nationale contre le Cancer. L'essai clinique PANDAS s'adresse aux patients atteints d'adénocarcinome du pancréas à la limite de la résécabilité. L'objectif est d'évaluer chez ces patients l'apport avant chirurgie d'une radiochimiothérapie concomitante à une chimiothérapie néoadjuvante par mFolfirinox. Cette combinaison de traitements devrait permettre une chirurgie sans résidus et donc d'éviter la récidive locale. Les patients ayant consentis à la participation à la collection biologique sont inclus dans TransPANDAS, étude ancillaire translationelle dont le but est de déterminer l'apport des biopsies liquides dans cette cohorte de patients (analyse de l'ADN et de l'ARN circulant). L'étude est toujours en cours de recrutement et les analyses n'ont pas encore débuté.
Collaborateurs : Pr. J.L. Merlin (PI), Pr. T. Conroy , Dr. A. Harlé (ICL, Nancy).
Financement : Ligue contre le cancer inter-régional Grand Est Bourgogne Franche Comté

Pancréas CGE (2016-en cours)
Notre laboratoire participe en collaboration avec le Cancéropôle de l'Est à l'analyse de l'ADN tumoral circulant des prélèvements sanguins recueillis dans la cohorte « Pancréas CGE ». Cette cohorte est constituée de patients atteints d'un adénocarcinome pancréatique résécable. Le but est d'évaluer la place de la biopsie liquide comme marqueur pronostic et comme marqueur de suivi des patients. L'étude est toujours en cours de recrutement mais je viens d'obtenir un financement pour débuter l'analyse des échantillons des 24 premiers patients (séquençage de la tumeur et des 4 premiers prélèvements sanguins).
Collaborateurs : Pr. B. Heyd (PI, CHU Besançon), Pr. C. Borg (CHU, Besançon), Pr. F. Ghiringhelli (CGFL, Dijon), Pr. P. Rat (CHU, Dijon), Pr. A. Ayav (CHRU, Nancy), Pr. T. Conroy (ICL, Nancy), Pr. A. Kianmanesh (CHU, Reims), Pr. P. Dufour (CPS, Strasbourg), Pr. P. Bachellier (CHU, Strasbourg), Dr. R. Boidot (CGFL, Dijon), Dr. E. Pancreach' (CHU, Strasbourg), Dr. A. Harlé (ICL, Nancy).
Financement : Cancéropôle Est

ColoBEAM (2016-2018)
278 patients atteints d'un cancer colorectal métastatique ont été inclus dans l'étude prospective multicentrique ColoBEAM. Le but de cet essai clinique était de déterminer la place de la biopsie liquide et de la technique OncoBEAM (PCR BEAMing) comme substitut à

la biopsie tissulaire pour l'évaluation des statuts *KRAS*, *NRAS* et *BRAF* à des fins théranostiques. Les premiers résultats ont été soumis sous forme d'abstract à l'ASCO 2019. La publication définitive des résultats sera soumise fin 2020.
Collaborateurs : Dr. C. Gavoille (PI, ICL, Nancy), Pr. O. Bouché (CHU, Reims), Pr. M. Ben Abdelghani (CPS, Strasbourg), Dr. J. Plaza (CHR, Metz), Dr. Lambert (ICL, Nancy), Dr. D. Spaeth (Polyclinique Gentilly, Nancy), Dr. A. Boudrant (CHR, Châlon sur Saône), Pr. F. Ghiringhelli (CGFL, Dijon), Dr. A.L. Villing (CH, Auxerre), Pr. C. Borg (CHU Besançon), Dr. F. S. Jones (Sysmex Inostics, USA), Dr. D. Edelstein (Sysmex Inostics, USA), Pr. J.L. Merlin (ICL, Nancy), Dr. A. Harlé (ICL, Nancy).
Financement : Merck, Sysmex-Inostics, Institut de Cancérologie de Lorraine

CICLADES (2017-en cours)
L'étude prospective multicentrique CICLADES propose d'évaluer l'intérêt de la biopsie liquide pour le monitoring notamment des mutations des gènes *ESR1*, *PIK3CA* et *AKT* ainsi que de gènes impliqués dans la résistance aux anti CDK4/6 et dans les voix de signalisations des MAPKinases, pendant le suivi en soin courant de patientes atteintes d'un cancer du sein métastatique traitées par un inhibiteur de l'aromatase associé ou non à un anti CDK4/6.
Collaborateurs : Dr. V. Massard (PI, ICL, Nancy), Dr. M. Gardner (Hôpital Claude Bernard, Metz), Dr. R. Longo (CHR, Metz-Thionville), Dr. C. Becuwe (Centre d'oncologie Gentilly, Nancy), Dr A.M. Savoye (Institut Jean Godinot, Reims), Dr. K. Prulhiere (Polyclinique de Courlancy, Reims), Dr. F. Clatot (Centre Henri Becquerel, Rouen), Dr. M. Achille (Polyclinique de l'Orangerie, Strasbourg), Dr. Y. Tazi (Clinique Saint Anne, Strasbourg), Dr. P. Barthelemy (CHU, Strasbourg), Pr. J.L. Merlin (ICL, Nancy), Dr. P. Gilson (ICL, Nancy), Dr. A. Harlé (ICL, Nancy).
Financement : Astra Zeneca, Institut de Cancérologie de Lorraine

OPTIMEL (2018-en cours)
L'étude prospective multicentrique OPTIMEL vise à étudier l'intérêt du suivi de l'ADN tumoral circulant ainsi que le suivi thérapeutique pharmacologique comme marqueurs de réponse et de rechute chez les patients atteints d'un mélanome métastatique *BRAF* muté et traités par l'association d'inhibiteurs de kinase dabrafenib et trametinib. L'analyse intermédiaire des résultats est programmée pour fin Octobre 2020.
Collaborateurs : Dr. L. Geoffrois (PI, ICL, Nancy), Dr. F. Granel-Brocard (CHRU, Nancy), Dr. N. Gambier (CHRU, Nancy), Dr. J. Scala Bertola (CHRU, Nancy), Pr. J.L. Merlin (ICL, Nancy), Dr. P. Gilson (ICL, Nancy), Dr. A. Harlé (ICL, Nancy).
Financement : Novartis, Institut de Cancérologie de Lorraine

ITMO Cancer (2017-en cours)
Le programme ITMO Cancer ("Mathematical modelling for targeted therapies resistance prediction") est une étude sur 3 ans en collaboration avec les mathématiciens et les statisticiens de l'INRIA et de l'Institut Elie Cartan de Lorraine (IECL). Cette étude vise à mettre au point un modèle mathématique capable de prédire la sensibilité à une molécule de thérapie ciblée ou la survenue d'une résistance chez les patients atteints de cancer à partie de données biologiques obtenues par le séquençage d'ADN tumoral circulant en biopsie liquide veineuse. La preuve de concept sera réalisée sur une cohorte de 60 patientes atteintes d'un cancer du sein et 50 patients atteints d'un mélanome métastatique.

Collaborateurs : Dr. N. Champagnat (INRIA, IECL, Nancy), Pr. P. Vallois (IECL, Nancy), Dr. A. Gegout-Petit (IECL, Nancy), Dr. U. Herbach (IECL, Nancy), Dr. A. Kurtzmann (IECL, Nancy), Dr. A. Gueudin (IECL, Nancy), Dr. C. Fritsch (INRIA, Nancy), Pr. J.L. Merlin (ICL, Nancy), Dr. P. Gilson (ICL, Nancy), Dr. A. Harlé (ICL, Nancy).
Financement : Appel d'offre ITMO Cancer INSERM

BOVARY Pilot (2020 – en cours)
Le protocole BOVARY-pilot est une étude monocentrique de faisabilité dont le but est la détection d'anomalies moléculaires au niveau de l'ADN tumoral circulant (SNV, indels, grands réarrangements, perte d'hétérozygotie, translocations et CNV) chez les patientes atteintes d'un cancer de l'ovaire de haut-grade et potentiellement éligible à un traitement par inhibiteur de PARP. Les gènes BRCA1 et BRCA2 ainsi que les gènes impliqués dans la déficience de la recombinaison homologue seront analysés à la fois sur des prélèvements sanguins et tissulaires. En cas de résultats positifs, l'étude multicentrique BOVARY sera mise en place afin d'inclure plus de patientes et ainsi valider la possibilité de l'utilisation en théranostique de la biopsie liquide pour la mise en place d'un traitement par anti-PARP.
Collaborateurs : Dr. C. Gavoille (PI, ICL, Nancy), Dr. A. Harlé (ICL, Nancy), Dr. P. Gilson (ICL, Nancy), Pr. J.L. Merlin (ICL, Nancy).
Financement : Tesaro-GSK, Institut de Cancérologie de Lorraine

MOSAPAC (2020 – en cours)
MOSAPAC - Etude translationnelle PRODIGE 24 - ACCORD 24 (2019-2022)
493 patients ayant bénéficiés d'une résection R0 ou R1 d'un adénocarcinome du pancréas dans les 3 à 12 semaines précédentes ont été randomisés entre FOLFIRINOX (12 cycles) et Gemcitabine (6 cycles) dans l'étude franco-canadienne de phase III PRODIGE 24 avec comme objectif principal le taux de survie sans récidive à 3 ans. Pour chaque patient ayant consenti à l'étude translationnelle, un bloc de tissu FFPE a été archivé et centralisé. 428 prélèvements de tissus tumoraux fixés au formol et inclus en paraffine sont attendus pour cette étude. Pour chaque prélèvement, une étude en tissue micro-array, RNA seq et DNA Seq d'un panel de plus 400 gènes est prévu. L'objectif de ce travail est d'améliorer les connaissances des aberrations moléculaires des PDAC afin de développer de nouvelles stratégies thérapeutiques, notamment des thérapeutiques ciblées et de l'immunothérapie, et également d'être en mesure de mieux identifier les patients avec les meilleures chances de réponse aux traitements.
Collaborateurs : Pr. T. Conroy (PI, ICL, Nancy), Dr. J. Cros (Hôpitaux Universitaires Paris Nord Val de Seine, Clichy), Pr. J.B. Bachet (Université Paris Sorbonne Cité, Centre Universitaire des Saints-Pères, Paris), Dr. Juan Ivanova (IPC, Marseille), Dr. E. Lonchamp (Unicancer, Paris), Pr. D. Malka (IGR, Villejuif), Dr. K. Litchfield (Francis Crick Institute, Londres, Royaume-Uni), Dr. R, Nicolle (CIT, Paris), Dr. A. Harlé (ICL, Nancy) et Groupe Unicancer UCGI.
Financement : PRTK 2019

Tous ces projets sont en cours ou en cours de finalisation, aussi les données seront soumises à publication fin 2020 pour l'étude ColoBEAM, courant 2021 pour l'étude OPTIMEL et fin 2021 pour MOSAPAC. Les autres résultats devraient être disponibles à l'horizon 2022.

Cancers HPV-induits

Le papillomavirus humain est un virus impliqué dans un nombre croissant de tumeurs des voies aéro-digestives supérieures (VADS) et des voies génitales. A ce jour, plus de 245 types d'HPV ont été décrits dont certains dits de haut-risque sont identifiés comme facteur de risque de cancer des VADS ou des voies génitales (Harlé *et al.*, Sci Rep, 2018). L'étude de la littérature retrouve une implication d'HPV dans 50 % des tumeurs amygdaliennes et dans une grande majorité des cancers du col. Les génotypes d'HPV les plus retrouvés dans ces cancers en Europe sont HPV16 et HPV18, à hauteur de 90 %. Les modes de transmission des HPV au niveau des VADS et plus particulièrement de la cavité buccale sont encore mal connus. Toutefois, il est admis que les femmes ayant un antécédent de cancer du col utérin, ainsi que leurs partenaires, présentent un risque plus élevé de développer un cancer de la cavité buccale.
L'ICL a initié en 2015 un programme d'études faisant appel au séquençage de nouvelle génération pour la détermination exhaustive des infections à HPV dans les cancers viro-induits aussi bien à partir de biopsies tumorales ou de biopsie liquide, la technique CatpHPV, développée en collaboration avec l'Institut Curie et le CERBA.

Le programme consacré à l'analyse viro-moléculaire des cancers HPV induits implique une partie du personnel du Département de Biopathologie de l'ICL, ainsi que des médecins oncologues ou spécialistes d'organes et des chirurgiens. Les collaborateurs sont ensuite détaillés par projet. Tous les projets sont systématiquement discutés, élaborés et suivis par la cellule data-biostatistiques et le service de promotion des essais cliniques de l'ICL.

CaptHPV (2015-2018)
Le protocole CaptHPV repose sur l'utilisation de la technique de NGS double capture « CaptHPV » mise au point à l'Institut Curie permettant la détection exhaustive de 245 génotypes différents d'HPV mais également du ou des sites d'insertion du génome viral dans les cellules infectées (Holmes *et al.*, NPJ Genom Med, 2016). Cette technique permet la caractérisation complète du génome viral et de son interaction avec la cellule cancéreuse. Cette technique fonctionne à la fois sur ADN extrait de tissus cryopréservées mais également fixés au formol et inclus en paraffine (FFPE). La technique a également été montrée suffisamment sensible pour la détection d'ADN HPV à partir de prélèvements sanguins. Cent patients atteints d'un cancer potentiellement HPV-induit (cancer de l'oropharynx, du col de l'utérus, de la vulve, du canal anal ou du pénis) ont été inclus dans le protocole CaptHPV. L'objectif de l'étude est de déterminer la concordance entre le statut HPV retrouvé dans le tissu tumoral et au niveau plasmatique, incluant la signature insertionnelle du virus. La publication est en cours de soumission.
Collaborateurs : Dr. X. Sastre-Garau (anciennement ICL, Nancy et CH, Créteil), Dr. A. Harlé (ICL, Nancy), Dr. J. Thomas (ICL, Nancy), Dr. A. Leroux (ICL, Nancy), Pr. G. Dolivet (ICL, Nancy), Dr. J. Guillet (ICL, Nancy), Dr. R. Mastronicola (ICL, Nancy), Dr. L. Leufflen (ICL, Nancy), Dr. C. Charra (ICL, Nancy), Dr. J.M. Costa (CERBA), Dr. F. Martin (CERBA).
Financement : Institut National du Cancer, CERBA, Institut de Cancérologie de Lorraine

CaptHPV vulves (2016-2018)
Cette étude repose également sur l'utilisation de la technique CaptHPV pour proposer d'établir de nouvelles connaissances sur l'implication d'HPV et son insertion dans le génome humain chez les patientes atteintes d'un cancer de la vulve. 65 patientes ont été incluses dans cette étude. Les données ont été analysées et l'article publié (Thomas *et al.*, J Low Genit Tract Dis, 2020).
Collaborateurs : Dr. X. Sastre-Garau (anciennement ICL, Nancy et CH, Créteil), Dr. A. Harlé (ICL, Nancy), Dr. J. Thomas (ICL, Nancy), V. Chesnais (Life&Soft), S. Diry (Life&Soft).
Financement : Institut National du Cancer, CERBA, Institut de Cancérologie de Lorraine

CapthHPV DK
Cette étude repose également sur l'utilisation de la technique CaptHPV pour la caractérisation de cancers gynécologiques HPV-induits de patientes de Dakar avec notamment les données de génotypes viral et les particularités d'insertions. L'intérêt principal est de pouvoir analyser des échantillons tumoraux autres que HPV16 ou HPV18 induits, la prévalence des souches HPV au Sénégal étant différentes de celles rencontrées en Europe.
Collaborateurs : Pr. M. Diop (PI, institut Joliot-Curie de l'hôpital Aristide le Dantec, Dakar, Sénégal), Dr. X. Sastre-Garau (anciennement ICL, Nancy et CH, Créteil), Dr. A. Harlé (ICL, Nancy), Dr. J. Thomas (ICL, Nancy), Dr. A. Leroux (ICL, Nancy), Dr. J.M. Costa (CERBA), Dr. F. Martin (CERBA).
Financement : Institut National du Cancer, CERBA, Institut de Cancérologie de Lorraine

Les données de ces différentes études ont donné lieu à plusieurs publications, notamment l'identification de *NFIX* comme gène préférentiel d'insertion d'HPV16 dans les cancers du canal anal (Jeannot *et al.*, Genes Chromosomes Cancer, 2018), ou encore de signatures moléculaires spécifiques d'insertion d'HPV dans les cancers de la vulve (Thomas *et al.*, J Low Genit Tract Dis, 2020). Nous avons également démontré l'intérêt de la technique CaptHPV dans le diagnostic de récidive de cancer, notamment en publiant en 2019 le cas clinique d'un patient atteint d'un cancer du canal anal, traité en 2017, puis souffrant d'un cancer de la langue en 2018 (Harlé *et al.*, BMC Cancer, 2019). L'analyse exhaustive des séquences virales d'HPV insérées dans les cellules tumorales a ainsi pu démontrer que cette nouvelle tumeur de la langue était en réalité une métastase du cancer du canal anal traité un an plus tôt, les deux tumeurs partageant toutes deux exactement la même signature insertionnelle du virus dans le génome tumoral (Figure 6).

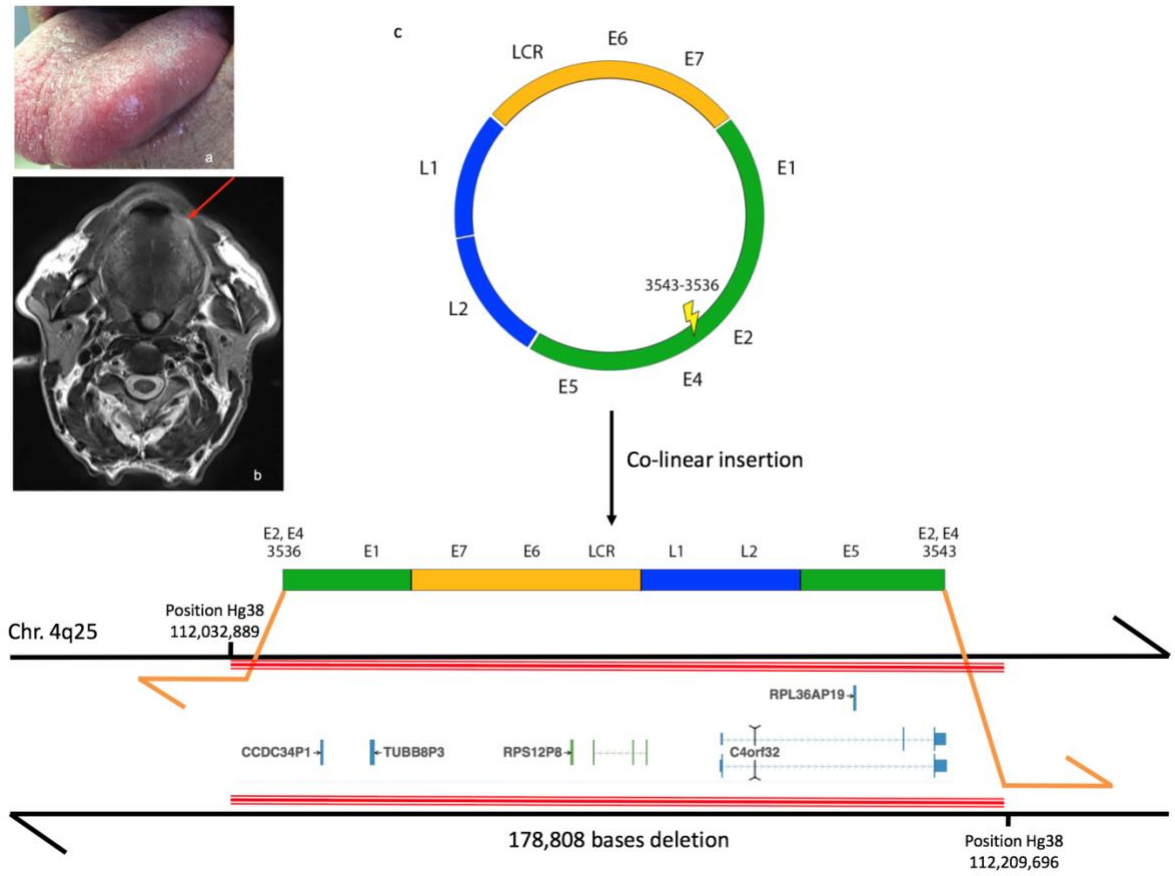


Figure 6. a) Aspect clinique du cancer localisé sur le côté gauche de la langue. b) image IRM montrant la masse de 11x11mm. c) Signature insertionnelle d'HPV16 dans le génome des cellules tumorales. Ici, l'insertion d'HPV16 retrouvée aussi bien au niveau de la tumeur du canal anal que de la langue, fait apparaitre des breakpoints communs aux deux localisations, avec une délétion des gènes indiqués entre les deux traits rouges.

Plus récemment, nous avons discuté de l'intérêt de l'étude d'HPV et son implication dans les cancers ORL dans une publication invitée dans le journal Frontiers in Oncology (Sastre-Garau *et al.*, Frontiers in Oncology, 2020).

Enfin, les résultats de l'analyse des données de l'étude CaptHPV visant à évaluer l'intérêt de la biopsie liquide dans le diagnostic de cancer invasif HPV-induit ont été soumis et l'article en cours de reviewing. L'étude des données a permis de mettre en évidence, en plus d'une caractérisation exhaustive des insertions d'HPV dans le génome des cellules tumorales, une très bonne sensibilité et une très bonne spécificité du test CaptHPV pour la détection d'ADN d'HPV dans le sang. Les données font ainsi état d'une sensibilité de 72% pour les cancers de Stade I et de plus de 94% pour les cancers de stades II à IV (Tableau 2).

Population	Nb of cases	HPV positive cases		Sensitivity	Specificity
		Tumors	Blood		
All cases	134	80	77	95.0% (76/80)	98.1% (53/54)
Cervix	73	62	60	96.8% (60/62)	100% (11/11)
Oropharynx	25	3	4	100% (3/3)	95.5% (21/22)
Oral cavity	19	2	2	100% (2/2)	100% (17/17)
Anal canal	12	12	10	83.3% (10/12)	-
Vulva	5	1	1	100% (1/1)	100% (4/4)
T1	19	11	8	72.7% (8/11)	100% (8/8)
T2	65	47	46	97.9% (46/47)	100% (18/18)
T3	30	18	19	100% (18/18)	91.7% (11/12)
T4	16	3	3	100% (3/3)	100% (13/13)
Tx	4	1	1	100% (1/1)	100% (3/3)
N0	47	29	26	89.7% (26/29)	100% (18/18)
N+	38	12	13	100% (12/12)	96.1% (25/26)
Nx	4	1	1	100% (1/1)	100% (3/3)
M0	126	79	76	94.9% (75/79)	97.9% (46/47)
M1	5	1	1	100% (1/1)	100% (4/4)
Mx	3	0	0	-	100% (3/3)

Tableau 2 : Résultats de sensibilité et de spécificité du test CaptHPV en biopsie liquide selon la localisation du cancer et le stade de la maladie.

4. Projet scientifique 2021-2025

J'envisage de continuer à développer mes programmes de recherche préclinique et translationnelle, dans le domaine de la génétique du cancer, et de l'hétérogénéité clonale, en particulier chez les patients atteints d'un cancer du pancréas. Par ailleurs, la coordination biologique du groupe de recherche « Génétique et Cancer » de l'Institut de Cancérologie de Lorraine m'a été confiée en septembre 2020. Parmi les missions attachées à cette coordination, l'accent sera mis sur l'intégration des apports de la génétique moléculaire dans la pratique quotidienne de l'oncologie médicale et la mise en place d'un continuum soin-recherche.

Cette activité permettra la constitution de collections biologiques, qui, à moyen terme, serviront à mieux documenter les données de génétique moléculaire de différents types de cancers, avec, d'une part, l'analyse exhaustive des tissus tumoraux, et d'autre part, par analyse de l'ADN tumoral libre circulant (concept de « biopsie liquide »), l'évaluation de l'hétérogénéité tumorale, ainsi que la recherche de maladie résiduelle chez les patients atteints de cancer en cours de traitement ou en cours de surveillance.

4.1. Implication de la protéine DDB2 dans les cancers du pancréas

Les travaux fondamentaux sur l'intérêt de la protéine DDB2 dans les cancers du pancréas font partie intégrante du sujet de la thèse d'Université de Madame Andréa Witz de Madame Julie Dardare dont je codirige les travaux de thèse à l'école Doctorale BioSE. Nous avons démontré au laboratoire que l'expression de DDB2 dans des lignées de cellules tumorales pancréatiques, avait un rôle dans le doublement cellulaire, la migration et l'invasion cellulaire, l'expression de marqueurs de la transition épithélio-mésenchymateuse, ainsi que sur la sensibilité à la chimiothérapie. Ces résultats questionnent sur le rôle de cette protéine au niveau de la régulation de voies de signalisation, mais également sur les applications thérapeutiques qui pourraient en découler à moyen terme (*cf.* Publications sélectionnées n˚4 et 5).

4.1.1. Exploration des voies de signalisation impliquées, voie du TGFβ et protéine DDB2

Nous souhaitons étudier le rôle de l'expression de la protéine DDB2 dans la régulation de différentes voies de signalisation. Il est prévu dans un premier temps d'explorer la voie du TGFβ et de la voie des MAPKinases compte tenu des données disponibles actuellement dans la littérature reprises dans l'article de revue rédigé par notre étudiante (Dardare *et al.*, Int J Mol Sci, 2020). Pour cela, nous commencerons par une étude *in vitro* sur les modèles cellulaires déjà établis au laboratoire (CRISPR KO et CRISPR activateur) avec une approche en RNAseq et des validations en RT-PCR. Notre hypothèse est que DDB2 pourrait jouer un rôle dans la régulation de ces deux voies de signalisation, particulièrement dans le cancer du pancréas et que cette expression viendrait modifier les

capacités de prolifération, de migration et de transition épithélio-mésenchymateuse. Il nous parait ainsi utile d'étudier les connexions potentiellement existantes entre la régulation de DDB2 et ces deux voies de signalisation.

La seconde étape sera de mettre au point une étude translationnelle rétrospective visant à évaluer le niveau d'expression de DDB2 et de ses cofacteurs, toujours au niveau des voies du TGFβ et des MAPKinases, en utilisant des échantillons tumoraux cryopréservées ou FFPE de patients atteints d'un PDAC de différents stades issus de la tumorothèque de l'Institut de Cancérologie de Lorraine, mais également des différents programmes « cancer du pancréas » dont je coordonne la partie biologique (MOSAPAC, TransPANDAS, Pancréas-CGE) et ainsi de confirmer ou non nos observations réalisées *in vitro*. Cette étude très exploratoire a pour but avant tout de mieux comprendre les différents mécanismes de régulation et la place de la protéine DDB2 dans les voies de signalisation particulièrement impliquées dans le cancer du pancréas.

4.1.2. Activation de DDB2 comme cible thérapeutique dans les cancers du pancréas

Nos observations actuelles au laboratoire nous permettent d'établir qu'une expression de la protéine DDB2 dans les cancers du pancréas serait plutôt synonyme d'un cancer au pronostic plus favorable, avec notamment un caractère moins prolifératif et moins agressif, ainsi qu'une meilleure réponse à la chimiothérapie. La question d'utiliser DDB2 comme cible thérapeutique se pose alors. Il n'est aujourd'hui techniquement pas possible chez l'homme d'administrer un traitement permettant de modifier directement l'expression d'un gène au niveau d'une tumeur. De récents travaux, notamment dans les pathologies oculaires ont utilisé la technique CRISPR-Cas9 dans le cadre d'un essai clinique et d'autres essais, notamment dans le domaine des syndromes génétiques sont en cours. Dans l'attente de pouvoir utiliser cette approche en toute sécurité chez l'homme, il nous parait plus probable et plus immédiat d'identifier les protéines responsables de la régulation de DDB2 et qui pourraient être ciblée : Il serait en effet plus simple, dans l'état actuel des connaissances, de développer un « inhibiteur de l'inhibiteur de DDB2 ». Cela implique donc d'améliorer les connaissances sur la régulation de DDB2 ou voire si DDB2 n'est finalement pas la cible finale responsable des phénomènes que nous observons, mais simplement un intermédiaire servant de simple régulateur. Nous souhaitons donc, dans le cadre de cette thématique, pousser les investigations et dans la mesure du possible, mieux comprendre la biologie de DDB2.

4.1.3. Évaluation de DDB2 comme marqueur pronostique et diagnostique des cancers du pancréas

Au vu des résultats expérimentaux obtenus au laboratoire, nous pensons que l'étude de l'expression de DDB2 pourrait potentiellement être utile en routine clinique comme marqueur pronostic et/ou prédictif de réponse aux chimiothérapies chez les patients

atteints de cancer du pancréas. Nous faisons l'hypothèse que l'absence d'expression de DDB2 est synonyme de moins bon pronostic et de moins bonne réponse à la chimiothérapie et inversement. Pour répondre à cette question, nous disposons de plusieurs cohortes de patients (TransPANDAS, MOSAPAC et Pancréas-CGE). L'étude rétrospective des résultats génomiques tumoraux mis en regard des données de survies sans-progression et globales devraient nous permettre d'affirmer ou non cette hypothèse. Si les résultats de cette étude se révèle positifs, alors il serait envisagé d'étendre l'étude de ce biomarqueur à d'autres cohortes et potentiellement en faire un biomarqueur de routine, permettant de distinguer d'une part les patients qui seront de bon pronostic et bons répondeurs à la chimiothérapie, et les patients de mauvais pronostic avec à priori une mauvaise réponse à la chimiothérapie. Il deviendrait ainsi envisageable à plus long terme d'imaginer un essai clinique prospectif avec intensification ou désintensification des doses de chimiothérapie en fonction de l'expression de la protéine DDB2.

4.2. Hétérogénéité tumorale, évolution clonale et marqueurs prédictifs de réponse aux thérapeutiques chez les patients atteints d'un cancer du pancréas

Lors de ma mobilité internationale au Francis Crick Institute de Londres dans l'équipe du Professeur Charles Swanton (juillet 2018-juilet 2019), j'ai eu l'opportunité d'élargir mes connaissances dans le domaine de l'hétérogénéité tumorale et de l'évolution clonale des tumeurs selon un modèle Darwinien. J'ai ainsi pu participer aux programmes TRACERx notamment dans les cancers bronchiques, les cancers rénaux et les mélanomes. Il n'existe pas à ce jour de programme similaire pour les patients atteints d'un cancer du pancréas. Je souhaite ainsi, en collaboration avec l'équipe du Crick Institute, développer les connaissances sur ces différents aspects dans le cancer du pancréas.

4.2.1. Exploration de l'hétérogénéité tumorale et de l'évolution clonale par l'ADN tumoral circulant

Dans le programme TRACERx, une stratégie de multi-échantillonnage tissulaire a été appliquée aussi bien au niveau du tissu primitif qu'au niveau des lésions métastatiques (Jamal-Hanjani *et al.*, PLoS Biol, 2014). Bien que très informatif sur les clones et les sous-clones présents au niveau de la tumeur primitive et des métastases, cette approche de multi-échantillonnage n'est pas applicable en routine clinique, aussi bien compte tenu du caractère invasif de la procédure de biopsie, qu'au niveau financier pour les laboratoires. Dans une étude récente, en collaboration avec le Crick Institute (Litchfield *et al.*, Cell Rep, 2020), nous avons démontré qu'il était possible d'approcher l'hétérogénéité tumorale sans avoir recours au le multi-échantillonnage tissulaire, mais en utilisant la technique dite du RepSeq (Representative Sequencing). Cette technique consiste à analyser l'intégralité des résidus opératoires, en les mixant et en extrayant l'ADN, après diagnostic par les anatomopathologistes. Ainsi, nous avons pu démontrer qu'il est possible d'identifier les populations clonales et sous-clonales en une seule analyse et ce, avec une très bonne

concordance. Dans cette étude, nous avons également analysé plusieurs prélèvements sanguins des patients ayant participé à cette étude, et environ 80% des variants détectés au niveau du tissu et a été détecté également au niveau du plasma.

Pour aller plus loin, nous développons actuellement au laboratoire une stratégie permettant d'augmenter le nombre de variants détectables au niveau plasmatique, afin de rendre aussi fidèlement possible, la composition clonale des tumeurs. C'est notamment ce que je propose dans le protocole Pancréas-CGE, pour lequel j'ai obtenu en Septembre un financement suite à l'appel à projet du Cancéropôle de l'Est. Nous allons ainsi procéder, en collaboration avec le Dr. Chris Sale (Nonacus, Birmingham, UK) au séquençage de la tumeur initiale à l'aide d'un panel de plus de 500 gènes, puis nous allons sélectionner 25 *loci* par patient, sélectionnés parmi les altérations clonales et sous clonales, pour créer un panel personnalisé par patient. Le recours à un panel de petite taille par patient devrait à priori, nous permettre d'améliorer considérablement la capacité de détection des altérations clonales et sous clonales et ainsi observer la dynamique des clones tumoraux au cours de la maladie. Pour ce programme, j'ai sollicité mon ancien collègue du Crick Institute, le Dr. Kevin Litchfield, ainsi que mes collaborateurs de l'Institut Elie Cartan de Lorraine (UMR CNRS 7502 Université de Lorraine) avec lesquels nous travaillons déjà sur la modélisation mathématique de la dynamique clonale chez les patients atteints de cancer.

4.2.2. Prédiction du retentissement antigénique des altérations somatiques et réponse à l'immunothérapie

L'immunothérapie a été une véritable révolution dans le traitement de certains cancers comme les cancers bronchiques (Carbone *et al.*, N Engl J Med, 2017) ou le mélanome (Larkin *et al.*, N Engl J Med, 2019). Les résultats sont cependant décevant dans le traitement des patients atteints d'un cancer du pancréas à l'exception de quelques cas cliniques publiés dans la littérature. En routine clinique, seule l'expression de PD-L1 est aujourd'hui utilisée, dans certaines indications, comme marqueur potentiel de réponse à l'immunothérapie et la validation de l'utilisation du TMB (Tumour Mutational Burden) a été présenté en Septembre 2020 au congrès de l'ESMO.

Il existe cependant des études publiées prometteuses, notamment par l'équipe du Crick Institute avec laquelle j'ai collaboré, démontrant que les altérations somatiques présentes au niveau tumoral et leur retentissement antigénique sont prédictives de la réponse à l'immunothérapie (Turajlic *et al.*, Lancet Oncol, 2017). Ainsi, la présence d'indels, plutôt que de SNVs au niveau du génome tumoral, est prédictive de la production d'un plus grand nombre de néoantigènes, ce qui entraine un phénotype immunogénique plus fort, plus à même de stimuler le système immunitaire, et donc permettre une meilleure réponse à l'immunothérapie. De plus, il pourrait exister un lien entre ces néoantigènes et invasion immunitaire, ce qui, là aussi permettrait de mieux identifier les patients qui pourraient bénéficier d'un traitement par immunothérapie (Rosenthal *et al.*, Nature, 2019).

Ces aspects sont inconnus aujourd'hui dans le cancer du pancréas, aussi, je souhaite explorer ce domaine de recherche chez ces patients. L'analyse génomique des tumeurs des patients inclus dans l'étude MOSAPAC, TransPANDAS et Pancréas-CGE devrait être particulièrement informative sur ce sujet. Ainsi, en collaboration avec le Dr. Jérôme Cros de l'APHP et le Dr. Kevin Litchfield du Francis Crick Institute, je souhaite également développer cette thématique au sein du laboratoire.

4.2.3. Détermination de la maladie résiduelle et prédiction précoce des rechutes chez les patients atteints de cancer

L'un des apports de l'étude de l'ADN tumoral circulant est la possibilité de tester de façon itérative les patients. En effet, il n'est ni éthique ni cliniquement possible de recourir de façon systématique à une biopsie tissulaire ou dans une moindre mesure à un examen d'imagerie pour la surveillance d'un patient atteint de cancer et présentant une réponse au traitement. La biopsie liquide est donc un moyen idéal et peu invasif, pouvant être répété mensuellement.

Il a été démontré que l'ADN tumoral circulant était le candidat idéal pour la détection de maladie résiduelle ou pour prédire une rechute, parfois jusqu'à 6 mois avant l'imagerie (Abbosh *et al.*, Nat Rev Clin Oncol, 2018). Le défi principal pour la détection de la maladie résiduelle réside en la détection de très faibles quantités d'ADN tumoral circulant dans le plasma ou dans les autres liquides biologiques. Il est ainsi estimé que la fréquence allélique maximale détectable lors de la présence d'un nodule de $1cm^3$ est de 0,008%, ce qui, à l'heure actuelle n'est pas réalisable en routine. Nous souhaitons donc pour cette raison développer l'approche technique décrite dans le paragraphe 4.2.1. de sélection de *loci* après séquençage large du tissu, afin d'améliorer au maximum la possibilité de détection de la maladie résiduelle et également prédire de manière très précoce la rechute. Nous pensons que ce nouvel outil sera utile pour les oncologues, qui pourront alors anticiper la reprise d'un traitement. Une anticipation de reprise de traitement devra être validé par un essai clinique prospectif randomisé, que je proposerai au groupe gastrointestinal d'Unicancer (UCGI) au sein duquel je suis responsable du bureau de recherche translationnel.

4.2.4. Evolution métastatique du cancer du pancréas

Les mécanismes d'évolution métastatique des cancers du pancréas sont aujourd'hui peu connus. Le pancréas étant un organe profond, la majorité des patients sont diagnostiqués à un stade métastatique, ce qui est associé à un pronostic sombre. C'est dans ce contexte que je souhaite développer un programme permettant de mieux comprendre les mécanismes moléculaires impliqués chez les patients atteints d'un cancer du pancréas et notamment le phénomène de doublement de génome (Bielski *et al.*, Nat Genet, 2018). Le doublement de génome a été décrit comme particulièrement important dans l'évolution et le pronostic des cancers avancés, mais également plus récemment dans l'accumulation d'altérations délétères au niveau du génome tumoral et l'évolution clonale (López *et al.*, Nat Genet, 2020).

La cohorte de 350 patients que nous allons étudier dans l'étude MOSAPAC permettra de mieux documenter ce phénomène, en particulier d'identifier l'intérêt du doublement de génome sur la survenue ou non d'une récidive chez les patients atteints d'un cancer du pancréas traités par chimiothérapie FOLFIRINOX ou Gemcitabine puis opérés.

Il faudra par la suite déterminer les mécanismes moléculaires mis en œuvre et dans quelle mesure ces résultats pourront être appliqués prospectivement en clinique, notamment pour la détection précoce du cancer du pancréas par des moyens non-invasifs comme la biopsie liquide.

5. Financements

Période	Titre du projet	Porteur	Co-porteur	Source	Montant
2021-2023	**Pancréas CGE** - Exploration génomique des cancers du pancréas : Évaluation de la biopsie liquide comme outil de mesure de l'évolution clonale et de l'hétérogénéité	A. Harlé		Cancéropôle de l'Est	15,000€
2021-2022	**CICLADES CE** - Biopsie liquide pour l'évaluation de l'hétérogénéité clonale et de l'évolution tumorale chez les patientes atteints d'un cancer du sein hormono-dépendant traité par anti-CDK4/6	A. Harlé		Cancéropôle de l'Est	20,000€
2020-2022	**BOVARY Pilot** - Determination of the feasibility of tumour somatic mutations detection in blood of patients with ovarian cancer	A. Harlé		Tesaro-GSK	54,000€
2020-2023	**MOSAPAC** - Signature moléculaire des adénocarcinomes du pancréas	J. Cros	A. Harlé (Coordination des analyses génomiques)	PRTK 2019	400,000€
2018-2020	**DDB2 Pancréas** – Étude de l'intérêt de la protéine DDB2 dans les cancers du pancréas	A. Harlé		CNRS CRAN	6,000€
2017-2021	**ITMO Cancer** – Modélisation mathématique de l'évolution clonale en ADN tumoral circulant	N. Champagnat	A. Harlé (Coordination des analyses génomiques)	ITMO Cancer	300,000€
2017-2020	**CICLADES** – Circulating DNA ESR1 Gene Mutations Analysis	J.L. Merlin	A. Harlé (Coordination des analyses génomiques)	Astra Zeneca	226,000€
2017-2020	**OPTIMEL** - Therapeutic Drug Monitoring of BRAF-mutated Advanced Melanoma	A. Harlé		Novartis	30,000€
2016-2020	**ColoBEAM** - Detection of KRAS, NRAS et BRAF Mutations in Plasma Circulating DNA From Patients With Metastatic Colorectal Cancer	J.L. Merlin	A. Harlé (Coordination des analyses génomiques)	Sysmex-Inostics, Merck	270,000€
2016-2020	**CaptHPV** - Validation of the Method "CaptHPV" for the Diagnosis of Human Papillomavirus Associated Infiltrating Carcinomas	X. Sastre-Garau	A. Harlé (Coordination des analyses génomiques)	Institut National du Cancer (INCa)	50,000€
2016-2020	**TransPANDAS** – Étude ancillaire PRODIGE44	J.L. Merlin	A. Harlé (Coordination des analyses génomiques)	Ligue contre le cancer 54	150,000€

6. Liste des encadrements d'étudiants

Stage d'initiation à la recherche

Date	Nom	Prénom	Direction	Co-direction	Sujet	Publication
22/02/2016	Billard	Pauline	x		Etude de l'expression de ERa36 et impact sur la signalisation cellulaire dans les cancers du sein	-
30/05/2016	Flammang	Mylène			Détermination de la sensibilité analytique de la plateforme Idylla pour la détermination des mutations BRAF sur ctDNA	-
15/06/2015	Franczak	Claire	x		Rechercher des mutations BRAF en technique Idylla	**Franczak C.**, Salleron J., Dubois C., Filhine-Trésarrieu P., Leroux A., Merlin J.L., **Harlé A**. Comparison of five different assays for the detection of BRAF mutations in formalin-fixed paraffin embedded tissues of patients with metastatic melanoma. Mol Diagn Ther. 2017 Apr;21(2):209-216.
01/12/2017	Geoffroy	Karen	x		Détermination de la sensibilité de la technique Idylla pour la détection des ctDNA dans le mélanome	Franczak C., Witz A., **Geoffroy K.**, Demange J., Rouyer M., Husson M., Massard V., Gavoille C., Lambert A., Gilson P., Gambier N., Scala-Bertola J., Merlin J.L., **Harlé A**. Evaluation of KRAS, NRAS and BRAF mutations detection in plasma using an automated system for patients with metastatic colorectal cancer. PLoS One. 2020 Jan 15;15(1):e0227294.
15/06/2015	Larcher	Lise		x	Etude des récepteurs ERa36 dans les cancers du sein	-
23/08/2011	Lion	Maëva	x		Les mutations PIK3CA dans les cancers du sein infiltrant HER2+, intérêt des techniques PCR-ARMS Scorpion et PCR-HRM	**Harlé A**., **Lion M**., Lozano N., Husson M., Harter V., Genin P., Merlin J.L. Analysis of PIK3CA exon 9 and 20 mutations in breast cancers using PCR-HRM and PCR-ARMS: correlation with clinicopathological criteria. Oncol Rep. 2013 Mar;29(3):1043-52. **Harlé A.**, **Lion M**., Lozano N., Merlin J.L. Signification clinique, diagnostique et intérêt théranostique des mutations du gène PIK3CA dans le cancer du sein. Bull Cancer. 2013 Oct;100(10):947-54.
18/05/2015	Rogerat	Laurann e	x		Détection des ADNct dans les cancers colorectaux métastatiques	-
07/10/2019	Rouzé	Valentin	x		Étude de la méthylation du prmoteur de MGMT en technique de PCR HRM	-

29/03/2017	Witz	Andréa	x		Mise au point de CIQ pour l'analyse des ctDNA dans les tumeurs solides	Franczak C., **Witz A.,** Geoffroy K., Demange J., Rouyer M., Husson M., Massard V., Gavoille C., Lambert A., Gilson P., Gambier N., Scala-Bertola J., Merlin J.L.**, Harlé A**. Evaluation of KRAS, NRAS and BRAF mutations detection in plasma using an automated system for patients with metastatic colorectal cancer. PLoS One. 2020 Jan 15;15(1):e0227294.

