HAL
open science

L'utilisation d'une plateforme de séquençage à haut débit dans l'étude des gènes BRCA1 et BRCA2

Agnès Collet

▶ **To cite this version:**

Agnès Collet. L'utilisation d'une plateforme de séquençage à haut débit dans l'étude des gènes BRCA1 et BRCA2. Sciences du Vivant [q-bio]. 2013. hal-01734309

HAL Id: hal-01734309

https://hal.univ-lorraine.fr/hal-01734309v1

Submitted on 14 Mar 2018

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.

UNIVERSITÉ DE LORRAINE

AVERTISSEMENT

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

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

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

Contact : ddoc-thesesexercice-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

UNIVERSITE DE LORRAINE FACULTE DE MEDECINE DE NANCY

2013

MEMOIRE

DU DIPLOME D'ETUDES SPECIALISEES (D.E.S.)

DE BIOLOGIE MEDICALE

Présenté et soutenu publiquement

par Agnès COLLET

le 15 octobre 2013

L'UTILISATION D'UNE PLATEFORME DE SEQUENCAGE A HAUT DEBIT DANS L'ETUDE DES GENES *BRCA1* ET *BRCA2*

Conformément aux dispositions du décret n°90-810 du 10 septembre 1990,

ce mémoire tient lieu de :

THESE

POUR LE DIPLOME D'ETAT DE DOCTEUR EN MEDECINE

Examinateurs de la thèse :

Président du jury de thèse : Monsieur le Professeur Philippe JONVEAUX

Juges : Monsieur le Professeur Bruno LEHEUP

Monsieur le Professeur Jean-Yves JOUZEAU

Monsieur le Docteur Claude HOUDAYER, Directeur de thèse

Madame le Docteur Elisabeth LUPORSI

UNIVERSITE DE LORRAINE FACULTE DE MEDECINE DE NANCY

2013

MEMOIRE

DU DIPLOME D'ETUDES SPECIALISEES (D.E.S.)

DE BIOLOGIE MEDICALE

Présenté et soutenu publiquement

par Agnès COLLET

le 15 octobre 2013

L'UTILISATION D'UNE PLATEFORME DE SEQUENCAGE A HAUT DEBIT DANS L'ETUDE DES GENES *BRCA1* ET *BRCA2*

Conformément aux dispositions du décret n°90-810 du 10 septembre 1990,

ce mémoire tient lieu de :

THESE

POUR LE DIPLOME D'ETAT DE DOCTEUR EN MEDECINE

Examinateurs de la thèse :

Président du jury de thèse : Monsieur le Professeur Philippe JONVEAUX

Juges : Monsieur le Professeur Bruno LEHEUP

Monsieur le Professeur Jean-Yves JOUZEAU

Monsieur le Docteur Claude HOUDAYER, Directeur de thèse

Madame le Docteur Elisabeth LUPORSI

Président de l'Université de Lorraine :
Professeur Pierre MUTZENHARDT

Doyen de la Faculté de Médecine :
Professeur Henry COUDANE

Vice-Doyen « Pédagogie »	:	**Mme la Professeure Karine ANGIOI**
Vice-Doyen Mission « Sillon lorrain »	:	**Mme la Professeure Annick BARBAUD**
Vice-Doyen Mission « Finances »	:	**Professeur Marc BRAUN**

Assesseurs :

- 1er Cycle :	**Professeur Bruno CHENUEL**
- 2ème Cycle :	**Professeur Marc DEBOUVERIE**
- 3ème Cycle : • « *DES Spécialités Médicales, Chirurgicales et Biologiques* »	**Professeur Jean-Pierre BRONOWICKI**
• « *DES Spécialité Médecine Générale* »	**Professeur Paolo DI PATRIZIO**
- Commission de Prospective Universitaire :	**Professeur Pierre-Edouard BOLLAERT**
- Développement Professionnel Continu :	**Professeur Jean-Dominique DE KORWIN**
- Filières professionnalisées :	**M. Walter BLONDEL**
- Formation Continue :	**Professeur Hervé VESPIGNANI**
- Recherche :	**Professeur Didier MAINARD**
- Relations Internationales :	**Professeur Jacques HUBERT**
- Universitarisation des études paramédicales et gestion des mono-appartenants :	**M. Christophe NEMOS**
- Vie Étudiante :	**Docteur Stéphane ZUILY**
- Vie Facultaire :	**Mme la Docteure Frédérique CLAUDOT**
- Étudiants :	**M. Xavier LEMARIE**

DOYENS HONORAIRES

Professeur Adrien DUPREZ - Professeur Jean-Bernard DUREUX - Professeur Jacques ROLAND - Professeur Patrick NETTER

==========

PROFESSEURS HONORAIRES

Jean-Marie ANDRE - Daniel ANTHOINE - Alain AUBREGE - Gérard BARROCHE - Alain BERTRAND - Pierre BEY
Marc-André BIGARD - Patrick BOISSEL – Pierre BORDIGONI - Jacques BORRELLY - Michel BOULANGE
Jean-Louis BOUTROY - Jean-Claude BURDIN - Claude BURLET - Daniel BURNEL - Claude CHARDOT - François CHERRIER Jean-Pierre CRANCE - Gérard DEBRY - Jean-Pierre DELAGOUTTE - Emile de LAVERGNE - Jean-Pierre DESCHAMPS
Jean DUHEILLE - Adrien DUPREZ - Jean-Bernard DUREUX - Gérard FIEVE - Jean FLOQUET - Robert FRISCH
Alain GAUCHER - Pierre GAUCHER - Hubert GERARD - Jean-Marie GILGENKRANTZ - Simone GILGENKRANTZ
Oliéro GUERCI - Pierre HARTEMANN - Claude HURIET - Christian JANOT - Michèle KESSLER - Jacques LACOSTE
Henri LAMBERT - Pierre LANDES - Marie-Claire LAXENAIRE - Michel LAXENAIRE - Jacques LECLERE - Pierre LEDERLIN Bernard LEGRAS - Jean-Pierre MALLIÉ - Michel MANCIAUX - Philippe MANGIN - Pierre MATHIEU - Michel MERLE
Denise MONERET-VAUTRIN - Pierre MONIN - Pierre NABET - Jean-Pierre NICOLAS - Pierre PAYSANT - Francis PENIN Gilbert PERCEBOIS - Claude PERRIN - Guy PETIET - Luc PICARD - Michel PIERSON - Jean-Marie POLU - Jacques POUREL Jean PREVOT - Francis RAPHAEL - Antoine RASPILLER – Denis REGENT - Michel RENARD - Jacques ROLAND
René-Jean ROYER - Daniel SCHMITT - Michel SCHMITT - Michel SCHWEITZER - Claude SIMON - Danièle SOMMELET
Jean-François STOLTZ - Michel STRICKER - Gilbert THIBAUT- Augusta TREHEUX - Hubert UFFHOLTZ - Gérard VAILLANT Paul VERT - Colette VIDAILHET - Michel VIDAILHET - Michel WAYOFF - Michel WEBER

==========

A. PROFESSEURS ÉMÉRITES

Professeur Daniel ANTHOINE - Professeur Gérard BARROCHE Professeur Pierre BEY - Professeur Patrick BOISSEL
Professeur Michel BOULANGE – Professeur Jean-Louis BOUTROY - Professeur Jean-Pierre CRANCE Professeur Jean-Pierre DELAGOUTTE - Professeur Jean-Marie GILGENKRANTZ - Professeure Simone GILGENKRANTZ Professeure Michèle KESSLER - Professeur Pierre MONIN - Professeur Jean-Pierre NICOLAS - Professeur Luc PICARD Professeur Michel PIERSON - Professeur Michel SCHMITT - Professeur Jean-François STOLTZ - Professeur Michel STRICKER Professeur Hubert UFFHOLTZ - Professeur Paul VERT - Professeure Colette VIDAILHET - Professeur Michel VIDAILHET Professeur Michel WAYOFF

==========

PROFESSEURS DES UNIVERSITÉS - PRATICIENS HOSPITALIERS

(Disciplines du Conseil National des Universités)

42ème Section : MORPHOLOGIE ET MORPHOGENÈSE

1ère sous-section : *(Anatomie)*
Professeur Gilles GROSDIDIER - Professeur Marc BRAUN
2ème sous-section : *(Cytologie et histologie)*
Professeur Bernard FOLIGUET – Professeur Christo CHRISTOV
3ème sous-section : *(Anatomie et cytologie pathologiques)*
Professeur François PLENAT – Professeur Jean-Michel VIGNAUD

43ème Section : BIOPHYSIQUE ET IMAGERIE MÉDECINE

1ère sous-section : *(Biophysique et médecine nucléaire)*
Professeur Gilles KARCHER – Professeur Pierre-Yves MARIE – Professeur Pierre OLIVIER
2ème sous-section : *(Radiologie et imagerie médecine)*
Professeur Michel CLAUDON – Professeure Valérie CROISÉ-LAURENT
Professeur Serge BRACARD – Professeur Alain BLUM – Professeur Jacques FELBLINGER - Professeur René ANXIONNAT

44ème Section : BIOCHIMIE, BIOLOGIE CELLULAIRE ET MOLÉCULAIRE, PHYSIOLOGIE ET NUTRITION

1ère sous-section : *(Biochimie et biologie moléculaire)*
Professeur Jean-Louis GUÉANT – Professeur Jean-Luc OLIVIER – Professeur Bernard NAMOUR
2ème sous-section : *(Physiologie)*
Professeur François MARCHAL – Professeur Bruno CHENUEL – Professeur Christian BEYAERT
3ème sous-section : *(Biologie Cellulaire)*
Professeur Ali DALLOUL
4ème sous-section : *(Nutrition)*
Professeur Olivier ZIEGLER – Professeur Didier QUILLIOT - Professeure Rosa-Maria RODRIGUEZ-GUEANT

45ème Section : MICROBIOLOGIE, MALADIES TRANSMISSIBLES ET HYGIÈNE

1ère sous-section : *(Bactériologie – virologie ; hygiène hospitalière)*
Professeur Alain LE FAOU - Professeur Alain LOZNIEWSKI – Professeure Evelyne SCHVOERER
2ème sous-section : *(Parasitologie et Mycologie)*
Professeure Marie MACHOUART
3ème sous-section : *(Maladies infectieuses ; maladies tropicales)*
Professeur Thierry MAY – Professeur Christian RABAUD

46ème Section : SANTÉ PUBLIQUE, ENVIRONNEMENT ET SOCIÉTÉ

1ère sous-section : *(Épidémiologie, économie de la santé et prévention)*
Professeur Philippe HARTEMANN – Professeur Serge BRIANÇON - Professeur Francis GUILLEMIN
Professeur Denis ZMIROU-NAVIER – Professeur François ALLA
2ème sous-section : *(Médecine et santé au travail)*
Professeur Christophe PARIS
3ème sous-section : *(Médecine légale et droit de la santé)*
Professeur Henry COUDANE
4ème sous-section : *(Biostatistiques, informatique médicale et technologies de communication)*
Professeur François KOHLER – Professeure Eliane ALBUISSON

47ème Section : CANCÉROLOGIE, GÉNÉTIQUE, HÉMATOLOGIE, IMMUNOLOGIE

1ère sous-section : *(Hématologie ; transfusion)*
Professeur Pierre FEUGIER
2ème sous-section : *(Cancérologie ; radiothérapie)*

Professeur François GUILLEMIN – Professeur Thierry CONROY - Professeur Didier PEIFFERT
Professeur Frédéric MARCHAL
3ème sous-section : *(Immunologie)*
Professeur Gilbert FAURE – Professeur Marcelo DE CARVALHO-BITTENCOURT
4ème sous-section : *(Génétique)*
Professeur Philippe JONVEAUX – Professeur Bruno LEHEUP

48ème Section : ANESTHÉSIOLOGIE, RÉANIMATION, MÉDECINE D'URGENCE, PHARMACOLOGIE ET THÉRAPEUTIQUE

1ère sous-section : *(Anesthésiologie - réanimation ; médecine d'urgence)*
Professeur Claude MEISTELMAN – Professeur Hervé BOUAZIZ - Professeur Gérard AUDIBERT
Professeur Thomas FUCHS-BUDER – Professeure Marie-Reine LOSSER
2ème sous-section : *(Réanimation ; médecine d'urgence)*
Professeur Alain GERARD - Professeur Pierre-Édouard BOLLAERT - Professeur Bruno LÉVY – Professeur Sébastien GIBOT
3ème sous-section : *(Pharmacologie fondamentale ; pharmacologie clinique ; addictologie)*
Professeur Patrick NETTER – Professeur Pierre GILLET
4ème sous-section : *(Thérapeutique ; médecine d'urgence ; addictologie)*
Professeur François PAILLE – Professeur Faiez ZANNAD - Professeur Patrick ROSSIGNOL

49ème Section : PATHOLOGIE NERVEUSE ET MUSCULAIRE, PATHOLOGIE MENTALE, HANDICAP ET RÉÉDUCATION

1ère sous-section : *(Neurologie)*
Professeur Hervé VESPIGNANI - Professeur Xavier DUCROCQ – Professeur Marc DEBOUVERIE
Professeur Luc TAILLANDIER - Professeur Louis MAILLARD
2ème sous-section : *(Neurochirurgie)*
Professeur Jean-Claude MARCHAL – Professeur Jean AUQUE – Professeur Olivier KLEIN
Professeur Thierry CIVIT - Professeure Sophie COLNAT-COULBOIS
3ème sous-section : *(Psychiatrie d'adultes ; addictologie)*
Professeur Jean-Pierre KAHN – Professeur Raymund SCHWAN
4ème sous-section : *(Pédopsychiatrie ; addictologie)*
Professeur Daniel SIBERTIN-BLANC – Professeur Bernard KABUTH
5ème sous-section : *(Médecine physique et de réadaptation)*
Professeur Jean PAYSANT

50ème Section : PATHOLOGIE OSTÉO-ARTICULAIRE, DERMATOLOGIE ET CHIRURGIE PLASTIQUE

1ère sous-section : *(Rhumatologie)*
Professeure Isabelle CHARY-VALCKENAERE – Professeur Damien LOEUILLE
2ème sous-section : *(Chirurgie orthopédique et traumatologique)*
Professeur Daniel MOLE - Professeur Didier MAINARD - Professeur François SIRVEAUX – Professeur Laurent GALOIS
3ème sous-section : *(Dermato-vénéréologie)*
Professeur Jean-Luc SCHMUTZ – Professeure Annick BARBAUD
4ème sous-section : *(Chirurgie plastique, reconstructrice et esthétique ; brûlologie)*
Professeur François DAP - Professeur Gilles DAUTEL - Professeur Etienne SIMON

51ème Section : PATHOLOGIE CARDIO-RESPIRATOIRE ET VASCULAIRE

1ère sous-section : *(Pneumologie ; addictologie)*
Professeur Yves MARTINET – Professeur Jean-François CHABOT – Professeur Ari CHAOUAT
2ème sous-section : *(Cardiologie)*
Professeur Etienne ALIOT – Professeur Yves JUILLIERE
Professeur Nicolas SADOUL - Professeur Christian de CHILLOU DE CHURET
3ème sous-section : *(Chirurgie thoracique et cardiovasculaire)*
Professeur Jean-Pierre VILLEMOT – Professeur Thierry FOLLIGUET
4ème sous-section : *(Chirurgie vasculaire ; médecine vasculaire)*
Professeur Denis WAHL – Professeur Sergueï MALIKOV

52ème Section : MALADIES DES APPAREILS DIGESTIF ET URINAIRE

1ère sous-section : *(Gastroentérologie ; hépatologie ; addictologie)*
Professeur Jean-Pierre BRONOWICKI – Professeur Laurent PEYRIN-BIROULET
3ème sous-section : *(Néphrologie)*
Professeure Dominique HESTIN – Professeur Luc FRIMAT
4ème sous-section : *(Urologie)*
Professeur Jacques HUBERT – Professeur Pascal ESCHWEGE

53ème Section : MÉDECINE INTERNE, GÉRIATRIE ET CHIRURGIE GÉNÉRALE
1ère sous-section : *(Médecine interne ; gériatrie et biologie du vieillissement ; médecine générale ; addictologie)*
Professeur Jean-Dominique DE KORWIN – Professeur Pierre KAMINSKY - Professeur Athanase BENETOS
Professeure Gisèle KANNY – Professeure Christine PERRET-GUILLAUME
2ème sous-section : *(Chirurgie générale)*
Professeur Laurent BRESLER - Professeur Laurent BRUNAUD – Professeur Ahmet AYAV

54ème Section : DÉVELOPPEMENT ET PATHOLOGIE DE L'ENFANT, GYNÉCOLOGIE-OBSTÉTRIQUE, ENDOCRINOLOGIE ET REPRODUCTION
1ère sous-section : *(Pédiatrie)*
Professeur Jean-Michel HASCOET - Professeur Pascal CHASTAGNER - Professeur François FEILLET
Professeur Cyril SCHWEITZER – Professeur Emmanuel RAFFO – Professeure Rachel VIEUX
2ème sous-section : *(Chirurgie infantile)*
Professeur Pierre JOURNEAU – Professeur Jean-Louis LEMELLE
3ème sous-section : *(Gynécologie-obstétrique ; gynécologie médicale)*
Professeur Philippe JUDLIN – Professeur Olivier MOREL
4ème sous-section : *(Endocrinologie, diabète et maladies métaboliques ; gynécologie médicale)*
Professeur Georges WERYHA – Professeur Marc KLEIN – Professeur Bruno GUERCI

55ème Section : PATHOLOGIE DE LA TÊTE ET DU COU
1ère sous-section : *(Oto-rhino-laryngologie)*
Professeur Roger JANKOWSKI – Professeure Cécile PARIETTI-WINKLER
2ème sous-section : *(Ophtalmologie)*
Professeur Jean-Luc GEORGE – Professeur Jean-Paul BERROD – Professeure Karine ANGIOI

3ème sous-section : *(Chirurgie maxillo-faciale et stomatologie)*
Professeur Jean-François CHASSAGNE – Professeure Muriel BRIX
==========

B. PROFESSEURS DES UNIVERSITÉS

61ème Section : GÉNIE INFORMATIQUE, AUTOMATIQUE ET TRAITEMENT DU SIGNAL
Professeur Walter BLONDEL

64ème Section : BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE
Professeure Sandrine BOSCHI-MULLER

==========

PROFESSEURS DES UNIVERSITÉS DE MÉDECINE GÉNÉRALE
Professeur Jean-Marc BOIVIN

PROFESSEUR ASSOCIÉ DE MÉDECINE GÉNÉRALE
Professeur associé Paolo DI PATRIZIO

==========

C. MAÎTRES DE CONFÉRENCES DES UNIVERSITÉS - PRATICIENS HOSPITALIERS

42ème Section : MORPHOLOGIE ET MORPHOGENÈSE
1ère sous-section : *(Anatomie)*
Docteur Bruno GRIGNON – Docteure Manuela PEREZ
2ème sous-section : *(Cytologie et histologie)*
Docteur Edouard BARRAT - Docteure Françoise TOUATI – Docteure Chantal KOHLER
3ème sous-section : *(Anatomie et cytologie pathologiques)*
Docteure Aude MARCHAL – Docteur Guillaume GAUCHOTTE

43ème Section : BIOPHYSIQUE ET IMAGERIE MÉDECINE
1ère sous-section : *(Biophysique et médecine nucléaire)*
Docteur Jean-Claude MAYER - Docteur Jean-Marie ESCANYE
2ème sous-section : (Radiologie et imagerie médecine)
Docteur Damien MANDRY

44ème Section : BIOCHIMIE, BIOLOGIE CELLULAIRE ET MOLÉCULAIRE, PHYSIOLOGIE ET NUTRITION
1ère sous-section : ***(Biochimie et biologie moléculaire)***
Docteure Sophie FREMONT - Docteure Isabelle GASTIN – Docteur Marc MERTEN
Docteure Catherine MALAPLATE-ARMAND - Docteure Shyue-Fang BATTAGLIA
2ème sous-section : ***(Physiologie)***
Docteur Mathias POUSSEL – Docteure Silvia VARECHOVA
3ème sous-section : ***(Biologie Cellulaire)***
Docteure Véronique DECOT-MAILLERET

45ème Section : MICROBIOLOGIE, MALADIES TRANSMISSIBLES ET HYGIÈNE
1ère sous-section : ***(Bactériologie – Virologie ; hygiène hospitalière)***
Docteure Véronique VENARD – Docteure Hélène JEULIN – Docteure Corentine ALAUZET
3ème sous-section : ***(Maladies Infectieuses ; Maladies Tropicales)***
Docteure Sandrine HENARD

46ème Section : SANTÉ PUBLIQUE, ENVIRONNEMENT ET SOCIÉTÉ
1ère sous-section : ***(Epidémiologie, économie de la santé et prévention)***
Docteur Alexis HAUTEMANIÈRE – Docteure Frédérique CLAUDOT – Docteur Cédric BAUMANN
2ème sous-section ***(Médecine et Santé au Travail)***
Docteure Isabelle THAON
3ème sous-section ***(Médecine légale et droit de la santé)***
Docteur Laurent MARTRILLE
4ère sous-section : ***(Biostatistiques, informatique médicale et technologies de communication***
Docteur Nicolas JAY

47ème Section : CANCÉROLOGIE, GÉNÉTIQUE, HÉMATOLOGIE, IMMUNOLOGIE
2ème sous-section : ***(Cancérologie ; radiothérapie : cancérologie (type mixte : biologique)***
Docteure Lina BOLOTINE
4ème sous-section : ***(Génétique)***
Docteur Christophe PHILIPPE – Docteure Céline BONNET

48ème Section : ANESTHÉSIOLOGIE, RÉANIMATION, MÉDECINE D'URGENCE, PHARMACOLOGIE ET THÉRAPEUTIQUE
3ème sous-section : ***(Pharmacologie fondamentale ; pharmacologie clinique)***
Docteure Françoise LAPICQUE – Docteur Nicolas GAMBIER – Docteur Julien SCALA-BERTOLA

50ème Section : PATHOLOGIE OSTÉO-ARTICULAIRE, DERMATOLOGIE ET CHIRURGIE PLASTIQUE
1ère sous-section : ***(Rhumatologie)***
Docteure Anne-Christine RAT
3ème sous-section : ***(Dermato-vénéréologie)***
Docteure Anne-Claire BURSZTEJN
4ème sous-section : ***(Chirurgie plastique, reconstructrice et esthétique ; brûlologie)***
Docteure Laetitia GOFFINET-PLEUTRET

51ème Section : PATHOLOGIE CARDIO-RESPIRATOIRE ET VASCULAIRE
3ème sous-section : ***(Chirurgie thoracique et cardio-vasculaire)***
Docteur Fabrice VANHUYSE
4ème sous-section : ***(Chirurgie vasculaire ; médecine vasculaire)***
Docteur Stéphane ZUILY

53ème Section : MÉDECINE INTERNE, GÉRIATRIE et CHIRURGIE GÉNÉRALE
1ère sous-section : ***(Médecine interne ; gériatrie et biologie du vieillissement ; médecine générale ; addictologie)***
Docteure Laure JOLY

54ème Section : DÉVELOPPEMENT ET PATHOLOGIE DE L'ENFANT, GYNÉCOLOGIE-OBSTÉTRIQUE,
ENDOCRINOLOGIE ET REPRODUCTION
5ème sous-section : ***(Biologie et médecine du développement et de la reproduction ; gynécologie médicale)***
Docteur Jean-Louis CORDONNIER

==========

MAÎTRE DE CONFÉRENCE DES UNIVERSITÉS DE MÉDECINE GÉNÉRALE
Docteure Elisabeth STEYER

==========

D. MAÎTRES DE CONFÉRENCES

5ème Section : SCIENCES ÉCONOMIQUES
Monsieur Vincent LHUILLIER

19ème Section : SOCIOLOGIE, DÉMOGRAPHIE
Madame Joëlle KIVITS

60ème Section : MÉCANIQUE, GÉNIE MÉCANIQUE, GÉNIE CIVIL
Monsieur Alain DURAND

61ème Section : GÉNIE INFORMATIQUE, AUTOMATIQUE ET TRAITEMENT DU SIGNAL
Monsieur Jean REBSTOCK

64ème Section : BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE
Madame Marie-Claire LANHERS – Monsieur Pascal REBOUL – Monsieur Nick RAMALANJAONA

65ème Section : BIOLOGIE CELLULAIRE
Monsieur Jean-Louis GELLY - Madame Ketsia HESS – Monsieur Hervé MEMBRE
Monsieur Christophe NEMOS - Madame Natalia DE ISLA - Madame Nathalie MERCIER – Madame Céline HUSELSTEIN

66ème Section : PHYSIOLOGIE
Monsieur Nguyen TRAN

==========

E. MAÎTRES DE CONFÉRENCES ASSOCIÉS

Médecine Générale
Docteure Sophie SIEGRIST - Docteur Arnaud MASSON - Docteur Pascal BOUCHE

==========

F. DOCTEURS HONORIS CAUSA

Professeur Charles A. BERRY (1982)
Centre de Médecine Préventive, Houston (U.S.A)
Professeur Pierre-Marie GALETTI (1982)
Brown University, Providence (U.S.A)
Professeure Mildred T. STAHLMAN (1982)
Vanderbilt University, Nashville (U.S.A)
Professeur Théodore H. SCHIEBLER (1989)
Institut d'Anatomie de Würtzburg (R.F.A)
Université de Pennsylvanie (U.S.A)
Professeur Mashaki KASHIWARA (1996)
Research Institute for Mathematical Sciences de Kyoto (JAPON)

Professeure Maria DELIVORIA-PAPADOPOULOS (1996)
Professeur Ralph GRÄSBECK (1996)
Université d'Helsinki (FINLANDE)
Professeur James STEICHEN (1997)
Université d'Indianapolis (U.S.A)
Professeur Duong Quang TRUNG (1997)
Université d'Hô Chi Minh-Ville (VIÊTNAM)
Professeur Daniel G. BICHET (2001)
Université de Montréal (Canada)
Professeur Marc LEVENSTON (2005)
Institute of Technology, Atlanta (USA)

Professeur Brian BURCHELL (2007)
Université de Dundee (Royaume-Uni)
Professeur Yunfeng ZHOU (2009)
Université de Wuhan (CHINE)
Professeur David ALPERS (2011)
Université de Washington (U.S.A)
Professeur Martin EXNER (2012)
Université de Bonn (ALLEMAGNE)

Remerciements

AU JURY

A Monsieur le Professeur Philippe Jonveaux, Président de Thèse

Professeur des Universités – Praticien Hospitalier de Génétique

Merci d'avoir accepté de juger ce travail.

Merci de m'avoir orientée vers ce sujet vaste et passionnant que représente le séquençage à haut débit.

Merci de m'avoir donné le courage de me lancer, il y a deux ans, vers la spécialisation en génétique. Je n'ai jamais regretté ce choix, bien au contraire.

Merci pour votre écoute et pour votre encadrement. Vos conseils dans l'organisation de ma formation m'auront toujours été précieux.

Que ce travail soit l'expression de ma sincère reconnaissance et de mon profond respect.

A Monsieur le Professeur Bruno Leheup, Juge

Professeur des Universités – Praticien Hospitalier de Génétique

Merci d'avoir accepté de faire partie de ce jury de thèse.

Je vous suis tout particulièrement reconnaissante d'être parvenu à vous rendre disponible pour pouvoir finalement y prendre part.

Merci de m'avoir acceptée dans votre service au semestre dernier et merci pour tout ce que vous avez pu m'enseigner d'un point de vue clinique.

Ces 6 mois auront été très enrichissants et c'est aujourd'hui un honneur pour moi de vous compter parmi les membres de ce jury.

A Monsieur le Professeur Jean-Yves Jouzeau, Juge

Professeur des Universités – Praticien Hospitalier de Bioanalyse du médicament

Merci d'avoir bien voulu « rendre service aux internes de biologie médicale » en acceptant de prendre part à la composition de ce jury de thèse.

C'est un honneur pour moi de pouvoir vous présenter mon travail.

A Monsieur le Docteur Claude Houdayer, Juge et Directeur de Thèse

Maitre de Conférence des Universités – Praticien Hospitalier de Génétique

Merci de m'avoir accueillie au sein de votre laboratoire de Génétique à l'Institut Curie de Paris.

Merci de m'avoir confié ce sujet de thèse qui me paraissait si vaste et complexe au départ et qui s'avère être si passionnant et complexe pour finir. J'espère sincèrement que ce travail sera l'occasion de pouvoir continuer à collaborer avec votre laboratoire à ce sujet.

Merci pour votre encadrement, votre écoute et votre patience. Ce fut un plaisir de pouvoir profiter de vos connaissances et de votre disponibilité.

Je garderai un excellent souvenir de ce semestre, aussi bien sur le plan professionnel qu'humain. Ce fut pour moi un réel bonheur de venir chaque jour prendre part à l'activité de votre équipe.

Je quitterai ce laboratoire avec regret.

A Madame le Docteur Elisabeth Luporsi, Juge

Spécialiste en Oncologie Médicale et Oncogénétique. Docteur en statistiques

Merci d'avoir immédiatement accepté avec autant d'enthousiasme de faire partie de mon jury de thèse.

J'espère que mon travail sera à la hauteur de vos attentes.

A MA FAMILLE

A ma maman, pour son amour, sa patience et son soutien indéfectibles…

Je ne serai jamais assez reconnaissante pour tout ce que tu as fait, et continues, de faire pour moi.

Après toutes ces années plus ou moins éprouvantes, c'est fini ☺ (sauf peut-être en ce qui concerne mes appels téléphoniques interminables).

Place désormais aux séances de rattrapage.

A mon compagnon Romain, avec qui j'ai traversé toutes ces années d'études… Merci d'être à mes côtés depuis tout ce temps.

Ca y est, on y est (presque, en ce qui te concerne). Il est désormais grand temps de vivre pour nous deux.

A ma mamie, notre Reine Mère, à qui je suis fière de pouvoir montrer où je suis arrivée aujourd'hui. Merci pour ton soutien et ton dévouement.

A ma marraine, ma deuxième mère. A Alfred.

Merci pour votre gentillesse et votre écoute. Merci d'être toujours présents à nos côtés.

A mon cousin Alexis, à ma cousine Héloïse. J'ai toujours plaisir à nous voir grandir ensemble.

A Aloïs, A Stéphanie

A Léo

Aux parents de Romain et à son frère

A L'EQUIPE DU LABORATOIRE DE GENETIQUE DE L'INSTITUT CURIE

Merci à tous pour ce semestre passé en votre compagnie et pour tout ce que vous m'avez appris.

Merci pour votre aide et votre soutien tout au long de mes manip'… (et parfois de mes déconvenues).

Merci pour votre bonne humeur et votre humour, ce fut un vrai plaisir de travailler dans ces conditions.

Merci tout particulièrement à Dominique Stoppa-Lyonnet et Claude Houdayer pour m'avoir aussi bien accueillie, à Etienne pour son écoute et ses suggestions.

A Lisa pour son aide, ses relectures, sa bonne humeur et son optimisme (ça va aller), à Virginie pour sa gentillesse et pour m'avoir tant appris, à Julien pour avoir pris le temps de loucher avec moi sur son écran de PC, à Anthony pour m'avoir écoutée me plaindre pendant 3 semaines durant ma quête du transcrit perdu, à Marine pour avoir initié l'excursion Nancéienne, à Henrique et Khadija pour m'avoir dévoilé les secrets du PGM, à Flo pour sa gentillesse, à Cédrick pour son aide dans mes balbutiements, à Catherine Dehainault (un Romain et une Héloïse… je compatis), à Florian, Gaétan, Catherine Dubois, Ambre, Sophie, Delphine (et notre combat d'anthologie) et à toutes les autres personnes du labo qui ont rendu ce stage si agréable.

Merci à Thomas et surtout à Quentin pour son aide précieuse durant cet été à la plateforme NGS.

Merci Coca Zéro, Nespresso et les chouquettes.

A L'EQUIPE DU LABORATOIRE DE GENETIQUE DE NANCY

Merci à toute l'équipe du laboratoire pour votre accueil il y a déjà 2 ans et pour tout ce que vous m'avez appris.

Merci à Monsieur Jonveaux, Mylène Valduga, Christophe Philippe et Céline Bonnet pour ce que vous m'avez transmis.

Merci à Lila. Je te dois énormément. Merci pour ton aide constante et ta gentillesse et ce, dès mon arrivée au laboratoire.

Merci à Laetitia pour ton dynamisme, ton humour et tout ce que tu as pu me transmettre.

A L'EQUIPE DE CONSULTATION DE L'HOPITAL D'ENFANTS DE BRABOIS

AUX PERSONNES QUE J'AI PU CROISER DURANT MON INTERNAT A NANCY

Et dont je garderai un agréable souvenir…

A Monsieur le Professeur Thomas Lecompte, merci d'avoir su me dire ce que j'avais besoin d'entendre dans mes débuts hésitants en biologie. Ce fut un plaisir de travailler à vos côtés et de profiter de votre savoir et de votre humour.

A Madame le Docteur Valérie Eschwege qui m'a transmis le gout de toujours chercher à comprendre. Merci pour ton accessibilité et ta gentillesse.

A Madame le Docteur Anne Kennel

Aux autres internes qui sont parfois venus se perdre en labo (et tant mieux) : Caroline, Mikael et Arthur

A MES AMIS

A ma Benj, avec qui j'aurais tant partagé, du bon comme du mauvais. Merci d'avoir toujours été là.

A Claire, à nos séances révisions/Kaamelott, à nos grandes discussions comme à nos fous rires. Merci aussi pour toutes ces années. (P.S : c'est à ton tour)

A Aurélie

A Thomas

A Cécilia, Céline, Mathieu, Marion, Anaïs et tout le reste des « anciens » (désolé) de Ludres

A Polo (Y-oo), Maf', Anne, Adrien, Vincent, Anne-Claire, Ludo, Julie

A mes amis de Forma-Sup : Le chef en tête, bien entendu.

A Audrey, ses parents, Christophe et, surtout, la jolie petite Emma.

A tous mes amis de l'Etrier de Lorraine : un merci tout particulier à Mam' Jourde, au soldat Cosson (courage) et à Marino pour leur aide.

A mes plus ou moins anciens co-internes, avec une pensée particulière pour Laure, Amélie, Caro, Lucille, Anthony, Anne-Lise et mon p'tit Paul (oui je te compte là-dedans !)

A Jean-Christophe, je m'excuse de n'avoir retenu aucune de vos propositions de sujet de thèse (pourtant toutes plus intéressantes les unes que les autres)

A Thomash'

A MES BESTIOLES

A ma Moun,' qui nous aura accompagnés, Romain et moi, durant (presque) toutes ces années. Ciao Patate. Tu nous manques.

A ma Suzzle

A ma Stitch, dont je rêvais depuis si longtemps.

SERMENT

« Au moment d'être admise à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité. Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux. Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité. J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences. Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admise dans l'intimité des personnes, je tairai les secrets qui me sont confiés. Reçue à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs. Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.

Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonorée et méprisée si j'y manque ».

« Dans la vie, rien n'est à craindre, tout est à comprendre. »

Marie Curie

« C'était compliqué, alors j'ai essayé d'expliquer ce qu'était le Graal,

pour que tout le monde comprenne.

C'était difficile, alors j'ai essayé de rigoler

pour que personne ne s'ennuie.

J'ai raté, mais je veux pas qu'on dise que j'ai rien foutu,

parce que c'est pas vrai. »

Alexandre Astier, *Kaamelott*, Livre V, écrit par Alexandre Astier

Table des Matières

Abréviations 23

Introduction 26

I. Généralités sur le cancer 26

II. Epidémiologie du cancer du sein en France 27

III. Les cancers du sein et de l'ovaire à prédisposition génétique 28

A. Les prédispositions génétiques au(x) cancer(s) 28

B. BRCA1 et BRCA2 29

1. *BRCA1* 29

a. Structure du gène *BRCA1* 30

b. Protéine BRCA1 30

2. *BRCA2* 31

a. Structure du gène *BRCA2* 31

b. Protéine BRCA2 31

3. Fonctions de *BRCA1* et *BRCA2* 32

a. Régulation de la transcription 32

b. Détection, signalisation et réparation des dommages de l'ADN 33

c. Autres fonctions de BRCA1 et BRCA2 34

BRCA1 et le remodelage de la chromatine 34

BRCA1, BRCA2 et la régulation du cycle cellulaire 34

Activité ubiquitine ligase de BRCA1 34

4. Pathologie moléculaire des gènes *BRCA1 et BRCA2* 35

a. Les mutations ponctuelles 35

b. Les réarrangements génomiques 36

5. Cas des variants de signification inconnue (VSI) 36

C. Les autres gènes responsables 37

1. Les autres gènes associés à un risqué tumoral élevé 38

a. *TP53* et le syndrome de Li-Fraumeni 38

b. *PTEN* et la maladie de Cowden. 39

c. *STK11* et le syndrome de Peutz-Jeghers 39

d. *CDH1* 39

2. Les autres gènes associés à un risque tumoral modéré 39

a. *ATM* 39

b. *CHEK2* 39

c. *BRIP1* et *PALB2* 40

d. *RAD51C* 40

3. Autres gènes associés à un risque tumoral faible 40

IV. L'épissage 41

A. Transcription et maturation des ARN pré-messagers 41

B. L'épissage : mécanismes et régulation 41

C. L'épissage alternatif 44

D. Epissage alternatif et pathologie 46

E. Nonsense mediated mRNA decay : le NMD 47

F. Les transcrits alternatifs de *BRCA1* et *BRCA2* 48

1. Principaux transcrits alternatifs observés pour *BRCA1* 48

2. Principaux transcrits alternatifs observés pour *BRCA2* 49

V. Next-Generation of Sequencing (NGS) : les technologies de séquençage à haut-débit 50

A. Principales technologies de NGS 50

1. Le pyroséquençage 50

2. Le séquençage par synthèse 51

3. Le séquençage par ligation 51

4. Le séquençage par nano-mesure de pH 52

B. Les séquenceurs dits « de paillasse » 52

C. Les méthodes de 3ème génération 53

D. Analyse Bio-informatique 56

1. L'alignement 58

2. Détection des variants 61

3. Applications dans la détection des transcrits 62

1ère partie : Application diagnostique du séquençage à haut débit par PGM pour *BRCA1* et *BRCA2* 64

I. Patients 64

II. Protocole de séquençage PGM 65

A. Enrichissement 65

B. Préparation des banques 66

C. PCR Emulsion (PCRe) 67

D. Séquençage à haut débit 69

III. Analyse des données de séquençage à l'aide du logiciel NextGENe ® (SOFTGENETICS ®) 73

A. Prise en charge des anomalies détectées 73

B. Contrôle des variants 76

IV. Exemple concret d'une série diagnostique 76

V. Discussion 79

A. Optimisation du processus actuel 79

B. Difficultés liées aux fournisseurs 80

C. Difficultés liées au traitement bio-informatique des données 80

2ème partie : Utilisation de la plateforme PGM en recherche ; étude de l'impact des variants d'épissage pour *BRCA1* et *BRCA2* 82

I. Introduction 82

II. Matériel et Méthodes 83

A. Patients et mutations 84

B. RT-PCR 86

C. Préparation des banques 86

1. *BRCA1* 86

2. *BRCA2* 88

D. Séquençage (cf. Annexe 4) 88

1. BRCA1 88

2. *BRCA2* 89

E. Analyse des résultats 90

1. Viewer Transcriptome de Next GENe 91

2. « *Transcript Report* » 92

III. Résultats et Discussion 93

A. Analyse des résultats obtenus pour la délétion génomique de l'exon 3 de BRCA1 94

1. Analyse « transcriptome » des données obtenues par le séquençage des amplicons de BRCA1 de manière isolée. 94

2. Analyse « transcriptome » des résultats obtenus par séquençage des mélanges d'amplicons. 95

B. Analyse « transcriptome » des autres patients inclus 97

C. Difficultés d'analyse liées au module « transcriptome » de NextGENe 98

1. « Transcript report » inexploitable 98

2. Variations importantes des taux de couverture 99

3. Anomalies d'alignement..... 99

4. Bilan de l'analyse transcriptome 100

D. Alignement génomique des read dans Next GENE : Exemple de la mutation c.9117G>A de *BRCA2*..... 100

E. Etudes et visualisation des *read* à l'aide d'ALAMUT 104

1. Visualisation des mutations et de leur impact sur l'épissage : exemple de la mutation c.4484G>T entrainant un saut hors phase de l'exon 14 104

2. Visualisation des transcrits alternatifs classiquement décrits pour les gènes *BRCA1* et *BRCA2* 106

3. Autres anomalies détectées : visualisations répétées d'inclusion de séquences introniques pouvant s'apparenter à des « pseudo-exons » 106

IV. Conclusions et perspectives 109

A. Projet « Sure Select » (Agilent) 111

B. Projet « AmpliSeq RNA » (Life Technologies)..... 112

Conclusion Générale 113

Bibliographie 116

Annexes 119

1. Protocole de RT-PCR utilisé pour l'étude des ARN de *BRCA1* 119

2. Protocole de RT-PCR utilisé pour L'étude des ARN de *BRCA2* 119

3. Design des Amorces utilisées pour l'amplification « maison » des ADNc de BRCA1 et BRCA2 120

4. Protocoles de séquençage à haut débit par PGM – laboratoire de Génétique – Institut Curie..... 122

Abréviations

ADN : acide désoxyribonucléique

ADNc : ADN complémentaire

ARHGEF6 : Rho Guanine nucleotide Exchange Factor 6

ARN : Acide ribonucléique

ARN : acide ribonucléique

ARNm : ARNmessager

ARNm : ARN messager

ATP : Adénosine Tri-Phosphate

BRCA1 : gène breast cancer 1

BRCA2 : gène breast cancer 2

CGH-array : Hybridation génomique comparative sur microréseau

cm : Centimètre

CNV : Copy number variation

CNV : Copy number variation

DMSO : Dimétylsulfoxide

dNTP : Désoxyribonucléoside triphosphate

DS : Déviation standard

EDTA : Acide éthylène diamine tétraacétique

kb : Kilobase

kg : Kilogramme

Mb : Megabase

mg : Milligramme

miARN : micro ARN

min : Minute

ml : Millilitre

NAHR : Recombinaison homologue non allélique

NCBI : National Center for Biotechnology Information

ng : Nanogramme

NGS : Séquençage de nouvelle génération

NHEJ : Non homologous end joining

NMD : Nonsense mediated mRNA decay

nt : nucléotide

OMIM : Online Mendelian Inheritance in Man

p : Bras court de chromosome

pb : Paires de bases

pb : paire de base

PBS : Phosphate buffered saline

PCR : Réaction en chaîne par polymérisation (polymerase chain reaction)

PCRe : PCR en émulsion

PE : Paired-end librairie

PGM : Personal Genome Machine

q : Bras long de chromosome

qPCR : PCR quantitative en temps réel

RefSeq : Base de données Reference Sequence

RT : Reverse transcriptase

RNASeq : Séquençage d'ARN

SDS : Sequence Detection System

SE : Single-end librairie

SNP : Single Nucleotide Polymorphism

SNV : Single Nucleotide Variant

TE : Tris-HCl et EDTA

TGFB1 : Transforming growth factor, beta 1

Tm : Température de demi-dissociation

VSI : Variant de Signification Inconnue

λ : Longueur d'onde

Introduction

I. Généralités sur le cancer

Le cancer est un groupe hétérogène de maladies : plus de 100 sous types de cancers sont connus chez l'humain dont plusieurs peuvent affecter un même organe[1].

Les recherches de ces 20 dernières années n'ont cessé d'essayer de confirmer l'idée générale considérant le cancer comme une maladie génétique de la cellule[2].

La transformation tumorale d'une cellule est un processus clonal procédant par étapes successives (le plus souvent des mutations) et permettant à la cellule d'acquérir de nouvelles fonctions qui lui confèrent un avantage sélectif par rapport aux autres cellules. Par exemple, le dérèglement du cycle cellulaire, l'inhibition de l'apoptose et de la différentiation ou l'acquisition de capacités angiogéniques ou métastatiques, sont autant d'étapes permettant d'acquérir des capacités de survie anormales et d'augmenter le potentiel de prolifération et de dissémination des cellules touchées[3,4].

Ce dérèglement peut avoir plusieurs sources : l' « hérédité » ou l' « inné » (les prédispositions génétiques), ou l' « acquis » par les répercussions sur l'ADN de phénomènes extérieurs (action d'un ou plusieurs composés mutagènes, réarrangements chromosomiques, amplification ou perte de matériel génomique et tout autre mécanisme pouvant mener à une transcription anormale d'une partie de l'acide désoxyribonucléique (ADN))[5,6].

Ces étapes cellulaires d'évolution clonale sont gouvernées par des mutations ou des modifications épigénétiques ayant des répercussions dans une ou plusieurs voies cellulaires impliquées dans les processus cités précédemment (apoptose, différentiation, régulation du cycle cellulaire, etc).

Ces mutations peuvent être :

- Soit « activatrices » : elles conduisent dans ce cas à la surexpression ou l'hyperactivité de la protéine concernée (amplification génique, mutation faux-sens) ou à la création d'une nouvelle protéine de fusion.
  Ces mutations ciblent les « oncogènes » et l'atteinte d'un seul allèle suffit à conférer un avantage sélectif à la cellule.
- Soit « inactivatrices » : ces mutations (codon stop prématuré, insertion ou délétion, anomalies d'épissage, faux-sens, réarrangements de grande taille) sont le plus souvent accompagnées d'une inactivation du second allèle au sein de la tumeur (généralement par perte de la région chromosomique correspondante). Ces mutations ciblent les « gènes suppresseurs de tumeur ».

II. Epidémiologie du cancer du sein en France

Le cancer du sein est le type de cancer le plus fréquemment diagnostiqué chez la femme : environ 1,38 millions de nouveaux cas diagnostiqués en 2008[7].

En 2012, on recensait 48763 nouveaux cas en France ce qui le place au 2ème rang des cancers les plus fréquents dans la population générale française et au 1er rang chez les femmes.

Les taux d'incidence standardisés ont fortement augmenté jusqu'en 2000 avant de se stabiliser puis de diminuer à partir de 2005.

Ces variations importantes rendent difficiles la modélisation et l'estimation des taux d'incidence et doivent être interprétés avec prudence. En effet, le taux d'incidence standardisé a augmenté de 1,4 % par an en moyenne entre 1980 et 2012 mais on observe une diminution de 1,5 % par an si on se focalise sur la période située entre 2005 et 2012 (56,3 cas pour 100 000 personnes-années en 1980, 97,8 en 2005 et 88,0 en 2012). La diminution de l'incidence concerne principalement les tumeurs diagnostiquées à un stade précoce (T1/T2-N0-M0) et est surtout observée dans le groupe des femmes âgées de 50 à 74 ans. Un facteur important de cette baisse de l'incidence depuis 2005 pourrait être la diminution de la

prescription de traitements hormonaux de la ménopause mais d'autres facteurs tels l'effet de saturation du dépistage organisé ou individuel doivent être discutés.

Le cancer du sein est un cancer de très bon pronostic. La mortalité est restée relativement stable jusqu'aux alentours de 1995 malgré une forte augmentation de l'incidence durant cette période, puis a diminué significativement jusqu'en 2012. On constate en effet une diminution moyenne de la mortalité de 0,6 % par an entre 1980 et 2012, et de 1,5 % par an entre 2005 et 2012 et donc un risque de décéder de ce cancer entre 0 et 74 ans en diminution (rapport de juillet 2013 de l'INCa et de l'InVS)[8].

En effet, sa survie nette est de 89 % à 5 ans pour les cancers diagnostiqués entre 2001 et 2004. Elle s'est améliorée au cours du temps puisqu'elle était de 81 % pour les cancers diagnostiqués entre 1989 et 1991. Le diagnostic à un stade de plus en plus précoce de ces cancers grâce au dépistage et l'amélioration de la prise en charge thérapeutique contribuent à l'augmentation de la survie.

III. Les cancers du sein et de l'ovaire à prédisposition génétique

A. Les prédispositions génétiques au(x) cancer(s)

La notion de prédisposition au(x) cancer(s) correspond à une augmentation du risque de cancer chez une personne par rapport à la population générale. Il s'agit de situations où le quota de mutations nécessaires à la transformation cellulaire est atteint plus rapidement puisqu'une étape est déjà franchie de manière constitutionnelle au niveau d'un gène clé[9].

Cette notion a été introduite en 1971 par Knudson dans le cadre du rétinoblastome (tumeur de la rétine chez le jeune enfant) puis complété par Comings en 1973 ; il s'agit d'une inactivation bi-allélique des gènes suppresseurs de tumeur (théorie du « double-hits » de Knudson). Le gène *RB1* mis en cause a d'ailleurs été le 1er gène suppresseur de tumeur à être mis en évidence.

Knudson avait alors expliqué l'existence de deux types de rétinoblastomes : héréditaires (1/3 des cas) et sporadiques, par un modèle basé sur la nécessité de deux évènements consécutifs au développement de ce type de tumeur. Ainsi, dans les cas héréditaires on retrouve :

- Une 1ère mutation, germinale, héritée et donc présente dans toutes les cellules de l'individu
- Une 2nde mutation, somatique et donc acquise par perte ou mutation de l'allèle normal restant jusque-là.

Dans les cas sporadiques par contre il faut que ces deux altérations viennent successivement toucher chacune des deux copies normales du gène *RB1* dans la même cellule cible.

Les cancers du sein d'origine familiale ne représentent qu'une faible proportion des cancers mammaires totaux : environ 5 à 7% des cas. Dans cette proportion, environ ¼ des cas peut être attribué aux mutations germinales touchant les deux gènes principaux de prédisposition que sont *BRCA1* (MIM# 113705) et *BRCA2* (MIM# 600185).

B. BRCA1 et BRCA2

Les mutations germinales de *BRCA1* et *BRCA2* entrainent respectivement un risque relatif cumulé de 60% et 55% pour le cancer du sein à l'âge de 70 ans, de 59% et 16,5% pour le cancer de l'ovaire et de 83% et 62% pour les cancers du sein controlatéraux[10].

Les risques de cancers du sein et de l'ovaire sont augmentés dans de plus faibles proportions dans le cas des mutations de *BRCA2* mais on y retrouve une proportion plus importante de cancers mammaires masculins.

En plus du risque de cancer du sein et de l'ovaire, les mutations de *BRCA1* et *BRCA2* sont aussi associées à un risque accru de cancers de localisations différentes (notamment pour le cancer du pancréas et *BRCA2*)[11].

Les gènes *BRCA1* et *BRCA2* présentent tous les deux un grand nombre d'exons avec la particularité d'avoir chacun dans leur partie centrale un exon 11 de grande taille. Aucune homologie de séquence n'a été retrouvée ni dans ces deux exons, ni même dans ces deux gènes.

1. *BRCA1*

A partir d'une trentaine de familles et de l'hypothèse d'un modèle de transmission autosomique dominant préalablement établi (par étude de ségrégation), Mary-Claire King est

parvenue à localiser *BRCA1* en 17q21 en 1990. D'autres études en partenariat avec plusieurs laboratoires ont ensuite permis d'identifier précisément le gène[10].

BRCA1 est un gène contrôle de l'intégrité génomique impliqué à de nombreux niveaux cellulaires : réparation de l'ADN par recombinaison homologue, point de contrôle du cycle cellulaire, régulation de la transcription. La description du gène a été établie par Miki et al. en 1994[12].

a. Structure du gène *BRCA1*

Sa séquence est composée de 5592 nucléotides et code une protéine de 1863 acides aminés.

L'ADN génomique s'étend sur 81 kb et comporte 24 exons dont deux considérés comme non codants :

- L'exon 1 : en amont du codon d'initiation de traduction ATG (situé dans l'exon 2). Il correspond à la majeure partie de la région 5'UTR du gène
- L'exon 4 : identifié initialement comme un exon, il s'agit en fait d'une séquence Alu non codante. Il n'est d'ailleurs même pas représenté le plus souvent.

L'exon 11 représente à lui seul 3426 pb (60% de la séquence codante de BRCA1). Cette taille importante rend d'ailleurs sont étude délicate. Les autres exons quant à eux sont de plus petites tailles.

Les séquences introniques (de tailles variant de 403 pb à 9,2 kb) représentent 91% du gène et contiennent une forte quantité de séquences répétées de type Alu.

b. Protéine BRCA1

Il s'agit d'une protéine de 1863 acides aminés dont seules les régions N- et C- terminales sont fortement conservées entre les espèces mammifères[13].

Elle comprend :

- un domaine RING en N-terminal (acides aminés 24 à 64) caractérisé par la présence de 8 résidus cystéines et histidines responsables de la liaison de deux atomes de Zn^{2+}. Ce motif fait partie d'un domaine plus étendu (acides aminés 1-109) incluant des résidus en N- et C- terminal du motif. Ces acides aminés forment des hélices alpha antiparallèles indispensables aux interactions protéiques dans cette région.

- 2 domaines BRCT retrouvés à l'extrémité C-terminal. Ils sont constitués d'environ 100 acides aminés. Ils sont repliés en 4 feuillets Beta parallèles entourés de deux hélices alpha d'une coté et d'une hélice alpha unique de l'autre. L'analyse de séquences protéiques a permis de démontrer que ce genre de domaines était souvent retrouvé dans une superfamille de protéines impliquées dans la réponse aux dommages de l'ADN et la régulation du cycle cellulaire.

2. *BRCA2*

Wooster et al. (1995) ont identifié le gène *BRCA2* par clonage positionnel de la région chromosomique 13q13-q13 impliquée dans des histoires familiales de cancer du sein dans des familles islandaises[14]. Tavtigian et al. (1996) ont quant à eux déterminé la séquence codante complète du gène et sa structure exonique ainsi que son mode d'expression[15].

a. Structure du gène *BRCA2*

La séquence codante du gène *BRCA2* est composée de 10254 nucléotides et code une protéine de 3418 acides aminés.

L'ADN génomique s'étend sur 84 kb et comporte 27 exons avec trois grands exons : l'exon 10 (1,1 kb), l'exon 11 (4,9 kb) et l'exon 27 (1,5 kb). Les autres exons ont une taille variant de 41 pb à 428 pb.

Le codon initiateur ATG est situé dans l'exon 2 et l'extrémité 5'UTR est, ici aussi, représentée par l'exon 1.

b. Protéine BRCA2

Le gène *BRCA2* encode une protéine de 3418 acides aminés. Celle-ci présente très peu d'homologie avec d'autres protéines et aucune avec BRCA1. On note la présence d'un motif d'environ 70 acides aminés (appelé BRC) répété 8 fois dans la partie centrale de la protéine.

Une découverte majeure dans le mécanisme de fonctionnement cellulaire de BRCA2 réside dans la découverte de l'interaction entre BRCA2 et RAD51 par le biais de ces répétitions[16].

Cependant, ces motifs variant légèrement d'une répétition à l'autre, tous ne présentent pas la même affinité pour la recombinase RAD51 (les répétitions BRC5 et BRC6 présentant l'affinité la plus faible).

La portion C-terminale comprend une région de 800 acides aminés environ qui forme des structures secondaires, dont des motifs similaires à des domaines de liaison à l'ADN simple brin.

Le premier motif de cette région comprend 190 acides aminés qui forment principalement des hélices alpha (domaine « hélical »).

Il est suivi de trois domaines d'environ 110 acides aminés (OB1,OB2 et OB3) présentant une homologie avec le repliement OB présent chez plusieurs protéines liant l'ADN simple brin.

Pour finir, elle comprend également deux signaux de localisation nucléaire (SLN) et un site de phosphorylation par des kinases cyclines dépendantes (CDK) qui semble réguler l'association de RAD51 à un deuxième motif de liaison à cet endroit[17]

3. Fonctions de *BRCA1* et *BRCA2*

BRCA1 et *BRCA2* sont exprimés de façon ubiquitaire chez l'être humain. Cette expression corrèle de près avec les degrés de prolifération cellulaire et varie au cours du cycle cellulaire avec des niveaux maximum à la transition G1/S[18–20].

BRCA1 et BRCA2 sont des protéines dites « caretakers » dont l'inactivation crée un état permissif dans lequel la cellule accumule des défauts cellulaires qui amènent une instabilité chromosomique.

Elles sont impliquées dans des fonctions essentielles telles que la régulation de la transcription, la détection et la réparation des brins de l'ADN, le contrôle du cycle cellulaire et le remodelage de la chromatine.

a. Régulation de la transcription

Concernant *BRCA1*, il entre en jeu dans la machinerie transcriptionnelle de base en interagissant avec le complexe de l'ARN polymérase II par l'intermédiaire de l'ARN hélicase A[21]. Il peut réguler l'activité de l'holoenzyme ARN polymérase II en modifiant l'état de phosphorylation de son domaine C-terminal par la kinase CAK. De plus, BRCA1 associée à BARD1 (BRCA1-associated RING domain 1) pourrait utiliser son activité ubiquitine ligase pour bloquer l'initiation de la synthèse de l'ARNm en ubiquitinant le complexe de pré-initiation transcriptionnel[22].

Il a également été démontré que BRCA1 pourrait agir comme co-régulateur sélectif de la transcription de gènes spécifiques. Plusieurs facteurs de transcription (à la fois co-répressurs et co-activateurs) interagissent avec BRCA1 dont p53, ESR1, CtIP et ZNF350.

En s'associant à p53, BRCA1 le stabilise et stimule son activité transcriptionnelle. Ceci induit un sous-groupe de gènes régulés par p53, et il semblerait que l'association de BRCA1 à p53 redirige l'activité de celle-ci des voies apoptotiques vers la réparation de l'ADN et l'arrêt du cycle cellulaire.

BRCA1 a aussi été associée à une répression de la signalisation par ER-α, probablement à travers la liaison compétitive de co-facteurs (p300/CBP, cycline D1, BRD7). Cette répression cause l'inhibition de la croissance cellulaire induite par l'œstrogène.

Par contre, le rôle de BRCA2 demeure obscur. Des études ont proposé que les acides aminés 23 à 105 pourraient activer la transcription via leurs interactions avec des protéines co-activatrices (P/CAF et GRIP1).

b. Détection, signalisation et réparation des dommages de l'ADN

Les cellules déficientes en BRCA1 et BRCA2 sont inefficaces à la réparation des dommages causés à l'ADN. En fonction du type cellulaire et du moment du cycle, la cellule dispose de différentes voies de réparation de l'ADN :

- La réparation par recombinaison homologue (homologous recombination, HR)
- La réparation par la jonction des extrémités non-homologues (non-homologous end joining, NHEJ)

La réparation homologue nécessite une chromatide sœur intacte comme modèle. Elle est considérée comme la voie la plus fiable (elle entre donc en jeu aux phases cellulaires S et G2).

Les fonctions de BRCA1 dans la réparation de l'ADN consistent surtout dans un rôle de médiateur entre les protéines de détection des bris et les protéines de réparation. En plus d'être impliqué dans la voie de recombinaison homologue, certaines études suggèrent aussi une implication de BRCA1 dans la réparation de type NHEJ. L'évidence de la participation de BRCA1 dans plusieurs aspects de la réponse aux cassures de l'ADN se reflète par ses interactions protéiques et sa présence dans plusieurs complexes différents.

BRCA2 est quant à elle surtout impliquée au niveau de la voie de réparation HR et la voie NHEJ semble intacte chez les cellules déficientes en BRCA2. BRCA2 est notamment étroitement impliqué dans la réparation de l'ADN par le biais de son interaction avec la recombinase RAD51 et de sa régulation. BRCA2 serait nécessaire à la localisation et au fonctionnement adéquat de RAD51 au niveau des foyers de dommages de l'ADN.

BRCA2 interagit avec RAD51 via ses motifs répétés BRCs qui pourraient lui permettre de séquestrer RAD51 en dehors des périodes de réparation. Par contre, une fois la voie de réparation HR enclenchée, il serait responsable de la relocalisation de RAD51 au niveau de la chromatine. Cette interaction serait ensuite stabilisée par une seconde interaction des multimères de RAD51 avec l'extrémité C-terminale de BRCA2. RAD51 permet l'invasion de la seconde molécule d'ADN à la recherche de la séquence homologue qui servira de modèle à la réparation[23].

c. Autres fonctions de BRCA1 et BRCA2

En plus de ces activités, BRCA1 et BRCA2 possèdent d'autres fonctions cellulaires généralement moins bien connues.

BRCA1 et le remodelage de la chromatine

L'activité de régulation transcriptionnelle de BRCA1 semble s'effectuer à la fois par le biais de co-activateurs et co-répresseurs, mais aussi par le recrutement de protéines impliquées dans le remodelage de la chromatine.

BRCA1, BRCA2 et la régulation du cycle cellulaire

L'activité de BRCA1 dans la régulation de plusieurs points de restriction du cycle cellulaire est effectuée en partie par le contrôle transcriptionnels de gènes clés (exemple : *GADD45A*) et par son association à des protéines de régulation. L'implication de BRCA2 dans ces mécanismes de régulation du cycle cellulaire est moins claire.

Activité ubiquitine ligase de BRCA1

BRCA1 s'associe de manière constitutive à BARD1 ce qui est indispensable à l'activité ubiquitine ligase de BRCA1. L'hétérodimère BRCA1-BARD1 participerait à plusieurs complexes cellulaires qui pourraient « ubiquitiner » différents substrats comme RNA pol II, Erα, CtIP et TFIIE.

4. Pathologie moléculaire des gènes *BRCA1 et BRCA2*

Les gènes possèdent une structure spécifique nécessaire à leur expression adéquate. Leurs séquences nucléotidiques peuvent être la cible de différentes modifications soit bénignes, soit délétères ou encore de conséquences indéterminées dans l'état actuel de nos connaissances (cf. III. B. 5.).

Dans le cas des gènes *BRCA1* et *BRCA2*, il s'agit de gènes suppresseurs de tumeurs dont les mutations sont de type « perte de fonction ». Dans le développement du cancer du sein (et/ou de l'ovaire) lié à *BRCA1* ou *BRCA2*, le 1er allèle est inactivé par mutation constitutionnelle et le 2ème allèle est généralement délété au niveau de la tumeur.

L'étude constitutionnelle des gènes *BRCA1* et *BRCA2* chez des patients à haut risque de prédisposition génétique a permis de mettre en évidence un très grand nombre de mutations différentes affectant ces gènes. En France, la base de données développée par Rosette Lidereau (Institut Curie, Paris) et Christophe Béroud (INSERM UMR-S910 Marseille)[24] regroupe aujourd'hui près de 4000 mutations différentes (1675 pour *BRCA1* et 2196 pour *BRCA2*)(www.umd.be)[25]. Il a été rapporté quelques mutations récurrentes sur ces gènes liés à des effets fondateurs (par exemple, la mutation c.3481_3491del11 de *BRCA1* décrites dans de nombreuses familles originaires de l'est de la France)[26] mais la grande majorité d'entre elles sont des mutations dites « privées ».

On reconnait généralement deux types d'altération de séquence : les mutations ponctuelles (substitutions nucléotidiques et délétions ou insertions de quelques nucléotides) et les réarrangements génomiques (translocations, inversions, délétions ou duplications d'un exon au gène entier, voir plus).

a. Les mutations ponctuelles

Leur effet dépend de leur position, celles-ci pouvant survenir au niveau :

- **Des séquences régulatrices en amont du gène ou dans la région promotrice**
  Ceci peut entrainer une modification quantitative de l'expression du gène (par altération de séquences consensus transcriptionnelles)
- **Des régions codantes**

- Les substitutions peuvent affecter le codon d'initiation de la traduction, entrainer la substitution d'un acide aminé par un autre au niveau de la protéine (variation faux-sens), changer un nucléotide sans pour autant amener à un changement d'acide aminé (variation neutre), amener à la création d'un codon stop prématuré (mutation non-sens).
- Les insertions ou délétions de petite taille peuvent décaler le cadre de lecture (***frameshift***). Lorsque le nombre de nucléotides insérés ou délétés n'est pas un multiple de 3, le cadre de lecture de l'acide ribo-nucléique messager (ARNm) est alors décalé et ceci aboutit le plus souvent à l'apparition d'un codon stop prématuré entrainant la synthèse d'une protéine tronquée ou l'absence de protéine.

  Elles peuvent aussi ne pas décaler le cadre de lecture (***inframe***) quand le nombre de nucléotidés insérés ou délétés est un multiple de 3. Dans ce cas la traduction de l'ARNm altéré aboutit à une protéine présentant un ou plusieurs acides aminés ajoutés ou supprimés.

- **Des introns**

  Les mutations introniques peuvent perturber l'épissage par altération des séquences consensus régulatrices (voir plus loin, paragraphe IV).

- **Des régions en aval du gène**

  Celles-ci peuvent perturber l'arrêt de la transcription ou la stabilité de l'ARNm.

b. Les réarrangements génomiques

De façon similaire, les grands réarrangements intéressant les régions promotrices peuvent altérer l'expression d'un gène (par exemple en emportant les séquences consensus).

De façon plus large, en fonction de leur taille et de leur position, ils aboutissent à une protéine tronquée ou à l'absence de protéine.

5. Cas des variants de signification inconnue (VSI)

Parmi l'ensemble des variations identifiées sur les gènes BRCA1 et *BRCA*2, environ 70% de celles-ci sont reconnues clairement comme étant délétères. Pour les autres, *BRCA1* et *BRCA2*

sont actuellement des gènes exemplaires dans les difficultés d'interprétation des variants de signification inconnue (VSI ou VUS pour *Variants of Unknown Significance*). 10 à 30% des variants identifiés en routine diagnostique ne fournissent pas d'interprétation lisible en génomique (données COVAR non publiées)[27]. Nous ne sommes donc pas toujours capables de prédire de façon certaine la répercussion de ces variations sur la fonction protéique. Un des enjeux majeurs pour le diagnostic génétique est de pouvoir déterminer l'impact réel de ce type de variation dans le risque de cancer. Plusieurs clés d'interprétation sont possibles pour élucider l'impact de ces VSI : cooccurrence avec des variants délétères, présence ou non dans un domaine fonctionnel de la protéine, fréquence, écart physico clinique, co-ségrégation, tests fonctionnels et impact sur l'épissage.

Comme nous l'avons vu précédemment pour les mutations de *BRCA1* et *BRCA2*, les VSI peuvent aussi avoir différents impacts en fonction de leur position :

- Impact sur la protéine (par modification d'un acide aminé)
- Impact sur l'épissage (si situé dans un site consensus)[3]

Cependant, l'étude des ARN est délicate car il s'agit d'une molécule fragile et accessible uniquement en faible quantité. Le développement d'outils bio-informatiques destinés à identifier les séquences contrôlant l'épissage et à prédire leurs altérations est une aide précieuse pour orienter le classement des VSI.

C. Les autres gènes responsables

On estime que seulement 20 % environ des histoires familiales de cancer du sein seul (c'est-à-dire non associé à un cancer de l'ovaire) sont expliquées par des altérations des gènes *BRCA1* et *BRCA2*. Ceci laisse donc penser qu'il existe sans doute un large panel d'autres gènes responsables encore incomplètement exploré.

Aujourd'hui, de nombreux gènes ont été identifiés comme liés à un risque plus ou moins modéré de développer un cancer du sein. Nous pouvons d'ailleurs les classer en fonction de leur risque de développement de cancer du sein : élevé, moyen ou faible[28](Figure 1 et 2).

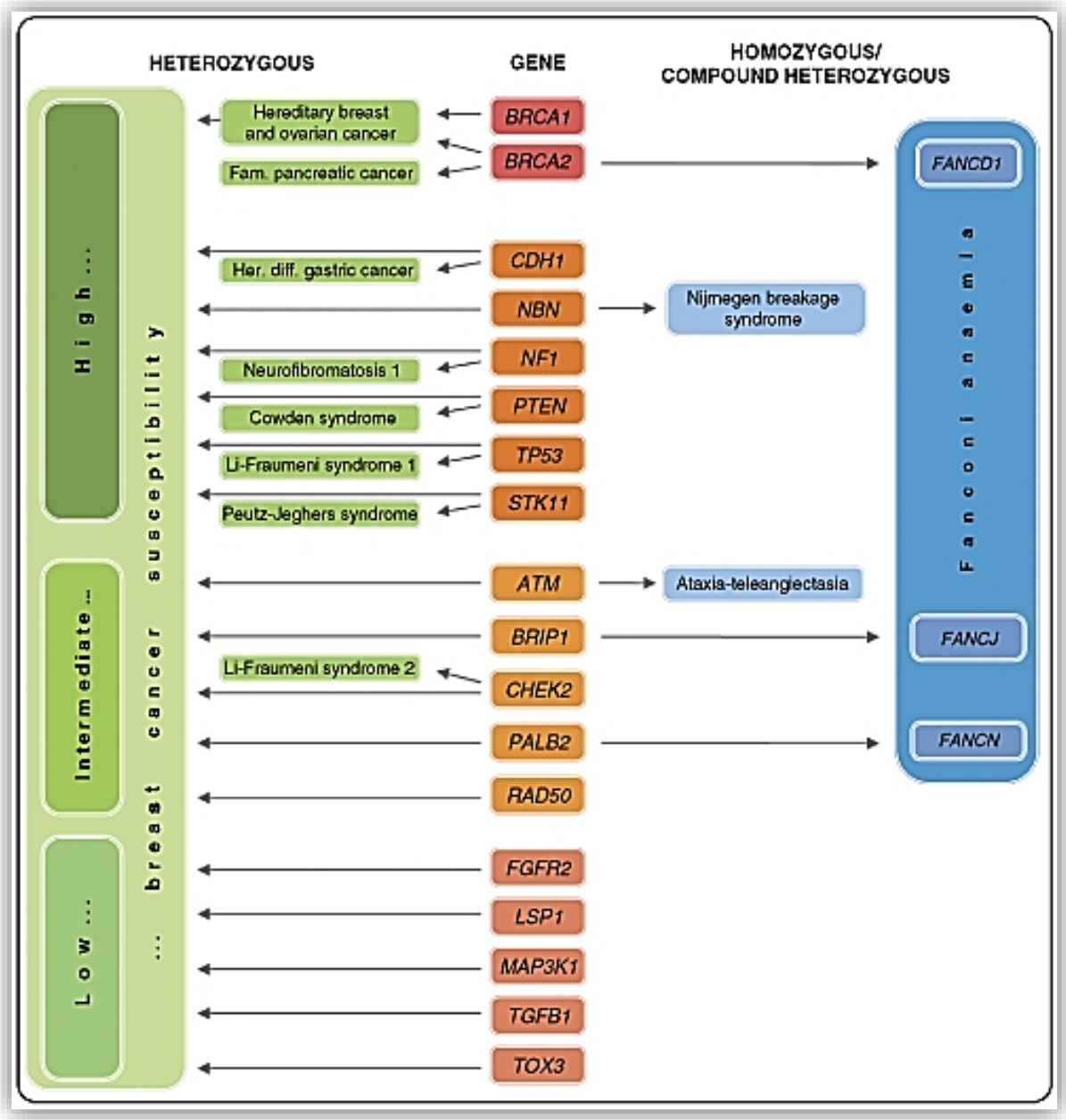


Figure 1 : Corrélation géno-phénotypique pour les mutations hétérozygotes ou homozygote ou hétérozygote composite des gènes de prédisposition aux cancers du sein et pour les SNP de susceptibilité aux cancers du sein de faible pénétrance[27]

1. Les autres gènes associés à un risqué tumoral élevé

a. *TP53* et le syndrome de Li-Fraumeni

Ce syndrome prédispose à de multiples cancers d'apparition précoce touchant principalement le sein, le cerveau, les muscles, les surrénales et le sang. Le risque de cancer du sein associé est de l'ordre de 40% à 45 ans (RR de 4).

b. *PTEN* et la maladie de Cowden.

C'est une maladie très rare (1/100 000) caractérisée par des hamartomes multiples et disséminés et par un risque accru de développer certaines tumeurs malignes.

c. *STK11* et le syndrome de Peutz-Jeghers

Ce syndrome également très rare (1/100 000) appartient au groupe des polyposes digestives hamartomateuses. Il est également associé à un risque accru de développement de cancers, notamment digestifs. Concernant le cancer du sein, le risque cumulé à 65 ans est alors supérieur à 50% (RR > 5).

d. *CDH1*

Ce gène suppresseur de tumeur code pour la E-cadhérine, protéine d'adhérence celullaire. Ses mutations sont liées aux carcinomes gastriques diffus et aux carcinomes mammaires de type lobulaire (risque d'environ 40% à 80 ans, RR de 4).

2. Les autres gènes associés à un risque tumoral modéré

Ces autres gènes sont impliqués à l'état hétérozygote dans un risque tumoral plus modéré (RR de l'ordre de 2 à 4).

a. *ATM*

Les mutations de ce gène à l'état homozygote sont responsables de l'Ataxie-Télangiectasie, maladie récessive rare caractérisée par une ataxie cérébelleuse progressive, des télangiectasies surtout visibles au niveau de la conjonctive et une augmentation du risque de cancer.

Les patients porteurs d'une altération d'*ATM* à l'état hétérozygote ont un risque de développer un cancer du sein augmenté de 2 par rapport à la population générale.

b. *CHEK2*

Il en associé à des prédispositions génétiques de cancer du sein, notamment dans le cas de la mutation 1100delC

c. *BRIP1* et *PALB2*

Ils font partie des gènes impliqués dans la maladie de Fanconi, sont également appelés *FANC-J* et *FANC-N*.

L'étude de *PALB2* dans les prédispositions au cancer du sein a fait l'objet de nombreuses publications récentes. Celles-ci ont pu démontrer un risque relatif augmenté de 2 à 4 selon le type d'étude[29,30].

d. *RAD51C*

Ce gène paralogue de *RAD51* a été récemment identifié comme étant un nouveau gène de prédisposition au cancer du sein. Il est lui aussi lié à l'anémie de Fanconi[31]

3. Autres gènes associés à un risque tumoral faible

Des études plus récentes tendent à identifier plusieurs gènes de faible susceptibilité (Figure 2) mais malgré tout, il reste environ 50% des cas d'histoires familiales de cancer du sein qui restent inexpliqués[32]. Le risque réel conféré par ces variants reste encore flou à l'heure actuelle et la généralisation des observations pour ce type d'effet très faible est complexe.

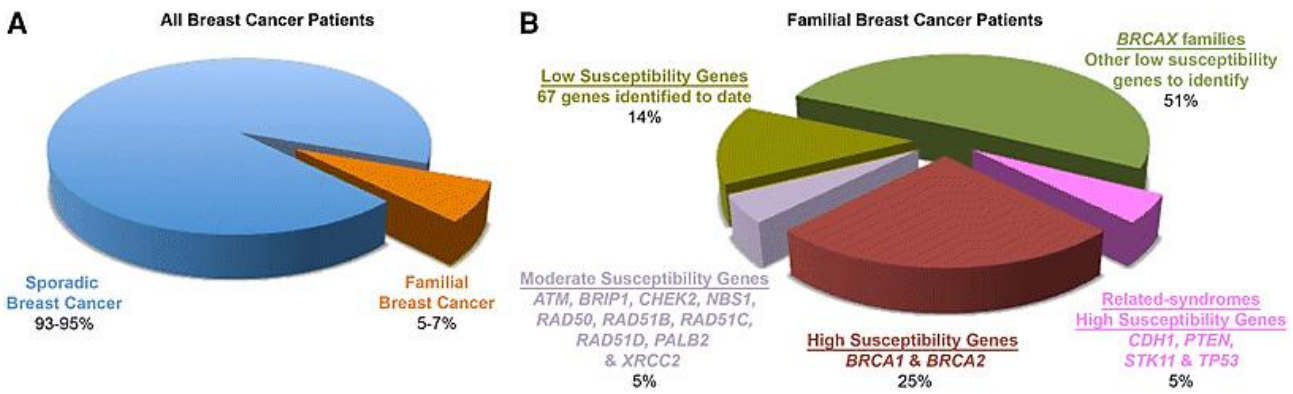


Figure 2 : Distribution des patients atteints de cancer du sein.[33]

A) Les cancers du sein familiaux représentent un pourcentage mineur par rapport à l'ensemble des cas de cancers du sein.

B) Proportion des cas familiaux de cancers du sein dus à des mutations constitutionnelles dans les gènes de prédisposition de forte, modérée et faible pénétrance.
BRCA1 et BRCA2 expliquent à eux seuls 25% des histoires familiales.
Cependant la majorité des familles ne présentent pas de mutation dans les gènes de susceptibilité aux cancers connus et présentent donc probablement des variations dans un ou plusieurs gènes de faible pénétrance non connus. (familles BRCAX)

IV. L'épissage

A. Transcription et maturation des ARN pré-messagers

La transcription est l'étape nucléaire au cours de laquelle sont synthétisés les différents types d'ARN à partir de l'ADN. La liaison de facteurs de transcription au niveau de certaines régions promotrices du génome (boite TATA) permet le recrutement de la machinerie d'initiation incluant les polymérases (polymérase II) permettant la synthèse des brins d'ARN. L'élongation de l'ARN se poursuit jusqu'à ce que la machinerie transcriptionnelle rencontre un site de terminaison (site de polyadénylation).

Les ARN ainsi obtenus sont nommés « ARN pré-messagers » (ou pré-ARNm). Ce pré-ARNm est une copie de l'ADN génomique contenant l'information nécessaire à la synthèse de la protéine que code le gène. Il devra subir des étapes de maturation afin d'obtenir ses « ARN messagers » aptes à être exportés vers le cytoplasme : ajout d'une coiffe en 5', épissage des introns, clivage à l'extrémité 3', ajout d'une queue de polyadénosines (queue poly[A]).

B. L'épissage : mécanismes et régulation

L'épissage permet l'excision des séquences non-codantes (introns) contenues dans l'ARN pré-messager et la jonction des séquences codantes (exons) qui persisteront dans l'ARN messager. Il concerne au moins 95 % des transcrits et se déroule en 2 étapes successives de trans-estérification[34]. Trois déterminants de séquence sont requis : les sites d'épissage en 5' (site donneur) et 3' (site accepteur) situés aux jonctions « exon-intron », respectivement en 5'et 3 ' de l'intron, et le site de branchement en amont du site 3' accepteur (Figure 3).

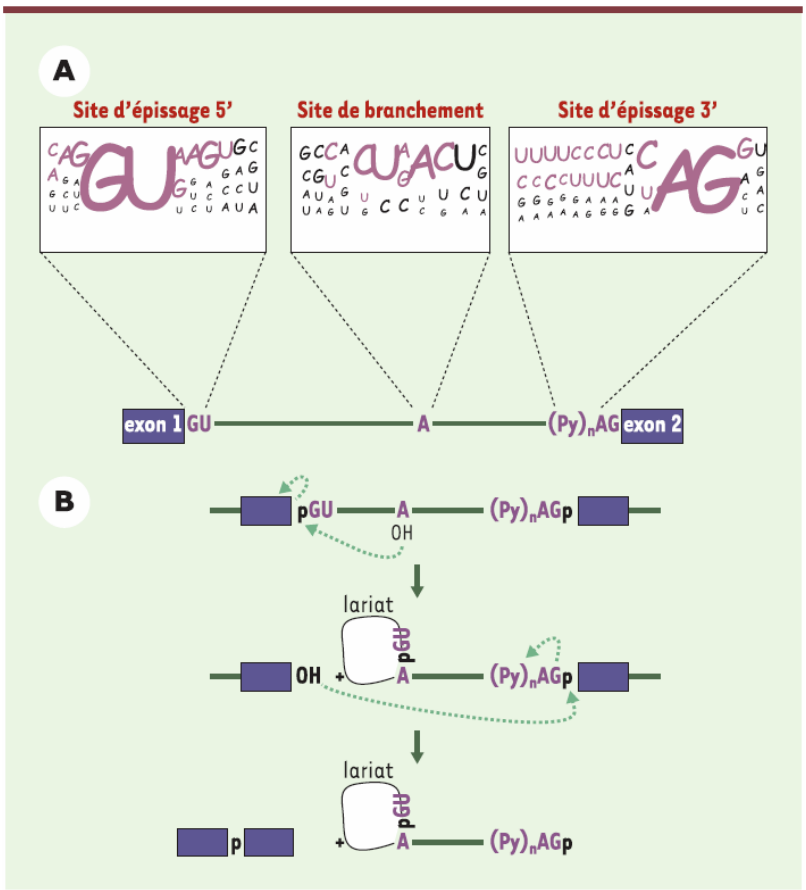


Figure 3 : Profil des séquences consensus au site dépissage 5', au site de branchement et au site d'épissage 3'.

A. À chaque position, la fréquence des 4 nucléotides A, C, G et T est représentée par la taille des caractères. La fréquence a été déterminée à partir de 1 683 séquences humaines[35] . Les séquences les plus conservées sont les dinucléotides GU et AG, respectivement en 5' et en 3' de l'intron, et le nucléotide A au niveau du site de branchement.
Les rectangles représentent les exons et les traits pleins les introns.
(Py)n : étendue de polypyrimidines.

B. Réaction d'excision-épissage d'un pré-ARNm. L'épissage d'ARN prémessager se déroule en deux étapes successives, impliquant chacune une réaction de trans-estérification. Trois éléments sont essentiels à ces réactions : les sites d'épissage 5' (ou site donneur) et 3' (ou site accepteur), situés à la jonction exon-intron, respectivement en 5' et 3' de l'intron, et le site de branchement, en amont du site 3' accepteur. Lors de la première étape, le groupement 2' hydroxyle de l'adénosine du site de branchement attaque le phosphate du site d'épissage 5'. Cette réaction engendre un exon 5' libre et un lariat intermédiaire. Lors de la deuxième étape, le 3' hydroxyle de l'exon 5' attaque le phosphate du site d'épissage 3', produisant ainsi l'ARNm et le lariat d'intron.

La réaction d'épissage a lieu au sein du spliceosome. C'est un complexe multiprotéique constitué de plus de 150 polypeptides et de cinq ribonucléoprotéines (small nuclear ribonucleoproteins), elles-mêmes constituées d'une molécule d'ARN (ARNsn) complexée à

plusieurs protéines. Il a pour fonction de reconnaitre les sites d'épissage puis d'effectuer les étapes couplées d'élimination des introns et re-ligation des exons.

Un autre répertoire de séquences en *cis* joue également un rôle déterminant lors de l'épissage, il s'agit des séquences « exonic splicing enhancers » (ESE), « exonic splicing silencers » (ESS), « intronic splicing enhancers » (ISE) et « intronic splicing silencers » (ISS). Ces séquences auxiliaires présentent toutes des caractéristiques communes : elles sont courtes (moins de 10 nucléotides), variables en terme de séquence, individuellement faibles et souvent présentes en plusieurs copies.

ESE et ESS sont des séquences exoniques qui favorisent (ESE) ou répriment (ESS) l'épissage de l'exon qui les contient et ce par interaction avec des facteurs de transcription spécifiques[36,37].

Les séquences introniques homologues peuvent se situer en amont, en aval ou de chaque côté de l'exon qu'elles régulent (Figure 4).

Qu'elles soient exoniques ou introniques, toutes ces séquences présentent des partenaires et des modes de fonctionnement assez proches. La régulation de l'épissage résulte d'une compétition entre les phénomènes activateurs et inhibiteurs dont une partie des mécanismes reste encore à préciser. La complexité de cette modulation prend tout son sens quand on aborde la notion d'épissage alternatif.

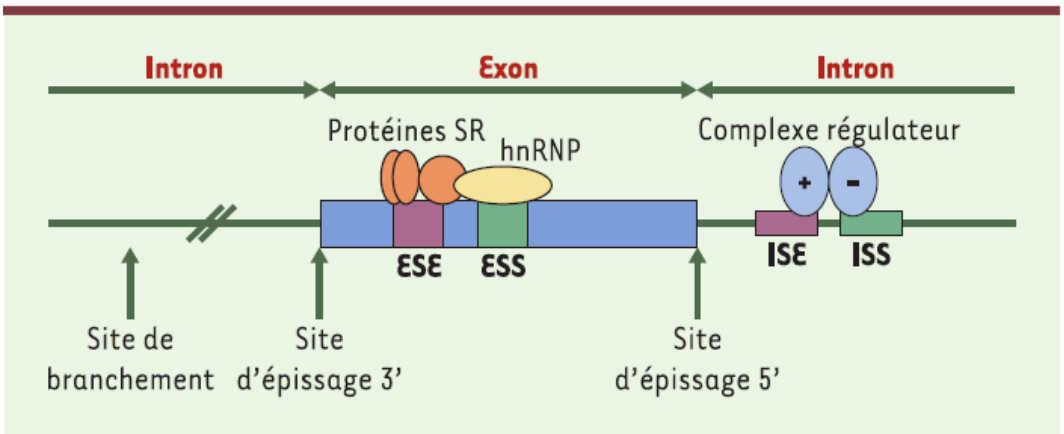


Figure 4 : La fonction des protéines SR (serine-rich) (représentées en orange) peut être réglée au niveau transcriptionnel, en fonction du tissu ou du stade de développement, par épissage alternatif, par phosphorylation-déphosphorylation, ce qui influence leur localisation subcellulaire et leur activité. Elles sont stockées dans des compartiments du noyau appelés speckles ou SFC (splicing factors compartments). Elles doivent être recrutées par le complexe d'épissage et dirigées à la périphérie des SFC vers les sites de transcription. Une fois engagées dans le splicéosome, elles sont les cibles de phosphatases Ser/Thr, PP1, PP2A, PP2Cγ, dont certaines font partie intégrante du splicéosome. Les protéines hnRNP (heterogeneous nuclear ribonucleoproteins) (représentées en jaune) figurent parmi

les protéines les plus abondantes du noyau ; certaines d'entre elles sont confinées dans le nucléoplasme (hnRNP C et U), tandis que d'autres (hnRNP A, B, K, J) font la navette entre le noyau et le cytoplasme. L'ESE (exonic splicing enhancer) (représenté en violet au niveau de l'exon) est le siège de nombreuses mutations qui modifient le site de liaison de protéines SR et conduisent à l'exclusion d'exon. Des algorithmes de prédiction d'ESE fonctionnels sont disponibles en ligne (exemple : http://exon.cshl.edu/ESE/index.html). L'ESS (exonic splicing silencer) (représenté en vert au niveau de l'exon) agit, entres autres, par liaison à hnRNP I (ou PTB) et hnRNP A/B ou H. HnRNP A1 peut également induire la répression par liaison à un site exonique et par le recrutement coopératif d'autres hnRNP, ce qui neutralise l'activité d'enhancers proximaux, voire l'assemblage du splicéosome. L'ISE (intronic splicing enhancer) et l'ISS (intronic splicing silencer) (représentés respectivement en violet ou en vert au niveau de l'intron) auraient des rôles antagonistes, lesquels peuvent s'exercer au sein d'éléments introniques bipartites[34].

C. L'épissage alternatif

L'épissage alternatif est un mécanisme naturel permettant la synthèse d'ARNm différents à partir d'un même pré-ARNm (et donc à partir d'un même gène) en utilisant différentes combinaisons d'exons et en générant ainsi des isoformes protéiques variées.

Certains exons sont toujours inclus dans les ARNm matures, il s'agit des exons « constitutifs ». D'autres sont régulés et peuvent parfois être supprimés ou « sautés » ou partiellement inclus (Figure 5).

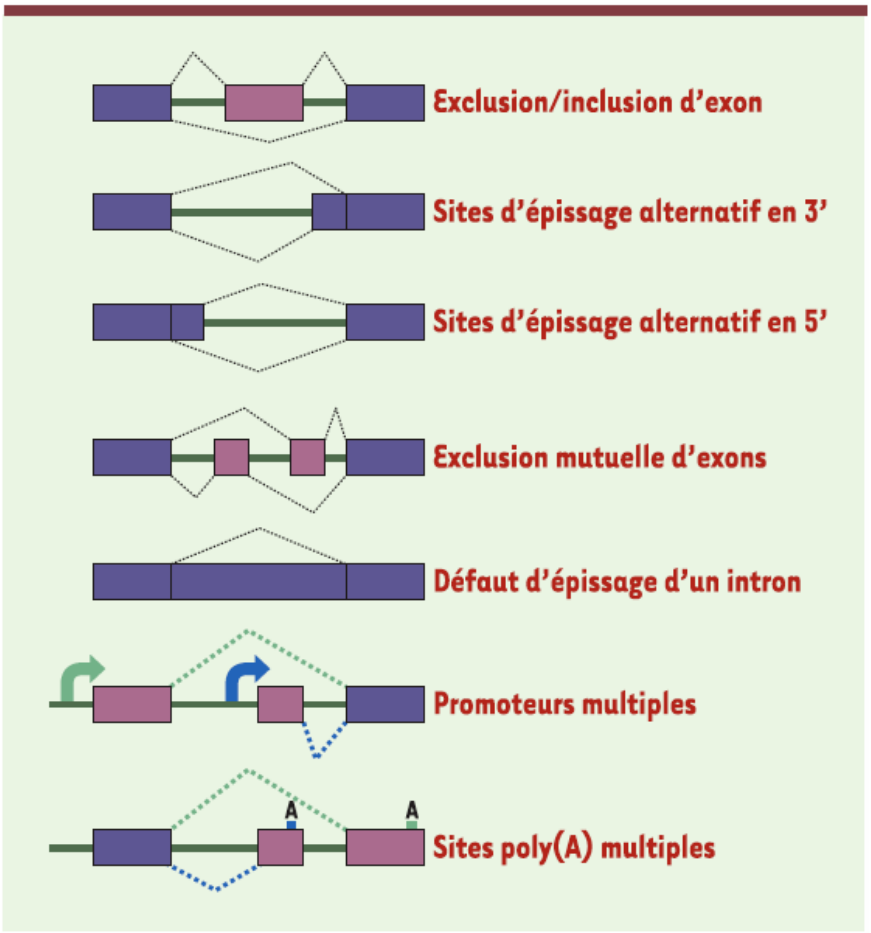


Figure 5 : En bleu les exons constitutifs, en violet les exons alternatifs, et en traits pleins les introns. Différentes combinaisons d'assemblage des exons, voire de rétention d'intron sont possibles. En outre, l'utilisation de promoteurs transcriptionnels peut dicter les choix d'épissage. Enfin, l'épissage alternatif peut conduire à l'utilisation de signaux de sites de polyadénylation distincts[3].

Des mécanismes permettent aussi d'allonger ou raccourcir certains exons en utilisant des sites d'épissages alternatifs (différents des sites constitutifs) permettant l'inclusion de séquences introniques plus ou moins importantes dans l'ARNm.

Les sites d'épissage alternatifs sont généralement moins forts que les sites constitutifs (c'est-à-dire plus éloignés par rapport à la séquence consensus) et plus assujettis à leur régulation[38]. Il semblerait que les exons alternatifs soient souvent entourés de régions introniques plus longues et beaucoup plus conservées que les exons constitutifs[39]. Les exons alternativement épissés sont également plus courts, présentent un degré de conservation plus important entre les exons orthologues et leur longueur est plus souvent un multiple de 3 (conservation du cadre de lecture)[40]. Les introns pouvant être éventuellement inclus dans l'ARNm final sont quant à eux significativement plus courts ques les autres classes d'introns[41] et possèdent eux

aussi des sites d'épissage plus faibles. La prédiction des évènements d'épissage alternatif basée sur l'identification de sites d'épissage et d'éléments de régulation potentiels reste un enjeu important et différentes approches basées sur des données expérimentales et des outils informatiques ont été développées[42](Figure 6).