Licence Professionnelle

Date	Nom	Prénom	Direction	Co-direction	Sujet	Publication
28/08/2015	Cavazza	Elisa		x	Caractérisation moléculaire de carcinomes ovariens de haut-grade: vers l'identification de cibles thérapeutiques	Chen S., **Cavazza E**., Barlier C., Salleron J., Filhine-Tresarrieu P., Gavoille C., Merlin J.L., **Harlé A**. Beside P53 and PTEN: Identification of molecular alterations of the RAS/MAPK and PI3K/AKT signaling pathways in high-grade serous ovarian carcinomas to determine potential novel therapeutic targets. Oncol Lett. 2016 Nov;12(5):3264-3272.
18/08/2017	Demange	Jessica	x		Mise au point de la technique CaptHPV sur des prélèvements de cancer de la vulve	Thomas J., Leufflen L., Chesnais V., Diry S., **Demange J.**, Depardieu C., AmineBani M., Marchal F., Charra-Brunaud C., Merlin J.L., Leroux A., Sastre-Garau X., **Harlé A.** Identification of Specific Tumor Markers in Vulvar Carcinoma Through Extensive Human Papillomavirus DNA Characterization Using NGS Method. J Low Genit Tract Dis. 2020 Jan;24(1):53-60. **Harlé A**, Guillet J, Thomas J, **Demange J**, Dolivet G, Peiffert D, Leroux A, Sastre-Garau X. HPV insertional pattern as a personalized tumor marker for the optimized tumor diagnosis and follow-up of patients with HPV-associated carcinomas: A case report. BMC Cancer. 2019 Mar 28;19(1):277.
26/08/2014	Julien	Amélia		x	Optimisation et validation d'une PCR HRM pour la détection des mutations des exons 2, 3 et 4 des gènes KRAS et NRAS chez les patients atteints d'un cancer colorectal métastatique	-
16/11/2017	Meras	Morgane	x		Rôle de la protéine DDB2 dans la cancérogenèse pancréatique	Gilson P., **Meras M.**, Ariztegui M., Becuwe P., Merlin J.L., **Harlé A**. Exploring the role of the DDB2 protein as a potent molecular target in pancreatic ductal adenocarcinoma. MAP 2019, November 7th-9th 2019, London, United Kingdeom, Annals of Oncology 30 (Supplement_7), mdz413. 044
12/08/2016	Wilhelm	Cassandra	x		Détection de l'AND tumoral circulant dans les cancers colorectaux métastatiques	-

Master1

Date	Nom	Prénom	Direction	Co-direction	Sujet	Publication
31/8/2020	Coste	Mathilde	x		Intérêt de l'ADN tumoral circulant dans la prise en charge des patients atteints d'un mélanome en stade précoce ou avancé	-

Master2

Date	Nom	Prénom	Direction	Co-direction	Sujet	Publication
31/08/2020	Dardare	Julie	x		Recherche de nouvelles cibles thérapeutiques dans les adénocarcinomes canalaires pancréatiques	**Dardare J.**, Witz A., Merlin J.L., Gilson P., **Harlé A.** SMAD4 and the TGFβ pathway in patients with pancreatic ductal adenocarcinoma. Int J Mol Sci. 2020 May 16;21(10):E3534.
20/06/2019	Franczak	Claire	x		Qualification et quantification de l'ADN libre circulant extrait de plasma et de culot globulaire : optimisation de l'extraction des ADN tumoraux circulants	-
19/06/2018	Lambert	Aurélien	x		Induction de mutations BRCA pouvant être ciblées par des thérapies existantes dans les lignées cellulaires de cancer du pancréas exocrine avec la méthode CRISPR-Cas9	-
30/06/2016	Massard	Vincent		x	Etude prospective monocentrique d'optimisation technique de la détection des mutations KRAS, NRAS et BRAF par séquençage de nouvelle génération sur AND tumoral circulant	-
30/06/2017	Pflumio	Carole		x	Mise au point de la détection d'ADN tumoral circulant en séquençage haut-débit méthode de capture : Patients présentant un cancer du pancréas à la limite de la résécabilité	-
19/06/2018	Serre	Delphine	x		Mise au point de la détection d'ADN tumoral circulant en séquençage haut-débit par méthode de capture	-
30/08/2019	Witz	André	x		Mise au point de la technique crispr-cas9 pour l'induction de mutations deleteres des genes BRCA1 et BRCA2 dans des modèles cellulaires de cancer du pancréas	-

Thèse d'exercice Pharmacie

Date	Nom	Prénom	Direction	Co-direction	Sujet	Publication
10/09/2018	Becker	Natalène	x		La prise en charge du patient atteint d'un cancer de la prostate et traité par thérapie	-

					anticancéreuse orale délivrée à l'officine ou à l'hôpital	
28/11/2016	Beurrier	Mathilde	x		Diététique et nutrition chez le patient atteint de cancer : Importance de la prise en charge et rôle du pharmacien d'officine	Prix de thèse Pharmagest
25/05/2018	Filhine-Tresarieu	Pierre	x		Les ADN tumoraux circulants : rappel bibliographique et mise en place d'un pipeline bioinformatique	**Harlé A**., **Filhine-Tresarrieu P**., Husson M., Boidot R., Rouyer M., Dubois C., Leroux A., Merlin J.L. Rare RAS mutations in metastatic colorectal cancer detected during routine RAS genotyping using next generation sequencing. Target Oncol. 2016 Jun;11(3):363-70. Chen S., Cavazza E., Barlier C., Salleron J., **Filhine-Tresarrieu P**., Gavoille C., Merlin J.L., **Harlé A**. Beside P53 and PTEN: Identification of molecular alterations of the RAS/MAPK and PI3K/AKT signaling pathways in high-grade serous ovarian carcinomas to determine potential novel therapeutic targets. Oncol Lett. 2016 Nov;12(5):3264-3272. Franczak C, **Filhine-Tressarieu P**, Broséus J, Gilson P, Merlin JL, **Harlé A**. Clinical interests of circulating tumor DNA in oncology. Arch Med Res. 2018 Jul;49(5):297-305. Franczak C., **Filhine-Tressarieu P**., Gilson P., Merlin J.L., Au L., **Harlé A**. Technical considerations for circulating tumor DNA detection in oncology. Expert Rev Mol Diagn. 2019 Feb;19(2):121-135.
15/05/2020	Floquet	Shadé		x	Carcinome Epidermoïde Cutané : changement de paradigme grâce à l'immunothérapie ?	-
20/09/2019	Franczak	Claire	x		Tumeurs solides et circulation sanguine : Influence des composants sanguins dans le processus tumoral et optimisation de l'analyse des ADN tumoraux circulants	**Franczak C**, Filhine-Tressarieu P, Broséus J, Gilson P, Merlin JL, **Harlé A**. Clinical interests of circulating tumor DNA in oncology. Arch Med Res. 2018 Jul;49(5):297-305. **Franczak C**., Filhine-Tressarieu P., Gilson P., Merlin J.L., Au L., **Harlé A**. Technical considerations for circulating tumor DNA detection in oncology. Expert Rev Mol Diagn. 2019 Feb;19(2):121-135.
En cours	Gaggioli	Antoine	x		Traitement et prise en charge à l'officine des patients atteints de tumeurs solides avec fusion du gène NTRK	-
27/05/2020	Geoffroy	Karen	x		Création et utilisation de virus oncolytiques antagonisant la protéine immunosuppressive MFGE8 pour le traitement immunothérapeutique du glioblastome	-
14/06/2019	Giannelli	Anabelle	x		Le dépistage du cancer broncho-pulmonaire	-
04/09/2015	Gillotin	Emmanuelle			Les dérivés de la vitamine A : Intérêt en dermatologie et en oncologie	-

20/04/2018	Hauvuy	Jean-Hughues	x		Les effets de l'activité physique en cancérologie	-
En cours	Humbert	Geoffrey	x		Cancer bronchique non a petites cellules de stade avance et avec rearrangement alk : comment une meilleure caracterisation de la tumeur, au diagnostic comme a progression, permettra de definir une sequence therapeutique personnalisee. proposition d'un algorithme de traitement.	-
24/06/2017	Huntzinger	Cécile	x		Rôle du pharmacien d'officine dans la prise en charge des cancers de l'enfant	-
29/06/2020	Marotel	Ludovic		x	L'histoire des gaz moutarde, de la substance toxique a la substance therapeutique, exemple du cyclophosphamide	-
En cours	Moncada	Julien	x		Les nouveaux challenges de l'industrie pharmaceutique en quete de performance : Le cas de la promotion du medicament a l'ere du digital et de son pilotage en france dans le cadre d'une pathologie : le cancer colorectal	-
03/02/2020	Rogerat	Laurann e	x		Intérêt de l'ADN tumoral circulant dans la prise en charge clinique des cancers : dépistage, diagnostic, suivi de traitement et risque de récidive	-
21/09/2017	Saribekyan	Mariann a		x	Les ADNs tumoraux circulants et leur rôle dans les cancers colorectaux métastatiques : Applications en théranostique	-
28/05/2018	Thibo	Orlane	x		L'amélioration de la survie des patientes atteintes d'un cancer du sein à travers l'évolution des thérapies ciblées : de l'hormonothérapie à l'immunothérapie	-
29/05/2017	Thirion	Anaïs	x		Prise en charge des effets indésirables chez les patients atteints de cancer et traités par radiothérapie	-
08/01/2018	Vaucourt	Jérôme	x		Education thérapeutique du patient sur les anticancéreux par voie orale en hématologie à l'hôpital de Mercy : Participation à la préparation et à la mise en place d'outils éducatifs	-

Thèse d'exercice Médecine

Date	Nom	Prénom	Direction	Co-direction	Sujet	Publication
21/12/2018	Busca	Stéphane	x		Utilisation de la fréquence relative des substitutions somatiques d'un résultat de séquençage exomique comme nouvel outil diagnostique des cancers d'origine pulmonaire	**Busca S.**, Salleron J, Boidot R., Merlin J.L., **Harlé A**. Somatic substitution signature as an innovative tool in lung cancer diagnosis. Sci Rep. 2019 Oct 10;9(1):14561.

Thèse de Doctorat d'Université

Date	Nom	Prénom	Direction	Co-direction	Sujet	Publication
Novembre 2019	Witz	Andréa		x	Recherche de nouvelles cibles thérapeutiques dans les adénocarcinomes canalaires pancréatiques (PDAC).	Gilson P., Drouot G., **Witz A.**, Merlin J.L., **Harlé A**. Emerging roles of DDB2 in cancer. Int J Mol Sci. 2019 Oct 18;20(20). pii: E5168.
Novembre 2020	Dardare	Julie		x	Identification de signatures génomiques de réponse à la chimiothérapie dans les cancers du pancréas	-

7. Liste des publications et communications

Publications dans des journaux internationaux à comité de lecture

1 Sastre-Garau X., **Harlé A**. Pathology of HPV-associated Head & Neck Carcinomas: Recent Data and Perspectives for the Development of Specific Tumor Markers. Front. Oncol. 10:528957. – *Rang C, IF 4.137 – 16 points SIGAPS*

2 Gilson P., Levy J., Rouyer M., Demange J., Husson M., Bonnet C., Salleron J., Leroux A., Merlin J.L., Harlé A. Evaluation of 3 molecular-based assays for the detection of microsatellite instability in formalin-fixed paraffin embedded tissues of patients with endometrial and colorectal cancers. Sci Rep. 2020 Oct 2;10(1):16386. *– Rang B, IF 4.011 – 24 points SIGAPS*

3 Suybeng V., Koeppel F., **Harlé A.**, Rouleau E. Comparison of pathogenicity prediction tools on somatic variants. J Mol Diag, 2020, *in press - Rang B, IF 4.426 – 24 points SIGAPS*

4 Dardare J., Witz A., Merlin J.L., Gilson P., **Harlé A.** SMAD4 and the TGFβ pathway in patients with pancreatic ductal adenocarcinoma. Int J Mol Sci. 2020 May 16;21(10):E3534 - *Rang C, IF 4.183 – 16 points SIGAPS*

5 Litchfield K., Stanislaw S., Spain L., Gallegos L. Rowan A., Schnidrig, D., Rosenbaum H., **Harlé A.**, Au L., Hill S., Tippu Z., Thomasc J., Xu H., Horswell S. Barhoumi A. Jones C., Leith K., Burgess D., Watkins T., Lim E., Birkbak N., Lamy P., Nordentoft I., Dyrskjøt L., Hazell S. Jamal-Hanjani M., Larkin J., Swanton C., Alexander N., Turajlic S. Representative Sequencing: Unbiased Sampling of Solid Tumor Tissue. Cell Rep. 2020 May 5;31(5):107550 - *Rang B, IF 7.815 – 6 points SIGAPS*

6 Franczak C., Witz A., Geoffroy K., Demange J., Rouyer M., Husson M., Massard V., Gavoille C., Lambert A., Gilson P., Gambier N., Scala-Bertola J., Merlin J.L., **Harlé A.** Evaluation of KRAS, NRAS and BRAF mutations detection in plasma using an automated system for patients with metastatic colorectal cancer. PLoS One. 2020 Jan 15;15(1):e0227294 - *Rang C, IF 2.776 – 16 points SIGAPS*

7 Thomas J., Leufflen L., Chesnais V., Diry S., Demange J., Depardieu C., AmineBani M.,Marchal F., Charra-Brunaud C., Merlin J.L., Leroux A., Sastre-Garau X., **Harlé A.** Identification of Specific Tumor Markers in Vulvar Carcinoma Through Extensive Human Papillomavirus DNA Characterization Using NGS Method. J Low Genit Tract Dis. 2020 Jan;24(1):53-60 - *Rang D, IF 1.458 – 12 points SIGAPS*

8 **Harlé A.** Cell-Free DNA in the Management of Colorectal Cancer. Recent Results Cancer Res. 2020;215:253-261. – *Pas de rang, pas d'IF, 4 points SIGAPS*

9 Gilson P., Drouot G., Witz A., Merlin J.L., **Harlé A.** Emerging roles of DDB2 in cancer. Int J Mol Sci. 2019 Oct 18;20(20). pii: E5168. - *Rang C, IF 4.183 – 16 points SIGAPS*

10 Busca S., Salleron J, Boidot R., Merlin J.L., **Harlé A.** Somatic substitution signature as an innovative tool in lung cancer diagnosis. Sci Rep. 2019 Oct 10;9(1):14561 - *Rang B, IF 4.011 – 24 points SIGAPS*

11 Gilson P., Franczak C., Dubouis L., Husson M., Rouyer M., Demange J., Perceau M., Leroux A., Merlin J.L., **Harlé A.** Evaluation of KRAS, NRAS and BRAF hotspot mutations detection for patients with metastatic colorectal cancer using direct DNA pipetting in a fully-automated platform and Next-Generation Sequencing for laboratory workflow optimisation. PLoS One. 2019 Jul 2;14(7):e0219204 - *Rang C, IF 2.776 – 16 points SIGAPS*

12 Thiebaut C., Chesnel A., Merlin J.L., Chesnel M., Leroux A., **Harlé A.**, Dumond H. Dual epigenetic regulation of ERα36 expression in breast cancer cells. Int. J. Mol. Sci. 2019, 20(11), 2637 - *Rang C, IF 4.183 – 12 points SIGAPS*

13 Franczak C., Kandathil S.M., Gilson P., Husson M., Rouyer M., Demange J., Leroux A., Merlin J.L., **Harlé A.** Uncommon mutational profiles of metastatic colorectal cancer detected during routine genotyping using next generation sequencing. Sci Rep. 2019 May 8;9(1):7083 - *Rang B, IF 4.011 – 24 points SIGAPS*

14 Harlé A, Guillet J, Thomas J, Demange J, Dolivet G, Peiffert D, Leroux A, Sastre-Garau X. HPV insertional pattern as a personalized tumor marker for the optimized tumor diagnosis and follow-up of patients with HPV-associated carcinomas: A case report. BMC Cancer. 2019 Mar 28;19(1):277 - *Rang D, IF 2.933 – 12 points SIGAPS*

15 Franczak C., Dubouis L., Gilson P., Husson M., Rouyer M., Demange J., Leroux A., Merlin J.L., **Harlé A.** Integrated routine workflow using Next-Generation Sequencing and a fully-automated platform for the detection of KRAS, NRAS and BRAF mutations in formalin-fixed paraffin embedded samples with poor DNA quality in patients with colorectal carcinoma. PLoS One. 2019 Feb 27;14(2):e0212801 - *Rang C, IF 2.776 – 16 points SIGAPS*

16 Franczak C., Filhine-Tressarieu P., Gilson P., Merlin J.L., Au L., **Harlé A.** Technical considerations for circulating tumor DNA detection in oncology. Expert Rev Mol Diagn. 2019 Feb;19(2):121-135 - *Rang C, IF 3.099 – 16 points SIGAPS*

17 Lambert A, Salleron J, Lion M, Rouyer M, Lozano N, Leroux A, Merlin JL, **Harlé A.** Comparison of three real-time PCR assays for the detection of PIK3CA somatic mutations in formalin-fixed paraffin embedded tissues of patients with breast carcinomas. Pathol Oncol Res. 2018 Nov 13 - *Rang D, IF 2.433 – 12 points SIGAPS*

18 Franczak C, Filhine-Tressarieu P, Broséus J, Gilson P, Merlin JL, **Harlé A.** Clinical interests of circulating tumor DNA in oncology. Arch Med Res. 2018 Jul;49(5):297-305 - *Rang D, IF 1.895 – 12 points SIGAPS*

19 Jeannot E., **Harlé A.**, Holnes A., Sastre-Garau X. Nuclear factor I X is a recurrent target for HPV16 insertions in anal carcinomas. Genes Chromosomes Cancer. 2018 Dec;57(12):638-644 - *Rang C, IF 2,490 – 12 points SIGAPS*

20 Vogin G., Merlin J.L., Rousseau A., Peiffert D., **Harlé A.**, Husson M., El Hajj L., Levitchi M., Simon T., Simon J.M. Absence of correlation between radiation-induced CD8 T-lymphocyte apoptosis and sequelae in patients with prostate cancer accidentally overexposed to radiation. Oncotarget. 2018 Aug 24;9(66):32680-32689 – *Pas de rang, pas d'IF, 4 poins SIGAPS*

21 Serre D., Salleron J., Husson M., Leroux A., Gilson P., J.L. Merlin, Geoffrois L., **Harlé A.** Accelerated BRAF mutation analysis using a fully automated PCR platform improves the management of patients with metastatic melanoma. Oncotarget. 2018 Aug 14;9(63):32232-32237 - *Pas de rang, pas d'IF, 4 poins SIGAPS*

22 **Harlé A.**, Guillet J., Thomas J., Sastre-Garau X., Rouyer M., Ramacci C., Gilson P., Dubois C., Dolivet G., Leroux A., Salleron J., Merlin J.L. Evaluation and validation of HPV real-time PCR assay for the detection of HPV DNA in oral cytobrush and FFPE samples. Sci Rep. 2018 Jul 27;8(1):11313 – *Rang B, IF 4.011 – 24 points SIGAPS*

23 **Harlé A.**, Franczak C. Authors' Reply to Uguen: "Comparison of Five Different Assays for the Detection of BRAF Mutations in Formalin-Fixed Paraffin Embedded Tissues of Patients with Metastatic Melanoma". Mol Diagn Ther. 2017 Dec;21(6):695-696 – *Rang C, IF 3,058 – 16 points SIGAPS*

24 Franczak C., Salleron J., Dubois C., Filhine-Trésarrieu P., Leroux A., Merlin J.L., **Harlé A.** Comparison of five different assays for the detection of BRAF mutations in formalin-fixed paraffin embedded tissues of patients with metastatic melanoma. Mol Diagn Ther. 2017 Apr;21(2):209-216 – *Rang C, IF 3,058 – 16 points SIGAPS*

25 Chen S., Cavazza E., Barlier C., Salleron J., Filhine-Tresarrieu P., Gavoille C., Merlin J.L., **Harlé A.** Beside P53 and PTEN: Identification of molecular alterations of the RAS/MAPK and PI3K/AKT signaling pathways in high-grade serous ovarian carcinomas to determine potential novel therapeutic targets. Oncol Lett. 2016 Nov;12(5):3264-3272 – *Rang E, IF 1,871 – 8 points SIGAPS*

26 Lion M., **Harlé A.**, Salleron J., Ramacci C., Campone M., Merlin J.L. Trastuzumab as pre-operative monotherapy does not inhibit HER 2 downstream signaling in HER 2 positive breast cancer. Oncol Lett. 2016 Sep;12(3):2028-2032 – *Rang E, IF 1,871 – 6 points SIGAPS*

27 Rauch P, Merlin JL, Leufflen L, Salleron J, Harlé A, Olivier P, Marchal F. Limited effectiveness of patent blue dye in addition to isotope scanning for identification of sentinel lymph nodes: Cross-sectional real-life study in 1024 breast cancer patients. Int J Surg. 2016 Aug 5. pii: S1743-9191(16)30285-0 – *Rang C, IF 3,158 – 4 points SIGAPS*

28 **Harlé A.**, Salleron J., Franczak C., Dubois C., Filhine-Tressarieu P., Leroux A., Merlin J.L. Detection of BRAF mutations using a fully automated platform and comparison with High Resolution Melting, Real-time allele specific amplification, immunohistochemistry and Next Generation Sequencing assays, for patients with metastatic melanoma. PLoS One. 2016 Apr 25;11(4):e0153576 0 – *Rang C, IF 2,776 – 16 points SIGAPS*

29 Pape E., Michel D.,Scala-Bertola J.,Schiestel T., **Harlé A.**,Bouchet S., Contet A.,Pochon C., Gambier N. Effect of esomeprazole on the oral absorption of dasatinib in a patient with Philadelphia-positive acute lymphoblastic leukemia. Br J Clin Pharmacol. 2016 Jun;81(6):1195-1196 – *Rang B, IF 3,867 –6 points SIGAPS*

30 **Harlé A.**, Filhine-Tresarrieu P., Husson M., Boidot R., Rouyer M., Dubois C., Leroux A., Merlin J.L. Rare RAS mutations in metastatic colorectal cancer detected during routine RAS genotyping using next generation sequencing. Target Oncol. 2016 Jun;11(3):363-70 – *Rang C, IF 3,683 – 16 points SIGAPS*

31 **Harlé A.**, Salleron J., Perkins G., Pilati C., Blons H., Laurent-Puig P., Merlin J.L. Expression of pEGFR and pAKT as response-predictive biomarkers for RAS wild-type patients to anti-EGFR monoclonal antibodies in metastatic colorectal cancers. Br J Cancer. 2015 Aug 11;113(4):680-5 – *Rang B, IF 5,416 – 24 points SIGAPS*

32 Merlin J.L., **Harlé A.**, Lion M., Ramacci C., Leroux A. Expression and activation of P38-MAP kinase in invasive ductal breast cancers: relation with estrogen receptor, HER2 expression and downstream signaling phosphorylated proteins expression. Oncol Rep. 2013 Oct;30(4):1943-8 – *Rang D, IF 3,041 – 9 points SIGAPS*

33 Chretien A.S., **Harlé A.**, Meyer-Lefebvre M., Rouyer M., Husson M., Ramacci C., Leroux A., Genin P., Merlin J.L. Optimization of routine KRAS mutation testing procedure for rational individualized first line targeted therapy selection in metastatic colorectal cancer. Cancer Med. 2013 Feb;2(1):11-20 – *Rang D, IF 3,357 – 9 points SIGAPS*

34 **Harlé A.**, Lion M., Lozano N., Husson M., Harter V., Genin P., Merlin J.L. Analysis of PIK3CA exon 9 and 20 mutations in breast cancers using PCR-HRM and PCR-ARMS: correlation with clinicopathological criteria. Oncol Rep. 2013 Mar;29(3):1043-52 – *Rang D, IF 3,041 – 12 points SIGAPS*

35 **Harlé A.**, Rouyer M., Harter V., Genin P., Leroux A., Merlin J.L. Comparison of COBAS 4800 KRAS, TaqMan PCR and High-Resolution Melting PCR assays for the detection of KRAS somatic mutations in formalin-fixed paraffin embedded colorectal tumors. Virchows Arch. 2013 Mar;462(3):329-35 – *Rang C, IF 2,868 – 16 points SIGAPS*

Publications dans des journaux nationaux à comité de lecture

1 Massard V., **Harlé A.**, Uwer L., Merlin J.L. ESR1 Mutations: From Benchside to Bedside. Oncologie (2019) 21:29-32 – *Rang E, IF 0,043 – 6 points SIGAPS*

2 Loriot MA, Ciccolini J, Thomas F, Barin-Le-Guellec C, Royer B, Milano G, Picard N, Becquemont L, Verstuyft C, Narjoz C, Schmitt A, Bobin-Dubigeon C, Harle A, Paci A, Poinsignon V, Quaranta S, Evrard A, Hennart B, Broly F, Fonrose X, Lafay-Chebassier C, Wozny AS, Masskouri F, Boyer JC, Etienne-Grimaldi MC. Dihydropyrimidine dehydrogenase (DPD) deficiency screening and securing of fluoropyrimidine-based chemotherapies: Update and recommendations of the French GPCO-Unicancer and RNPGx networks. Bull Cancer. 2018 Feb 24. pii: S0007-4551(18)30053-5 – *Rang E, IF 0,729 – 2 points SIGAPS*

3 **Harlé A.**, Merlin J.L. Biomarqueurs prédictifs de réponse aux thérapies ciblées en oncologie. Ann Biol Clin (Paris). 2013 Nov;71:89-97 – *Rang E, IF 0,503 – 8 points SIGAPS*

4 **Harlé A.**, Lion M., Lozano N., Merlin J.L. Signification clinique, diagnostique et intérêt théranostique des mutations du gène PIK3CA dans le cancer du sein. Bull Cancer. 2013 Oct;100(10):947-54 – *Rang E, IF 0,729 – 2 points SIGAPS – Rang E, IF 0,729 – 8 points SIGAPS*

5 **Harlé A.**, Lion M., Husson M., Dubois C., Merlin J.L. Validation de méthode selon la norme ISO 15189 et le SH GTA 04 : Application à l'extraction et au dosage quantitatif de l'ADN par méthode spectrophotométrique. Ann Biol Clin (Paris). 2013 Sep-Oct;71(5):608-14 – *Rang E, IF 0,503 – 8 points SIGAPS*

6 **Harlé A.**, Dubois C., Rouyer M., Merlin J.L. Validation de méthode selon la norme ISO 15189 et le SH GTA 04 : Application à la détection des mutations KRAS par méthode TaqMan. Ann Biol Clin (Paris). 2013 Sep-Oct;71(5):603-7 – *Rang E, IF 0,503 – 8 points SIGAPS*

7 **Harlé A.**, Chapelle D., Merlin J.L. Rôle du pharmacien d'officine dans la prise en charge des patients atteints de tumeurs solides et traités par un inhibiteur de kinase. Journal de Pharmacie Clinique. 2013 ;32(2):71-8 – Pas de rang, pas d'IF, non référencé SIGAPS.

Communications orales invitées

1 A NGS-based blood test for the diagnosis of invasive HPV-associated carcinomas with extensive viral genomic characterizations. Genomics Live, 10 au 12 Novembre 2020, London, United-Kingdom (online).
2 Microsatellite instability (MSI) status assessment in patients with colorectal and endometrial cancers. BioRad webinar, 5 Novembre 2020.
3 A NGS-based blood test for the diagnosis of invasive HPV-associated carcinomas with extensive viral genomic characterizations. 3rd International Liquid Biopsy Symposium, 30 Octobre 2020, Freiburg, Germany (online).
4 Adoption of broad NGS profiling for HRR Deficiency: Looking beyond BRCA, SNPs and indels. Genomeweb, 22 Octobre 2020.
5 Anti-PARP et signatures de deficit en recombinaison homologue. OncoStream, 8 Octobre 2020.
6 Somatic substitution signature s an innovative tool in lung cancer diagnosis. Journée R2CL, 17 Janvier 2020, Nancy, France
7 How to integrate Liquid Biopsies into Routine care. Genomics Live, 4th and 5th December 2019, Basel, Switzerland.
8 Mode d'action des anti-PARP. Symposium Tesaro, 14 Novembre 2019, Nancy, France.
9 Mode d'action des anti-PARP. Symposium Tesaro, 23 Octobre 2019, Reims, France
10 ADN tumoral circulant : Applications cliniques et hétérogénéité tumorale. Modélisation de l'hétérogénéité tumorale et thérapies ciblées, 21 Octobre 2019, Nancy, France.
11 Marqueurs de réponse à l'immunothérapie dans les cancers des voies aérodigestives supérieures. Symposium BMS, 11 Octobre 2019, Dijon, France.
12 Turning ctDNA in routine practice. Liquid biopsy congress, 10 Octobre 2019, London, United-Kingdom
13 Translocations ALK et recherche de mutations de résistance aux anti-ALK en biopsie liquide. Symposium Merck-Pfizer, 2 Avril 2019, Nancy, France.
14 Détection des mutations des gènes KRAS, NRAS et BRAF chez les patients atteints d'un cancer colorectal métastatique. Journée de formation Merck Afrique du Nord, 28 Mars 2019, Alger, Algérie
15 Détection des mutations des gènes KRAS, NRAS et BRAF chez les patients atteints d'un cancer colorectal métastatique. Journée de formation Merck Afrique du Nord, 27 Mars 2019, Oran, Algérie
16 Détection des mutations des gènes KRAS, NRAS et BRAF chez les patients atteints d'un cancer colorectal métastatique. Journée de formation Merck Afrique du Nord, 26 Mars 2019, Constantine, Algérie
17 Detection of somatic alterations in FFPE and plasma samples of patients with solid tumours using capture-based sequencing. Workshop Sophia Genetics, 13 Novembre 2018, Bruxelles, Belgique
18 L'ADN tumoral circulant outil d'évaluation de l'hétérogénéité tumorale et de la maladie résiduelle . Médecine de précision et biopsise liquides Roche Diagnostics, 12 Octobre 2018, Paris, France.
19 Evaluation de la plateforme Idylla pour la détection des mutations des gènes KRAS, NRAS et BRAF dans le plasma de patients atteints d'un mCRC. User meeting Biocartis, 4 Octobre 2018, Paris, France.
20 Place de l'ADN circulant en cancérologie digestive. Séminaire de cancérologie digestive, Réseau de cancérologie ONCOLOR, 20 Octobre 2017, Vittel, France.
21 Chair de la session "Resistance to treatment, clonal evolution". Tumor Liquid Biopsy Symposium, 26 et 27 Septembre 2017, Nancy, France.
22 NGS for cfDNA detection: Let's talk about sensitivity. Tumor Liquid Biopsy Symposium, 26 et 27 Septembre 2017, Nancy, France.
23 Biopsies liquides et cancer du pancras : vers un dépistage ? Actualités en cancérologie digestive, Institut de Cancérologie de Lorraine., 30 Mars 2017, Nancy, France.
24 Aspects pré-analytiques en biologie moléculaire. Atelier de renforcement des capacités des pays francophones de la région Africaine de l'OMS dans la lutte contre le cancer, 14 au 16 Mars 2017, Abidjan,

Côte d'Ivoire.

25 Technologie BEAMing pour l'analyse de l'ADN tumoral circulant : Point de vue du biologiste. 4ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie - OncoTrans 2015, 25 et 26 Juin 2015, Dijon, France.

26 Biomarqueurs et angiogenèse. Approches innovantes en oncologie thoracique. Cancéropôle Grand-Est, 14 et 15 Novembre 2014, Strasbourg, France.

27 Actualités en cancérologie digestive. Institut de Cancérologie de Lorraine, 2 Octobre 2014, Nancy, France.

28 Molecular Biomarker Testing for RAS mutations in mCRC. Roche Molecular mCRC preceptorship, 22 et 23 Septembre 2014, Paris, France.

29 Molecular biomarker testing: Case studies. Roche Molecular mCRC preceptorship, 23 et 23 Septembre 2014, Paris, France.

30 Génétique somatique en oncologie. 2ème Journée du Diagnostic Moléculaire, 18 Octobre 2013, Paris, France.

31 Nouveaux aspects réglementaires et points critiques dans l'accréditation COFRAC d'un laboratoire de Biopathologie - Expérience du CLCC de Nancy. Réunion du GPCO, 14 Juin 2013, Paris, France.

32 Prélèvements en recherche clinique, objectifs et intérêts. Rencontres scientifiques de l'Institut de Cancérologie de Lorraine, 22 Février 2013, Nancy, France.

33 Validation de méthode selon la norme ISO 15189. Réunion de restitution des CQ Nationaux BRAF, 17 Janvier 2013, Paris, France.

34 Biothérapies – Biosimilaires. Quels défis ? 4èmes échanges régionaux en Oncologie Amgen, 15 Juin 2012, Lyon, France.

35 Physiopathologie du PTI. Séminaire de formation UTIP Pharmaciens d'officine, 22 Novembre 2012, Nancy, France.

36 Démarche d'accréditation ISO 15189 et résultats enquête de satisfaction plateforme KRAS. Symposium Merck, 12 Décembre 2012, Nancy, France.

Communications orales

1 Gilson P., Meras M., Becuwe P., Merlin J.L., **Harlé A.** Rôles de la protéine DDB2 dans la cancérogenèse des adénocarcinomes pancréatiques exocrines. 5ème journée d'échanges en cancérologie en Lorraine, 21 Janvier 2019, Nancy, France.

2 Gilson P., Serre D., Salleron J., Husson M., Leroux A., Merlin J.L., Geoffrois L., **Harlé A.** Diagnostic ciblé rapide des mutations du gène BRAF par une plateforme automatisée de PCR en temps réel pour améliorer la prise en charge des patients atteints de mélanome métastatique. 5ème journée d'échanges en cancérologie en Lorraine, 21 Janvier 2019, Nancy, France.

3 Franczak C., Dubouis L., Witz A., Geoffroy K., Flammang M., Rogerat L., Demange J., Rouyer M., Husson M., Serre D., Massard V., Gavoille C., Lambert |A., Gilson P., Gambier N., Scala-Bertola J., Merlin J.L., **Harlé A.** Évaluation d'une plateforme automatisée de PCR en temps réel pour la détection des mutations sur les gènes KRAS, NRAS et BRAF dans le plasma de patients atteints d'un cancer colorectal métastatique. 5ème journée d'échanges en cancérologie en Lorraine, 21 Janvier 2019, Nancy, France

4 Franczak C., Geoffroy K., Demange J., Rouyer M., Husson M., Lacomme S., Gauchotte G., Gilson P., **Harlé A.**, Vignaud J.M., Leroux A., Merlin J.L. Évaluation d'une plateforme automatisée de PCR en temps réel pour la détection des mutations du gène EGFR chez les patients atteints d'un cancer bronchique non à petites cellules. 5ème journée d'échanges en cancérologie en Lorraine, 21 Janvier 2019, Nancy, France

5 **Harlé A.**, Champagnat N. Utilisation de l'ADN tumoral circulant pour la modélisation dynamique des résistances aux thérapies ciblées. 1ère Journée autour des mathématiques et de l'informatique en analyse de données et d'imagerie en oncologie., 12 Juin 2018, Nancy, France

6 Thomas J., Depardieu C., Leroux A., Leufflen L., Charra C., Demange J., Ramacci C., Merlin J.L., Sastre-Garau X., **Harlé A.** Cancers de la vulve : aspects anatomocliniques, virologiques et immunophénotypiques. 4ème journée d'échanges en cancérologie en Lorraine, 17 Novembre 2017, Nancy, France.

7 **Harlé A.**, Serre D., Salleron J., Husson M., Leroux A., Geoffrois A., Merlin J.L. Détection rapide des mutations BRAF en système automatisé et amélioration de la prise en charge des patients atteints d'un mélanome métastatique, résultats de l'étude MELFAST. 5ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie - OncoTrans 2017, 2 et 3 Février 2017, Besançon, France.

8 Flammang M., Salleron J., Franczak C., Dubois C., Filhine-Trésarrieu P., Leroux A., Merlin J.L., **Harlé A.** Détection des mutations BRAF en théranostique des mélanomes métastatiques : Utilisation de la plateforme IdyllaTM pour l'analyse de tissus FFPE et de plasma. Carrefour de Pathologie, 7 Novembre 2016, Paris, France.

9 Thiébaut C., Morel C., **Harlé A.**, Chesnel A., Leroux A., Chamard-Jovenin C., Boukhobza T., Merlin J.L., Dumond H. From ERα66 to ERα36: a new predictive marker for cancer progression and therapeutic response in breast tumors ?. 21st World Congress on Advances in Oncology and 19th International Symposium on Molecular Medicine, Octobre 2016, Athènes, Grèce.

10 **Harlé A.**, Dubois C., Ramacci C., Depardieu C., Charra-Brunaud C., Peiffert D., Ameloot-Yaribakht S., Leufflen L., Marchal F., Dolivet G., Salleron J., Fernandes L., Gillon V., Leroux A., Merlin J.L., Sastre-Garau X. Diagnostic sérique des cancers HPV-associés par capture et séquençage de l'AND viral circulant. 3ème journée d'échanges en cancérologie en Lorraine, 14 Octobre 2016, Nancy, France.

11 Chen S., Cavazza E., Barlier C., Filhine-Tressarieu P., Gavoille C., Merlin J.L., **Harlé A.** Investigation of druggable molecular targets in high grade ovarian tumors using Next Generation Sequencing. Forum du Cancéropôle Grand-Est, 12 et 13 Novembre 2015, Strasbourg, France.

12 **Harlé A.**, Filhine-Tressarieu P., Rouyer M., Massard V., Boidot R., Merlin J.L. Detection of KRAS, NRAS, BRAF, PIK3CA and EGFR somatic mutations in circulating tumor DNA using Next Generation Sequencing (NGS) for patients with metastatic colorectal cancer. 2ème Journée de la recherché de l'Institut de Cancérologie de Lorraine, 6 Novembre 2015, Nancy, France.

13 **Harlé A.**, Boidot R., Rouyer M., Chevrier C., Massard V., Ghiringhelli F., Merlin J.L. Détection des mutations somatiques de KRAS, NRAS et BRAF en NGS : Application à l'ADN tumoral circulant. 4ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie - OncoTrans 2015, 25 & 26 Juin 2015, Dijon, France.

14 **Harlé A.** Génotypage des mutations KRAS, NRAS, BRAF et PIK3CA a partir d'échantillons de tumeurs FFPE par technique de séquençage bidirectionnel haut-débit. Journée d'Echanges sur la Recherche en Cancérologie en Lorraine. La personnalisation du diagnostic et des traitements, 26 Septembre 2014, Nancy, France.

15 Merlin J.L., **Harlé A.** Génétique moléculaire et fonctionnalité de la signalisation tumorale appliquées au diagnostic à la recherche de marqueurs prédictifs de réponse. Journée d'Echanges sur la Recherche en Cancérologie en Lorraine. La personnalisation du diagnostic et des traitements, 26 Septembre 2014, Nancy, France.

16 **Harlé A.**, Guillet J., Rouyer M., Merlin J.L. Development and validation of a real-time PCR assay using cobas® for the detection and genotyping of human papillomavirus in cervical or buccal smears and in formalin-fixed embedded head and neck tumors. 3ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie OncoTrans 2013, 27 et 28 Juin 2013, Reims, France.

17 Guillet J., **Harlé A.**, Bevilacqua S., Dolivet G., Merlin J.L. Détection des HPV sur des frottis en phase liquide et des prélèvements solides inclus en paraffine par une nouvelle méthode basée sur la RT-PCR. 8ème journée de la Recherche Biomédicale, 24 Mai 2013, Nancy, France.

18 Merlin J.L., **Harlé A.** Molecular diagnosis, response predictive biomarkers and therapeutic drug monitoring for personalized therapy with oral kinase inhibitors in oncology. 8ème Congrès de Physiologie, de Pharmacologie et de Thérapeutique, 22 au 24 Avril 2013, Angers, France.

19 Chrétien A.S., **Harlé A.**, Meyer-Lefebvre M., Rouyer M., Husson M., Ramacci C., Genin P., Leroux A., Merlin J.L. Analyse des mutations KRAS dans les cancers colorectaux métastatiques : Optimisation des techniques et des délais pour la sélection rationnelle d'une première ligne thérapeutique. Carrefour de Pathologie, 21

au 25 Novembre 2011, Paris, France.

20 **Harlé A.** Préventions des infections nosocomiales chez le neutropénique. Rencontres d'hématologie, 2009, Nancy, France.

21 Merlin J.L., Ramacci C., Rouyer M., Giacommetti G., **Harlé A.**, Chergui F., Dolivet G., Leroux A. Phosphoprotein array for molecular diagnosis in breast and head and neck squamous cell carcinoma. 1er Forum du Cancéropôle Grand Est, 19 Octobre 2007, Vittel, France.

Communications affichées

1 Merlin J.L., Gilson P., Husson M., **Harlé A.** Targeted PCR vs NGS for molecular diagnostic in solid tumors and liquid biopsies. How to choose in real-life. ASCO 2020 annual meeting, May 29th-31st 2020, Chicago, IL, USA, J Clin Oncol 38 (15_suppl), e15576-e15576

2 Gilson P., Meras M., Ariztegui M., Becuwe P., Merlin J.L., **Harlé A.** Exploring the role of the DDB2 protein as a potent molecular target in pancreatic ductal adenocarcinoma. MAP 2019, November 7th-9th 2019, London, United Kingdeom, Annals of Oncology 30 (Supplement_7), mdz413. 044

3 Dufraing K., Garrec C., Egele C., Keppens C., Chiesa J., Lamy A., Lacroix L., **Harlé A.**, Pascal R., Van Casteren K., Fetique D., Dequeker E., Bellocq J.P., Rouleau E. Two years of BRCA1 and BRCA2 somatic external quality assessment with Gen&Tiss scheme in France. ESMO 2019 congress, September 29th - October 1st 2019, Barcelona, Spain

4 Dufraing K., Leroy K., Denis M., **Harlé A.**, Lacroix L., Lamy A., Vanwelden K., Dequeker E., Rouleau E. First national External Quality Assessement for the interpretation of somatic variants: assessment of 25 variants in colorectal, lung, ovarian cancers and melanoma in France. ESMO 2019 congress, September 29th - October 1st 2019, Barcelona, Spain

5 **Harlé A.**, Gavoille C., Bouché O., Ben Abdelghani M., Plaza J., Lambert A,, Spaeth D., Boudrant A., Ghiringhelli F., Villing A.L., Borg C., Rouyer M., Husson M., Gilson P., Polet F., Jones F.S., Edelstein D., Baumard F., Salleron J., Merlin J.L. cfDNA for accurate determination of RAS and BRAF mutations using OncoBEAM liquid biopsy in metastatic colorectal cancer patients: Results of the real-world multicentric ColoBEAM study. ASCO 2019 annual meeting, May 31st - June 4th 2019, Chicago, IL, USA. J Clin Oncol 37, 2019 (suppl; abstr 3542)

6 Au L., Horswell S., **Harlé A.**, Xu H., Rowan A., Hubank M., Litchfield K., Nicol D., O'brien T., Chowbhury S., Rudman S., Gore M., Larkin J., Swanton C., Turajlic S., and the TRACERx Renal Consortium Cell-free DNA profiles reflect clonal and sub-clonal lineages of clear-cell renal cell carcinoma and tracks clinical trajectory: TRACERx Renal. 14th European International Kidney Cancer Symposium, 29-30 March 2019, Valamar Lacroma, Dubrovnik, Croatia

7 Massard V., Uwer L., Salleron J., Deblock M., Kieffer A., Rios M., Gilson P., Lesur A., **Harlé A.**, Merlin J.L. CICLADES: Monitoring of ESR1, PIK3CA and AKT1 ctDNA mutations during real-life follow-up of patients with advanced breast cancer treated with endocrine therapy. SABCS, 4 au 8 Décembre 2018, San Antonio, TX, USA. Cancer Research 79 (4 Supplement), OT1-03-02-OT1-03-02

8 **Harlé A.**, Dietmaier W., Vogl I., Neumann K., Haumaier F., Beggs A., Pestinger V., Grand D., Tallet A., Merlin J.L. Detection of ALK, RET, ROS1, NTRK1 and MET rearrangements and actionable mutations using Next Generation Sequencing in patients with NonSmall Cell Lung Cancer. MAP2018, 14 et 15 Septembre 2018, Paris, France. Annals of Oncology, Volume 29, Issue suppl_6, 1 September 2018, mdy318.017,

9 Massard V., **Harlé A.**, Salleron J., Uwer L., Lesur A., Merlin J.L. Suivi des mutations des gènes ESR1, PIK3CA et AKT1 sur ADN tumoral circulant en vie réelle chez des patientes présentant un cancer du sein avancé en cours d'hormonothérapie - Etude en cours. SFSPM, 8 au 10 Novembre 2017, Lille, France

10 **Harlé A.**, Gavoille C., Massard V., Clément-Duchêne C., Sastre-Garau X., Champagnat N., Geoffrois L., Merlin J.L. Ongoing cfDNA-based clinical trials at Institut de Cancérologie de Lorraine. Tumor Liquid Biopsy Symposium, 25 Septembre 2017, Nancy, France.

11 Thiébaut C., Chamard-Jovenin C., Chesnel A., Morel C., Grillier-Vuissoz I, **Harlé A.**, Boukhobza T., Merlin J.L. ERα36 expression: a key breast cancer risk factor? 2nd EACR-OECI Conference, 13 au 16 Mars 2017, Amsterdam, Pays-Bas

12 Merlin J.L., Rauch P., Leufflen L., Salleron J., **Harlé A.**, Olivier P., Marchal F. Limited effectiveness of patent blue dye in addition to isotope scanning for identification of sentinel lymph nodes: Cross-sectional real-life study 1024 breast cancer patients. San Antonio Breast Cancer Symposium, 6 au 10 Décembre 2016, San Antonio, TX, USA.

13 Chen S., Cavazza E., Barlier C., Filhine-Tresarrieu P., Gavoille C., **Harlé A.**, Merlin J.L. Identification of druggable targets in high-grade epithelial ovarian cancer (EOC) using next generation sequencing (NGS)–based molecular diagnostic. ASCO 2016 annual meeting, 3 au 7 Juin 2016, Chicago, IL, USA (abstr e17076)

14 Morel C., **Harlé A.**, Chesnel A., Chamard-Chovenin C., Jung A., Abecassis J., Merlin J.L., Dumond H. From Erα66 to Erα36: A new marker for cancer progression and therapeutic response in breast tumors?. Forum du Cancéropôle Grand-Est, 13 et 14 Novembre 2015, Strasbourg, France.

15 Morel C., **Harlé A.**, Chesnel A., Chamard-Chovenin C., Jung A., Abecassis J., Merlin J.L., Dumond H. From Erα66 to Erα36: A new marker for cancer progression and therapeutic response in breast tumors? Journée de la recherché de l'Institut de Cancérologie de Lorraine, 6 Novembre 2015, Nancy, France.

16 **Harlé A.**, Husson M., Rouyer M., Leroux A., Merlin J.L. Detection of rare somatic mutational profiles in metastatic colorectal cancer during routine RAS sequencing using next generation sequencing. 4ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie - OncoTrans 2015, 25 et 26 Juin 2015, Dijon, France. Prix de la meilleure communication poster.

17 Merlin J.L., Husson M., Rouyer M., Leroux A., **Harlé A.** Detection of rare somatic mutational profiles in metastatic colorectal cancer (mCRC) during routine RAS sequencing using next generation sequencing (NGS). ASCO 2015 annual meeting, 28 Mai au 2 Juin 2015, Chicago, IL, USA. J Clin Oncol 33 : 2015 suppl abstr 11055

18 Harlé A, Husson M, Rouyer M, Leroux A, Merlin JL. Testing of KRAS, NRAS, BRAF and PIK3CA somatic mutations on FFPE tumors samples using bidirectionnal ultra-deep pyrosequencing. AACR annual meeting., 5 au 9 Avril 2014, San Diego, CA, USA

19 **Harlé A.**, Lion M., Rouyer R., Lozano N., Merlin J.L. Comparative evaluation of three real-time PCR assays for the detection of PIK3CA somatic mutations in formalin-fixed paraffin embedded breast carcinomas. AACR Molecular Targets and Cancer Therapeutics, 19 au 23 Octobre 2013, Boston, MA, USA.

20 Merlin J.L., **Harlé A.**, Lion M., Chrétien A.S., Ramacci C., Leroux A. Activation of P38MAP kinase in estrogen positive invasive breast cancers: Relation with HER2 and downstream signaling phosphorylated proteins expression. 3ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie OncoTrans 2013, 27 et 28 Juin 2013, Nancy, France.

21 Merlin J.L., **Harlé A.**, Lion M., Chrétien A.S., Ramacci C., Leroux A. Activation of P38MAP kinase in estrogen positive invasive breast cancers: Relation with HER2 and downstream signaling phosphorylated proteins expression. ASCO annual meeting 2013, 31 Mai au 4 Juin 2013, Chicago, IL, USA. J Clin Oncol 31, 2013 (suppl; abstr e11560)

22 Merlin J.L., **Harlé A.**, Lion M., Chrétien A.S., Ramacci C., Leroux A. Expression of phosphorylated proteins from PI3-kinase and MAP-kinase signaling pathways in infiltrating breast cancer: relation with histopathologic and molecular subtypes. ESMO IMPAKT Breast cancer conference 2013, 2 au 4 Mai 2013, Bruxelles, Belgique. Ann Oncol 24 : 2013 (suppl ; abstr 735)

23 **Harlé A.**, Lion M., Lozano N., Husson M., Harter V., Genin P., Merlin J.L. PIK3CA exon 9 and 20 mutations analysis of breast cancers using PCR-HRM and PCR ARMS: Relation with clinicopathologic criteria. 14èmes journées du GPCO, 15 et 16 Novembre 2012, Paris, France.

24 **Harlé A.**, Merlin J.L. Comparaison des techniques COBAS KRAS 4800 et PCR TaqMan pour la détection des mutations somatiques du gène KRAS - Application dans le processus d'accréditation des laboratoires de biologie médicale selon la norme ISO 15189. 14èmes journées du GPCO, 15 et 16 Novembre 2012, Paris, France.

25 **Harlé A.**, Merlin J.L. Comparison of COBAS 4800 KRAS, PCR TaqMan and PCR High Resolution Melting assays for the detection of KRAS somatic mutations in Formalin Fixed Paraffin Embedded colorectal

tumors. ASCO EORTC NCI Markers In Cancer 2012, 11 au 13 Octobre 2012, Miami, FL, USA. J Clin Oncol 30: 2012 (suppl 30; abstr 103)

26 **Harlé A.**, Lozano N., Lion M., Husson M., Rouyer M., Ramacci C., Genin P., Merlin J.L. Highly sensitive determination of PIK3CA exon 9 and 20 hotspot mutations in breast tumors. ASCO annual meeting 2012, 1er au 5 Juin 2012, Chicago, IL, USA. J Clin Oncol 30, 2012 (suppl; abstr TPS10634)

27 **Harlé A.**, Lion M., Lozano N., Nicolaizeau C., Rouyer M., Husson M., Ramacci C., Genin P., Merlin J.L. Analyse des mutations PIK3CA : Nouveau biomarqueur de personnalisation thérapeutique dans les cancers du sein et dans les cancers colorectaux. 7ème Journée Claude Huriet de la Recherche clinique, 2 Mars 2012, Nancy, France.

28 **Harlé A.**, Lozano N., Nicolaizeau C., Chrétien A.S., Husson M., Ramacci C., Rouyer M., Genin P., Leroux A., Merlin J.L. Validation of two methods of detection of PIK3CA mutations on exons 9 and 20 by PCR-HRM and real-time PCR-ARMS Scorpion assay. 2ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie OncoTrans 2011, 8 et 9 Septembre 2011, Nancy, France.

29 Chretien A.S., **Harlé A.**, Meyer-Lefebvre M., Rouyer M., Husson M., Ramacci C., Leroux A., Genin P., Merlin J.L. Optimization of routine KRAS mutation testing procedure for rational individualized first line targeted therapy selection in metastatic colorectal cancer. 2ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie OncoTrans 2011, 8 et 9 Septembre 2011, Nancy, France.

30 **Harlé A.**, Scala-Bertola J., Gambier N., Jouzeau J.Y., Gillet P., Merlin J.L., Royer-Morrot M.J. Development of a simple method for the simultaneous determination of Imatinib, Norimatinib, Dasatinib and Nilotinib in human plasma by HPLC Photodiode Array Detector. 2ème Colloque Inter-Régional Grand-Est de Recherche Translationnelle en Oncologie OncoTrans 2011, 8 et 9 Septembre 2011, Nancy, France.

31 Merlin J.L., Genin P., **Harlé A.**, Giacometti G., Chergui F., Bouali S., Chrétien A.S., Ramacci C., Rouyer M., Dolivet G., Leroux A. Phosphoprotein array in tumor biopsies for molecular diagnosis of PI3 kinase and MAP kinase signalling functionality. AACR annual meeting 2007, 14 au 18 Avril 2007, Los Angeles, CA, USA. Molecular Cancer Therapeutics, 2007 Dec, 6(12) : 3540S-3540S

32 Chrétien A.S., **Harlé A.**, Genin P., Bouali S., Chergui F., Ramacci C., Rouyer M., Leroux A., Merlin J.L. Recherche de marqueurs prédictifs de la réponse aux thérapies ciblées anti-proteine kinases : Diagnostic moléculaire de la fonctionnalité des voies de signalisation par Bioplex® phosphoprotein array. 2ème Journée Claude Huriet de la Recherche clinique, Mars 2007, Nancy, France.

8. Sélection de 5 publications

Publication 1 : Representative Sequencing: Unbiased Sampling of Solid Tumor Tissue (Cell Rep 2020, Litchfield *et al.*)

J'ai choisi tout d'abord cette publication, réalisée en collaboration avec les chercheurs du Crick Institute lors de ma mobilité. Ce travail est en lien avec la thématique « hétérogénéité tumorale et évolution clonale » que je souhaite développer dans le cadre de mes projets de recherche translationnelle.

Publication 2 : Somatic substitution signature as an innovative tool in lung cancer diagnosis (Sci Rep 2019, Busca *et al.*)

Cette publication a été rédigée par un de mes étudiants, encadrée lors de sa thèse d'exercice en Médecine. Cet article est en lien également avec la thématique « hétérogénéité tumorale et évolution clonale » que je souhaite développer dans le cadre de mes projets de recherche translationnelle.

Publication 3 : Expression of pEGFR and pAKT as response-predictive biomarkers for RAS wild-type patients to anti-EGFR monoclonal antibodies in metastatic colorectal cancers (Br J Canc 2014, Harlé *et al.*)

Cette publication a été écrite pour l'obtention de mon diplôme de Docteur d'Université. Cette étude fait, je pense, le lien entre ce qui peut être fait en recherche translationnelle et en recherche plus fondamentale, notamment l'exploration des voies de signalisation chez les patients atteints de cancer.

Publication 4 : SMAD4 and the TGFβ pathway in patients with pancreatic ductal adenocarcinoma (Int J Mol Sci 2020, Dardare *et al.*)

Cette publication a été rédigée par une de mes étudiantes, encadrée lors de son Master2 Recherche. Ce papier de revue est en lien avec la thématique de recherche fondamentale que je souhaite continuer à développer au laboratoire.

Publication 5 : Emerging roles of DDB2 in cancer (Int J Mol Sci 2019, Gilson *et al.*)

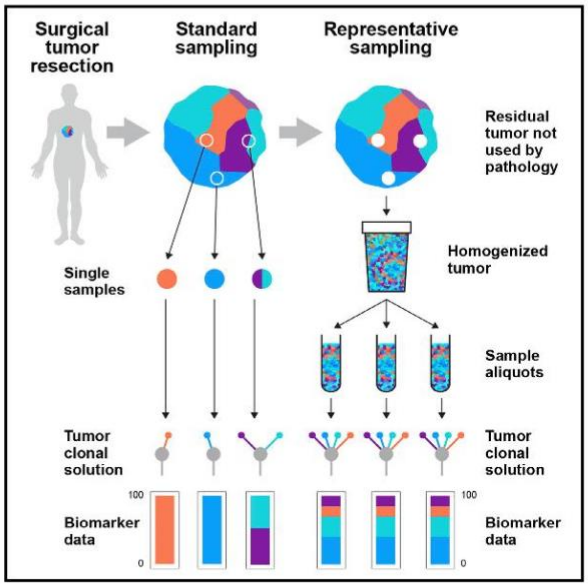
J'ai coordonné l'écriture de cet article de revue qui est en lien avec la thématique de recherche fondamentale que je souhaite continuer à développer au laboratoire.

Cell Reports

Resource

Representative Sequencing: Unbiased Sampling of Solid Tumor Tissue

Graphical Abstract



Highlights

- Representative sequencing (Rep-Seq) is a new method for tumor molecular profiling
- Rep-Seq homogenizes residual tumor tissue not taken for standard pathology
- Representative sampling of tumors generates accurate tumor mutation burden scores
- Rep-Seq detects more mutations and accurately resolves clonal from subclonal variants

Authors

Kevin Litchfield, Stacey Stanislaw, Lavinia Spain, ..., Charles Swanton, Nelson R. Alexander, Samra Turajlic

Correspondence

charles.swanton@crick.ac.uk (C.S.), nelson.alexander@roche.com (N.R.A.), samra.turajlic@crick.ac.uk (S.T.)

In Brief

Solid tumors are under-sampled in the clinic, such that only 0.0005% of initial tumor volume is used as input for diagnostic testing. Litchfield et al. apply the principles of representative sampling to implement an unbiased tumor sampling approach that improves the reproducibility and accuracy of next-generation sequencing.

Litchfield et al., 2020, Cell Reports *31*, 107550
May 5, 2020 © 2020 The Authors.
https://doi.org/10.1016/j.celrep.2020.107550

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

Representative Sequencing: Unbiased Sampling of Solid Tumor Tissue

Kevin Litchfield,[1,13] Stacey Stanislaw,[2,13] Lavinia Spain,[1,3] Lisa L. Gallegos,[2] Andrew Rowan,[1] Desiree Schnidrig,[1] Heidi Rosenbaum,[4] Alexandre Harle,[1,5] Lewis Au,[1,3] Samantha M. Hill,[2,6] Zayd Tippu,[3] Jennifer Thomas,[3] Lisa Thompson,[7] Hang Xu,[1] Stuart Horswell,[8] Aoune Barhoumi,[2] Carol Jones,[2] Katherine F. Leith,[2] Daniel L. Burgess,[4] Thomas B.K. Watkins,[1] Emilia Lim,[1] Nicolai J. Birkbak,[1,9] Philippe Lamy,[9] Iver Nordentoft,[9] Lars Dyrskjøt,[9] Lisa Pickering,[3] Stephen Hazell,[10] Mariam Jamal-Hanjani,[11,12] PEACE Consortium, James Larkin,[3] Charles Swanton,[1,11,12,*] Nelson R. Alexander,[2,*] and Samra Turajlic[1,3,14,*]

[1]Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK
[2]Roche Tissue Diagnostics, 1910 E. Innovation Park Drive, Tucson, AZ 85755, USA
[3]Renal and Skin Units, The Royal Marsden Hospital, London SW3 6JJ, UK
[4]Roche Sequencing Solutions, Madison, 500 S. Rosa Road, Madison, WI 53719, USA
[5]Université de Lorraine, CNRS UMR 7039 CRAN, Institut de Cancérologie de Lorraine, Service de Biopathologie, 54000 Nancy, France
[6]Department of Cancer Biology, University of Arizona Cancer Center, Tucson, AZ 85724, USA
[7]The Centre for Molecular Pathology, The Royal Marsden Hospital, London SW3 6JJ, UK
[8]Department of Bioinformatics and Biostatistics, The Francis Crick Institute, London NW1 1AT, UK
[9]Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark
[10]Histopathology Department, Royal Marsden NHS Foundation Trust, London and Sutton, UK
[11]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK
[12]Department of Medical Oncology, University College London Hospitals, London, UK
[13]These authors contributed equally
[14]Lead Contact
*Correspondence: charles.swanton@crick.ac.uk (C.S.), nelson.alexander@roche.com (N.R.A.), samra.turajlic@crick.ac.uk (S.T.)
https://doi.org/10.1016/j.celrep.2020.107550

SUMMARY

Although thousands of solid tumors have been sequenced to date, a fundamental under-sampling bias is inherent in current methodologies. This is caused by a tissue sample input of fixed dimensions (e.g., 6 mm biopsy), which becomes grossly under-powered as tumor volume scales. Here, we demonstrate representative sequencing (Rep-Seq) as a new method to achieve unbiased tumor tissue sampling. Rep-Seq uses fixed residual tumor material, which is homogenized and subjected to next-generation sequencing. Analysis of intratumor tumor mutation burden (TMB) variability shows a high level of misclassification using current single-biopsy methods, with 20% of lung and 52% of bladder tumors having at least one biopsy with high TMB but low clonal TMB overall. Misclassification rates by contrast are reduced to 2% (lung) and 4% (bladder) when a more representative sampling methodology is used. Rep-Seq offers an improved sampling protocol for tumor profiling, with significant potential for improved clinical utility and more accurate deconvolution of clonal structure.

INTRODUCTION

Over the past decade, clinical researchers have demonstrated the inability of an individual biopsy or formalin-fixed paraffin-embedded (FFPE) block to capture the genetic diversity of a solid tumor (Turajlic et al., 2018; Jamal-Hanjani et al., 2017; Warrick et al., 2019). This issue stems from the reliance on single-site samples, which are used as the current standard protocol for tumor sequencing in both research and clinical contexts. Sampling only once, from a single spatial location, will miss major expanding tumor subclones in other distant tumor locations, creating a bias that cannot be resolved through excess sequencing depth. Moreover, this widespread sampling bias is inherently obscured because the unsampled fixed tumor tissue not submitted for paraffin embedding is considered surgical waste and is incinerated. The pitfalls of using a non-representative sampling method are well documented in the "theory of sampling" developed by Pierre Gy (Gy, 1988) and have been practically demonstrated across multiple fields, ranging from food contamination to electoral polling and the mining industry (Rohde et al., 2015; Crespi, 1988; David, 1988). In the case of tumor sequencing, this bias arises within the context of spatially biased intratumor heterogeneity (ITH) as an important feature of cancer, combined with the increasing utility of molecular profiling as a tool to stratify patients for therapy (AACR Project GENIE Consortium, 2017). Failure to address this issue risks undermining the clinical utility of genomic medicine in cancer, through reduced sensitivity to detect prognostic and predictive markers, a lack

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

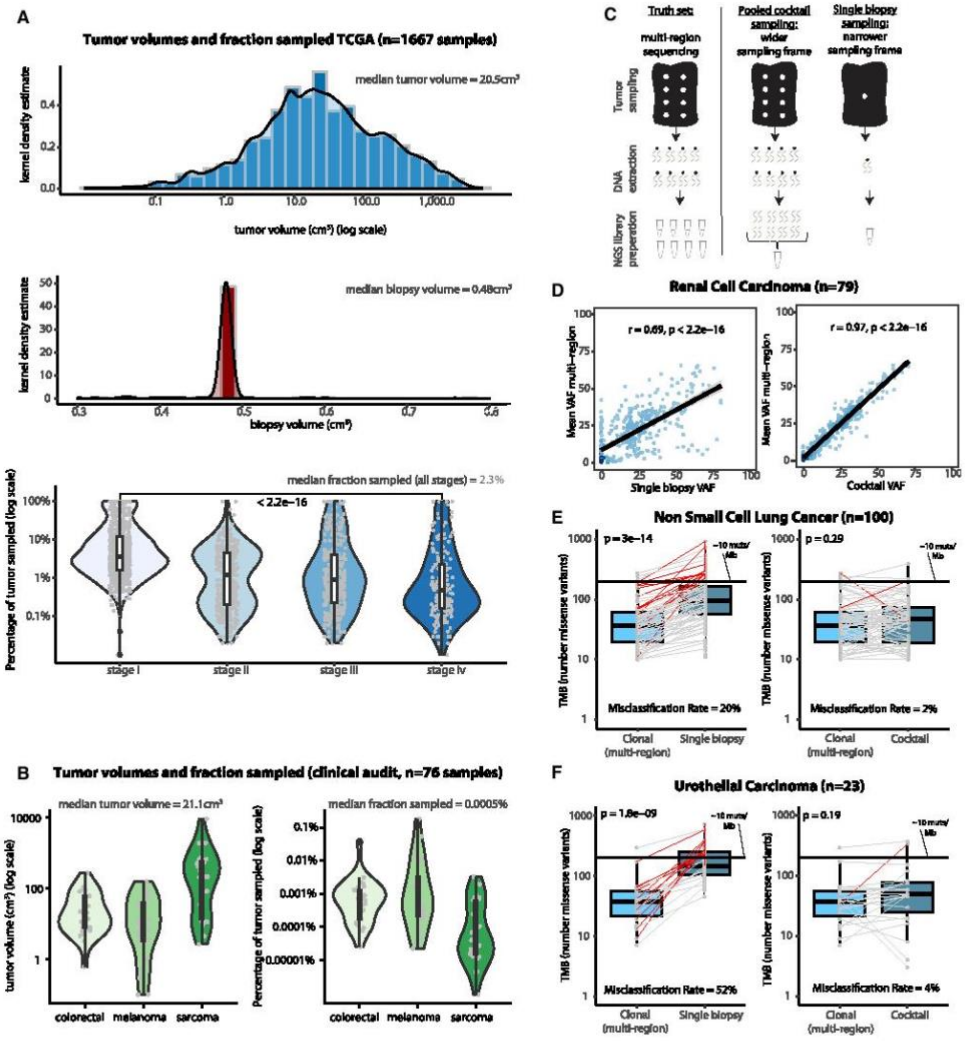


Figure 1. Current Tumor Sequencing Methods Lead to Under-sampling, Which Can Be Resolved through a Wider Sampling Frame

(A) Top: a density plot of the distribution of tumor volumes from The Cancer Genome Atlas (TCGA), with tumor volume (cubic centimeters) plotted on the x axis with log scale. Middle: a density plot of the volume of biopsy tissue used as input material for sequencing in the same cohort of TCGA samples. Bottom: the proportion of tissue sampled (i.e., values from the middle panel divided by those from the top panel for each case), split by tumor stage.

(B) Left: the estimated tumor volume (cubic centimeters) from a clinical audit, split by tissue site. Right: the proportion of tissue sampled, split by tissue site, in the clinical audit.

(C) The experimental design used for pilot experiment of pooled "cocktail" samples.

(D) Left: the variant allele frequency (VAF) correlation between cocktail sequencing (x axis) compared with true VAFs estimated from multi-biopsy sequencing (y axis); right: the variant allele frequency (VAF) correlation between single-region sequencing (x axis) compared with true VAFs estimated from multi-biopsy sequencing (y axis).

(legend continued on next page)

of reproducibility in sequencing results among samples, misassignment of subclonal variants as clonal (Pearson et al., 2016), and unreliable estimates of tumor mutation burden (TMB). A frequently proposed solution is to conduct multi-region sampling and profiling, which although able to address these issues in a research setting is cost prohibitive and too labor intensive for wide-scale adoption in clinical practice. Furthermore, the recurrent question of how many biopsies should be taken is unanswerable because of the high variability in ITH among patients (Turajlic et al., 2018). We hypothesized that a new sampling methodology could be implemented to circumvent these limitations, through the use of residual tumor tissue to create a more representative sample that captures the heterogeneity of the total tumor mass. We show that this new method better captures the diversity of the tumor and leads to improved biomarker results, without the need for increased sequencing depth or for multiple samples to be profiled per tumor. Furthermore, as the method described herein removes the physical bias of single-sample profiling and makes use of residual fixed tumor material that would otherwise be clinical waste, it empowers clinicians and researchers with a practical solution to overcome sampling bias within current tumor sequencing protocols.

RESULTS

Audit of Current Sampling Protocols

To examine the extent of current under-sampling bias in cancer genomics, we first analyzed pan-cancer sequencing data from The Cancer Genome Atlas (TCGA). Clinical annotation files were extracted for each solid tumor cohort from the Broad Institute's TCGA GDAC Firehose repository, from which tumor length and width data were available for 1,667 samples across six tumor types (STAR Methods). For each sample, total tumor tissue volume and research sample volume were estimated from the clinical data, and the two values were compared to assess the proportion of total tumor mass sampled for each case (STAR Methods). This revealed that current research protocols sample on average only 2.3% of the tumor mass (median value for all stages, n = 1,667 samples), decreasing to 0.5% for stage IV tumors (median value, n = 181) (Figure 1A). We next investigated this pattern within a routine clinical context, through audit of randomly selected cases undergoing molecular profiling as part of standard of care at a major cancer center (see STAR Methods). In total, 76 cases were audited, across three different solid tumor types. This revealed a median clinical tumor sampling proportion of only 0.0005% (range 0.000001%–0.2%) (Figure 1B), reflecting the minimal input material provided from standard FFPE tumor sections. The remaining ≥99.9% of tumor tissue is left unsampled for molecular profiling purposes, leading to a high level of under-sampling in a clinical context. Thus, within both research and clinical settings, current tumor sequencing protocols are associated with a high under-sampling bias. Furthermore, we note that sampling bias is also likely to be affected by the level of heterogeneity and the purity of tumor cells in the sample (Figure S1).

A More Representative Sample Leads to Increased Ability to Detect Variants and Improved Accuracy in TMB Estimation

To investigate the effect of spatial bias in single-biopsy sampling, we first conducted a pilot experiment, creating a single, more representative sample by pooling extracted DNA from 1,184 multi-region biopsies, taken from 79 primary renal cell carcinomas (RCCs), to create "cocktail" solutions per tumor (Figure 1C). Pooled cocktail samples were subject to next-generation sequencing (NGS) (median depth 674×), and mutation calls were compared with previously generated single-biopsy (reflecting current clinical practice; median depth 608×) and multi-region biopsy (truth set) data (median depth 612×) (Turajlic et al., 2018). Multi-region variant allele frequencies (VAFs) were determined by taking the mean across all regions per tumor, and single-region samples were selected by random sampling (see STAR Methods). All samples were processed through an identical protocol (see STAR Methods). Across all 79 tumors, the cocktail samples discovered a median of 100% (range 30%–100%) of the truth-set mutations compared with single biopsies, which achieved a median discovery rate of 73% (range 15%–100%), supporting the hypothesis that a more representative sample leads to improved ability to detect variants ($p = 6.6 \times 10^{-11}$, paired Wilcoxon test; Figure S2). In addition, VAFs derived from cocktail samples demonstrated a strong correlation with true VAF values from multi-region sequencing ($r = 0.97$), a superior correlation compared with that achieved by single-biopsy sampling ($r = 0.69$) (Figure 1D). This suggests that using a single, more representative sample provides a more accurate estimation of true cellular mutational prevalence across the total tumor mass (the true search space), an important consideration for both prognostic and predictive biomarkers. Within this context, TMB is emerging as a robust predictive biomarker for immune checkpoint inhibitor (CPI) therapy, validated across multiple tumor types (Forde et al., 2018; Hellmann et al., 2018; Snyder et al., 2014; Mariathasan et al., 2018). In particular clonal mutation load (clonal TMB) has been shown to be the key driver of CPI response (McGranahan et al., 2016; Miao et al., 2018), with subclonal mutations likely playing a neutral or negative role in achieving a sustained anti-tumor immune response (Gejman et al., 2018). To explore this further in the context of tumor sampling, we analyzed multi-region TRACERx non-small cell lung cancer (NSCLC) (Jamal-Hanjani et al., 2017) and urothelial carcinoma (UC) (Lamy et al., 2016; Thomsen et al., 2016) datasets and found high levels of intratumor TMB variability. In total, 20% of NSCLC cases (n = 100) and 52% of UC cases (n = 23) had one or more single biopsies with high TMB but low clonal TMB overall (on the basis of the prospectively validated 10.0

(E) Left: data from 100 non-small cell lung cancer cases, with clonal tumor mutation burden (TMB) values compared with single-biopsy TMB values (paired values within the same case are joined by gray line). Right: the same data, this time with clonal TMB compared with TMB estimates from *in silico* "cocktail" sampling. Red lines denote cases for which one paired value is below the 10 muts/Mb threshold and one is above. The 10 muts/Mb threshold is defined as 200 missense mutations.

(F) Urothelial carcinoma data from 23 patients, following the same format as (E).

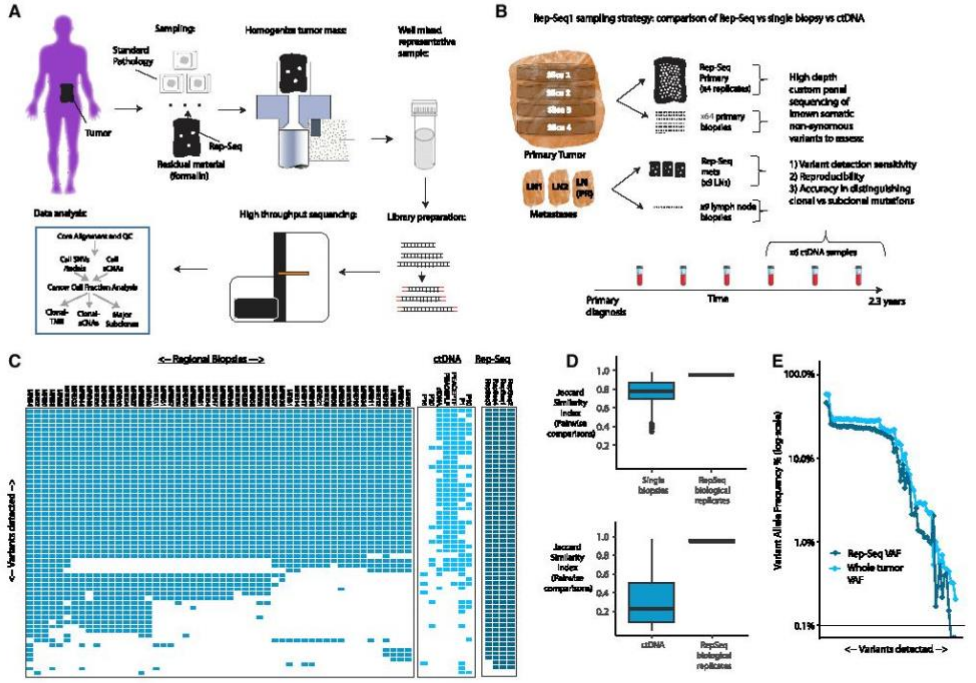


Figure 2. Representative Sequencing Method

(A) The methodological workflow for Representative Sequencing (Rep-Seq).

(B) The sampling strategy for case RS1, for validation of the Rep-Seq method against extensive biopsy sampling.

(C) The map of non-synonymous variants discovered in RS1, across tumor biopsies, ctDNA samples, and Rep-Seq biological replicates. PEACPLF is a cfDNA taken from pleural fluid, and PEACPTF is a cfDNA taken from peritoneal fluid, both at the time of rapid autopsy, and ctDNA is a pooled sample of P1/P10/P16/P20 timepoints.

(D) Jaccard similarity index results, as a measure of reproducibility, for single-biopsy sequencing versus Rep-Seq (left) and ctDNA sequencing versus Rep-Seq (right).

(E) The variant allele frequency (VAF) of mutations detected in RS1 Rep-Seq (dark blue) compared with VAF for the same mutation in the whole RS1 tumor (calculated as an average across all biopsies).

mutations/Mb threshold for CPI therapy; Forde et al., 2018; Figures 1E and 1F). These variability rates, in contrast, were reduced to 2% (NSCLC) and 4% (UC) when we generated *in silico* "cocktail" solutions and predicted clonal TMB (from the single cocktail sample; see STAR Methods), highlighting the potential clinical utility of a more representative sample in reducing the risk for TMB misclassification (Figures 1E and 1F). Finally, we assessed the ability of cocktail sequencing to determine clonal versus subclonal somatic copy number alterations (SCNAs), using the primary RCC dataset (n = 79). Known RCC clonal events (Turajlic et al., 2018) (e.g., loss of 3p25.3 and gain of 5q35.3) were found to have higher logR (log_2 ratio of coverage for tumor versus matched normal samples) than other (predominantly) subclonal alterations (Figure S2), indicating that clonal and subclonal SCNA events may be distinguishable in a pooled sequencing approach.

Homogenization of Residual Tumor Material to Create a Truly Representative Sample

Next, we next sought to develop an updated tumor sampling methodology that was consistent with the theory of sampling (Gy, 1988) and suitable for clinical adoption (i.e., not reliant on resource-intensive multiple-biopsy sampling and DNA extraction, as required for cocktail sampling). Here we demonstrate a new method called "representative sequencing" (Rep-Seq), which comprises homogenization of solid tumor masses into well-mixed solutions, coupled with NGS (Figure 2A). Tumor masses were sourced as the entire residual tumor material not taken for pathology assessment following surgery, material that would otherwise have been treated as clinical waste and destroyed (per College of American Pathologists and Royal College of Pathologists guidelines) or bio-banked. Residual samples represented on average 54.8% of the total tumor volume and

provided on average 223.5 g of tissue per tumor (on the basis of our pilot cohort of cases; see Table S1). These values are a significant increase over the average of 0.0005% sampled, and 0.0001 g of tumor tissue, currently used as sampling input in standard molecular profiling approaches. Following dissection of the formalin-fixed residual tumor tissue away from the surrounding normal tissue, the residual tumor mass is homogenized into a representative solution. From this well-mixed solution, samples are drawn for DNA extraction, library preparation, and sequencing (Figure 2A). Rep-Seq was implemented on a proof-of-concept basis in 11 tumors, from four different cancer types (Table S1).

Reproducibility and Sensitivity of Rep-Seq Compared with Biopsy Sequencing

The first tumor processed was RS1, a large clear cell RCC (ccRCC) tumor (17 cm maximal dimension), selected to allow extensive sampling for cross-validation purposes. In total, 64 fresh-frozen individual biopsies were taken from the primary tumor, and the remaining formalin-fixed residual mass (1,258 g of tissue) was homogenized under the Rep-Seq protocol. To define the variant landscape in this tumor, whole-exome sequencing (WES) was first conducted on a selection of 7 spatially disparate primary biopsies and an aliquot of the Rep-Seq sample (median depth 162×), leading to the discovery of a total of 76 non-synonymous mutations (single-nucleotide variants [SNVs] and small-scale insertion/deletions). These 76 mutations were subsequently captured in a targeted custom panel and successfully sequenced to high depth (median depth > 10,000×) in the 64 primary biopsies, 11 biopsies taken from two lymph node (LN) metastases, four separate aliquots drawn from the representative homogenate of the primary tumor sample (biological replicates), six circulating tumor DNA (ctDNA) samples collected at different time points, and three homogenized LN Rep-Seq samples (one LN was not biopsied) (Figure 2B). This integrated dataset was used to comprehensively evaluate the Rep-Seq methodology (Figure 2C). First we evaluated the reproducibility of each method, comparing the Jaccard similarity index between pairwise combinations of tumor biopsies, versus pairwise biological replicates of Rep-Seq, and pairwise combinations of ctDNA time points. Median pairwise similarity between biopsies was 0.78, suggesting that approximately 20–25% of mutations discovered in individual biopsies cannot be reproduced in subsequent biopsy samples. In contrast, median similarity among Rep-Seq replicates was higher at 0.95 (Figure 2D), with almost identical mutation lists discovered by each replicate (Figure 2C). Pairwise similarity among plasma ctDNA samples was low (overall median similarity index = 0.24; Figures 2C and 2D), reflecting the technical challenges of ctDNA profiling. The ongoing temporal evolution of the tumor across ctDNA time points should also be recognized, but we note substantial difference even among samples drawn close time points (e.g., P16 and P20 were only 21 days apart but shared no mutations in common). Last, we assessed the sensitivity of variant detection in case RS1 and note that variants as low as 0.15% VAF were successfully detected in the Rep-Seq sample (Figure 2E). In addition, the VAFs from Rep-Seq closely followed the overall tumor VAFs, as measured from the set of 64 biopsies (overall tumor VAF was calculated as total alt read count/total read count, summed across all biopsies) (Figure 2E). Only 3 mutations (of the 76 captured) were not detectable in the Rep-Seq sample, all with VAFs below 0.3%, close to the technical limit of ~0.1% VAF for variant detection using NGS protocols.

Clonal Structure Prediction from Rep-Seq Compared with Biopsy Sequencing

Next, we investigated the utility of Rep-Seq in determining clonal structure, given that measures of clonal diversity have been shown to associate with prognosis (Turajlic et al., 2018; McGranahan et al., 2016). Cancer cell fraction (CCF) estimates were first calculated for all mutations (n = 76) within the RS1 primary tumor biopsy set (n = 52 biopsies used, n = 12 excluded because of low purity; see STAR Methods) and grouped together into mutational clusters to infer a high-confidence benchmark clonal structure. Four distinct tumor clones were detected: clone A (truncal clone, mutations in every cancer cell, CCF = 100.0%, n = 41 mutations) and (sub)clones B (CCF = 45.6%, n = 2 mutations, and 14q loss as a known RCC driver SCNA event; Turajlic et al., 2018), C (CCF = 52.8%, n = 6 mutations), and D (CCF = 17.1%, n = 2 mutations) (Figure 3A). The remaining mutations were predominantly lower frequency and could not be reliably clustered into subclones. The clustering process was repeated for the Rep-Seq sample alone (n = 1), and the clonal solution was then compared between methods (Figure 3A). Rep-Seq correctly clustered all 41 clonal mutations together into truncal clone A (CCF = 100.0%), as well as identifying two major (sub) clones, B (CCF = 40.6%) and C (CCF = 33.3%) (Figure 3A) (we note that the sets of mutations in subclones B and C were not fully identical between the multi-region and Rep-Seq solutions). The remaining mutations in the Rep-Seq sample were again predominantly lower frequency and could not be reliably clustered, reflecting the challenge in accurately grouping together low-CCF mutations. CCF estimates from both methods were validated against physical mapping of spatial biopsy locations and mutational presence back to images of the sampled tumor: (sub)clone B was found in 20 of 52 primary biopsies (38.4%), and (sub)clone C was found in 32 of 52 (61.5%) (Figure 3B). This confirmed the presence of two major, spatially distinct subclones. Intriguingly, LN metastases LN1 and LN2 were exclusively seeded by clone B, whereas spatially proximal peri-renal LN sample LN(PR) was polyclonal, with clones B and C present. Individual mutation CCFs within each sample were next considered, in order to assess how well separated truncal events (clone A) were from subclonal mutations (clones B and C). Within the Rep-Seq sample, CCF estimates for clonal events (clone A) were clearly separated from (sub)clones B and C, reflecting the rapid convergence of CCF estimates in Rep-Seq toward true values (Figure 3C, right). In contrast, CCF distributions in single-sample biopsies overlapped among clones A, B, and C, with sub(clones) B and C frequently appearing (incorrectly) as clonal in individual biopsies with CCFs of ~100% ("illusion of clonality"; Burrell and Swanton, 2016). On average across the 52 primary biopsies, 17% (range 6%–35%) of clonal variants suffered from an illusion of clonality, if they were considered a single-region sample (Figure 3D) (all 76 mutations included; see STAR Methods). An illusion of clonality persisted even with multi-region biopsy

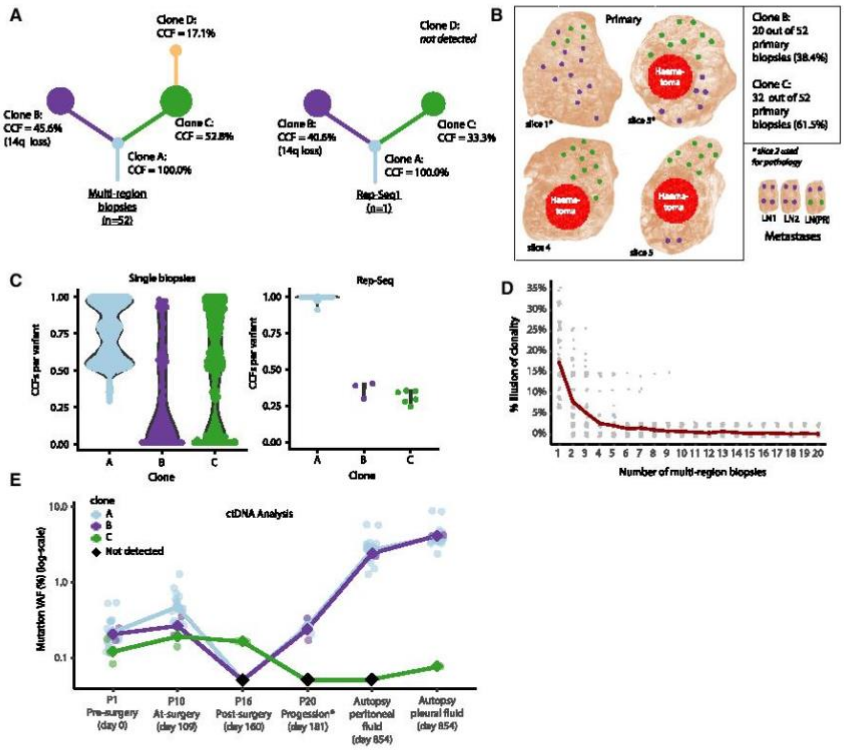


Figure 3. Clonal Tracking by Rep-Seq and ctDNA in Renal Cell Carcinoma

(A) The phylogenetic tree for case RS1, as derived using extensive multiple-biopsy sequencing (left) and a single Rep-Seq sample.

(B) The clonal distribution across the four slices of the primary RS1 tumor and lymph node metastases.

(C) The cancer cell fraction (CCF) estimates for mutations in tumor clones A, B, and C in each biopsy sample (left) and in the Rep-Seq sample (right).

(D) The illusion of clonality simulation data, with the number of simulated biopsy samples plotted on the x axis and the percentage of variants that incorrectly appear to be clonal (illusion of clonality) on the y axis (from 100 simulated sample combinations).

(E) ctDNA data for RS1, with the VAFs plotted (y axis) for mutations in clones A, B, and C.

sampling, with simulation showing that two random biopsies yielded an illusion of clonality rate of 9% (range 0%–25%), three of 6% (range 0%–25%), four of 4% (range 0%–15%), and five of 3% (range 0%–15%) (Figure 3D). Furthermore these results are likely to represent a conservative under-estimation of true illusion of clonality rates, as 3-dimensional sampling was conducted along the z axis in this case. We note that if standard 2-dimensional tumor slice sampling had been conducted only on slice 4, which was monoclonal for clone C, no number of biopsies would have prevented clonal illusion (Figure 3B), and critically, clone B (which metastasized to the LNs) would have been completely missed.

Given the broad applicability of liquid biopsies, and previous work demonstrating that both clonal and subclonal mutations can be identified (Abbosh et al., 2017), a pertinent question is to what extent ctDNA samples from plasma represent true clonal diversity compared with a more representative sampling of the primary tumor. Taking advantage of a well-characterized primary tumor, and five longitudinal ctDNA time points, we investigated this question. At pre-surgery time points (P1 and P10), variants from clones A, B, and C were all detectable at VAFs of 0.1%–1.0%, but many variants were missed, including some clone A truncal events, reflecting the technical challenges of profiling ctDNA. Mean VAF within clone A was overall higher than within (sub)clones B and C, but large inconsistency was observed in terms of individual variants analyzed across time points (Figure 3E). At post-surgery time point P16, the highest VAF variant was from clone C, and similarly at time point P20, a clone B mutation was highest, both above any clone A truncal mutations (Figure 3E). At later stage time points, however (post-mortem

sampling), all clonal mutations were detected, with highly consistent VAF frequency (Figure 3E). Correlation coefficients of VAFs from ctDNA, compared with the count of how many biopsies a mutation was present in, ranged from r = −0.17 (time point P16) to r = 0.78 (pleural fluid obtained post-mortem). Rep-Seq CCFs had the highest correlation with biopsy count data (r = 0.90). This suggests that inferring *de novo* clonal structure from ctDNA alone remains challenging, but tracking clonal markers that have been previously identified from tumor tissue remains highly informative (e.g., for minimal residual disease [MRD] tracking). We note, in the context of MRD tracking, that a larger panel of variants is likely to increase sensitivity to detect relapse at earlier time points (e.g., in the RS1 data at time point p16, an MRD panel based on mutations from a single biopsy would have missed disease relapse 53% of the time, compared with a 0% miss rate using a Rep-Seq-based panel; Table S2). In terms of the clonal dynamics of RS1, an interesting pattern was observed, with clone C dying out at later ctDNA time points and becoming undetectable (Figure 3E). This supports the predicted metastatic seeding pattern, as determined by biopsy and Rep-Seq profiling, that is, that clone B achieved distal metastatic seeding, whereas clone C was contained within the (peri-)renal area (Figure 3B). As an additional validation, copy number analysis was conducted on 31 further biopsies sampled from 20 distinct anatomical sites of metastatic disease present at the time of RS1 autopsy. All sites contained loss of 14q, a driver event found only in clone B in the primary tumor (Figure S3). We note that in single-region biopsy sequencing, metastasizing clone B would have been missed 32 times out of 52 (61.5%).

Rep-Seq Implemented on Lymph Node Samples in the Context of Metastatic Melanoma

To understand the utility of Rep-Seq profiling on LN residual material (as opposed to RS1, for which a large primary tumor was also available), we homogenized two LN samples (internal iliac LN and right inguinal LN) from a patient with metastatic melanoma (case RS2). In parallel, single-biopsy LN (right inguinal LN) sequencing was additionally conducted (biopsy taken prior to homogenization), as well as multi-region biopsies sequenced from a further eight distinct anatomical sites of metastases, in order to validate the Rep-Seq methodology (total n = 17 samples). All samples underwent WES (median coverage ~200×), with tissue sampled at the time of rapid autopsy (STAR Methods). The first question was to compare the results from LN Rep-Seq with those from single-sample LN sequencing, which would be the standard sampling approach in the context of a LN dissection. CCF clustering analysis was completed for the LN single sample (see STAR Methods; Figure 4A), and the calculated tree was monoclonal in structure (100% of mutations were predicted clonal, n = 90). Results from the homogenized Rep-Seq LN sample, in contrast, calculated a polyclonal tumor structure, with only 63% of mutations (n = 58) being clonal and the remaining mutations (n = 34) being subclonal, clustered into four distinct subclones (Figures 4B and 4C). To verify which solution was correct (mono- versus polyclonal), a joint clustering analysis was completed across all biopsy samples (n = 16 from LN and eight distal metastases) to yield a high-confidence clonal solution (Figure 4D). The tumor was indeed found to be polyclonal, with a total of seven distinct tumor subclones detected (Figure 4D). A major branch encompassing multiple distinct subclones (Figure 4D, far left of tree) was found to be shared across pelvic, liver, paravertebral, abdominal wall, loin soft tissue, and supra-renal metastases but was absent in two distinct right groin soft tissue masses (Figure 4D). The right groin masses were each characterized by distinct tumor subclones (Figure 4D). A number of individual sites were also found to be polyclonal in nature (e.g., liver, paravertebral mass) containing both the major left-hand subclone branch and the distinct subclones present in the right groin masses (Figure 4D). These polyphyletic patterns would be consistent with either polyclonal dissemination from the primary or metastasis-to-metastasis seeding. We note that not all tumor subclones were detectable using the Rep-Seq LN methodology, but the representative sampling methodology was accurate enough to clearly distinguish this as a polyclonal tumor, with a lower proportion of mutations being clonal. This finding was supported by extensive sampling of multiple metastases. The single-site sample, in contrast, incorrectly predicted a monoclonal tumor structure, with all mutations being clonal in nature. Clinically, we note that this patient received three lines of immune CPI therapy (adjuvant nivolumab, ipilimumab, and pembrolizumab) and failed to respond throughout, with progressive disease recorded for all rounds and sites of CPI treatment (Figure 4E). Although only descriptive in nature, this lack of response would be consistent with the hypothesis of a heterogeneous subclonal neoantigen repertoire being associated with poor response to immunotherapy (Wolf et al., 2019). For completeness, we additionally obtained archival FFPE blocks for the RS2 primary tumor sample, and at median ~200× sequencing coverage, none of the metastasizing clones were detectable, highlighting the challenge of melanoma primary tissue sampling where the tumor size is so small.

Rep-Seq Implemented on an Extended Cohort of Primary Solid Tumors across Four Tissue Types

The Rep-Seq method was additionally conducted in a further ten cases as a technical feasibility exercise. Cases RS3 and RS4 were ccRCC tumors and predominantly monoclonal in structure (Figure S4). Cases RS5–RS11 were tumors of breast, colorectal, and lung origin, and tumor-specific driver mutations were successfully detected in all specimens (Table S3). The highest mutation burden was observed in RS8 (colon), with 980 and 251 non-synonymous SNVs and indels and presence of cosmic mutational signatures 6 and 15 associated with mismatch repair deficiency (Figure 5A), which was confirmed by immunohistochemistry staining showing loss of *MLH1*. Signature analysis of the three NSCLC tumors (RS9, RS10, and RS11) showed evidence of signature 4 (smoking associated) uniquely in these three tumors only (Figure 5A). Hence the mutational signatures derived from Rep-Seq were consistent with expected patterns.

Tumor Purity-Enriched Rep-Seq Protocol

Finally, an additional benefit of tissue homogenization is the preservation of the cellular integrity of the tissue and therefore the ability to add a cell-sorting step prior to DNA purification, to enrich for higher tumor purity. As a proof of principle, flow sorting was conducted on RS12 (colorectal tumor), preferentially selecting isolated tumor nuclei on the basis of the presence of

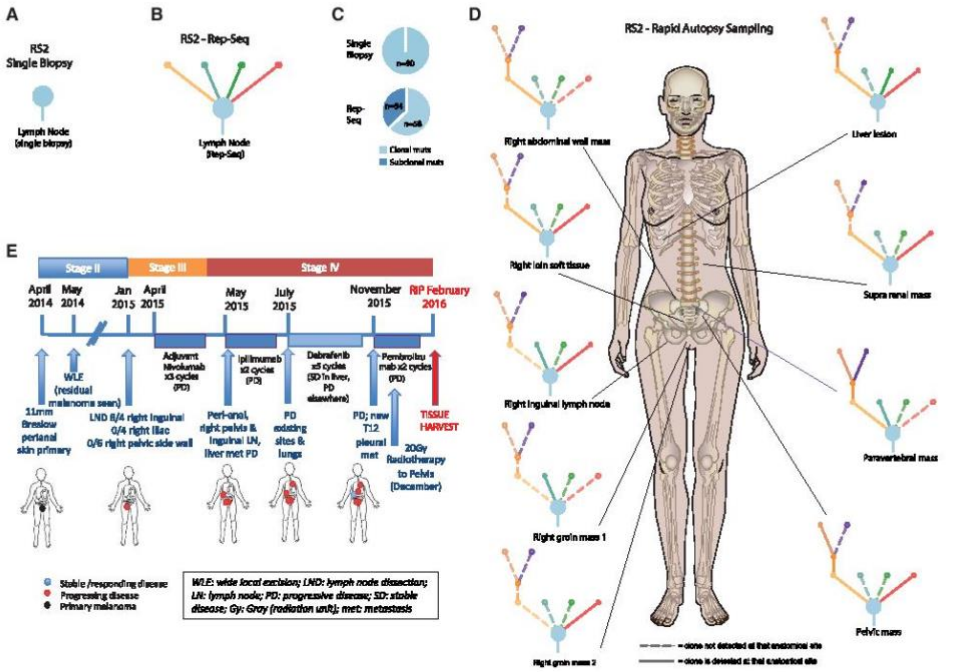


Figure 4. Clonal Tracking by Rep-Seq in Metastatic Melanoma

(A) The phylogenetic tree for case RS2 using single-biopsy lymph node sequencing (right inguinal LN).

(B) The phylogenetic tree for case RS2 using Rep-Seq LN sequencing.

(C) The proportion of mutations, clonal and subclonal, in RS2 on the basis of single-biopsy (top) and Rep-Seq (bottom) profiling.

(D) The extended phylogenetic tree for case RS2, on the basis of joint clustering of LN and eight distal metastases samples. Clone presence is marked on the basis of anatomical site of disease, with absent clones in a given site shaded in lighter color. Note that the right inguinal LN now becomes subclonal in composition, because of the benefit of multi-region joint clustering. Previously in (A), these same mutations are incorrectly clustered to the truncal clone (illusion of clonality) because of the limitations of single-biopsy sampling.

(E) The clinical history of case RS2. The best response by RECIST version 1.1 criteria is denoted in parentheses. SD, stable disease; PD, progressive disease.

cytokeratins 8 and 18 and high forward/side scattering. WES was conducted, first on a standard sample from the Rep-Seq protocol (non-sorted Rep-Seq, depth 221×), which showed tumor purity of 0.44. Strong enrichment was observed in the flow-sorted Rep-Seq sample (depth 215×), with purity of 0.89, which resulted in an approximate doubling of the effective tumor cell sequencing coverage (from 90× to 184×), for the same overall bulk sample sequencing depth and cost (Figure 5B). In terms of variant discovery, 365 non-synonymous SNVs were observed in common between both standard and flow-sorted samples, and then an additional 68 mutations (an increase of 19%) were found uniquely in the flow-sorted sample, presumably because of the increased sensitivity from higher effective tumor depth (Figure 5B). Only 5 mutations (1.4% of the total, all with VAFs < 5%) were found with the opposite pattern (i.e., in the standard sample but missing in the flow-sorted sample), which suggests in this case that flow sorting did not create systematic bias such that certain tumor subclones are excluded. The additional sensitivity in mutation detection enabled a larger set of subclones to be defined (Figure 5C).