On estime que 1000 à 20 000 gènes humains sont sujets à un épissage alternatif[34].Des études ont montré que seulement 19% des épissages alternatifs conduiraient à une protéine tronquée[37], la grande majorité (70% à 88%) changeant uniquement la nature de la protéine. Cette observation résulte en grande partie du fait que l'épissage alternatif se trouve souvent associé à l'apparition d'un codon stop prématuré pouvant certes être à l'origine de production de protéines tronquées, mais surtout du mécanisme de nonsense-mediated mRNA decay (NMD).

Sites donneurs et accepteurs
Splice Site Prediction http://www.fruitfly.org/seq_tools/splice.html
MaxEntScan http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html
GeneSplicer http://www.tigr.org/tdb/GeneSplicer/gene_spl.html

ESE
ESE finder http://exon.cshl.org/ESE/index.html
Rescue ESE http://genes.mit.edu/burgelab/rescue-ese/

Figure 6 : Les différents sites d'épissage recherchés ont des séquences consensus relativement conservées, ce qui autorise leur identification par différents algorithmes. Ces algorithmes permettent également d'évaluer quantitativement l'impact de modifications nucléotidiques au sein de ces séquences car un score est attribué pour chaque base, pour chaque position, par comparaison avec l'hypothèse la plus probable.
Concernant la liste présentée (non exhaustive) d'« outils web » de prédiction, il est possible d'interroger simultanément ces différents sites, ce qui simplifie grandement la procédure (données disponibles sur demande). Il existe d'autres logiciels dédiés à l'identification des séquences réglant la transcription[3].
Le logiciel Alamut qui sera utilisé par la suite propose une fenêtre dédiée aux prédictions d'épissage regroupant différents scores en une seule et même vue[43].

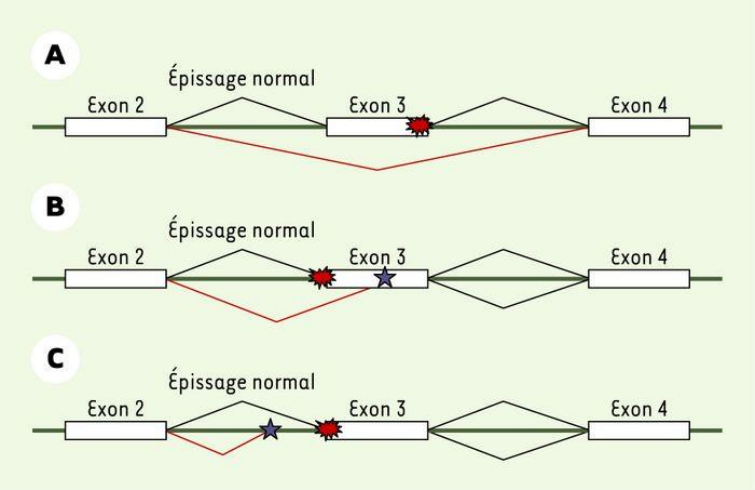
D. Epissage alternatif et pathologie

Les altérations des sites génomiques consensus et leurs conséquences sur l'épissage sont désormais bien connues. Grandes délétions, insertions ou délétions d'une ou plusieurs bases et modifications nucléotidiques sont à même d'altérer la fonction de ces sites avec des conséquences diverses (Figure 7) :

- Abolition du site d'épissage physiologique avec saut de l'exon concerné (A)
- Abolition du site d'épissage physiologique avec révélation d'un site cryptique qui prend le relais du site sauvage. En fonction de la localisation du site cryptique il s'en suit une délétion exonique (B) ou une rétention intronique (C).

Ces mécanismes peuvent survenir de manière isolée ou s'ajouter l'un à l'autre.

Les modifications nucléotidiques peuvent aussi ne pas toucher ces séquences consensus mais néanmoins altérer l'épissage en créant un site cryptique plus fort qui sera préférentiellement utilisé par rapport au site sauvage.



-

Figure 7 : Les exons sont représentés par des rectangles, les introns par un trait vert. La position de la mutation est indiquée par un symbole rouge ; les traits rouges schématisent l'anomalie de la transcription[3].

A. Altération du site donneur de l'exon 3, saut de l'exon 3.
B. Abolition du site accepteur de l'exon 3, avec révélation d'un site cryptique exonique (étoile bleue) : délétion exonique partielle.
C. Abolition du site accepteur de l'exon 3, avec révélation d'un site cryptique intronique (étoile bleue) : rétention intronique.

E. Nonsense mediated mRNA decay : le NMD

Parmi les dispositifs de contrôles de qualité établis tout au long du processus de synthèse des protéines, la dégradation de l'ARNm après un codon stop prématuré ou anormal (NMD) constitue une réaction essentielle dans la régulation génétique. Ce mécanisme permet l'élimination d'ARNm anormaux qui engendreraient des protéines aux effets délétères de type gain de fonction ou dominant négatif s'ils évoluaient jusqu'à l'étape de traduction. Ce

mécanisme est très souvent observé si le codon stop est situé à plus de 50-55 nucléotides en amont de la dernière jonction exon-exon.

Il existe cependant des cas où il n'est pas déclenché : en cas de codon stop anormal proche du codon stop normal (à une distance inférieure à 50-54 pb de ce dernier)[44] ou à proximité du codon d'initiation de la traduction (AUG).

Le déclenchement du NMD est complexe et il apparaitrait désormais que son rôle ne se limiterait pas uniquement à celui de gardien de l'intégrité des ARNm, mais qu'il aurait également une importance dans la régulation d'expression de certains gènes et la dégradation de 5 à 10% d'ARNm normaux[45].

L'effet probablement bénéfique du NMD *in vivo* pose cependant des problèmes majeurs en diagnostic, l'anomalie étant masquée *in vitro*. Il a ainsi longtemps été un obstacle majeur à la stratégie diagnostique basée sur l'étude de l'ARN, mais son effet est prévenu par l'ajout de puromycine aux prélèvements d'intérêt[46].

F. Les transcrits alternatifs de *BRCA1* et *BRCA2*

De nombreux transcrits alternatifs ont été découverts dans différents tissus pour ces deux gènes mais leurs régulations et leurs rôles exacts restent cependant peu connus[47]. Il est très probable que bon nombre d'entre eux ne soient pas fonctionnels au niveau physiologique (car décrits à partir de tissus tumoraux altérés) mais quelques variants prédominants semblent quand même ressortir pour ces deux gènes. Etant données les difficultés méthodologiques à trier pertinemment toutes ces données, nous avons retenuspour notre étude les principaux variants sélectionnés par le « groupe épissage » (données non publiées, ne concernant pas l'exon 11 d'étude délicate)

1. Principaux transcrits alternatifs observés pour *BRCA1*

- Δ exon 5
- Δ 22pb de l'exon 5
- Δ 3 premiers nucléotides de l'exon 8
- Δ exon 9-10
- Δ 3 premiers nucléotides de l'exon 13

- Δ 3 premiers nucléotides de l'exon 14

Il semblerait tout de même d'après d'autres études qu'il existe d'autres transcrits à prendre en compte, mais ceux-ci restent à préciser (Figure 8)[47].

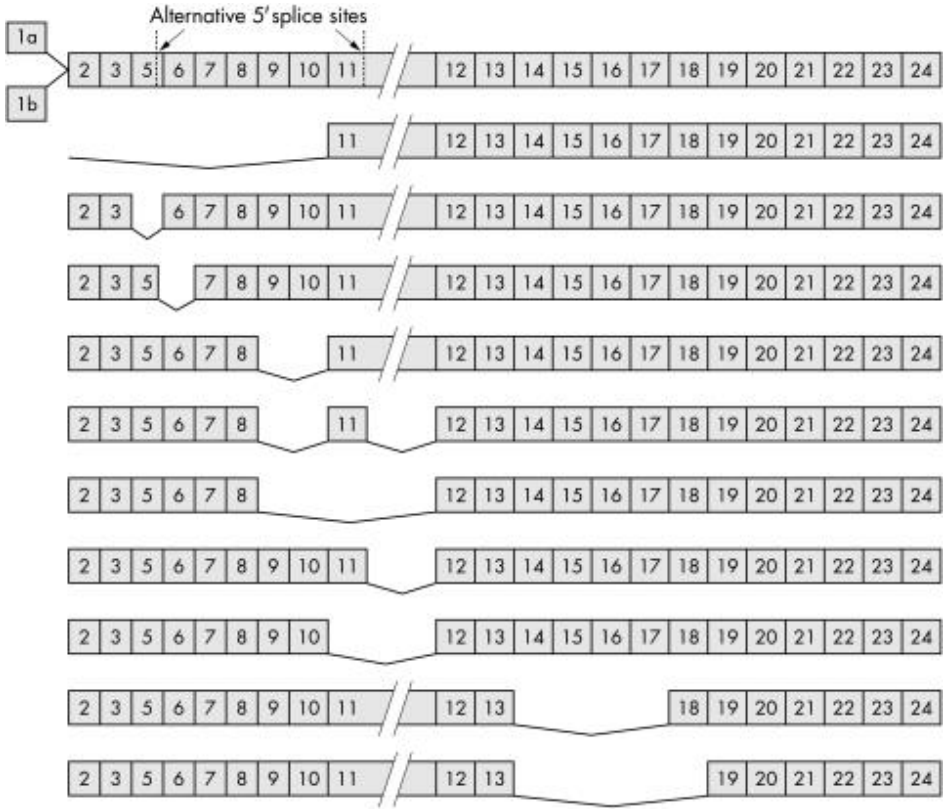


Figure 8 : Structures des différents variants d'épissage des ARNm de BRCA1 conservant le cadre de lecture et capables d'aboutir à une synthèse protéique[47].

2. Principaux transcrits alternatifs observés pour *BRCA2*

- Δ exon 3
- Δ exon 6-7
- Δ exon 12
- Δ exon 17-18

V. Next-Generation of Sequencing (NGS) : les technologies de séquençage à haut-débit

La méthode automatisée de séquençage Sanger est considérée comme une technologie de « 1ère génération ». Cette méthode a été la référence pendant une vingtaine d'années et a permis l'accomplissement de bon nombre de découvertes majeures (comme celle du 1er séquençage complet d'un génome humain en 2004[48]). Le séquençage Sanger est actuellement dépassé par la nouvelle génération de séquenceurs permettant du séquençage à très grande échelle et de manière beaucoup plus rapide avec un coût à la base beaucoup plus faible. Ainsi le 1er séquençage complet d'un génome humain a été réalisé en 15 ans alors qu'il peut désormais l'être en 10 jours sur une plateforme Illumina HiSeq 2000.

A. Principales technologies de NGS

Les principales technologies de séquençage à haut débit actuellement disponibles sont celles proposées par Roche 454 Life Sciences, Illumina et Ion Torrent. Elles utilisent la détection de signaux chimiques ou lumineux lors de l'incorporation de bases nucléotidiques pendant la synthèse du brin complémentaire pour déterminer la séquence. De manière générale, l'ADN est fragmenté à la taille souhaitée pour son étude, chaque fragment est ensuite lié à des adaptateurs et cloné pour amplifier le signal de fluorescence ou chimique. Les trois méthodes différent par les réactifs, les substrats utilisés et par la technologie de détection des signaux d'incorporation des différentes bases[49](Figure 9).

1. Le pyroséquençage

Dans cette technologie mise au point par Roche 454 Life Sciences, une seule séquence d'intérêt est couplée à une microbille *via* un adaptateur et amplifiée grâce à une étape de PCR émulsion. Chaque microbille est placée dans un micro-puit avec des billes supplémentaires couplées à la sulfurylase et la luciférase. Une amorce est hybridée à l'ADN puis chacune des bases est incorporée dans un ordre prédéfini à partir de l'extrémité 3'. Un signal est mesuré uniquement lorsque la base est intégrée à la séquence cible par la polymérase et qu'un pyrophosphate est libéré. L'ATP sulfurylase vient alors transformer ce pyrophosphate libre en

ATP qui est ensuite utilisé (d'où le terme pyroséquençage) et couplé à une luciférine par une luciférase. Cette réaction libère une oxyluciférine et un signal lumineux. La caméra « Charge Coupled Device » (CCD) permet de capturer les signaux émis et de déduire la séquence en fonction de l'ordre d'incorporation des bases et de l'intensité lumineuse sur le chromogramme.

Cette technologie a l'avantage de produire de longues séquences (330 bases en moyenne), qui facilitent l'assemblage *de novo* de génome sans utiliser de séquence de référence.

2. Le séquençage par synthèse

Initialement développée par Solexa, il s'agit aujourd'hui de la technique utilisée par la plateforme Illumina Genome Analyzer. Cette technologie utilise une phase solide d'amplification où les adaptateurs sont liés à chaque extrémité de la séquence cible. Celle-ci est ensuite amplifiée par multiplications clonales des séquences (et la formation de cluster grâce à des ponts d'amplification). Les quatre types de bases sont ajoutés successivement et celles non-hybridées sont éliminées. Ce type de séquençage (contrairement au précédent) permet d'incorporer un nucléotide à la fois et le cycle est ainsi répété jusqu'à la taille attendue du fragment. Aujourd'hui le système HiSeq 2000 possède un rendement de 6 milliards de séquences pairées pour un total d'environ 600 Gb en 11 jours (avec un taille moyenne de séquence de 100 pb). Cette technologie est aujourd'hui très largement employée (séquençage, re-séquençage, Chip-Seq, RNA-Seq).

3. Le séquençage par ligation

Cette technologie utilise une ligase et des amorces marquées afin de déterminer l'ordre des nucléotides d'un brin d'ADN. Des amorces de différentes longueurs contenant un dinucléotide connu sont utilisées conjointement avec de l'ADN ligase. Ces amorces sont identifiées avec des sondes fluorescentes et celles demeurant non hybridées sont éliminées. S'ensuit alors une détection par fluorescence des amorces liguées à la séquence cible. Ce cycle peut se répéter en utilisant des amorces clivées permettant à la sonde fluorescente de se détacher et de régénérer un groupe phosphate en 5' de l'amorce. Le processus de ligation, détection et clivage se répète ainsi plusieurs fois jusqu'à l'obtention d'une longueur prédéterminée. La méthode de ligation du séquenceur SOLiD (Life Technologies/Applied Biosystems) produisant

des fragments courts (autour de 75 pb) a surtout été utilisée dans des projets de reséquençage de génome et de transcriptome. Aujourd'hui cependant les coûts et le débit du SOLID ne peuvent rivaliser avec ceux de la technologie Illumina et la part de marché de cette machine tend à diminuer sérieusement.

4. Le séquençage par nano-mesure de pH

Le système Ion Torrent est unique puisqu'il utilise des variations de pH afin de détecter l'incorporation successive des nucléotides. Il détecte la libération des protons à chaque intégration d'un nucléotide au cours de l'élongation. En incorporant séquentiellement chaque nucléotide, cette méthode permet de détecter ceux qui ont été incorporés au brin complémentaire.

Le Proton est un séquenceur haut débit basé sur cette technologie des semi-conducteurs (similaire à celle du Ion – Torrent –PGM). Il permet de répondre à des demandes très variées de séquençage et de génotypage par séquençage avec une montée en puissance rapide : si la puce PI délivre déjà 10 Gb et 70 millions de « reads » de 200 bases (idéal, par exemple, pour le séquençage de petits génomes), la puce PII permet de réaliser plus de 280 de millions de « reads » soit plus de 70 Gb (de quoi séquencer un génome humain à 20 X). Celle-ci devrait permettre le séquençage *de novo* sur de plus gros génomes et faciliter les analyses de transcriptomes. De plus, l'analyse des données est facilitée grâce à une interface web reliée à un serveur autonome. Celle-ci permet de lancer de façon intuitive et autonome les analyses les plus classiques (« mapping » sur référence, détection des SNP…..). Malheureusement, la technique ne semble actuellement pas encore assez au point pour un transfert en utilisation diagnostique.

B. Les séquenceurs dits « de paillasse »

Il s'agit de séquenceurs de 2ème génération de débit plus faible et de cout d'utilisation moins important par rapport aux précédents. Ils ont plutôt pour vocation de séquencer un grand nombre d'échantillons sur des régions de taille plus réduites et permettent en quelque sorte la « démocratisation » du séquençage à haut-débit. Les trois leaders du marché proposent leur appareil :

- Life Technologie : Ion Torrent PGM et Proton
- Illumina : MiSeq
- Roche : GS Junior

Le « Personal Genome Machine » ou PGM peut générer 1 Gb de séquences via l'utilisation de puces 318 avec une taille moyenne de fragments séquencés d'environ 250 pb en moins de 2 heures.

L'Ion Torrent PGM 318 s'avère être compétitif dans cette catégorie, le MiSeq produisant des séquences plus courtes (environ 150 pb) et le GS Junior affichant un débit plus faible (35 Mb).

C. Les méthodes de 3ème génération

Ces nouvelles méthodes utilisent pour la plupart la détection d'un signal chimio luminescent à chaque incorporation de nucléotide afin de séquencer directement les fragments d'ADN un à un ce qui rend l'amplification de l'ADN inutile. Ces innovations diminueraient la durée et simplifieraient considérablement les étapes de préparation des banques d'ADN. De plus, l'élimination de ces étapes permettrait aussi de diminuer les sources de biais possibles pouvant être introduits par les phases de PCR (erreurs de polymérisation ou recombinaisons par PCR). Un autre avantage de ce techniques réside dans le fait que des ADN faiblement concentrés voir même dégradés pourraient ainsi tout de même devenir exploitables[50].

Par exemple, le système HeliScope (Helios Genetic Analysis) produit de 600 millions à 1 milliard de séquences d'une longueur de 35 pb en 30 heures. Une autre technologie a été récemment développée par les laboratoires Pacific Biosciences (PacBio RS SMS). Cette méthode permet l'incorporation continue de nucléotides marqués pendant la synthèse du brin d'ADN. L'ADN polymérase est attachée à la surface d'un détecteur ZMW (Zero Mode Waveguide) qui assimile l'ordre des nucléotides phosphatés quand ceux-ci sont inclus dans le brin généré. Cette technologie permet d'atteindre des longueurs de séquence de 10000 pb mais la polymérase peut parfois se dégrader avant d'atteindre cette taille à cause des lectures au laser (contrairement à l'HeliScope). La plateforme PacBio possède le taux d'erreur le plus élevé de toutes les plateformes NGS (environ 15%) mais ceci est contrebalancé par le fait qu'un même fragment d'ADN est séquencé plusieurs fois.

Ainsi, les dernières versions du System PacBio RS produit 35 à 45 Mb de données par cellule d'une longueur moyenne de 1500 pb avec une sensibilité de 99,999% à une couverture de 30X[51].

La technologie de séquençage de 3ème génération par nanopores promet de révolutionner plusieurs applications à commencer par le séquençage *de novo* en laissant peu d'espoir à la technologie de Pacific Biosciences. Elle devrait remplacer PacBio dans les stratégies de séquençage hybride (qui consiste « à coupler » un séquenceur de 2ème génération permettant d'être très profond et d'une technologie de 3ème génération permettant de générer des reads très longs ce qui permet au final un assemblage de meilleure qualité).

Après avoir engendré de la curiosité, de l'impatience, puis déçu avec une arrivée sans cesse repoussée, il semble qu'Oxford Nanopore doive prouver l'efficacité de sa technologie.

Le principe de cette technologie est de permettre de séquencer sans pré-amplification (il s'agit d'un séquenceur de 3ème génération) une molécule d'ADN en en hydrolysant les nucléotides élémentaires, qui passeront dans des pores. Ces pores seront couplés à des capteurs saisissant la signature singulière de chaque base constitutive de la molécule à séquencer. Jusqu'à présent. La difficulté technique majeure empêchant le développement et la diffusion de cette technique réside encore dans l'incapacité à contrôler la vitesse à laquelle diffus chaque brin d'ADN à travers le nanopore.

Ces techniques sont encore à l'étape de mise au point à l'heure actuelle et très peu de laboratoires en sont équipés pour le moment.

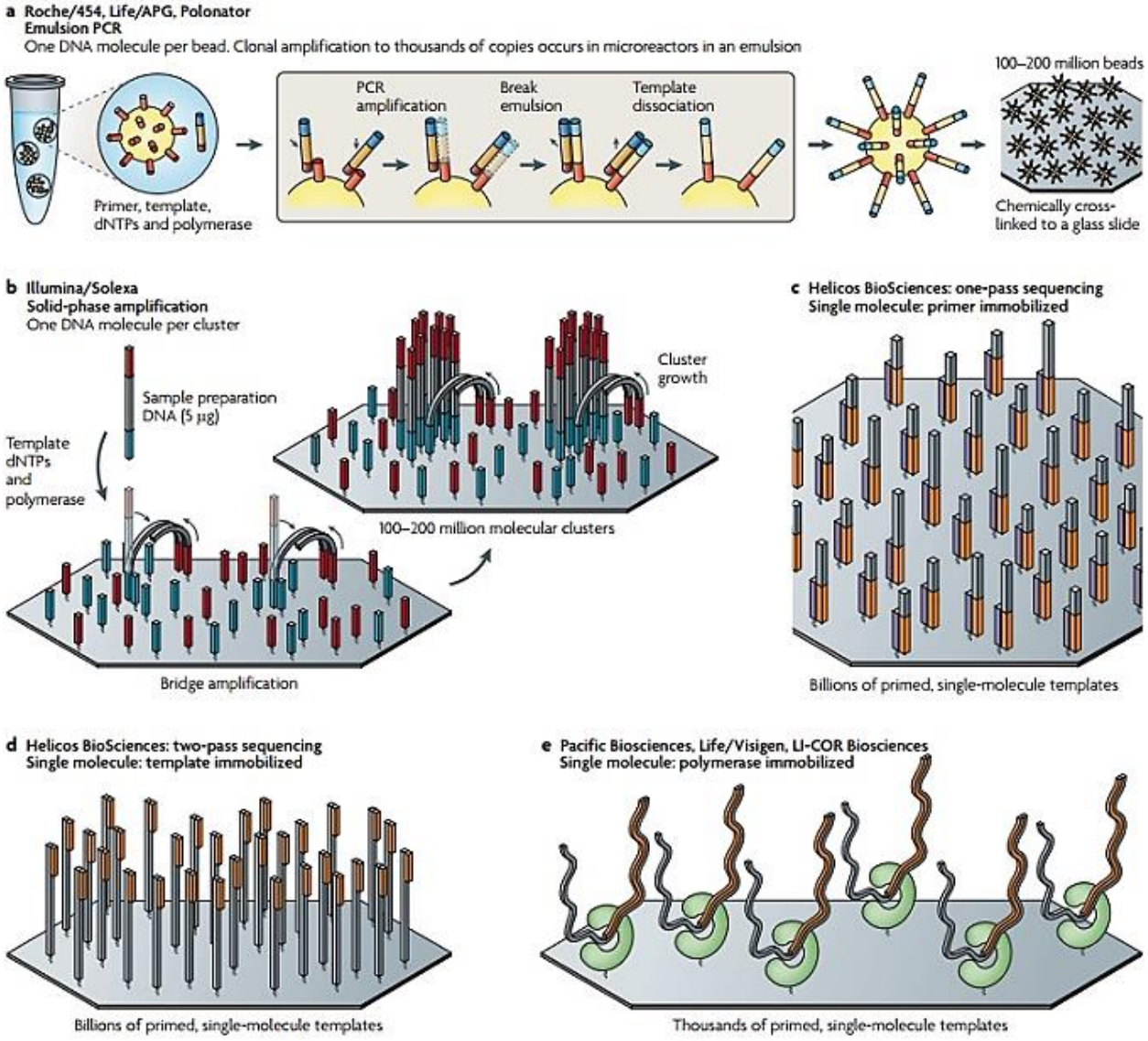


Figure 9 : Présentation des différentes stratégies de fixation des échantillons à séquencer[49].

a) Dans le cas de l'émulsion PCR (PCRe), un mélange réactionnel spécifique est créé de manière à obtenir une émulsion mêlant eau et huile où des micro-goutelettes aqueuse contiennent une séquence d'ADN et une bille. L'amplification PCR a lieu au sein de ces unités réactionnelles de manière à obtenir des billes présentant plusieurs milliers de copies d'une même séquence d'ADN à leur surface. Les billes peuvent ensuite être chimiquement attachées sur une lame en verre ou déposées sur une puce à micro-puits.

b) La phase d'amplification de la technologie Solid et composée de deux étapes : une première phase où mes fragments à séquencer sous forme simple brin sont immobilisés sur une plaque de verre à l'aide d'amorces s'hybridant avec un des adaptateurs, et une deuxième étape d'amplification des fragments qui se fait par pontage à l'aide de l'ADN polymérase. Cela permet l'amplification localisée d'un fragment unique en plusieurs millions de fragments identiques.

c) 3 approches différentes permettant la fixation d'un ADN simple brin sont représentées : fixation par un primer(d), fixation par une séquence (e) et fixation via une polymérase (f)

D. Analyse Bio-informatique

La quantité de data générée par ces nouvelles techniques a rendu l'analyse des résultats manuelle impossible et a nécessité le développement d'outils bio-informatiques dédiés. En plus des pipelines d'analyse académiques initialement construits se sont développés progressivement des logiciels « clés en main ». Les étapes de traitements informatiques des données sont incontournables à l'évaluation de la qualité des séquences produites, à leur alignement génomique et aux filtrages des « anomalies » ainsi détectées (Tableau 1).

Le traitement des données NGS est difficile sur une station locale. Il nécessite des ressources logistiques importantes aussi bien en capacité de calculs, en disponibilités de logiciels spécialisés (la lecture des fichiers est impossible avec les logiciels de bureautique classiques) qu'en capacité de stockage.

Définitions	
Lecture ou *read*	Terme générique désignant les séquences sortant du séquenceur
Mismatch	Incohérence entre deux nucléotides
Gap	Trou dans l'alignement de séquence
Mappabilité	Caractère d'une région génomique à être couverte en lectures (région répétées = faible mappabilité, région unique = forte mappabilité)
Indel	Insertion/délétion sur un génome de référence
Génome de référence	Séquence connue, supposée être la plus proche possible des données étudiées, utilisée comme pour ancrer et organiser l'information des lectures simples
Couverture	Fraction de la référence couverte par au moins une lecture.
Profondeur de lecture	Nombre moyen de lectures par base à une position donnée (elle est exprimée en nombre de X)
Format de fichier .FASTA	Fichier sous format texte composée de deux parties : il commence avec une ligne simple de description de la séquence suivie de lignes correspondant à la séquence elle-même
Format de fichier .FASTQ	Fichier texte combinant les séquences et les scores de qualité des bases.
Format de fichier .SAM	Fichier sous format texte tabulé regroupant les data d'alignements de séquences
Format de fichier .BAM	Version binaire d'un fichier .SAM, format conseillé pour le partage des fichiers (fichier de référence)

Tableau 1 : principales définitions utiles en bio-informatiques

L'analyse bio-informatique est un processus complexe se découpant en plusieurs étapes (Figure 10).

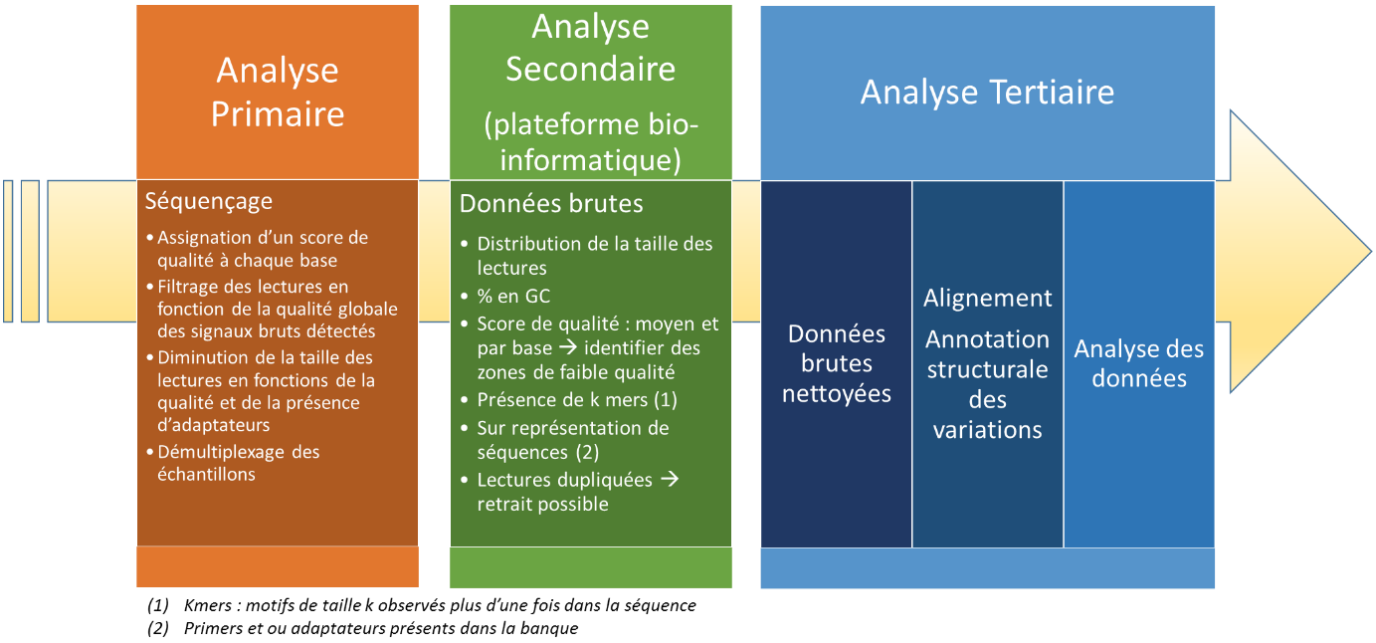


Figure 10 : Principales étapes de l'analyse bio-informatique

1. L'alignement

L'alignement est une étape cruciale dans l'application du NGS ; il a pour but de transformer une information simple (lecture indépendante) en une information globale, structurée et réduite.

Deux stratégies d'alignement sont envisageables (et peuvent très bien être utilisées de manière combinée) :

- **Alignement *de novo*** (Figure 11) : appliqué quand le génome n'est pas connu, il permet à l'aide de calculs intensifs de reconstruire un génome par recoupement des régions chevauchantes des différents *read*.

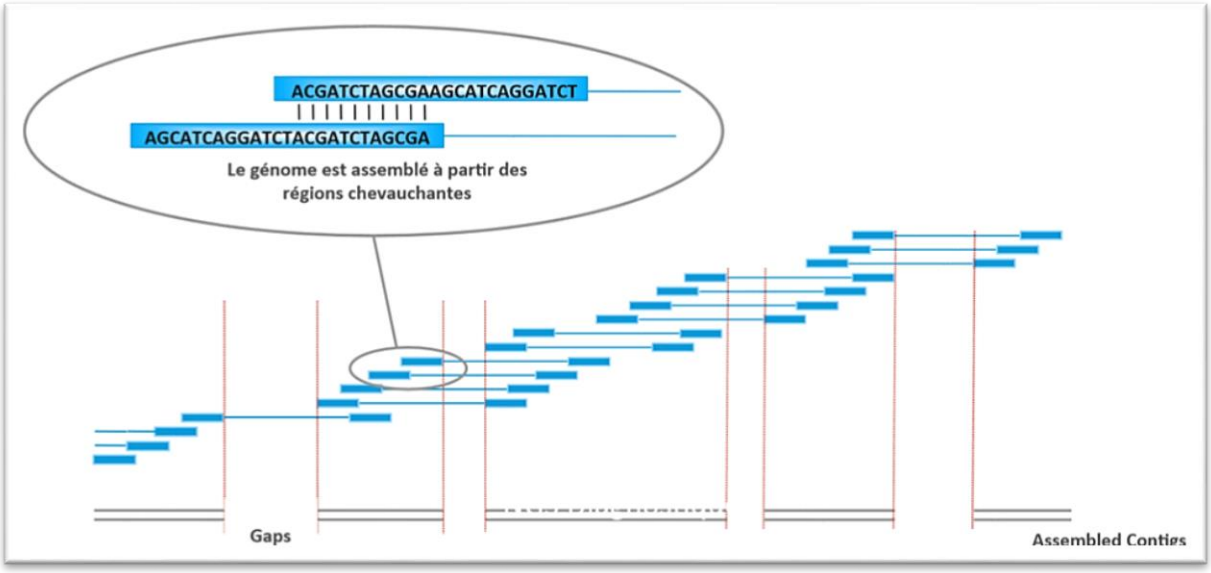


Figure 11 : Alignement de novo (Présentation Power Point de l'équipe de Bio-informatique de l'Institut Curie, janvier 2013, Nicolas Servant, Séverine Lair)

- **Alignement sur un génome de référence** : les lectures sont directement analysées et alignées sur un génome de référence connu

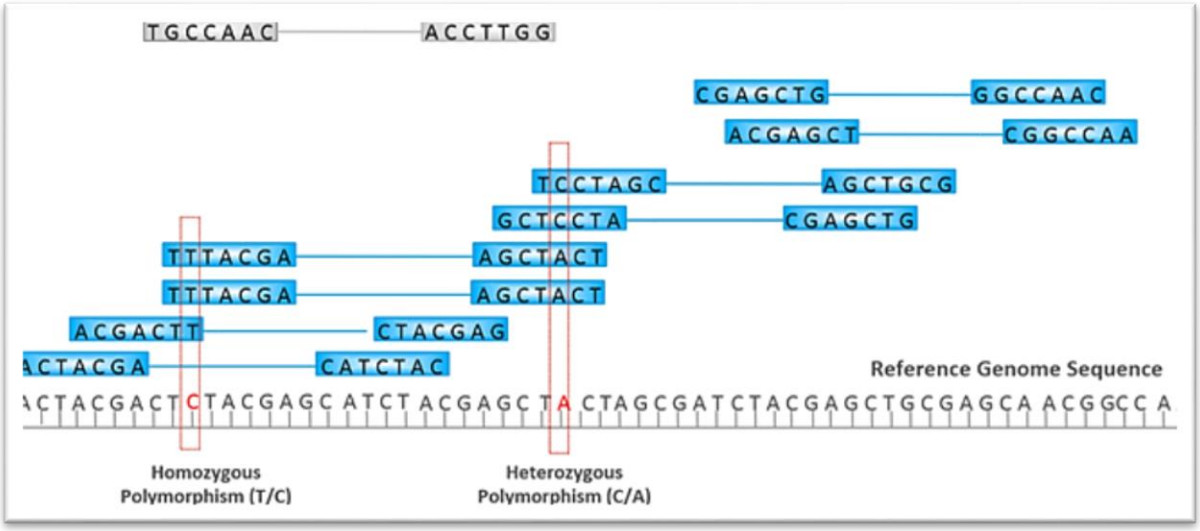


Figure 12 : Alignement sur un génome de référence (Présentation Power Point de l'équipe de Bio-informatique de l'Institut Curie, janvier 2013, Nicolas Servant, Séverine Lair)

Les nouveaux algorithmes d'analyse doivent prendre en compte les spécificités et caractéristiques des données actuellement produites :

- lectures (ou *read*) produites en très grande quantité (millions) et de tailles différentes (de 35 pb à 500 pb)

- prise en compte des qualités de lectures, des erreurs de séquençage (environ 1%), des régions génomiques répétées et l'évolution rapide des technologies et des formats

Au fil du temps, de nombreux outils informatiques d'alignement de plus en plus performants ont vu le jour. Il persiste malgré tout quelques voies d'amélioration (gestion des homopolymères, des petites *indel*, des régions répétées…)

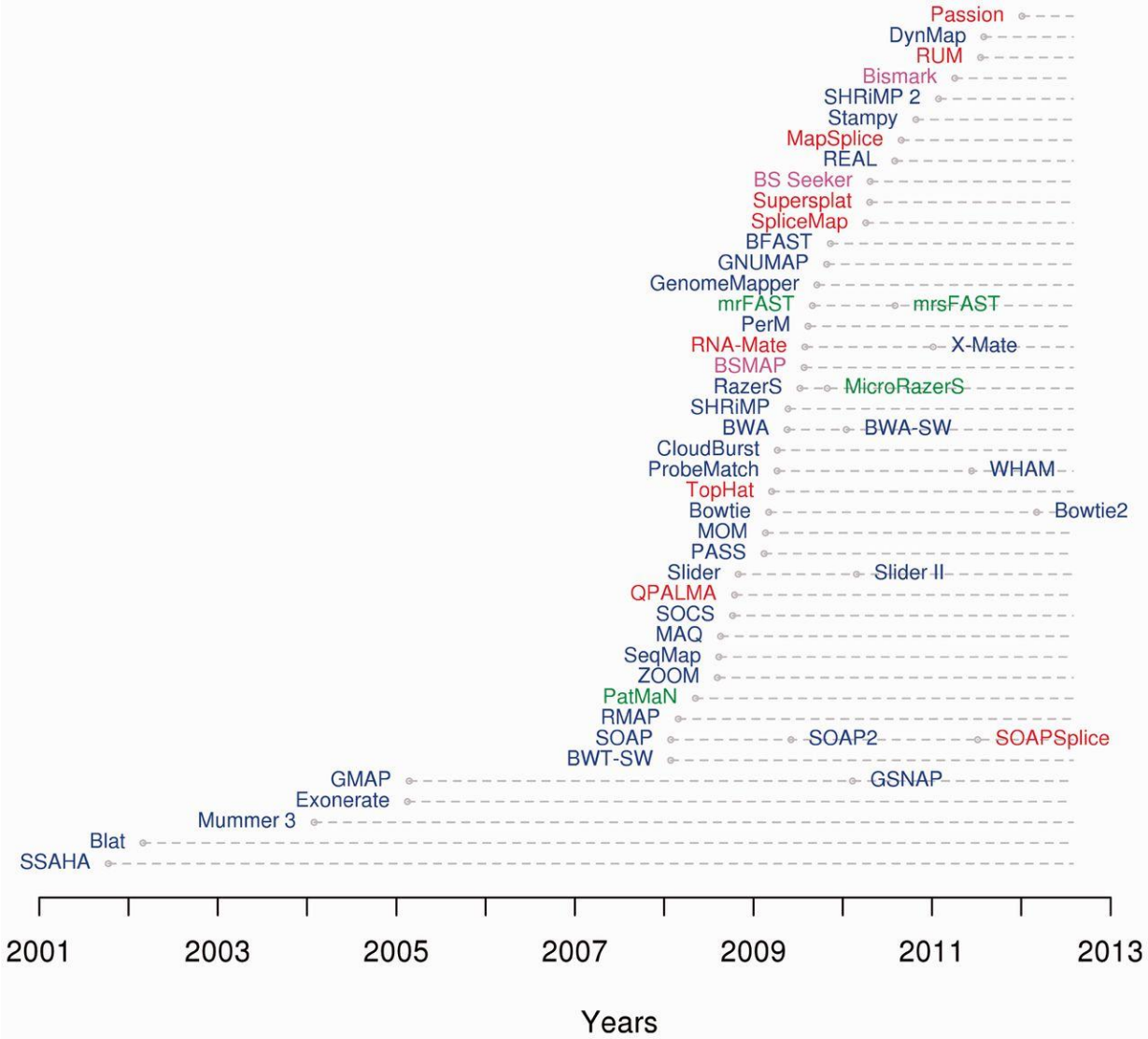


Figure 13 : Chronologie d'apparition des outils d'alignements (depuis 2001)[52].

- *En bleu : Outils d'alignement d'ADN*
- *En rouge : Outils d'alignement d'ARN*
- *En vert : Outils d'alignement de micro ARN (miARN)*
- *En violet : Outils d'alignement de séquence bisulfite (étude des profils de méthylation)*
- *Les pointillés gris signalent les liens existants entre certains logiciels (nouvelles versions ou extensions).*

Le panel de logiciels proposés est donc très large et en constant développement pour répondre de mieux en mieux aux besoins du NGS.

2. Détection des variants

Le but de cette étape est de détecter de manière efficace les anomalies au niveau des régions codantes des gènes d'intérêt pouvant être responsables de la pathologie. Plusieurs logiciels (académiques ou commerciaux) sont disponibles à cet effet (Figure 14).

Il s'agit d'identifier les différences nucléotidiques par rapport au génome humain de référence tout en répondant aux exigences du diagnostic :

- Couverture suffisante à toutes les positions interrogées
- Sensibilité proche de 100%
- Peu de faux positifs malgré tout (spécificité) même si la sensibilité prévaut dans ce contexte

Software	Available from	Calling method	Prerequisites
SOAP2	http://soap.genomics.org.cn/index.html	Single-sample	High-quality variant database (for example, dbSNP)
realSFS	http://128.32.118.212/thorfinn/realSFS/	Single-sample	Aligned reads
Samtools	http://samtools.sourceforge.net/	Multi-sample	Aligned reads
GATK	http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit	Multi-sample	Aligned reads
Beagle	http://faculty.washington.edu/browning/beagle/beagle.html	Multi-sample LD	Candidate SNPs, genotype likelihoods
IMPUTE2	http://mathgen.stats.ox.ac.uk/impute/impute_v2.html	Multi-sample LD	Candidate SNPs, genotype likelihoods
QCall	ftp://ftp.sanger.ac.uk/pub/rd/QCALL	Multi-sample LD	'Feasible' genealogies at a dense set of loci, genotype likelihoods
MaCH	http://genome.sph.umich.edu/wiki/Thunder	Multi-sample LD	Genotype likelihoods

Figure 14 : Liste non exhaustive d'outils d'annotation des variants disponibles[53].
Une liste plus complète est disponible à l'adresse suivante : http://seqanswers/wiki/Software/list[54]

3. Applications dans la détection des transcrits

La profondeur de lecture du NGS (30 à > 1000 *read* pour chaque base) permet de séquencer un panel très large d'ARN en une seule et même fois (y compris les transcrits rares). Par contre, les *read* obtenus sont souvent de très petite taille (de 35 à 330 pb selon la technologie de séquençage utilisée). L'enjeu de la bio-informatique dans ce contexte est de permettre la reconstruction d'un transcrit complet à partir de ces petits fragments tout en parvenant à distinguer les transcrits de différentes tailles (Figure 15). Le problème est que les différents transcrits présents partagent la plupart du temps les même exons et qu'il est difficile de savoir à quel transcrit précis appartient un *read* donné[55].

La précision de l'évaluation du niveau d'expression du gène est influencée par les méthodes de construction des librairies de ADNc, les technologies de séquençage et les techniques de pré-traitement des données.

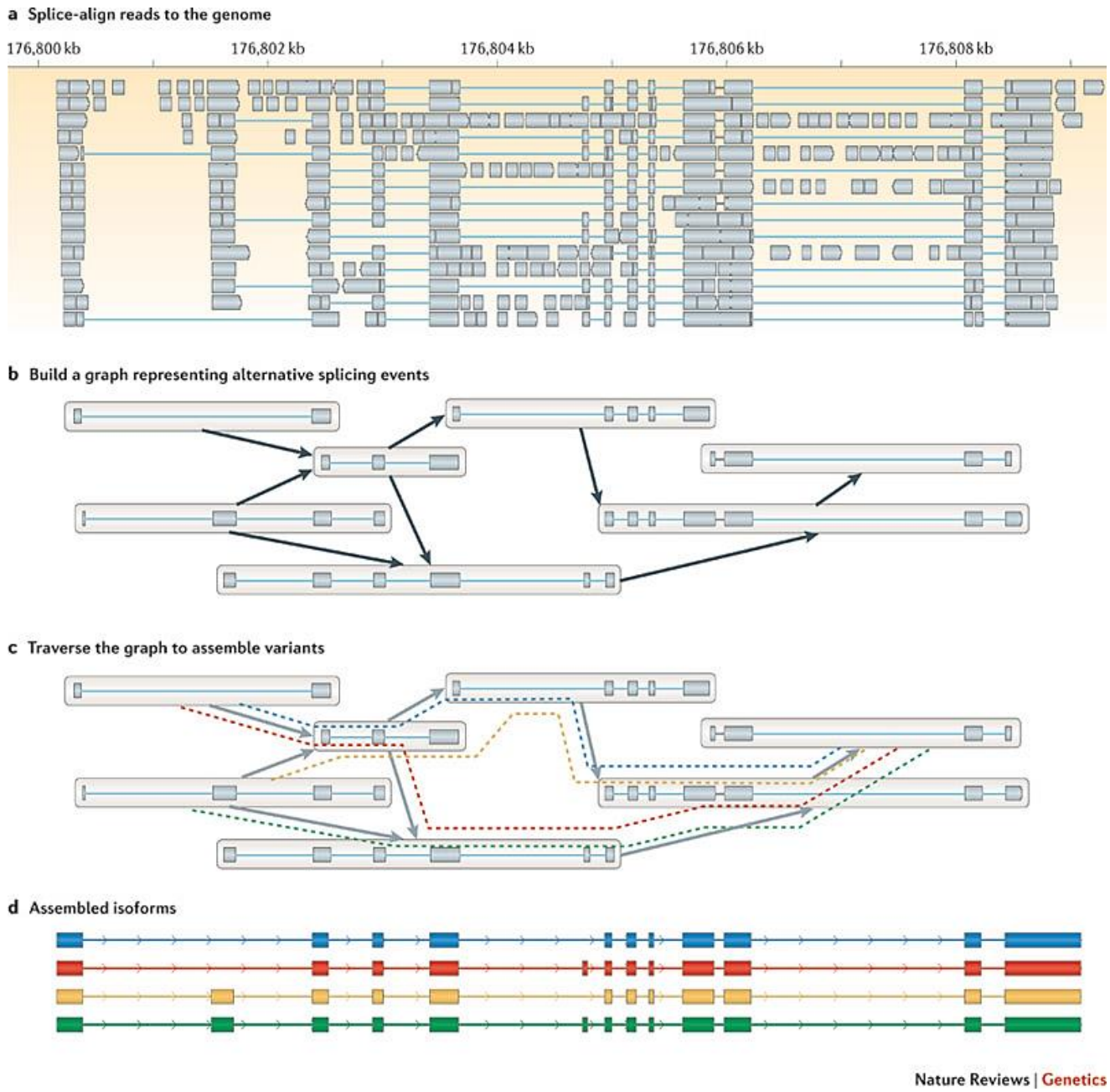


Figure 15 : Représentation des différentes étapes théoriques des stratégies d'analyse de transcriptome par alignement sur une séquence de référence. (exemple du gène GRMZM2G060216, géne du maïs)[55].

a. Les read sont d'abord alignés sur une référence
b. Un graphique de connectivité est ensuite construit afin de représenter toutes les isoformes possibles à un locus
c. Et d. au final, les différentes combinaisons établies par le graphique sont suivies pour joindre les read ensemble afin d'obtenir une représentation des différentes isoformes possibles

1ère partie : Application diagnostique du séquençage à haut débit par PGM pour *BRCA1* et *BRCA2*

I. Patients

Tous les patients pour qui est réalisée cette étude ont auparavant reçu un conseil génétique adapté à l'issue duquel ils ont fourni leur consentement éclairé. Il s'agit de patients répondant aux critères actuels d'indications d'un test génétique (Tableau 2)

Histoire personnelle	
1	adénocarcinome mammaire avant 36 ans
2	adénocarcinome type médullaire quel que soit l'âge au diagnostic
3	adénocarcinome de l'ovaire avant 61 ans
4	adénocarcinome mammaire + cancer ovaire
5	cancer du sein masculin
6	adénocarcinome triple négatif avant 51 ans
et/ou Histoire familiale	
1	3 cas de cancer du sein unis par des liens de 1er ou 2nd degré passant par un homme
2	2 cas de cancer du sein unis par des liens de 1er ou 2nd degré passant par un homme avec critères d'âge (différents selon les centres)
3	1 cas de cancer du sein et 1 ovaire unis par des liens de 1er ou 2nd degré passant par un homme

Tableau 2 : Indications actuelles des études génétiques

Les ADN sont extraits automatiquement à partir de prélèvement sanguins à l'aide de l'automate Quickgene 610-L de Fujifilm puis aliquotés à 50 ng/µL (par dosage au Nanodrop). Ils sont ensuite séquencés par série de 30 patients.

II. Protocole de séquençage PGM

Le protocole de séquençage par PGM est un processus complexe comprenant de nombreuses étapes nécessitant environ 15 jours de manipulations pour parvenir à l'étape de séquençage.

Ces différentes étapes vont être détaillées par la suite.

A. Enrichissement

Pour chaque patient, l'ADN extrait subit une étape d'enrichissement (Figure 16) des séquences codantes et des jonctions introns/exons des gènes *BRCA1* et *BRCA2* par PCR multiplex à l'aide du kit Multiplicom© *BRCA1/2*.

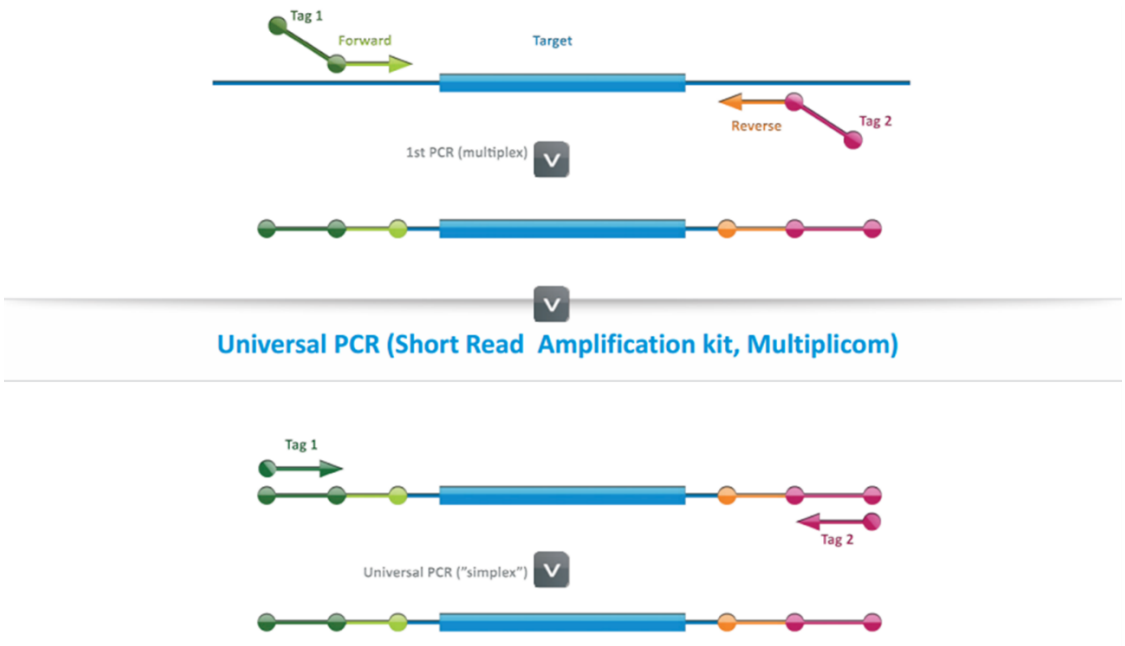


Figure 16 : principes d'enrichissement par PCR multiplex (Multiplicom)
La 1ère étape a pour but de fixer des amorces universelles (vert foncé et rose) à l'aide d'amorces spécifiques (vert clair et orange) de la séquence cible (« target », en bleu)
*Tous les produits de PCR multiplexes sont ensuite amplifiés par une étape de **PCR universelle** à l'aide de primers universels. Pour chaque patient, on obtient ainsi 5 multiplex différents d'amplicons.*
Les différents fragments peuvent ensuite être amplifiés à l'aide d'une PCR dite « universelle ».

La qualité des fragments ainsi obtenus est contrôlée par électrophorèse capillaire à l'aide d'un automate ABI3130XL (présence ou non de tous les amplicons attendus, efficacité et homogénéité des PCR multiplex).

A l'issue de ces étapes, les différents produits de PCR multiplex sont mélangés en proportions équivalentes pour chaque patient. Les pools sont alors purifiés à l'aide de billes Agencourt (Figure 17).

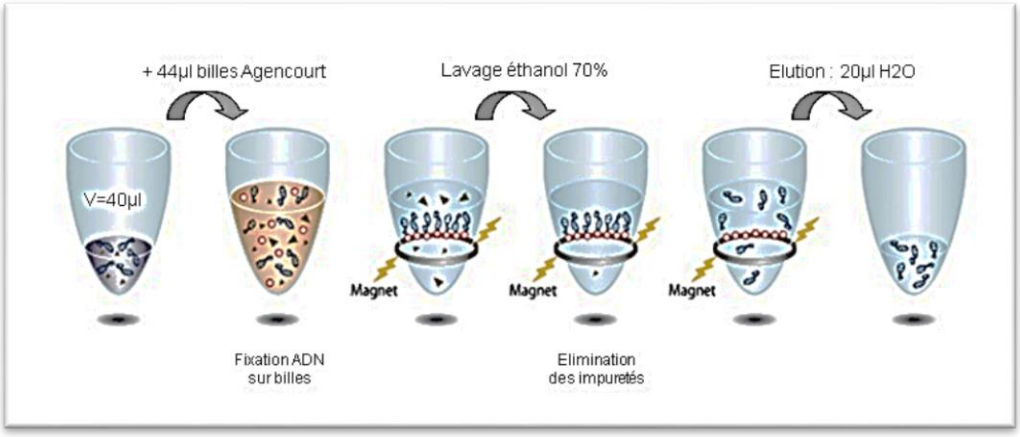


Figure 17 : étapes de Purification à l'aide billes Agencourt

B. Préparation des banques

Les différents pools purifiés sont dosés par spectrophotométrie (Genios-Tecan) à l'aide d'un kit Quant-iT dsDNA HS Assay kit. Ce dosage permet ensuite de déterminer le volume du pool à prélever pour chaque patient afin d'obtenir une quantité de 100ng d'ADN pour la préparation des banques à l'aide du système AB Library Builder (Life Technologies).

Les banques sont fragmentées de manière à obtenir des fragments d'une taille finale avoisinant les 300 pb. Les fragments sont liés aux adaptateurs P1 et A pour permettre par la suite l'amplification de tous les fragments puis aux séquences « code barre » (séquences d'environ 10 nucléotides permettant d'identifier les ADN par la suite quand ils seront tous mélangés pour le séquençage).

Les librairies sont ensuite mélangées et subissent une PCR d'amplification (à l'aide d'amorces complémentaires aux adaptateurs A et P1) visant à favoriser les constructions pertinentes

(c'est-à-dire contenant un adaptateur et un code barre) puis une étape de purification à l'aide de billes Agencourt.

Les librairies obtenues sont contrôlées par électrophorèse sur puce à l'aide du LabChip (Caliper) afin de vérifier leurs tailles moyennes, leurs concentrations et la qualité de leurs purifications (absence de primers en excès)(Figure 18). Si besoin, on procèdera à une nouvelle purification des échantillons.

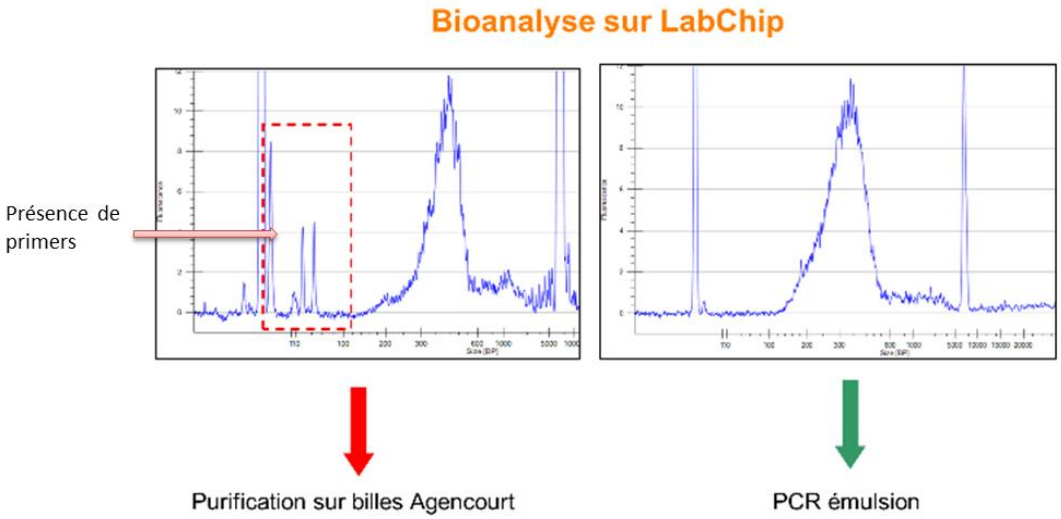


Figure 18 : Contrôle qualité des banques. S'il persiste encore des primers après purification (encadré rouge), une nouvelle étape de purification sera réalisée (présentation Power Point – laboratoire de Génétique – Institut Curie).

Pour finir, les librairies amplifiées sont mélangées entre elles de manière équimolaire afin de pouvoir être soumises à l'étape de PCR émulsion.

C. PCR Emulsion (PCRe)

Cette étape automatisée permet l'amplification clonale (OneTouch V2) (Figure 19) suivie d'un enrichissement en « *Ion Sphere Particles* » (ISPs) (Figure 20), sphères à la surface desquelles un fragment de librairie est amplifié (OneTouch ES). L'amplification clonale est réalisée au cours de la PCRe et contribue à atteindre un seuil de détection du signal nécessaire et suffisant au moment du séquençage. Malgré une optimisation du ratio 1/1 (1 ISP/1 Fragment d'ADN), plusieurs configurations de microréacteurs au sein des goutelettes sont envisageables. Seule la configuration de monoclonalité est souhaitée car elle seule sera source de données de séquençage. Les autres configurations généreront des données qui seront filtrées lors de

l'analyse primaire effectuée par la « *Torrent suite* » (serveur du séquenceur PGM permettant l'analyse bio-informatique des données brutes et le *base calling, cf. paragraphe suivant*).

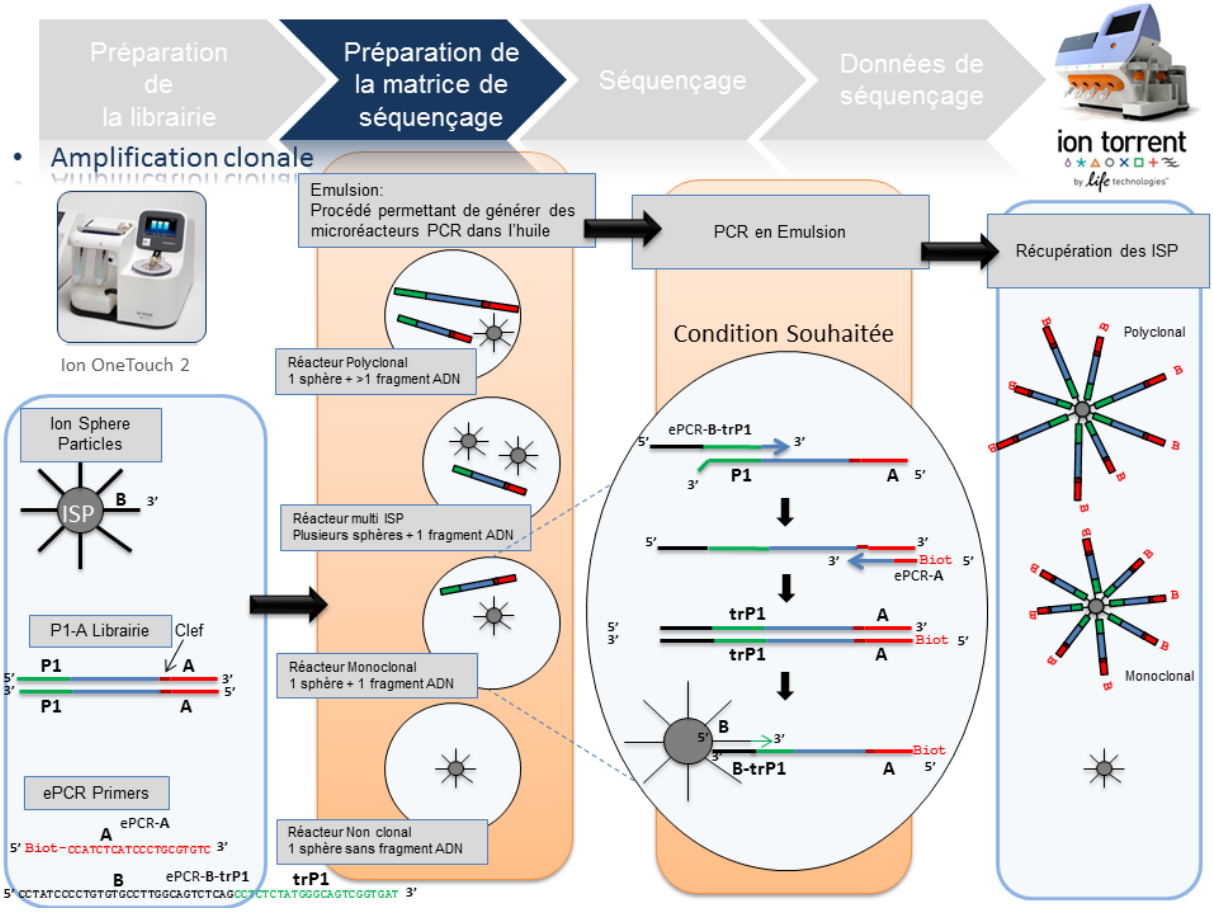


Figure 19 : Résumé des étapes de la PCRe.
Les conditions optimales sont résumées dans le 3ème cadre en partant de la gauche : au sein d'une seule et même micro-goutelette aqueuse se trouvent un fragment d'ADN, une sphère et les primers de la PCRe. L 'adaptateur P1 (en vert), permettra la fixation des fragments d'ADN au niveau de la surface des sphères. L'adaptateur A (en rouge) permettra par la suite la capture des sphères couplées aux fragments d'ADN[56].

L'amorce ePCR-A couplée à la biotine permettra l'enrichissement ultérieur par un système de capture sur billes liées à la streptavidine.

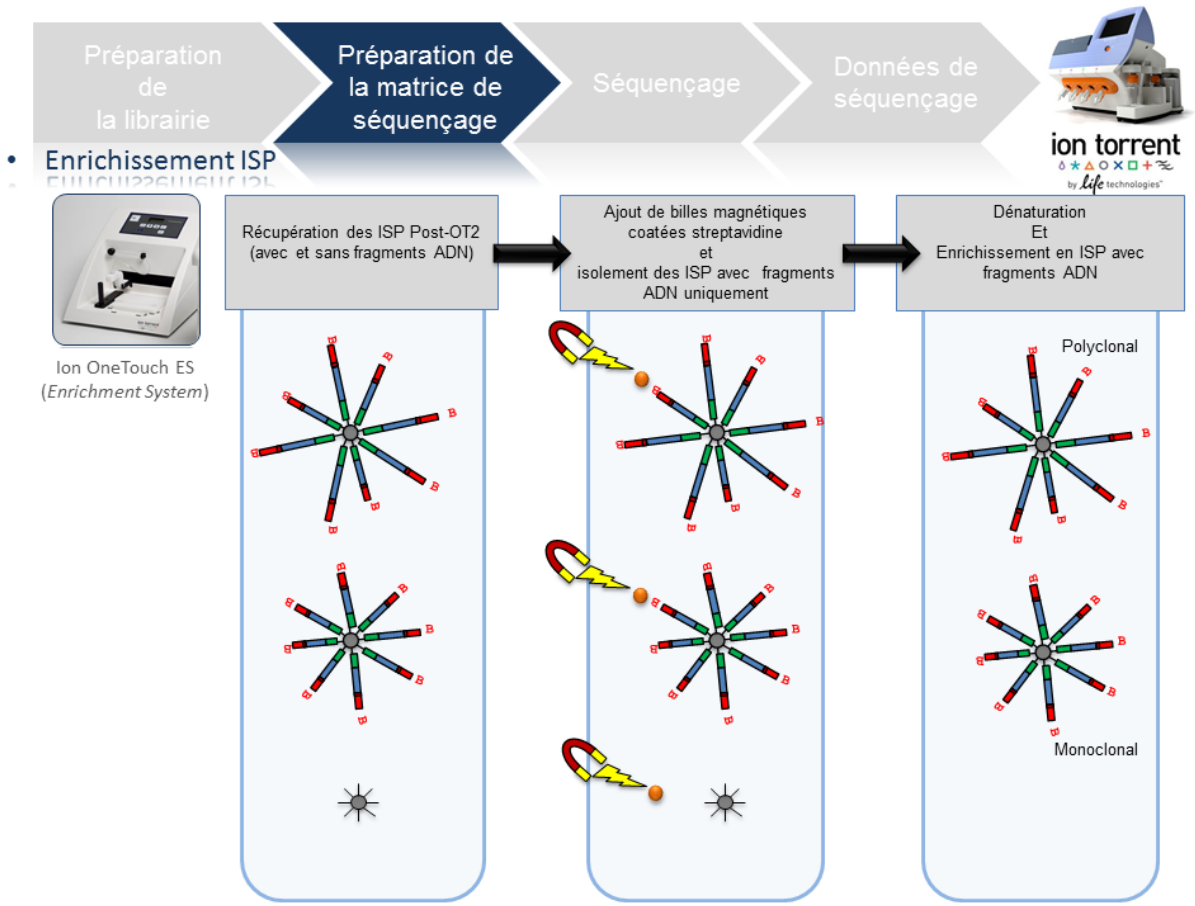


Figure 20 : Enrichissement en ISP. Les sphères couplées à des fragments d'ADN amplifiés sont chargées en Biotine via leur extrémité « A » (en rouge). Elles sont isolées à l'aide de billes magnétiques couvertes de streptavidine (en orange)[56].

D. Séquençage à haut débit

En amont de l'étape de séquençage, une étape d'initialisation du PGM est requise et permet notamment une homogénéisation des valeurs de pH aux environs de 7,8 au sein des différents réactifs de l'appareil (« Auto pH ») (Figure 21).

Une fraction de billes « tests de fragments » est également ajoutée aux ISP purifiées (elles permettent un contrôle de qualité du séquençage).

La matrice de séquençage couplée aux amorces de séquençage et à la polymérase est chargée sur la puce Ion Torrent selon un protocole bien spécifique. Les puces se déclinent selon 3

capacités de séquençage (Chip 314 >10Mb, Chip 316 >100Mb, Chip 318 >1Gb). A noter qu'une version « v2 » pour chacune des puces précitées existe et est indispensable pour toute application de séquençage nécessitant la chimie « 400 pb ». Le séquençage multiparallélisé revient donc au décryptage simultané des fragments d'ADN couplés aux ISP. A chaque polymérisation de nucléotides non modifiés, la libération d'ions H^+ entraine une variation de pH, elle-même détectée au niveau de la couche mince (technologie des semi-conducteurs) située au fond de chaque puits où elle est transformée sous forme de variation électrique. L'ensemble des données brutes générées est transcrit sous forme de ionogrammes.

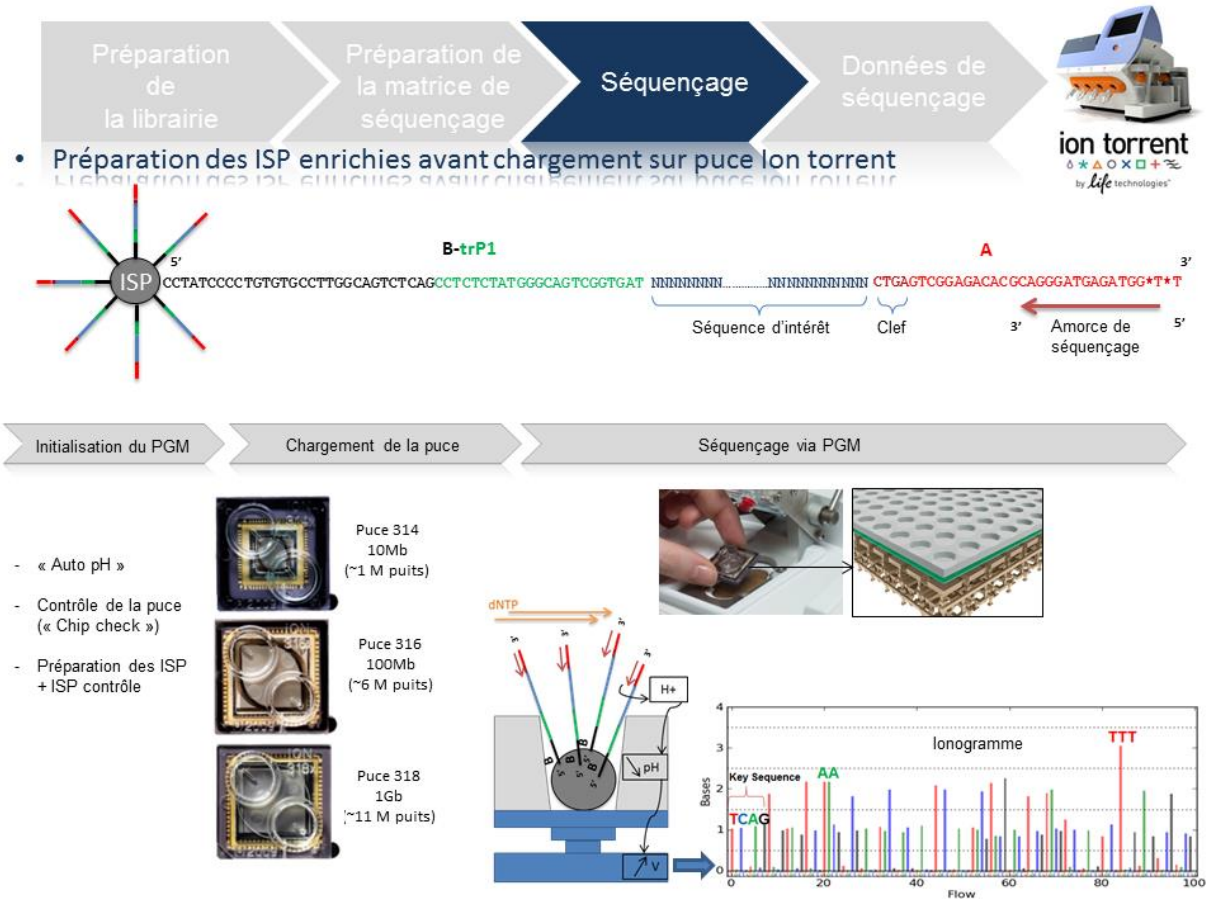


Figure 21 : principe du séquençage par PGM

- *Les puces sont élaborées de manière à ce qu'un puits ne puisse contenir qu'une seule sphère.*
- *Le principe du PGM repose sur la mesure en temps réel de l'émission de protons produits lors de l'incorporation d'une base pendant l'élongation.*
- *Ces signaux chimiques (variations de pH) sont transformés en signaux électriques (variations de voltage) puis traduits sous forme de ionogramme : à chaque cycle, une base différente (A,T,G,ou C) est injectée ; si elle est incorporée, une différence de potentiel est mesurée. Cette valeur est proportionnelle au nombre de base*
- *La séquence clé « TCAG » permet d'aligner les résultats sur ses niveaux d'expression (intensité de signal correspondant à 1 base incorporée)[56].*

A l'issue du « *run* » de séquençage, les fichiers regroupant l'ensemble des données brutes (ionogrammes) sont transférés du PGM vers le *Torrent Server* où ils seront traités par le programme *Torrent Suite*. L'algorithme de **«** *base calling* » permet la conversion des données sous forme de lettres en séquences (A,T,C,G) formant le « *read* » (séquence au format .FASTA) associé à un score de qualité pour chaque base incorporée (Score Phred codé en ASCII), les deux types de données étant associés dans un fichier .FASTQ (qui tend à devenir le format de référence).

Les scores de qualité Phred sont reliés de façon logarithmique à la probabilité d'erreur d'identification d'une base (Tableau 3).

Score de qualité Phred	Probabilité d'une identification incorrecte	Précision de l'identification d'une base
10	1 pour 10	90%
20	1 pour 100	99%

Tableau 3 : Les scores de qualité Phred

Un prétraitement est également appliqué sur la base des « *read* » générés et qui équivaut au nombre d'ISPs vivantes ou « *Live ISPs* » (On parle d'ISPs vivantes pour les ISPs associées à la cleé) :

- trimming : élimination des adaptateurs et/ou portions de read de mauvaise qualité (quality value QV>12)

- filtres : élimination des « *read* » de petites tailles, de mauvaise qualité, des polyclonaux

L'ensemble de ces informations est repris au travers du « *report* » généré à l'issue du séquençage et du prétraitement (Figure 22). Y sont également renseignés, le nombre de *read* générés ainsi que leur taille moyenne.

Run Report for auto_TD29

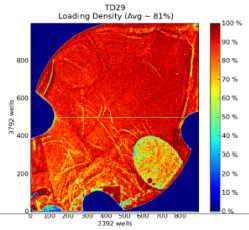
Run Summary

1.2 G
Total Bases

70
Key Signal

81%
ISP Loading

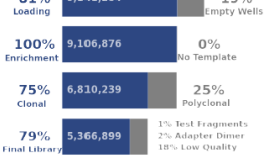
ISP Density



5,366,899
Total Reads

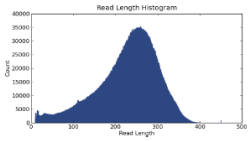
59%
Usable Reads

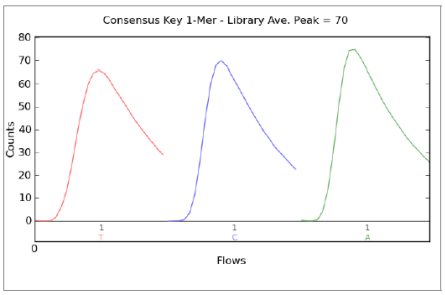
ISP Summary



226 bp
Mean Read Length

Read Length



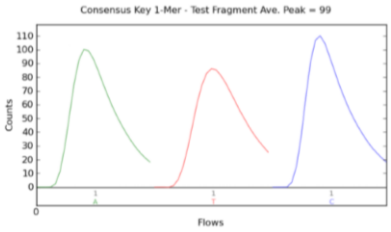


Addressable Wells	11,302,658	
With ISPs	9,141,204	80.9%
Live	9,106,876	99.6%
Test Fragment	35,492	00.4%
Library	9,071,384	99.6%
Library ISPs	**9,071,384**	
Filtered: Polyclonal	2,296,637	25.3%
Filtered: Low Quality	1,282,024	14.1%
Filtered: Primer Dimer	125,824	01.4%
Final Library ISPs	5,366,899	59.2%

Test Fragment Report

Test Fragment Summary

Test Fragment	Percent (50AQ17 / Num)
TF_A	95%



Based on Predicted Per-Base Quality Scores - Independent of Alignment

Total Number of Bases [Mbp]	1,212.97
‣ Number of Q20 Bases [Mbp]	1,094.55
Total Number of Reads	5,365,332
Mean Length [bp]	226
Longest Read [bp]	480

Figure 22 : Extraits d'un « report » généré à l'issue d'un run diagnostique. On y trouve :

- *Une image du chargement de la puce selon une échelle colorimétrique signalée sur la droite de l'image (rouge foncé = 100% de chargement) et le % de chargement de la puce.*
- *La quantité de données générée (en vert)*
- *La longueur moyenne des read séquencés (en jaune)*
- *La valeur des tests fragments (en rose)*
- *Le nombre de bases de bonne qualité de séquençage (Q20) (en bleu)*

III. Analyse des données de séquençage à l'aide du logiciel NextGENe ® (SOFTGENETICS ®)

Les données ainsi obtenues sont ensuite analysées à l'aide du programme d'analyse NextGENe puis triées à partir d'un arbre décisionnel établi au sein du laboratoire de l'institut sur la base d'une série test de 62 patients mutés pour lesquels les variations de *BRCA1* et *BRCA2* étaient connues[57]. Il a ensuite été validé à partir d'une série de 77 patients. Des tests et simulations ont permis de déterminer l'ensemble des paramètres d'analyse permettant d'assurer une sensiblité optimale à l'application du NGS en routine diagnostique (voir ci-dessous).

A. Prise en charge des anomalies détectées

Les mutations détectées à la sortie du PGM sont triées par NextGENe en fonction de leur type et de leur qualité (QV >12). Le PGM ne produisant pas beaucoup de faux positifs au niveau des substitutions, un seul filtre configuré dans NextGENe permet de détecter toutes les substitutions d'intérêt (figure 23).

Il est par contre nécessaire d'appliquer d'autres niveaux de filtres pour affiner le tri des indels et des homopolymères. L'enjeu est de permettre un affinage des résultats issus du PGM qui est connu pour produire un nombre conséquent de faux positifs de type fausse insertion ou fausse délétion.

Si le but est d'améliorer la spécificité du traitement des données, il est bien entendu indispensable de maintenir une sensibilité élevée en parallèle. Le tri des *indel* et des homopolymères se base sur différents niveaux de filtres (détaillés dans la Figure 3 et sa légende) :

- Un filtrage des données en fonction de la récurrence de l'anomalie détectée
- Un filtrage des données en fonction du taux de mutation
- Un filtrage éventuel des données en fonction du déséquilibre « Foward/Reverse » (et de la qualité de séquençage du run)

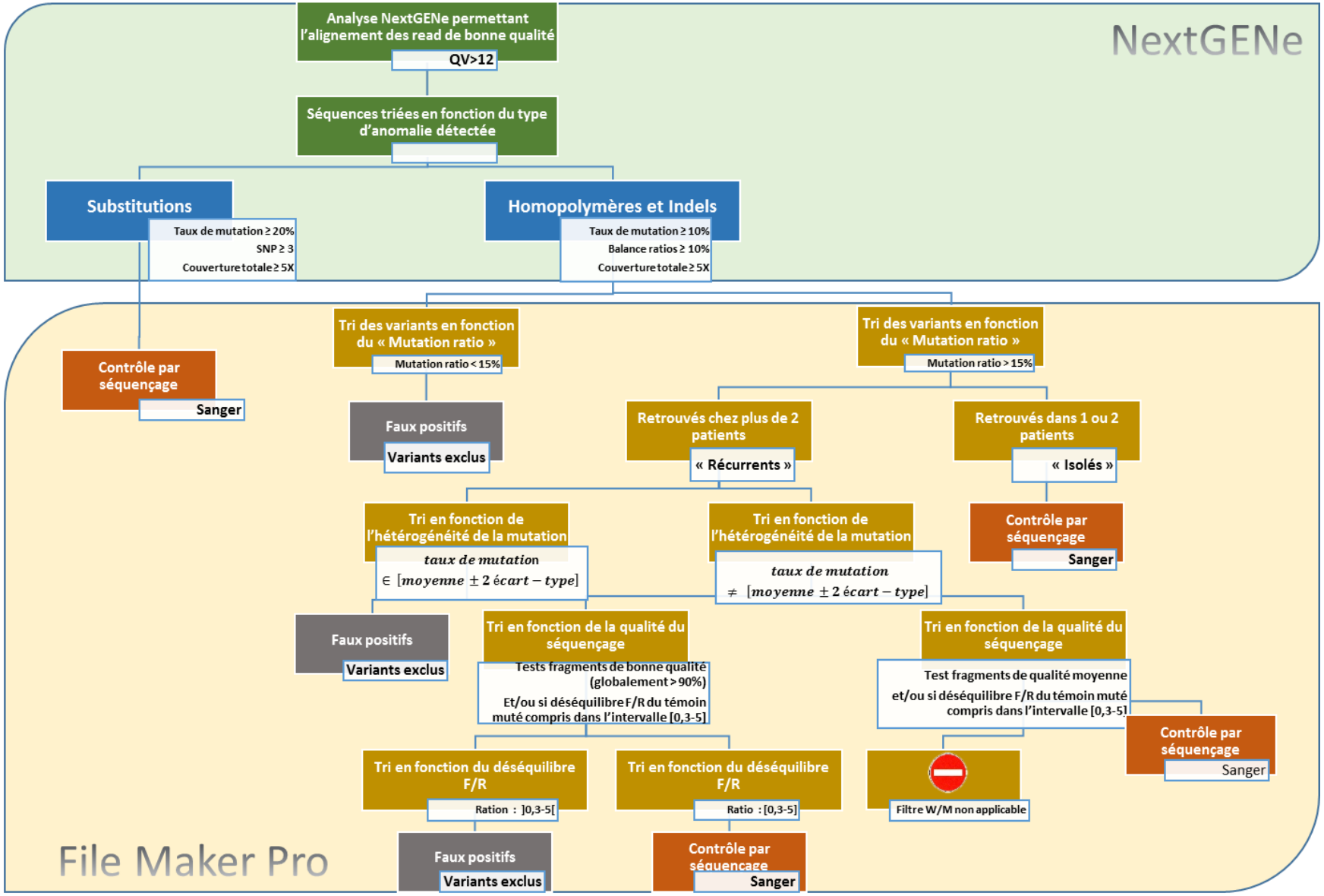


Figure 23 : Algorithme de traitement des variants détectées par PGM pour BRCA1 et BRCA2 en diagnostic

A) ***Prise en charge des substitutions*** *: Récapitulatif des paramètres de filtrage configurés dans NextGENe.*

- *« Mutation percentage » ou taux de mutation: Une variation entre le read aligné et la séquence de référence à une position donnée doit survenir à une fréquence excédant cette valeur pour être considérée comme une mutation (≥ 20%)*
- *« SNP allele » : Nombre de read portant une subsitution à partir duquel la variation est considérée comme étant une mutation (≥ 3)*
- *« Total coverage » ou couverture totale : Le nombre total de read à une position donnée doit dépasser ce taux minimal de couverture pour qu'une mutation soit retenue à cette position (≥ 5X)*
- *« Except for homozygous » : cette option est sélectionnée par défaut. Ceci signifie que ces paramètres d'exclusion ne s'appliquent plus si l'anomalie est retrouvée à l'état homozygote.*

B) ***Prise en charge des homopolymères et des indels*** *: L'enjeu des différents niveaux de filtrage appliqués est de permettre un affinage des résultats issus du PGM qui produit un nombre conséquent d'artéfacts de type fausses insertions ou délétions (sans pour autant sacrifier la sensibilité devant rester élevée (critère fondamental à une approche diagnostique).*

- *Un 1er filtrage des données issues du PGM est aussi effectué au moment de l'alignement des séquences par NextGENe. Ainsi sont éliminées les indels :*
  - *Dont le taux de mutation est < 10%*
  - *Dont la couverture totale est < 5X*
  - *Dont le « balance ratios » est inférieur à 10%*
    - *Le « balance ratio » est un critère composite permettant de trier en même temps les indels en fonction*
      - *De l'équilibre « Forward/Reverse » de la détection de la mutation dans les deux sens (si < 10% → séquence rejetée)*
      - *De l'équilibre « Forward/Reverse » à la position génomique de la mutation chez l'allèle muté comparativement à celui retrouvé au niveau de l'allèle sauvage à la même position*
- *Suite à ce 1er niveau de filtre « NextGENe » il persiste encore un nombre important d'indels détectées dont de nombreux « faux positifs ». D'autres techniques de triage des données sont alors mises en place pour diminuer le nombre de séquences à contrôler à tort.*
  - ***Si l'homopolymère ou l'indel est « isolé »***
    - *il peut s'agir très probablement d'une vraie mutation*
    - *cette anomalie est systématiquement contrôlée par une méthode de séquençage Sanger classique.*
  - ***Si l'homopolymère ou l'indel détecté est récurrent*** *(chez au moins 3 patients), de nouveaux filtres sont appliqués*
    - *la moyenne des différents taux de mutation retrouvés pour un même type d'anomalie est calculée. Un intervalle [moyenne des taux de mutations ± 2 écart-type] est constitué*
      - *Sont retenues les mutations présentant un taux de mutation compris dans cet intervalle (élimination des variants aux taux de mutations trop hétérogènes pour être vraisemblables)*
    - *Parmi les derniers variants persistants, on pourra appliquer un nouveau niveau de filtre en fonction de la qualité du run (déterminée par les valeurs des tests fragments et la détection du témoin muté par l'application de ce filtre supplémentaire). Ce filtre consiste en l'analyse des déséquilibres de brins Forward/Reverse en établissant pour une mutation le ratio*

$$\frac{\frac{Forward}{Reverse} WildType}{\frac{Forward}{Reverse} Muté}$$

Seront conservés les homopolymères et les indels pour lesquels ce ratio se situera dans l'intervalle [0,3-5]. (ces bornes ont été déterminées en effectuant ce calcul sur l'ensemble des variants observés jusqu'alors)

Des observations récentes ont démontré que ce niveau de filtre pouvait être pris en défaut dans le cas de runs de qualité inférieure et il reste à l'heure actuelle en cours d'évaluation

B. Contrôle des variants

Tous les variants restant sont contrôlés par séquençage Sanger. D'après les statistiques établies à partir des données diagnostiques actuelles du laboratoire de génétique de l'Institut Curie, 1,5 à 2 contrôles par patient sont effectués et on détecte en moyenne trois vraies mutations par série de 30 patients. Sur 1000 patients ayant eu un séquençage de *BRCA1* et *BRCA2* on retrouve en moyenne 13% de mutations et 14% de VSI)

IV. Exemple concret d'une série diagnostique

La série 139 est composée de 30 patients, ainsi que de 2 témoins :

- Témoin n°1 :
    - duplication des exons 3 à 8 de *BRCA1*
    - IVS7+36delCTT sur *BRCA1*
    - c.4383A>C sur *BRCA1*
    - c.5744C>T sur *BRCA2*
- Témoin n°2 : c.1961delA sur *BRCA1*

Elle a été séquencée le 18 août 2013. 2910 anomalies ont été détectées par le logiciel NextGENe (348 substitutions, 2559 homopolymères et *indel*) (Figure 24).

A la fin du filtrage des données, 60 contrôles PCR ont été effectués (soit 2 contrôles Sanger/patient en moyenne).

Sur ces 60 contrôles, 10 variants ont été validés soit :

- 3 VSI :
    - c.280C>T (VSI faux sens)
    - c.6978T>C (VSI neutre)
    - c.4378T>G (VSI faux sens)
- 5 mutations délétères :
    - Une délétion c.4038_4041del4 (entrainant un décalage du cadre de lecture)
    - 2 insertions (entrainant toutes deux un décalage du cadre de lecture) :
        - c.7680dupT

        - c.6079dupA
    - 2 mutations introniques ayant une répercussion sur l'épissage
        - IVS18-1G>C
        - IVS13+1G>A
- Et 2 variants s'avérant finalement être des polymorphismes déjà décrits
    - IVS23-10C>A
    - IVS25+9A>C

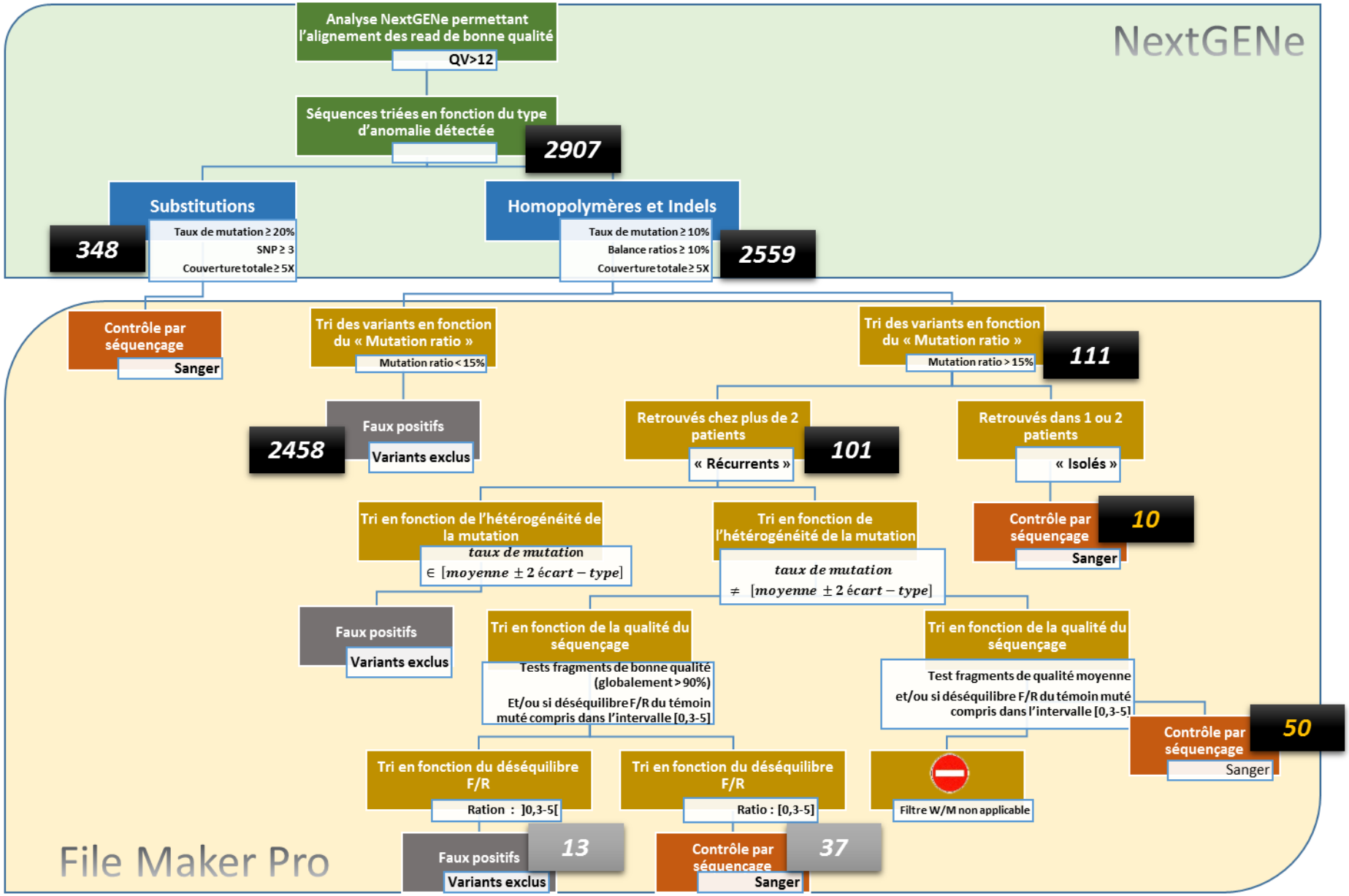

NextGENe
Analyse NextGENe permettant l'alignement des read de bonne qualité
QV>12
Séquences triées en fonction du type d'anomalie détectée
2907
Substitutions
Taux de mutation ≥ 20%
SNP ≥ 3
Couverture totale ≥ 5X
348
Homopolymères et Indels
Taux de mutation ≥ 10%
Balance ratios ≥ 10%
Couverture totale ≥ 5X
2559
Contrôle par séquençage
Sanger
Tri des variants en fonction du « Mutation ratio »
Mutation ratio < 15%
Faux positifs
Variants exclus
2458
Tri des variants en fonction du « Mutation ratio »
Mutation ratio > 15%
111
Retrouvés chez plus de 2 patients
« Récurrents »
101
Retrouvés dans 1 ou 2 patients
« Isolés »
Contrôle par séquençage
Sanger
10
Tri en fonction de l'hétérogénéité de la mutation
taux de mutation ∈ [moyenne ± 2 écart − type]
Faux positifs
Variants exclus
Tri en fonction de l'hétérogénéité de la mutation
taux de mutation ≠ [moyenne ± 2 écart − type]
Tri en fonction de la qualité du séquençage
Tests fragments de bonne qualité (globalement > 90%)
Et/ou si déséquilibre F/R du témoin muté compris dans l'intervalle [0,3-5]
Tri en fonction de la qualité du séquençage
Test fragments de qualité moyenne
et/ou si déséquilibre F/R du témoin muté compris dans l'intervalle [0,3-5]
Contrôle par séquençage
Sanger
50
Tri en fonction du déséquilibre F/R
Ration :]0,3-5[
Tri en fonction du déséquilibre F/R
Ratio : [0,3-5]
Filtre W/M non applicable
Faux positifs
Variants exclus
13
Contrôle par séquençage
Sanger
37
File Maker Pro

Figure 24 : Détails du traitement des variants détectés à l'aide de l'algorithme d'analyse appliqué en routine pour la série 139.
Le filtrage en fonction du « déséquilibre Foward/Reverse » n'a pas été utilisé pour cette série.