DISCUSSION

Here we present Rep-Seq as a new method to achieve unbiased tumor sampling, drawing DNA molecules from a well-mixed homogenized solution of all residual surgical tumor material. This method removes the spatial bias inherent in current single-biopsy approaches and significantly increases the probability of detecting the genomic heterogeneity of solid tumors in a single sample protocol. Successful delivery of precision medicine in oncology is contingent on a reliable identification of biomarkers. ITH affects the interpretation of prognostic biomarkers and predictive biomarkers of response to targeted and immunotherapy, most notably TMB. Rep-Seq offers a clear opportunity to

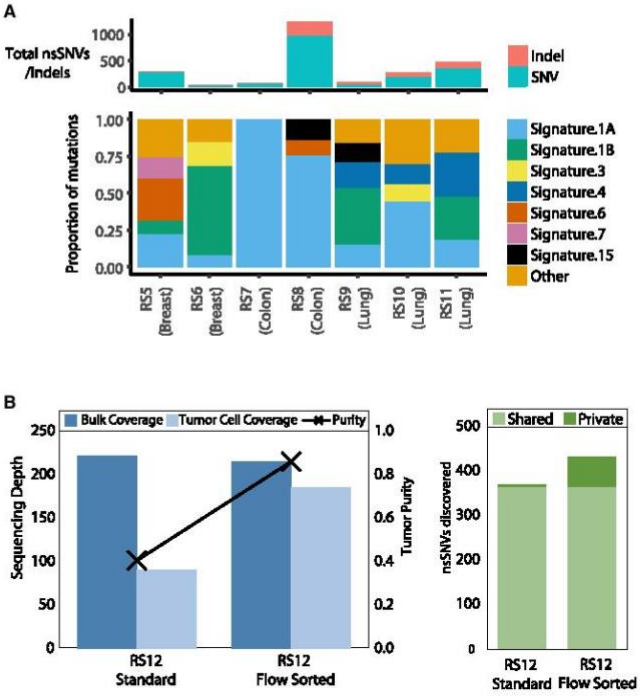


Figure 5. Extended Rep-Seq Cohort and Tumor Purity Enrichment

(A) Non-synonymous mutation count for cases RS5–RS11 (top), along with mutational signature analysis results (bottom).

(B) Left: RS12 coverage data for the overall bulk mixed cell population and then an estimated tumor cell coverage, along with purity estimates. Data are shown for standard Rep-Seq, and then tumor purity-enriched Rep-Seq, for comparison purposes. Right: the number of non-synonymous variants detected in normal and purity-enriched Rep-Seq samples. Variant counts are color coded on the basis of either being "shared" (i.e., present in both samples) or "private" (i.e., variant is present in only that sample and absent from the other).

(C) Phylogenetic trees for the normal and flow-sorted RS12 tumor subclones.

overcome the issue of ITH in these contexts and could feasibly be adopted in routine clinical practice.

In our analysis we first show the extent of under-sampling in current research studies, with data from TCGA demonstrating an average tissue sampling proportion of 2.3%. This bias is even more pronounced in a routine clinical context, with our audit data from standard molecular profiling at a major cancer center demonstrating that only five cancer cells in every million (0.0005%) are being sampled. This rate of under-sampling raises the risk for misinterpreting clinically relevant tumor sequencing results, as demonstrated by a high level of variability in TMB scores between single sites and the true clonal mutation load. Indeed, in NSCLC and UC, we show that 20% and 52% of cases, respectively, have at least one biopsy classified as high TMB, when the overall clonal TMB is low. This misclassification risk can be reduced through using a more representative sample, with discordance rates being reduced to 2% and 4%, respectively, using *in silico* "cocktail" analysis. We note, however, that the *in silico* "cocktail" samples may have slightly higher tumor purity than the main Rep-Seq methodology, as the latter takes all residual tumor material as input rather the biopsy-sampled regions.

To address the issue of spatial sampling bias in a manner consistent with the theory of sampling, as well as the challenge of multiple-biopsy sampling being too labor intensive for routine clinical use, we next sought to develop an updated tumor sampling methodology. We demonstrate Rep-Seq as a new method able to overcome both these points, through sampling a large volume of tumor tissue (average sampled proportion 55%), but maintaining a single (n = 1) DNA sample for downstream NGS library preparation and sequencing. The source of material in Rep-Seq is unique by making use of excess residual tumor tissue from surgery, which would otherwise be destroyed as clinical waste or bio-banked for research use. Homogenization of this residual tissue makes an ideal source for DNA extraction and sequencing, given the more representative nature of the sample, and generates a substantial amount of material that can be retained for future

research use. Such practices of homogenized sampling have been in routine practice in other fields for many decades, such as the sampling of mineral deposits. This fact, combined with the large amount tissue processed as clinical waste, highlights the sub-optimal nature of current molecular profiling sampling approaches in oncology. Here we demonstrate the utility of Rep-Seq through extensive sampling in a cohort of pilot cases, primarily as a proof-of-concept study. We show that the reproducibility in variant detection between replicates in Rep-Seq is high (Jaccard similarity index = 0.95) and superior to that of single-biopsy sequencing (0.78). In case RS1, sampling of >50 biopsies, from five tumor slices, validated the accuracy of Rep-Seq clonality predictions. In contrast, single-biopsy sequencing suffered from a high rate of illusion of clonality, with 17% of mutations appearing clonal when in reality they were subclonal. Conversely, single-biopsy sequencing also frequently missed important events, with 62% of biopsies missing the metastasizing clone. And notably all biopsies from tumor slice 4 missed the lethal clone, indicating that even extensive multi-region sampling is still susceptible to spatial bias, unless implemented on a 3-dimensional basis. Ultimately, any biopsy-based sequencing approach will retain spatial bias to some degree when tumor tissue is left unsampled, which hence emphasizes the benefit of a homogenization-based sampling approach.

In case RS2, we used Rep-Seq to profile LN metastases from a patient with primary melanoma, which revealed polyclonal disease. This was confirmed by extensive biopsy sequencing of eight further sites of distal metastases, which validated polyclonality, with multiple distinct subclones present across anatomically separate sites. Results from single-biopsy LN sequencing, in contrast, predicted a monoclonal tumor, which was shown to be incorrect. With increasing evidence now supporting clonal neoantigens as a key driver of immunotherapy response (Miao et al., 2018; Gejman et al., 2018), the accurate delineation of clonal versus subclonal alterations is of clinical relevance. This is descriptively illustrated in case RS2, in which lack of response to three lines of immunotherapy was observed, in the context of a high subclonal neoantigen repertoire. We additionally implemented Rep-Seq on a further ten cases (RS3–RS12), from lung, colorectal, and breast cancer types, showing mutation rates and signatures in line with expected results. Finally, as a further extension of the Rep-Seq protocol, we demonstrate flow sorting as a method to achieve tumor purity enrichment. For case RS12, purity was increased from 44% to 89%, with no observed shift or bias introduced to the sequencing results. Purity enrichment complements well the homogenization-based Rep-Seq approach and could significantly reduce sequencing costs (the same equivalent coverage of tumor cells could have been achieved in half the cost for case RS12).

We acknowledge small cohort size as a limitation of this work and confirm that a larger prospective study of Rep-Seq (n = 500 cases) is under way (ClinicalTrials.gov identifier NCT03832062). In addition, although Rep-Seq yields accurate and unbiased estimates of clonal and major subclonal mutations, this comes with the trade-off of losing resolution to detect lower frequency tumor (sub)clones. This trade-off may be acceptable in a clinical context, in which lower frequency mutations may be less directly actionable than widely expanded clonal or major subclonal driver events. Furthermore, we show with deep custom panel sequencing in case RS1 that low-frequency mutations can be reliably detected in Rep-Seq down to 0.3% VAF, with sensitivity determined by current limits in NGS accuracy. As more sensitive sequencing assays (e.g., duplex sequencing) become adopted, it is likely that Rep-Seq can be used to search for ultra-low-frequency events. We also acknowledge that Rep-Seq will not be applicable in all patient contexts, for example, patients with advanced-stage disease who undergo only biopsy sampling (no other surgeries). However, in metastatic cases undergoing resection, as well as primary surgery, there is broad potential for bio-banking and homogenization of residual material using the Rep-Seq approach. Finally, there is also potential scope for hybrid approaches, in which a large part of the residual material is homogenized, but then a portion (e.g., a transverse section) is also retained for spatially dependent analysis (e.g., imaging, spatial transcriptomics).

In summary, here we implement Rep-Seq as a new tumor sampling methodology combined with NGS, which adopts a more representative sampling approach via homogenization of residual tumor tissue. This method offers a clinically practical solution to the dramatic under-sampling bias inherent in current molecular profiling workflows. We find the reproducibility of results in Rep-Seq to be significantly higher than for current single-biopsy sequencing approaches (at same equivalent sequencing depth), and greater accuracy was also achieved in determining clonal from subclonal variants. We note that these results are predicted from the theory of sampling, a statistically driven rule set for sampling solid masses that are heterogeneous. These results offer potential clinical utility in the context of both prognostic (e.g., greater sensitivity to detect metastasis driving subclones) and predictive biomarkers (e.g., improved clonal TMB estimates).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Clinical studies
- METHOD DETAILS
  - Clinical audit of current molecular profiling practices
  - Regional biopsy and cocktail sample preparation
  - Grossing and homogenization of residual tumor tissue
  - Genomic DNA purification from tissue and cfDNA
  - Target-enriched NGS Library construction and sequencing
  - Flow sorting method to increase tumor purity
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Tumor volume sampling analysis
  - Processing of sequencing data
  - Analysis of pooled cocktail sequencing data
  - Analysis of in-silico pooled cocktail sequencing data

- Custom panel design
- Jaccard reproducibility analysis
- Clustering and phylogenetic analysis
- Illusion of clonality simulation
- Analysis of purity enriched data
- Statistical methods

● DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.107550.

CONSORTIA

The members of PEACE Consortium are Chris Abbosh, Kai-Keen Shiu, John Bridgewater, Daniel Hochhauser, Martin Forster, Siow-Ming Lee, Tanya Ahmad, Dionysis Papadatos-Pastos, Sam Janes, Peter Van Loo, Katey Enfield, Nicholas McGranahan, Ariana Huebner, Sergio Quezada, Stephan Beck, Peter Parker, Henning Walczak, Tariq Enver, Rob Hynds, Mary Falzon, Ian Proctor, Ron Sinclair, Chi-wah Lok, Zoe Rhodes, David Moore, Teresa Marafioti, Elaine Borg, Miriam Mitchison, Reena Khiroya, Giorgia Trevisan, Peter Ellery, Mark Linch, Sebastian Brandner, Crispin Hiley, Selvaraju Veeriah, Maryam Razaq, Heather Shaw, Gert Attard, Mita Afroza Akther, Cristina Naceur-Lombardelli, Lizi Manzano, Maise Al-Bakir, Simranpreet Summan, Nnenna Kanu, Sophie Ward, Uzma Asghar, Emilia Lim, Faye Gishen, Adrian Tookman, Paddy Stone, Caroline Stirling, Andrew Furness, Kim Edmonds, Nikki Hunter, Sarah Sarker, Sarah Vaughan, Mary Mangwende, Karla Pearce, Lavinia Spain, Scott Shepherd, Haixi Yan, Ben Shum, Eleanor Carlyle, Steve Hazell, Annika Fendler, Fiona Byrne, Nadia Yousaf, Sanjay Popat, Olivia Curtis, Gordon Stamp, Antonia Toncheva, Emma Nye, Aida Murra, Justine Korteweg, Nahid Sheikh, Debra Josephs, Ashish Chandra, James Spicer, Ula Mahadeva, Anna Green, Ruby Stewart, Lara-Rose Iredale, Tina Mackay, Ben Deakin, Debra Enting, Sarah Rudman, Sharmistha Ghosh, Lena Karapagniotou, Elias Pintus, Andrew Tutt, Sarah Howlett, Vasiliki Michalarea, James Brenton, Carlos Caldas, Rebecca Fitzgerald, Merche Jimenez-Linan, Elena Provenzano, Alison Cluroe, Grant Stewart, Colin Watts, Richard Gilbertson, Ultan McDermott, Simon Tavare, Emma Beddowes, Patricia Roxburgh, Andrew Biankin, Anthony Chalmers, Sioban Fraser, Karin Oien, Andrew Kidd, Kevin Blyth, Matt Krebs, Fiona Blackhall, Yvonne Summers, Caroline Dive, Richard Marais, Fabio Gomes, Mat Carter, Jo Dransfield, John Le Quesne, Dean Fennell, Jacqui Shaw, Babu Naidu, Shobhit Baijal, Bruce Tanchel, Gerald Langman, Andrew Robinson, Martin Collard, Peter Cockcroft, Charlotte Ferris, Hollie Bancroft, Amy Kerr, Gary Middleton, Joanne Webb, Salma Kadiri, Peter Colloby, Bernard Olisemeke, Rodelaine Wilson, Ian Tomlinson, Sanjay Jogai, Christian Ottensmeier, David Harrison, Massimo Loda, Adrienne Flanagan, Mairead McKenzie, Allan Hackshaw, Jonathan Ledermann, Abby Sharp, Laura Farrelly, and Hayley Bridger.

ACKNOWLEDGMENTS

We greatly thank the staff at the Royal Marsden Pathology Lab for their kind collaboration and efforts to support this project, as well as the Renal and Skin Unit Research Team, also at the Royal Marsden NHS Foundation Trust. We gratefully acknowledge all members of the PEACE Consortium. We kindly thank Melinda Day and Patrick Brunhoeber for their early collaboration and support on this project, and we thank Joe Brock for his excellent illustration utilised in Figure 4. We dedicate this work to the memory of Martin Gore. Funding for this study was provided by Roche Tissue Diagnostics. K.L. is funded by the Medical Research Council (MR/P014712/1), the Rosetrees Trust (A2437), Cancer Research UK (C69256/A30194), and the CRUK Lung Cancer Centre of Excellence. A.H. is funded by Fondation de France. C.S. is a Royal Society Napier Research Professor. This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001169, FC001202), the Medical Research Council (FC001169, FC001202), and the Wellcome Trust (FC001169, FC001202). C.S. is funded by Cancer Research UK (TRACERx, PEACE, and CRUK Cancer Immunotherapy Catalyst Network), the CRUK Lung Cancer Centre of Excellence, the Rosetrees Trust, the NovoNordisk Foundation (ID16584), and the Breast Cancer Research Foundation (BCRF). This research is supported by a Stand Up To Cancer (SU2C)-LUNGevity-American Lung Association Lung Cancer Interception Dream Team Translational Research Grant (SU2C-AACR-DT23-17). SU2C is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C. C.S. receives funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) Consolidator Grant (FP7-THESEUS-617844), European Commission ITN (FP7-PloidyNet 607722), an ERC Advanced Grant (PROTEUS) from the ERC under the European Union's Horizon 2020 research and innovation program (835297), and Chromavision from the European Union's Horizon 2020 research and innovation program (665233). J.L. is supported by the National Institute for Health Research (NIHR) RM/ICR Biomedical Research Centre for Cancer. S.T. is funded by Cancer Research UK (C50947/A18176), the NIHR Biomedical Research Centre at the Royal Marsden Hospital and Institute of Cancer Research (A109), the Kidney and Melanoma Cancer Fund of The Royal Marsden Cancer Charity, the Rosetrees Trust (A2204), and Ventana Medical Systems Inc. (10467 and 10530). We thank the patients and their families.

AUTHOR CONTRIBUTIONS

Study Design, N.R.A., S.T., and C.S.; Laboratory Experiments, S.S., A.R., H.R., L.G., S. Hill, A.H., and D.L.B.; Clinical/Pathology Work, S.T., L.S., L.A., J.T., L.T., C.J., S. Hazell, and J.L.; Data Analysis, K.L., H.X., D.S., and S. Horswell; Data Interpretation, all authors; Manuscript Writing, K.L., S.T., N.R.A., S.S., H.R., L.G., and S.H.

DECLARATION OF INTERESTS

S.T. reports grants from Roche Tissue Diagnostics. K.L., S.T., and C.S. report speaker fees from Roche Tissue Diagnostics. K.L., S.S., S.T., and N.R.A. report a patent pending using Rep-Seq profiling to support detection of MRD. N.R.A., S.S., L.G., A.B., K.F.L., and S.H. have patents pending on representative sampling of solid tumors. C.S. receives grant support from Pfizer, AstraZeneca (AZ), Bristol-Myers Squibb (BMS), Roche-Ventana, Boehringer Ingelheim, and Ono. C.S. has consulted for Pfizer, Novartis, GlaxoSmithKline (GSK), Merck Sharpe & Dohme (MSD), BMS, Celgene, AZ, Illumina, Genentech, Roche-Ventana, GRAIL, Medicxi, and the Sarah Cannon Research Institute. C.S. is a shareholder of Apogen Biotechnologies, Epic Bioscience, and GRAIL and has stock options in and is co-founder of Achilles Therapeutics. Outside of the submitted work, K.L., S.T., and C.S. have a patent pending on indel burden and CPI response and a patent pending on targeting of frameshift neoantigens for personalized immunotherapy. J.L. reports institutional research support from BMS, MSD, Novartis, Pfizer, Achilles Therapeutics, Roche, Nektar Therapeutics, Covance, Immunocore, Pharmacyclics, and Aveo and consultancy support from Achilles, AZ, Boston Biomedical, BMS, Eisai, EUSA Pharma, GSK, Ipsen, Imugene, Incyte, iOnctura, Kymab, Merck Serono, MSD, Nektar, Novartis, Pierre Fabre, Pfizer, Roche/Genentech, Secarna, and Vitaccess. S.T. has received speaking fees from Astra Zeneca, Novartis, and Ipsen. S.T., K.L., and C.S. have the following patents filed: Clear Cell Renal Cell Carcinoma Biomarkers P113326GB.

Received: June 2, 2019
Revised: September 17, 2019
Accepted: April 1, 2020
Published: May 5, 2020

REFERENCES

AACR Project GENIE Consortium (2017). AACR Project GENIE: powering precision medicine through an international consortium. Cancer Discov. 7, 818–831.

Abbosh, C., Birkbak, N.J., Wilson, G.A., Jamal-Hanjani, M., Constantin, T., Salari, R., Le Quesne, J., Moore, D.A., Veeriah, S., Rosenthal, R., et al.; TRACERx Consortium; PEACE Consortium (2017). Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature *545*, 446–451.

Andreatta, M., and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics *32*, 511–517.

Aran, D., Sirota, M., and Butte, A.J. (2015). Systematic pan-cancer analysis of tumour purity. Nat. Commun. *6*, 8971.

Budczies, J., Allgäuer, M., Litchfield, K., Rempel, E., Christopoulos, P., Kazdal, D., Endris, V., Thomas, M., Fröhling, S., Peters, S., et al. (2019). Optimizing panel-based tumor mutational burden (TMB) measurement. Ann. Oncol. *30*, 1496–1506.

Burrell, R.A., and Swanton, C. (2016). Re-evaluating clonal dominance in cancer evolution. Trends Cancer *2*, 263–276.

Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219.

Crespi, I. (1988). Pre-Election Polling: Sources of Accuracy and Error (Russell Sage).

David, M. (1988). Handbook of Applied Advanced Geostatistical Ore Reserve Estimation (Elsevier).

Fang, H., Bergmann, E.A., Arora, K., Vacic, V., Zody, M.C., Iossifov, I., O'Rawe, J.A., Wu, Y., Jimenez Barron, L.T., Rosenbaum, J., et al. (2016). Indel variant analysis of short-read sequencing data with Scalpel. Nat. Protoc. *11*, 2529–2548.

Faustino-Rocha, A., Oliveira, P.A., Pinho-Oliveira, J., Teixeira-Guedes, C., Soares-Maia, R., da Costa, R.G., Colaço, B., Pires, M.J., Colaço, J., Ferreira, R., and Ginja, M. (2013). Estimation of rat mammary tumor volume using caliper and ultrasonography measurements. Lab Anim. (NY) *42*, 217–224.

Forde, P.M., Chaft, J.E., and Pardoll, D.M. (2018). Neoadjuvant PD-1 blockade in resectable lung cancer. N. Engl. J. Med. *379*, e14.

Gejman, R.S., Chang, A.Y., Jones, H.F., DiKun, K., Hakimi, A.A., Schietinger, A., and Scheinberg, D.A. (2018). Rejection of immunogenic tumor clones is limited by clonal fraction. eLife *7*, e41090.

Gerstung, M., Beisel, C., Rechsteiner, M., Wild, P., Schraml, P., Moch, H., and Beerenwinkel, N. (2012). Reliable detection of subclonal single-nucleotide variants in tumour cell populations. Nat. Commun. *3*, 811.

Gy, P. (1988). Heterogeneite, Echantillonnage, Homogeneisation (Masson).

Hedley, D.W., Friedlander, M.L., Taylor, I.W., Rugg, C.A., and Musgrove, E.A. (1983). Method for analysis of cellular DNA content of paraffin-embedded pathological material using flow cytometry. J. Histochem. Cytochem. *31*, 1333–1335.

Hellmann, M.D., Ciuleanu, T.E., Pluzanski, A., Lee, J.S., Otterson, G.A., Audigier-Valette, C., Minenza, E., Linardou, H., Burgers, S., Salman, P., et al. (2018). Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. N. Engl. J. Med. *378*, 2093–2104.

Jamal-Hanjani, M., Wilson, G.A., McGranahan, N., Birkbak, N.J., Watkins, T.B.K., Veeriah, S., Shafi, S., Johnson, D.H., Mitter, R., Rosenthal, R., et al.; TRACERx Consortium (2017). Tracking the evolution of non-small-cell lung cancer. N. Engl. J. Med. *376*, 2109–2121.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics *25*, 2283–2285.

Lamy, P., Nordentoft, I., Birkenkamp-Demtröder, K., Thomsen, M.B., Villesen, P., Vang, S., Hedegaard, J., Borre, M., Jensen, J.B., Høyer, S., et al. (2016). Paired exome analysis reveals clonal evolution and potential therapeutic targets in urothelial carcinoma. Cancer Res. *76*, 5894–5906.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E.E., III, Koeppen, H., Astarita, J.L., Cubas, R., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature *554*, 544–548.

McGranahan, N., Furness, A.J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science *351*, 1463–1469.

Miao, D., Margolis, C.A., Vokes, N.I., Liu, D., Taylor-Weiner, A., Wankowicz, S.M., Adeegbe, D., Keliher, D., Schilling, B., Tracy, A., et al. (2018). Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. Nat. Genet. *50*, 1271–1281.

Nilsen, G., Liestøl, K., Van Loo, P., Moen Vollan, H.K., Eide, M.B., Rueda, O.M., Chin, S.F., Russell, R., Baumbusch, L.O., Caldas, C., et al. (2012). Copynumber: efficient algorithms for single- and multi-track copy number segmentation. BMC Genomics *13*, 591.

Pearson, A., Smyth, E., Babina, I.S., Herrera-Abreu, M.T., Tarazona, N., Peckitt, C., Kilgour, E., Smith, N.R., Geh, C., Rooney, C., et al. (2016). High-level clonal FGFR amplification and response to FGFR inhibition in a translational clinical trial. Cancer Discov. *6*, 838–851.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Raynaud, F., Mina, M., Tavernari, D., and Ciriello, G. (2018). Pan-cancer inference of intra-tumor heterogeneity reveals associations with different forms of genomic instability. PLoS Genet. *14*, e1007669.

Rohde, A., Hammerl, J.A., Appel, B., Dieckmann, R., and Al Dahouk, S. (2015). Sampling and homogenization strategies significantly influence the detection of foodborne pathogens in meat. BioMed Res. Int. *2015*, 145437.

Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B.S., and Swanton, C. (2016). DeconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol. *17*, 31.

Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Côté, A., and Shah, S.P. (2014). PyClone: statistical inference of clonal population structure in cancer. Nat. Methods *11*, 396–398.

Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et al. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. *33*, 1152–1158.

Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. *27*, 491–499.

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J.M., Desrichard, A., Walsh, L.A., Postow, M.A., Wong, P., Ho, T.S., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. N. Engl. J. Med. *371*, 2189–2199.

Talevich, E., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. PLoS Comput. Biol. *12*, e1004873.

Thomsen, M.B., Nordentoft, I., Lamy, P., Høyer, S., Vang, S., Hedegaard, J., Borre, M., Jensen, J.B., Ørntoft, T.F., and Dyrskjøt, L. (2016). Spatial and temporal clonal evolution during development of metastatic urothelial carcinoma. Mol. Oncol. *10*, 1450–1460.

Turajlic, S., Xu, H., Litchfield, K., Rowan, A., Horswell, S., Chambers, T., O'Brien, T., Lopez, J.I., Watkins, T.B.K., Nicol, D., et al.; TRACERx Renal Consortium (2018). Deterministic evolutionary trajectories influence primary tumor growth: TRACERx Renal. Cell *173*, 595–610.e11.

Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et al. (2010).

Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. USA *107*, 16910–16915.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164.

Warrick, J.I., Sjodahl, G., Kaag, M., Raman, J.D., Merrill, S., Shuman, L., Chen, G., Walter, V., and Degraff, D.J. (2019). Intratumoral heterogeneity of bladder cancer by molecular subtypes and histologic variants. Eur. Urol. *75*, 18–22.

Wolf, Y., Bartok, O., Patkar, S., Eli, G.B., Cohen, S., Litchfield, K., Levy, R., Jimenez-Sanchez, A., Trabish, S., Lee, J.S., et al. (2019). UVB-induced tumor heterogeneity diminishes immune response in melanoma. Cell *179*, 219–235.e21.

Wong, S.Q., Li, J., Tan, A.Y., Vedururu, R., Pang, J.M., Do, H., Ellul, J., Doig, K., Bell, A., MacArthur, G.A., et al.; CANCER 2015 Cohort (2014). Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Med. Genomics *7*, 23.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse anti-cytokeratin 8/18 antibody	Ventana Medical Systems	cat # 760-4344; RRID:AB_10583331
Goat-anti-Mouse antibodyconjugated with Alexa Fluor 488	Invitrogen	cat # A11001; RRID:AB_2534069
Goat-anti-Mouse antibody conjugated with Alexa Fluor 647	Invitrogen	cat # A-21236; RRID:AB_141725
Chemicals, Peptides, and Recombinant Proteins		
BD FACS Aria	Bectin Dickinson	cat # 656700
CC1 buffer	Ventana Medical Systems	cat # 950-124
antibody diluent buffer	Ventana Medical Systems	cat # 251-018
autoMACS buffer	Miltenyi Biotech	cat # 130-091-221
phosphate buffered saline	Fisher Scientific	cat # 14190
Tween 20	Fisher Scientific	cat # AC233362500
DAPI	Sigma	cat #D9542
Pepsin	Sigma	cat # P7012
Proteinase K	VWR	cat # 0706
Critical Commercial Assays		
cobas cfDNA Sample Preparation Kit	Roche	7247737190
AVENIO ctDNA Enrichment Kit	Roche	8061041001
AVENIO ctDNA Library Prep Kit	Roche	8061050001
HyperCap Target Enrichment Kit, 96 Reactions	Roche	8286345001
SeqCap EZ MedExome Enrichment Kit	Roche	7681330001
NimbleGen SeqCap Hybridization and Wash Kit	Roche	5634253001
HiSeq Rapid SBS Kit v2	Illumina	FC-402-4021
HiSeq SBS Kit V4 250 cycle kit	Illumina	FC-401-4003
xGen Dual Index UMI Adapters	IDT	N/A
SeqCap EZ Prime Choice Probes- Onco	Roche	4000030990- 08247498001
SeqCap EZ Share Choice- Onco	Roche	4000007080- 08332975001
SeqCap EZ Custom Design	Roche	8332975001
SeqCap Adaptor Kit A	Roche	7141530001
HiSeq® 2500 Sequencing System	Illumina	SY-401-2501
cBot System	Illumina	SY-301-2002
20 micron cell strainer	Pluriselect	cat # 43-50020-03
IKA disposable grinding chamber	IKA-Works	cat # MT 40.100
IKA Works Tube Mill Control system	IKA-Works	cat # 0004180001
Deposited Data		
Human reference genome NCBI build 37, GRCh37	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
Rep-Seq data for cases RS1-RS12	This paper	European Phenome Genome Archive: EGAS00001004246
Software and Algorithms		
Burrows-Wheeler Aligner (BWA) v0.7.15	Li and Durbin, 2009	http://bio-bwa.sourceforge.net/
Samtools v1.3.1	Li and Durbin, 2009	http://samtools.sourceforge.net/
Picard 1.81		http://broadinstitute.github.io/picard/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mutect v1.1.7	Cibulskis et al., 2013	http://software.broadinstitute.org/cancer/cga/mutect
VarScan v2.4.1	Koboldt et al., 2009	http://varscan.sourceforge.net/
Scalpel v0.5.3	Fang et al., 2016	https://github.com/hanfang/scalpel-protocol
Annovar	Wang et al., 2010	http://annovar.openbioinformatics.org/en/latest/
CNVkit v0.7.3	Talevich et al., 2016	https://github.com/etal/cnvkit
R package 'Copynumber'	Nilsen et al., 2012	http://bioconductor.org/packages/release/bioc/html/copynumber.html
ABSOLUTE v1.0.6	Carter et al., 2012	http://software.broadinstitute.org/cancer/cga/absolute
bedtools package	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
AlleleCounter	N/A	https://github.com/cancerit/alleleCount
ASCAT	Van Loo et al., 2010	https://github.com/Crick-CancerGenomics/ascat

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Samra Turajlic (samra.turajlic@crick.ac.uk). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Clinical studies

Cases RS1, RS3 and RS4 were diagnosed with renal cell carcinoma, and were consented for research under the TRACERx Renal study (National Health Service Research Ethics Committee approval 11/LO/1996), as previously described[5]. Autopsy samples from RS1 and RS2 were obtained through the PEACE Study (NIHR 18422), where samples are harvested within ~48 hours from death (see secondary author list for the full list of PEACE consortium investigators). Residual surgical material for cases RS5, RS6, RS7, RS8, RS9, RS10, RS11 and RS12 were was obtained from commercial providers of research specimens (GLAS Consultants, Winston-Salem, NC. (IRB#: 120160685) and The MT Group, Van Nuys, CA (MTG-015)) from U.S. hospitals, under IRB approval. Sex/age of each case is as follows: RS1 M/54, RS2 M/41, RS3 M/75, RS4 F/75, RS5 F/55, RS6 F/81, RS7 F/90, RS8 M/65, RS9 M/83, RS10 F/68, RS11 F/75, RS12 (n.a.).

METHOD DETAILS

Clinical audit of current molecular profiling practices

Institutional review board approval was obtained for a service evaluation to quantify the tumor volume routinely profiled as a diagnostic standard of care in resected colorectal, melanoma and sarcoma tumors (SE725), where surgery and molecular profiling were both undertaken at the Royal Marsden NHS Foundation Trust was reviewed. Cases were included if there were > 2 macroscopic tumor dimensions recorded on the histopathology report and if information was available on the number and thickness of slides used for molecular profiling. The audit data is shown in Table S4.

Regional biopsy and cocktail sample preparation

Multi and single region biopsy sampling of surgically resected tumor tissue was conducted using the same method as previously described (Turajlic et al., 2018). Cocktail samples (as displayed in Figure 1) were created for each tumor, by pooling extracted DNA taken from all single-regions taken for each tumor, in equimolar ratios. A median of 9 single—region samples were pooled per cocktail. For the autopsy samples from RS1 and RS2, $3mm^3$ sections were dissected from snap frozen tumor core biopsies (6mm diameter) and tissue disrupted using the Tissue Raptor. They were then processed through a QIAGEN QIAshredder and DNA was purified using the QIAGEN All Prep DNA/RNA Mini kit according to manufacturer's instructions. Germline DNA was extracted from whole blood or a buffy coat.

Grossing and homogenization of residual tumor tissue

Following diagnostic histologic sampling and removal of fresh biopsies, three distinct clinical surgical waste tissues from a kidney radical nephrectomy (RS1) containing: 1) a primary tumor, 2) para-aortic lymph node cluster, and 3) renal hilar nodes were fixed in

10% neutral buffered formalin for 24 h to mimic the standard clinical workflow. After fixation, samples were exchanged into phosphate-buffered saline (PBS, 14190, Fisher Scientific, USA) for 24 h, then stored in ethanol for until dissected. Tumor tissue was identified by a pathologist through macroscopic evaluation and physical palpation, and all identifiable tumor was dissected away from the surrounding tissue. An area of normal tissue (at least 5 cm from the tumor) was also dissected by a pathologist and retained. Lymph nodes detected during gross examination of the RS1 specimen included hilar and peri-renal nodes were also dissected as an independent tissue samples. All dissected tissue was weighed prior to homogenization.

Residual primary tumor tissue was split into two-625 g portions; each portion was combined with 600 mL autoMACS Running Buffer (Miltenyi Biotec Inc., 130-091-221) and homogenized in a liquidizer (Cookworks, BL9292E-GS) for 3 min at the highest setting. The primary tumor homogenates (2.5 l total) were manually combined and mixed in a plastic container, divided back and re-liquidized for additional homogenization and mixing, and pooled together into a large plastic container. Segregated tissues of renal hilar lymph nodes, peri-renal node, normal kidney tissue, and para-aortic lymph nodes were each homogenized independently in autoMACS Running Buffer (1:1, mass: volume) with an IKA Tube Mill (IKA Works Inc. Wilmington North Carolina, 0004180001) for 2 min at 15,000 rpm using single-use blending containers. When tissue mass exceeded the capacity of an individual blender container, homogenates for the same sample were pooled by mixing as described above. Samples of each tissue homogenate were stored in methanol (1:1, v:v) at 4°C. Case RS2 was lymph-node material from a patient with primary melanoma, and cases RS3 and RS4 were primary renal tumors, and all were processed through the same homogenization protocol as RS1. For RS2 primary melanoma tissue, the FFPE blocks were trimmed by razor-blade to remove excess paraffin wax that did not contain any tissue. Remaining wax embedded tissue was removed from FFPE plastic cassette and minced into approximately 1mm cubed pieces using razor blade. Tissue pieces were deparaffinised by 5 sequential 1.5 hour 52 degrees' Celsius xylene washes until all but trace wax was removed and only tissue remained. Xylene was removed from the sample using a room temperature 10 minute acetone wash followed by three sequential 20 minute 100% ethanol washes. Ethanol was removed, and the sample hydrated for blending, in two 10 minute sequential washes of 1x PBS.

For cases RS5 to RS12, each specimen had been subjected to standard sampling for diagnosis and staging purposes. These cases were considered surgical waste and slated for incineration, thus were stored in formalin for four to six weeks. Upon arrival, tissue was transferred to PBS for 12-24h. Tumor tissue was identified by a pathologist through macroscopic evaluation and physical palpation, and all identifiable tumor was dissected away from the surrounding tissue. An area of normal tissue (at least 5 cm away from the tumor) was also dissected by a pathologist and retained. All dissected tissue was weighed prior to homogenization. Dissected tumor and normal tissue were homogenized separately in single-use blender containers (IKA Works Inc. Wilmington North Carolina, 0004180001), or single use consumer grade blenders (Hamilton Beach, 51102, Glen Allen, VA) in autoMACS buffer (1:1, m:v) for 2 min at 15,000 rpm or at the highest setting. Resulting homogenates were stored at 4°C until further processed.

Genomic DNA purification from tissue and cfDNA

An aliquot of each tissue homogenate (1200 µl) was collected by centrifugation at 5000 r*cf.* for 2 min, rinsed with TE buffer pH 8.0 (VWR, AAJ62745-EQE) twice, and incubated in 5 mL protease digestion buffer [9.75 mL TE buffer pH 8.0, 60 mg Proteinase K (VWR, 0706), and 0.25 mL 20% SDS in aqueous solution (Amresco, 0837)] at 56°C for 2-16 h. Digested tissue (100 µl) was used for genomic DNA purification by High Pure PCR Purification Kit (Roche Applied Sciences, Mannheim Germany, 11 732 668 001) according to manufacturer's protocol. Purified genomic DNA was quantified using a NanoDrop 8000 (Thermo Fisher Scientific) and stored at −20°C. cfDNA was isolated from plasma using cobas cfDNA Sample Preparation kit (Roche, 07247737190).

Target-enriched NGS Library construction and sequencing

Illumina compatible indexed NGS libraries were constructed from genomic DNA from tissue using the SeqCap EZ HyperCap Workflow User's Guide, v1.0 (Roche Sequencing Solutions) with notable parameters specified below. Briefly, 1 µg of purified genomic DNA was enzymatically fragmented for 33-40 min at 37°C and prepared for adaptor ligation using the KAPA HyperPlus library prep kit according to manufacturer's instructions (Roche Sequencing Solutions, KK8514). SeqCap sequencing adaptor final reaction concentration was 2 µM, and adaptor ligation reaction time was extended to 14-18 h, at 16°C. No pre-capture PCR was used following ligation reaction purification. SeqCap EZ library probe baits for either MedExome (07681330001), Onco_EZ (08333076001), or a custom RS1 specific (see data analysis section below for design criteria) target-enrichment panel (Roche Sequencing Solutions) and 2 nM blocking oligos (Roche Sequencing Solutions), were incubated for 18-22 h at 47°C following manufacturer instructions. Post-capture PCR was performed using KAPA HiFi HotStart ReadyMix and LM-PCR oligos for 14 cycles. Post-capture purified library concentrations were determined by Qubit (ThermoFisher) and fragment size distribution analyzed by Bioanalyzer 2100 (Agilent). Amplified enriched libraries were each diluted to 2nM and stored at −20°C prior to pooling for sequencing. Pooled libraries were sequenced using MiSeq and HiSeq instruments (Illumina) according to manufacturer's recommendations for paired-end sequencing using (Illumina) runs with 101 base paired-end reads. cfDNA sequencing libraries were constructed using the AVENIO ctDNA Targeted Kit (Roche, 08061076001) by following the AVENIO ctDNA Analysis Kits Reagent Workflow User Guide v1.0.0. Amplified, adaptor-ligated samples were concentrated together with the Hybridization Supplement using a Vacufuge plus instrument (Eppendorf). Each sample was resuspended in the appropriate Enhancing Oligo, the custom RS1-specific panel, and Hybridization master mix. Enrichment, hybridization cleanup and amplification were performed according to manufacturer instructions. Samples (equal mass) were pooled, and sequenced using Hiseq (Illumina), according to instructions, with 151 base paired-end reads. Multi-region, cocktail and single regions samples, from 79 renal cell carcinomas as displayed in Figure 1, underwent renal driver

Panel_v6 library preparation and sequencing, using methods as previously described (Turajlic et al., 2018). Multi-region whole exome library preparation and sequencing for RS1 was conducted by external laboratory (Eurofins Scientific), using Agilent SureSelect Human All Exon v5 kits. Multi-region whole exome library preparation for RS2 was conducted at the UCL Pathogen Genomics Unit, using Agilent SureSelect Human All Exon v5 kits, and sequenced by external laboratory (GENWIZ). Sequencing coverage metrics, and library kits, for all samples profiled is provided in Table S5.

Flow sorting method to increase tumor purity

A representative sample from formalin fixed residual tumor tissue from RS12 was generated by homogenization in an IKA blender in autoMACS buffer (1:1 mass to volume). Aliquots of the homogenate (1 g) were further dissociated to individual nuclei by adapting a previously described method (Hedley et al., 1983). Briefly, tissue was collected by centrifugation, resuspended in CC1 (Ventana Medical Systems, Tucson, AZ) buffer (5:1 mass to volume), and heated at 80°C for 30 min. Tissue was washed once with PBS, and resuspended in PBS containing 1 mg/ml proteinase K (1:1 mass to volume) (VWR USA) and incubated at 50°C for 10 min. The sample was exchanged into 5 mg/ml pepsin in 150 mM NaCl, pH 1.5 (Sigma, USA) and incubated 30 min at 37°C. The sample was adjusted to pH 8 with 5 M NaOH, and exchanged into PBS, 0.5% BSA and 0.5% Tween 20 (Fisher Scientific, USA (AC233362500) prior to filtration through a 20-μm filter (Pluriselect, San Diego, CA) to collect nuclei.

Nuclei were then collected by centrifugation at 400 x g and exchanged into antibody diluent for 30 min at 20°C. Samples were exchanged into mouse anti-cytokeratin 8/18 primary antibody (Ventana Medical Systems, Tucson, AZ) for 1 hour at 4°C, washed three times in 0.5 mL PBS, 0.1% BSA and 0.1% Tween 20, and incubated for 30 min at 4°C in goat-anti-mouse antibodies conjugated to Alexa Fluor 488 or Alexa Fluor 647 (2 μg/ml) (Invitrogen, San Diego, CA) and DAPI (3 μM) (Sigma, USA). Stained samples were washed and filtered prior to analysis and sorting using a BD FACS Aria (656700, Bectin Dickinson) equipped with a 355 nm, 60 mW laser and 450/50 nm filter for DAPI; 488 nm, 60 mW laser and 530/30 nm filter for AF 488; and 633 nm, 100 mW laser and 670/30 nm filter for AF 647. No compensation was used. DAPI was used for doublet discrimination. RS11 tumor nuclei were enriched by FACS after gating to include cytokeratin positive (CK+), high side-scatter (SSC) nuclei and exclude cytokeratin negative (CK-), low SSC nuclei. The flow cytometry gating strategy is shown in (Figure S5).

QUANTIFICATION AND STATISTICAL ANALYSIS

Tumor volume sampling analysis

For the the clinical audit data, all samples had data on width (W) and length (L) dimensions available, and tumor volume (T_V) was estimated using the following formula:

$$T_V = (W^2 \times L)/2$$

(taken from the literature as the most accurate tumor volume measurement approach (Faustino-Rocha et al., 2013)).

Biopsy volume (B_V) was calculated based on the 2D surface area analysis of 8 typical slides, with each slide scanned using the Aperio AT2 whole slide scanner at 40x. Each image was annotated by hand, following the perimeter of the tissue, and the surface area calculated via using the Aperio ImageScope software. The average surface area was 3.37cm^2 and this value was multiplied by slide thickness (10 μm), and the total number of slides used, to obtain B_V estimates per tumor. We note that in cases where multiple slides were used for molecular profiling, (up to 5 were used), each slide was taken from the same block (i.e., all from one fixed spatial location). The proportion of total tumor volume sampled in each case is then simply calculated as B_V / T_V. For the cancer genome atlas (TCGA) dataset analysis, we extracted summary clinical annotation files for each solid tumor cohort from the Broad Institute TCGA GDAC Firehose repository. Tumor dimension data was available for n = 1667 samples, across 6 tumor types: ACC, KICH, KIRC, KIRP, PAAD and THCA. Tumor volume (T_V) was calculated as per above using the formula: $T_V = (W^2 \times L) / 2$. In cases where only one dimension was given (i.e., the maximal dimension) this was assumed to be the tumor length, and the tumor width was estimated using a L:W ratio of 1:0.8, with the 0.8 standard value estimated as the median ratio value observed across all cases with available length and width data. Biopsy sample volumes were calculated from exact length (L), width (W) and depth (D) dimensions, as given in the clinical annotation files, with biopsy shape assumed to be cuboid and biopsy volume (B_V) calculated as $B_V = L \times W \times D$. Where biopsy dimensions were missing in the clinical annotation files, a standard biopsy volume (B_V) of 0.48cm^3 was assumed, based on the median value from all other bases where data was available. The proportion of total tumor volume sampled in each case is then simply calculated as B_V / T_V. In Figure 1A, significance was assessed with a Wilcoxon rank-sum test. In Figure S1 the intratumor heterogeneity scores, for cases overlapping with the proportion of tumor sampled dataset, were taken from (Raynaud et al., 2018) (number of clones score). Purity estimates, again for TCGA cases overlapping with the proportion of tumor volume sampled analysis, were taken from (Aran et al., 2015) (consensus purity estimates).

Processing of sequencing data

Paired-end reads in FastQ format sequenced by MiSeq and Hiseq were aligned to the reference human genome (build hg19), using the Burrows-Wheeler Aligner (BWA) v0.7.15. with seed recurrences (-c flag) set to 10000 (Li and Durbin, 2009). Intermediate processing of Sam files was performed using Samtools v1.3.1, deduplication was performed using Picard 1.81 (http://broadinstitute.github.

io/picard/) and local indel realignment with GATK v3.6. For whole exome and renal driver Panel_v6 sequencing datasets, single Nucleotide Variant (SNV) calling was performed using Mutect v1.1.7 and small scale insetion/deletions (INDELs) were called running VarScan v2.4.1 in somatic mode with a minimum variant frequency (–min-var-freq) of 0.005, a tumor purity estimate (–tumor-purity) of 0.75 and then validated using Scalpel v0.5.3 (scalpel-discovery in - -somatic mode) (intersection between two callers taken) (Cibulskis et al., 2013, Fang et al., 2016, Koboldt et al., 2009). SNVs called by Mutect were further filtered using the following criteria: i) variant allele frequency (VAF) $\leq$ 1% in the corresponding germline sample, ii) variants that falling into mitochondrial chromosome, haplotype chromosome, HLA genes or any intergenic region were not considered, iii) presence of both forward and reverse strand reads supporting the variant. For custom RS1 panel sequencing data, sequencing was conducted at high depth using unique molecular barcode (UMI) indexes, and UMI-tools (Smith et al., 2017) was used to group PCR duplicates and de-duplicate reads to yield one read per group. SNVs were then called using deepSNV (Gerstung et al., 2012), as Mutect is known to not be calibrated for higher sequencing depth levels. Varscan and Scalpel were used to call RS1 custom panel INDELs as described above. All variants were annotated using Annovar (Wang et al., 2010). To estimate somatic copy number alterations, CNVkit v0.7.3 was performed with default parameter on paired tumor-normal sequencing data (Talevich et al., 2016). Outliers of the derived logR calls from CNVkit were detected and modified using Median Absolute Deviation Winsorization before case-specific joint segmentation to identify genomic segments of constant logR (Nilsen et al., 2012). Tumor sample purity, ploidy and absolute copy number per segment were estimated using ABSOLUTE v1.0.6 (Carter et al., 2012). Neoantigen predictions were derived by first determining the 4-digit HLA type for each patient, along with mutations in class I HLA genes, using POLYSOLVER (Shukla et al., 2015). Next, all possible 9, 10 and 11-mer mutant peptides were computed, based on the detected somatic non-synonymous SNV and INDEL mutations in each sample. Binding affinities of mutant and corresponding wild-type peptides, relevant to the corresponding POLYSOLVER-inferred HLA alleles, were predicted using NetMHCpan (v3.0) and NetMHC (v4.0) (Andreatta and Nielsen, 2016). Neoantigen binders were defined as $IC_{50} < 50$ nM or rank < 2.0. Signature analysis was conducted on all non-synonymous mutations using package deconstructSigs (Rosenthal et al., 2016). We additionally checked for evidence of formalin induced artifact variants in the Rep-Seq data, given the protocol involves formalin exposed material. Formalin fixed paraffin embedded (FFPE) samples can contain artifacts, typically arising due to hydrolytic deamination of cytosine to form uracil, or thymine if the cytosine is methylated. Such artifacts are normally visible as an excess of C > T/G > A mutations at lower variant allele frequency (Wong et al., 2014). Analysis of this in the whole exome sequencing data from Rep-Seq cases showed no evidence of excess low frequency formalin induced artifact, with the proportion of low frequency (below 5% VAF) C > T mutations being 34.0%, closely comparable to the average across all base changes (33.3%) (Figure S5).

Analysis of pooled cocktail sequencing data

The final set of cocktail samples included 79 tumors with matched processed reference datasets from multi-region sequencing. The number of biopsies per cocktail sample ranged from 2 to 75 with a median number of 8 biopsies per tumor and a total number of 1,184 individual biopsies. As a reference dataset of true variants, we used previously published multi-region sequencing variant calls from the same cases, which represented the sum of all variants detected in each tumor (Turajlic et al., 2018). In our analysis we compared the overall performance of single-region and cocktail sequencing, in detecting somatic mutations from the known truth set. The single region sample was selected as one random single-region biopsy per tumor, from the overall multi-region dataset. We first evaluated the performance of the cocktail sequencing approach compared to multi-region and single region sequencing, by comparing the number of somatic variants detected per tumor with each approach. To reflect the average performance of single region biopsies, we calculated the mean number of variants detected through single-region sequencing per tumor. We next determined the detection rates of true variants in the cocktail and the single-region samples using the multi-region sequencing data as a reference. Significance was assessed with a paired Wilcoxon Test. Finally, in order to establish the accuracy of the cocktail sequencing approach, we next determined the correlation between the variant allele frequencies (VAF) of all somatic mutations detected through multi-region sequencing and the VAFs from the cocktail samples as well as a randomly selected single-region biopsy per tumor. The multi-region VAFs were calculated as the mean VAFs across all regions included in the cocktails. The correlations were calculated with a Spearman's rank-order correlation test.

Analysis of in-silico pooled cocktail sequencing data

For the analysis presented in Figures 1E and 1F, we made use of two additional multi-region sequencing datasets: i) the TRACERx100 non-small cell lung cancer dataset (Jamal-Hanjani et al., 2017) and ii) a cohort of urothelial carcinomas (Lamy et al., 2016, Thomsen et al., 2016). Each multi-region sample had undergone whole exome sequencing, as described previously described (Lamy et al., 2016, Thomsen et al., 2016, Jamal-Hanjani et al., 2017). Sequencing reads were reverted back to fastq format using bam2fastq, and alignment/variant calling was conducted as described above. The *in-silico* cocktail samples were created for each case by down-sampling the reads in each biopsy sample using picard-tools "DownsampleSam," and then remerging the down-sampled BAM files to make a merged *in-silico* cocktail sample. The down-sampling proportion was selected to ensure the merged *in-silico* cocktail sample had equivalent coverage to a single biopsy samples for that tumor (e.g., if a tumor had five different biopsy samples sequenced, a proportion of 0.2 of each BAM was taken to create a merged file of equivalent depth to a single biopsy). For each dataset, TMB was measured as the number of missense variants per tumor, with a high TMB threshold of 200 missense mutations used, which prior analysis has shown is equivalent to the 10 mutations/Mb threshold (synonymous and non-synonymous) derived from

higher coverage panel datasets (Budczies et al., 2019). The TMB values were calculated using three different methods: A) Single biopsy-TMB, based on the number of mutations in each individual biopsy sample. B) clonal-TMB, based on the full multi-region data for each tumor, with mutations being judged as clonal by simple ubiquity across all tumor regions. C) Cocktail-TMB, based on the number of predicted clonal mutations within the *in-silico* cocktail, with clonality determined by clustering variant allele frequency values using R function daisy, with the distance matrix computed using Gower's formula, and the number of clusters set to two (such that mutations were grouped into either clonal/subclonal groups).

Custom panel design

To conduct in-depth validation of the representative sequencing method high coverage profiling was conducted in case RS1, using a custom panel. The panel design was based on whole exome sequencing results from: i) 7 biopsies taken from the RS1 primary tumor (before homogenization) and ii) an aliquot of the RS1 homogenized solution. SNV and INDEL mutations were called across the 8 samples as described above, and a total of 76 non-synonymous mutations were detected (Table S6). These 76 mutations were successfully captured in a targeted custom panel, and sequenced to high depth (median > 10,000x) in the 64 primary biopsies, 11 biopsies taken from 2 lymph node metastases, 4 biological primary Rep-Seq replicates, 6 circulating tumor (ct) DNA samples collected at different time points, and 3 homogenized lymph node Rep-Seq samples. As an additional validation, all 76 RS1 variants were also validated using Ion Torrent sequencing, as orthogonal confirmation.

Jaccard reproducibility analysis

The reproducibility of variant discovery between RS1 tumor biopsies, Rep-Seq biological replicates and ctDNA samples was assessed using the Jaccard similarity coefficient. Each pairwise combination between samples (within each group) was considered, e.g., Biopsy1 (A) versus Biopsy2 (B), Biopsy1 (A) versus Biopsy3 (B), etc. Jaccard similarity coefficient was calculated using the standard formula (J):

$$J = M_{11}/(M_{01} + M_{10} + M_{11})$$

where M_{11} represents the total number of variants present in both samples A and B, M_{10} represents the total number of variants present in A but not B and M_{01} represents the total number of variants present in B but not A.

Clustering and phylogenetic analysis

Clustering analysis was performed on cases RS1 (custom panel data) and RS2 (whole exome data) using PyClone Dirichlet process clustering (Roth et al., 2014). For each mutation, the observed alternative allele count, reference count and total local tumor copy number was used as input, together with the purity for each sample. PyClone was run with 10,000 iterations and a burn-in of 1000, and default parameters, with–var_prior total_copy_number. For RS1, two separate PyClone runs were conducted, the first for the primary multi-region biopsies dataset. Of the total n = 64 primary biopsies sequenced, n = 52 passed quality control for clustering analysis, with n = 12 biopsies excluded due to lower purity (measured based on purity being too low to call the known clonal 3p copy number loss event correctly). The second RS1 PyClone clustering run was conducted just for Rep-Seq homogenate sample alone (n = 1), using the same parameters. Similarly, three separate PyClone runs were conducted for case RS2, for single biopsy LN sample (alone), Rep-Seq LN sample (alone) and then all biopy samples (from LN and eight other sites of metastases).

Illusion of clonality simulation

To assess the risk of illusion of clonality, a biopsy sampling approach was simulated, for 1 up to 20 biopsies taken, using the RS1 datatset. For each biopsy number (n = 1-20), a random sample of biopsies of size n was drawn from the total set of 64 primary biopsies profiled for RS1. Within the random sampled set, the number of mutations which appeared to be clonal (based on being ubiquitously present in all biopsies in the sampled set) was calculated. This list was then compared to the known list of truly clonal mutations (from the full 64 set), and percentage of variants which were incorrectly classified as clonal was recorded. This process was repeated for 100 iterations for each n, to give a distribution, from which mean and standard deviation values were calculated.

Analysis of purity enriched data

For RS12, whole exome sequencing case conducted using the standard Rep-Seq protocol, and then repeated with the additional step of flow sorted purity enrichment. Variant calling was completed, and purity estimates calculated in both samples in the same way, as detailed above. The number of variants discovered in each sample, and then those in common across samples were calculated and plotted in Figure 5.

Statistical methods

Statistical package R v3.3.2 or higher was used for all analysis. All statistical tests were two-sided. The tests used for each analysis are listed above in the relevant methods section.

DATA AND CODE AVAILABILITY

Sequencing data that supports this study has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the European Bioinformatics Institute (EBI); under accession number EGAS00001004246. The TRACERx lung exome sequence data used during the study is available through the Cancer Research UK & University College London Cancer Trials Centre (ctc.tracerx@ucl.ac.uk) for non-commercial research purposes, and access will be granted upon review of a project proposal that will be evaluated by a TRACERx data access committee and entering into an appropriate data access agreement subject to any applicable ethical approvals.

SCIENTIFIC REPORTS
nature research

OPEN

Somatic substitution signature as an innovative tool in lung cancer diagnosis

Received: 22 May 2019
Accepted: 18 September 2019
Published online: 10 October 2019

Stéphane Busca[1], Julia Salleron[2], Romain Boidot[3,4], Jean-Louis Merlin[1,5] & Alexandre Harlé[1,5]

Diagnosis of lung cancer can sometimes be challenging and is of major interest since effective molecular-guided therapies are available. Compounds of tobacco smoke may generate a specific substitutional signature in lung, which is the most exposed organ. To predict whether a tumor is of lung origin or not, we developed and validated the EASILUNG (Exome And SIgnature LUNG) test based on the relative frequencies of somatic substitutions on coding non-transcribed DNA strands from whole-exome sequenced tumors. Data from 7,796 frozen tumor samples (prior to any treatment) from 32 TCGA solid cancer groups were used for its development. External validation was carried out on a local dataset of 196 consecutive routine exome results. Eight out of the 12 classes of substitutions were required to compute the EASILUNG signature that demonstrated good calibration and good discriminative power with a sensitivity of 83% and a specificity of 72% after recalibration on the external validation dataset. This innovative test may be helpful in medical decision-making in patients with unknown primary tumors potentially of lung origin and in the diagnosis of lung cancer in smokers.

Current cancer therapies are generally based on anatomical site. Identifying a tumor's origin is therefore crucial for treatment selection. Lung cancer is the most common major cancer diagnosed worldwide[1] and is mostly represented by two subtypes, small cell lung carcinoma and non-small cell lung carcinoma (NSCLC), accounting for 15% and 85% of all lung cancer, respectively[2]. Squamous cell carcinoma (SCC) represents 25 to 30% of lung cancers and adenocarcinoma is the most common histological subtype of lung cancer, representing 40% of all lung cancers. A cancer diagnosed at a metastatic stage with no identifiable primary tumor is termed a cancer of unknown primary origin (CUP). A diagnosis of CUP is evocated when the clinical picture of metastatic disease is associated with one or more tissue biopsy results that are inconsistent with a determined primary tumor. Two to four percent of newly diagnosed cases of carcinoma are CUP[3]; histological subtypes of these CUP are adenocarcinomas (40–60%), undifferentiated carcinomas (15–30%), and SCCs (15–20%)[4]. Patients with CUP syndrome have a median survival of 8–11 months and only 25% survive at 1 year[5].

The Human Genome Variation Society (HGVS) nomenclature defines 12 different somatic substitution classes (C > A, G > T, C > G, G > C, C > T, G > A, A > C, T > G, A > G, T > C, A > T, and T > A) that can be read on the coding non-transcribed DNA strand. Some childhood cancers, like acute myeloid leukemia (AML), are reported to carry very few somatic substitutions, whereas cancers such as lung carcinomas and malignant melanoma, which are related to chronic exposure to mutagens such as tobacco or ultraviolet light, present with a high number of somatic substitutions[6]. The prevalence of somatic substitutions, quantified in substitutions per megabase (sub/Mb), can be calculated by dividing the number of somatic substitutions by the size of the DNA sequence targeted region in Mb; this is also known as the tumor mutational burden. A mutational signature is a characteristic combination of mutation types arising from specific mutagenesis processes, such as DNA replication infidelity, exposure to exogenous and endogenous genotoxins, defective DNA repair pathways, and DNA enzymatic editing[7]. Tissues directly exposed to tobacco smoke (*e.g.*, lung), as well as some tissues not directly exposed (*e.g.*, bladder), show elevated levels of DNA adducts in smokers and thus evidence of exposure to carcinogenic components

[1]Service de Biopathologie, Institut de Cancérologie de Lorraine, F-54519, Vandœuvre-lès-Nancy, France. [2]Data Biostatistics Unit, Institut de Cancérologie de Lorraine, F-54519, Vandœuvre-lès-Nancy, France. [3]Department of Tumor Biology and Pathology, Georges-Francois Leclerc Cancer Center – UNICANCER, Dijon, France. [4]Institut National de la Santé et de la Recherche Médicale (Inserm), LNC UMR1231, Dijon, France. [5]Université de Lorraine CNRS UMR 7039 CRAN, F-54519, Vandœuvre-lès-Nancy, France. Correspondence and requests for materials should be addressed to A.H. (email: a.harle@nancy.unicancer.fr)

of tobacco smoke[8,9].Benzo[a]pyrene (BaP) present in tobacco smoke is the most extensively studied compound. After two successive oxidation reactions mediated by cytochrome P450 enzymes (largely CYP1A1), BaP is converted to benzo[a]pyrenediolepoxide (BPDE). BPDE is considered an ultimate carcinogen because, unlike its BaP precursor, it has the ability to directly attack and form covalent adducts with DNA bases, which may then generate G > T/C > A somatic substitutions. Importantly, exposure of a cell to polycyclic aromatic hydrocarbons like BaP and other compounds of tobacco smoke increases synthesis of CYP1A1, thereby accelerating these reactions in cells[10,11], and may generate a specific substitutional signature in lung, which is the most exposed organ.

The aim of our study was to develop and validate an innovative genomics-based test, the EASILUNG (Exome And SIgnature LUNG) test, able to predict whether or not a tumor is of lung origin using only the prevalence and the relative frequencies of somatic substitutions from whole-exome sequencing (WES) results.

Methods

Development dataset. We used the full publicly available TCGA data where somatic mutations are provided as simple somatic mutations (SSM) from WES results. SSM consists of somatic substitutions, deletions of ≤200 bp, insertions of ≤200 bp, and multiple base substitutions (≥2 bp and ≤200 bp). Mutation annotation format (MAF) files containing open access SSM data of the cancer genomes used in this study were obtained from all 33 TCGA groups (February 1, 2016). SSM data from all genome sequencing center (GSC) platforms available were downloaded, corresponding to 130 MAF files containing somatic mutation data from 28,437 WES results from 9,522 distinct tumors sequenced by one to seven GSCs. All TCGA data were generated from frozen tumor samples prior to any treatment. The AML group (2 MAF files containing 394 WES results from 197 distinct samples), WES results with any substitutions (35 WES results), and MAF files that did not include the size of the targeted region of the exome (76 MAF files containing 16,514 WES results) were excluded from the development dataset; these conditions were not mutually exclusive. All WES results from all TCGA solid cancer groups (32/33 TCGA groups) from all GSC platforms that included the size of the targeted region of the exome in MAF files or in previously related TCGA publications were evaluated for the development dataset[12–29], corresponding to 54 of the 130 MAF files containing somatic mutations in 11,720 WES results from 7,796 of the 9,342 distinct tumors. Each TCGA tumor had a unique barcode; if several GSCs had sequenced the same tumor and there were multiple WES results for the same tumor, a randomization was carried out to select only one WES result for each tumor. In total, the development dataset included 7,796 WES results from 7,796 distinct tumor samples (745 lung cancer and 7,051 non-lung cancer). All details are provided in Supplementary Methods.

Substitution classes and prevalence of somatic substitutions. Somatic substitutions were analyzed on the coding non-transcribed strand and their absolute frequencies were classified into the 12 classes according to the HGVS nomenclature (C > A, G > T, C > G, G > C, C > T, G > A, A > C, T > G, A > G, T > C, A > T and T > A). Targeted region size (in Mb) and absolute frequencies of the 12 somatic substitution classes read on the coding non-transcribed strand were collected. Insertion, deletion, and multiple base substitutions were not considered. The total number of somatic substitutions was obtained by adding absolute frequencies of the 12 somatic substitution classes. Prevalence of somatic substitutions for each tumor was quantified in sub/Mb and was calculated by dividing the total number of somatic substitutions by the size in Mb of the exome targeted region, which varied between 36 to 64 Mb across TCGA GSCs and groups of solid cancers (Supplementary Methods).

Relative frequencies of substitution classes. Relative frequencies of somatic substitutions for each class (%CA, %GT, %CG, %GC, %CT, %GA, %AC, %TG, %AG, %TC, %AT, and %TA) were determined as the absolute frequency of each somatic substitution class (respectively C > A, G > T, C > G, G > C, C > T, G > A, A > C, T > G, A > G, T > C, A > T, and T > A) divided by the total number of somatic substitutions.

External validation dataset. A total of 196 consecutive tumor samples sequenced between January 2015 and August 2017 available from the Molecular Tumor Board of Centre Georges-François Leclerc (CGFL, Dijon, France) was used for the external validation dataset. Informed consent was obtained for each patient and the study was approved by the Centre Georges-François Leclerc ethical board. All methods were performed in accordance with the relevant guidelines and regulations.

All tumors had WES results and matched histopathological characteristics. All tumors and paired germline DNA were sequenced in 2 x 150 cycles on NextSeq. 500 using the Agilent SureSelect XT Human All exon v5 kit (targeted region = 50 Mb). All samples were FFPE or snap-frozen human tumors (primary or metastasis) from different histological classes and localizations. Initial site and diagnoses were determined using histopathological information. For all samples, normal DNA from the same individuals had also been sequenced to establish the somatic origin of variants and exclude germline single nucleotide polymorphisms. For each tumor, a file containing two columns with the allele of reference and mutated allele for each single-base substitution read in the coding non-transcribed strand was sent to our institution (Institut de Cancérologie de Lorraine, France), blinded of all other parameters.

Identification of the predictive model and establishing the EASILUNG signature. Patients with a somatic substitution prevalence of <1 sub/Mb were classified as having non-lung cancer. For patients with a somatic substitution prevalence of ≥1 sub/Mb, the EASILUNG signature was computed using the 12 substitution classes. They were described by the median and the interquartile range, since the distributions of these parameters were not normal according to the Kolmogorov-Smirnov test. The log-linearity assumption of the logistic model was checked. This was carried out by categorizing each classes of substitution into 10 groups (corresponding to deciles) and by looking at the plot of the logit of observed percentages of lung cancer in each class. When this assumption was not verified, the continuous parameter was transformed into binary variables. This was performed by using the AUC. The threshold maximizing the Youden index was chosen.

A multivariable logistic regression was performed with the 12 substitution classes as independent parameters and the lung cancer as the dependent parameter (full model). The simplification of this full model to avoid overfitting was done using a multivariable logistic regression with backward selection at the level $p = 0.2$. The stability of the selected model was investigated using the bootstrap resampling method[30]. A multivariable logistic regression was performed using the substitution classes selected by the bootstrap resampling method (final model). Results of the final model were expressed as Odds ratio and 95% confidence interval. Odds ratio correspond to the exponential function of the regression coefficient ($e\beta$). These regression coefficients (β) were multiplied by 100 and rounded to the nearest integer to obtain the EASILUNG score points. These integers were then added to obtain an overall lung cancer signature for each patient, namely the linear predictor. The model's discriminative performance was assessed by the AUC, and the calibration by the Hosmer-Lemeshow chi-square test. The threshold of the predicted risk score maximizing the Youden index was chosen and the sensitivity and the specificity with their 95% confidence intervals were computed.

The overall performance of the classification was assessed by taking into account the patients with a somatic substitution prevalence of <1 sub/Mb by classifying them as having non-lung cancer, because reports of lung cancers with a <1 sub/Mb prevalence are rare in smokers[6]. The sensitivity and the specificity with their 95% confidence intervals were computed on the total population. The sensitivity corresponded to the number of cases with a score greater than the pre-specified threshold divided by the number of lung cancers. The specificity corresponded to the number of cases with a somatic substitution prevalence of <1 sub/Mb plus the number of cases with a score less than pre-specified threshold divided by the number of non-lung cancer.

External validation. The EASILUNG score was computed on the validation set. The discriminative performance of the EASILUNG test on the validation set was assessed by the sensitivity and the specificity using the threshold identified with the development dataset. The calibration of the EASILUNG signature was assessed by computing the logistic regression coefficients (the intercept A and the slope B of the regression) using a logistic regression with the presence of lung cancer as dependent variable and unmodified EASILUNG signature as independent parameter. Perfect predictions correspond to an intercept A of 0 and slope B of 1. If this condition is not satisfied, a recalibration has to be performed by computed "$A + B \times$ EASILUNG score" giving a recalibrated EASILUNG score[31]. After the recalibration, the discriminative performance was assessed by the AUC and the calibration by the Hosmer-Lemeshow chi-square test. A new threshold maximizing the Youden index was chosen and the sensitivity and the specificity with their 95% confidence intervals were computed.

Results

Score development. The development dataset included 7,796 WES results from solid tumors from The Cancer Genome Atlas (TCGA). Among the 7,796 samples, 745 (9.56%) were tumors where the primary origin was lung, from the lung adenocarcinoma group ([LUAD]) or the lung SCC group ([LUSC]) and 7,051 (90.44%) were from the 30 other non-lung primary location groups (Supplementary Table 1). Among these 7,796 samples, 2,888 samples had a somatic substitution prevalence of <1 sub/Mb, including 43 lung cancer cases (1.5%). We considered one lung tumor with a somatic substitution prevalence lower than 1 sub/Mb to be non-lung cancer (because reports of lung cancers with a <1 sub/Mb prevalence are rare). Table 1 describes the 12 substitution classes for the 4,908 tumors with a somatic substitution prevalence of ≥1 sub/Mb, including 702 (14.3%) lung tumors and 4,206 (85.70%) non-lung tumors. The bivariate analyses demonstrated that all classes of substitutions were predictive of lung cancer (Table 2). A relative frequency of C > A (%CA) greater than 12% or of G > T (%GT) greater than 13% were associated with a greater risk of lung cancer (respectively with an odds ratio and 95%CI of 14.44 [11.95;17.46] and 9.17 [7.66;10.98]) whereas the presence of C > G or G > C were associated with a lower risk of lung cancer (respectively 0.45 [0.38;0.54] and 0.50 [0.42;0.60]). The full multivariate model to predict lung cancer including the 12 substitution classes is reported in Table 2.

After backward selection with bootstrap validation, eight classes of substitutions were significantly associated with lung cancer and were retained for the EASILUNG test (Table 3). Thus, the final score is defined as Score = +175(if %CA > 12) + 62(if %GT > 13) − 9*(%CT) − 6*(%GA) − 17*(%AC) − 74(if %AG > 5) −73(if %TC > 5). Table 4 details the signature based on these eight classes of substitutions with two examples of how it was calculated. To illustrate the application of the EASILUNG test (Example 1, Table 4), when we considered a tumor with %CA = 27, %GT = 19, %CT = 8, %GA = 12, %AC = 0, %TG = 0, %AG = 3, and %TC = 5, the EASILUNG score was 93 and the estimated predicted probability that the tumor was lung cancer was 0.89, according to Fig. 1.

EASILUNG test performance on an TCGA development dataset. The median EASILUNG score was −485 (IQR from −582 to −353). The score was significantly higher for lung cancer: 0 (from −139 to 2) vs −511 (from −597 to −411) for non-lung cancer ($p < 0.0001$). Area under the ROC curve (AUC) of the final model was 0.931 [95%CI 0.920; 0.942]. Descriptive characteristics of the EASILUNG score on the development dataset and the probability of lung cancer are provided in Fig. 1. The calibration assessed by the Hosmer-Lemeshow chi-square test was equal to 10.89 ($p = 0.2079$). The best EASILUNG score threshold to separate lung and non-lung cancer was set at −303: a score greater than −303 favored lung cancer. The associated sensitivity and specificity were respectively 85.7% [83.2%; 88.3%] and 89.6% [88.7%; 90.6%].

Among the 7,796 samples, when we considered the 2,888 samples with a somatic substitution prevalence of <1 sub/Mb to be non-lung cancer, the respective overall sensitivity and overall specificity of the EASILUNG-test were 80.0% [77.1%; 82.8%] (596/745) and 94.3% [93.7%; 94.8%] (6,646/7,051).

Performance of the EASILUNG test thresholded at −303 on the development dataset are presented for lung adenocarcinoma and lung SCC according to patients' smoking status, and for all other solid cancer groups and subgroups in Supplementary Table 2.

	All (n = 4908)	Other location (n = 4206)	Lung cancer (n = 702)	p-value
%CA >12	13.47% (661)	7.01% (295)	52.14% (366)	<0.0001
%GT >13	15.14% (743)	9.49% (399)	49% (344)	<0.0001
%CG ≥1	82.46% (4047)	84.38% (3549)	70.94% (498)	<0.0001
%GC ≥1	81.32% (3991)	83.05% (3493)	70.94% (498)	<0.0001
%CT, median (IQR)	22 (15; 30)	24 (18; 31)	11 (0; 16)	<0.0001
%GA, median (IQR)	22 (15; 30)	25 (18; 32)	11 (0; 16)	<0.0001
%AC, median (IQR)	1 (0; 2)	1 (0; 3)	1 (0; 1)	<0.0001
%TG, median (IQR)	1 (0; 3)	1 (0; 3)	0 (0; 1)	<0.0001
%AG >5	43.44% (2132)	47.81% (2011)	17.24% (121)	<0.0001
%TC >5	42.97% (2109)	47.0% (1977)	18.8% (132)	<0.0001
%AT >3	26.67% (1309)	24.82% (1044)	37.75% (265)	<0.0001
%TA >3	26.55% (1303)	24.37% (1025)	39.6% (278)	<0.0001

Table 1. Distribution of the relative frequencies of the 12 substitution classes in the development dataset for the 4,908 tumors with a somatic substitution prevalence of ≥1 sub/Mb. Results presented as percentage and frequency %(n) unless otherwise indicated. IQR: interquartile range; %CA = relative frequency of C > A substitutions (%); %GT = relative frequency of G > T substitutions (%); %CG = relative frequency of C > G substitutions (%); %GC = relative frequency of G > C substitutions (%); %CT = relative frequency of C > T substitutions (%); %GA = relative frequency of G > A substitutions (%); %AC = relative frequency of A > C substitutions (%); %TG = relative frequency of T > G substitutions (%); %AG = relative frequency of A > G substitutions (%); %TC = relative frequency of T > C substitutions (%); %AT = relative frequency of A > T substitutions (%); %TA = relative frequency of T > A substitutions (%).

Relative frequencies of substitutions in tumors	Bivariate Analyses		Full Mutivariate Model	
	Odds ratio and 95% CI	p-value	Odds ratio and 95% CI	p-value
%CA >12 vs ≤12	14.44 [11.95; 17.46]	0.001	4.86 [3.59; 6.58]	0.001
%GT >13 vs ≤13	9.17 [7.66; 10.98]	<0.001	1.62 [1.2; 2.19]	0.001
%CG ≥1 vs 0	0.45 [0.38; 0.54]	<0.001	1.52 [0.75; 3.06]	0.24
%GC ≥1 vs 0	0.50 [0.42; 0.60]	<0.001	1.58 [0.78; 3.22]	0.20
%CT	0.84 [0.83; 0.85]	<0.001	0.91 [0.88; 0.93]	<0.001
%GA	0.84 [0.83; 0.85]	<0.001	0.93 [0.91; 0.95]	<0.001
%AC	0.67 [0.63; 0.72]	<0.001	0.82 [0.76; 0.90]	<0.001
%TG	0.65 [0.61; 0.70]	<0.001	0.83 [0.76; 0.90]	<0.001
%AG >5 vs ≤5	0.23 [0.19; 0.28]	<0.001	0.43 [0.33; 0.58]	<0.001
%TC >5 vs ≤5	0.26 [0.21; 0.32]	<0.001	0.44 [0.33; 0.59]	<0.001
%AT >3 vs ≤3	1.84 [1.55; 2.17]	<0.001	0.99 [0.74; 1.33]	0.96
%TA >3 vs ≤3	2.04 [1.72; 2.40]	<0.001	1.15 [0.86; 1.53]	0.36

Table 2. Bivariate analysis and full multivariate model with the 12 substitution classes using logistic regression for the 4,908 tumors with a somatic substitution prevalence of ≥1 sub/Mb. CI = confidence interval; vs = versus; %CA = relative frequency of C > A substitutions (%); %GT = relative frequency of G > T substitutions (%); %CG = relative frequency of C > G substitutions (%); %GC = relative frequency of G > C substitutions (%); %CT = relative frequency of C > T substitutions (%); %GA = relative frequency of G > A substitutions (%); %AC = relative frequency of A > C substitutions (%); %TG = relative frequency of T > G substitutions (%); %AG = relative frequency of A > G substitutions (%); %TC = relative frequency of T > C substitutions (%); %AT = relative frequency of A > T substitutions (%); %TA = relative frequency of T > A substitutions (%).

Test performance on external validation dataset. Characteristics of 196 samples and distribution of the relative frequencies of the 12 substitution classes are presented in Supplementary Tables 3 and 4. Among 196 available tumors with WES data, we considered one lung tumor with a somatic substitution prevalence lower than 1 sub/Mb to be non-lung cancer. The EASILUNG score was computed on the remaining 195 samples. Before recalibration by using the threshold of −303, the associated sensitivity and specificity were respectively 48.9% [34.6%; 63.2%] and 95.3% [91.8%; 98.7%]. Descriptive characteristics of the EASILUNG score on the validation dataset and the probability of lung cancer are provided in Fig. 2. The new calibration of EASILUNG score was 2.094 + 0.008 × EASILUNG score. After calibration on the validation dataset, the AUC of the EASILUNG score was 0.849 [0.781; 0.917]. The calibration assessed by the Hosmer-Lemeshow chi-square test was equal to 8.63 (p = 0.3745) with close agreement between predicted and observed risk of lung cancer, with no apparent over- or underprediction (Supplementary Fig. 1). When using the new threshold of −467 calculated after the calibration, the associated overall sensitivity and specificity were respectively 83.0% [72.2%; 93.7%] and 72.3% [65.1%;

	Odds Ratio[‡] and 95% CI	β Coefficient[‡]	EASILUNG score points[◊]
%CA >12 vs ≤12	5.78 [4.36; 7.66]	1.7547	175
%GT >13 vs ≤13	1.86 [1.4; 2.47]	0.6209	62
%CT	0.92 [0.9; 0.94]	−0.0863	−9
%GA	0.94 [0.92; 0.96]	−0.0631	−6
%AC	0.84 [0.77; 0.92]	−0.1719	−17
%TG	0.84 [0.78; 0.92]	−0.1704	−17
%AG >5 vs ≤5	0.48 [0.36; 0.63]	−0.7433	−74
%TC >5 vs ≤5	0.48 [0.36; 0.64]	−0.7317	−73

Table 3. Final logistic regression defining the EASILUNG score. [‡]Odds ratio correspond to the exponential function of the regression coefficient (e^{β}). [◊]β Coefficient*100 rounded to the nearest integer CI = confidence interval; %CA = relative frequency of C > A substitutions (%); %GT = relative frequency of G > T substitutions (%); %CG = relative frequency of C > G substitutions (%); %GC = relative frequency of G > C substitutions (%); %CT = relative frequency of C > T substitutions (%); %GA = relative frequency of G > A substitutions (%); %AC = relative frequency of A > C substitutions (%); %TG = relative frequency of T > G substitutions (%); %AG = relative frequency of A > G substitutions (%); %TC = relative frequency of T > C substitutions (%); %AT = relative frequency of A > T substitutions (%); %TA = relative frequency of T > A substitutions (%).

Substitution classes	%CA[Δ]	%GT[◊]	%CT	%GA	%AC	%TG	%AG[*]	%TC[‡]	
EASILUNG signature =	**+175**	**+62**	**−9*(%CT)**	**−6*(%GA)**	**−17*(%AC)**	**−17*(%TG)**	**−74**	**−73**	
Example 1[(a)]									
Relative frequency	*27*	*19*	*8*	*12*	*0*	*0*	*3*	*5*	
EASILUNG score =	+175	+62	−72	−72	+0	+0	+0	+0	=93
Example 2[(b)]									
Relative frequency	*2*	*2*	*27*	*37*	*4*	*1*	*6*	*5*	
EASILUNG score =	+0	+0	−243	−222	−68	−17	−74	+0	=−624

Table 4. Calculating the EASILUNG score with two examples from the TCGA. %CA = relative frequency of C > A substitutions (%); %GT = relative frequency of G > T substitutions (%); %CT = relative frequency of C > T substitutions (%); %GA = relative frequency of G > A substitutions (%); %AC = relative frequency of A > C substitutions (%); %TG = relative frequency of T > G substitutions (%); %AG = relative frequency of A > G substitutions (%); %TC = relative frequency of T > C substitutions (%). [Δ]For samples with %CA > 12. [◊]For samples with %GT > 13. *For samples with %AG > 5. [‡]For samples with %TC > 5. [(a)]Example 1/Tumor sample barcode = TCGA-75–7027–01A-11D-1945–08 (lung cancer) with 27% C > A substitutions (+175 since > 12), 19% G > T substitutions(+62 since > 13), 8% C > T substitutions (−9*8), 12% G > A substitutions (−6*12), 0% A > C substitutions (−17*0), 0% T > G substitutions (−17*0), 3% A > G substitutions (0 since ≤5), 5% T > C substitutions (0 since ≤5). [(b)]Example 2/Tumor sample barcode = TCGA-75–7027–01A-11D-1945–08 (non-lung cancer) with 2% C > A substitutions (+0 since ≤12), 2% G > T substitutions(+0 since ≤ 13), 27% C > T substitutions (−9*27), 37% G > A substitutions (−6*37), 4% A > C substitutions (−17*4), 1% T > G substitutions (−17*1), 6% A > G substitutions (−74 since>5), 5% T > C substitutions (0 since ≤5).

79.5%]. Se and Sp were respectively 88.2%[77.4%; 99.1%] and 70.9%[62.9%; 78.8%] for the 161 FFPE samples and 64.3%[39.2%; 89.4%] and 80.9%[64.2%; 97.7%] for the 35 fresh frozen samples without any difference on sensitivities (p = 0.095) nor specificities (p = 0.339). The associated performance of the EASILUNG test on the validation dataset is detailed in Supplementary Table 5.

Discussion

Lung carcinogenesis by tobacco smoke induce a characteristic pattern of substitutions, our goal was to focus on somatic substitutions to exhibit how these parameters may be helpful to discriminate lung cancer from other cancers, especially in patient exposed to tobacco smoke. EASILUNG test use the relative frequency of somatic substitutions from WES results as a diagnostic tool. Using TCGA data, Alexandrov *et al.*[6] published a comprehensive signature classification based on the different patterns of genomic substitutions in cancer. In their study, they describe signatures with 96 parameters, corresponding to the type of substitution in their trinucleotide context, and in particular a tobacco-induced signature. These 96-parameter signatures are descriptive and are not actually used clinically in diagnosis, because each is present in several classes of cancer, *e.g.* the signature proposed for tobacco is present in the same way in lung cancers, head and neck SCCs, and in liver hepatocellular carcinoma[6], without specified thresholds and therefore without assessment performed for the sensitivity and specificity of these signatures. Based on only 8 variables, here we propose a score that can be easily applied to whole-exome results in cases where the somatic substitution prevalence is ≥1 sub/Mb. The 8 substitutions retained for the score are the signature of tobacco related lung cancer.

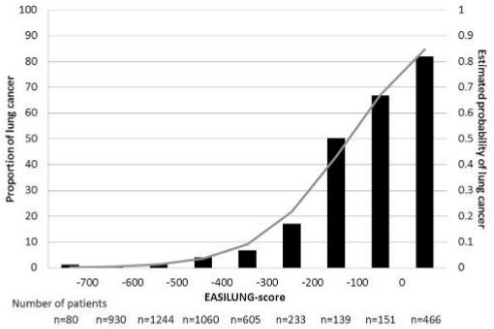


Figure 1. Description of the EASILUNG score in the development dataset. This includes the 4,908 tumors with a somatic substitution prevalence of ≥1 sub/Mb and shows the estimated probability of lung cancer calculated from the EASILUNG score according to the values of the score. Logit (lung cancer) = 1.216 + 0.010 × EASILUNG score. Probability of lung cancer = 1/(1 + exp [−logit(lung cancer)]).

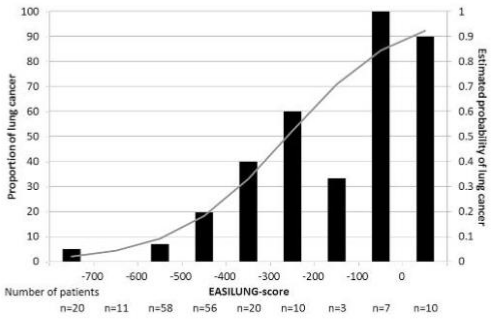


Figure 2. Description of the EASILUNG score in the validation dataset. This includes the 195 tumors with a somatic substitution prevalence of ≥1 sub/Mb and shows the estimated probability of lung cancer calculated from the EASILUNG signature according to the values of the score. Logit (lung cancer) = 2.094 + 0.008 × EASILUNG score. Probability of lung cancer = 1/(1 + exp [−logit(lung cancer)]).

We stated the threshold of 1 sub/Mb as a hypothesis for this work, as we assumed that tobacco related cancers were more likely to carry at least one somatic mutation per megabase, as described in several publications. This threshold allows an easy selection of potential candidate tumors for the application of the score. The interest of this threshold is to have a minimal number of substitutions allowing a statistical analysis using 12 variables (at least 40 substitutions considering exome targeted region >40 Mb). Furthemore, tobacco-induced lung cancer with <1 sub/Mb is uncommon, in opposite to lung cancer in non-smokers. Finally, this threshold minimizes the artifacts of non-smokers lung tumors data in a score which is related to tobacco smoke. By a posteriori assessment, we noticed that this threshold is the value that allows to reach the best NPV (98.5%). Moreover, the threshold of 1 sub/Mb on an exome sequence makes possible to have an acceptable number of substitutions - at least forty if the exome targeted region is >40 Mb - making the analysis of the distribution of 12 variables statistically significant.

We decided to use the 12 somatic substitution classes in order to consider transcriptional strand biases which is not possible with the regular 6-subtypes approach and also avoid to use the 96-subtypes model to develop a tool easy to use in everyday practice. Strand bias of tobacco-induced mutagenesis could be indicative of specific repair processes that are active in the cell; the mutational signature associated with lung cancer exhibited predominantly G > T/C > A substitutions with a transcriptional strand bias (more G > T transversions than complementary C > A somatic substitutions in the coding non-transcribed strand), suggesting the formation of bulky adducts on guanine and previously described by *Alexandrov et al.*[6]. The efficiency of DNA damage and DNA maintenance processes differ between the transcribed and non-transcribed gene strands. A well-known cause of this phenomenon is transcription-coupled nucleotide excision repair (NER) that operates predominantly on the transcribed strand of genes and is initiated by RNA polymerase II when it encounters bulky DNA helix-distorting lesions[32]. The high rate of G > T/C > A somatic substitutions in lung cancer is probably an

imprint of the bulky DNA adducts generated by polycyclic hydrocarbons found in tobacco smoke and their removal by transcription-coupled NER[33]. Furthermore, in our study high C > A relative frequency (%CA > 12%) was the most discriminative parameter, with the highest odds ratio, followed by high G > T relative frequency (%GT > 13%). The higher prevalence of G > T substitutions on the non-transcribed strand compared to complementary C > A substitutions in lung cancer reflects this strand bias and confirms that bulky adduct damage to guanine may be the cause of the observed mutations.

The AML cancer group was not included simply because there is no possible diagnostic confusion between AML and solid cancers in routine clinical practice. Furthermore, AML genomes have fewer mutations than most other adult cancers, with an average of only 13 mutations found[34]. Because of the low density of somatic substitutions in AML (<1 sub/Mb), to reject AML the specificity of our test would have be equal to 100%. Inclusion of AML would have caused inflation of the specificity of our test without any clinical relevance.

No somatic substitution was described in 35 tumors (1 lung SCC and 34 non-lung tumors), accounting for approximately one tumor out of 1000 in TCGA. We assumed that these tumors were simply carrying too few mutations or may have somatic mutations with low allele frequencies that may be excluded from the variants list[35,36]. We thus decided to exclude these 35 tumors from our validation dataset, even though our calculated specificity would have been artificially better, since all of these 35 tumors would have been considered as negative by the EASILUNG test (<1 sub/Mb).

We chose to make distinct TCGA histological subgroups, smoking status subgroups and immunophenotypical subgroups (see Supplementary Table 2) relevant in the differential diagnosis, because some TCGA groups are very heterogeneous. Heterogeneity of the results in lung cancer subgroups according to smoker status (Supplementary Table 2) suggest a high potential for our test in smokers: in lung adenocarcinoma group, positivity of the test is higher than 90% in current reformed smoker for ≤15 years and current smoker (includes daily smokers and non-daily smokers or occasional smokers) subgroups and decrease to 38% in lifelong Non-smoker subgroup (less than 100 cigarettes smoked in lifetime). We also specified anatomical subdivision in the head and neck SCC group because epidemiological and etiological heterogeneities exist in these subgroups although the histological type is still SCC. The larynx is the anatomic region closer to the trachea than other anatomic regions of head and neck SCC. It is therefore clinically relevant to evaluate the results of our test in these different anatomical subgroups. Positive EASILUNG tests in SCCs localized in the larynx (30/61; 49%), which suggests this location is more closely similar to lung SCCs than other head and neck SCC locations: oral cavity (6/153; 4%), oropharynx (2/27; 7%), and hypopharynx (0/3; 0%). The gradient in test positivity through the airways in the head and neck SCC group may reflect the existence of an underlying, parallel gradient in the extent of exposure of cells of the respiratory tract to BaP and other tobacco smoke carcinogens. These results can be linked with a previous study[33] about a gradient in the prevalence of TP53 G > T transversions in cancers of smokers, from low in the oral cavity, to intermediate in the larynx, and high in various histological types of lung cancers. The EASILUNG score is not limited in discriminating smoking lung cancer from non-smoking cancers. This score is also able to discriminate smoking lung cancer from other localizations independently of the tobacco status. For example (Supplementary Table 2), patients with lung cancer that are current reformed smoker for ≤15 years or current smoker (includes daily smokers and non-daily smokers or occasional smokers) have a positive score in 93% and 91% respectively and patients with head and neck cancer or bladder which are also often smoking-related have a positive score in only 14% and 6% respectively.

Some RNA-based molecular tests studying mRNA expression exist to help identify cases not resolved by immunohistochemistry and classified as CUP. More recently, an alternative strategy was to propose a treatment targeting the supposed primary tumor according to the molecular signature, based on gene expression analyzed by a 92-gene RT-PCR assay[37]. In this study, the sensitivity was 63% for lung adenocarcinoma from frozen samples and 0% for formalin fixed paraffin-embedded (FFPE) samples. One of the hypotheses explaining the limitations of diagnostic methods studying mRNA expression is that a genetic signature predisposing to metastasis coexists in primary tumors with the tissue-specific genetic signature. This "metastatic" genetic program that transcends distinctions between the original tissues has been described and could lead to errors in the molecular classification of various tumors at primary sites such as breast or bladder[38]. The selection of genes relies heavily on the number and type of primary solid tumors studied. In most published series, the "training set" included fewer than 1000 tumor specimens (typically 100–400). In addition, gene expression profiles are heterogeneous in pulmonary adenocarcinomas[39]. By studying DNA, our approach is not impacted by these limitations; this explains our good results and suggests the superiority of our approach in lung cancers, particularly in adenocarcinomas. By analyzing lung adenocarcinomas and lung SCC, we cover the two most prevalent lung cancer histologies and our results suggest that the EASILUNG test remained acceptably sensitive and strongly specific on FFPE samples. The mutational artefact potentially generated by formalin fixation[40] was not a limitation for our test, as demonstrated by our local sample results. Indeed, formalin crosslinking with cytosine nucleotides on either strand can result in incorrect incorporation of adenine in place of guanosine, causing an artificial C > T/G > A mutation[41,42] that could impact negatively the score (Table 2).

A known major issue of external validation studies is that they are often based on small and local datasets. This means that there is often heterogeneity in the performance of prediction models and that multiple external validation studies are needed to fully appreciate the generalizability of a model. Our external validation shows that the EASILUNG test will be suitable for other datasets in different platforms without modifying the signature but after a recalibration in each platform before a routine use: this recalibration per platform is required to define the specific threshold to be used. Our diagnosis approach could allow the detection of lung cancer using circulating tumor DNA (ctDNA) detection. ctDNA is a promising biomarker for noninvasive assessment of cancer. In a previous study, ctDNA have been detected in 100% of patients with stage II– to IV NSCLC and in 50% of patients with stage I, with 96% specificity for mutant allele fractions down to ~0.02%[43]. Exome-wide analysis of circulating

tumor DNA could complement current invasive biopsy approaches to identify single somatic substitutions spectra and density associated with lung cancer in a future blood-based screening strategy.

In conclusion, we developed an innovative genomics-based tool which may be helpful in diagnosis of lung cancer and in medical decision-making for patients with a history of tobacco smoking with unknown primary tumors potentially of lung origin. The external validation of our test confirmed its great potential. Multiple external validation studies and prospective validation studies on exome results are needed to fully realize the generalizability of our model. Clinical trials are required to evaluate its benefits to patients.

Data Availability

The local validation sample datasets analyzed during the current study are not fully publicly available: due to reasonable privacy and security concerns, the data are not easily redistributable to researchers other than those engaged in Institutional Review Board-approved research collaborations with the named medical center.

References

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012: Globocan 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
2. Sher, T., Dy, G. K. & Adjei, A. A. Small Cell Lung Cancer. *Mayo Clin. Proc.* **83**, 355–367 (2008).
3. Levi, F., Te, V. C., Erler, G., Randimbison, L. & La Vecchia, C. Epidemiology of unknown primary tumours. *Eur. J. Cancer Oxf. Engl.* **1990**(38), 1810–1812 (2002).
4. Zaun, G., Schuler, M., Herrmann, K. & Tannapfel, A. CUP Syndrome—Metastatic Malignancy with Unknown Primary Tumor. *Dtsch. Ärztebl. Int.* **115**, 157 (2018).
5. Pavlidis, N., Briasoulis, E., Pentheroudakis, G. & ESMO Guidelines Working Group. Cancers of unknown primary site: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. Off. *J. Eur. Soc. Med. Oncol.* **21** Suppl 5, v228-231 (2010).
6. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
7. Forbes, S. A. *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
8. Hecht, S. S. DNA adduct formation from tobacco-specific N-nitrosamines. *Mutat. Res. Mol. Mech. Mutagen.* **424**, 127–142 (1999).
9. Phillips, D. H. & Venitt, S. DNA and protein adducts in human tissues resulting from exposure to tobacco smoke. *Int. J. Cancer* **131**, 2733–2753 (2012).
10. ADME. Atomic views of a human P450. *Nat. Rev. Drug Discov.* **2**, 685–685 (2003).
11. Miller, E. C. Some current perspectives on chemical carcinogenesis in humans and experimental animals: Presidential Address. *Cancer Res.* **38**, 1479–1496 (1978).
12. Collisson, E. A. *et al.* Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
13. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
14. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
15. Hammerman, P. S. *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
16. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
17. Cancer Genome Atlas Research Network *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
18. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
19. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).
20. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
21. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
22. Davis, C. F. *et al.* The Somatic Genomic Landscape of Chromophobe Renal Cell Carcinoma. *Cancer Cell* **26**, 319–330 (2014).
23. Cancer Genome Atlas Research Network. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690 (2014).
24. Cancer Genome Atlas Network. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
25. Cancer Genome Atlas Research Network *et al.* Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
26. Cancer Genome Atlas Network. Genomic Classification of Cutaneous Melanoma. *Cell* **161**, 1681–1696 (2015).
27. Cancer Genome Atlas Research Network *et al.* Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma. *N. Engl. J. Med.* **374**, 135–145 (2016).
28. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011–1025 (2015).
29. Zheng, S. *et al.* Comprehensive Pan-Genomic Characterization of Adrenocortical Carcinoma. *Cancer Cell* **29**, 723–736 (2016).
30. Sauerbrei, W. & Schumacher, M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat. Med.* **11**, 2093–2109 (1992).
31. Cox, D. R. Two Further Applications of a Model for Binary Regression. *Biometrika* **45**, 562 (1958).
32. Hanawalt, P. C. & Spivak, G. Transcription-coupled DNA repair: two decades of progress and surprises. *Nat. Rev. Mol. Cell Biol.* **9**, 958–970 (2008).
33. Pfeifer, G. P. *et al.* Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. *Oncogene* **21**, 7435–7451 (2002).
34. Network, T. C. G. A. R. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
35. Balaparya, A. & De, S. Revisiting signatures of neutral tumor evolution in the light of complexity of cancer genomic data. *Nat. Genet.* https://doi.org/10.1038/s41588-018-0219-4 (2018).
36. Williams, M. J. et al. Reply to 'Revisiting signatures of neutral tumor evolution in the light of complexity of cancer genomic data'. *Nat. Genet.* https://doi.org/10.1038/s41588-018-0210-0 (2018).
37. Hainsworth, J. D. *et al.* Molecular gene expression profiling to predict the tissue of origin and direct site-specific therapy in patients with carcinoma of unknown primary site: a prospective trial of the Sarah Cannon research institute. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **31**, 217–223 (2013).
38. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**, 49–54 (2003).
39. Garber, M. E. *et al.* Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* **98**, 13784–13789 (2001).