- *Sur fond noir : nombre d'anomalies filtrées à chaque étape (en jaune apparaissent le nombre d'anomalies contrôlées par Sanger)*
- *Sur fond gris : nombre de mutations qui auraient été exclues ou contrôlées si le filtre « déséquilibre Foward/Reverse » avait été appliqué.*

V. Discussion

Le processus d'études diagnostiques de *BRCA1* et *BRCA2* est constamment en cours d'amélioration, le but étant de gagner en qualité de séquençage et de tri des données (diminution du taux de faux positifs à contrôler en séquençage Sanger), en rapidité d'analyse et de rendu des résultats, tout en faisant face aux aléas de modification et de disponibilité des différents outils. Les obstacles fréquemment rencontrés dans l'application quotidienne de cette technologie sont nombreux. Ils représentent une partie importante du temps consacré à l'utilisation de cet outil et il semble inévitable de les mentionner également dans ce type de sujet.

A. Optimisation du processus actuel

Si le processus actuellement mis en place s'avère déjà très efficace, certaines étapes pourraient tout de même être, si ce n'est supprimées, au moins améliorées.

L'étape de préparation des banques sur le Library Builder reste une étape chronophage (1 jour ½ car nécessitant la préparation et le lancement de 3 séries). L'idéal serait de pouvoir profiter d'une technique liant directement les codes-barres aux amplicons et d'une capacité de séquençage ne nécessitant pas l'étape de fragmentation préalable.

Des runs-tests sont actuellement en préparation, l'un visant à tester le kit Ampliseq permettant de lier directement les codes-barres aux amplicons, l'autre visant à évaluer l'efficacité d'un séquençage en 400 pb d'une banque n'ayant subi aucune étape de fragmentation au préalable.

B. Difficultés liées aux fournisseurs

Le NGS est une technologie en plein développement et fait donc l'objet d'ajustements perpétuels de la part des différents fournisseurs : les kits commercialisés pour la préparation des PCR émulsions et pour le séquençage PGM changent souvent (arrêt de la production du kit 300 pb), les nouvelles versions de NextGENe sont commercialisées sans réelle information des changements qu'elles apportent… A cela s'ajoute des difficultés d'approvisionnement avec des ruptures de stock récurrentes.

Ces difficultés se font d'autant plus ressentir qu'il existe un manque important de communication de la part des fournisseurs quant à leurs produits, leurs évolutions futures et les retours que les autres utilisateurs peuvent leur en faire.

Durant ces derniers mois, une diminution de la qualité de séquençage a été observée sur certains *run* (isolés) avec des tests fragments de qualité plus basse. A l'heure actuelle, aucune réponse quant à la signification de ces résultats et l'origine possible de cette diminution de qualité ne nous a été fournie par Life Technologies (et ce malgré plusieurs relances).

Ces aléas matériels compliquent sévèrement l'utilisation de ces produits dans un circuit de diagnostic de routine avec des exigences en termes de qualité (Norme ISO 15189) et en termes de délais peu compatibles avec ce genre de réajustements. L'instauration d'un nouveau matériel nécessite à chaque fois l'investissement important de techniciens qualifiés pour les tester et les valider en un temps limité afin de ne pas (trop) impacter sur les délais de rendus de résultats.

C. Difficultés liées au traitement bio-informatique des données

L'analyse bio-informatique des résultats est une partie prépondérante dans l'utilisation de la technologie NGS. Idéalement, elle devrait trouver une place prépondérante en amont de tout projet. Malheureusement, les ressources effectives disponibles sont le plus souvent insuffisantes pour assumer l'ensemble des tâches qui leur revient (diagnotiques et de recherche). L'activité du laboratoire repose sur deux pipeline de traitement bio-informatique : l'un académique, l'autre commercial (logiciel NextGENe).

Le pipeline académique permettait notamment la mise en place d'une étude systématique des grands réarrangements de BRCA1 et BRCA2 à partir de l'analyse des fragments des différents amplicons. Cette approche très prometteuse n'est pourtant plus applicable depuis le départ du post doctorant en charge de ce projet. Ce problème récurrent d'expertise bio-informatique personne-dépendante très précise et difficilement transmissible est un aspect important à prendre en compte.

Concernant l'outil commercial NextGENe, celui-ci semble très efficace dans son utilisation diagnostique. Il faut néanmoins rester vigilant concernant les mises à jour effectuées sans réelle information quant aux changements opérés (modifications éventuelles des paramètres de filtrages ou d'alignement).

La manipulation bio-informatiques des données, étape essentielle dans l'utilisation de la technologie NGS reste un point critique sans solution idéale.

2ème partie : Utilisation de la plateforme PGM en recherche ; étude de l'impact des variants d'épissage pour *BRCA1* et *BRCA2*

I. Introduction

Les anomalies de l'épissage, et notamment celles pouvant découler des variations nucléotidiques de signification inconnue (VSI) représentent un enjeu majeur dans l'étude des risques de prédisposition aux cancers. Des outils bio-informatiques destinés à identifier les séquences contrôlant l'épissage et à prédire leurs altérations se sont développés pour répondre à cet enjeu. Ce développement est motivé par la nature variée et complexe des anomalies de l'épissage qui sont difficiles à appréhender dans le cadre du diagnostic. La problématique diagnostique est, en effet, bien distincte de celle de la recherche où les analyses sont généralement faites sur des cas isolés ou de petites séries, sans date limite, ni même nécessité de rendu de résultats. Le diagnostic en génétique moléculaire se fait sur de grandes séries, avec des délais de réalisation, et le résultat est utilisé pour le conseil génétique et le suivi médical du patient. Il s'agit donc, pour le biologiste, de relever le défi de la complexité et de le concilier avec la finalité diagnostique[3].

L'interprétation de ces anomalies est parfois délicate, car elles sont associées à une pénétrance et à une expressivité variable selon les individus porteurs. L'interprétation du caractère causal est encore plus complexe quand il existe des membres non atteints dans la famille.

Une difficulté de l'analyse des transcrits est l'existence d'un épissage alternatif physiologique, qu'il est indispensable de connaître, afin de ne pas interpréter comme délétère un saut d'exon(s) reflétant, en fait, un transcrit alternatif. La prise en compte des anomalies de l'épissage pour le diagnostic génétique est donc aussi importante que délicate.

La complexité de l'interprétation des VSI pourrait être facilitée par la disponibilité d'une technique d'étude sensible, puissante et rapide des transcrits obtenus pour un variant donné.

Le but principal de cette étude est d'analyser la pertinence du recours à la technologie NGS pour évaluer le retentissement de certains variants de *BRCA1* et *BRCA2* (connus, dans un 1er temps) sur l'épissage. Du fait de leur taille et d'un possible mécanisme d'épissage particulier, les exons 11 de *BRCA1* et *BRCA2* n'ont pas été étudiés.

Les différents objectifs sont :

- de parvenir à détecter les anomalies d'épissage attendues et
- de pouvoir ensuite les quantifier pour avoir une idée de l'impact réel et de la pénétrance de la mutation
- de pouvoir évaluer s'il existe, ou non, des variations d'expression entre les ARNm habituellement étudiés en diagnostic (issus de cellules lymphoblastoïdes) et ceux s'exprimant au sein du tissu d'intérêt (tissu mammaire)

II. Matériel et Méthodes

Trois approches ont été élaborées : la première utilise une amplification « maison » des ADNc de chaque patient et le séquençage des banques sur PGM (les résultats de ce design d'étude seront détaillés par la suite dans ce manuscrit), la deuxième utilise le kit de capture et d'amplification d'ARN Sure Select et la troisième le kit d'amplification Ampliseq.

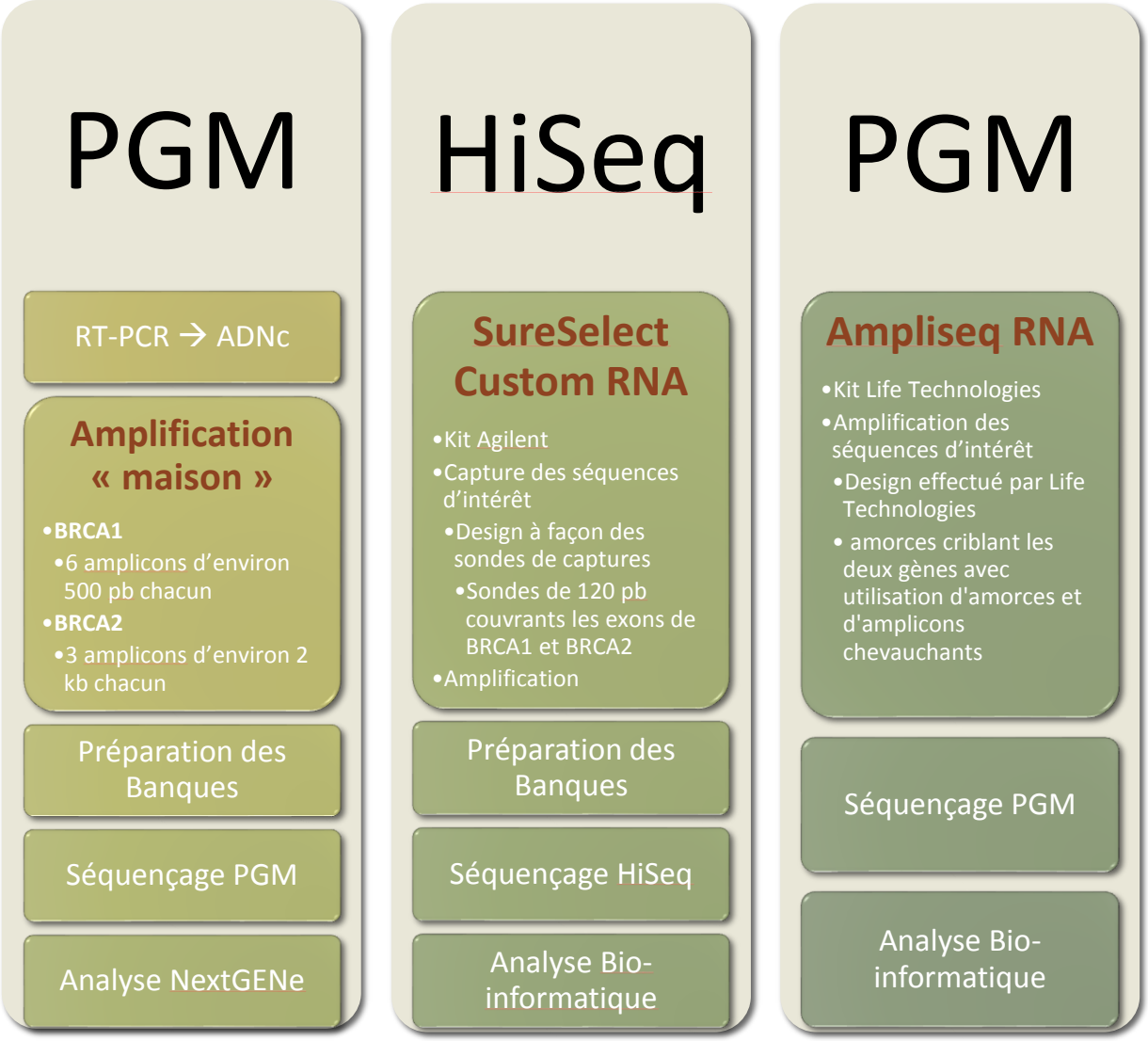


Figure 25 : Présentation des 3 bras de l'étude de l'impact des variants d'épissage par NGS proposés.

A. Patients et mutations

8 patients mutés pour BRCA1 et 8 patients mutés pour BRCA2 ont été sélectionnés (Tableau 4) selon les critères suivants :

- Mutation connue comme affectant l'épissage (l'étude ARN avait déjà été effectuée par le passé en séquençage Sanger, les variations d'épissage attendues étaient donc définies au préalable)
- ARN issu de lignées lymphoblastoïdes disponibles

    - Sans traitement par la puromycine
    - Avec traitement par la puromycine
- ARN issu de tissu mammaire sain disponible

Un autre patient présentant une délétion génomique de l'exon 3 a également été inclus à l'étude. Il a été utilisé comme « contrôle » dans la mesure où il présente de manière constitutive un transcrit « normal » issu de l'allèle non muté et un transcrit exempt de l'exon 3 issu de l'allèle où l'exon est délété. Il devrait donc permettre de contrôler la validité de la technique utilisée.

Gène	Numéro d'inclusion	Variant génomique	Variation d'épissage observée
BRCA1	1	c.4484G>T	Saut hors phase de l'exon 14
	2	c.5074G>C	Saut de l'exon 17
	3	c.5077_5080delGCTGins10	Saut de l'exon 18 en phase
	4	IVS05+3A>G	Anomalie d'épissage de l'intron 5
	5	IVS17-8T>G	Utilisation d'un site de novo d'épissage (7 nucléotides en amont en 3')
	6	IVS19+2delT	Saut de l'exon 19
	7	c.211A>G	Saut de l'exon 5 + utilisation d'un site cryptique 22 nucléotides en amont en 5'
	8	IVS08+3A>T	Saut hors phase de l'exon 8
	9	*Délétion génomique de l'exon 3*	
BRCA2	10	c.7805G>C	Saut de l'exon 16 + saut des 10 dernières bases de l'exon 10
	11	c.9117G>T	Saut de l'exon 23
	12	IVS05+3A>T	Saut hors phase de l'exon 5
	13	IVS07+3A>G	Saut de l'exon 7
	14	IVS15+1G>T	Saut hors phase de l'exon 15
	15	IVS22+2T>G	Saut de l'exon 22
	16	IVS02-7T>A	Augmentation de l'épissage alternatif de l'exon 3
	17	c.231T>G	Saut partiel de l'exon 3 et augmentation de l'épissage alternatif de l'exon 3

Tableau 4 : Tableau des mutations inclues et étudiées et anomalies d'épissage attendues.

B. RT-PCR

Les ADNc ont été obtenus à partir des ARN de chaque patient par RT-PCR selon 2 protocoles différents (enzyme SuperScript II pour BRCA1 et enzyme MuLV pour BRCA2).

(cf. Annexes 1 et 2)

C. Préparation des banques

Les préparations des banques ont été réalisées selon deux approches différentes pour *BRCA1* et *BRCA2*

1. *BRCA1*

Des primers ont été sélectionnés de manière à amplifier préférentiellement des séquences de ADNc d'environ 500 pb chacune (Tableau 5). On obtenait ainsi 6 amplicons différents couvrant l'ensemble du gène, à l'exception de l'exon 11 (de taille trop importante pour être couvert par un amplicon).

n° de l'amplicon	*exons couverts par l'amplicon*	*positions des primers*	*séquences des primers*
1	2-3-4-5-6-7	5'utr	CACCCTCTGCTCTGGGTAAA
		fin de l'exon 8	TGTAGACAGACGTCTTTTGAGGTT
2	8-9-10	Début de l'exon 7	TCAGCTTGACACAGGTTTGG
		Début de l'exon 11	ATGGCTCCACATGCAAGTTT
3	12-13-14	Fin de l'exon 11	GAACGGGCTTGGAAGAAAAT
		Exon 15	TTCTGAAGACTCCCAGAGCAA
4	15-16	Exon 14	TGCTGACAAGTTTGAGGTGTCT
		Début de l'exon 17	TGGTGTTTTCTGGCAAACTT
5	17-18-19	Début de l'exon 16	ACCCCTTACCTGGAATCTGG

6	19-20-21-22-23	Fin de l'exon 20	TGGACCTTGGTGGTTTCTTC
		Fin de l'exon18	CGGGAGGAAAATGGGTAGTT
		3' UTR	AAGCTCATTCTTGGGGTCCT

Tableau 5 : détails des amplicons et des primers utilisés

Les différents amplicons sont chevauchant sur le reste du gène sur au moins un exon à chaque fois (Figure 26).

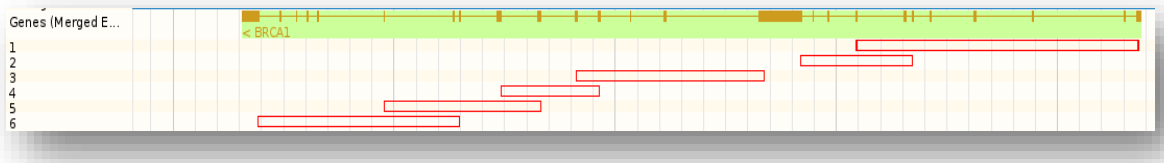


Figure 26 : Représentation de la séquence de référence de BRCA1 en orange avec visualisation des exons (rectangles) et des introns (lignes). La disposition des 6 amplicons (numérotés de 1 à 6, en rouge) est alignée par rapport à la référence en fonction des exons amplifiés.
www.ensembl.org

Une préparation des amplicons à l'aide du Library Builder (automate de préparation des librairies commercialisé par Life Technologues et également utilisé en routine, cf. 2ème partie II. B.) a permis de les fragmenter à une taille souhaitée de 300 pb et de les lier à leurs codes-barres respectifs.

Les librairies ont été préparées de manière à obtenir pour chaque type de prélèvement et chaque patient un mélange de ces 6 amplicons à séquencer simultanément.

Une exception a cependant été faite pour la délétion génomique de l'exon 3 « contrôle » pour laquelle deux types de librairies ont été préparées :

- La 1ère a été réalisée de manière « habituelle » : les 6 amplicons obtenus pour chaque prélèvement ont été mélangés puis traités ensemble pour la préparation de la librairie (un seul code barre est attribué pour chaque échantillon dans ce cas-là)
- La 2ème a été réalisée de manière « isolée » pour chaque amplicon pour un seul prélèvement (ARN extrait de lignées et traité par puromycine) : les 6 amplicons ont été séparément fragmentés et liés aux code-barres afin de pouvoir séquencer et analyser par la suite chaque amplicon de manière distincte (un code-barres différent a été utilisé pour chaque amplicon).

2. *BRCA2*

Les primers ont été sélectionnés de manière à constituer 3 amplicons d'environ 2 kb chacun qui couvrent à nouveau l'ensemble du gène, à l'exception de l'exon 11 (Tableau 6).

n° de l'amplicon	*exons couverts par l'amplicon*	*positions des primers*	*séquences des primers*
1	2-3-4-5-6-7-8-9-10	Exon 2	CACCTCTGGAGCGGACTTAT
		Début de l'exon 11	GCAGGCATGACAGAGAATCA
2	12-13-14-15-16-17-18-19-20-21	Fin de l'exon 11	ACTGCCAAGTCATGCCACAC
		Exon 22	CAATACGCAACTTCCACACG
3	19-20-21-22-23-24-25-26	Exon 18	ATGGAAAGGGATGACACAGC
		Exon 27	CTGGAAAGGTTAAGCGTCAA

Tableau 6 : Détails des amplicons et des primers utilisés

Les fragments obtenus après amplification étant de taille supérieure, une fragmentation manuelle a été effectuée de manière à obtenir des tailles finales d'environ 450 pb (le Library Builder ne permettant pas de réaliser une fragmentation automatisée pour des tailles de fragments supérieures à 300 pb). La ligation des codes-barres a également été effectuée manuellement.

D. Séquençage (cf. Annexe 4)

1. BRCA1

Après avoir effectué la PCRe à l'aide d'un kit 400, le séquençage a été effectué sur PGM en 400 pb (Figure 27).

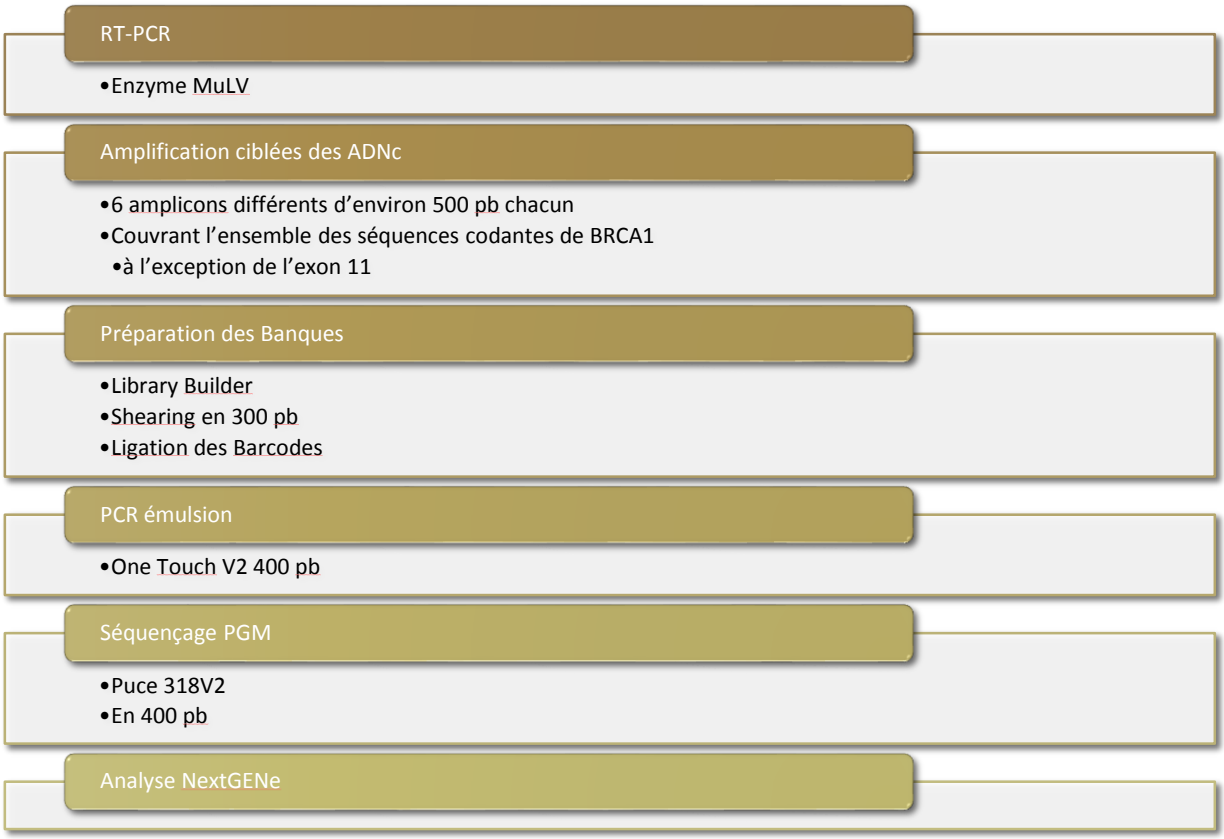


Figure 27 : récapitulatif des étapes de l'étude des transcrits de BRCA1

2. *BRCA2*

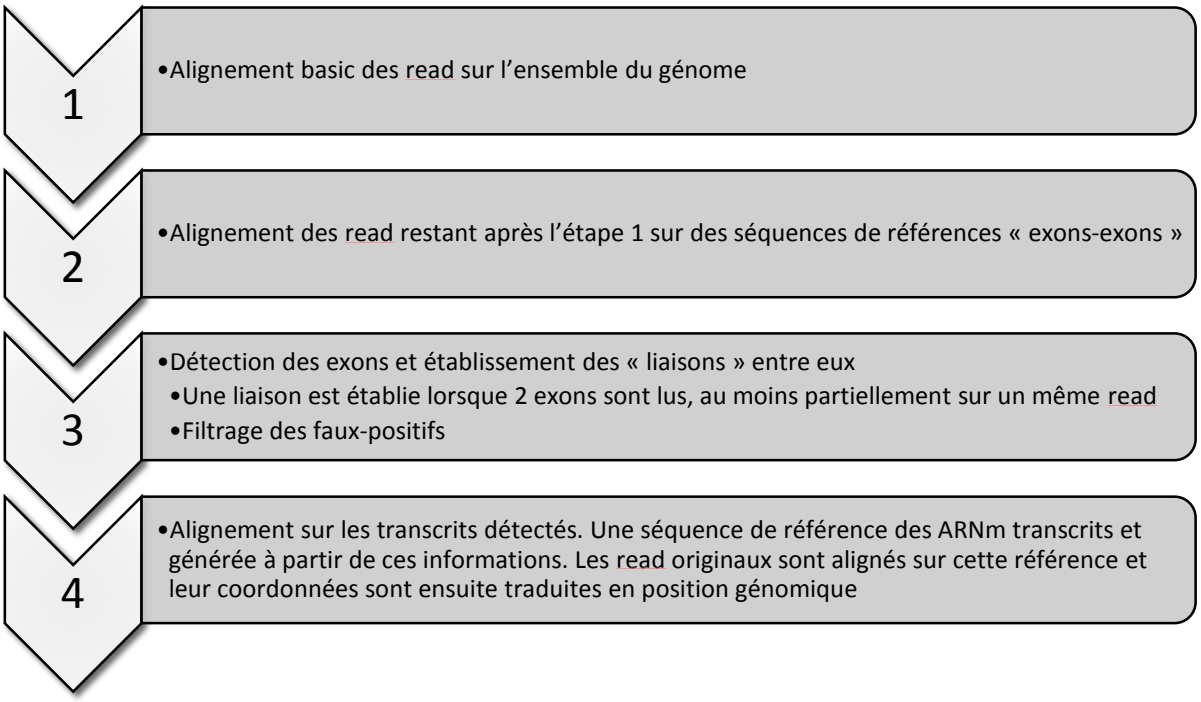
Le séquençage des banques a été effectué en 200 pb, le laboratoire n'ayant pas été fourni en kits 400 pb au moment du run.

Figure 28 : Récapitulatif des étapes pour l'étude des transcrits de BRCA2

E. Analyse des résultats

Les data de séquençage ont été récupérées à la sortie du PGM sous forme de *read*, c'est-à-dire de séquence. Elles ont alors été analysées à l'aide du logiciel NextGENe et du module dédié « Transcriptome » proposé dans la version commerciale du logiciel. Cette option permet l'alignement et la reconstruction des transcrits détectés par séquençage à haut-débit.

Ce module procède en 4 étapes d'alignements non modulables :



1. Viewer Transcriptome de Next GENe

Les résultats sont visibles dans un *viewer* spécifique (Figure 29) permettant de visualiser dans une fenêtre centrale les reconstructions des transcrits potentiels. Un code couleur est utilisé pour annoter les séquences détectées.

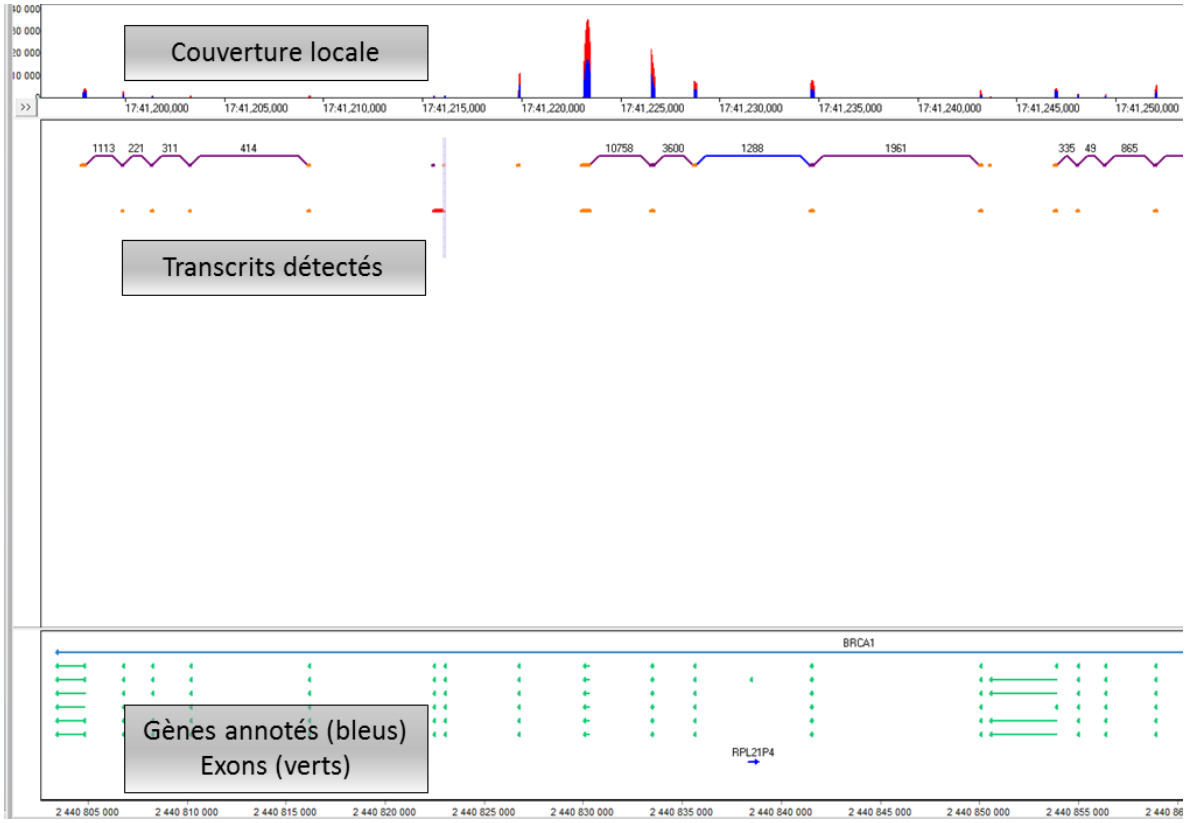


Figure 29 : Viewer du module « Transcriptome » proposé par Next GENe
Dans la fenêtre du haut : visualisation des profondeurs de couvertures obtenues (en bleu dans le sens foward, en rouge dans le sens reverse)
Dans la fenêtre centrale : Visualisation des transcrits détectés, les « rectangles » représentant les exons, reliés entre eux par des liaisons portant le nombre de fois où elles ont pu être établies. Ces transcrits sont représentés à l'aide d'un code couleur :

- *En violet : sont représentées les séquences exoniques ou liaisons correspondant à la séquence de référence*
- *En bleu : sont représentées les séquences exoniques ou liaison divergeant de la séquence de référence*
- *En orange : sont représentés les exons ayant un début ou une fin inhabituelle (site d'épissage alternatif)*
- *En rouge : sont représentées des rétentions de séquence intronique*
- *En rose : sont normalement annotés les sauts d'exon*
- *En gris : sont mentionnés début de 1er exon ou fin de dernier exon alternatif*

Dans la fenêtre du bas : sont représentés le gène de référence en bleu et les différents exons (en vert) retrouvés en fonction des transcrits décrits.

2. « *Transcript Report* »

A cette vue s'ajoute un « *transcript report* » (Figure 30) : tableau exportable fournissant des informations à propos des exons détectés (comme leur localisation), de la profondeur de

couverture, du nombre de liaisons établies entre deux exons, du type de variant et du type d'isoforme détecté.

Chr All Gene All

Page 1 of 1 First<< Previous< 1 >Next >>Last to Page 1 Go

Index	Chr	Start	End	Length	Gene	Exon(s)	Link N	PE Link	Avg Coverage	<-Cov	->Cov	Type	Isoform	Protein
1	1	89,290,619	89,291,122	504	PKN2	NA	NA	NA	24	NA	NA	Insertion	NM_006256.2	NP_0062
2	7	1,908,214	1,909,040	827	MAD1L1	NA	NA	NA	20	NA	NA	Insertion	NM_003550.2	NP_0035
3	8	22,546,067	22,546,441	375	EGR3	1	NA	NA	2	NA	NA	Alt Splice Site	NM_001199880.1	NP_0011
4	8	26,179,211	26,179,659	449	PPP2R2A	NA	NA	NA	11	NA	NA	Insertion	NM_001177591.1	NP_0011
5	14	78,373,676	78,374,049	374	ADCK1	NA	NA	NA	5	NA	NA	Insertion	NM_001142545.1	NP_0011
6	15	38,818,284	38,818,724	441	RASGRP1	15	NA	NA	46	NA	NA	Alt Splice Site	NM_005739.3	NP_0057
7	17	41,215,891	41,215,968	78	BRCA1	7	NA	NA	93	NA	NA	Exon	NM_007298.3	NP_0092
8	17	41,215,968	41,219,625	3658	BRCA1	7;8	8	NA	NA	93	364	Known Link	NM_007298.3	NP_0092
9	17	41,219,625	41,220,013	389	BRCA1	8	NA	NA	364	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
10	17	41,222,945	41,223,255	311	BRCA1	9	NA	NA	3960	NA	NA	Exon	NM_007298.3	NP_0092
11	17	41,223,255	41,228,505	5251	BRCA1	9;10	78	NA	NA	3960	3064	Known Link	NM_007298.3	NP_0092
12	17	41,242,961	41,243,049	89	BRCA1	13	NA	NA	2199	NA	NA	Exon	NM_007298.3	NP_0092
13	17	41,256,882	41,256,985	104	BRCA1	19	NA	NA	289	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
14	17	41,258,473	41,258,550	78	BRCA1	20	NA	NA	20	NA	NA	Exon	NM_007298.3	NP_0092
15	17	41,258,550	41,267,747	9198	BRCA1	20;21	13	NA	NA	20	99	Known Link	NM_007298.3	NP_0092
16	17	41,276,034	41,276,132	99	BRCA1	22	NA	NA	153	NA	NA	Exon	NM_007298.3	NP_0092
17	17	41,276,132	41,277,288	1157	BRCA1	22;5	137	NA	NA	153	102	Known Link	NM_007298.3	NP_0092
18	17	41,277,288	41,277,308	21	BRCA1	5	NA	NA	102	NA	NA	Exon	NM_007298.3	NP_0092
19	17	41,222,945	41,223,269	325	BRCA1	9	NA	NA	8568	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
20	17	41,223,269	41,228,505	5237	BRCA1	9;10	236	NA	NA	8568	964	Novel Link	NM_007298.3	NP_0092
21	17	41,256,130	41,256,278	149	BRCA1	18	NA	NA	964	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
22	17	41,256,278	41,256,882	605	BRCA1	18;19	3	NA	NA	964	289	Novel Link	NM_007298.3	NP_0092
23	17	41,215,346	41,215,973	628	BRCA1	6	NA	NA	27	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
24	17	41,222,875	41,223,255	381	BRCA1	9	NA	NA	10686	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
25	17	41,223,255	41,226,344	3090	BRCA1	9;10	14	NA	NA	10686	2187	Novel Link	NM_007298.3	NP_0092
26	17	41,256,130	41,256,282	153	BRCA1	18	NA	NA	2187	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
27	17	41,256,282	41,256,885	604	BRCA1	18;19	57	NA	NA	2187	149	Novel Link	NM_007298.3	NP_0092
28	17	41,222,875	41,223,271	397	BRCA1	9	NA	NA	3960	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
29	17	41,234,421	41,234,845	425	BRCA1	12	NA	NA	3016	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
30	17	41,226,343	41,226,552	210	BRCA1	10	NA	NA	4788	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
31	17	41,226,344	41,226,548	205	BRCA1	10	NA	NA	7680	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
32	17	41,228,502	41,228,639	138	BRCA1	11	NA	NA	2693	NA	NA	Alt Splice Site	NM_007298.3	NP_0092
33	17	41,228,639	41,234,421	5783	BRCA1	11;12	2	NA	NA	2693	3902	Novel Link	NM_007298.3	NP_0092

Figure 30 : Transcript report résumant les caractéristiques des exons, liaisons et anomalies détectés.

III. Résultats et Discussion

Tous les résultats de séquençage ont été analysés à l'aide du module « transcriptome » de NextGENe.

Les résultats ont été traités en deux temps : l'analyse s'est tout d'abord concentrée sur l'interprétation des données obtenues pour les amplicons séquencés séparément les uns des autres pour le « contrôle » (délétion génomique de l'exon 3). L'étude des résultats obtenus

pour les « *pool* » d'amplicons a eu lieu dans un second temps seulement (une fois les premières données décryptées)

A. Analyse des résultats obtenus pour la délétion génomique de l'exon 3 de BRCA1

Les quantités de données générées par l'analyse transcriptome étant très abondantes et de manipulation ardue, nous avons tout d'abord décidé d'observer les résultats obtenus distinctement pour chacun des amplicons chez ce patient avant d'essayer d'analyser les résultats des « mélanges » d'amplicons effectués pour chaque type de prélèvement.

1. Analyse « transcriptome » des données obtenues par le séquençage des amplicons de BRCA1 de manière isolée.

Cette analyse a permis d'observer la délétion hétérozygote de l'exon 3 avec visualisation d'un transcrit altéré et d'un transcrit normal pour le 1er amplicon (figure…).

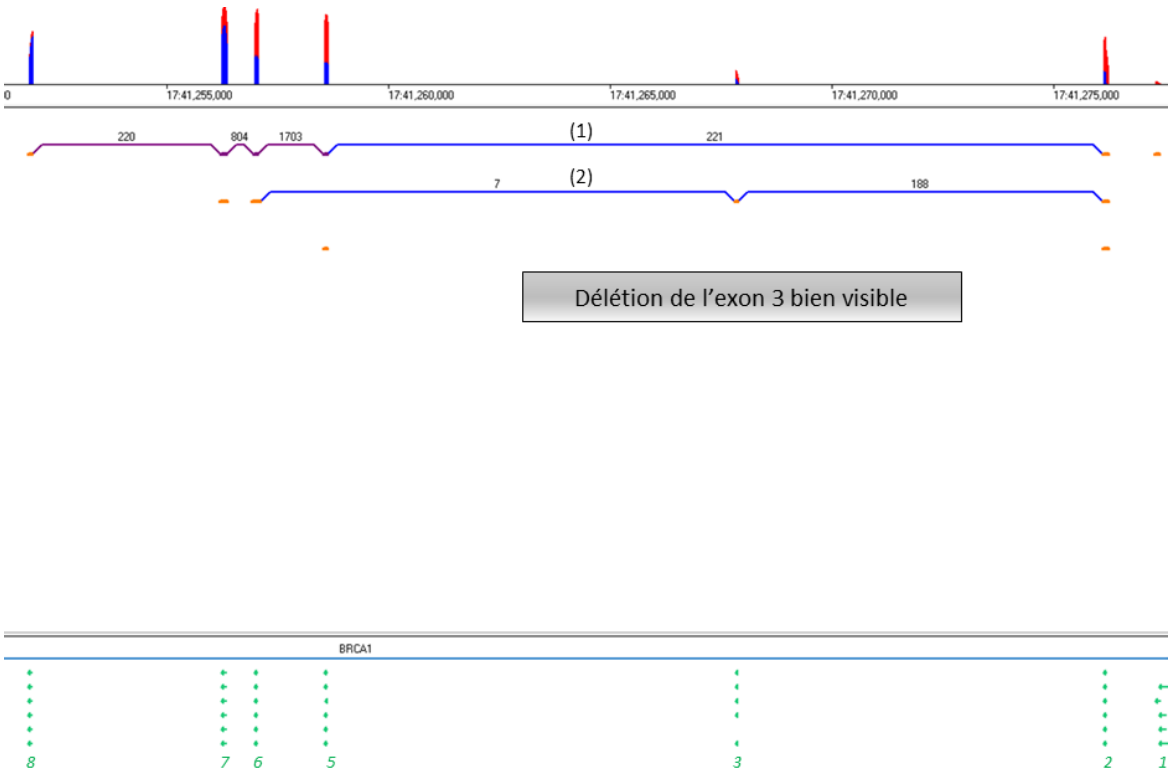


Figure 31 : Visualisation des transcrits détectés pour la délétion de l'exon 3 de BRCA1. Les exons de référence de BRCA1 ont été annotés et numérotés en vert. Deux transcrits ont été détectés : le 1er ne comportant aucune séquence exonique en regarde de l'exon 3 (1), le 2ème comportant bien une séquence correspondant à cet exon et représentant donc le transcrit « normal » (2).

2. Analyse « transcriptome » des résultats obtenus par séquençage des mélanges d'amplicons.

Les 3 prélèvements (ARN issus de lignée lymphoblastoïde non traitée ou traitée par puromycine et ARN extraits de tissu mammaire sain) ont été analysés de cette manière.

Dans les 2 premiers prélèvements, les deux types de transcrits attendus ont bel et bien été retrouvés.

Par contre, lors de l'analyse des transcrits au niveau du tissu mammaire sain, seul le transcrit anormal a été annoté dans Next GENe. Ceci pourrait laisser supposer qu'il s'agissait peut-être de tissu mammaire d'origine tumorale (Figure 32).

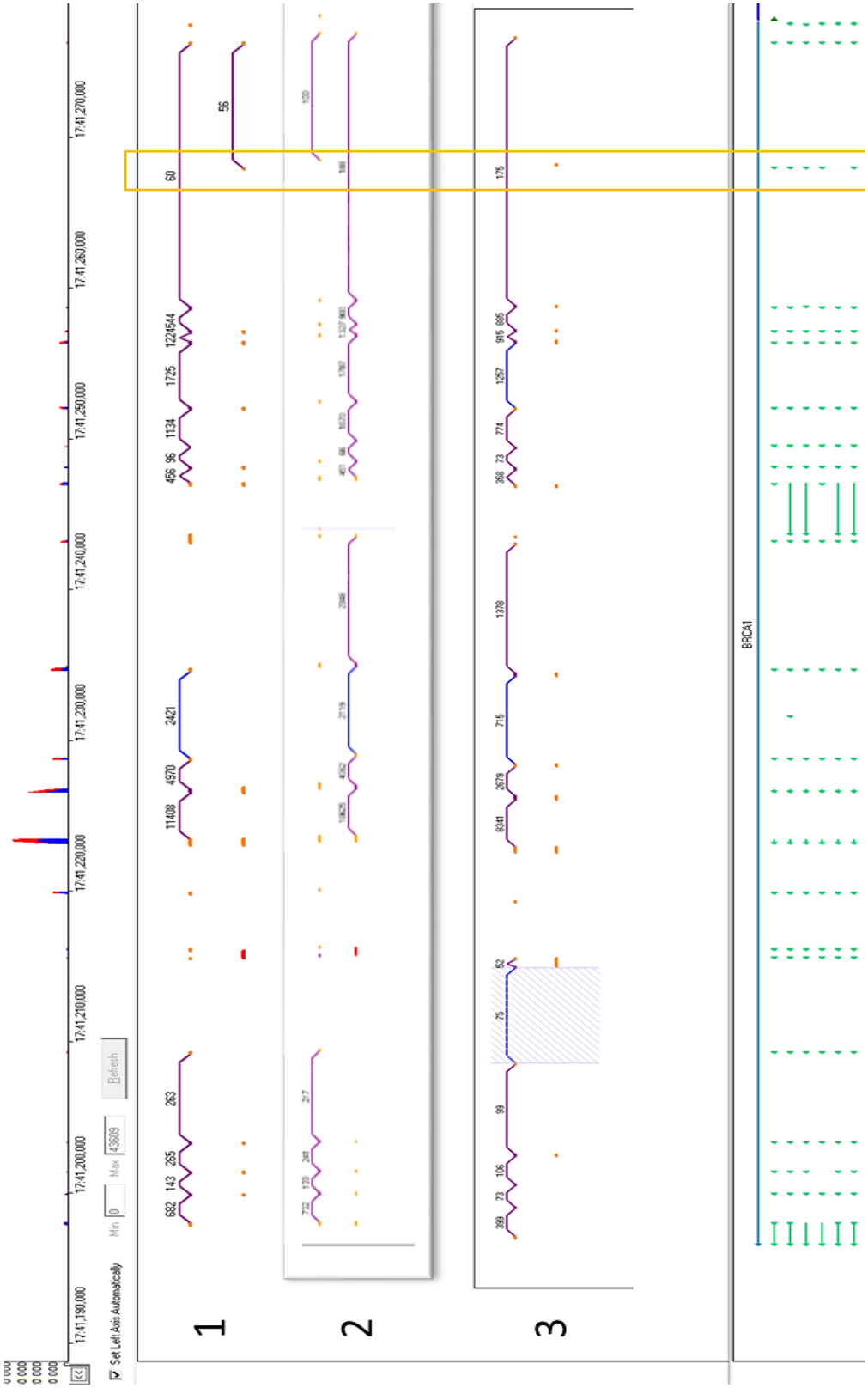


Figure 32 : condensation des résultats obtenus pour les 3 types de prélèvement différents pour la délétion génomique de l'exon 3. La position théorique de l'exon 3 est indiquée par l'encadrement jaune. En (1) sont représentés les transcrits détectés dans les échantillons issus de lignées lymphoblastoïdes non traités par puromycine, en (2) les transcrits détectés dans les échantillons issus de lignées lymphoblastoïdes traités par puromycine et en (3) les transcrits détectés dans les échantillons issus du tissu sain mammaire.