40. Williams, C. *et al*. A High Frequency of Sequence Alterations Is Due to Formalin Fixation of Archival Specimens. *Am. J. Pathol.* **155**, 1467–1471 (1999).
41. Srinivasan, M., Sedmak, D. & Jewell, S. Effect of Fixatives and Tissue Processing on the Content and Integrity of Nucleic Acids. *Am. J. Pathol.* **161**, 1961–1971 (2002).
42. Chen, G., Mosier, S., Gocke, C. D., Lin, M.-T. & Eshleman, J. R. Cytosine Deamination Is a Major Cause of Baseline Noise in Next-Generation Sequencing. *Mol. Diagn. Ther.* **18**, 587–593 (2014).
43. Newman, A. M. *et al*. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20**, 548–554 (2014).

Acknowledgements
TCGA consortium. Centre Georges-François Leclerc (Dijon, France).

Author Contributions
S.B. conceived the original idea, extracted the TCGA data, and wrote the manuscript. J.S. conceived and wrote the statistical part of the manuscript, planned statistical analysis, and drafted figures. S.B., J.S., and A.H. and J.L.M. contributed with study design and revised the paper. R.B. contributed with acquisition of data from the external validation samples. A.H. directed the overall research.

Additional Information
Competing Interests: The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019

FULL PAPER

BJC

British Journal of Cancer (2015), 1–6 | doi: 10.1038/bjc.2015.250

Keywords: metastatic colorectal cancer; cetuximab; panitumumab; RAS; EGFR; AKT; phosphoproteins; pEGFR; pAKT

Expression of pEGFR and pAKT as response-predictive biomarkers for *RAS* wild-type patients to anti-EGFR monoclonal antibodies in metastatic colorectal cancers

A Harlé*[,1,2,3], J Salleron[4], G Perkins[5], C Pilati[5], H Blons[5], P Laurent-Puig[5] and J L Merlin[1,2,3]

[1]*Université de Lorraine, Faculté de Pharmacie, 54001 Nancy, France;* [2]*CNRS UMR 7039 CRAN, 54506 Vandoeuvre les Nancy, France;* [3]*Institut de Cancérologie de Lorraine, Service de Biopathologie, 54519 Vandoeuvre les Nancy, France;* [4]*Institut de Cancérologie de Lorraine, Cellule Data Biostatistique, 54519 Vandoeuvre les Nancy, France and* [5]*Institut National de la Santé et de la Recherche Médicale (INSERM), Unité Mixte de Recherche (UMR)-S1147, Personalized Medicine, Pharmacogenomics, Therapeutic Optimization, Université Paris Descartes, 45 rue des Saints Pères, 75006 Paris, France*

Background: *RAS* wild-type ($RAS_{w/t}$) tumours have been associated with better outcomes in patients with metastatic colorectal cancer (mCRC) treated with anti-EGFR monoclonal antibodies (mAb). We investigated the expression of EGFR downstream proteins under their active phosphorylated forms as potential markers in response to these patients.

Methods: One-hundred tumour samples were collected from patients with mCRC refractory to FOLFOX and/or FOLFIRI and treated by a combination of chemotherapy with anti-EGFR mAb. The outcomes were measured on response evaluation criteria in solid tumour (RECIST), progression-free survival (PFS) and overall survival (OS). All samples were assessed for *RAS* and *BRAF* mutations, and the key phosphorylated proteins of EGFR downstream signalling were quantitatively analysed using the BioPlex Protein array.

Results: Among the 60 $RAS_{w/t}$ patients, 45.0% achieved a complete or partial response when treated with anti-EGFR mAb. Expression of pAKT, pERK1/2 and pMEK1 was significantly lower in $RAS_{w/t}$ patients ($P=0.0246$; $P=0.004$; $P=0.0110$, respectively). The response rate was significantly higher for $RAS_{w/t}$ patients who express pEGFR and pAKT ($P=0.0258$; $P=0.0277$, respectively).

Conclusions: Overexpression of pEGFR and pAKT may predict the response rate in $RAS_{w/t}$ patients treated with anti-EGFR mAb. On the basis of our results, we hypothesise that the association of anti-EGFR mAb and anti-AKT therapies could be of interest.

Colorectal cancer is the third most common cancer worldwide with more than one million patients diagnosed each year, among 50% of whom will develop metastatic disease (Jemal *et al*, 2011; Tougeron *et al*, 2013). Recent efforts to improve the treatment of metastatic colorectal cancer (mCRC) have led to the development of monoclonal antibodies (mAbs) such as cetuximab and panitumumab that inhibit the activation of the epidermal growth factor receptor (EGFR) and its downstream pathways (namely RAS/RAF/MAPK and PI3K/AKT/mTOR) that promote cell growth, proliferation, inhibition of apoptosis, invasion and metastasis. Evidence has first showed that patients with codon 12 or 13 *KRAS*-mutant tumours receive little or no benefit from anti-EGFR therapies as single agents or combined with chemotherapy (Di Fiore *et al*, 2007; Khambata-Ford *et al*, 2007; De Roock *et al*, 2008; Lievre *et al*, 2008). More recent studies revealed that the presence of mutations on exons 2, 3 and 4 of *KRAS* or *NRAS* were also predictive of

*Correspondence: Dr A Harlé; E-mail: a.harle@nancy.unicancer.fr

Received 13 January 2015; revised 3 April 2015; accepted 16 June 2015

© 2015 Cancer Research UK. All rights reserved 0007 – 0920/15

resistance to anti-EGFR therapies (Douillard *et al*, 2013; Sorich *et al*, 2014). In other words, cetuximab or panitumumab shows benefits only to '*RAS* wild-type' patients, that is, patients with tumours that are not bearing mutations on exons 2, 3 and 4 of *KRAS* and *NRAS*. These findings led the European Medicine Agency to the use of cetuximab and panitumumab for patients with *RAS* wild-type tumours.

However, from studies including *RAS* wild-type patients, the response rates to cetuximab or panitumumab therapy ranged from only 40 to 60% (Linardou *et al*, 2008), which results in a large fraction of patients being without any known causes for treatment failure. The presence of alterations in other genes such as *PIK3CA* or *BRAF* in the EGFR-dependent signalling pathways (Laurent-Puig *et al*, 2009; Custodio and Feliu, 2013; Douillard *et al*, 2013) is responsible for some of the non-responding cases. Moreover, overexpression or alterations on proteins such as PTEN, PI3K and AKT, involved in the RAS/RAF/MAPK or PI3K/AKT/mTOR signalling pathways, can have a significant impact on cell proliferation or apoptosis. Absence or overexpression of proteins under their phosphorylated forms may be of interest to predict response to anti-EGFR mAb. Ligand fixation of the EGF receptor activates the pathway. The fixation of ligand on the receptor activates the kinase activity that is responsible for the activation of downstream proteins by phosphorylation. The study of the expression of the downstream proteins under their phosphorylated forms is important to understand mechanisms of resistance to anticancer drugs and more precisely in our study of anti-EGFR drugs. In a previous study, we showed that the expression level of EGFR downstream signalling phosphoproteins (pMEK1 and pP70S6K) could predict longer progression-free survival (PFS) and overall survival (OS) in KRAS exon 2 wild-type mCRC using multivariate analysis (Perkins *et al*, 2010).

In this paper, we describe the impact of *RAS* mutations (that is, *KRAS* and *NRAS* exon 2, 3 and 4 mutations) and of the expression of proteins involved in RAS/RAF/MAPK and PI3K/AKT/mTOR pathways in their active phosphorylated forms on the response to anti-EGFR mAbs in *RAS* wild-type patients with an advanced colorectal cancer. The 42 patient population described in our previous study (Perkins *et al*, 2010) has been extended to 100 patients, and the research of mutations has been also extended from *KRAS* exon 2 to *RAS* (*KRAS*, *NRAS*) exons 2, 3 and 4.

PATIENTS AND METHODS

Patients. We analysed total protein extracts prepared from 100 frozen tumour samples from patients with mCRC treated by a combination of chemotherapy with anti-EGFR therapy, from the collection that has been already described (Lievre *et al*, 2008; Laurent-Puig *et al*, 2009). Characteristics of the patients are shown in Table 1. All patients were refractory to FOLFOX and/or FOLFIRI regimen and were treated with anti-EGFR antibodies. In summary, 95% of the patients were treated by a combination of anti-EGFR and irinotecan, 2% were treated by a monotherapy using cetuximab or panitumumab, 1% were treated by a combination of anti-EGFR with a chemotherapy associating 5-fluorouracil and oxaliplatin and 1% were treated by 5-fluorouracil. The outcomes were measured on RECIST, PFS and OS. Evaluation was measured every 2 to 3 months according to French standard of care. Patients with complete and partial responses were considered as responders and patients with stable disease or progression as non-responders. Tumours were collected at the time of surgery. All samples were available either as frozen samples or as formalin-fixed paraffin-embedded tissues. This study has been approved by the Ile-de-France ethical committee number 2 (ID-RCB-AO1058-47/2008-135).

Table 1. Characteristics of patients analysed for EGFR pathway protein using the Bioplex protein assay

Median age (years; range)	61 (22–80)
Gender	
Male	58
Female	41
NA	1
Number of chemotherapy line used before anti-EGFR	
0	0%
1	5%
2	46%
3	36%
4	8%
5	3%
6	2%
Death	
Yes	83%
No	17%
Median OS (month–range)	10.0 (0.1–48.1)
Median PFS (month–range)	4.1 (0–20.0)
Response	
Complete response	2%
Partial response	26%
Stable disease	34%
Progressive disease	37%
NA	1%
WHO	
0	28%
1	48%
2	11%
3	1%
NA	12%
Mutation	
KRAS	36%
NRAS	4%
BRAF	2%

Abbreviations: EGFR = epidermal growth factor receptor; PFS = progression-free survival; OS = overall survival; WHO = World Health Organization.

Protein extraction. Whole-cell protein extraction was performed from 5-mm frozen tumour tissues using the Kit RIPA lysis Buffer 1X (Tebubio, Le Perray en Yvelines, France) prepared with inhibitors (PMSF, protease inhibitor cocktail and sodium orthonavate; Dutscher, Brumath, France) according to the manufacturer's recommendations.

Phosphoprotein analysis. The expression of phosphorylated-EGFR ($pEGFR_{(Tyr)}$) and the key phosphorylated proteins of EGFR downstream signalling pathway such as $pAKT_{(Ser473)}$, $pERK1\text{-}2_{(Thr202/Tyr204,\ Thr185/Tyr187)}$, $pGSK3_{(Ser21/Ser9)}$, $pIGF1R_{(Tyr1131)}$, $pMEK1_{(Ser217/Ser221)}$, $pP70S6K_{(Thr421/Ser424)}$ and $pP90RSK_{(Thr359/Ser363)}$ were quantitatively analysed in freshly frozen samples using BioPlex Protein assay (BPA; Bio-Rad, Marnes-la-coquette, France) as previously described and validated (Chergui *et al*, 2009). This technique is on the basis of multiplex sandwich bead immunoassays. Protein extracts were transferred into 96-well dishes and diluted with buffered solution. Fluorescent capture beads coupled to antibodies directed against the phosphoproteins $pEGFR_{(Tyr)}$, $pAKT_{(Ser473)}$, $pERK1\text{-}2_{(Thr202/Tyr204,\ Thr185/Tyr187)}$, $pGSK3_{(Ser21/Ser9)}$, $pIGF1R_{(Tyr1131)}$, $pMEK1_{(Ser217/Ser221)}$, $pP70S6K_{(Thr421/Ser424)}$ and $pP90RSK_{(Thr359/Ser363)}$ were mixed and added into each well and incubated overnight. Following incubation, the plates were washed and incubated with biotinylated antibodies fixing each target protein. Streptavidin–phycoerythrin solution was then added. The analysis consisted of a double laser fluorescence detection allowing

simultaneous identification of the target protein through the red fluorescence emission signal of the bead and quantification of the target protein through the fluorescence intensity of phycoerythrin. Results were recorded as the mean fluorescence intensities and compared with negative controls. Positive controls consisting of standard protein extracts from cell lines were added to each series. All results were normalised through the different batches of analyses by the same mutated tumour sample. The expression level of each phosphoprotein was recorded as fluorescence arbitrary unit. BPA assay requires 15–20 mg of tissue containing more than 50% of tumour material or equivalent amount of protein extract (that is, 25 mg total protein per assay in triplicate). All samples were assessed in triplicate to ensure reproducibility of the data.

Determination of KRAS, NRAS and BRAF mutations. *KRAS* exon 2 and 3 and *BRAF* exon 15 mutation statuses were assessed by allelic discrimination using TaqMan qPCR probes as previously described (Lievre *et al*, 2008; Laurent-Puig *et al*, 2009). KRAS exon 4 and NRAS exon 2, 3 and 4 mutation statuses were assessed using Sanger sequencing and next-generation sequencing. Patients with wild-type tumours for exons 2, 3 and 4 of *KRAS* and *NRAS* were considered as «*RAS* wild-type».

Statistical analysis. Statistical analysis was performed using the SAS software (SAS Institute, Cary, NC, USA 25513; version 9.2). A *P*-value <0.05 was considered as statistically significant.

Quantitative variables were described by the mean, median and interquartile range, and qualitative variables by frequency and percentage. The normality of the distribution of the expression of each phosphoprotein was assessed with the Shaprio–Wilk test.

Overall survival and PFS were described using the Kaplan–Meier method and compared between *RAS* mutation status with bivariate Cox proportional hazard model. Comparisons of each phosphoprotein according to the *RAS* mutation status were performed by a Mann–Whitney *U*-test.

In the subgroup of RAS wild-type patients, the discriminant power of each phosphoprotein to predict the response rate was estimated with the area under the receiver operating characteristic (AUC) and their 95% confidence intervals. An AUC equal to 0.5 corresponds to no discrimination; an AUC between 0.5 and 0.7 means a low discriminant power; an AUC between 0.7 and 0.8 corresponds to an acceptable discrimination and above 0.8 to an excellent discrimination. Each phosphoprotein was then dichotomised according to the best cutoff maximising the sensitivity and the specificity. These thresholds were computed owing to the Youden Index.

The percentage of response rate was calculated for each level of the new binary variables and compared with a χ^2-test or a Fisher Exact test. The odds ratio and their 95% confidence interval were computed with a bivariate logistic regression.

The phosphoproteins with a *P*-value less than 0.2 in bivariate analysis were introduced in a multivariate logistic regression with backward selection (Harrell *et al*, 1996). The stability of the selected model was investigated using the bootstrap-resampling method (Sauerbrei and Schumacher, 1992). The discriminant power of the final model was estimated using the AUC with 95% CI. We addressed the optimism bias using a bootstrap-resampling method. The stability of the score was estimated by cross-validation (Altman and Royston, 2000).

RESULTS

Among the 100 tumour samples, 40 were found with a *RAS* mutation (36% with *KRAS* mutation and 4% with *NRAS* mutation) and 60% were *RAS* wild type. Among the *RAS* wild-type patients, 45.0% achieved a complete or partial response, and 55.0% had a stable disease or progression ($P<0.001$) when treated with anti-EGFR. Patients with a *RAS* mutation had significant lower PFS (HR = 3.04 (1.91; 4.83)) and OS (HR = 2.49 (1.56; 3.97)). Progression-free survival and OS were significantly higher in *RAS* wild-type patients (Figure 1). Expression levels of pAKT, pERK1/2 and pMEK1 were significantly lower in *RAS* wild-type patients than in *RAS*-mutated patients ($P=0.0246$; $P=0.004$; $P=0.0110$, respectively), and no significant difference was observed between *RAS* wild-type and *RAS*-mutated tumours in the expression of pEGFR, pGSK3, pIGFR, pP70S6K and pP90SRK (Table 2).

In *RAS* wild-type patients, the AUC of the expression of each phosphoprotein according to the response rate were under 0.80. After dichotomising the expression of each phosphoprotein with the best threshold to maximise the sensitivity and the specificity, the response rate was significantly higher for tumours that express pEGFR and pAKT above the calculated threshold ($P=0.0258$ and $P=0.0277$, respectively). No significant relation was found between the response rate and the level of expression of the other phosphoproteins (Table 3a).

These two phosphoproteins remained statistically significant in multivariate analysis to predict response after bootstrap validation (Table 3b). The AUC of the final model was 0.673 (95% CI (0.540; 0.805)). After correction for the optimism bias, the AUC was 0.654. The cross-validation covariate was associated with a (beta) value of 1.08 close to 1.

No statistically significant difference was found between expression of phosphoproteins and PFS or OS in *RAS* wild-type tumour-bearing patients (data not shown).

DISCUSSION

The development of molecular biology and knowledge about signalling pathways in cancer have been a real opportunity to

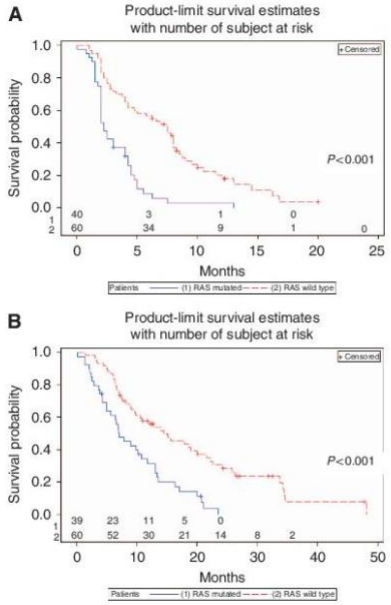


Figure 1. Kaplan–Meier's curve of *RAS* wild-type and *RAS*-mutated patients for PFS (**A**) and OS (**B**).

Table 2. Expression of phosphoproteins in all samples, RAS-mutated samples and RAS wild-type samples

	All*	*RAS* mutated*	*RAS* wild type*	*P*
pAKT	104.01; 98.5 (86.04; 124.55)	110.79; 109.46 (93.31; 126.27)	99.49; 92.83 (82.10; 122.07)	**0.0246**
pEGFR	139.35; 109.39 (86.69; 141.04)	174.69; 105.38 (88.58; 146.54)	115.80; 110.11 (85.16; 139.13)	0.9355
pERK1/2	245.73; 138.98 (101.59; 263.85)	350.22; 178.23 (137.87; 395.01)	176.07; 113.83 (92.04; 169.58)	**0.0004**
pGSK3	133.03; 101.05 (66.38; 183.82)	135.62; 97.79 (66.38; 168.41)	131.30; 105.88 (65.32; 184.71)	0.7676
pIGFR	83.65; 74.83 (64.29; 96.68)	83.52; 70.81 (63.89; 91.62)	83.74; 77.13 (64.70; 101.09)	0.4024
pMEK1	272.19; 131.82 (80.42; 280.74)	411.25; 209.52 (87.89; 402.14)	179.48; 104.09 (77.70; 201.28)	**0.0110**
pP70S6K	123.31; 112.49 (89.14; 135.21)	131.39; 124.47 (93.58; 136.83)	117.93; 105.24 (85.00; 135.21)	0.1690
pP90SRK	74.37; 75.61 (49.71; 95.38)	81.03; 80.27 (48.73; 101.78)	69.93; 71.78 (50.11; 88.26)	0.1303

Expression of pAKT, pERK1/2 and pMEK1 is significantly lower in *RAS* wild-type patients. *Mean; median (lower quartile; upper quartile).

Table 3. Expression of phoshoproteins and response rate

(a)

pProt	AUC and 95% CI	Threshold[a]	CR or PR	*P*-value	OR and 95% CI
pAKT	0.588 (0.436; 0.74)	<120 $\geqslant 120$	36.36% (16) 68.75% (11)	**0.0258**	1 3.85 (1.13; 13.07)
pEGFR	0.634 (0.491; 0.776)	<100 $\geqslant 100$	25.00% (5) 55.00% (22)	**0.0277**	1 3.67 (1.12; 12.03)
pERK1/2	0.497 (0.348; 0.646)	<110 $\geqslant 110$	38.45% (10) 50.00% (17)	0.373	1 1.60 (0.57; 4.51)
pGSK3	0.504 (0.355; 0.653)	<130 $\geqslant 130$	38.24% (13) 53.85% (14)	0.228	1 1.88 (0.67; 5.31)
pIGFR	0.513 (0.365; 0.662)	<60 $\geqslant 60$	33.33% (3) 47.06% (24)	0.445	1 1.78 (0.40; 7.89)
pMEK1	0.497 (0.348; 0.646)	<100 $\geqslant 100$	38.46% (10) 50.00% (17)	0.373	1 1.60 (0.57; 4.51)
pP70S6K	0.585 (0.439; 0.731)	<105 $\geqslant 105$	36.67% (11) 53.33% (16)	0.194	1 1.97 (0.70; 5.54)
pP90SRK	0.577 (0.43; 0.723)	<70 $\geqslant 70$	35.71% (10) 53.13% (17)	0.176	1 2.04 (0.72; 5.77)

(b)

Multivariate

pProt	OR and 95% CI
pAKT	1 3.70 (1.04;13.18)
pEGFR	1 3.53 (1.03;12.11)

Abbreviations: AUC = area under the receiver operating characteristic; CI = confidence interval; CR = complete response; OR = odds ratio; PR = partial response. (a) Patients with expression of pAKT over 120 and pEGFR over 100 achieved significantly better response rate. (b) These two phosphoproteins remained statistically significant in multivariate analysis to predict response after bootstrap validation. Statistically significant p-values are presented in bold.
[a]Optimal threshold according to the Youden index.

improve treatment for patients with mCRC. The history of *RAS* and the use of anti-EGFR therapies are characteristic of the importance of identifying response-predictive biomarkers and molecular testing. Unluckily, a large proportion of *RAS* wild-type patients do not reach benefit from anti-EGFR-targeted therapies and it is important to explore new targets and identify patients who will benefit from targeted therapies. The selection of these patients is crucial to avoid unworthy toxicities or inefficient treatment, but also for economic issues.

In our study, the repartition of *RAS* mutation in the studied population is consistent with previously published data (Sorich *et al*, 2014). The observed response rate is also consistent with known data with few or no response in *RAS*-mutant tumour (Bokemeyer *et al*, 2012). Complete or partial responses were achieved in $\sim$45.0% of *RAS* wild-type tumour-bearing patients, which is higher than that described in previously published data for patients treated for second or further line with an association for anti-EGFR and FOLFIRI (Iwamoto *et al*, 2014). This difference can probably be related to the lower median age of our population and the better WHO status.

The level of expression of phosphoproteins in the *RAS*-mutated patients was found to be significantly higher than in RAS wild-type patients as we expected. The accumulation of RAS protein under its activated form causes the activation of RAF/RAS/MAPK and PI3K/AKT/mTOR signalling pathways, resulting in an overexpression of the downstream proteins involved in these pathways (Newbold, 1984).

In our study, the level of expression of phosphoproteins was measured with the BPA and not with immunohistochemistry. BPA is a validated method and is comparable to western blot analysis (Chergui *et al*, 2009). Most of the studies published in the literature are using western blot or immunohistochemistry, which are

qualitative and semiquantitative assays, respectively. BPA is a quantitative assay and allows more precision in the measurement of the level of phosphorylated proteins. However, immunohistochemistry allows the determination of the expression according to the localisation of the target—that is, nucleus, cytoplasm, stroma and so on—whereas BPA is using the whole tumour and is unable to determine the localisation of the overexpression in the cell. This difference of technical approach may be responsible to the discrepancies we will discuss further in this paragraph.

Overexpression of pEGFR and pAKT was significantly associated with higher response rate in *RAS* wild-type patients; however, no statistically significant relation was found between the expression of phosphoproteins and PFS and OS. These date do not confirm our previously published data (Perkins *et al*, 2010). In the study published by Perkins *et al* (2010), PFS was related to a lower expression of pMEK and pP70S6K in wild-type *KRAS* exon 2 tumour-bearing patients. Among the 42 patients included in this study, only 23 were wild-type *KRAS* tumour-bearing. The lack of subjects in the wild-type *KRAS* class and the impact of *KRAS* exon 3, 4 and *NRAS* exon 2, 3 and 4 mutations may explain this difference with our present data. Moreover, although being in the same order as sample sizes in other papers investigating phosphorylated proteins as putative response-predictive biomarkers in mCRC (Scartozzi *et al*, 2012; Cejas. *et al*, 2012), the statistical power would have been better with a larger sample size and these results have to be confirmed in the future on a largest population.

In our paper, phosphoproteins have been extracted and analysed from primary tumours; however, some studies suggest that the correlation of the expression of EGFR and AKT between the primary and the metastasis does not appear in 20–40% of cases (Scartozzi *et al*, 2012; Cejas. *et al*, 2012). However, in these two papers pEGFR and pMAPK have been investigated using IHC considering weakly stained samples as negative. In our paper, phosphorylated protein expression was quantified in frozen-specimen protein extracts and considered as positive when exceeding 3 × the background noise value, as validated in Chergui *et al* (2009). In addition, in these two papers, the analyses were performed in FFPE specimens, which does not warrant optimal preservation of phosphoproteins as the freezing procedure we used. According to these published data, it would be of interest to compare the expression of pEGFR and pAKT in primary and metastasis in a larger series of samples and to compare these data to response rate, PFS and OS.

The relation between the overexpression of pEGFR and response rate has already been described (Razis *et al*, 2014) and can be explained by the avidity of the tumour cell for its ligand: the more the cell is dependent on its ligand, the more EGFR will be expressed and the more anti-EGFR will be effective against tumour cells. Considering this ligand avidity, it is not surprising that a better response rate was achieved with anti-EGFR with *RAS* wild-type tumours overexpressing pEGFR. More surprisingly, a significant relation has been found between higher response rate and overexpression of pAKT. It is known that activation of AKT is dependent on the activation of EGFR (Mirzoeva *et al*, 2009). It has also been described that expression of pAKT was associated with lower response rate and lower PFS (Scartozzi *et al*, 2012); however, the studied population was selected between 2007 and 2011, and no indication mentions whether the patients treated by anti-EGFR were only *RAS* wild-type or not. In our study, the expression of pAKT was associated with a higher response rate, but no statistically significant difference in PFS and OS was found (HR = 1.004; 95% CI (0.995; 1.013) and HR = 1.006; 95% CI (0.996; 1.016), respectively). In the study of Saglam *et al* (2007), an increasing in the pAKT expression has been described during colorectal cancer progression. We hypothesise that the overexpression of pEGFR induces better response rate for *RAS* wild-type patients treated with anti-EGFR; however, this overexpression of pEGFR induces an overexpression of pAKT that leads to progression and annihilates anti-EGFR benefits and explains that no statistically significant difference in PFS was found. As already suggested, AKT may be an interesting target in colorectal cancers (De Roock *et al*, 2011; Jeong *et al*, 2014). Ongoing trials are currently testing an allosteric inhibitor of AKT, the MK-2206 (Agarwal *et al*, 2014; Jeong *et al*, 2014) in patients with mCRC. On the basis of our results, we hypothesise that associating anti-EGFR and anti-AKT therapies may be of interest in RAS wild-type patients.

In conclusion, these data suggest that the overexpression of pEGFR and pAKT may be response-predictive biomarkers for RAS wild-type patients treated with anti-EGFR, and, on the basis of our results, we hypothesise that the association of anti-EGFR and anti-AKT therapies could be of interest.

ACKNOWLEDGEMENTS

This work has been supported by the translational research programme INCa 2009-1-RT-03 and the FP7 APODECIDE.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Agarwal E, Chaudhuri A, Leiphrakpam PD, Haferbier KL, Brattain MG, Chowdhury S (2014) Akt inhibitor MK-2206 promotes anti-tumor activity and cell death by modulation of AIF and Ezrin in colorectal cancer. *BMC Cancer* **14**: 1471–2407.

Altman DG, Royston P (2000) 'What do we mean by validating a prognostic model?'. *Stat Med* **19**(4): 453–473.

Bokemeyer C, Van Cutsem E, Rougier P, Ciardiello F, Heeger S, Schlichting M, Celik I, Kohne CH (2012) Addition of cetuximab to chemotherapy as first-line treatment for KRAS wild-type metastatic colorectal cancer: pooled analysis of the CRYSTAL and OPUS randomised clinical trials. *Eur J Cancer* **48**: 1466–1475.

Cejas P, Lopez-Gomez M, Aguayo C, Madero R, Moreno-Rubio J, de Castro Carpeño J, Belda-Iniesta C, Barriuso J, Moreno García V, Díaz E, Burgos E, Gonzalez-Barón M, Feliu J (2012) 'Analysis of the concordance in the EGFR pathway status between primary tumors and related metastases of colorectal cancer patients:implications for cancer therapy.. *Curr Cancer Drug Targets* **12**(2): 124–131.

Chergui F, Chretien AS, Bouali S, Ramacci C, Rouyer M, Bastogne T, Genin P, Leroux A, Merlin JL (2009) Validation of a phosphoprotein array assay for characterization of human tyrosine kinase receptor downstream signaling in breast cancer. *Clin Chem* **55**: 1327–1336.

Custodio A, Feliu J (2013) Prognostic and predictive biomarkers for epidermal growth factor receptor-targeted therapy in colorectal cancer: beyond KRAS mutations. *Crit Rev Oncol Hematol* **85**: 45–81.

De Roock W, De Vriendt V, Normanno N, Ciardiello F, Tejpar S (2011) KRAS, BRAF, PIK3CA, and PTEN mutations: implications for targeted therapies in metastatic colorectal cancer. *Lancet Oncol* **12**: 594–603.

De Roock W, Piessevaux H, De Schutter J, Janssens M, De Hertogh G, Personeni N, Biesmans B, Van Laethem JL, Peeters M, Humblet Y, Van Cutsem E, Tejpar S (2008) KRAS wild-type state predicts survival and is associated to early radiological response in metastatic colorectal cancer treated with cetuximab. *Ann Oncol* **19**: 508–515.

Di Fiore F, Blanchard F, Charbonnier F, Le Pessot F, Lamy A, Galais MP, Bastit L, Killian A, Sesboue R, Tuech JJ, Queuniet AM, Paillot B, Sabourin JC, Michot F, Michel P, Frebourg T (2007) Clinical relevance of KRAS mutation detection in metastatic colorectal cancer treated by Cetuximab plus chemotherapy. *Br J Cancer* **96**: 1166–1169.

Douillard JY, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, Humblet Y, Bodoky G, Cunningham D, Jassem J, Rivera F, Kocakova I,

Ruff P, Blasinska-Morawiec M, Smakal M, Canon JL, Rother M, Williams R, Rong A, Wiezorek J, Sidhu R, Patterson SD (2013) Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. *N Engl J Med* **369**: 1023–1034.

Harrell Jr FE, Lee KL, Mark DB (1996) Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* **15**: 361–387.

Iwamoto S, Hazama S, Kato T, Miyake Y, Fukunaga M, Matsuda C, Bando H, Sakamoto J, Oba K, Mishima H (2014) Multicenter phase II study of second-line cetuximab plus folinic acid/5-fluorouracil/irinotecan (FOLFIRI) in KRAS wild-type metastatic colorectal cancer: the FLIER study. *Anticancer Res* **34**: 1967–1973.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *Ca* **61**: 69–90.

Jeong WJ, Cha PH, Choi KY (2014) Strategies to overcome resistance to epidermal growth factor receptor monoclonal antibody therapy in metastatic colorectal cancer. *World J Gastroenterol* **20**: 9862–9871.

Khambata-Ford S, Garrett CR, Meropol NJ, Basik M, Harbison CT, Wu S, Wong TW, Huang X, Takimoto CH, Godwin AK, Tan BR, Krishnamurthi SS, Burris 3rd HA, Poplin EA, Hidalgo M, Baselga J, Clark EA, Mauro DJ (2007) Expression of epiregulin and amphiregulin and K-ras mutation status predict disease control in metastatic colorectal cancer patients treated with cetuximab. *J Clin Oncol* **25**: 3230–3237.

Laurent-Puig P, Cayre A, Manceau G, Buc E, Bachet JB, Lecomte T, Rougier P, Lievre A, Landi B, Boige V, Ducreux M, Ychou M, Bibeau F, Bouche O, Reid J, Stone S, Penault-Llorca F (2009) Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. *J Clin Oncol* **27**: 5924–5930.

Lievre A, Bachet JB, Boige V, Cayre A, Le Corre D, Buc E, Ychou M, Bouche O, Landi B, Louvet C, Andre T, Bibeau F, Diebold MD, Rougier P, Ducreux M, Tomasic G, Emile JF, Penault-Llorca F, Laurent-Puig P (2008) KRAS mutations as an independent prognostic factor in patients with advanced colorectal cancer treated with cetuximab. *J Clin Oncol* **26**: 374–379.

Linardou H, Dahabreh IJ, Kanaloupiti D, Siannis F, Bafaloukos D, Kosmidis P, Papadimitriou CA, Murray S (2008) Assessment of somatic k-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced non-small-cell lung cancer and metastatic colorectal cancer. *Lancet Oncol* **9**: 962–972.

Mirzoeva OK, Das D, Heiser LM, Bhattacharya S, Siwak D, Gendelman R, Bayani N, Wang NJ, Neve RM, Guan Y, Hu Z, Knight Z, Feiler HS, Gascard P, Parvin B, Spellman PT, Shokat KM, Wyrobek AJ, Bissell MJ, McCormick F, Kuo WL, Mills GB, Gray JW, Korn WM (2009) Basal subtype and MAPK/ERK kinase (MEK)-phosphoinositide 3-kinase feedback signaling determine susceptibility of breast cancer cells to MEK inhibition. *Cancer Res* **69**: 565–572.

Newbold R (1984) Cancer: mutant ras proteins and cell transformation. *Nature* **310**: 628–629.

Perkins G, Lievre A, Ramacci C, Meatchi T, de Reynies A, Emile JF, Boige V, Tomasic G, Bachet JB, Bibeau F, Bouche O, Penault-Llorca F, Merlin JL, Laurent-Puig P (2010) Additional value of EGFR downstream signaling phosphoprotein expression to KRAS status for response to anti-EGFR antibodies in colorectal cancer. *Int J Cancer* **127**: 1321–1331.

Razis E, Pentheroudakis G, Rigakos G, Bobos M, Kouvatseas G, Tzaida O, Makatsoris T, Papakostas P, Bai M, Goussia A, Samantas E, Papamichael D, Romanidou O, Efstratiou I, Tsolaki E, Psyrri A, De Roock W, Bafaloukos D, Klouvas G, Tejpar S, Kalogeras KT, Pectasides D, Fountzilas G (2014) EGFR gene gain and PTEN protein expression are favorable prognostic factors in patients with KRAS wild-type metastatic colorectal cancer treated with cetuximab. *J Cancer Res Clin Oncol* **140**: 737–748.

Saglam O, Garrett CR, Boulware D, Sayegh Z, Shibata D, Malafa M, Yeatman T, Cheng JQ, Sebti S, Coppola D (2007) Activation of the serine/threonine protein kinase AKT during the progression of colorectal neoplasia. *Clin Colorectal Cancer* **6**: 652–656.

Sauerbrei W, Schumacher M (1992) A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med* **11**: 2093–2109.

Scartozzi M, Giampieri R, Maccaroni E, Mandolesi A, Biagetti S, Alfonsi S, Giustini L, Loretelli C, Faloppi L, Bittoni A, Bianconi M, Del Prete M, Bearzi I, Cascinu S (2012) Phosphorylated AKT and MAPK expression in primary tumours and in corresponding metastases and clinical outcome in colorectal cancer patients receiving irinotecan-cetuximab. *J Transl Med* **10**: 1479–5876.

Sorich MJ, Wiese MD, Rowland A, Kichenadasse G, McKinnon RA, Karapetis CS (2014) Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized controlled trials. *Ann Oncol* **2014.**

Tougeron D, Lecomte T, Pages JC, Villalva C, Collin C, Ferru A, Tourani JM, Silvain C, Levillain P, Karayan-Tapon L (2013) Effect of low-frequency KRAS mutations on the response to anti-EGFR therapy in metastatic colorectal cancer. *Ann Oncol* **24**: 1267–1273.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License

International Journal of
Molecular Sciences

MDPI

Review

SMAD4 and the TGFβ Pathway in Patients with Pancreatic Ductal Adenocarcinoma

Julie Dardare, Andréa Witz, Jean-Louis Merlin, Pauline Gilson and Alexandre Harlé *

Université de Lorraine, CNRS UMR7039 CRAN, Service de Biopathologie, Institut de Cancérologie de Lorraine, 54519 Vandoeuvre-lès-Nancy, France; j.dardare@nancy.unicancer.fr (J.D.); a.witz@nancy.unicancer.fr (A.W.); jl.merlin@nancy.unicancer.fr (J.-L.M.); p.gilson@nancy.unicancer.fr (P.G.)
* Corresponding author: a.harle@nancy.unicancer.fr; Tel.: +33-383656119

Received: 27 April 2020; Accepted: 14 May 2020; Published: 16 May 2020

Abstract: Pancreatic ductal adenocarcinoma (PDAC) is the fourth leading cause of cancer death worldwide. PDAC is an aggressive disease with an 11-month median overall survival and a five-year survival of less than 5%. Incidence of PDAC is constantly increasing and is predicted to become the second leading cause of cancer in Western countries within a decade. Despite research and therapeutic development, current knowledge about PDAC molecular mechanisms still needs improvements and it seems crucial to identify novel therapeutic targets. Genomic analyses of PDAC revealed that transforming growth factor β (TGFβ) signaling pathways are modified and the *SMAD4* gene is altered in 47% and 60% of cases, respectively, highlighting their major roles in PDAC development. TGFβ can play a dual role in malignancy depending on the context, sometimes as an inhibitor and sometimes as an inducer of tumor progression. TGFβ signaling was identified as a potent inducer of epithelial-to-mesenchymal transition (EMT), a process that confers migratory and invasive properties to epithelial cells during cancer. Therefore, aberrant TGFβ signaling and EMT are linked to promoting PDAC aggressiveness. TGFβ and SMAD pathways were extensively studied but the mechanisms leading to cancer promotion and development still remain unclear. This review aims to describe the complex role of SMAD4 in the TGFβ pathway in patients with PDAC.

Keywords: SMAD4; TGFβ; pancreatic cancer; PDAC; EMT

1. Introduction

Pancreatic ductal adenocarcinoma (PDAC) is the most common malignancy of the pancreas, with an extremely poor prognosis with a five-year survival rate of 5% and a median survival of less than 11 months [1]. Due to the lack of specific clinical symptoms and biomarkers for PDAC detection, as well as the difficulty in imaging the tumors in early stages, PDAC is often diagnosed in the later stages [2]. PDAC is currently the fourth leading cause of cancer death worldwide and is projected to become the second leading cause of cancer-related deaths before 2030 in the United States [3].

PDAC is an aggressive disease with early distant metastasis, as well as perineural and vascular local growth, making it eligible for surgical resection in only 10%–20% of patients [4]. Moreover, 69% to 75% of patients that benefitted from surgery finally relapse within two years, and 80% to 90% relapse within five years [4]. For patients with unresectable or potentially resectable PDAC, most common therapies used in cancer treatment including chemotherapy or radiotherapy are limited due to a remarkable treatment resistance. Gemcitabine was the standard of care for more than two decades in patients with pancreatic cancer, after demonstrating significant improvement in overall survival (5.6 months as compared with 4.4 months for fluorouracil-based regimen; $p = 0.0022$) [5]. The FOLFIRINOX protocol including fluorouracil, leucovorin, irinotecan, and oxaliplatin is now the

first-line option for patients with metastatic pancreas cancer or in adjuvant settings with a better efficacy compared to gemcitabine [6,7]. Nab-paclitaxel also represents an alternative in association with gemcitabine in metastatic PDAC patients, with a median overall survival of 8.5 months (hazard ratio (HR) 0.72; 95% confidence interval (CI) 0.62, 0.83; $p < 0.001$) [8].

The POLO clinical trial (Pancreas Cancer Olaparib Ongoing) evaluated the efficacy of maintenance therapy with olaparib, a poly(adenosine diphosphate–ribose) polymerase (PARP) inhibitor, in patients with germline *BRCA* mutation and metastatic pancreatic cancer. Median progression-free survival was significantly better in the olaparib group compared to placebo (7.4 months versus 3.8; HR 0.53; 95% CI 0.35, 0.82; $p = 0.004$) [9]. However, targeted therapies still have a limited role due to the lack of understanding of the complex molecular biology of PDAC.

Genomic analyses of PDAC revealed a large panel of molecular alterations particularly affecting Kirsten rat sarcoma viral oncogene (*KRAS*), tumor protein P53 *(TP53)*, Mothers Against Decapentaplegic Homolog 4 (*SMAD4)*, and cyclin-dependent kinase inhibitor 2A *(CDKN2A)* genes [10,11]. *SMAD4* gene is inactivated in approximately 60% of PDAC cases [12], and it is an effector of the transforming growth factor β (TGFβ) signaling pathway which is also altered in 47% of PDAC cases [13]. Given the multiple roles of TGFβ in cancer and its impact on PDAC, it seems interesting to focus on its effector SMAD4. This review aims to summarize current knowledge concerning SMAD4 in TGFβ pathway in patients with PDAC and to discuss the potential SMAD4 molecular targeting in therapeutic.

2. Methods

2.1. Search Strategy

This review was conducted through a systematic review according to the directions denoted by the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA). To investigate the entirety of the published studies on SMAD4 and the TGFβ pathway in patients with pancreatic ductal adenocarcinoma, a comprehensive literature search of the electronic database PubMed was performed up to April 2020. Studies were selected using the following search terms: "PDAC" and "TGFβ" and "SMAD4".

3. Results

3.1. TGFβ Signaling Pathways

TGFβ is a ubiquitously expressed cytokine belonging to a family composed of two branches: the TGFβ branch, represented by ligands such as TGFβ, activin, nodal, or myostatin, and the bone morphogenetic protein (BMP) branch, represented by ligands such as BMP, and growth and differentiation factor (GDF) [14]. TGFβ firstly provides a versatile means of driving developmental programs in mammalian; then, it acts in adult homeostasis with regulation of tissue repair, wound healing, and immune response. There is also a wide panel of cell-type specific biological TGFβ activities such as differentiation, cell-cycle arrest, migration, adhesion, apoptosis, or cancer biology [15–17]. TGFβ is also a potent inducer of epithelial-to-mesenchymal transition (EMT), a well-coordinated process during embryonic development and a pathological feature in neoplasia and fibrosis [15]. TGFβ signaling pathway is activated through either a SMAD-dependent or a SMAD-independent process.

3.1.1. The Canonical TGFβ Signaling Pathway

TGFβ signaling is mediated in most cells through three cell-surface proteins: the two serine/threonine kinase receptors TGFβ receptor I (TβR-I) and TGFβ receptor II (TβR-II), and the TGFβ receptor III (TβR-III). TGFβ ligands can bind directly to TβR-II. TβR-II is constitutively active and can phosphorylate TβR-I, leading to its activation and the propagation of the signal through the phosphorylation of SMAD proteins in their C-terminal serines (SXS) motif [16]. The protein phosphatase 2a (PP2a) can be recruited to dephosphorylate TβR-I, thereby preventing further

signaling [18]. TβR-I is able to activate different members of the receptor regulated SMAD (R-SMAD) family [19]. In the presence of ligands from the TGFβ branch, TβR-I induces the activation of SMAD2 and 3, whereas it activates SMAD1, 5, and 8 in the presence of ligands from the BMP branch [20]. Recognition of R-SMADs may be facilitated by auxiliary proteins, such as the SMAD anchor for receptor activation (SARA) that immobilizes SMAD2 and SMAD3 near the cell surface [21]. Once activated, the R-SMADs undergo a change of conformation that increases their affinity for SMAD4, a co-mediator SMAD (Co-SMAD) distributed throughout the cytoplasm. R-SMAD and Co-SMAD form a heterotrimeric complex, which is translocated and accumulated into the nucleus. SMAD proteins act as transcription factors by binding to the SMAD binding element (SBE) due to their Mad-Homology 1 (MH1) domain [21]. SMAD activates transcriptional TGFβ responses through the Mad-Homology 2 (MH2) domain [22] in association with a transcriptional co-activator, such as CREB Binding Protein (CBP)/p300 (Figure 1) [20,21,23].

Activated R-SMADs can be submitted to a second phosphorylation in their interdomain linker region by cyclin-dependent kinases (CDK) 8 and 9, which favors interactions with a co-activator and enhances their transcriptional activity [19]. On the other hand, CDK8 and CDK9 negatively regulate SMAD levels by recruiting glycogen synthase kinase 3 (GSK3) for proteasomal degradation by E3 ubiquitin ligases [24]. This SMAD-dependent signaling pathway can also be negatively regulated by the inhibitory SMAD (I-SMAD) family, including SMAD6 and 7. I-SMADs are expressed through ligand-induced signaling and act as an auto-inhibition feedback to regulate TGFβ signaling pathway [14]. SMAD7 expression can also be regulated through activation of epidermal growth factor receptor (EGFR), interferon γ signaling by STAT (signal transducer and activator of transcription) protein, or activation of NF-κB by Tumor Necrosis Factor α (TNFα) [14,25,26]. SMAD7 binds to the activated receptors in competition with R-SMADs and leads to ubiquitination and degradation of the receptors with the help of E3 ubiquitin ligases termed SMAD ubiquitination regulatory factors (Smurfs) 1 and 2 [21]. Smurfs also interact directly with R-SMADs in the cytoplasm; Smurf 1 is more specific to BMP responses by targeting SMAD1 and SMAD5, whereas Smurf 2 has the ability to interact with all R-SMADs [25]. SMAD6 can also negatively regulate TGFβ signaling pathway through interfering with the heterotrimerization of the BMP-activated SMADs and SMAD4 [25].