B. Analyse « transcriptome » des autres patients inclus

De la même manière, ces analyses ont été effectuées sur l'ensemble des échantillons disponibles. La mutation c.4484G>T de BRCA1 entrainant un saut hors phase de l'exon 14 est réprésentative des difficultés rencontrées (Figure 33).

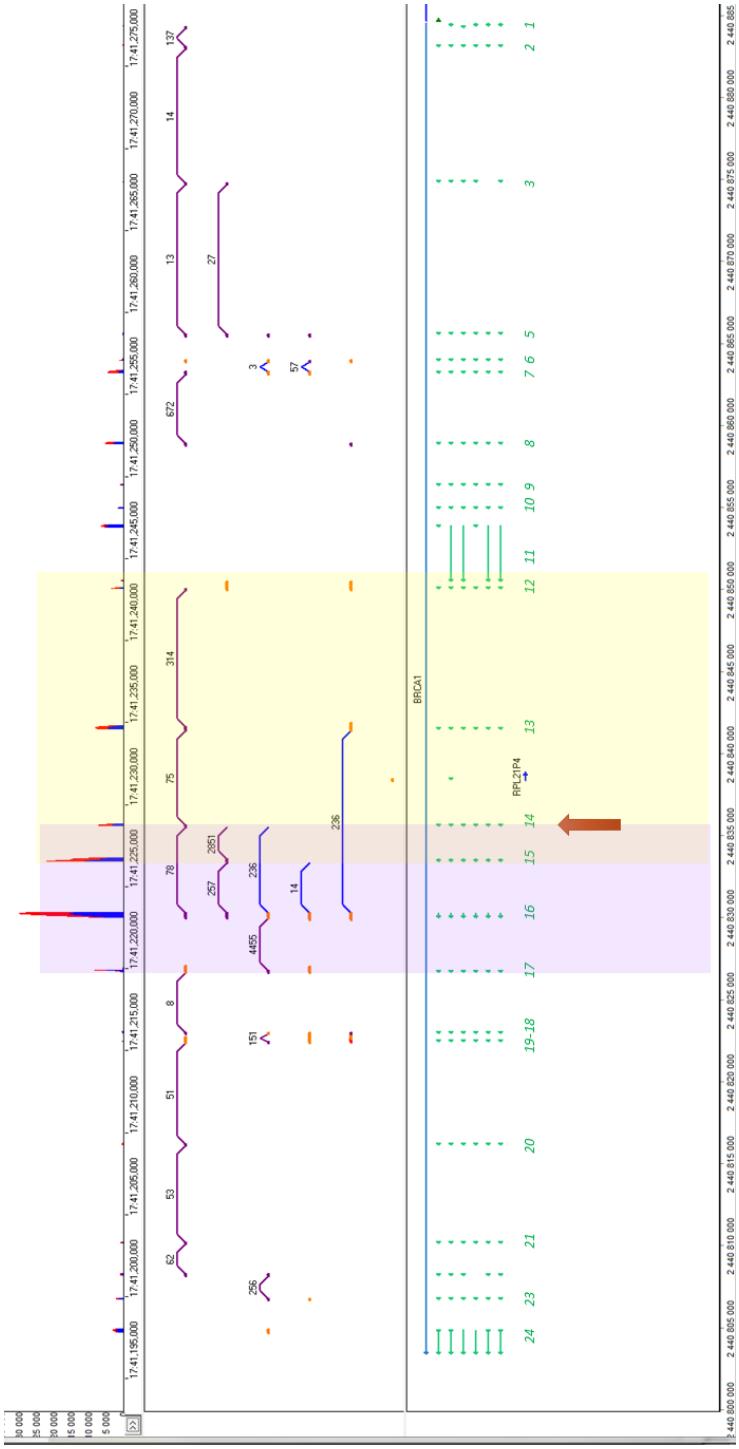


Figure 33 : représentation des résultats obtenus lors de l'analyse transcriptome des séquences provenant de cellules lymphoblastoïdes (non traitées par la puromycine) du patient présentant la mutation c.44B4G>T (entrainant un saut hors phase de l'exon 14).

- Flèche orange : indique où se situent normalement les séquences correspondant à l'exon 14 dans le viewer.
- En jaune et en violet : limites des régions couvertes par les deux amplicons incluant l'exon 14
Sur la dernière ligne des transcrits reportés, on observe un saut d'exon englobant les exons 14 et 15.

L'analyse pour cette mutation a débuté avec la reprise des résultats obtenus pour l'échantillon issus de cellules lymphoblastoïdes non traité par la puromycine. Si le saut de l'exon 14 était bel et bien retrouvé sur la dernière ligne des transcrits, celui-ci était associé à un saut de l'exon 15 qui n'a, a priori, pas lieu d'être.

Dans l'échantillon issu de cellules lymphoblastoïdes traité par puromycine, ce saut de l'exon 14 associé au saut de l'exon 15 a également été retrouvé. Cependant, dans ce cas-là, ce saut d'exon était annoté comme « normal » par le logiciel… Or aucun transcrit référencé ne décrit cette variation comme étant une variante possible d'épissage normal. Les transcrits du tissu mammaire sain ont été analysés en dernier, mais aucun transcrit inhabituel n'a été décrit.

D'autre part, aucun transcrit alternatif connu (par exemple le transcrit Δ9-10 de *BRCA1*) n'a été annoté par l'analyse transcriptome.

Le module transcriptome s'est très vite avéré être insuffisant pour l'interprétation des résultats. De manipulation fastidieuse et de qualité d'analyse imprécise, seuls quelques échantillons pour chaque gène ont pu être étudiés manuellement. Ceci a cependant permis, à l'aide d'exemples concrets, de cerner les inconvénients et les insuffisances du programme.

C. Difficultés d'analyse liées au module « transcriptome » de NextGENe

L'utilisation de l'outil dédié « transcriptome » proposé par NextGENe semblait très prometteuse dans le cadre des objectifs fixés par cette étude. Elle s'est cependant très vite avérée ardue et inadaptée.

1. « Transcript report » inexploitable

Le tableau résumant les résultats obtenus pour chaque échantillon ne peut malheureusement pas être utilisé pour un traitement numérique et standardisé des résultats. En effet, l'annotation des exons est fausse avec confusion parfois entre certains exons proches (par exemple, les exons 10 et 11 sont fréquemment traités, à tort, comme un seul et même exon).

L'équipe de NextGENe a reconnu l'existence d'anomalies d'annotation des gènes trasncrits en sens inverse. Ces « bugs » devraient faire l'objet de corrections dans les versions du logiciel à venir.

2. Variations importantes des taux de couverture

Même au sein d'un seul et même amplicon, les taux de couvertures des différents exons peuvent être très hétérogènes. Il en va de même d'un amplicon à l'autre pour un même prélèvement. Aucune hypothèse ne permet d'expliquer ces variations mais il semblerait qu'elles puissent en partie provenir d'un tri et d'un trimming importants des *read* sur lesquels nous n'avons pas d'influence (Figure 34.

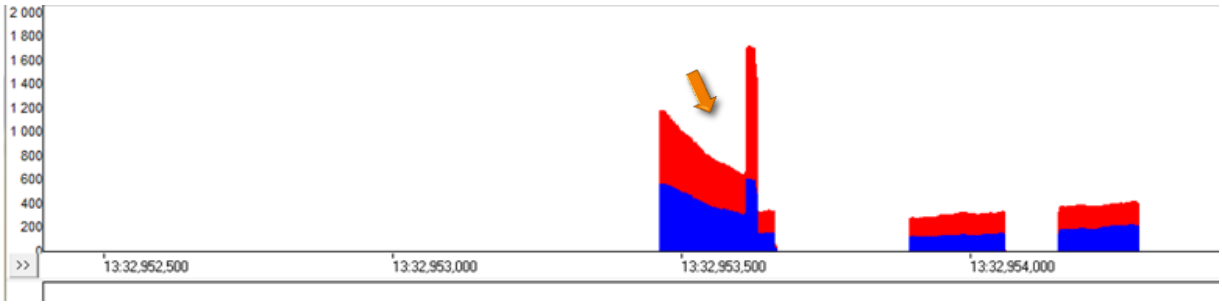


Figure 34 : exemple de l'exon 22 de BRCA2 (à gauche) présentant ici une répartition très inégale de sa profondeur de couverture (exemple tiré des résultats obtenus pour la mutation c.9117G>A de BRCA2).

- *En abscisse : position génomique*
- *En ordonnée : profondeur de couverture (en nombre de X)*
- *En rouge : lecture forward (ou sens)*
- *En bleu : lecture reverse (ou anti-sens)*

Ces variations de couverture sur un seul et même exon n'ont pas lieu d'être en théorie et suggère donc un trimming des données important. L'élagage important sur l'exon 22 (flèche orange) peut s'expliquer par la présence de séquences d'homopolymères et/ou par des difficultés d'alignement

3. Anomalies d'alignement

A plusieurs reprises, l'alignement de NextGENe semble être pris en défaut :

- Fausses rétentions introniques entre deux exons proches
- Exons mentionnés en « orange » (site alternatif d'épissage) avec débordement de l'alignement de part et d'autre de l'exon dans les séquences introniques mais ce, sans profondeur de lecture

- Anomalies apparaissant faussement normales (car en violet) et vice versa, liaison normales représentées en « bleu » (donc détectées comme « inhabituelles » par rapport aux transcrits de référence)
- Représentation d'un exon 11 complet et normal dans certains transcrits alors que celui-ci n'est même pas couvert par le design des amplicons.
- L'alignement systématique sur le transcrit référencé le plus ressemblant empêche parfois la visualisation de certaines anomalies.
- Absence de représentation des transcrits alternatifs classiquement décrits (par exemple le Δ9-10 pour *BRCA1* ou le Δ3 pour *BRCA2*)

L'alignement de NextGENe semble beaucoup trop approximatif pour permettre une détection et une quantification des différents transcrits possibles.

4. Bilan de l'analyse transcriptome

Le module « transcriptome » actuellement commercialisé par Soft Genetics dans NextGENe apparait donc clairement insuffisant pour une application diagnostique comme initialement espéré. Les données sont difficilement exploitables dans ce module : la manipulation de l'outil informatique se révèle longue et fastidieuse et l'accès au *read* séquencé en lui-même, compliqué.

Devant l'incohérence entre les anomalies attendues (car identifiées en Sanger) et celles retrouvées (ou non) en transcriptome, l'hypothèse de l'existence réelle des *read* comprenant les anomalies attendues mais non annotées par l'analyse « transcriptome » a été testée. L'alignement génomique basique des *read* a donc été effectué afin de pouvoir confirmer l'échec de l'alignement du module « transcriptome » et la validité de nos lectures.

D. Alignement génomique des read dans Next GENE : Exemple de la mutation c.9117G>A de *BRCA2*

La mutation c.9117G>A de *BRCA2* est connue pour entrainer un saut de l'exon 23.

Ce saut d'exon ainsi que le transcrit sauvage étaient visibles à l'analyse transcriptome (Figure 35).

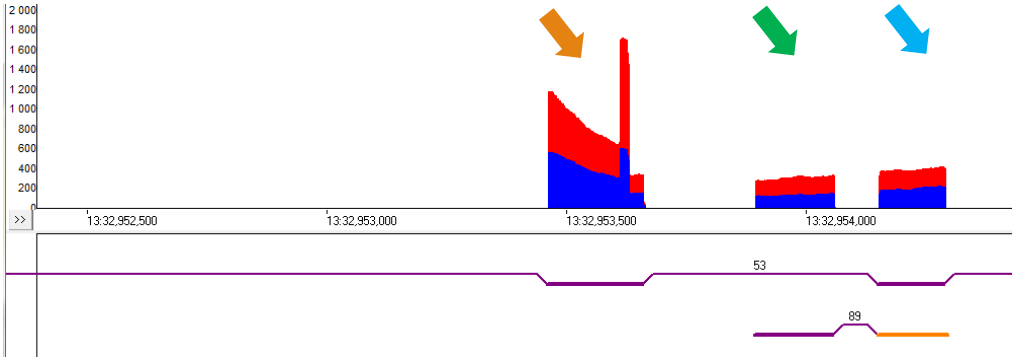


Figure 35 : Annotation des transcrits détectés pour la mutation c.9117G>A de BRCA2 entrainant un saut de l'exon 23 (flèche verte).
On note :

- *La présence du transcrit anomal sur la ligne du dessus (saut de l'exon 23, flèche verte) avec jonction directe entre les exons 22 et 24 (flèche bleue)*
- *La présence d'une partie du transcrit sauvage sur la ligne du dessous (jonction entre les exons 23 et 24)*

L'alignement sur la séquence génomique de référence de *BRCA2* a également permis de visualiser ces anomalies au niveau des *read* (Figure 36).

Figure 36 : Alignement génomique des read obtenus pour la mutation c.9117G>A
En bleu : sont surlignées les séquences correspondant au début de l'exon 24. Celles-ci parviennent à être très bien alignées sur la séquence de référence correspondante. Sur la gauche des séquences alignées sur l'exon 24, on peut suivre le reste du read que NextGENE tente d'aligner (en vain) sur la séquence intronique normalement adjacente au début de l'exon 24.
En violet : sont surlignées les séquences adjacentes correspondant à la séquence de la fin de l'exon 23
En jaune : sont surlignées les séquences adjacentes correspondant à la séquence de la fin de l'exon 22

Les anomalies d'épissage attendues ont donc bien été séquencées et sont retrouvées à l'échelle du *read*.

Il en est de même pour les transcrits alternatifs classiquement décrits : si ceux-ci étaient jamais détectés dans l'analyse « transcriptome », ils étaient pourtant tous présents dans les séquences des *read* analysés (Figure 37).

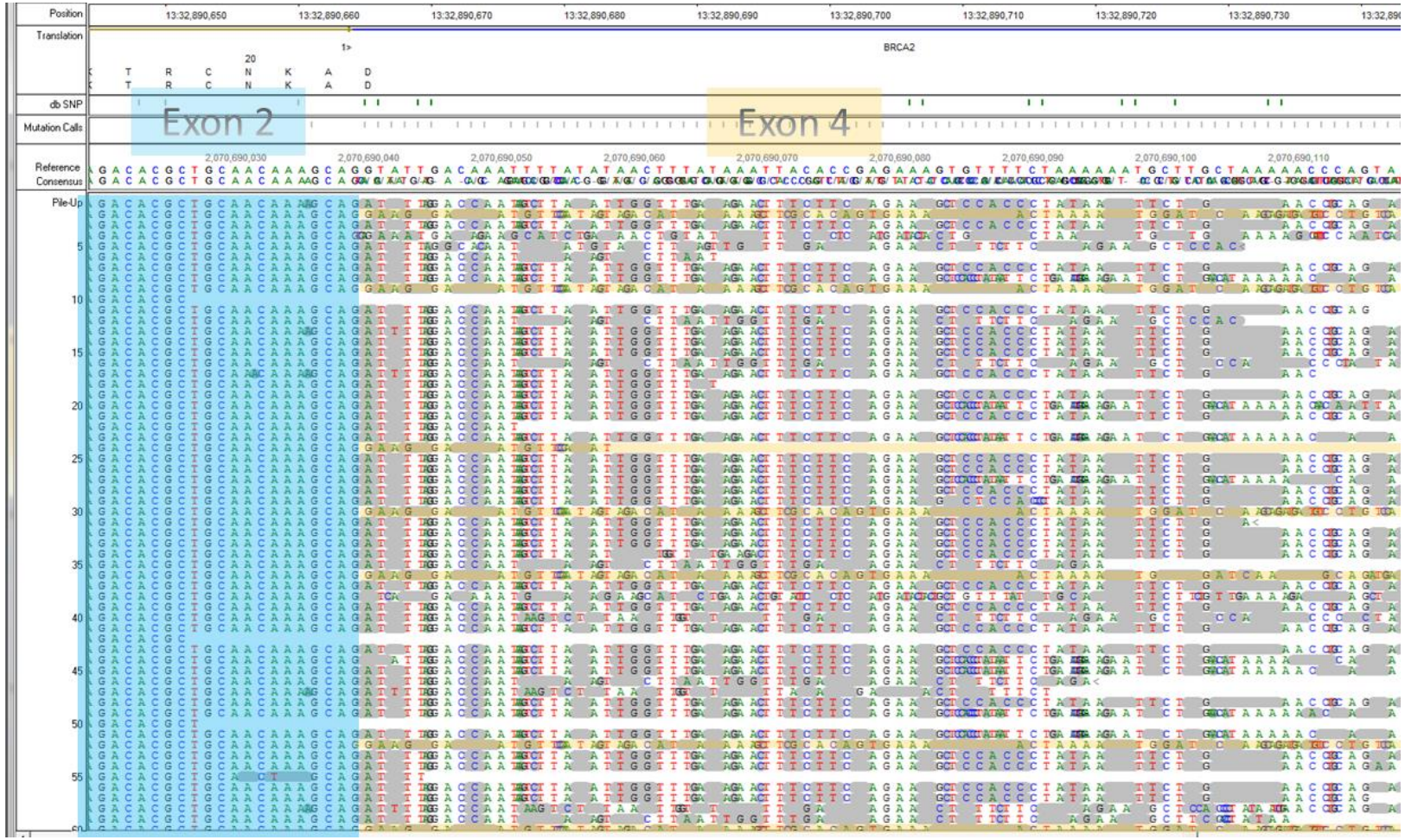


Figure 37 : Figure : Exemple du transcrit alternatif Δ exon 3 de BRCA2. Visualisation des deux formes d'épissage (exon 3 inclus ou non) à l'aide du module d'alignement génomique de Next GENe :
En bleu : sont surlignées toutes les séquences s'alignant bien sur la référence de l'exon 2
En jaune : sont surlignées les suites des read correpondant à la séquence nucléotidique de l'exon 4 (ce sont donc les lectures correspondant au transcrit alternatif Δ3 de BRCA2)

Si ces transcrits alternatifs sont donc bien présents et séquencés, ils ne sont cependant jamais représentés dans les résultats du module «transcriptome ».

Les mêmes observations ont pu être faites pour les autres transcrits alternatifs de *BRCA1* et *BRCA2*.

E. Etudes et visualisation des *read* à l'aide d'ALAMUT

Si l'alignement génomique sous Nest GENe a permis d'identifier les lectures attendues, leur visualisation reste pauvre (Figure 36 et Figure 37). Il est donc apparu intéressant d'opter pour une visualisation des lectures alignées sous Alamut (Interactive Biosoftware). Le logiciel ALAMUT permet l'affichage des séquences de *read* sur une référence de transcrits donnée. L'interface d'ALAMUT a donc été utilisée pour visualiser les résultats obtenus pour *BRCA1* et permettre de confirmer, ou non, nos réserves quant à l'efficience du module d'analyse « transcriptome » proposé par NextGENe.

1. Visualisation des mutations et de leur impact sur l'épissage : exemple de la mutation c.4484G>T entrainant un saut hors phase de l'exon 14

Dans chacun des prélèvements étudiés pour ce patient, l'anomalie d'épissage attendue était bien vérifiée puisqu'il y avait des *read* correspondant aux transcrits « normaux » (comprenant l'exon 14 à la suite du 13) et des *read* correspondant aux transcrits « anormaux » attendus (l'exon 15 succédant à l'exon 13)(Figure 38.

Par contre, aucun *read* présentant un saut combiné des exons 14 et 15 n'a pu être visualisé (contrairement à ce qui pouvait être annoté dans NextGENe, Figure 33). Ceci prouve bien les erreurs d'alignement effectuées par le module « transcriptome ».

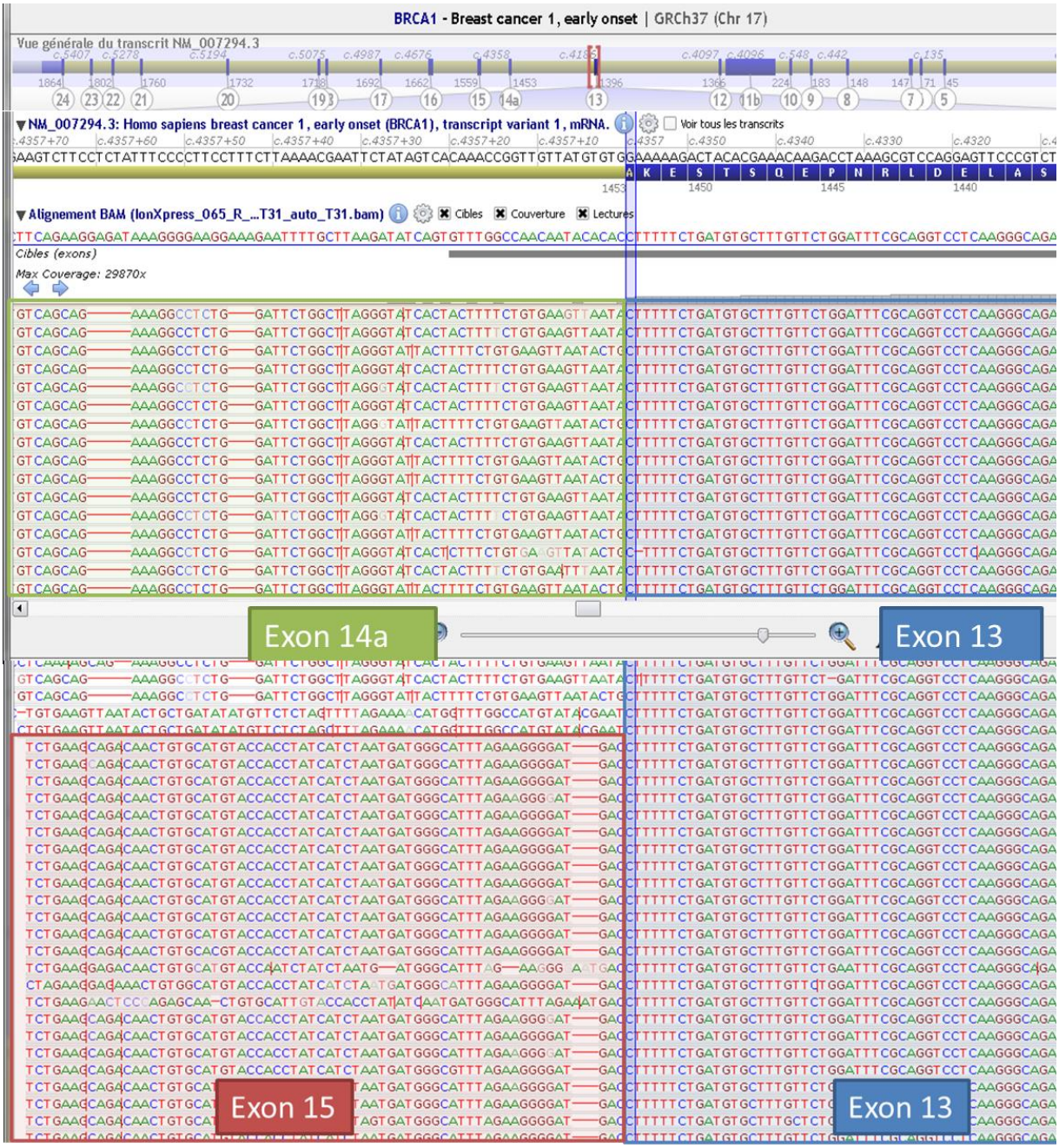


Figure 38 : visualisation des read obtenus pour la mutation c.4484G>T de BRCA1 dans ALAMUT
La 1ère ligne (« vue générale du transcrit NM_007294.3) de l'interface représente le gène BRCA1 dans son ensemble avec les exons représentés sous forme de rectangles bleus et les introns en jaune.
La région d'intérêt entre crochet rouge est développée plus précisément en dessous
La deuxième ligne (« NM_007294.3 ») représente la séquence nucléotidique de référence centrée sur le début de l'exon 13 (cf colonne bleue)
L'interface « alignement BAM » représente tous les read obtenus lors du séquençage. Ceux-ci sont alignés sur la séquence génomique de référence. Cette vue permet de suivre (sur la gauche) toutes les séquences des read s'alignant sur l'exon 13 (encadrées en bleu). Ces read se poursuivent donc soit par des séquences correspondant à celle de l'exon 14a (encadrées en vert), soit à celle de l'exon 15 (encadrées en rouge). Aucun read ne présente la succession de séquences correspondant à une jonction « exon 13-exon 16 ».

Les autres anomalies visualisées de la même manière (délétion de l'exon 3, mutation c.9117G>A de *BRCA2* entrainant un saut de l'exon 23 etc) amènent toutes à la même conclusion : les anomalies attendues sont toutes séquencées, mais ne parviennent pas à ressortir convenablement lors de l'analyse « transcriptome ».

Pour rappel, lors de l'étude du prélèvement issu du tissu mammaire sensé être sain, NextGENe ne rapportait qu'un transcrit anormal avec saut de l'exon 3 pour ce prélèvement (Figure 32). Cependant, lorsqu'on aligne ces mêmes *read* à l'aide d'Alamut, on retrouve bien les deux types différents de transcrits (avec et sans l'exon 3). L'analyse transcriptome de NextGENe a donc été une fois de plus prise en défaut et il s'agissait bien de tissu mammaire sain (présence des transcrits sauvages et anormaux).

2. Visualisation des transcrits alternatifs classiquement décrits pour les gènes *BRCA1* et *BRCA2*

Ces transcrits sont bien retrouvés lors de l'alignement des séquences dans Alamut. Ils ressortent dans des proportions attendues, quelle que soit la nature de l'échantillon initial. Malgré une réalité physique indéniable de ces transcrits alternatifs, le module transcriptome n'arrive donc pas à les représenter dans son analyse finale.

3. Autres anomalies détectées : visualisations répétées d'inclusion de séquences introniques pouvant s'apparenter à des « pseudo-exons »

A l'alignement des *read* obtenus pour les différents patients dans Alamut, plusieurs rétentions de séquences introniques ont été observées. Elles apparaissent sur quelques *read* avec une profondeur de couverture plus faible (taux variant entre 30x et 250x environ pour des profondeurs atteignant les 2000X à 5000X pour les exons flanquants)(Figure 39).

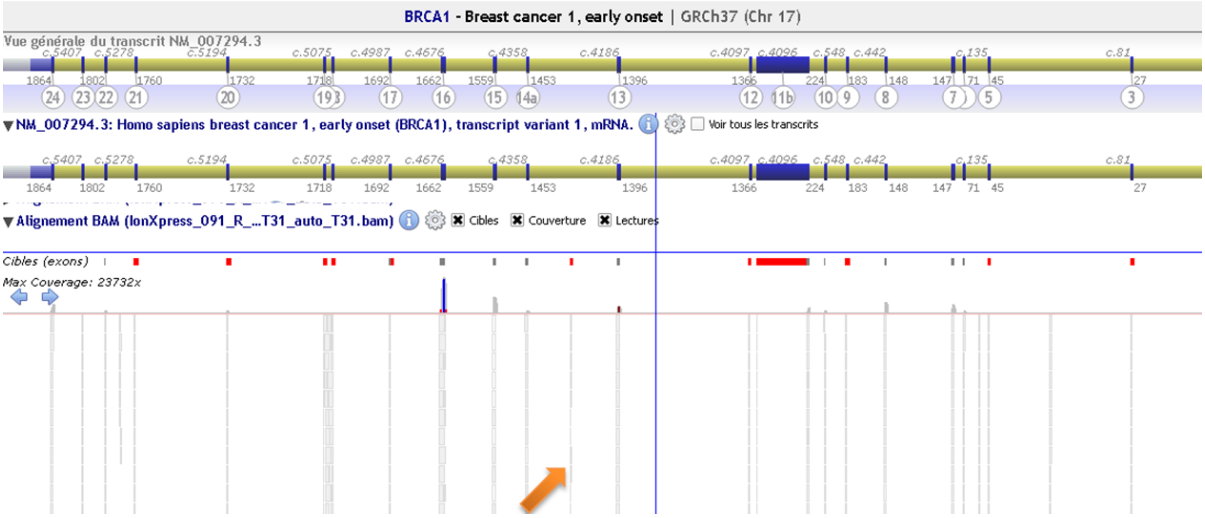


Figure 39 : Visualisation de la répartition de l'ensemble des read obtenus pour un type de prélèvement (ici il s'agit au départ tu tissu mammaire sain de la delétion génomique de l'exon 3). La flèche orange indique la présence de read alignés sur l'intron 13.

Ces séquences s'apparentent à des pseudos exons, avec la présence de sites donneurs et accepteurs potentiels de part et d'autre (aux scores prédictifs généralement élevés)(Figure 40). Elles apparaissent de manière plus ou moins variable d'un patient à un autre mais :

- Les différents types de rétention introniques sont retrouvés selon le même schéma dans les 3 prélèvements pour chaque patient.
- Proportionnellement, elles apparaissent généralement un peu plus abondantes dans le prélèvement traité par puromycine (en gardant bien entendu à l'esprit que cette technique ne permet pas dans ce cas une approche quantitative). Ceci est en adéquation avec le fait que le NMD y est inhibé et que les transcrits alternatifs devraient donc y être détectés de manière plus importante.
- Certaines d'entre elles sont retrouvées fréquemment parmi les différents patients analysés (par exemple la rétention intronique survenant entre les exons 13 et 14).

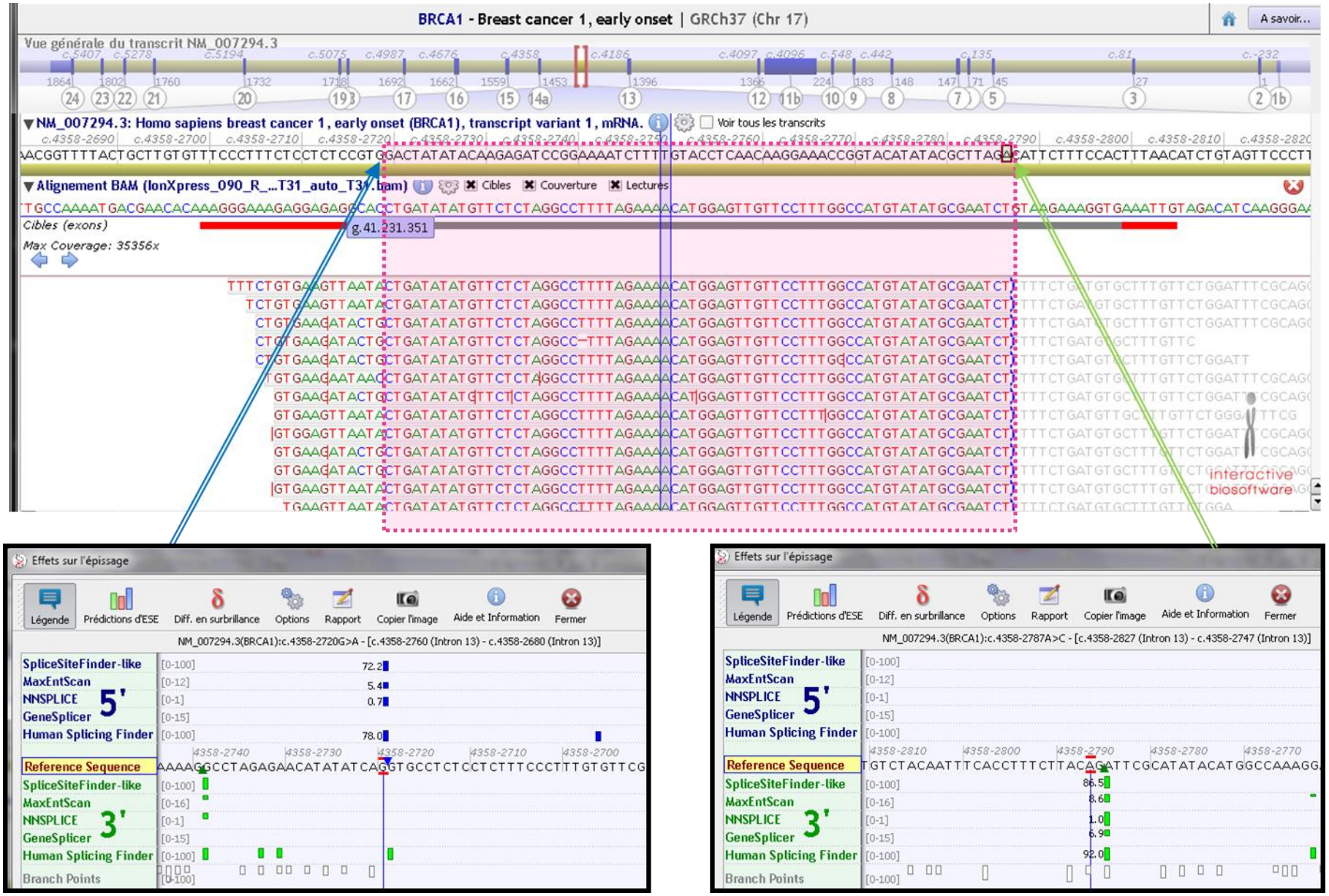

BRCA1 - Breast cancer 1, early onset | GRCh37 (Chr 17)
A savoir...
Vue générale du transcrit NM_007294.3
NM_007294.3: Homo sapiens breast cancer 1, early onset (BRCA1), transcript variant 1, mRNA.
Voir tous les transcrits
Alignement BAM (IonXpress_090_R_...T31_auto_T31.bam)
Cibles
Couverture
Lectures
Cibles (exons)
g.41.231.351
Max Coverage: 35356x
interactive biosoftware
Effets sur l'épissage
Légende
Prédictions d'ESE
Diff. en surbrillance
Options
Rapport
Copier l'image
Aide et Information
Fermer
NM_007294.3(BRCA1):c.4358-2720G>A - [c.4358-2760 (Intron 13) - c.4358-2680 (Intron 13)]
SpliceSiteFinder-like
MaxEntScan
NNSPLICE
GeneSpliceer
Human Splicing Finder
5'
3'
Reference Sequence
Branch Points
NM_007294.3(BRCA1):c.4358-2787A>C - [c.4358-2827 (Intron 13) - c.4358-2747 (Intron 13)]

Figure 40 : exemple d' « exonisation » d'une séquence intronique dans les transcrits issus de BRCA1 (exemple de la délétion génomique de l'exon 3) visualisée dans Alamut

- *En rose : sont encadrées les séquences des read s'alignant sur la séquence intronique en question. L'encadrement comprends inclut également la séquence de référence pour faciliter la visualisation des début et fin du « pseudo-exon »*

En bas, on observe deux fenêtres d'analyse de l' « effet sur l'épissage » correspondant aux deux extrémités de la séquence « exonisée ». Cette fenêtre de prédiction a pour but d'appréhender d'éventuels sites d'épissage cryptiques en résumant les scores attribués à une position donnée par différents outils prédictifs de sites d'épissage.

- *A gauche (flèche bleue) : on note la présence d'un potentiel site accepteur en 3' (de score prédictifs forts) à l'extrémité de la séquence exonisée*
- *A droite (flèche verte) : on note la présence d'un potentiel site donneur en 5' (de scores prédictifs élevés) à l'extrémité de la séquence exonisée.*

Ces rétentions introniques, ou « pseudo-exons » posent donc la question de leur réalité physique et de leur significativité biologique. Il pourrait bien s'agir de véritables variants d'épissage qui jusque-là ne pouvaient pas être détectés par Sanger car présents en proportion trop faible et donc noyés dans le bruit de fond. Il faudrait bien entendu pouvoir pousser les investigations pour pouvoir affiner leur description, asseoir leur réelle existence et déterminer quelle importance physiologique leur donner. Il sera donc essentiel de vérifier lors des analyses à venir si ces séquences ressortent dans les études par NGS futures.

IV. Conclusions et perspectives

Les premiers résultats obtenus par séquençage des ADNc par NGS sont encourageants. En effet, les anomalies d'épissage attendues sont bien séquencées comme telles. Malheureusement les outils d'analyse ne permettent pas dans l'état actuel des choses de les visualiser et de les analyser. La plateforme de bio-informatique de l'institut Curie a repris ces séquences afin de mettre au point un autre alignement des données issues de séquençage de ADNc plus efficace et permettre une visualisation plus aisée des transcrits détectés (Alamut ou autre).

Les difficultés d'interprétation des *read* issus du séquençage des ADNc sont inhérentes aux spécificités de la technique de séquençage à haut débit. Les séquences ne dépassant pas des tailles de 100 à 200 bases, il est difficile de savoir quelle séquence appartient à quel variant d'épissage et de parvenir à reconstruire les transcrits existants.

Une approche certainement intéressante à tester serait d'étudier les mutations affectant l'épissage en ciblant plusieurs amplicons dans la région d'intérêt (contrairement à la couverture exhaustive du gène qui avait été privilégiée ici). On obtiendrait ainsi une profondeur de couverture plus importante au niveau de la région cible avec des chevauchements multiples pouvant faciliter l'analyse bio-informatique et la reconstruction des transcrits. Le fait de retrouver à chaque fois les anomalies attendues au niveau des read prouve bien que l'on parvient à séquencer ce que l'on espérait observer. Par contre, un ajustement de la conception des amplicons et des banques semble inévitable pour parvenir à affiner ce type d'analyse.

Le design des amplicons ici utilisé n'était sans doute pas optimal et la préparation des banques pourraient être optimisée en favorisant des amplicons de grandes tailles se chevauchant sur plusieurs exons et en préparant également des banques de fragments plus longs (séquençage à l'aide de kits 400 pb).

Une mise au point de la technique devrait ainsi pouvoir permettre d'étudier l'impact des variants d'épissage en s'affranchissant du traitement des prélèvements par la puromycine. En effet, nous avons pu observer que les anomalies étaient détectées dans les échantillons non traités (ARN issus de lignées et de tissus mammaires) et qu'elles pouvaient l'être à un seuil suffisant pour éventuellement envisager la suppression de cette étape de prise en charge des prélèvements ARN. La présence résiduelle d'anomalies d'épissage détectées par NGS dans les prélèvements non traités par la puromycine signifie également que le NMD n'est pas complet. L'absence de transcrits anormaux généralement décrite dans les analyses de séquençage Sanger à partir des prélèvements non traités par puromycine serait donc en réalité la conséquence d'un défaut de sensibilité de la technique Sanger dans la détection de transcrits présents en faible quantité.

L'objectif de pouvoir quantifier les transcrits alternatifs n'est pas envisageable avec cette approche PGM dans l'état actuel de préparation des banques : les qualités de séquençage et les zones de couverture sont beaucoup trop variables d'un amplicon à un autre, mais aussi d'un patient à un autre. D'autres approches en projet pourraient cependant limiter les étapes de préparation des banques ou, du moins, les uniformiser et peut-être permettre une approche beaucoup plus quantitative.

Deux autres projets sur ce sujet sont donc en cours (Figure 41). Malheureusement, aucune donnée provenant de ces autres approches n'est disponible au moment de la rédaction de cette thèse.

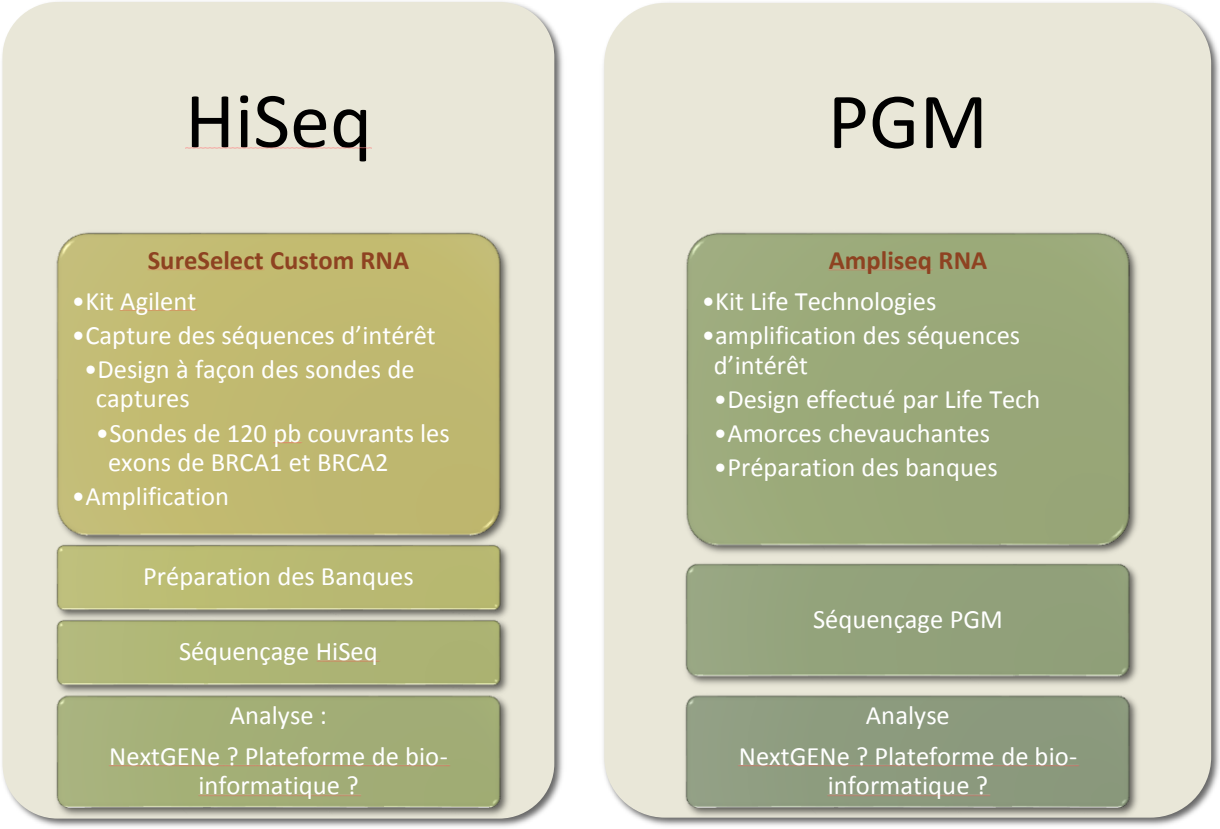


Figure 41 : Autres projets en cours.

A. Projet « Sure Select » (Agilent)

Le kit RNA Sure Select custom d'Agilent permet la capture et l'amplification des ARN d'intérêt avec un design des sondes de capture réalisé à façon. La taille de ces sondes est fixée par le fournisseur à 120 pb et n'est pas modulable.

Afin de n'induire aucun biais de sélection de certaines formes d'épissage, les sondes de capture ont été établies à partir des séquences génomiques de chaque exon et aucune séquence chevauchant les jonctions « exon-exon » n'a été prise en compte. En raison de l'existence de certains exons posant un problème de taille inférieure à celle proposée pour les sondes (<120 pb), les séquences cibles ont été élargies de -75 nt à + 75 nt de part et d'autre

de chaque exon. Cependant, après discussion avec l'équipe d'Agilent, la sensibilité de ces sondes devraient permettre la capture des exons cibles, même si l'hybridation est partielle.

Les banques ainsi préparées seront ensuite séquencées à l'aide du séquenceur HiSeq d'Illumina puis analysées à l'aide de ressources disponibles via la plateforme de bio-informatiques. Les résultats obtenus pourront ensuite être visualisés dans Alamut.

B. Projet « AmpliSeq RNA » (Life Technologies)

Le kit commercialisé par Life Technologies permet l'amplification des ARN d'intérêt et la préparation simultanée des banques par ligation directe des codes-barres.

Le design des primers d'amplification a été réalisé par une équipe de Life Technologies (« *White Glove* »). Contrairement à l'approche précédente, elle a placé des amorces tout le long de la séquence ADNc en autorisant des amorces chevauchantes « exon-exon » et la construction d'amplicons également chevauchants.

Le séquençage sera effectué par PGM et l'approche d'analyse bio-informatique reste aussi à préciser.

Conclusion Générale

Le paysage de la génétique subit actuellement une révolution importante avec l'arrivée récente des nouvelles technologies diagnostiques que propose le NGS. Parallèlement à cela, on observe une augmentation croissante des demandes de tests génétiques dans le cadre des prédispositions génétiques aux cancers.

Un rapport récent du plan cancer insiste sur la nécessité d'observer des délais de rendus de résultats de plus en plus restreints afin d'optimiser et d'adapter la prise en charge du patient et de ses proches. L'application de la technologie de NGS en diagnostic de routine s'inscrit pleinement dans ces objectifs. La mise à profit de cet outil dans un cadre diagnostique aux critères de qualité très stricts n'est cependant pas chose aisée et nécessite un investissement et des réajustements permanents. En permettant une approche diagnostique plus globale et précise, elle aboutit à la découverte de nombreux variants actuellement de signification inconnue qui représentent un enjeu majeur dans l'interprétation des résultats. Les méthodes d'analyse des transcrits et d'évaluation de l'impact de certains variants que nous essayons de mettre au point à partir de techniques plus rapides et moins couteuses que le RNAseq pourraient être très utiles dans l'interprétation d'une partie de ces variants.

Ce projet s'avère avoir été un peu vaste pour être traité et cerné en un sujet de thèse et sur un seul semestre. Mais à la vue des premiers résultats prometteurs et compte tenus des enjeux diagnostiques auxquels il pourrait répondre, il mérite réellement d'être poursuivi et approfondi.

L'établissement de liens de causalité entre les variants détectés, leur impact sur la transcription ou sur la protéine et leur causalité effective nécessitera bien entendu de poursuivre également des études épidémiologiques en parallèle.

Dans ce champs d'application, le biologiste se retrouve plus que jamais au centre d'enjeux et de dialogues clinico-biologiques. Chaque interprétation de variant doit se faire à la lumière du contexte clinique (familial et épidémiologique) et des connaissances scientifiques du moment. Celle-ci est réalisée à l'aide d'outils de plus en plus performants mais également de manipulation de plus en plus complexe.

Il apparait donc essentiel d'instaurer une formation adaptée en bio-informatique auprès des professionnels médicaux concernés. Si un support bio-informatique spécialisé (dédié ou délocalisé) reste indispensable, le biologiste confronté à l'utilisation d'outils NGS est obligé d'acquérir des notions solides à ce sujet pour pouvoir utiliser cette technologie de manière optimale. Cette facette désormais inévitable dans l'utilisation des outils de diagnostic génétique de nouvelle génération mérite une attention grandissante dès l'étape de formation des internes en génétique (aussi bien biologistes que médicaux) et il apparait urgent de pouvoir proposer des enseignements de bio-informatiques plus adaptées aux besoins futurs de notre spécialité.

Bibliographie

1. Grizzi, F. & Chiriva-Internati, M. Cancer: looking for simplicity and finding complexity. *Cancer Cell Int.* **6,** 4 (2006).
2. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nat. Med.* **10,** 789–799 (2004).
3. Houdayer, C. & Stoppa-Lyonnet, D. Anomalies de la transcription et diagnostic en génétique constitutionnelle. *MS Médecine Sci.* **21,** 170–174 (2005).
4. Stoppa-Lyonnet, D., Stern, M. H., Soufir, N. & Lenoir, G. [Cancer genetic predisposition: current events and perspectives in 2010]. *Pathol. Biol. (Paris)* **58,** 324–330 (2010).
5. Garnis, C., Buys, T. P. H. & Lam, W. L. Genetic alteration and gene expression modulation during cancer progression. *Mol. Cancer* **3,** 9 (2004).
6. Sawan, C., Vaissière, T., Murr, R. & Herceg, Z. Epigenetic drivers and genetic passengers on the road to cancer. *Mutat. Res.* **642,** 1–13 (2008).
7. Ferlay, J., Parkin, D. M. & Steliarova-Foucher, E. Estimates of cancer incidence and mortality in Europe in 2008. *Eur. J. Cancer Oxf. Engl. 1990* **46,** 765–781 (2010).
8. Binder-Foucard. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012.
9. Stoppa-Lyonnet, D. & Lenoir, G. Prédispositions génétiques aux cancers : actualités et perspectives en 2005. *MS Médecine Sci.* **21,** 962–968 (2005).
10. Mavaddat, N. *et al.* Cancer risks for BRCA1 and BRCA2 mutation carriers: results from prospective analysis of EMBRACE. *J. Natl. Cancer Inst.* **105,** 812–822 (2013).
11. Bartsch, D. K., Gress, T. M. & Langer, P. Familial pancreatic cancer--current knowledge. *Nat. Rev. Gastroenterol. Hepatol.* **9,** 445–453 (2012).
12. Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266,** 66–71 (1994).
13. Szabo, C. I. *et al.* Human, canine and murine BRCA1 genes: sequence comparison among species. *Hum. Mol. Genet.* **5,** 1289–1298 (1996).
14. Wooster, R. *et al.* Identification of the breast cancer susceptibility gene BRCA2. *Nature* **378,** 789–792 (1995).
15. Tavtigian, S. V. *et al.* The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nat. Genet.* **12,** 333–337 (1996).
16. Wong, A. K., Pero, R., Ormonde, P. A., Tavtigian, S. V. & Bartel, P. L. RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2. *J. Biol. Chem.* **272,** 31941–31944 (1997).
17. Esashi, F. *et al.* CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* **434,** 598–604 (2005).
18. Chodosh, L. A. Expression of BRCA1 and BRCA2 in normal and neoplastic cells. *J. Mammary Gland Biol. Neoplasia* **3,** 389–402 (1998).
19. Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12,** 68–78 (2012).
20. Foulkes, W. D. & Shuen, A. Y. In brief: BRCA1 and BRCA2. *J. Pathol.* **230,** 347–349 (2013).

21. Anderson, S. F., Schlegel, B. P., Nakajima, T., Wolpin, E. S. & Parvin, J. D. BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. *Nat. Genet.* **19,** 254–256 (1998).
22. Horwitz, A. A., Affar, E. B., Heine, G. F., Shi, Y. & Parvin, J. D. A mechanism for transcriptional repression dependent on the BRCA1 E3 ubiquitin ligase. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 6614–6619 (2007).
23. Suwaki, N., Klare, K. & Tarsounas, M. RAD51 paralogs: roles in DNA damage signalling, recombinational repair and tumorigenesis. *Semin. Cell Dev. Biol.* **22,** 898–905 (2011).
24. Caputo, S. *et al.* Description and analysis of genetic variants in French hereditary breast and ovarian cancer families recorded in the UMD-BRCA1/BRCA2 databases. *Nucleic Acids Res.* **40,** D992–1002 (2012).
25. The UMD central database. at <http://www.umd.be/UMD_Central/>
26. Ferla, R. *et al.* Founder mutations in BRCA1 and BRCA2 genes. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. ESMO* **18 Suppl 6,** vi93–98 (2007).
27. COsegregation of VARiants in the BRCA1/2 Genes - Full Text View - ClinicalTrials.gov. at <http://clinicaltrials.gov/show/NCT01689584>
28. Ripperger, T., Gadzicki, D., Meindl, A. & Schlegelberger, B. Breast cancer susceptibility: current knowledge and implications for genetic counselling. *Eur. J. Hum. Genet. EJHG* **17,** 722–731 (2009).
29. Casadei, S. *et al.* Contribution of inherited mutations in the BRCA2-interacting protein PALB2 to familial breast cancer. *Cancer Res.* **71,** 2222–2229 (2011).
30. Rahman, N. *et al.* PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat. Genet.* **39,** 165–167 (2007).
31. Meindl, A. *et al.* Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nat. Genet.* **42,** 410–414 (2010).
32. Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat. Genet.* **45,** 353–361, 361e1–2 (2013).
33. Melchor, L. & Benítez, J. The complex genetic landscape of familial breast cancer. *Hum. Genet.* **132,** 845–863 (2013).
34. Corcos, L. & Solier, S. Épissage alternatif, pathologie et thérapeutique moléculaire. *MS Médecine Sci.* **21,** 253–260 (2005).
35. Burge: 20 Splicing of Precursors to mRNAs by the Spliceosomes. at <http://cshmonographs.org/index.php/monographs/article/view/5123>
36. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* **3,** 285–298 (2002).
37. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72,** 291–336 (2003).
38. Garg, K. & Green, P. Differing patterns of selection in alternative and constitutive splice sites. *Genome Res.* **17,** 1015–1022 (2007).
39. Sorek, R. & Ast, G. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13,** 1631–1637 (2003).
40. Kim, N. & Lee, C. Bioinformatics detection of alternative splicing. *Methods Mol. Biol. Clifton NJ* **452,** 179–197 (2008).
41. Zheng, C. L., Fu, X.-D. & Gribskov, M. Characteristics and regulatory elements defining constitutive splicing and different modes of alternative splicing in human and mouse. *RNA New York N* **11,** 1777–1787 (2005).

42. Holste, D. & Ohler, U. Strategies for identifying RNA splicing regulatory motifs and predicting alternative splicing events. *PLoS Comput. Biol.* **4,** e21 (2008).
43. Houdayer, C. In silico prediction of splice-affecting nucleotide variants. *Methods Mol. Biol. Clifton NJ* **760,** 269–281 (2011).
44. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23,** 198–199 (1998).
45. Silva, A. L. & Romão, L. The mammalian nonsense-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision? *FEBS Lett.* **583,** 499–505 (2009).
46. Andreutti-Zaugg, C., Scott, R. J. & Iggo, R. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. *Cancer Res.* **57,** 3288–3293 (1997).
47. Orban, T. I. & Olah, E. Emerging roles of BRCA1 alternative splicing. *Mol. Pathol. MP* **56,** 191–197 (2003).
48. Consortium, I. H. G. S. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
49. Metzker, M. L. Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11,** 31–46 (2010).
50. Orlando, L. *et al.* True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Res.* **21,** 1705–1719 (2011).
51. Egan, A. N., Schlueter, J. & Spooner, D. M. Applications of next-generation sequencing in plant biology. *Am. J. Bot.* **99,** 175–185 (2012).
52. Fonseca, N. A., Rung, J., Brazma, A. & Marioni, J. C. Tools for mapping high-throughput sequencing data. *Bioinformatics* bts605 (2012). doi:10.1093/bioinformatics/bts605
53. Nielsen, R., Paul, J. S., Albrechtsen, A. & Song, Y. S. Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* **12,** 443–451 (2011).
54. Software/list - SEQwiki. at <http://seqanswers.com/wiki/Software/list>
55. Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nat. Rev. Genet.* **12,** 671–682 (2011).
56. Survol de la technologie de séquençage PGM Ion Torrent | biorigami. at <http://www.biorigami.com/?p=4643>
57. Tarabeux, J. *et al.* Streamlined ion torrent PGM-based diagnostics: BRCA1 and BRCA2 genes as a model. *Eur. J. Hum. Genet. EJHG* (2013). doi:10.1038/ejhg.2013.181

Annexes

1. Protocole de RT-PCR utilisé pour l'étude des ARN de *BRCA1*

	CI	CF	µL (X 1)
BUFFER	5X	1X	4
DTT	0,1M	0,01M	2
DNTP	5mM each	0,5mM each	2
RANDOM PRIMERS	0,5µg/µL	0,15µg/µL (1X)	6 (ou 1 à 6X)
SUPERSCRIPT II RT	200U/µL	100U	0,5
RNASIN	40U/µL	20U	0,5
MIX			15 (ou 10)
ARN (1µG)			5 (ou 10)
FINAL VOLUME			20

Conditions Thermocycleur

TEMPERATURE	TEMPS
25	10
42	30
99	5
4	∞

2. Protocole de RT-PCR utilisé pour L'étude des ARN de *BRCA2*

	CONCENTRATION FINALE	**VOLUME (µL)**
MGCL2 (25MM)	3,5 mM	**5,6**
BUFFER+MGCL2 (10X+15MM)	1x+1,5mM	**4**
ARN	2µg/40µL	
H20	maxi 8,4	
DNTP (10MM)	4mM	**16**
RNASE INHIBITOR (20U/µL)	1U/µL	**2**
MULV RT (50U/µL)	2,5U/µL	**2**
RANDOM HEXAMERS (50µM)	2,5 µM	**2**
VOLUME FINAL		**40**

Conditions Thermocycleur

TEMPERATURE	TEMPS
42°C	15 min
99°C	5 min
5°C	5 min
4°C	∞

3. Design des Amorces utilisées pour l'amplification « maison » des ADNc de BRCA1 et BRCA2

	BRCA1									
ions des pri	séquences des primers									
5'utr	CACCCTCTGCTCTGGGTAAA	Start: 193	Length: 20 bp	Tm: 60.2 °C	GC: 55.0 %	ANY: 4.0	SELF: 1.0	Product Size: 578 bp	Pair Any: 5.0	Pair End: 2.0
fin ex8	TGTAGACAGACGTCTTTTGAGGTT	Start: 770	Length: 24 bp	Tm: 59.4 °C	GC: 41.7 %	ANY: 8.0	SELF: 0.0			
fin_ex6	TCAGCTTGACACAGGTTTGG	Start: 511	Length: 20 bp	Tm: 59.9 °C	GC: 50.0 %	ANY: 4.0	SELF: 0.0	Product Size: 542 bp	Pair Any: 4.0	Pair End: 0.0
deb_ex11	ATGGCTCCACATGCAAGTTT	Start: 1052	Length: 20 bp	Tm: 60.5 °C	GC: 45.0 %	ANY: 6.0	SELF: 2.0			
fin_ex11	GAACGGGCTTGGAAGAAAAT	Start: 4275	Length: 20 bp	Tm: 60.4 °C	GC: 45.0 %	ANY: 4.0	SELF: 2.0	Product Size: 514 bp	Pair Any: 5.0	Pair End: 1.0
ex15	TTCTGAAGACTCCCAGAGCAA	Start: 4788	Length: 21 bp	Tm: 60.1 °C	GC: 47.6 %	ANY: 5.0	SELF: 0.0			
ex14	TGCTGACAAGTTTGAGGTGTCT	Start: 4651	Length: 22 bp	Tm: 60.0 °C	GC: 45.5 %	ANY: 4.0	SELF: 1.0	Product Size: 600 bp	Pair Any: 7.0	Pair End: 0.0
deb_ex17	TGGTGTTTTCTGGCAAACTT	Start: 5250	Length: 20 bp	Tm: 58.3 °C	GC: 40.0 %	ANY: 6.0	SELF: 3.0			
deb_ex16	ACCCCTTACCTGGAATCTGG	Start: 4913	Length: 20 bp	Tm: 60.2 °C	GC: 55.0 %	ANY: 3.0	SELF: 1.0	Product Size: 567 bp	Pair Any: 3.0	Pair End: 1.0
fin_ex20	TGGACCTTGGTGGTTTCTTC	Start: 5479	Length: 20 bp	Tm: 59.9 °C	GC: 50.0 %	ANY: 4.0	SELF: 0.0			
fin_ex18	CGGGAGGAAAATGGGTAGTT	Start: 5355	Length: 20 bp	Tm: 60.2 °C	GC: 50.0 %	ANY: 2.0	SELF: 0.0	Product Size: 518 bp	Pair Any: 4.0	Pair End: 0.0
3utr	AAGCTCATTCTTGGGGTCCT	Start: 5872	Length: 20 bp	Tm: 60.1 °C	GC: 50.0 %	ANY: 4.0	SELF: 1.0			
	BRCA2									
ex2	CACCTCTGGAGCGGACTTAT	Start: 175	Length: 20 bp	Tm: 59.3 °C	GC: 55.0 %	ANY: 4.0	SELF: 2.0	Product Size: 2190 bp	Pair Any: 5.0	Pair End: 2.0
deb_ex11	GCAGGCATGACAGAGAATCA	Start: 2364	Length: 20 bp	Tm: 60.0 °C	GC: 50.0 %	ANY: 4.0	SELF: 1.0			
fin_ex11	ACTGCCAAGTCATGCCACAC	Start: 6959	Length: 20 bp	Tm: 62.2 °C	GC: 55.0 %	ANY: 5.0	SELF: 1.0	Product Size: 2192 bp	Pair Any: 4.0	Pair End: 1.0
ex_22	CAATACGCAACTTCCACACG	Start: 9150	Length: 20 bp	Tm: 60.2 °C	GC: 50.0 %	ANY: 3.0	SELF: 2.0			
ex_18	ATGGAAAGGGATGACACAGC	Start: 8253	Length: 20 bp	Tm: 59.9 °C	GC: 50.0 %	ANY: 3.0	SELF: 2.0	Product Size: 2278 bp	Pair Any: 4.0	Pair End: 0.0
ex_27	CTGGAAAGGTTAAGCGTCAA	Start: 10530	Length: 20 bp	Tm: 58.0 °C	GC: 45.0 %	ANY: 4.0	SELF: 2.0			

4. Protocoles de séquençage à haut débit par PGM – laboratoire de Génétique – Institut Curie

institutCurie
Ensemble, prenons le cancer de vitesse

AMPLIFICATION BRCA POUR SEQUENCAGE HAUT-DEBIT

Objet : *Enrichissement des séquences codantes et des jonctions intron/exon des gènes* BRCA1 *et* BRCA2 *par PCR multiplexes, avec le kit Multiplicom*

1. Domaine d'application

(Ce paragraphe précise les activités, services, produits, personnels, secteurs, périodes et/ou clients auxquels s'applique le mode protocole)

Ce protocole concerne…

2. Responsabilités

(Ce paragraphe précise les responsabilités concernant l'application du protocole et la vérification de sa bonne application)

3. Description détaillée

Le processus complet de diagnostic des gènes BRCA sur PGM se déroule en 4 phases :

I - Enrichissement > ce protocole décrit la procédure d'enrichissement utilisée avec le kit Multiplicom

http://www.multiplicom.com/products/brca-mastr-v21

II - Préparation des banques

III - Séquençage (PCRe + PGM)

IV - Analyse

Enrichissement (Kit Multiplicom)

1/ PCR multiplex

1) Doser les ADNs stock après leur décongélation au Nanodrop.
   Préparer sur une microplaque des ADNs dilués à 10ng/µl.

2) Mettre les 5 tubes de Primer mix (numérotés de 1 à 5) du kit BRCA MASTR Dx à décongeler sur la glace (environ 1h)

Note : Les kits BRCA MASTR Dx contiennent en plus des 5 Primer mix des tubes de réactif d'amplification supplémentaire (Amplification Reagent = AR1) à utiliser pour la PCR universelle.

Sous la hotte

3) Bien vortexer puis centrifuger les tubes de Primer mix

4) Préparer un mix pour chaque tube de primer (soit 5 mix au total), selon les proportions suivantes en fonction du nombre d'ADN à amplifier dans la série :

Réaction	**x1**	**x8,8**	**x17**	**x33**
Primer mix (µl)	10	88	170	330
Taq (µl)	0,075	0,66	1,275	2,475

Sur la paillasse ADN

5) Distribuer 50 ng d'ADN (soit 5µl d'ADN à 10ng/µl) pour chaque mix dans une plaque PCR.

Cette étape peut être effectuée au Robot TECAN (en Pré-PCR) : Distribuer 10 µl de chaque mix par échantillon

Cette étape peut être effectuée au Robot (en Post-PCR) :

Exemple de répartition des ADNs et des Mixs (notés M1 à M5) pour une série de 32 cas sur 2 plaques (amplification d'1 mix complet sur 1 plaque, pour éviter les biais d'amplification en fonction du thermocycleur):

Plaque 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	ADN1 M1	ADN9 M1	ADN17 M1	ADN25 M1	ADN1 M2	ADN9 M2	ADN17 M2	ADN25 M2	ADN1 M3	ADN9 M3	ADN17 M3	ADN25 M3
B	ADN2 M1	ADN10 M1	ADN18 M1	ADN26 M1	ADN2 M2	ADN10 M2	ADN18 M2	ADN26 M2	ADN2 M3	ADN10 M3	ADN18 M3	ADN26 M3

Plaque 2

	1	2	3	4	5	6	7	8
A	ADN1 M4	ADN9 M4	ADN17 M4	ADN25 M4	ADN1 M5	ADN9 M5	ADN17 M5	ADN2 M5
B	ADN2 M4	ADN10 M4	ADN18 M4	ADN26 M4	ADN2 M5	ADN10 M5	ADN18 M5	ADN2 M5

C	ADN3 M1	ADN11 M1	ADN19 M1	ADN27 M1	ADN3 M2	ADN11 M2	ADN19 M2	ADN27 M2	ADN3 M3	ADN11 M3	ADN19 M3	ADN27 M3
D	ADN4 M1	ADN12 M1	ADN20 M1	ADN28 M1	ADN4 M2	ADN12 M2	ADN20 M2	ADN28 M2	ADN4 M3	ADN12 M3	ADN20 M3	ADN28 M3
E	ADN5 M1	ADN13 M1	ADN21 M1	ADN29 M1	ADN5 M2	ADN13 M2	ADN21 M2	ADN29 M2	ADN5 M3	ADN13 M3	ADN21 M3	ADN29 M3
F	ADN6 M1	ADN14 M1	ADN22 M1	ADN30 M1	ADN6 M2	ADN14 M2	ADN22 M2	ADN30 M2	ADN6 M3	ADN14 M3	ADN22 M3	ADN30 M3
G	ADN7 M1	ADN15 M1	ADN23 M1	TM 1 M1	ADN7 M2	ADN15 M2	ADN23 M2	TM 1 M2	ADN7 M3	ADN15 M3	ADN23 M3	TM 1 M3
H	ADN8 M1	ADN16 M1	ADN24 M1	TM 2 M1	ADN8 M2	ADN16 M2	ADN24 M2	TM 2 M2	ADN8 M3	ADN16 M3	ADN24 M3	TM 2 M3

C	ADN3 M4	ADN11 M4	ADN19 M4	ADN27 M4	ADN3 M5	ADN11 M5	ADN19 M5	ADN2 M5
D	ADN4 M4	ADN12 M4	ADN20 M4	ADN28 M4	ADN4 M5	ADN12 M5	ADN20 M5	ADN2 M5
E	ADN5 M4	ADN13 M4	ADN21 M4	ADN29 M4	ADN5 M5	ADN13 M5	ADN21 M5	ADN2 M5
F	ADN6 M4	ADN14 M4	ADN22 M4	ADN30 M4	ADN6 M5	ADN14 M5	ADN22 M5	ADN3 M5
G	ADN7 M4	ADN15 M4	ADN23 M4	TM 1 M4	ADN7 M5	ADN15 M5	ADN23 M5	TM 1 M5
H	ADN8 M4	ADN16 M4	ADN24 M4	TM 2 M4	ADN8 M5	ADN16 M5	ADN24 M5	TM 2 M5

6) Lancer la PCR

98°C - 10 min
95°C - 45 s
60°C - 45 s
68°C - 2 min
} X20
72°C - 10 min
4°C - ∞

Important : Après PCR, conserver les réactions multiplexes à -20°C, ou enchainer immédiatement sur la PCR universelle.

2/ PCR Universelle

1) Mettre les tubes de mix pour la PCR universelle (bouchon vert) + les tubes de Réactifs d'amplification (AR1) à décongeler sur la glace (environ 1h)
   Ex : 1 tube = 448µl, donc un peu + de 5 tubes pour 162 réactions.

Pendant ce temps, il est possible de procéder à la dilution de la 1ère PCR en pièce Post-PCR (voir ci-après)

2) Bien vortexer puis centrifuger le tube de mix pour la PCR universelle
3) Préparer un mix pour l'ensemble des échantillons (dans un tube à hémolyse si > 1ml)
   Exemple pour une série de 32 cas, amplifiés pour 5 multiplexes : préparer un mix pour 160 réactions + 2 réactions de volume mort.

Nombre échantillons	**/**	**8**	**16**	**32**
Réaction	**x1**	**X41**	**X82**	**X162**
mix PCR universelle (µl)	14	574	1148	2268
Amplification Reagent AR1 (µl)	10	410	820	1620
Taq (µl)	0,125	5,125	10,25	20,25

En pièce Post-PCR

4) Préparer une dilution au 1/1000eme de la PCR multiplex.

Procéder en 2 étapes :

1ère dilution au 1/33ème (soit 3µl de produits PCR + 100 µl d'eau),

2ème dilution au 1/33ème (soit 3µl la 1ère dilution + 100 µl d'eau)

Bien mélanger sur agitateur (minimum 5min) entre chaque dilution.

5) Distribuer 24 µl de mix par puits

6) Déposer 1 µl de la 2ème dilution dans le mix

7) Lancer la PCR
   98°C - 10 min
   95°C - 45 s
   64°C - 45 s
   } X20
   68°C - 2 min
   72°C - 10 min
   4°C - ∞

3/ PCR de validation Genescan

1) Mettre les tubes de *labelling PCR mix* (bouchon jaune) à décongeler sur la glace (environ 1h)
   Ex : 1 tube = 24 réactions, donc ~7 tubes pour 160 réactions.
   Attention : produit fluorescent > à conserver à l'abri de la lumière

2) Bien vortexer puis centrifuger le tube de *labelling PCR mix*

3) Préparer un mix pour l'ensemble des échantillons (dans un tube à hémolyse si > 1ml)
Exemple pour une série de 32 cas, amplifiés pour 5 multiplexes : préparer un mix pour 160 réactions + 4 réactions de volume mort.

Nombre échantillons	**/**	**8**	**16**	**32**
Réaction	**x1**	**X42**	**X82**	**X164**
***labelling PCR mix* (µl)**	13	546	1066	2132
Taq (µl)	0,075	3,15	6,15	12,3

En pièce Post-PCR

4) Distribuer 13 µl de mix par puits

5) Déposer 2 µl de PCR universelle dans le mix

6) Lancer la PCR

98°C - 10 min
95°C - 45 s
60°C - 45 s
68°C - 2 min
} X5
72°C - 10 min
4°C - ∞

4/ Analyse de fragments

En pièce Post-PCR

1) Préparer le mix suivant (dans un tube à hémolyse si > 1ml)

Nombre échantillons	**/**	**8**	**16**	**32**
Réaction	**x1**	**X42**	**X82**	**X164**
HiDi formamide (µl)	19,5	819	1599	3198
GS500-ROX (µl)	0,5	21	41	82

2) Distribuer 20 µl de mix par puits
3) Diluer au 1/10ème la *labelling PCR* (2µl de PCR dans 20µl d'H2O)
4) Déposer 2 µl de la dilution au 1/10ème de *labelling PCR* dans chaque puits
5) Dénaturer 2 minutes à 92°C puis déposer immédiatement sur glace
6) Remplir la fiche de run et transmettre à la plate-forme de séquençage Sanger
   ATTENTION : penser à préciser sur la fiche de run que le temps d'injection est de 5s.
   (pour éviter la saturation des amplicons). Un module de run dédié a été créé, nommé « GSHD500_D_50cm_POP7_multiampli ».

5/ Interprétation analyse de fragments

1/ Vérification de la qualité des amplifications

Regarder les **données brutes** de chaque échantillon.

Utiliser les logiciels GeneMapper (Applied Biosystems) ou GeneMarker (SoftGenetics).

- Vérifier que le nombre de pics pour chaque multiplexe est correct.

Attention aux amplicons avec polymorphisme fréquent entrainant une délétion d'1 ou qq bases (cf annotation des profils page suivante) :

- BRCA1_ex07_01> IVS07+36del14 (plex 4, amplicon 9)
- BRCA1_ex09_01 > IVS08-58delT (plex 5, amplicon 2)
- BRCA2_ex11_22 > IVS11+80delTTAA (plex 5, amplicon 17)

- Vérifier l'intensité et l'homogénéité de chaque pic.

S'il n'y a pas d'amplification :

Faire un dosage fluo de la PCR universelle de l'échantillon correspondant, en le comparant avec un échantillon normal.

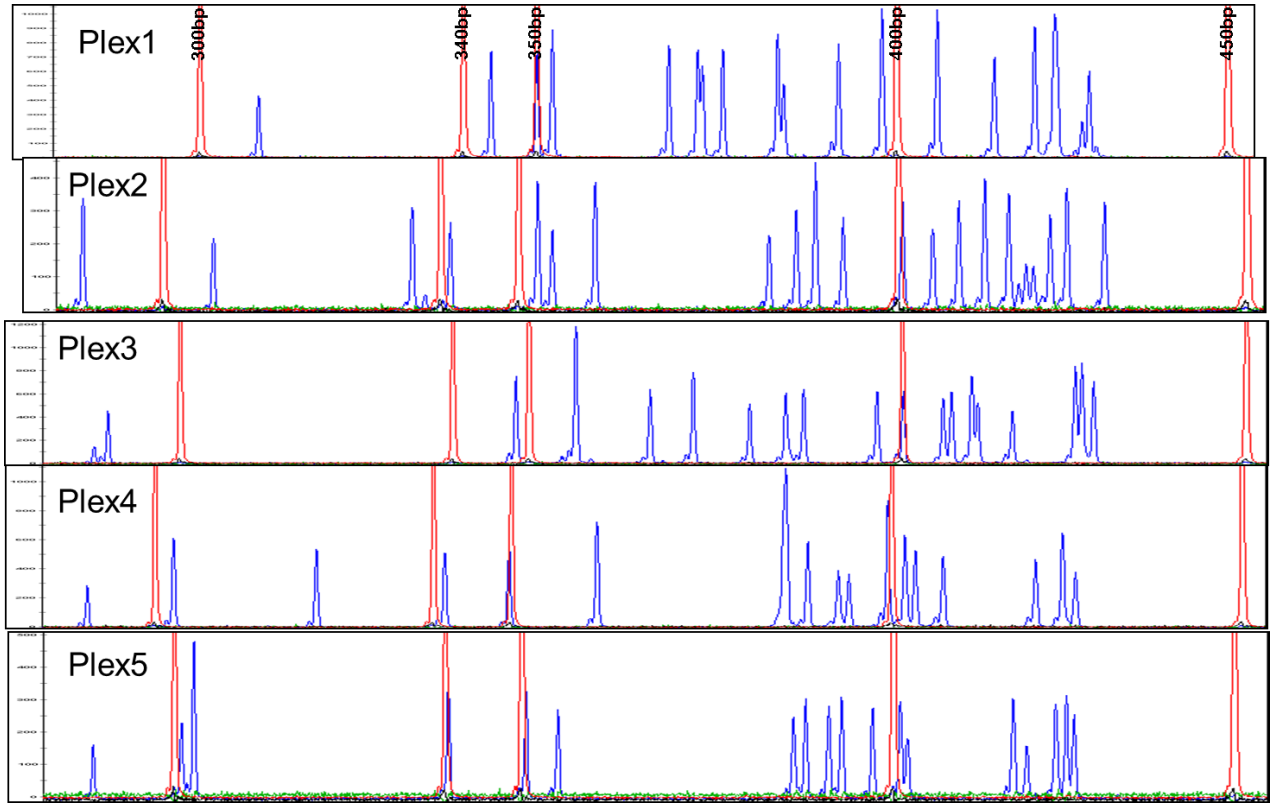
Si le dosage est conforme, reprendre le protocole à partir de l'analyse de fragment,

Si le dosage est non conforme, reprendre le protocole à partir de la PCR Multiplexe.

Si l'amplification est hétérogène :

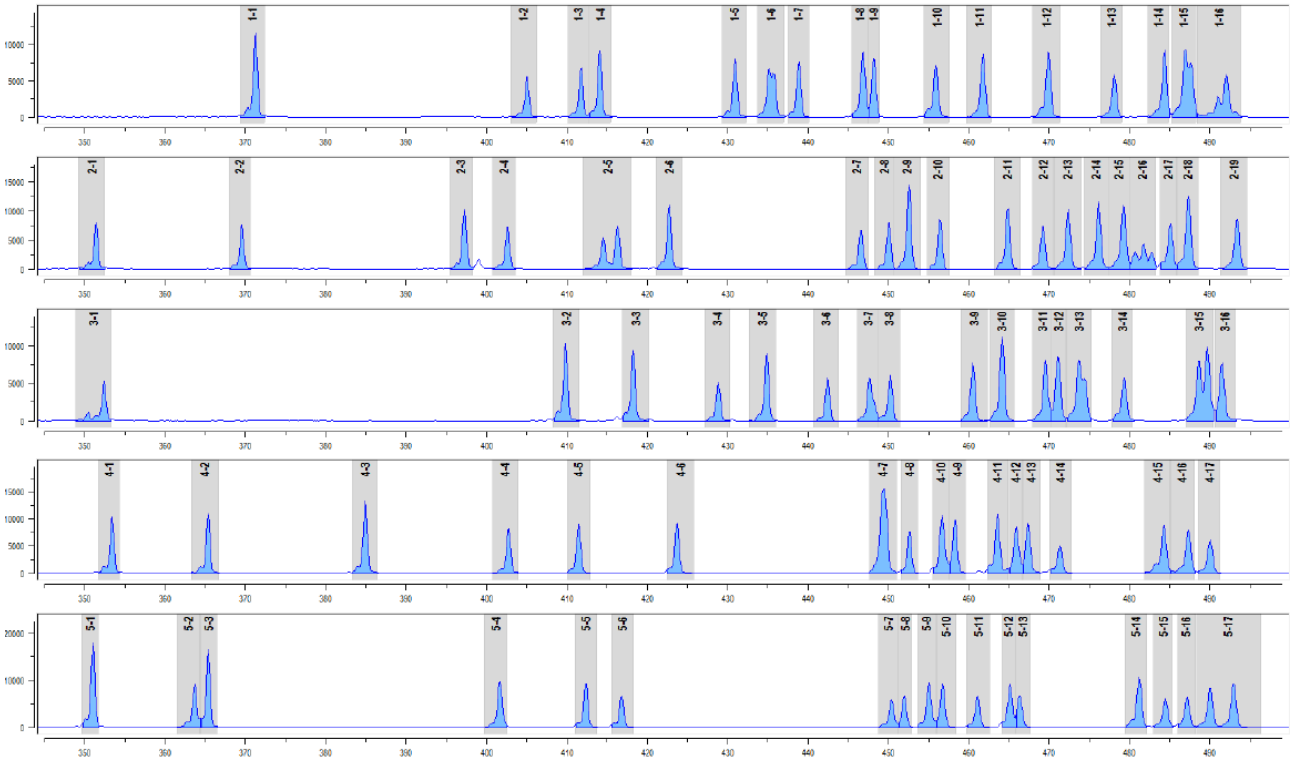
Reprendre le protocole à partir de la PCR Multiplexe.

Exemple de profils obtenus (images de GeneMapper) :



en bleu les amplicons, en rouge le marqueur de taille

BRCA MASTR v2.1 genescan profiles and annotation



Bin	Amplicon	Bin	Amplicon	Bin	Amplicon	Bin	Amplicon	Bin	Amplicon
1-1	BRCA2_ex09_01	2-1	BRCA1_ex20_01	3-1	BRCA2_ex06_01	4-1	BRCA1_ex03_01	5-1	BRCA1_ex22_01

1-2	BRCA1_ex23_01	2-2	BRCA1_ex19_01	3-2	BRCA1_ex14_01	4-2	BRCA2_ex11_13	5-2	BRCA1_ex09_01
1-3	BRCA1_ex11_12	2-3	BRCA1_ex16_02	3-3	BRCA2_ex19_01	4-3	BRCA1_ex17_01	5-3	BRCA2_ex11_14
1-4	BRCA2_ex11_07	2-4	BRCA2_ex25_02	3-3	BRCA1_ex02_01	4-4	BRCA1_ex13_01	5-4	BRCA2_ex21_01
1-5	BRCA2_ex13_01	2-5	BRCA2_ex02_01	3-4	BRCA2_ex10_04	4-5	BRCA1_ex16_01	5-5	BRCA1_ex10_01
1-6	BRCA1_ex11_07	2-5	BRCA2_ex11_09	3-5	BRCA1_ex11_05	4-6	BRCA2_ex11_11	5-6	BRCA2_ex12_01
1-6	BRCA2_ex11_19	2-6	BRCA1_ex21_01	3-6	BRCA1_ex06_01	4-7	BRCA1_ex11_06	5-7	BRCA2_ex22_01
1-7	BRCA2_ex15_01	2-7	BRCA2_ex16_01	3-7	BRCA1_ex11_08	4-7	BRCA2_ex14_02	5-8	BRCA1_ex11_10
1-8	BRCA2_ex11_01	2-8	BRCA2_ex08_01	3-8	BRCA1_ex12_01	4-8	BRCA2_ex10_03	5-9	BRCA2_ex03_02
1-9	BRCA2_ex11_10	2-9	BRCA1_ex11_01	3-9	BRCA2_ex11_12	4-9	BRCA1_ex07_01	5-10	BRCA2_ex11_02
1-10	BRCA2_ex24_01	2-10	BRCA1_ex08_01	3-10	BRCA2_ex11_05	4-10	BRCA1_ex05_01	5-11	BRCA2_ex11_06
1-11	BRCA2_ex14_01	2-11	BRCA1_ex11_14	3-11	BRCA1_ex18_01	4-11	BRCA2_ex05_01	5-12	BRCA2_ex11_17
1-12	BRCA2_ex11_15	2-12	BRCA1_ex15_01	3-12	BRCA1_ex11_13	4-12	BRCA2_ex27_03	5-13	BRCA1_ex11_02
1-13	BRCA2_ex11_21	2-13	BRCA2_ex27_02	3-13	BRCA2_ex11_08	4-13	BRCA1_ex11_03	5-14	BRCA2_ex27_01
1-14	BRCA1_ex11_04	2-14	BRCA2_ex11_18	3-13	BRCA2_ex11_16	4-14	BRCA2_ex25_01	5-15	BRCA1_ex24_01
1-15	BRCA2_ex18_01	2-15	BRCA2_ex11_04	3-14	BRCA2_ex23_01	4-15	BRCA1_ex11_09	5-16	BRCA2_ex04_01
1-15	BRCA2_ex10_01	2-16	BRCA2_ex10_05	3-15	BRCA2_ex18_02	4-16	BRCA2_ex11_03	5-17	BRCA2_ex26_01
1-16	BRCA2_ex03_01	2-17	BRCA2_ex10_02	3-15	BRCA2_ex07_01	4-17	BRCA1_ex11_11	5-17	BRCA2_ex11_22
		2-18	BRCA2_ex11_20	3-16	BRCA2_ex20_01				
		2-19	BRCA2_ex17_01						

2/ Quantification des multiplexes pour analyse des RGT

Utiliser le logiciel GeneMarker (SoftGenetics).

Attention : une fois les données analysées dans GeneMarker, la taille des pics ne correspond pas à l'intensité réelle ! Pour vérifier l'intensité, regarder les données brutes.

Ouvrir le logiciel

- Importer les fichiers .fsa à analyser : « Open Data » « Add Folder » choisir le run à analyser puis « ok » : les données brutes s'affichent.

- Lancer le run (alignement + normalisation interne) : « Run », sélectionner le template correspondant au gène étudié soit : Multiplicom plex1-2, multiplicom plex2, multiplicom plex3, multiplicom plex4, multiplicom plex5.

- Vérifier le type type de panel et de marqueur de taille utilisé soit GS500 et red. Cliquer sur « Next » : la fenêtre Data Process- MLPA analysis s'ouvre.

- Vérifier que le pourcentage « Global max » = 1, le pourcentage « Local Max » = 5 et le pourcentage « Left Stutter Peak Filter » = 25 (paramètres recommandés dans le manuel Gene Marker). Cliquer sur « next » : la fenêtre Additional Settings s'ouvre.

- Vérifier que le mode « Population Normalization » est bien sélectionné. Cliquer sur « ok ». Le logiciel aligne et normalise les différents échantillons.

Dans la fenêtre de gauche, faire un clic droit sur « allele call » puis « select max » pour visualiser les électrophorégrammes de l'ensemble de la série analysée.

- Vérifier que pour chacun d'eux, les pics correspondant aux différents exons/contrôles sont bien identifiés (rectangle avec le nom de l'exon/contrôle sur fond gris) Les pics non identifiés sont repérables par un rectangle sur fond rouge. Normalement, cela concerne uniquement les pics correspondant aux Q fragments, D fragments et chromosomes X et Y.

- Sauvegarder le projet puis choisir « MLPA Analysis » Vérifier que les bornes des ratios sont à 0,75 (del) et 1,15 (dup). « Ok ». La fenêtre « MLPA Analysis » s'ouvre.

- Créer un contrôle synthétique : dans la case « Control Sample », choisir « Synthetic Control Sample ». Puis, cliquer sur l'onglet « ctrl ». Cocher les échantillons servant à faire le contrôle synthétique et déselectionner les témoins mutés.

Pour chaque échantillon, les ratios correspondant aux sondes contrôles apparaissent sous forme d'un carré bleu, ceux correspondant aux différents exons du gène étudié sous forme d'un carré vert si ils sont dans la fourchette cible ou en rouge si ils sont hors de cette fourchette.

Générer un fichier .pdf en cliquant sur l'imprimante puis « Preview » « print » « cute pdf writer ».

Aller dans Project/run

- Cliquer sur "select an existing template or create one"
- Nommer le template

<u>Création d'un panel :</u>

- *Aller dans Tools/Panel après une première analyse des échantillons (avec le bon marqueur de taille)*
- *Fichier « new panel » pour créer un nouveau panel*
- *Cocher « automatically create » et « use all samples »*
- *Cliquer sur "ok"*
- *Pour le panel en cours supprimer les couleurs inutiles, ne garder que le bleu*
- *Cliquer sur « blue1 »*
- *Vérifier que tous les pics contrôles et échantillons sont présents. Supprimer les autres par un clic droit « delete allele »*
- *Annoter les pics contrôles en mettant « 1 » dans la case « control »*
- *Annoter les pics de l'échantillon en inscrivant le nom du pic dans « allele name)*
- *Cliquer sur Fichier/Save puis Fichier/Exit*

- Choisir le panel, le size standard et la couleur
- Analysis type = MLPA
- Cliquer sur « save » ➔ le template appaîtra toujours à gauche
- Cliquer sur « next » (« data Process ») et laisser les paramètres par défaut
- Cliquer sur OK puis « Apply to all » (et « call size again »)
- Cliquer sur "MLPA Analysis"
- Laisser "MLPA ratio" et régler les ratio sur del < 0.70 ≤ N ≤ 1.30 < Dup
- Dans la nouvelle fenêtre, cliquer sur le menu déroulant de « CTRL Sample » et sélectionner le « synthetic control sample »
- Cliquer sur « CTRL » et cocher les échantillons qui vont servir à faire le contrôle synthétique
- Analyse des graphes : Point bleu = contrôles
  Point vert = pics à analyser Normal

  Point rouge = pics à analyser Hors limites

6/ Mix des produits PCR

Pour chaque échantillon, combiner les produits de la PCR universelle dans les puits des colonnes 6 et 12 restés vides des plaques PCR, de la façon suivante :

Plex1 : 8 µl

Plex2 : 11 µl

Plex3 : 13 µl

Plex4 : 8 µl

Plex5 : 10 µl

(Total 50µl)

7/ Purification des produits PCR

Sortir les billes Agencourt au minimum 30min à température ambiante

Préparer éthanol 70% (35ml d'éthanol absolu + 15 ml d'eau pure)

6/ 1- Purification manuelle :

1) Déposer 40 µl du mix PCR dans un microtube 1,5ml *Lo-bind* et ajouter 44 µl de billes Agencourt
2) Mixer par aller-retour (10x)
3) Laisser 5 minutes à température ambiante
4) Déposer les tubes 3 minutes sur le portoir magnétique
5) Eliminer le surnageant
6) Ajouter 200 µl d'éthanol 70% (fraichement préparé) et tourner 2 ou 3 fois les tubes sur le portoir pour déplacer les billes (environ 30 secondes)
7) Eliminer le surnageant
8) Refaire étape 6 et 7
9) Laisser sécher le culot environ 5 minutes (bouchons ouverts)
10) En dehors du portoir magnétique, resuspendre les billes dans 20 µl d'eau
11) Bien vortexer, centrifuger puis remettre sur le portoir magnétique 1 minute
12) Prélever le surnageant et le mettre dans un tube 1.5ml *Lo-bind* neuf.

Attention : Ne pas prendre de billes

6/ 2- Purification au robot Beckman-Coulter :

1) Déposer 40 µl du mix PCR dans une plaque MicroAmp 96 puits (Life Tech - Applied)
2) Ouvrir le logiciel *« Biomek Software »* du robot Beckman-Coulter et le programme
3) Dans « *Agencourt user interface* » : sélectionner le nombre d'échantillons à purifier, et les volumes de réactifs, soit : 44 µl de billes Agencourt, 180 µl d'éthanol,
4) Remplir les réservoirs dédiés de billes Agencourt, d'éthanol 70% et d'eau
5) Dans « *Instrument Setup* », suivre les instructions du programme pour disposer sur la plateforme du robot :
   - la plaque de produits PCR sur une base dédiée
   - une nouvelle plaque MicroAmp 96 puits pour l'élution des échantillons sur une base dédiée
   - une plaque « poubelle »
   - une plaque de lavage contenant 1ml d'eau par puits utilisé

6) Ajouter le bon nombre de pointes pour l'exécution du programme
7) Lancer le programme (durée ~1h)
8) A la fin du programme, récupérer la plaque contenant les produits purifiés.