3.1.2. Non-canonical TGFβ Signaling Pathway

SMAD signaling is crucial for most but not all TGFβ-regulated transcriptional responses. SMAD-independent signaling cascades can also operate in a context-dependent manner including the mitogen-activated protein kinase (MAPK) and the phosphatidylinositol 3-kinase (PI3K) pathways, as well as small GTPases [25,27,28].

The activation of TβR-I phosphorylates Src homology domain 2-containing protein (Shc) and induces its association with the adaptor protein growth factor receptor-binding protein 2 (Grb2) and the GTP exchange factor SOS. This complex enables Ras to become its active form that can enhance the kinase cascade with c-Raf, MEK1 or MEK2, and Erk1 or Erk2 [29]. TGFβ also induces p38 and JNK pathways through the activation of the MAP kinase kinase kinase (MAP3K) TGFβ-activated kinase 1 (TAK1), which activates the MAP kinase kinases (MKKs) MKK3/6 and MKK4, respectively. TGFβ receptors induce the ubiquitination of the TNF receptor-associated factor 6 (TRAF6), which is a critical activator of TAK1. Then, TAK1 can activate p38 and JNK MAPK pathways [30,31]. TGFβ signaling pathway may induce NF-κB signaling due to TAK1 phosphorylation and activation of inhibitor of nuclear factor kappa B (IκB) [25]. TβR-II was shown to be a target for Src, which is an oncoprotein able to phosphorylate TβR-II on a tyrosine residue. Src coordinates the docking of Grb2 and Shc, and it contributes in this manner to TGFβ-induced activation of the p38 MAPK pathway [30,32].

In response to TGFβ, TβR-I was shown to interact with the catalytic subunit of PI3K, leading to the activation of Akt and mTOR (mammalian target of rapamycin) downstream proteins [18,27]. mTOR serine/threonine protein kinase can phosphorylate the two principle effectors S6 kinase 1 (S6K1) and the eukaryotic initiation factor 4E-binding protein 1 (4E-BP1), both acting on the translation of proteins that contribute to EMT [33].

TGFβ can activate Rho-like GTPases including RhoA, Rac1, and Cdc42 [25]. Rac and Cdc42 interact with MAP3K, which regulates JNK and p38 signaling [25]. Rho GTPases are required for the organization of the microtubules and actin cytoskeletons. Therefore, the TGFβ-induced activation of the small GTPases seems to contribute to the regulation of cell adhesion and migration [18]. Indeed, Rho and Rac were identified to mediate the pro-EMT, pro migratory, and invasive effects of TGFβ in normal and in cancer epithelial cells [34].

It appears clear that the non-canonical TGFβ pathway is linked to cancer features such as migration, invasion, and EMT. Interestingly, it was described that the non-canonical TGFβ pathway is inappropriately activated in metastatic cancer cells and interferes with the SMAD2/3-mediated tumor suppressing message [17].

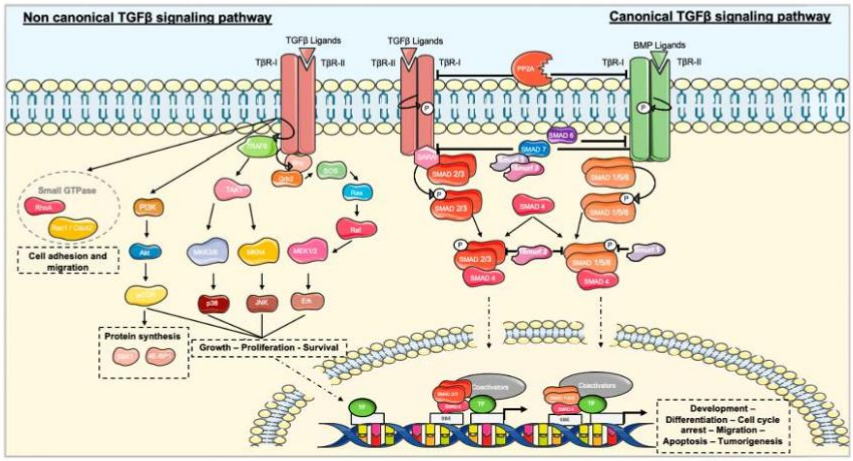


Figure 1. TGFβ signaling is activated through BMP or TGFβ ligands. Ligands bind the TGFβ receptor II (TβR-II) which recruits and phosphorylates TGFβ receptor I (TβR-I). In the canonical TGFβ pathway, TβR-I phosphorylates regulated SMAD (R-SMAD), SMAD2 and 3, or SMAD1, 5, and 8. SMAD4 forms a heterotrimeric complex with R-SMAD. Once translocated to the nucleus, SMAD proteins interact with transcription factors (TF) and a co-activator/co-repressor to regulate gene transcription. Signaling is stopped through inhibitory SMADs (I-SMAD), SMAD6 and 7, with the help of SMAD ubiquitination regulatory factors (Smurfs) 1 and 2. Phosphatase PP2A dephosphorylates receptors to stop the signal. Non-canonical TGFβ signaling regroups MAPKs including Erk, p38, JNK, PI3k/Akt, and the small GTPase pathway.

3.2. TGFβ and PDAC

TGFβ was identified as a major signaling pathway in PDAC. Recently, a study indicated that elevated serum soluble TGFβ levels at diagnosis predict poor progression-free survival (PFS) and overall survival (OS) in unresectable pancreatic cancer compared to patients with a low level of soluble TGFβ (OS 13.7 vs. 9.2 months; HR 2.602; $p = 0.004$; PFS 9.0 vs. 5.8 months; HR 2.010; $p = 0.034$) [35]. Another study in patients with locally advanced and metastatic pancreatic cancers reported that higher soluble TGFβ level exhibited poor OS than lower serum soluble TGFβ (HR 1.35; 95% CI 1.07, 1.69; $p = 0.011$) [36]. These observations corroborate the "TGFβ paradox"; TGFβ can be both a powerful tumor suppressor and a tumor promoter in a context-dependent manner [15].

3.2.1.TGFβ as a Tumor Suppressor

In normal and premalignant cells, TGFβ plays a role in homeostasis and promotes tumor suppression directly through cellular autonomous effects such as cytostasis, differentiation, or apoptosis (Figure 2) [15]. When cell proliferation rate is high, the TGFβ pathway is involved in

cytostatic effects, stimulates cellular differentiation, or triggers proapoptotic mechanisms. In cell-cycle regulation, TGFβ suppresses the expression of CDKs that are crucial for Gap 1 (G1) phase progression, TGFβ also enhances the expression of CDK inhibitors (CDKIs) such as p15 and p21. TGFβ represses the expression and function of c-Myc, a well-known transcription factor implied in cell proliferation [15,16]. TGFβ also drives precursor cells into a less proliferative stage, or it can lead cells to apoptosis [37].

It was suggested that the tumor-suppressive role of TGFβ is only effective when the TGFβ signaling pathway is not defective [38]. However, in PDAC alterations of TGFβ signaling through the mutation of genes involved in the pathway (e.g., *SMAD4*, *SMAD3*, *TβR-I*, *TβR-II*, *ACVR1B*, and *ACVR2A*), this role is present in 47% of cases [13], highlighting that TGFβ can sometimes acts as a tumor promoter.

3.2.2. TGFβ as a Tumor Promoter

TGFβ signaling was identified as a tumor suppressor in normal pancreatic cells and in stages I and II PDAC by inhibiting cell proliferation, but it also showed tumorigenic activity in late-stage malignancy [39]. A high level of TGFβ signaling in PDAC was associated with poor prognosis despite its clear tumor-suppressive effect [40].

In non-cancer situations, TGFβ initially prevents premalignant cells through its tumor suppressor activities; however, in advanced stages of malignant cancer, cells can turn TGFβ signaling to their advantage [15]. TGFβ indirectly promotes tumor suppression by acting on the cellular micro environment; however, in PDAC, TGFβ was shown to induce stromal proliferation in the tumor microenvironment and promote EMT (Figure 2) [41–43].

Pancreatic cancer is characterized by a desmoplastic stroma enriched with small blood vessels and extracellular matrix proteins partly produced by cancer-associated fibroblasts (CAFs) and inflammatory cells [44]. TGFβ contributes to PDAC desmoplasia by enhancing the conversion of fibroblasts or endothelial cells into myofibroblasts also known as CAFs [45,46]. Aggressiveness is amplified by immune cells and fibroblasts in the tumor microenvironment, which can produce high levels of TGFβ in a paracrine manner [46,47]. TGFβ can induce proangiogenic factors such as vascular endothelial growth factor (VEGF), allowing PDAC progression, invasion, and metastases [46].

PDAC-infiltrating lymphocytes were shown to improve prognosis for patients with high rates of cluster of differentiation 4 (CD4), CD8, and dendritic cells [48]. As an immunosuppressive cytokine, TGFβ inhibits the function and development of innate and adaptive immune systems including macrophages, natural killers, dendritic cells, and T cells. TGFβ also stimulates regulatory T cells, which inhibit the function of other lymphocytes [15,47,49]. These suppressive functions confer to TGFβ one of many cancer hallmarks with avoiding immune destruction. On the other hand, others suggested that defective TGFβ signaling, as observed in PDAC, could contribute to an excessive inflammation in favor of tumor progression [15,47].

TGFβ triggers EMT through induction of the expression of specific transcription factors Snail and Zeb1/2 [50]. EMT provides migratory and invasive behaviors to the cells due to cell adhesion modifications. This process involves a loss of epithelial features (i.e., E-cadherin and other components of epithelial cell junctions) and the acquisition of mesenchymal features leading to motility and invasive properties [50]. In PDAC, EMT represents an important process leading to the progression and metastasis of cancer cells.

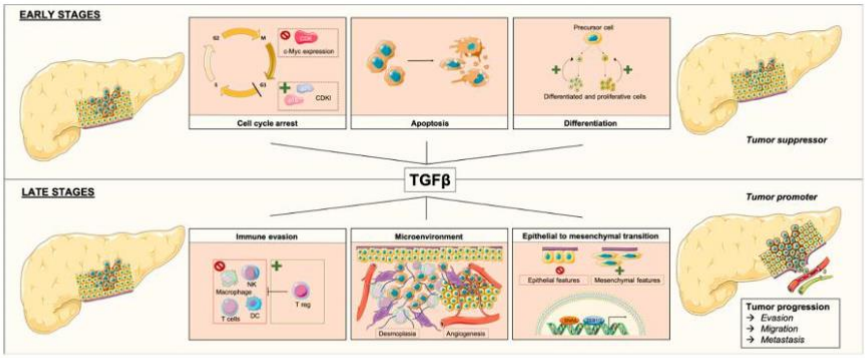


Figure 2. Effects of TGFβ depending of pancreatic ductal adenocarcinoma stages. At early stages, TGFβ acts as a tumor suppressor, by promoting cell-cycle arrest through repression of c-Myc expression and cyclin-dependent kinase (CDK), and through enhancement of CDK inhibitor (CDKI) expression, thereby enhancing apoptosis and stimulating differentiation of precursor cells into a less proliferative stage. At late stages, TGFβ acts as a tumor promoter through immune evasion by stimulation of regulatory T cells and by inhibition of natural killers (NKs), macrophages, T cells, and dendritic cells (DCs). TFGβ acts also on the microenvironment by promoting desmoplasia and angiogenesis, and it triggers epithelial-to-mesenchymal transition (EMT) through induction of expression of transcription factors Snail and Zeb1/2.

3.3. SMAD4

SMAD4 or DPC4 (deleted in pancreatic cancer locus 4) is an intracellular transcriptional mediator of the TGFβ signaling pathway encoded by the *SMAD4* gene which is located on the 18q21.2 human chromosome locus and is composed of 12 coding exons and 10 introns (GRCh38, NM_005359.5). The SMAD4 protein has a molecular weight of 60 kDa and is composed of 552 amino acids [12,51]. SMAD4 is composed of MH1 and MH2 domains connected by a linker region involved in DNA binding and its transcriptional activity [51].

SMAD4 activity and stability are regulated by three major signaling pathways through its linker region that contains a growth factor-regulated transcription activation domain integrating Wnt, fibroblast growth factor (FGF)/epidermal growth factor (EGF), and TGFβ signaling [52]. SMAD4 can be phosphorylated and activated through the Erk pathway in response to FGF and EGF. However, this MAPK pathway triggers phosphorylation of SMAD4 by GSK3, which creates a phosphodegron leading SMAD4 to its polyubiquitination and degradation by the E3-ligase β-TrCP. The Wnt pathway can regulate the phosphodegron, and it inhibits the binding of β-TrCP to SMAD4. In the presence of Wnt, SMAD4 phosphorylation by GSK3 is inhibited and TGFβ signaling is consequently prolonged [52].

3.4. SMAD4 and PDAC

The most common driver mutational events in PDAC occur in four cancer genes *KRAS*, *TP53*, *SMAD4*, and *CDKN2A*. *KRAS* is activated by mutation in almost 90% of cases, *TP53* is inactivated in 50%–75% of cases, and *CDK2NA* is inactivated in about 30% of patients with PDAC [10,53]. *SMAD4* was firstly identified as a tumor suppressor gene that is inactivated in 60% of PDAC cases due to homozygous deletion or mutation [12]. The *SMAD4* gene is located on the q arm of chromosome 18, which is frequently deleted hemi- or homozygously in PDAC [54]. Alterations of the *SMAD4* gene occur in late stages when the carcinoma is histologically recognizable [55]. In pancreatic cancers, homozygous deletion of SMAD4 is found in approximately 30% of cases, inactivation is found in 20% of cases, and allelic loss of chromosome 18q is found in almost 90% of cases [56]. Hotspot mutations (i.e., nonsense, missense, or frameshift indels) in the C-terminal MH2 domain of *SMAD4*

were described [56]. Nonsense substitutions were identified in 25% of cases, missense substitutions were identified in 37% of cases, and frameshift insertion and deletion were identified in 7% and 5% of cases, respectively (GRCh38. COSMIC v91). *SMAD4* mutations lead to a non-functional or truncated SMAD4 protein or to the protein degradation through the ubiquitin–proteasome pathway [57]. Notably, two SMAD4 missense mutations (Pro130Ser and Asn351His) were found to increase SMAD4 phosphorylation by GSK3 and lead to protein degradation [57].

3.4.1. SMAD4 and EMT

Loss of SMAD4 inhibits the tumor suppressor effects of TGFβ without affecting tumor response, promoting a more aggressive phenotype [58]. *SMAD4* knockdown in PDAC cell lines showed SMAD4 as a regulator of genes involved in cell proliferation, adhesion, and motility [59]. In another study using *SMAD4* silencing, SMAD4 was required in TGFβ-induced cell-cycle arrest and migration. Regarding EMT, authors found that *SMAD4* silencing did not abolish EMT, suggesting that EMT can be conducted not only through SMAD4 signaling [58].

TGFβ can induce EMT in a SMAD4-dependent manner through the expression of Snail and Zeb1 transcription factors and the decrease of E-cadherin epithelial marker [60]. SMAD4 deficiency in PDAC cells induces E-cadherin expression and is associated with an ineffectiveness of TGFβ to induce EMT [61]. However, a study of patients with PDAC examined the clinical association between SMAD4 expression and EMT status and revealed that SMAD4 suppression was associated with a mesenchymal phenotype [62], highlighting the complex mechanistic of EMT. An increase in in the non-canonical TGFβ pathways was demonstrated in SMAD4-defective PDAC cells [63]. This SMAD4-independent pathway activates the PI3K/Akt/mTOR pathway and induces small GTPases, which can contribute to EMT [33,34,53]. SMAD4 can suppress JNK activation; therefore, in the case of SMAD4 deletion, JNK activity is subsequently enhanced [64] and can be required in the EMT response [65].

TGFβ-induced EMT is commonly linked with tumor progression; however, a study demonstrated that TGFβ can also induce a SMAD4-dependent lethal EMT in PDAC cells. In SMAD4 wild-type cells, TGFβ induces the conversion of the transcription factor Sox4 from an enhancer of tumorigenesis into a promoter of apoptosis. In this process, Snail induces the repression of Klf5, which is associated with Sox4 to prevent Sox4-mediated apoptosis in oncogenesis [42]. This SMAD4-dependent mechanism can provide an explanation about the dual role of TGFβ in tumors.

3.4.2. Effect of SMAD4 Expression in Patients with PDAC

Loss of SMAD4 expression occurs late in PDAC tumor progression [62] and is correlated with a higher metastatic burden [66], as well as malignant phenotypes of PDAC, including degree of differentiation, tumor size, and lymph node metastasis [67]. Patients with SMAD4 loss are commonly associated with distant disease progression, whereas patients with wild-type SMAD4 harbored a local tumor pattern of progression [62].

The prognostic effects of SMAD4 loss in PDAC patients are controversial. It was found that *SMAD4* gene inactivation, by intragenic mutation or homozygous deletion, leads to poor prognosis for patients with surgically resected pancreatic adenocarcinoma [68]. SMAD4 loss was associated with shorter OS (18.3 months versus 30.1 months for preserved SMAD4; $p < 0.001$) and disease-free survival (DFS) (6.0 months versus 13.5 months for preserved SMAD4; $p < 0.001$) in patients with resectable pancreatic cancer [69]. In advanced pancreatic cancers treated by gemcitabine, SMAD4 loss had no impact on OS (HR 1.008; $p = 0.656$) but was associated with a prolonged PFS (HR 1.565, $p = 0.038$) [70]. Recently, an analysis of publicly available datasets from The Cancer Genome Atlas (TGCA) indicated that PDAC patients with a deleted *SMAD4* gene had poorer DFS; however, *SMAD4* alterations did not predict the OS [71]. In another study, SMAD4 expression status was neither associated with early death (OR 0.5; $p = 0.15$) nor associated with recurrence pattern (OR 0.9; $p = 0.9$) in patients with resected PDAC [72]. Further investigations are needed to understand the prognostic value of SMAD4.

4. Discussion

SMAD4 as a Potential Target in Patients with PDAC and Impact on Other Therapies

SMAD4 plays a central role in TGFβ signaling and seems to regulate the interruption of the cell cycle; it also improves apoptosis, decreases invasion and metastasis, and indirectly affects the prognosis of patients with PDAC. SMAD4 is, thus, designated as a potential target in the management of patients with PDAC.

A study showed that SMAD4-deficient cell lines are modestly less sensitive to gemcitabine and 2.0-fold and 4.5-fold more sensitive to cisplatin and irinotecan, respectively [73]. Controversially, a recent study found that SMAD4-deleted PDAC cells are more sensitive to gemcitabine, while *SMAD4* mutated cells had similar sensitivity to gemcitabine as *SMAD4* wild-type cells, suggesting that the SMAD4 copy number could be used as a therapeutic biomarker for PDAC treatment with gemcitabine [71]. This study also found that SMAD4-deleted PDAC cells are sensitive to other agents modulating the cell cycle including cytarabine, clofarabine, darinaparsin, and olaparib, due to the upregulation of cell cycle-related genes such as CDK1 [71]. Synthetic lethality-based screening methodologies identified two compounds (UA62001 and UA62784) to selectively target SMAD4-negative cells. In SMAD4-negative pancreatic cancer cells treated with the compound UA62001, the cell cycle was interrupted in synthesis (S) and Gap 2 (G2) – mitotic (M) phases [74], whereas UA62784 activated CDK1 and induced mitotic cell-cycle arrest and apoptosis [75]. These results confirmed results from Hsieh *et al.* and supported that SMAD4 determination status at initial diagnosis may be valuable to stratify patients for personalized chemotherapy (Figure 3).

SMAD4 deficiency can enhance the glycolytic capacity of cancer cells with upregulation of glucose transporter expression contributing to aerobic glycolysis, which is the main metabolic pathway in PDAC [76]. PDAC harboring mutant SMAD4 exhibited a high metabolic tumor burden [66]. A subsequent study identified the glycolytic enzyme phosphoglycerate kinase 1 (PGK1) as a target for TGFβ/SMAD4. In PDAC, SMAD4 loss mediates PGK1 upregulation, which enhances glycolysis and contributes to aggressive tumor behavior. This study identified PGK1 as a decisive oncogene in patient without SMAD4 and, therefore, a potential target in the development of therapeutics for SMAD4-deficient PDAC (Figure 3) [77].

Different studies used a microarray-based strategy and identified miR-301a-3p, miR-483-3p, and miR-421 to target SMAD4 in PDAC. Overexpression of these microRNAs (miRNAs) promoted cell invasion, migration, and colony formation [78–80]. miRNA inhibitors that prevent the binding of miRNA to their mRNA targets, may have a promising effect on PDAC progression (Figure 3). Further researches need to be conducted to evaluate the therapeutic and diagnostic potential of these SMAD4 miRNA regulators.

SMAD4 mutations can interfere with its regulation associated with the Wnt/GSK3 and Erk pathways. Mutations can enhance phosphorylation by GSK3 and lead to a loss of function by protein degradation. A study suggested a therapeutic approach with GSK3 inhibitor in order to stabilize SMAD4 and restore TGFβ signaling [57]. The authors demonstrated that GSK3 inhibition reactivated TGFβ signaling in SMAD4 mutated cells in colon cancer. The potential therapeutic application of GSK3 inhibitors may be promising and requires further study in PDAC models.

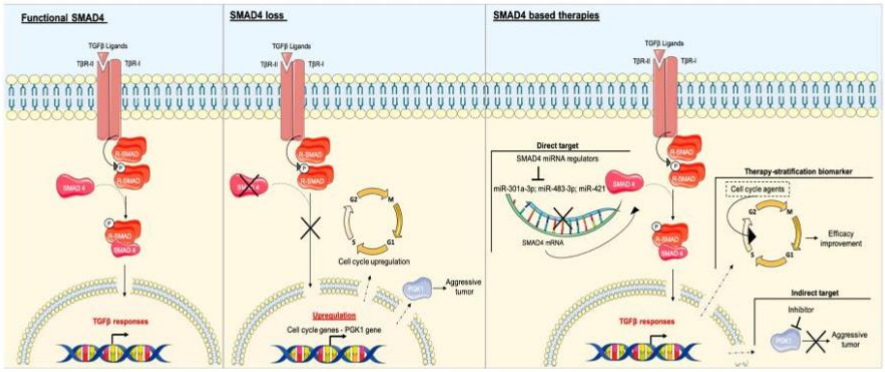


Figure 3. Effects of SMAD4 loss and potential therapeutic strategies in SMAD4 signaling. In normal cells, SMAD4 acts as an intracellular mediator of TGFβ. In SMAD4-null cells, loss of SMAD4 leads to upregulation of the cell cycle and phosphoglycerate kinase 1 (PGK1), leading to tumor promotion. Therapies based on SMAD4 can be direct by targeting SMAD4 miRNA regulators to modulate SMAD4 or indirect by targeting a downstream component. SMAD4 loss leads to the enhanced efficacy of agents modulating the cell cycle, whereby SMAD4 as a biomarker could be useful to stratify therapies.

5. Conclusions

PDAC is projected to become the second leading cause of cancer-related death in the next decade. This alarming situation prompts the need to understand molecular mechanisms of PDAC to develop new therapies. The TGFβ signaling pathway is a powerful tumor suppressor in early stages, promoting cell-cycle arrest, differentiation, and apoptosis through SMAD-dependent signaling. In late stages, TGFβ switches its role to become a tumor promoter, favoring desmoplasia, angiogenesis, immune evasion, and EMT through MAPK, PI3K, or small GTPase pathways. SMAD4 is a key partner to TGFβ canonical signaling. In PDAC, the *SMAD4* gene is inactivated through homozygous deletion or mutation in 60% of cases. Inactivation of *SMAD4* leads to a modification in TGFβ responses, promoting non-canonical TGFβ signaling which is linked to pro-tumorigenic responses. It is still controversial whether SMAD4 loss has a real influence on the prognosis of patients with PDAC, and further investigations are needed. SMAD4 is a potential direct or indirect therapeutic target for the management of patients with PDAC, and it may also have a potential role as a therapy stratification biomarker. Patients with PDAC may benefit from SMAD4-based therapeutic strategies, but the SMAD-independent TGFβ signaling pathway may bypass strategies, leading to an innate resistance mechanism.

Author Contributions: Writing—original draft preparation, J.D.; writing—review and editing, J.D., A.H., A.W., and P.G.; supervision, A.H. and P.G.; project administration, J.-L.M. and A.H. All authors read and approved the final manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Collisson, E.A.; Bailey, P.; Chang, D.K.; Biankin, A.V. Molecular subtypes of pancreatic cancer. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16*, 207–220, doi:10.1038/s41575-019-0109-y.
2. Kleeff, J.; Korc, M.; Apte, M.; Vecchia, C.L.; Johnson, C.D.; Biankin, A.V.; Neale, R.E.; Tempero, M.; Tuveson, D.A.; Hruban, R.H.; et al. Pancreatic cancer. *Nat. Rev. Dis. Primer* **2016**, *2*, 1–22, doi:10.1038/nrdp.2016.22.

3. Rahib, L.; Smith, B.D.; Aizenberg, R.; Rosenzweig, A.B.; Fleshman, J.M.; Matrisian, L.M. Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **2014**, *74*, 2913–2921, doi:10.1158/0008-5472.CAN-14-0155.
4. Lambert, A.; Schwarz, L.; Borbath, I.; Henry, A.; Van Laethem, J.-L.; Malka, D.; Ducreux, M.; Conroy, T. An update on treatment options for pancreatic adenocarcinoma. *Ther. Adv. Med. Oncol.* **2019**, *11*, 1758835919875568, doi:10.1177/1758835919875568.
5. Burris, H.A.; Moore, M.J.; Andersen, J.; Green, M.R.; Rothenberg, M.L.; Modiano, M.R.; Cripps, M.C.; Portenoy, R.K.; Storniolo, A.M.; Tarassoff, P.; et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: A randomized trial. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **1997**, *15*, 2403–2413, doi:10.1200/JCO.1997.15.6.2403.
6. Conroy, T.; Hammel, P.; Hebbar, M.; Ben Abdelghani, M.; Wei, A.C.; Raoul, J.-L.; Choné, L.; Francois, E.; Artru, P.; Biagi, J.J.; et al. FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. *N. Engl. J. Med.* **2018**, *379*, 2395–2406, doi:10.1056/NEJMoa1809775.
7. Conroy, T.; Desseigne, F.; Ychou, M.; Bouché, O.; Guimbaud, R.; Bécouarn, Y.; Adenis, A.; Raoul, J.-L.; Gourgou-Bourgade, S.; de la Fouchardière, C.; et al. FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. *N. Engl. J. Med.* **2011**, *364*, 1817–1825, doi:10.1056/NEJMoa1011923.
8. Von Hoff, D.D.; Ervin, T.; Arena, F.P.; Chiorean, E.G.; Infante, J.; Moore, M.; Seay, T.; Tjulandin, S.A.; Ma, W.W.; Saleh, M.N.; et al. Increased Survival in Pancreatic Cancer with nab-Paclitaxel plus Gemcitabine. *N. Engl. J. Med.* **2013**, *369*, 1691–1703, doi:10.1056/NEJMoa1304369.
9. Golan, T.; Hammel, P.; Reni, M.; Van Cutsem, E.; Macarulla, T.; Hall, M.J.; Park, J.-O.; Hochhauser, D.; Arnold, D.; Oh, D.-Y.; et al. Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer. *N. Engl. J. Med.* **2019**, *381*, 317–327, doi:10.1056/NEJMoa1903387.
10. Raphael, B.J.; Hruban, R.H.; Aguirre, A.J.; Moffitt, R.A.; Yeh, J.J.; Stewart, C.; Robertson, A.G.; Cherniack, A.D.; Gupta, M.; Getz, G.; et al. Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma. *Cancer Cell* **2017**, *32*, 185–203.e13, doi:10.1016/j.ccell.2017.07.007.
11. Waddell, N.; Pajic, M.; Patch, A.-M.; Chang, D.K.; Kassahn, K.S.; Bailey, P.; Johns, A.L.; Miller, D.; Nones, K.; Quek, K.; et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **2015**, *518*, 495–501, doi:10.1038/nature14169.
12. Hahn, S.A.; Schutte, M.; Hoque, A.T.; Moskaluk, C.A.; da Costa, L.T.; Rozenblum, E.; Weinstein, C.L.; Fischer, A.; Yeo, C.J.; Hruban, R.H.; et al. DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **1996**, *271*, 350–353, doi:10.1126/science.271.5247.350.
13. Bailey, P.; Chang, D.K.; Nones, K.; Johns, A.L.; Patch, A.-M.; Gingras, M.-C.; Miller, D.K.; Christ, A.N.; Bruxner, T.J.C.; Quinn, M.C.; et al. Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **2016**, *531*, 47–52, doi:10.1038/nature16965.
14. Moustakas, A.; Souchelnytskyi, S.; Heldin, C.H. Smad regulation in TGF-beta signal transduction. *J. Cell Sci.* **2001**, *114*, 4359–4369.
15. Massagué, J. TGFbeta in Cancer. *Cell* **2008**, *134*, 215–230, doi:10.1016/j.cell.2008.07.001.
16. Tian, M.; Neil, J.R.; Schiemann, W.P. Transforming growth factor-β and the hallmarks of cancer. *Cell. Signal.* **2011**, *23*, 951–962, doi:10.1016/j.cellsig.2010.10.015.
17. Wendt, M.K.; Tian, M.; Schiemann, W.P. Deconstructing the mechanisms and consequences of TGF-β-induced EMT during cancer progression. *Cell Tissue Res.* **2012**, *347*, 85–101, doi:10.1007/s00441-011-1199-1.
18. Mu, Y.; Gudey, S.K.; Landström, M. Non-Smad signaling pathways. *Cell Tissue Res.* **2012**, *347*, 11–20, doi:10.1007/s00441-011-1201-y.
19. Massagué, J. TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 616–630, doi:10.1038/nrm3434.
20. Massagué, J.; Gomis, R.R. The logic of TGFbeta signaling. *FEBS Lett.* **2006**, *580*, 2811–2820, doi:10.1016/j.febslet.2006.04.033.
21. Shi, Y.; Massagué, J. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **2003**, *113*, 685–700, doi:10.1016/s0092-8674(03)00432-x.
22. Macias, M.J.; Martin-Malpartida, P.; Massagué, J. Structural determinants of Smad function in TGF-β signaling. *Trends Biochem. Sci.* **2015**, *40*, 296–308, doi:10.1016/j.tibs.2015.03.012.
23. Feng, X.-H.; Derynck, R. Specificity and versatility in tgf-beta signaling through Smads. *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 659–693, doi:10.1146/annurev.cellbio.21.022404.142018.

24. Gao, S.; Alarcón, C.; Sapkota, G.; Rahman, S.; Chen, P.-Y.; Goerner, N.; Macias, M.J.; Erdjument-Bromage, H.; Tempst, P.; Massagué, J. Ubiquitin Ligase Nedd4L Targets Activated Smad2/3 to Limit TGF-β Signaling. *Mol. Cell* **2009**, *36*, 457–468, doi:10.1016/j.molcel.2009.09.043.
25. Derynck, R.; Zhang, Y.E. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature* **2003**, *425*, 577–584, doi:10.1038/nature02006.
26. Massagué, J. How cells read TGF-beta signals. *Nat. Rev. Mol. Cell Biol.* **2000**, *1*, 169–178, doi:10.1038/35043051.
27. Lamouille, S.; Derynck, R. Cell size and invasion in TGF-beta-induced epithelial to mesenchymal transition is regulated by activation of the mTOR pathway. *J. Cell Biol.* **2007**, *178*, 437–451, doi:10.1083/jcb.200611146.
28. Moustakas, A.; Heldin, C.-H. Non-Smad TGF-beta signals. *J. Cell Sci.* **2005**, *118*, 3573–3584, doi:10.1242/jcs.02554.
29. Lee, M.K.; Pardoux, C.; Hall, M.C.; Lee, P.S.; Warburton, D.; Qing, J.; Smith, S.M.; Derynck, R. TGF-beta activates Erk MAP kinase signalling through direct phosphorylation of ShcA. *EMBO J.* **2007**, *26*, 3957–3967, doi:10.1038/sj.emboj.7601818.
30. Kang, J.S.; Liu, C.; Derynck, R. New regulatory mechanisms of TGF-beta receptor function. *Trends Cell Biol.* **2009**, *19*, 385–394, doi:10.1016/j.tcb.2009.05.008.
31. Yamashita, M.; Fatyol, K.; Jin, C.; Wang, X.; Liu, Z.; Zhang, Y.E. TRAF6 mediates Smad-independent activation of JNK and p38 by TGF-beta. *Mol. Cell* **2008**, *31*, 918–924, doi:10.1016/j.molcel.2008.09.002.
32. Galliher, A.J.; Schiemann, W.P. Src Phosphorylates Tyr284 in TGF-β Type II Receptor and Regulates TGF-β Stimulation of p38 MAPK during Breast Cancer Cell Proliferation and Invasion. *Cancer Res.* **2007**, *67*, 3752–3758, doi:10.1158/0008-5472.CAN-06-3851.
33. Lamouille, S.; Derynck, R. Emergence of the phosphoinositide 3-kinase-Akt-mammalian target of rapamycin axis in transforming growth factor-β-induced epithelial-mesenchymal transition. *Cells Tissues Organs* **2011**, *193*, 8–22, doi:10.1159/000320172.
34. Ungefroren, H.; Witte, D.; Lehnert, H. The role of small GTPases of the Rho/Rac family in TGF-β-induced EMT and cell motility in cancer. *Dev. Dyn.* **2018**, *247*, 451–461, doi:10.1002/dvdy.24505.
35. Park, H.; Bang, J.-H.; Nam, A.-R.; Park, J.E.; Jin, M.H.; Bang, Y.-J.; Oh, D.-Y. The prognostic role of soluble TGF-beta and its dynamics in unresectable pancreatic cancer treated with chemotherapy. *Cancer Med.* **2020**, *9*, 43–51, doi:10.1002/cam4.2677.
36. Javle, M.; Li, Y.; Tan, D.; Dong, X.; Chang, P.; Kar, S.; Li, D. Biomarkers of TGF-β signaling pathway and prognosis of pancreatic cancer. *PLoS ONE* **2014**, *9*, e85942, doi:10.1371/journal.pone.0085942.
37. Derynck, R.; Akhurst, R.J. Differentiation plasticity regulated by TGF-β family proteins in development and disease. *Nat. Cell Biol.* **2007**, *9*, 1000–1004, doi:10.1038/ncb434.
38. Principe, D.R.; Doll, J.A.; Bauer, J.; Jung, B.; Munshi, H.G.; Bartholin, L.; Pasche, B.; Lee, C.; Grippo, P.J. TGF-β: Duality of function between tumor prevention and carcinogenesis. *J. Natl. Cancer Inst.* **2014**, *106*, djt369, doi:10.1093/jnci/djt369.
39. Friess, H.; Yamanaka, Y.; Büchler, M.; Ebert, M.; Beger, H.G.; Gold, L.I.; Korc, M. Enhanced expression of transforming growth factor β isoforms in pancreatic cancer correlates with decreased survival. *Gastroenterology* **1993**, *105*, 1846–1856, doi:10.1016/0016-5085(93)91084-U.
40. Wagner, M.; Kleeff, J.; Friess, H.; Büchler, M.W.; Korc, M. Enhanced Expression of the Type II Transforming Growth Factor-β Receptor Is Associated with Decreased Survival in Human Pancreatic Cancer. *Pancreas* **1999**, *19*, 370–376, doi:10.1097/00006676-199911000-00008.
41. Alvarez, M.A.; Freitas, J.P.; Mazher Hussain, S.; Glazer, E.S. TGF-β Inhibitors in Metastatic Pancreatic Ductal Adenocarcinoma. *J. Gastrointest. Cancer* **2019**, *50*, 207–213, doi:10.1007/s12029-018-00195-5.
42. David, C.J.; Huang, Y.-H.; Chen, M.; Su, J.; Zou, Y.; Bardeesy, N.; Iacobuzio-Donahue, C.A.; Massagué, J. TGF-β Tumor Suppression through a Lethal EMT. *Cell* **2016**, *164*, 1015–1030, doi:10.1016/j.cell.2016.01.009.
43. Nieto, M.A.; Huang, R.Y.-J.; Jackson, R.A.; Thiery, J.P. EMT: 2016. *Cell* **2016**, *166*, 21–45, doi:10.1016/j.cell.2016.06.028.
44. Nielsen, M.F.B.; Mortensen, M.B.; Detlefsen, S. Key players in pancreatic cancer-stroma interaction: Cancer-associated fibroblasts, endothelial and inflammatory cells. *World J. Gastroenterol.* **2016**, *22*, 2678–2700, doi:10.3748/wjg.v22.i9.2678.

45. Zeisberg, E.M.; Potenta, S.; Xie, L.; Zeisberg, M.; Kalluri, R. Discovery of Endothelial to Mesenchymal Transition as a Source for Carcinoma-Associated Fibroblasts. *Cancer Res.* **2007**, *67*, 10123–10128, doi:10.1158/0008-5472.CAN-07-3127.
46. Zhou, Q.; Xia, S.; Guo, F.; Hu, F.; Wang, Z.; Ni, Y.; Wei, T.; Xiang, H.; Shang, D. Transforming growth factor-β in pancreatic diseases: Mechanisms and therapeutic potential. *Pharmacol. Res.* **2019**, *142*, 58–69, doi:10.1016/j.phrs.2019.01.038.
47. Achyut, B.R.; Yang, L. Transforming Growth Factor-β in the Gastrointestinal and Hepatic Tumor Microenvironment. *Gastroenterology* **2011**, *141*, 1167–1178, doi:10.1053/j.gastro.2011.07.048.
48. Seo, Y.D.; Pillarisetty, V.G. T-cell programming in pancreatic adenocarcinoma: A review. *Cancer Gene Ther.* **2017**, *24*, 106–113, doi:10.1038/cgt.2016.66.
49. Pickup, M.; Novitskiy, S.; Moses, H.L. The roles of TGFβ in the tumour microenvironment. *Nat. Rev. Cancer* **2013**, *13*, 788–799, doi:10.1038/nrc3603.
50. Thiery, J.P.; Acloque, H.; Huang, R.Y.J.; Nieto, M.A. Epithelial-mesenchymal transitions in development and disease. *Cell* **2009**, *139*, 871–890, doi:10.1016/j.cell.2009.11.007.
51. Zhao, M.; Mishra, L.; Deng, C.-X. The role of TGF-β/SMAD4 signaling in cancer. *Int. J. Biol. Sci.* **2018**, *14*, 111–123, doi:10.7150/ijbs.23230.
52. Demagny, H.; Araki, T.; De Robertis, E.M. The Tumor Suppressor Smad4/DPC4 Is Regulated by Phosphorylations that Integrate FGF, Wnt, and TGF-β Signaling. *Cell Rep.* **2014**, *9*, 688–700, doi:10.1016/j.celrep.2014.09.020.
53. Polireddy, K.; Chen, Q. Cancer of the Pancreas: Molecular Pathways and Current Advancement in Treatment. *J. Cancer* **2016**, *7*, 1497–1514, doi:10.7150/jca.14922.
54. Furukawa, T.; Sunamura, M.; Horii, A. Molecular mechanisms of pancreatic carcinogenesis. *Cancer Sci.* **2006**, *97*, 1–7, doi:10.1111/j.1349-7006.2005.00134.x.
55. Kubiczkova, L.; Sedlarikova, L.; Hajek, R.; Sevcikova, S. TGF-β—An excellent servant but a bad master. *J. Transl. Med.* **2012**, *10*, 183, doi:10.1186/1479-5876-10-183.
56. Xia, X.; Wu, W.; Huang, C.; Cen, G.; Jiang, T.; Cao, J.; Huang, K.; Qiu, Z. SMAD4 and its role in pancreatic cancer. *Tumor Biol.* **2015**, *36*, 111–119, doi:10.1007/s13277-014-2883-z.
57. Demagny, H.; Robertis, E.M.D. Point mutations in the tumor suppressor Smad4/DPC4 enhance its phosphorylation by GSK3 and reversibly inactivate TGF-β signaling. *Mol. Cell. Oncol.* **2016**, *3*, e1025181, doi:10.1080/23723556.2015.1025181.
58. Levy, L.; Hill, C.S. Smad4 dependency defines two classes of transforming growth factor {beta} (TGF-{beta}) target genes and distinguishes TGF-{beta}-induced epithelial-mesenchymal transition from its antiproliferative and migratory responses. *Mol. Cell. Biol.* **2005**, *25*, 8108–8125, doi:10.1128/MCB.25.18.8108-8125.2005.
59. Jazag, A.; Ijichi, H.; Kanai, F.; Imamura, T.; Guleng, B.; Ohta, M.; Imamura, J.; Tanaka, Y.; Tateishi, K.; Ikenoue, T.; et al. Smad4 silencing in pancreatic cancer cell lines using stable RNA interference and gene expression profiles induced by transforming growth factor-β. *Oncogene* **2005**, *24*, 662–671, doi:10.1038/sj.onc.1208102.
60. Ahmed, S.; Bradshaw, A.-D.; Gera, S.; Dewan, M.Z.; Xu, R. The TGF-β/Smad4 Signaling Pathway in Pancreatic Carcinogenesis and Its Clinical Significance. *J. Clin. Med.* **2017**, *6*, doi:10.3390/jcm6010005.
61. Shichi, Y.; Sasaki, N.; Michishita, M.; Hasegawa, F.; Matsuda, Y.; Arai, T.; Gomi, F.; Aida, J.; Takubo, K.; Toyoda, M.; et al. Enhanced morphological and functional differences of pancreatic cancer with epithelial or mesenchymal characteristics in 3D culture. *Sci. Rep.* **2019**, *9*, 10871, doi:10.1038/s41598-019-47416-w.
62. Yamada, S.; Fujii, T.; Shimoyama, Y.; Kanda, M.; Nakayama, G.; Sugimoto, H.; Koike, M.; Nomoto, S.; Fujiwara, M.; Nakao, A.; et al. SMAD4 Expression Predicts Local Spread and Treatment Failure in Resected Pancreatic Cancer. *Pancreas* **2015**, *44*, 660–664, doi:10.1097/MPA.0000000000000315.
63. Chen, Y.-W.; Hsiao, P.-J.; Weng, C.-C.; Kuo, K.-K.; Kuo, T.-L.; Wu, D.-C.; Hung, W.-C.; Cheng, K.-H. SMAD4 Loss triggers the phenotypic changes of pancreatic ductal adenocarcinoma cells. *BMC Cancer* **2014**, *14*, 181, doi:10.1186/1471-2407-14-181.
64. Zhang, X.; Cao, J.; Pei, Y.; Zhang, J.; Wang, Q. Smad4 inhibits cell migration via suppression of JNK activity in human pancreatic carcinoma PANC-1 cells. *Oncol. Lett.* **2016**, *11*, 3465–3470, doi:10.3892/ol.2016.4427.
65. Moustakas, A.; Heldin, C.-H. Mechanisms of TGFβ-Induced Epithelial–Mesenchymal Transition. *J. Clin. Med.* **2016**, *5*, 63, doi:10.3390/jcm5070063.

66. Shi, S.; Ji, S.; Qin, Y.; Xu, J.; Zhang, B.; Xu, W.; Liu, J.; Long, J.; Liu, C.; Liu, L.; et al. Metabolic tumor burden is associated with major oncogenomic alterations and serum tumor markers in patients with resected pancreatic cancer. *Cancer Lett.* **2015**, *360*, 227–233, doi:10.1016/j.canlet.2015.02.014.
67. Wang, Z.; Li, Y.; Zhan, S.; Zhang, L.; Zhang, S.; Tang, Q.; Li, M.; Tan, Z.; Liu, S.; Xing, X. SMAD4 Y353C promotes the progression of PDAC. *BMC Cancer* **2019**, *19*, 1037, doi:10.1186/s12885-019-6251-7.
68. Blackford, A.; Serrano, O.K.; Wolfgang, C.L.; Parmigiani, G.; Jones, S.; Zhang, X.; Parsons, D.W.; Lin, J.C.-H.; Leary, R.J.; Eshleman, J.R.; et al. SMAD4 Gene Mutations Are Associated with Poor Prognosis in Pancreatic Cancer. *Clin. Cancer Res.* **2009**, *15*, 4674–4679, doi:10.1158/1078-0432.CCR-09-0227.
69. Oshima, M.; Okano, K.; Muraki, S.; Haba, R.; Maeba, T.; Suzuki, Y.; Yachida, S. Immunohistochemically detected expression of 3 major genes (CDKN2A/p16, TP53, and SMAD4/DPC4) strongly predicts survival in patients with resectable pancreatic cancer. *Ann. Surg.* **2013**, *258*, 336–346, doi:10.1097/SLA.0b013e3182827a65.
70. Ormanns, S.; Haas, M.; Remold, A.; Kruger, S.; Holdenrieder, S.; Kirchner, T.; Heinemann, V.; Boeck, S. The Impact of SMAD4 Loss on Outcome in Patients with Advanced Pancreatic Cancer Treated with Systemic Chemotherapy. *Int. J. Mol. Sci.* **2017**, *18*, doi:10.3390/ijms18051094.
71. Hsieh, Y.-Y.; Liu, T.-P.; Chou, C.-J.; Chen, H.-Y.; Lee, K.-H.; Yang, P.-M. Integration of Bioinformatics Resources Reveals the Therapeutic Benefits of Gemcitabine and Cell Cycle Intervention in SMAD4-Deleted Pancreatic Ductal Adenocarcinoma. *Genes* **2019**, *10*, 766, doi:10.3390/genes10100766.
72. Winter, J.M.; Tang, L.H.; Klimstra, D.S.; Liu, W.; Linkov, I.; Brennan, M.F.; D'Angelica, M.I.; DeMatteo, R.P.; Fong, Y.; Jarnagin, W.R.; et al. Failure patterns in resected pancreas adenocarcinoma: Lack of predicted benefit to SMAD4 expression. *Ann. Surg.* **2013**, *258*, 331–335, doi:10.1097/SLA.0b013e31827fe9ce.
73. Cui, Y.; Brosnan, J.A.; Blackford, A.L.; Sur, S.; Hruban, R.H.; Kinzler, K.W.; Vogelstein, B.; Maitra, A.; Diaz, L.A.; Iacobuzio-Donahue, C.A.; et al. Genetically Defined Subsets of Human Pancreatic Cancer Show Unique In Vitro Chemosensitivity. *Clin. Cancer Res.* **2012**, *18*, 6519–6530, doi:10.1158/1078-0432.CCR-12-0827.
74. Wang, H.; Han, H.; Von Hoff, D.D. Identification of an Agent Selectively Targeting DPC4 (Deleted in Pancreatic Cancer Locus 4)–Deficient Pancreatic Cancer Cells. *Cancer Res.* **2006**, *66*, 9722–9730, doi:10.1158/0008-5472.CAN-05-4602.
75. Wang, H.; Stephens, B.; Von Hoff, D.D.; Han, H. Identification and Characterization of a Novel Anticancer Agent With Selectivity Against Deleted in Pancreatic Cancer Locus 4 (DPC4)-Deficient Pancreatic and Colon Cancer Cells. *Pancreas* **2009**, *38*, 551–557, doi:10.1097/MPA.0b013e31819d7415.
76. Liang, C.; Qin, Y.; Zhang, B.; Ji, S.; Shi, S.; Xu, W.; Liu, J.; Xiang, J.; Liang, D.; Hu, Q.; et al. Metabolic plasticity in heterogeneous pancreatic ductal adenocarcinoma. *Biochim. Biophys. Acta* **2016**, *1866*, 177–188, doi:10.1016/j.bbcan.2016.09.001.
77. Liang, C.; Shi, S.; Qin, Y.; Meng, Q.; Hua, J.; Hu, Q.; Ji, S.; Zhang, B.; Xu, J.; Yu, X.-J. Localisation of PGK1 determines metabolic phenotype to balance metastasis and proliferation in patients with SMAD4-negative pancreatic cancer. *Gut* **2020**, *69*, 888–900, doi:10.1136/gutjnl-2018-317163.
78. Hao, J.; Zhang, S.; Zhou, Y.; Liu, C.; Hu, X.; Shao, C. MicroRNA 421 suppresses DPC4/Smad4 in pancreatic cancer. *Biochem. Biophys. Res. Commun.* **2011**, *406*, 552–557, doi:10.1016/j.bbrc.2011.02.086.
79. Hao, J.; Zhang, S.; Zhou, Y.; Hu, X.; Shao, C. MicroRNA 483-3p suppresses the expression of DPC4/Smad4 in pancreatic cancer. *FEBS Lett.* **2011**, *585*, 207–213, doi:10.1016/j.febslet.2010.11.039.
80. Xia, X.; Zhang, K.; Cen, G.; Jiang, T.; Cao, J.; Huang, K.; Huang, C.; Zhao, Q.; Qiu, Z. MicroRNA-301a-3p promotes pancreatic cancer progression via negative regulation of *SMAD4*. *Oncotarget* **2015**, *6*, doi:10.18632/oncotarget.4124.

© 2020 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/).

International Journal of
Molecular Sciences

MDPI

Review

Emerging Roles of DDB2 in Cancer

Pauline Gilson [1], Guillaume Drouot [2], Andréa Witz [1], Jean-Louis Merlin [1], Philippe Becuwe [2] and Alexandre Harlé [1,*]

[1] Institut de Cancérologie de Lorraine, Service de Biopathologie, Université de Lorraine, CNRS UMR 7039 CRAN, 54519 Vandœuvre-lès-Nancy CEDEX, France; p.gilson@nancy.unicancer.fr (P.G.); a.witz@nancy.unicancer.fr (A.W.); jl.merlin@nancy.unicancer.fr (J.-L.M.)

[2] Faculté des Sciences et Technologies, Université de Lorraine, CNRS UMR 7039 CRAN, 54506 Vandœuvre-lès-Nancy CEDEX, France; guillaume.drouot@orange.fr (G.D.); philippe.becuwe@univ-lorraine.fr (P.B.)

* Correspondence: a.harle@nancy.unicancer.fr; Tel.: +33-383656119

Received: 23 September 2019; Accepted: 15 October 2019; Published: 18 October 2019

Abstract: Damage-specific DNA-binding protein 2 (DDB2) was originally identified as a DNA damage recognition factor that facilitates global genomic nucleotide excision repair (GG-NER) in human cells. DDB2 also contributes to other essential biological processes such as chromatin remodeling, gene transcription, cell cycle regulation, and protein decay. Recently, the potential of DDB2 in the development and progression of various cancers has been described. DDB2 activity occurs at several stages of carcinogenesis including cancer cell proliferation, survival, epithelial to mesenchymal transition, migration and invasion, angiogenesis, and cancer stem cell formation. In this review, we focus on the current state of scientific knowledge regarding DDB2 biological effects in tumor development and the underlying molecular mechanisms. We also provide insights into the clinical consequences of DDB2 activity in cancers.

Keywords: DDB2; DNA repair; cancers; proliferation; migration; invasion

1. Introduction

DDB2 (damage-specific DNA-binding protein 2, also known as p48 subunit) is a 48-kDa protein exclusively localized in the nucleus of mammalian cells. DDB2 is ubiquitously present in human tissues, albeit differentially expressed. High DDB2 expression level has been described in liver, thymus, kidney, and testes, and a low expression level has been described in brain, lung, skin, heart, and muscles [1]. DDB2 is composed of seven WD40 repeat domains and a *N*-terminal helix–loop–helix motif [2]. The WD40 domains play a role as a module for sequence-specific protein–DNA or protein–protein interactions. Both domains appear critical for the biological functions of DDB2.

Numerous overlapping mechanisms have been found to be involved in the regulation of DDB2 expression. DDB2 basal expression is cell cycle-dependent in normal dividing cells with a DDB2 level that gradually increases in the mid-G1 phase and reaches a maximum at the G1/S boundary before dropping in the S-phase [3]. DDB2 transcription was shown to be transiently increased after UV-induced DNA damage in a p53-dependent manner [1]. The p53 protein cooperates with BRCA1 (Breast Cancer Associated protein 1) for its binding to the *ddb2* promoter and inducing *ddb2* transactivation [1] (Figure 1). The TAp63γ (Tumor protein 63 isoform gamma) isoform of the p63 protein that belongs to the p53 family and shares strong structural similarity with p53 is also able to activate DDB2 expression through recognition of the same region upstream of the transcription initiation site [4]. Structural analysis of the *ddb2* promoter shows multiple Sp1 (Transcription factor Sp1) -specific binding sites as usually found in gene promoters with a G–C rich sequence lacking a TATA box and suggesting a critical role for Sp1 for the basal expression of DDB2 [5] (Figure 1). NF1

(Neurofibromin 1) and E2F (Transcription factor E2F) elements are also identified, although probably having a smaller impact on DDB2 regulation [5]. Moreover, DDB2 activity is finely adjusted through post-transcriptional and post-translational mechanisms. An IRES (internal ribosome entry site) element located at the 5′ end of DDB2 mRNA stimulates the translational process of DDB2 in stress conditions such as serum starvation or exposure to doxorubicin [6]. Moreover, the 3′ untranslated region (3′UTR) of DDB2 mRNA harbors an uracil-rich sequence enabling its prompt export to the cytoplasm and thus DDB2 translational upregulation. During the DNA repair process, the Cul4A (Cullin4A) protein regulates DDB2 protein lifespan by means of its ubiquitin ligase functions [3]. The PARylation (Protein poly ADP-ribosylation) of the DDB2 protein in a PARP1-dependent (Poly [ADP-ribose] polymerase 1) manner stabilizes DDB2 and delays its degradation [7]. A PIP (PCNA-interacting protein) box sequence located in the *N*-terminal region of DDB2 enables the interaction between DDB2 and PCNA (proliferating cell nuclear antigen) for DDB2 proteolytic degradation even in the absence of DNA damage [8,9].

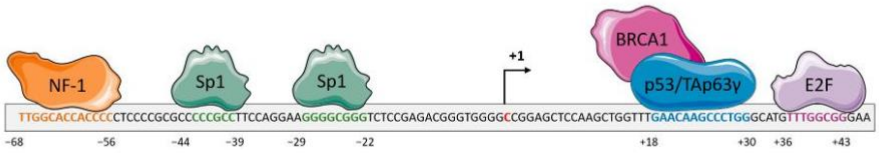


Figure 1. Schematic representation of the regulation of the gene encoding the DDB2 (Damage-specific DNA-binding protein 2) protein. The proximal promoter of *ddb2* gene harbors response elements for the transcription factors NF-1 (Neurofibromin 1) (orange) and Sp1 (Transcription factor Sp1) (green) upstream of the transcription initiation site. The proximal promoter also contains a response element for p53, in association with BRCA1 (Breast cancer type 1 susceptibility protein), or Tap63γ (Tumor protein 63 isoform gamma) (blue) proteins and the E2F (Transcription factor E2F) transcription factor downstream of the transcription initiation site. The binding of these proteins leads to the regulation of *ddb2* gene expression.

DDB2 was originally identified as a component of the UV-DDB human damage-specific DNA-binding heterodimeric complex along with DDB1 (damage-specific DNA-binding protein 1, also named p127 subunit) [2]. The X-ray structure of the UV-DDB complex showed that the interaction between DDB1 and DDB2 is mediated by the contact between the helix–loop–helix domain of DDB2 and two short repetitive β-propeller domains (WD40) of DDB1 [2]. UV-DDB forms a larger complex through the association of the DDB1 adaptor with the CRL ((CUL4A)-Regulator of Cullins-1 (ROC1 or RBX1) E3 ubiquitin ligase) complex [10]. In this structure, DDB1 serves as an adaptor molecule while DDB2 functions as a substrate receptor module that determines the specificity of targeted substrate [10]. DDB2 locally recognizes DNA damage sites and initiates the global genome nucleotide excision repair process (GG-NER) [2]. DDB2 specifically binds to the most frequent UV-damaged DNA lesions including (6-4) pyrimidine–pyrimidone photoproducts (6-4PPs) and cyclobutane pyrimidine dimers (CDPs) and appears essential for the repair of this latter [11]. The WD40 motifs of DDB2 facilitate the access of XPC (Xeroderma Pigmentosum group C) and other repair proteins to DNA lesions through histone modifications and chromatin decondensation that weaken DNA-histones interactions [10,12,13]. DDB2 fulfils this function through the recruitment of poly(ADP-ribose) polymerases (PARP) that append several units of ADP-ribose on histones [7]. As a part of the NER system, DDB2 also contribute to the repair of bulky DNA adducts other than UV-induced ones. For example, DDB2 was shown to recognize many forms of DNA damages, including those induced by platin-derived products, nitrogen mustard, psoralen, abasic sites, as well as single-stranded DNA [14]. However, less is known about the specific role of DDB2 in the repair of these DNA damages.

Beyond its most-studied role in DNA repair pathways, DDB2 appears as a multifunctional protein that participates in other essential biological processes such as gene transcription, cell cycle

progression, and protein degradation. Since DDB2 exhibits chromatin-remodeling functions, DDB2 has come to be recognized as a transcription regulator for a wide range of target genes. DDB2 especially participates in the activation of SWI/SNF (SWItch/Sucrose Non Fermenting) remodeling complexes [15] as well as STAGA (SPT3-TAFII31-GCN5L acetylase) and CBP (CREB (cAMP-response-element-binding protein)-binding protein)/p300histone acetyltransferase complexes [16] that contribute to chromatin unfolding and facilitate the recruitment of RNA polymerase to specific regions. DDB2 also stimulates the transcriptional activity of the transcription factor E2F1 and thus the expression of multiple target genes involved in cell cycle progression [3]. As a recognized member of the DDB1 and CUL4-associated factors (DCAF), DDB2 is in a close contact with the E3 ubiquitin ligase of the Cul4A protein and hence participates in the degradation of various proteins such as the cyclin-dependent kinase (CDK) inhibitor p27, a well-known negative regulator of the cell cycle [17]. DDB2 also contributes to the decay of p53 protein to maintain low levels of the CDK inhibitor $p21^{Waf1/Cip1}$ [18].

Several studies reported DDB2 as playing a novel function in the development and progression of various cancers. Herein, we focus on the current state of scientific knowledge regarding the biological effects of DDB2 in tumor development and the underlying molecular mechanisms. We also provide insights into the clinical consequences of DDB2 activity in cancers.

2. DDB2: A New Potent Tumor Suppressor?

DDB2 was first considered as a novel tumor suppressor based on the findings that mutations in the ddb2 gene result in an impairment in DDB2–DNA or DDB2–DDB1 interactions and subsequent NER activity defects [2]. Such deficiencies are observed in a subset of Xeroderma pigmentosum hereditary disease (Xeroderoma pigmentosum group E, XP-E) that displays an extreme sensitivity to UV radiation and a high predisposition to skin cancers [7]. Moreover, DDB2-deficient mice are prone to developing a wide panel of tumors, even in the absence of UV exposition [19]. Several studies highlighted an altered DDB2 expression, compared to non-malignant tissues, in many types of cancers [20], including prostate [21], colorectal [22,23], skin [24], head and neck [25], and ovarian [26] cancers. Furthermore, a correlation between low DDB2 expression level and poor outcomes was established among patients with melanoma, colon, ovarian, lung, breast, brain or head, and neck cancers, suggesting a critical role for DDB2 in tumor suppression [22,25–27].

3. DDB2 Has a Dual Activity on Cancer Cell Proliferation

High levels of DDB2 protein and mRNA are reported in ER (Estrogen receptor)-positive and non-invasive breast cancer models compared to ER-negative aggressive breast cancer cells and mammary non-malignant cells [16]. Enhanced DDB2 expression in DDB2-low level models upregulates in vitro cancer cell growth rate and clonogenicity. Such effects are abrogated by DDB2 knockdown in DDB2-overexpressed breast cancer models, suggesting the oncogenic role of DDB2 in mammary cancer cell growth [16]. DDB2 facilitates cell cycle progression, especially the entry in the S-phase, through the binding of the C-terminal domain and the co-activation of the transcription factor E2F1 [3]. By means of E2F1 transcriptional abilities, DDB2 indirectly regulates the expression of key genes involved in DNA replication and G1/S transition. The stimulating effect of DDB2 on cancer cell growth also involves the downregulation of manganese superoxide dismutase (MnSOD) [16]. MnSOD is a mitochondrial enzyme that detoxifies reactive oxygen species (ROS) to protect cells from oxidative damage. DDB2 interacts with the proximal *SOD2* (Superoxide dismutase 2) promoter, resulting in the loss of H3 histone acetylation, and in the recruitment of the AP-2 transcription factor, which is well known in the repression of the *SOD2* gene and downregulation of the encoded MnSOD [28]. By this mechanism, DDB2 attenuates the elimination of ROS that are known to activate several signaling pathways involved in breast cancer cell growth [28].

In contrast, DDB2 has shown antiproliferative properties in ovarian and prostate cancers in vitro. In an ovarian cancer cell model, DDB2 is reported to negatively regulate NEDD4L (Neural precursor cell expressed developmentally downregulated gene 4-like) cellular levels by inducing histone H3

trimethylation at the *NEDD4L* promoter region [29]. This limits the NEDD4L-dependent proteolytic degradation of the effector proteins SMAD2 (Mothers against decapentaplegic homolog) and SMAD3 and enhances the TGF-β (Transforming growth factor beta) signal transduction downstream, finally contributing to the inhibition of cancer cell proliferation [29].

NRIP/DCAF6 (Nuclear receptor-interacting protein 1/DDB1- and CUL4-associated factor 6) and DDB2 proteins are both members of the DDB1 and CUL4-associated factors DCAF family and androgen receptor (AR)-interacting proteins that physiologically compete to maintain AR expression level. DDB2 mediates the contact between AR and CUL4A–DDB1 E3 ligase complex for AR ubiquitination and proteasomal degradation in a p53-independent manner [30], while NRIP displaces DDB2 from the complex and stabilizes AR [21]. As the DDB2 expression level is found to be lower in prostate cancer tissues compared to non-neoplastic ones, this could interfere with AR homeostasis and induce subsequent AR-dependent prostate cancer growth.

Deregulation in the Wnt signaling pathway is usually co-opted during colon cancer development and is sought to be a driver event in this process. Negative regulatory effects of DDB2 in the Wnt signaling are mediated by the recruitment of β-catenin and the H3K27 methylase EZH2 (Enhancer of zeste homolog 2) to *Rnf43* (Ring Finger protein 43) promoter. This facilitates the activation of the RNF43 enzyme that eliminates Wnt receptors at the cell surface and downregulates Wnt signaling in colorectal cancer cells [31].

4. DDB2 Confers Resistance to Radiation and PARP Inhibitors and Sensitizes Cancer Cells to Chemotherapy-Induced Apoptosis

DDB2 participates in cellular responses to radiation-induced DNA damages [32]. Upon ionizing radiations, the DDB2 level transiently increases and facilitates homologous recombination repair of DNA double-strand breaks and prevents apoptosis. This is mediated by the phosphorylation of the G2-arrest-mediating factor Chk1 and confers radioresistance in the NSCLC (Non-small-cell lung carcinoma) cell model. A recent study shows that DDB2 also protects triple negative breast cancer cells from PARP inhibition and apoptosis through the regulation of DNA double-strand break repair by homologous recombination pathway [33]. DDB2 deficiency could thus sensitize TNBC cells to PARP inhibitors.

Given the role of DDB2 in DNA repair pathways, it could be argued that tumors with DDB2 deficiency should be more sensitive to DNA damaging treatments, such as platin products, as they lack some repair. However, DDB2 activity seems not needed to repair cisplatin-induced DNA crosslinks [34]. Besides, DDB2 participates in sensitizing ovarian cancer cells to cisplatin-mediated apoptosis through the downregulation of the *bcl-2* (B-cell lymphoma 2) transcriptional machinery in a HDAC1-dependent (Histone deacetylase 1) manner and the ubiquitylation of the antiapoptotic protein bcl-2 [34].

In the context of NER, DDB2 has a significant role in apoptotic processes through the regulation of p53 activity [35]. Upon exposure to DNA-damaging agents, DDB2-deficient cells fail to undergo apoptosis due to the implication of DDB2 as a member of the UV-DDB complex in the degradation of the CDK (cyclin-dependent kinase) inhibitor $p21^{Waf1/Cip1}$ [20]. In case of DDB2 depletion, $p21^{Waf1/Cip1}$ accumulates in cells associated with a blockade of apoptosis and cell cycle arrest. Cisplatin-resistant Hela cells are found to survive UV-induced apoptosis due to the accumulation of DDB2 associated with the upregulation of the antiapoptotic protein c-Flip (Cellular FLICE-like inhibitory protein) [36].

5. DDB2 Influences Epithelial to Mesenchymal Transition (EMT) and Cancer Cell Migration and Invasion

Several studies reported DDB2 as a negative regulatory factor of cancer cell migration and invasion. This notion is supported by the findings that DDB2 associates with the *NFKBIA* (NFKB Inhibitor Alpha) promoter region and enhances the expression of the encoded protein IκBα (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha), a cytoplasmic inhibitor of NF-κB (nuclear factor κB) [28]. Thus, DDB2 may reduce the NF-κB activity and its associated cancer-invasion effects in

breast cancers. DDB2 also modulates nanomechanical properties and stiffness of the mammary cancer cells associated with changes in the cortical actin–cytoskeleton organization and a loss of adhesion capacity [37].

DDB2 also represses head and neck squamous cell carcinoma and colon cancer progression by recruiting the histone H3K9 methylase suv39h (suppressor of variegation 3-9 homolog 1) to limit the transcriptional activity of Snail and ZEB1 (Zinc finger E-box Binding homeobox 1) factors and their epithelial-to-mesenchymal functions [25]. Upregulation of DDB2 levels induces the conversion of colon cancer cells with mesenchymal-type to epithelial-type. Conversely, the depletion of DDB2 in epithelial-type cells induces a morphological change in these cells with the appearance of mesenchymal characteristics. In the same way, DDB2 restricts the transcription of the pro-angiogenic growth factor VEGF (vascular endothelial growth factor) [25] and thus contributes to the inhibition of cancer cell invasion and metastasis formation.

However, another recent study suggests that DDB2 exerts migration and invasion stimulatory functions in gastric cancers. It reveals that DDB2 promotes the ubiquitination and degradation of a newly identified tumor suppressor PAQR3 (progestin and adipoQ receptor family member III) that limits cancer development and progression, notably inhibiting cancer cell migration, invasion, and angiogenesis [38].

6. DDB2 Affects Cancer Stem Cell Populations

Some studies reveal the protective role of DDB2 for ovarian cancer progression and recurrence by limiting the cancer stem cell (CSC) population. High levels of DDB2 are shown to halt the formation of ovarian tumor xenografts in vivo [26]. DDB2 exerts this regulatory function through the increase of IκBα protein levels, which results in the repression of the NF-κB signaling pathway and impinges the capacity of CSC to self-renew [26]. DDB2 also represses ovarian cancer cell dedifferentiation by inhibiting the transcription of aldehyde dehydrogenase 1 family member A1 (ALDH1A1) [39]. The impact of DDB2 on cancer stem cell populations is still under investigation in other cancer models and need to be explored.

7. Concluding Remarks and Perspectives

DDB2 was originally identified as a sensor of DNA damage that plays a critical role in DNA repair pathways. Sequence-specific interactions of DDB2 with DNAs or proteins make it a multifunctional protein with various functions in chromatin remodeling, gene transcription, cell cycle progression, and protein decay. More recently, the emerging role of DDB2 in the regulation of carcinogenesis has been established. DDB2 was first considered as a new tumor suppressor based on the finding that DDB2 deficiency leads to tumor formation. However, growing evidence about DDB2 activity suggests a regulatory mode more complex than envisioned. Actually, DDB2 seems to have a dual function in cancers, sometimes having tumor-suppressive properties and sometimes functioning as an oncogene. DDB2 exerts multilevel regulation of cancer development and progression, modulating cancer cell growth, migration and invasion, angiogenesis, and CSC formation (Figure 2)

These data suggest that DDB2 may have a potent role as a prognosis and predictive biomarker in cancers. However, they emphasize the need to improve our knowledge on DDB2 functions in the different cancer types, owing to the fact that DDB2 behaves differently according to cancer localization (Figure 3). Considering the importance of DDB2 in cancers, it becomes crucial to better understand its regulation and expression. The potent role of DDB2 as a direct or indirect drug target for the management of patients with cancer needs to be investigated.

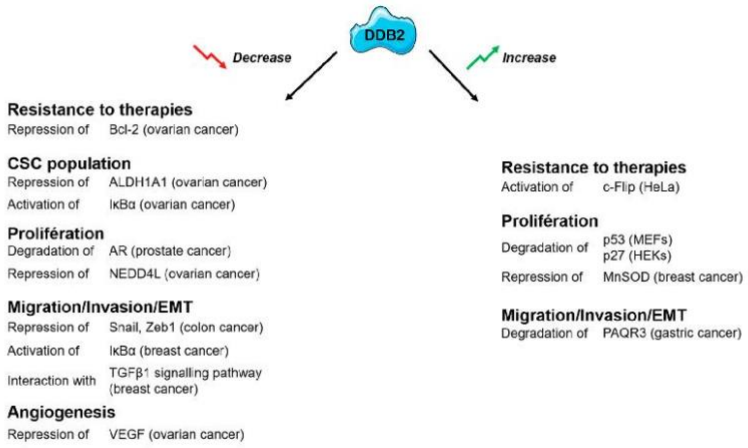


Figure 2. Roles of the DDB2 (Damage-specific DNA-binding protein 2) protein and its identified partners. AR (Androgen Receptor), ALDH1A1 (Aldehyde dehydrogenase 1 family, member A1), Bcl-2 (B-cell lymphoma 2), c-Flip (Cellular FLICE-like inhibitory protein, EMT (Epithelial–mesenchymal transition), HEKs (Human epidermal keratinocytes), IκBα (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha), MEFs (Mouse embryonic fibroblasts), MnSOD (Manganese superoxide dismutase), NEDD4L (Neural precursor cell expressed developmentally downregulated gene 4-like), PAQR3 (Progestin and adipoQ receptor family member 3), TGF-β1 (Transforming growth factor beta 1), VEGF (vascular endothelial growth factor), Zeb1 (Zinc finger E-box-binding homeobox 1).

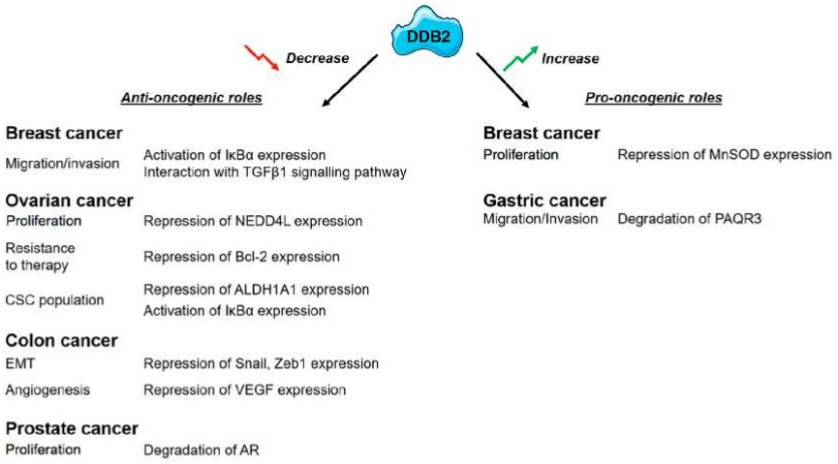


Figure 3. Identified impacts of the DDB2 protein activity in different solid tumors.

Author Contributions: Writing—original draft preparation, P.G. and G.D.; writing—review and editing, P.G., A.H., A.W., and P.B.; supervision, P.B. and A.H.; project administration, J.-L.M. and A.H. All authors read and approved the final manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

6-4PPs	6-4 pyrimidine-pyrimidone photoproducts
ALDH1A1	Aldehyde Dehydrogenase 1 family member A1
AR	Androgen receptor
Bcl-2	B-cell lymphoma 2
BRCA1	Breast Cancer Associated protein 1
c-Flip	Cellular FLICE-like inhibitory protein
CBP	CREB-binding protein
CDK	Cyclin-dependent kinase
CDPs	Cyclobutane pyrimidine dimers
CREB	cAMP-response-element-binding protein
CRL	Cullin4A-Regulator of Cullins-1 E3 ubiquitin ligase
CSC	Cancer stem cells
Cul4A	Cullin4A
DCAF	DDB1 and CUL4-associated factors
DDB1	Damage-specific DNA-binding protein 1
DDB2	Damage-specific DNA-binding protein 2
E2F	Transcription factor E2F
EMT	Epithelial to mesenchymal transition
ER	Estrogen receptor
EZH2	Enhancer of zeste homolog 2
GG-NER	Global genomic nucleotide excision repair
HDAC1	Histone deacetylase 1
HEKs	Human epidermal keratinocytes
IκBα	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
IRES	Internal ribosome entry site
MEFs	Mouse embryonic fibroblasts
MnSOD	Manganese superoxide dismutase
NEDD4L	Neural precursor cell expressed developmentally downregulated gene 4-like
NF-κB	Nuclear factor κB
NFKBIA	NFKB Inhibitor Alpha
NF1	Neurofibromin 1
NRIP	Nuclear receptor-interacting protein 1
NSCLC	Non-small-cell lung carcinoma
PAQR3	Progestin and adipoQ receptor family member III
PARP	Poly(ADP-ribose) polymerases
PARP1	Poly [ADP-ribose] polymerase 1
PARylation	Protein poly ADP-ribosylation
PCNA	Proliferating cell nuclear antigen
PIP	PCNA-interacting protein
Rnf43	Ring Finger protein 43
ROC1	Regulator of Cullins-1
ROS	Reactive oxygen species
SMAD	Mothers against decapentaplegic homolog
SOD2	Superoxide dismutase 2
Sp1	Transcription factor Sp1
STAGA	SPT3-TAFII31-GCN5L acetylase
Suv39h	Suppressor of variegation 3-9 homolog 1
SWI/SNF	SWItch/Sucrose Non Fermenting
TAp63γ	Tumor protein 63 isoform gamma
TGF-β	Transforming growth factor beta
UV-DDB	UV-damaged DNA-binding protein
VEGF	Vascular Endothelial Growth Factor

XPC	Xeroderma pigmentosum group C
XP-E	Xeroderma pigmentosum group E
ZEB1	Zinc finger E-box Binding homeobox 1

References

1. Tan, T.; Chu, G. p53 Binds and Activates the Xeroderma Pigmentosum DDB2 Gene in Humans but Not Mice. *Mol. Cell. Biol.* **2002**, *22*, 3247–3254. [CrossRef]
2. Scrima, A.; Konícková, R.; Czyzewski, B.K.; Kawasaki, Y.; Jeffrey, P.D.; Groisman, R.; Nakatani, Y.; Iwai, S.; Pavletich, N.P.; Thomä, N.H. Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* **2008**, *135*, 1213–1223. [CrossRef]
3. Nag, A.; Bondar, T.; Shiv, S.; Raychaudhuri, P. The xeroderma pigmentosum group E gene product DDB2 is a specific target of cullin 4A in mammalian cells. *Mol. Cell. Biol.* **2001**, *21*, 6738–6747. [CrossRef] [PubMed]
4. Liu, J.; Lin, M.; Zhang, C.; Wang, D.; Feng, Z.; Hu, W. TAp63γ enhances nucleotide excision repair through transcriptional regulation of DNA repair genes. *DNA Repair* **2012**, *11*, 167–176. [CrossRef] [PubMed]
5. Nichols, A.F.; Itoh, T.; Zolezzi, F.; Hutsell, S.; Linn, S. Basal transcriptional regulation of human damage-specific DNA-binding protein genes DDB1 and DDB2 by Sp1, E2F, N-myc and NF1 elements. *Nucleic Acids Res.* **2003**, *31*, 562–569. [CrossRef] [PubMed]
6. Dai, W.; Ma, W.; Li, Q.; Tao, Y.; Ding, P.; Zhu, R.; Jin, J. The 5′-UTR of DDB2 harbors an IRES element and upregulates translation during stress conditions. *Gene* **2015**, *573*, 57–63. [CrossRef] [PubMed]
7. Luijsterburg, M.S.; Lindh, M.; Acs, K.; Vrouwe, M.G.; Pines, A.; van Attikum, H.; Mullenders, L.H.; Dantuma, N.P. DDB2 promotes chromatin decondensation at UV-induced DNA damage. *J. Cell Biol.* **2012**, *197*, 267–281. [CrossRef] [PubMed]
8. Cazzalini, O.; Perucca, P.; Mocchi, R.; Sommatis, S.; Prosperi, E.; Stivala, L.A. DDB2 association with PCNA is required for its degradation after UV-induced DNA damage. *Cell Cycle Georget. Tex* **2014**, *13*, 240–248. [CrossRef]
9. Perucca, P.; Sommatis, S.; Mocchi, R.; Prosperi, E.; Stivala, L.A.; Cazzalini, O. A DDB2 mutant protein unable to interact with PCNA promotes cell cycle progression of human transformed embryonic kidney cells. *Cell Cycle* **2015**, *14*, 3920–3928. [CrossRef]
10. Marteijn, J.A.; Lans, H.; Vermeulen, W.; Hoeijmakers, J.H.J. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 465–481. [CrossRef]
11. Scrima, A.; Fischer, E.S.; Lingaraju, G.M.; Böhm, K.; Cavadini, S.; Thomä, N.H. Detecting UV-lesions in the genome: The modular CRL4 ubiquitin ligase does it best! *FEBS Lett.* **2011**, *585*, 2818–2825. [CrossRef] [PubMed]
12. Lan, L.; Nakajima, S.; Kapetanaki, M.G.; Hsieh, C.L.; Fagerburg, M.; Thickman, K.; Rodriguez-Collazo, P.; Leuba, S.H.; Levine, A.S.; Rapić-Otrin, V. Monoubiquitinated histone H2A destabilizes photolesion-containing nucleosomes with concomitant release of UV-damaged DNA-binding protein E3 ligase. *J. Biol. Chem.* **2012**, *287*, 12036–12049. [CrossRef] [PubMed]
13. Lans, H.; Marteijn, J.A.; Vermeulen, W. ATP-dependent chromatin remodeling in the DNA-damage response. *Epigenetics Chromatin* **2012**, *5*, 4. [CrossRef] [PubMed]
14. Tang, J.; Chu, G. Xeroderma pigmentosum complementation group E and UV-damaged DNA-binding protein. *DNA Repair* **2002**, *1*, 601–616. [CrossRef]
15. Zhao, Q.; Wang, Q.E.; Ray, A.; Wani, G.; Han, C.; Milum, K.; Wani, A.A. Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex. *J. Biol. Chem.* **2009**, *284*, 30424–30432. [CrossRef] [PubMed]
16. Kattan, Z.; Marchal, S.; Brunner, E.; Ramacci, C.; Leroux, A.; Merlin, J.L.; Domenjoud, L.; Dauça, M.; Becuwe, P. Damaged DNA binding protein 2 plays a role in breast cancer cell growth. *PLoS ONE* **2008**, *3*, e2002. [CrossRef]
17. Yan, Y.; Zhang, X.; Legerski, R.J. Artemis interacts with the Cul4A-DDB1DDB2 ubiquitin E3 ligase and regulates degradation of the CDK inhibitor p27. *Cell Cycle* **2011**, *10*, 4098–4109. [CrossRef]
18. Sharma, P.; Nag, A. CUL4A ubiquitin ligase: A promising drug target for cancer and other human diseases. *Open Biol.* **2014**, *4*, 130217. [CrossRef]

19. Yoon, T.; Chakrabortty, A.; Franks, R.; Valli, T.; Kiyokawa, H.; Raychaudhuri, P. Tumor-prone phenotype of the DDB2-deficient mice. *Oncogene* **2005**, *24*, 469–478. [CrossRef]
20. Bagchi, S.; Raychaudhuri, P. Damaged-DNA Binding Protein-2 Drives Apoptosis Following DNA Damage. *Cell Div.* **2010**, *5*, 3. [CrossRef]
21. Chen, H.-H.; Fan, P.; Chang, S.W.; Tsao, Y.P.; Huang, H.P.; Chen, S.L. NRIP/DCAF6 stabilizes the androgen receptor protein by displacing DDB2 from the CUL4A-DDB1 E3 ligase complex in prostate cancer. *Oncotarget* **2017**, *8*, 21501–21515. [CrossRef] [PubMed]
22. Yang, H.; Liu, J.; Jing, J.; Wang, Z.; Li, Y.; Gou, K.; Feng, X.; Yuan, Y.; Xing, C. Expression of DDB2 Protein in the Initiation, Progression, and Prognosis of Colorectal Cancer. *Dig. Dis. Sci.* **2018**, *63*, 2959–2968. [CrossRef] [PubMed]
23. Liu, J.; Li, H.; Sun, L.; Feng, X.; Wang, Z.; Yuan, Y.; Xing, C. The Differential Expression of Core Genes in Nucleotide Excision Repair Pathway Indicates Colorectal Carcinogenesis and Prognosis. *BioMed Res. Int.* **2018**, *2018*, 9651320. [CrossRef] [PubMed]
24. Stoyanova, T.; Roy, N.; Bhattacharjee, S.; Kopanja, D.; Valli, T.; Bagchi, S.; Raychaudhuri, P. p21 cooperates with DDB2 protein in suppression of ultraviolet ray-induced skin malignancies. *J. Biol. Chem.* **2012**, *287*, 3019–3028. [CrossRef] [PubMed]
25. Bommi, P.V.; Ravindran, S.; Raychaudhuri, P.; Bagchi, S. DDB2 regulates Epithelial-to-Mesenchymal Transition (EMT) in Oral/Head and Neck Squamous Cell Carcinoma. *Oncotarget* **2018**, *9*, 34708–34718. [CrossRef] [PubMed]
26. Han, C.; Zhao, R.; Liu, X.; Srivastava, A.; Gong, L.; Mao, H.; Qu, M.; Zhao, W.; Yu, J.; Wang, Q.E. DDB2 suppresses tumorigenicity by limiting the cancer stem cell population in ovarian cancer. *Mol. Cancer Res.* **2014**, *12*, 784–794. [CrossRef] [PubMed]
27. De Sousa, J.F.; Torrieri, R.; Serafim, R.B.; Di Cristofaro, L.F.; Escanfella, F.D.; Ribeiro, R.; Zanette, D.L.; Paçó-Larson, M.L.; da Silva, W.A., Jr.; Tirapelli, D.P.; et al. Expression signatures of DNA repair genes correlate with survival prognosis of astrocytoma patients. *Tumor Biol.* **2017**, *39*, 1010428317694552. [CrossRef] [PubMed]
28. Becuwe, P.; Ennen, M.; Klotz, R.; Barbieux, C.; Grandemange, S. Manganese superoxide dismutase in breast cancer: From molecular mechanisms of gene regulation to biological and clinical significance. *Free Radic. Biol. Med.* **2014**, *77*, 139–151. [CrossRef]
29. Zhao, R.; Cui, T.; Han, C.; Zhang, X.; He, J.; Srivastava, A.K.; Yu, J.; Wani, A.A.; Wang, Q.E. DDB2 modulates TGF-β signal transduction in human ovarian cancer cells by downregulating NEDD4L. *Nucleic Acids Res.* **2015**, *43*, 7838–7849. [CrossRef]
30. Chang, S.W.; Su, C.H.; Chen, H.H.; Huang, C.W.; Tsao, L.P.; Tsao, Y.; Chen, S.L. DDB2 is a novel AR interacting protein and mediates AR ubiquitination/degradation. *Int. J. Biochem. Cell Biol.* **2012**, *44*, 1952–1961. [CrossRef]
31. Huang, S.; Fantini, D.; Merrill, B.J.; Bagchi, S.; Guzman, G.; Raychaudhuri, P. DDB2 Is a Novel Regulator of Wnt Signaling in Colon Cancer. *Cancer Res.* **2017**, *77*, 6562–6575. [CrossRef] [PubMed]
32. Zou, N.; Xie, G.; Cui, T.; Srivastava, A.K.; Qu, M.; Yang, L.; Wei, S.; Zheng, Y.; Wang, Q.E. DDB2 increases radioresistance of NSCLC cells by enhancing DNA damage responses. *Tumour Biol.* **2016**, *37*, 14183–14191. [CrossRef] [PubMed]
33. Zhao, L.; Si, C.-S.; Yu, Y.; Lu, J.-W.; Zhuang, Y. Depletion of DNA damage binding protein 2 sensitizes triple-negative breast cancer cells to poly ADP-ribose polymerase inhibition by destabilizing Rad51. *Cancer Sci.* **2019**. [CrossRef] [PubMed]
34. Zhao, R.; Han, C.; Eisenhauer, E.; Kroger, J.; Zhao, W.; Yu, J.; Selvendiran, K.; Liu, X.; Wani, A.A.; Wang, Q.E. DNA Damage-Binding Complex Recruits HDAC1 to Repress Bcl-2 Transcription in Human Ovarian Cancer Cells. *Mol. Cancer Res.* **2014**, *12*, 370–380. [CrossRef]
35. Itoh, T.; O'Shea, C.; Linn, S. Impaired regulation of tumor suppressor p53 caused by mutations in the xeroderma pigmentosum DDB2 gene: Mutual regulatory interactions between p48(DDB2) and p53. *Mol. Cell. Biol.* **2003**, *23*, 7540–7553. [CrossRef]
36. Sun, N.K.; Sun, C.L.; Lin, C.H.; Pai, L.M.; Chao, C.C. Damaged DNA-binding protein 2 (DDB2) protects against UV irradiation in human cells and Drosophila. *J. Biomed. Sci.* **2010**, *17*, 27. [CrossRef]
37. Barbieux, C.; Bacharouche, J.; Soussen, C.; Hupont, S.; Razafitianamaharavo, A.; Klotz, R.; Pannequin, R.; Brie, D.; Bécuwe, P.; Francius, G.; et al. DDB2 (damaged-DNA binding 2) protein: A new modulator of nanomechanical properties and cell adhesion of breast cancer cells. *Nanoscale* **2016**, *8*, 5268–5279. [CrossRef]

38. Qiao, S.; Guo, W.; Liao, L.; Wang, L.; Wang, Z.; Zhang, R.; Xu, D.; Zhang, Y.; Pan, Y.; Wang, Z.; et al. DDB2 is involved in ubiquitination and degradation of PAQR3 and regulates tumorigenesis of gastric cancer cells. *Biochem. J.* **2015**, *469*, 469–480. [CrossRef]
39. Cui, T.; Srivastava, A.K.; Han, C.; Wu, D.; Wani, N.; Liu, L.; Gao, Z.; Qu, M.; Zou, N.; Zhang, X.; et al. DDB2 represses ovarian cancer cell dedifferentiation by suppressing ALDH1A1. *Cell Death Dis.* **2018**, *9*, 561. [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/).

9. Références bibliographiques

Abbosh, C., *et al.* (2018). "Early stage NSCLC - challenges to implementing ctDNA-based screening and MRD detection." Nat Rev Clin Oncol **15**(9): 577-586

Bailey, P., *et al.* (2016). "Genomic analyses identify molecular subtypes of pancreatic cancer." Nature **531**(7592): 47-52

Bielski, C. M., *et al.* (2018). "Genome doubling shapes the evolution and prognosis of advanced cancers." Nat Genet **50**(8): 1189-1195

Busca, S., *et al.* (2019). "Somatic substitution signature as an innovative tool in lung cancer diagnosis." Sci Rep **9**(1): 14561

Carbone, D. P., *et al.* (2017). "First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer." N Engl J Med **376**(25): 2415-2426

Chen, S., *et al.* (2016). "Beside P53 and PTEN: Identification of molecular alterations of the RAS/MAPK and PI3K/AKT signaling pathways in high-grade serous ovarian carcinomas to determine potential novel therapeutic targets." Oncol Lett **12**(5): 3264-3272

Conroy, T., *et al.* (2016). "Current standards and new innovative approaches for treatment of pancreatic cancer." Eur J Cancer **57**: 10-22

Conroy, T., *et al.* (2018). "FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer." N Engl J Med **379**(25): 2395-2406

Dardare, J., *et al.* (2020). "SMAD4 and the TGFβ Pathway in Patients with Pancreatic Ductal Adenocarcinoma." Int J Mol Sci **21**(10)

Franczak, C., *et al.* (2019). "Integrated routine workflow using next-generation sequencing and a fully-automated platform for the detection of KRAS, NRAS and BRAF mutations in formalin-fixed paraffin embedded samples with poor DNA quality in patients with colorectal carcinoma." PLoS One **14**(2): e0212801

Franczak, C., *et al.* (2018). "Clinical Interest of Circulating Tumor DNA in Oncology." Arch Med Res **49**(5): 297-305

Franczak, C*., et al.* (2019). "Uncommon mutational profiles of metastatic colorectal cancer detected during routine genotyping using next generation sequencing." Sci Rep **9**(1): 7083

Franczak, C*., et al.* (2017). "Comparison of Five Different Assays for the Detection of BRAF Mutations in Formalin-Fixed Paraffin Embedded Tissues of Patients with Metastatic Melanoma." Mol Diagn Ther **21**(2): 209-216

Franczak, C*., et al.* (2020). "Evaluation of KRAS, NRAS and BRAF mutations detection in plasma using an automated system for patients with metastatic colorectal cancer." PLoS One **15**(1): e0227294

Gilson, P*., et al.* (2019). "Emerging Roles of DDB2 in Cancer." Int J Mol Sci **20**(20)

Greaves, M. and C. C. Maley (2012). "Clonal evolution in cancer." Nature **481**(7381): 306-313

Harlé, A*., et al.* (2016). "Rare RAS Mutations in Metastatic Colorectal Cancer Detected During Routine RAS Genotyping Using Next Generation Sequencing." Target Oncol **11**(3): 363-370

Harlé, A*., et al.* (2019). "HPV insertional pattern as a personalized tumor marker for the optimized tumor diagnosis and follow-up of patients with HPV-associated carcinomas: a case report." BMC Cancer **19**(1): 277

Harlé, A*., et al.* (2018). "Evaluation and validation of HPV real-time PCR assay for the detection of HPV DNA in oral cytobrush and FFPE samples." Sci Rep **8**(1): 11313

Harlé, A*., et al.* (2016). "Detection of BRAF Mutations Using a Fully Automated Platform and Comparison with High Resolution Melting, Real-Time Allele Specific Amplification, Immunohistochemistry and Next Generation Sequencing Assays, for Patients with Metastatic Melanoma." PLoS One **11**(4): e0153576

Harlé, A*., et al.* (2015). "Expression of pEGFR and pAKT as response-predictive biomarkers for RAS wild-type patients to anti-EGFR monoclonal antibodies in metastatic colorectal cancers." Br J Cancer **113**(4): 680-685

Holmes, A*., et al.* (2016). "Mechanistic signatures of HPV insertions in cervical carcinomas." NPJ Genom Med **1**: 16004

Jamal-Hanjani, M*., et al.* (2014). "Tracking genomic cancer evolution for precision medicine: the lung TRACERx study." PLoS Biol **12**(7): e1001906

Jeannot, E.*, et al.* (2018). "Nuclear factor I X is a recurrent target for HPV16 insertions in anal carcinomas." Genes Chromosomes Cancer **57**(12): 638-644

Kattan, Z.*, et al.* (2008). "Damaged DNA binding protein 2 plays a role in breast cancer cell growth." PLoS One **3**(4): e2002

Lambert, A.*, et al.* (2019). "Comparison of Three Real-Time PCR Assays for the Detection of PIK3CA Somatic Mutations in Formalin-Fixed Paraffin Embedded Tissues of Patients with Breast Carcinomas." Pathol Oncol Res **25**(3): 1117-1123

Larkin, J.*, et al.* (2019). "Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma." N Engl J Med **381**(16): 1535-1546

Litchfield, K.*, et al.* (2020). "Representative Sequencing: Unbiased Sampling of Solid Tumor Tissue." Cell Rep **31**(5): 107550

López, S.*, et al.* (2020). "Interplay between whole-genome doubling and the accumulation of deleterious alterations in cancer evolution." Nat Genet **52**(3): 283-293

Luijsterburg, M. S.*, et al.* (2012). "DDB2 promotes chromatin decondensation at UV-induced DNA damage." J Cell Biol **197**(2): 267-281

Mandel, P. and P. Metais (1948). "[Not Available]." C R Seances Soc Biol Fil **142**(3-4): 241-243

Moore, G.*, et al.* (2020). "Top Notch Targeting Strategies in Cancer: A Detailed Overview of Recent Insights and Current Perspectives." Cells **9**(6)

Nag, A.*, et al.* (2001). "DDB2 induces nuclear accumulation of the hepatitis B virus X protein independently of binding to DDB1." J Virol **75**(21): 10383-10392

Rahib, L.*, et al.* (2014). "Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States." Cancer Res **74**(11): 2913-2921

Rosenthal, R.*, et al.* (2019). "Neoantigen-directed immune escape in lung cancer evolution." Nature **567**(7749): 479-485

Sastre-Garau, X. and A. Harlé (2020). "Pathology of HPV-Associated Head and Neck Carcinomas: Recent Data and Perspectives for the Development of Specific Tumor Markers." Frontiers in Oncology **10**(2505)

Scrima, A.*, et al.* (2008). "Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex." Cell **135**(7): 1213-1223

Serre, D.*, et al.* (2018). "Accelerated BRAF mutation analysis using a fully automated PCR platform improves the management of patients with metastatic melanoma." Oncotarget **9**(63): 32232-32237

Thomas, J.*, et al.* (2020). "Identification of Specific Tumor Markers in Vulvar Carcinoma Through Extensive Human Papillomavirus DNA Characterization Using Next Generation Sequencing Method." J Low Genit Tract Dis **24**(1): 53-60

Turajlic, S.*, et al.* (2017). "Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: a pan-cancer analysis." Lancet Oncol **18**(8): 1009-1021

Yan, Y.*, et al.* (2011). "Artemis interacts with the Cul4A-DDB1DDB2 ubiquitin E3 ligase and regulates degradation of the CDK inhibitor p27." Cell Cycle **10**(23): 4098-4109

Zhao, F.*, et al.* (2015). "Association of single nucleotide polymorphisms of DNA repair genes in NER pathway and susceptibility to pancreatic cancer." Int J Clin Exp Pathol **8**(9): 11579-11586

Zhao, Q.*, et al.* (2009). "Modulation of nucleotide excision repair by mammalian SWI/SNF chromatin-remodeling complex." J Biol Chem **284**(44): 30424-30432