Stockage des produits purifiés à -20°C

	Noms (Fonctions)	Dates de signature
Rédacteur(s)	Julien Tarabeux (Ingénieur) Virginie Moncoutier (THQ)	

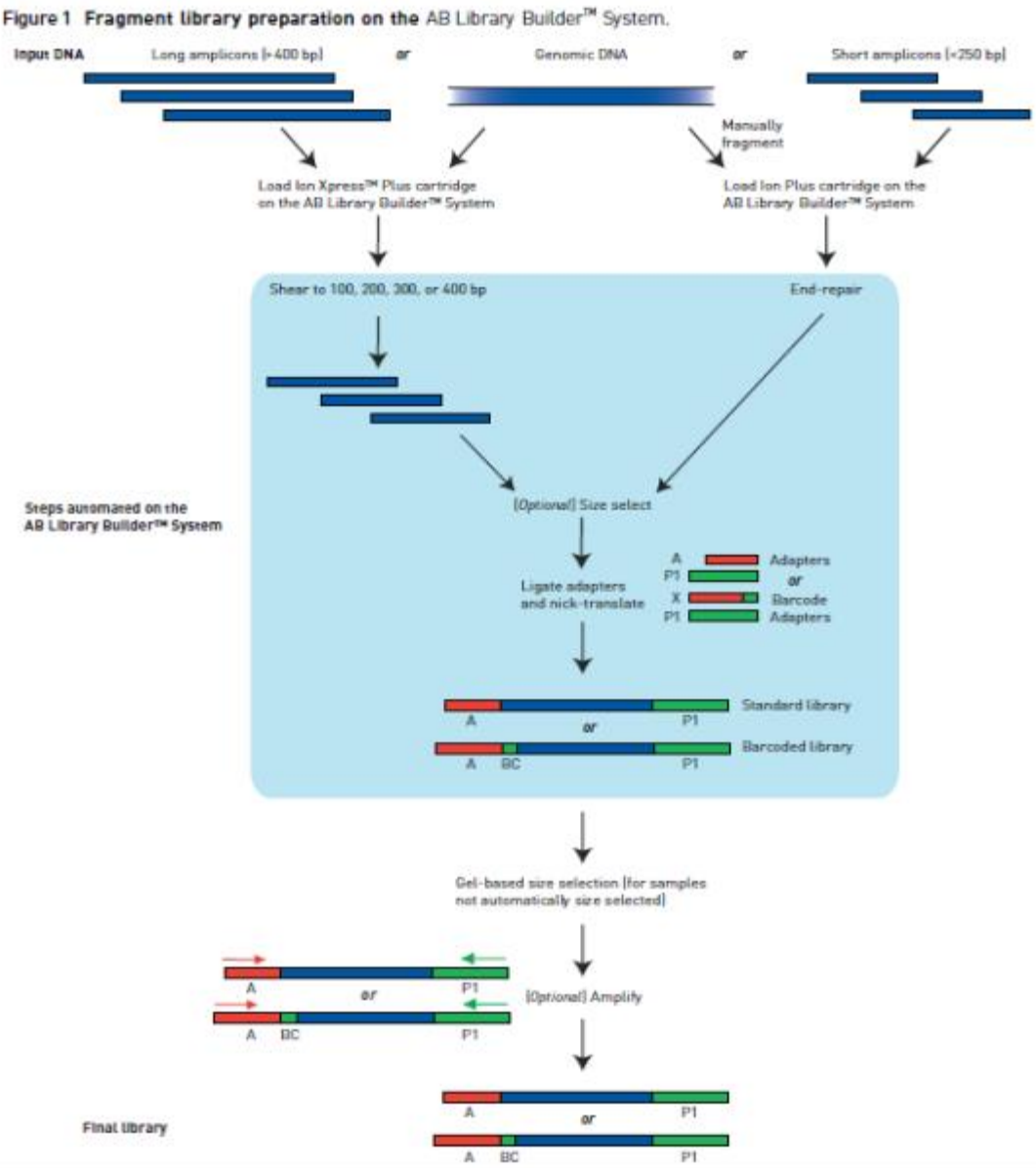
PREPARATION DES BANQUES BRCA POUR SEQUENCAGE HAUT-DEBIT

Objet : *Construction des banques d'ADN pour les gènes* BRCA1 *et* BRCA2*, avec l'automate «* Library Builder *» (*Life Technologies*), amplification, dosage et* pool.

4. Domaine d'application

Le Library Builder permet de préparer des banques d'ADN pour leur passage sur le séquenceur PGM.

Cette technique automatise plusieurs étapes : digestion les amplicons, sélection des amplicons selon la taille désirée (100, 200 ou 300pb) et la ligation des barcodes et l'adapteur P1 en 3' et 5' des amplicons.



Les banques d'ADN construites doivent ensuite être amplifiées, dosées et mélangées de façon équimolaire avant de préparer la PCR en émulsion et le run PGM.

5. Responsabilités

(Ce paragraphe précise les responsabilités concernant l'application du protocole et la vérification de sa bonne application)

6. Description détaillée

Le processus complet de diagnostic des gènes BRCA sur PGM se déroule en 4 phases :

I - Enrichissement

II - Préparation des banques > ce protocole décrit la procédure de préparation avec l'automate *Library Builder*, puis l'amplification, le dosage et le pool des banques avant la PCR en émulsion.

Ion Xpress™ Plus and Ion Plus Library Preparation for the AB Library Builder™ System

III - Séquençage (PCR émulsion + PGM)

IV - Analyse

Plan de préparation des banques (Library Builder)

1/ Dosage des échantillons > cf protocole dosage au QuBit

2/ Préparation des banques

3/ Amplification des banques

4/ Purification des banques amplifiées

5/ Contrôle et dosage des banques amplifiées > cf protocole dosage Bioanalyzer ou LabChip

6/ Détermination du facteur de dilution des banques

7/ Mélange équimolaire des différentes banques

Préparation des banques (Library Builder)

1/ Dosage des échantillons

1/ Dosage au QuBit Fluorometer (Invitrogen)

Quant-iT dsDNA BR Assay kit (pour ADN de conc. 2 à 1000ng/µl) réf. Q32850

Quant-iT dsDNA HS Assay kit (pour ADN de conc. 0,2-100 ng/µl) réf. Q32854

Tubes Invitrogen « Qubit Assay tubes », 0.5ml (réf. : Q32856)

Stockage des réactifs et des tampons à température ambiante ; stockage ADN standards à 4°C.

- Dosage par échantillon
- Cf protocole dédié

2/ Dosage au spectrophotomètre Genios (Tecan)

- Dosage par plaque de 96 puits
- Utiliser les mêmes réactifs que pour le dosage au QuBit
- Utiliser des plaques noires, exemple de référence : plaques 96 puits noires FluoroNuncTM, réf 734-2017 chez VWR.

1) Préparation de la solution de travail

- Evaluer le nombre d'échantillons à doser (= nombre d'ADN à doser + Gamme STD à 5 points + 1 tube en excès)

- Préparer la solution de travail (Working Solution = tampon + Reagent), à température ambiante, en fonction du nombre d'échantillons à doser :

Nombre d'ADN à doser	Nombre d'échantillons total	Volume de Buffer (en µl)	Volume de Reagent (en µl)
-	1	199	1
20	26	5174	26
32	38	7562	38

2) Préparation de la gamme standard

a) Pour le dosage des enrichissements Multiplicom après purif :

- Utiliser les Standards 1 et 2 du kit QuBit.

- Préparer une gamme à 5 points. Exemple en utilisant le kit HS, avec le **STD 1** = 0ng/µl, et le **STD 2** = 10ng/µl.

Concentration (ng/µl)	Volume de Working solution (en µl)	Volume de **STD 1** (en µl)	Volume de **STD 2** (en µl)
0	198	2	0
2,5	198	1,5	0,5
5	198	1	1

7,5	198	0,5	1,5
10	198	0	2

b) Pour le dosage des banques library builder après purif :

- Utiliser les Standards 1 et 2 du kit QuBit.HS

- Préparer une gamme à 5 points utilisant le kit HS, avec le **STD 1** = 0ng/µl, et le **STD 2 = 10ng/µl dilué au 1/10 avec le tampon TE (2µL STD2 + 8µL tampon TE)**

Concentration (ng/µl)	Volume de Working solution (en µl)	Volume de **STD 1** (en µl)	Volume de **STD 2 au 1/10** (en µl)
0	198	2	0
2,5	198	1,5	0,5
5	198	1	1
7,5	198	0,5	1,5
10	198	0	2

3) Préparation des échantillons à doser (dilution au 1/200e)

- Dans une plaque 96 puits noire, déposer 198µl de Working solution et 2µl d'ADN.

- Mélanger par aspiration / refoulement à la pipette multicanaux.

- Après homogénéisation, les échantillons doivent être incubés au minimum 2 minutes à température ambiante.

- Après cette incubation, la fluorescence est stable pendant 3 heures à température ambiante.

La température peut entraîner des variations de fluorescence.

4) Dosage des échantillons

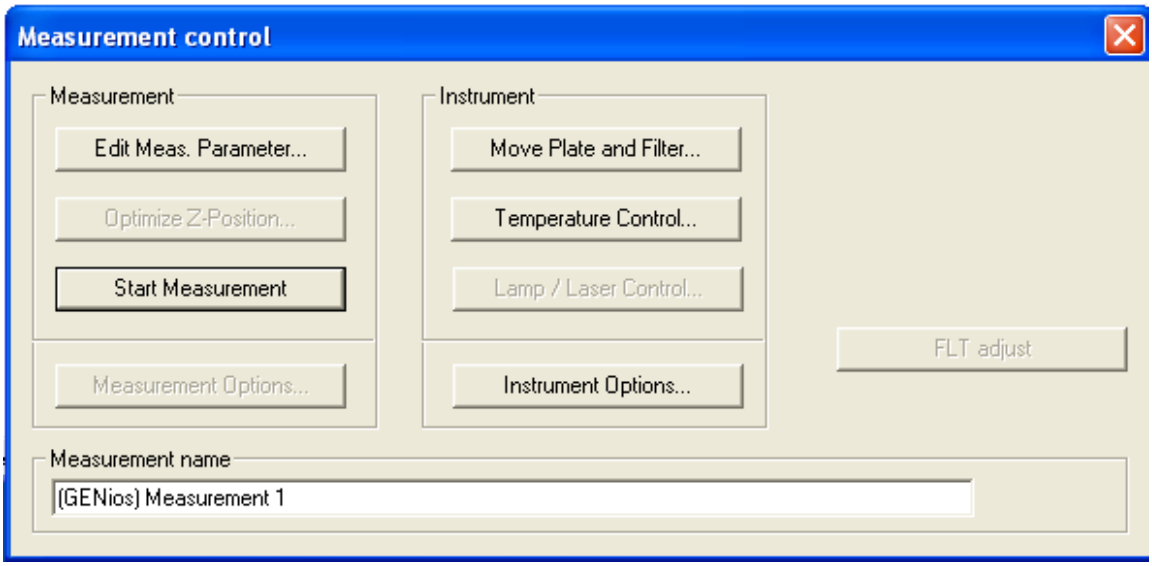
- Allumer le PC avec son mot de passe réseau, sous IC-NT01

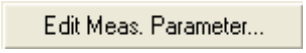
- Allumer le spectrophotomètre Genios

- Ouvrir le logiciel « RdrOle4 »

- Menu « Instrument » > Connect

- Menu « Instrument » > measurement control panel





- General : measurement mode = Fluorescence Intensity
- Plate : Plate definition file > Name = GRE96fb.pdf
  Select Part of the plate > par colonnes
- Meas. Params : Fluorescence wavelengths = fluorescein (valeurs par défaut)
- Shaking : shaking before measurement : Duration = 15sec
  Settle time = 120 sec

- Plate > out : déposer la plaque de dosage dans la machine. Plate > in.

- Filtre d'excitation = 485nm, filtre d'émission = 535nm.

A la fin des mesures, les valeurs s'affichent automatiquement à l'écran.

File > export RawData :

- Sauvegarder sous Excel.

- Exploiter les Rawdatas grâce à un graphique (nuage de points). **Attention à ce que les valeurs soient bien au format « nombre ».**

- Créer la courbe de tendance avec les valeurs de la gamme.

- Vérifier que la gamme soit linéaire, et faire les calculs de concentration grâce à l'équation de la courbe de tendance.

Utiliser les modèles xls pour calculer les concentrations des échantillons :

- Pour les enrichissements : Dosage genios enrichissement BRCA modèle v1
- Pour les banques LB : Dosage genios Banques BRCA modèle v1

Déterminer le volume nécessaire pour obtenir 100ng d'ADN pour construction des banques au LB.

- (cf : Dosage genios enrichissement BRCA modèle v1onglet s_BRCA-Dilutions pour LB

2/ Préparation des banques

Protocole issu de "Ion Xpress™ Plus and Ion Plus Library Preparation for the AB Library Builder™ System"

Références des kits utilisés:

- Ion Xpress™ Plus Fragment Library Kit for AB Library Builder™ System (Ion Torrent) réf. 4477598 (comprend les cartouches à conserver à -20°C, et les consommables)
- Ion Xpress Barcode Adapters 1-96 kit (10 réactions), réf. 4474517

Sortir les billes Agencourt à T° ambiante au minimum 30 minutes avant leur manipulation, et homogénéiser avant usage.

Décongeler le nombre nécessaire de cartouches de « *Library Builder* » : 1 cartouche par échantillon (maximum 2h à T° ambiante avant utilisation)

1/ Dilution des ADN

a/ Numéroter le nombre de microtubes de 1,5ml *LoBind* correspondant au nombre d'ADNs à préparer,

b/ Diluer dans chaque microtube 100 ng d'ADN dans 50 µl de tampon E1 (fourni dans le kit).

2/ Dilution des adaptateurs

a/ Préparer le nombre de microtubes de 1,5ml *LoBind* correspondant au nombre de Barcodes à préparer (= au nombre d'ADNs),

b/ Diluer dans chaque microtube les adaptateurs P1 et les barcodes (de 1 à 96 au choix) selon les proportions suivantes :

- Ion Xpress adaptateur P1 = 2 µl
- Ion Xpress Barcode = 2 µl (utiliser un n° de barcode différent pour chaque banque d'ADN)
- Tampon E1 = 46 µl

c/ Transférer les 50 µl d'adaptateurs dilués dans un tube de « *LdN buffer* » par banque (fourni dans le kit).

d/ Mélanger à la pipette par aspiration / refoulement puis effectuer une centrifugation rapide.

3/ Préparation du « *Library Builder* »

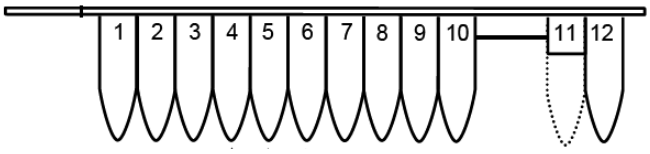
Avant d'allumer le « *Library Builder* », insérer la carte

Ne pas enlever la carte quand l'appareil est allumé.

Porte fermée, allumer l'appareil et patienter pendant son initialisation.

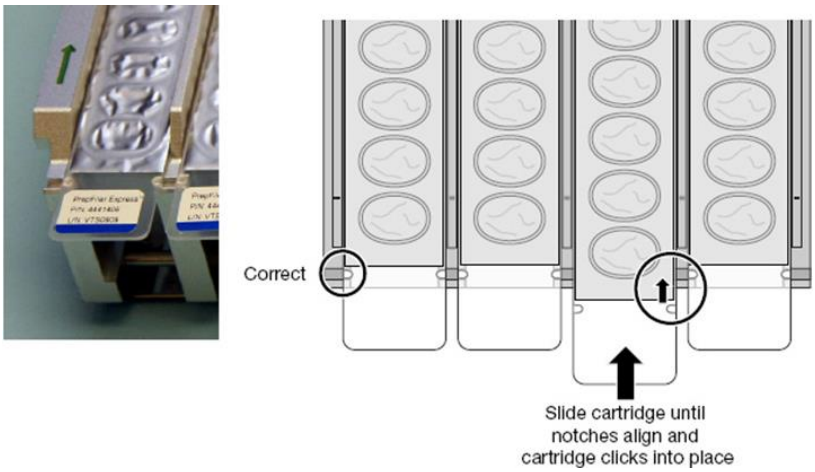
Ouvrir la porte, retirer le portoir et le rack de cartouches.

a/ Chargement des cartouches :



Vérifier que tous les réactifs des cartouches soient bien décongelés et au fond de chaque puits.

Placer en les glissant le nombre de cartouches correspondant dans le rack, jusqu'au clic.



Déposer - dans le puits 10 : 500 µl de billes Agencourt après avoir percé l'opercule aluminium

- dans le puits 11 : les tubes de « *LdN buffer* » ouverts, contenant les adaptateurs dilués

- dans le puits 12 : 50 µl d'ADN (quantité 100ng) après avoir retiré l'opercule aluminium.

b/ Chargement du portoir :

Déposer - en ligne E : les tubes d'élution (fournis dans le kit), bouchons ouverts et placés à côté

- en lignes T1 et T2 : les pointes avec leur protection (fournies dans le kit)

- en ligne S : les tubes de « *Ion shear Plus Enzyme* » (bouchons jaunes, après centrifugation rapide et conservés jusqu'au dernier moment sur glace) bouchons ouverts.

Placer le rack puis le portoir remplis dans l'appareil, fermer la porte.

Appuyer sur START.

Suivre les instructions sur l'écran :

- appuyer sur 1 = *auto Size Selection*

- appuyer sur 1 = *auto Shearing*

- appuyer sur 3 = *300 pb shear*

- confirmer que vous avez bien inséré tous les réactifs et consommables dans l'appareil,

- appuyer sur start.

Durée ~2h15.

Volume final de banques : 40 µl.

Une fois le *run* terminé, récupérer et fermer les tubes contenant les banques, jeter les consommables, nettoyer le rack et le portoir et éteindre l'appareil.

Si présence de billes restantes à la fin de cette étape, les éliminer en déposant les tubes sur le portoir magnétique, et récupérer l'éluat dans un nouveau tube.

Stockage des banques à -20°C.

3/ Amplification des banques

- Dans une plaque PCR MicroAmp 96 puits (Life Tech - Applied), combinez les réactifs suivants :

	1x	32,5x
Size-selected Librairie	20 µl	/
Platinum PCR SuperMix High Fidelity	100 µl	3250 µl
Primer Mix (10µM)*	5 µl	162,5 µl

* préparation du Primer Mix:

Amorces stock à 100µM > P1amp : 5'-CCACTACGCCTCCGCTTTCCTCTCTATG-3'

(Sigma-Proligo) > T_PCR_A : 5'-CCATCTCATCCCTGCGTGTC-3'

Purifiées HPLC

Exemple de préparation du Primer Mix à 10µM (soit 5µM de chaque amorce) pour 1ml final:

50µl amorce stock P1amp + 50µl amorce stock T_PCR_A + 900µl TE.

Lancer le programme suivant :

72°C - 20 min (= nick translation)
95°C - 5 min
95°C - 15 s
58°C - 15 s
70°C - 1 min
} X10
4°C - ∞

Stockage des banques amplifiées à -20°C.

4/ Purification des banques amplifiées

Sortir les billes Agencourt au minimum 30min à température ambiante avant leur manipulation, et homogénéiser avant usage.

Préparer de l'éthanol à 70% extemporané (35ml d'éthanol absolu + 15 ml d'eau pure)

4/ 1- Purification manuelle :

1) Transférer les 125 µl de produit PCR dans un microtube 1,5ml *Lo-bind*
2) Ajouter 187,5 µl de billes Agencourt
3) Mixer par aller-retour (10x)
4) Laisser 5 minutes à température ambiante
5) Déposer les tubes 3 minutes sur le portoir magnétique
6) Eliminer le surnageant
7) Ajouter 500 µl d'éthanol 70% (fraichement préparé) et tourner 2 ou 3 fois les tubes sur le portoir pour déplacer les billes (environ 30 secondes)
8) Eliminer le surnageant
9) Refaire étape 6 et 7
10) Laisser sécher le culot environ 5 minutes (bouchons ouverts)
11) En dehors du portoir magnétique, resuspendre les billes dans 20 µl de TE
12) Bien vortexer, centrifuger puis remettre sur le portoir magnétique 1 minute
13) Prélever le surnageant et le mettre dans un tube 1.5ml *Lo-bind* neuf.

Attention : Ne pas prendre de billes

4/ 2- Purification au robot Beckman-Coulter :

1) Ouvrir le logiciel *« Biomek Software »* du robot Beckman-Coulter et le programme
2) Dans « *Agencourt user interface* » :
   sélectionner le nombre d'échantillons à purifier,
   les volumes de réactifs, soit : 187,5 µl de billes Agencourt, 180 µl d'éthanol,
   le volume d'élution, soit 30 µl
   le volume de transfert, soit 25 µl
3) Remplir les réservoirs dédiés de billes Agencourt, d'éthanol 70% et d'eau Nuclease free
4) Dans « *Instrument Setup* », suivre les instructions du programme pour disposer sur la plateforme du robot :
   - la plaque de produits PCR sur une base dédiée
   - une nouvelle plaque MicroAmp 96 puits pour l'élution des échantillons sur une base dédiée
   - une plaque AB1127 pour le mélange échantillons / billes et lavages
   - une plaque « poubelle »
   - une plaque de lavage contenant 1ml d'eau par puits utilisé
5) Ajouter le bon nombre de pointes pour l'exécution du programme
6) Lancer le programme (durée ~1h)
7) A la fin du programme, récupérer la plaque contenant les produits purifiés.

Stockage des produits purifiés à -20°C.

5/ Contrôle et dosage des banques amplifiées

Jusqu'à 11 banques d'ADN à contrôler, utiliser une puce *Bioanalyzer* (Agilent)

A partir de 12 banques d'ADN à contrôler, utiliser une puce *LabChip* (Caliper).

5/ 1- Contrôle sur puce *Bioanalyzer* (Agilent)

réf kit et puces à compléter, stockage à +4°C

5/ 2- Contrôle sur puce *LabChip* (Caliper)

http://www.perkinelmer.com/Catalog/Product/ID/122000

Kit HT DNA HiSens Reagent, stockage à +4°C

Puce LabChip HT DNA 1K/12K/HiSens, stockage à +4°C

Sortir le kit et la puce *LabChip* au minimum 30min à température ambiante avant utilisation.

Préparation du Gel-Dye

- Ajouter 13 µL de *DNA Dye COncentrate* (bouchon bleu) dans un tube *DNA HiSens Gel Matrix* de 1.1ml (bouchon rouge) ; vortexer + spin

Note : il est possible de faire un mélange en divisant les volumes par 2 (soit 6.5µl de *DNA Dye COncentrate* (bouchon bleu) + 550µl *DNA HiSens Gel Matrix* (bouchon rouge))

- Transférer le mélange sur 1 ou 2 spin-filters en fonction du volume (env. 550 µL par filtre)
- Centrifuger 10 min à 9200 rcf TA, jeter le filtre

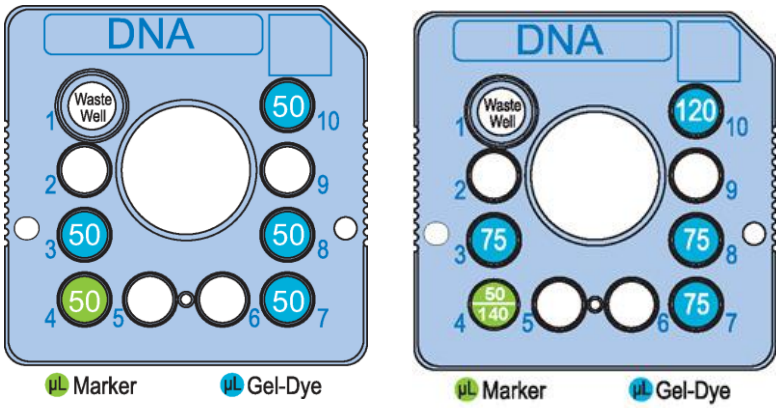
Le gel peut se conserver 3-4 semaines (à l'abri de la lumière et à 4°C)

Préparation de la puce

Toujours garder la puce dans la boite pendant sa préparation

- aspirer les puits actifs (1, 3, 4, 7, 8 et 10) sur les parois et non au milieu du puits

- rincer les puits actifs avec de l'eau ultrapure (remplir à fond les puits) puis aspirer (sur les parois), faire 2 lavages en tout
- remplir les puits 3, 7, 8 et 10 avec du Gel-Dye (le vortexer avant usage) en évitant les bulles (s'arrêter au 1er cran de la pipette):
    - pour un maximum de 48 dosages : 50 µL
    - pour un maximum de 96 dosages : 75 µL et 120 µL pour le puits 10
- remplir le puits 4 avec du *DNA HiSens Marker* :
    - pour un maximum de 48 dosages : 50 µL
    - pour un maximum de 96 dosages : 75 µL



Note : la puce, en condition normale de manipulation, a une capacité minimum de 2000 dosages.

Préparation de l'appareil *LabChip*

- allumer d'abord sur Z power button sur la prise, l'appareil puis l'ordinateur (l'appareil fait son focus au milieu au début)
- lancer le programme « LabChip Gx v.4.0 »
- remplir un Buffer Tube (fourni avec le kit) avec 750 µL H2O ultra-pure
- vérifier la fenêtre de lecture sur la puce, utiliser un papier kimwipe imbibé d'isopropanol à 70%, puis essuyer avec le chiffon fourni avec le kit,
- appuyer sur Chip pour faire sortir le racle pour la puce et celui de la plaque. Installer la puce, fermer le racle et le faire rentrer dans l'appareil. Mettre le Buffer Tube à son emplacement. Appuyer sur Eject pour fermer. Le capillaire plonge dans le Buffer Tube
- lancer le Prime : cliquer sur *Instrument* puis *Start Prime*. Sélectionner la puce HT DNA High Sensitivity (durée 10 min)

Préparation de la plaque

Afin de minimiser le volume de nos échantillons, un transfert vers une plaque 384 puits est nécessaire.

Le volume minimum à déposer est de 10 µL, procéder à une dilution éventuelle de l'échantillon. Pour les banques BRCA, il est d'usage de faire une dilution au ½ (soit 5µl d'écht° + 5µl d'eau).

Vérifier l'absence de bulles, centrifuger si besoin.

Préparation du Ladder Tube

A faire juste avant son utilisation.

Ladder Tube = 12 µL DNA Ladder (Jaune) + 108 µL H2O ultrapure

Fermer le tube avec du Parafilm et vortexer.

Lancement du Run

Installer la plaque 384 puits et le Ladder Tube

NE PAS APPUYER SUR LE BOUTON CHIP (une fois que la puce est en place, elle ne peut pas être délogée avant la fin de son utilisation)

- Veiller à appuyer sur le bouton EJECT pour ouvrir
- Installer la plaque ainsi que le Ladder Tube
- Appuyer sur le bouton EJECT pour fermer

- Cliquer sur *Run*
- Sélectionner la plaque : twin.tec 384 PCR Plate Eppendorf
- Sélectionner les puits contenant les échantillons à analyser
- Choisir l'ordre de lecture (par lignes ou par colonnes)
- Cliquer sur *Start*

(durée environ 1 minute par échantillon)

La puce peut être utilisée dans les 8 heures pour d'autres *runs*.

Nettoyage de la puce

Une fois le *Run* terminé procéder au nettoyage de la puce

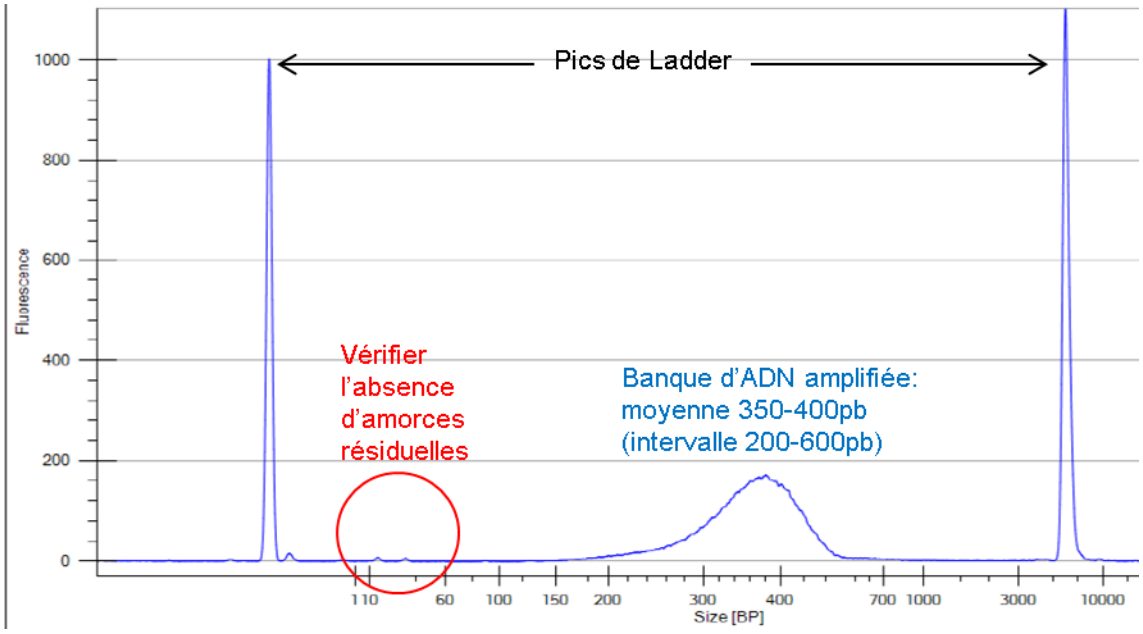
- Appuyer sur CHIP pour faire sortir le racle de la puce, retirer la puce et la mettre sur la boite de stockage
- Eliminer les réactifs de chaque puits par aspiration à l'aide de la pompe
- Faire le nettoyage des puits actifs (1, 3, 4, 7, 8 et 10) avec de l'eau 2 fois
- Mettre 100 µL de *DNA Chip Storage Buffer* dans les puits actifs
- Remettre la puce dans l'appareil et cliquer sur *Wash* (10 min)
- Enlever la puce et le remettre dans sa boite de stockage
- Rajouter 50µL de *DNA Chip Storage Buffer* dans les puits 1 et 4
- Recouvrir les puits avec un Parafilm® pour empêcher l'évaporation du tampon et conserver la puce à 4°C.

Eteindre l'appareil.

Interprétation des résultats

Dans le logiciel « LabChip Gx v.4.0 »

Exemple de profil obtenu :



Si le profil de migration présente des résidus d'amorces (pics d'intensité supérieure de 50% de l'intensité des banques), procéder à une nouvelle étape de purification avec les billes Agencourt. Pour une bonne purification, respecter un ratio échantillon / billes de 1,8 (exemple : pour 100 µl d'échantillon, prévoir 180 µl de billes Agencourt).

Détermination de la taille moyenne des banques et de la molarité

- Import Data file > Analysis > Analysis Settings > Smear Analysis 200-600pb

- Select dans « Property Displayed in Well Table): 1 ligne avec: Molarity (nmol/l)

1 ligne avec: Size (pb)

> Apply: les données apparaissent dans le tableau.

Note : si besoin de revenir sur d'anciennes données sauvegardées :

C : program files > Caliper Life Sciences > LabChip GX > Data > *dossier à la date du jour du dosage.*

Un raccourci a été créé sur le bureau du PC du LabChip.

Transfert des données

- Utiliser le serveur d'échange smercure3 pour le transfert des données entre les sections hopital et recherche.

smercure3 > Dossiers > BRCA_diag > Données LabChip.

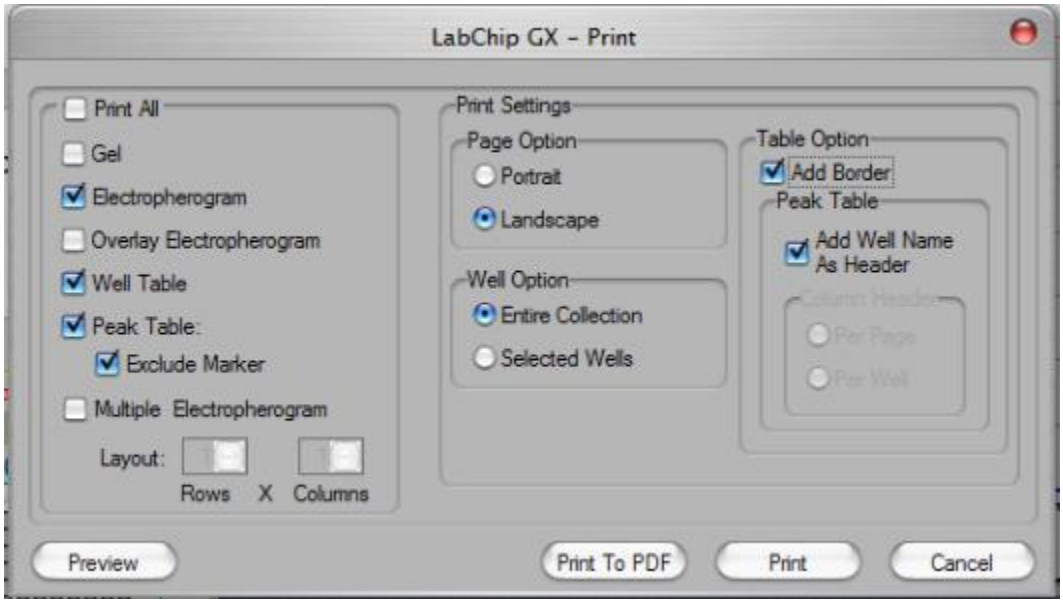
Un raccourci a été créé sur le bureau du PC du LabChip.

- Créer un fichier .csv sous excel, comprenant : « position du puits », « nom de l'écht° »

Exemple : N15, S130R01

- Tools > Sample Name Editor > import

- Print sous .pdf. Sélectionner "Electrophoregram", "Well Table", "Peak Table", "Entire Collection".



- Enregistrer les données sur le bureau du PC du LabChip, dans le dossier « Houdayer ».

- Transférer les données via smercure.

5/ 3- Dosage des banques au QuBit HS

C'est le dosage QuBit qui est utilisé pour le calcul de concentration des banques avant la PCR émulsion.

6/ Détermination du facteur de dilution des banques

La concentration molaire des échantillons en pmol/L (pM) est déterminée grâce :

- à la taille moyenne en pb des banques, obtenue par le logiciel d'analyse *LabChip* ou *Bioanalyzer*.

- au dosage en ng/µl obtenu par fluorescence au QuBit

Diluer les banques d'ADN pour obtenir une concentration moléculaire d'environ **$2x10^7$ molécules / µl**.

Déterminer le facteur de dilution permettant d'obtenir une concentration finale d'environ **33,2 pM**.

Un fichier excel est disponible pour obtenir ce résultat :

S:\Genetique\GenConstitutionnelle\Projets NGS\PGM \OneTouch_Calculator_v2.1.xls

Exemple de calcul de concentration molaire en pM: (dosage en ng/µl x 1000) / (0,00066 x taille en pb)

7/ Mélange équimolaire des différentes banques

Pool des écht°s avant PCRe (attention à faire au dernier moment car ne se conserve pas

	Noms (Fonctions)	**Dates de signature**
Rédacteur(s)	Julien Tarabeux (Ingénieur) Virginie Moncoutier (Tech) Khadija Abidallah (Tech) Henrique Tenreiro (Tech)	
Validateur(s)		
HISTORIQUE DES EVOLUTIONS		
Date	**Motif**	

ENRICHISSEMENT DES ISPS AVEC L'ION ONETOUCH™ ES

Objet : *Ce document décrit les différentes étapes permettant de récupérer uniquement les sphères sur lesquelles des fragments ADN s'y sont hybridés en captant celles-ci à l'aide de billes streptavidine.*

7. Domaine d'application

(Ce paragraphe précise les activités, services, produits, personnels, secteurs, périodes et/ou clients auxquels s'applique le mode opératoire)

Ce mode opératoire concerne…

8. Responsabilités

(Ce paragraphe précise les responsabilités concernant l'application du mode opératoire et la vérification de sa bonne application)

9. Description détaillée

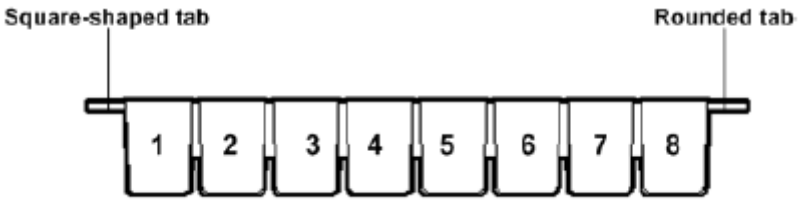
3.1 Distribution de l'échantillon et des différents réactifs dans la barrette

1. Préparer la solution de dénaturation (Melt-Off Solution) en mélangeant les différents réactifs comme indiqué dans le tableau ci-dessous :

Order	Component	Volume
1	Nuclease-Free Water	865 µL
2	1 M NaOH	125 µL
3	10% Tween® 20 in Nuclease-Free Water	10.0 µL
	Total	**1.0 mL**

2. Préparer les billes streptavidine (***Dynabeads® MyOne™ Streptavidin C1 Beads***) :
   a. Dans un tube LoBind, ajouter 130 µL de solution de lavage (***MyOne™ Beads Wash Solution).***

   b. Vortexer la bouteille contenant les billes streptavidine jusqu'à l'homogénéisation complète (environ 30s, vitesse max).
   c. Transférer 13 µL de billes (***Dynabeads® MyOne™ Streptavidin C1 Beads***) dans le tube contenant la solution de lavage. Faire plusieurs aspiration-refoulement pour récupérer le maximum de billes.
   d. Placer le tube sur le portoir magnétique (***DynaMag™-2 magnet***) environ 2 min et éliminer le surnageant.
   e. Ajouter 130 µL de solution de lavage (***MyOne™ Beads Wash Solution***, elle permet d'augmenter l'affinité entre les sphères et les billes). Faire plusieurs aspiration-refoulement pour resuspendre les billes.
3. Préparer l'***Ion OneTouch™ ES*** :
   a. Charger un nouveau cône, retirer le bras de son support et le placer au-dessus du cône, presser environ 1s le bras contre le cône et le remettre sur son socle.
   b. Disposer un nouveau tube PCR ouvert, à l'emplacement disponible à gauche du bras (socle noir). Le tube doit être identifié par le nom du Run.
4. S'assurer que les ISPs soient bien repris dans 100 µL de solution de lavage (***Ion OneTouch™ Wash Solution***). Si l'échantillon a été conservé à +4°C centrifuger les sphères à 15 500 g pendant 2min 30s, éliminer le surnageant et ajouter suffisamment de solution de lavage (***Ion OneTouch™ Wash Solution***) pour obtenir un volume final de 100 µL.
5. Récupérer une barrette dans le kit et la disposer sur la paillasse de sorte que l'extrémité arrondie soit à droite :



6. Resuspendre les ISPs à l'aide d'une pipette (aspiration-refoulement x100) et transférer 2µL de l'échantillon dans un tube PCR identifié par le nom du Run et le sigle UN pour Unenriched. L'aliquot servira de contrôle-qualité des ISPs. Conserver l'échantillon à +4°C.
7. Transférer la totalité des ISPs resuspendu dans le puits 1 (100µL).
8. Remplir les différents puits de la barrette comme indiqué dans le tableau ci-dessous :

Well number	Reagent to dispense in well
Well 1 (well closest to the square-shaped tab)	Entire template-positive ISP sample (100 µL; prepared in step 3 of this procedure) (U)
Well 2	130 µL of Dynabeads® MyOne™ Streptavidin C1 Beads resuspended in MyOne™ Beads Wash Solution (prepared in step 2 of this procedure) (B)
Well 3	300 µL of Ion OneTouch™ Wash Solution (W)
Well 4	300 µL of Ion OneTouch™ Wash Solution (W)
Well 5	300 µL of Ion OneTouch™ Wash Solution (W)
Well 6	Empty
Well 7	300 µL of freshly-prepared Melt-Off solution (M)
Well 8	Empty

9. Disposer la barrette sur l'appareil dans l'emplacement prévu à cet effet de manière à ce que l'extrémité arrondie soit toujours à droite, plaquer la barrette vers la droite :

3.2 Lancement du Run

1. Allumer l'appareil et attendre qu'il termine son initialisation. Le message "rdy" apparaitra sur l'écran une fois l'initialisation terminée.
2. Appuyer sur Start/stop. L'écran affiche "run". Le run dure environ 35min.

PROCEDER AU LAVAGE DES SPHERES DANS LES 15MIN APRES LA FIN DU RUN : l'évaporation et le contact prolongé avec la solution dénaturante peut endommager les sphères et l'ADN.

3. A la fin du run, l'appareil bip toutes les minutes. Appuyer sur Start/Stop pour arrêter l'alarme et réinitialiser l'appareil pour un prochain run.
4. Récupérer le tube PCR contenant les sphères et procéder rapidement à leur lavage.

3.3 Lavage et récupération des sphères enrichies

1. Centrifuger les ISPs (15 500 g ; 1min30s).
2. S'assurer qu'il ne reste pas de billes streptavidine. Dans ce cas, un culot rouge est visible.
3. Laver les ISPs enrichies avec l'***Ion OneTouch™ Wash Solution*** :
    a. Eliminer le surnageant en laissant environ 10 µL sans disperser le culot puis ajouter 200 µL d'***Ion OneTouch™ Wash Solution***. Régler une pipette à 50 µL pour resuspendre les sphères par aspiration-refoulement (10 fois) et vortexer 5 fois 1s le tube.
    b. Centrifuger le tube (15 500 g ; 1min30s)
    c. Réaliser les étapes a et b une deuxième fois.
    d. Eliminer le surnageant en laissant environ 50 µL sans disperser les sphères.
    e. Régler une pipette à 20 µL et resuspendre les sphères (10x aspiration-refoulement). Transférer 2 µL de l'échantillon dans un tube PCR identifié par le nom du Run et le sigle EN pour Enriched. L'aliquot servira de contrôle qualité de l'enrichissement. Conserver l'aliquot à +4°C.
    f. Procéder au séquençage ou conserver l'échantillon à +4°C (3 jours maximum).
4. Eliminer le cône et la barrette laissés sur l'***Ion OneTouch™ ES***.

SEQUENÇAGE SUR PGM – ION PGM SEQUENCING 300 KIT

Objet : *Ce document décrit les différenctes étapes permettant l'initialisation du PGM et le lancement d'un run.*

10. Domaine d'application

Ce mode opératoire s'applique aux laboratoires de Génétique de l'Institut Curie Paris.

11. Responsabilités

(Ce paragraphe précise les responsabilités concernant l'application du mode opératoire et la vérification de sa bonne application)

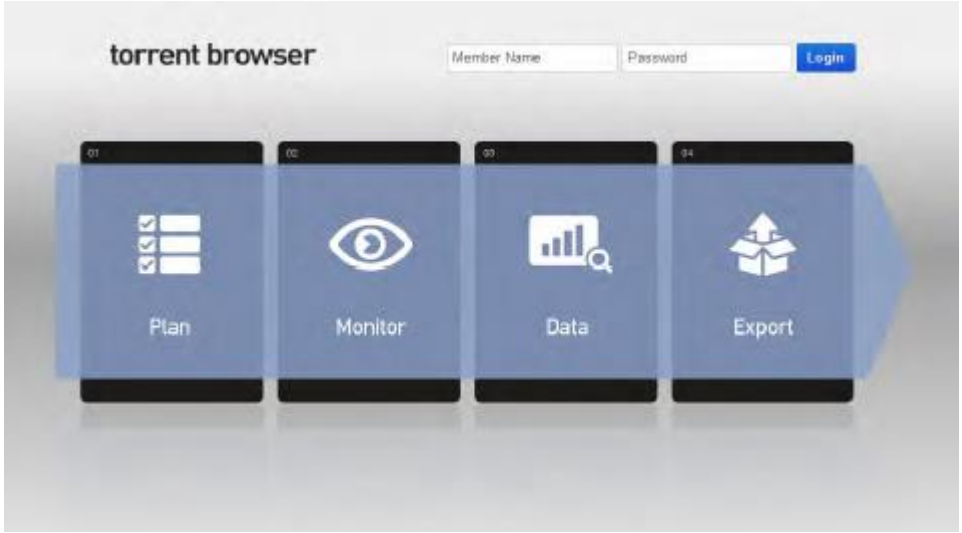
12. Description détaillée

- **Kits et réactifs**

Ion PGM™ Sequencing Supplies 300 Kit (Part no. 4480446)				
Component	**Color**	**Quantity**	**Volume**	Storage
Wash Bottle Sipper Tubes	Gray	8 for 250-mL bottles 4 for 2-L bottles	—	15°C to 30°C
Reagent Bottle Sipper Tubes	Blue	16	—	
Reagent Bottles w/ labels (50 mL)	—	25	—	
Wash 1 Bottle w/ label (250 mL)	Green	1	—	
Wash 2 Bottle w/ label (2 L)	Green	1	—	
Wash 3 Bottle w/ label (250 mL)	Green	1	—	
Ion PGM™ Sequencing Reagents 300 Kit (Part no. 4480447)				
Component	Cap color	Quantity	Volume	Storage
Ion PGM™ Sequencing 300 dGTP	Black	1 tube	100 µL	−30°C to −10°C
Ion PGM™ Sequencing 300 dCTP	Blue	1 tube	100 µL	
Ion PGM™ Sequencing 300 dATP	Green	1 tube	100 µL	
Ion PGM™ Sequencing 300 dTTP	Red	1 tube	100 µL	
Ion PGM™ Sequencing 300 Polymerase	Yellow	1 tube	36 µL	
Sequencing Primer	White	1 tube	144 µL	
Control Ion Sphere™ Particles	Clear	1 tube	60 µL	
Ion PGM™ Sequencing Solutions 300 Kit (Part no. 4480448)				
Component	Label color	Quantity	Volume	Storage
Ion PGM™ Sequencing 300 W2 Solution	Black	4 bottles	126.25 mL each	15°C to 30°C (store W2 Solution protected from light)
Ion PGM™ Cleaning Tablet	—	4	—	
Annealing Buffer	—	1 bottle	25 mL	
Ion PGM™ Sequencing 300 1X W3 Solution	—	2 bottles	100 mL each	

3.2 Création de la feuille de route

1. La création de la feuille de route se fait sur le serveur Torrent connecté au PGM.



2. Se connecter et cliquer sur *Plan*.
3. Cliquer sur *Plan Run* correspondant au protocole BRCA DIAG

4. Entrer le nom du run (TD pour un run lancé sur le Triton, ND pour un run lancé sur le Néréïde suivi du numéro du run. Exemple : TD7, ND9)
5. Entrer le nom de chaque échantillon selon le *barcode* qui lui est attribué (Exemple pour le run TD8 barcode 5 : TD8R5)
6. Se déconnecter du serveur

3.3 Nettoyage du PGM

Le séquenceur PGM nécessite un nettoyage avec de l'eau ultra-pure 18 MΩ ou une solution chlorite avant chaque nouvelle initialisation.

3.3.1 Fréquence de nettoyage

Nettoyage	Planning
Eau 18 MΩ	1. Avant chaque initialisation. 2. Après 1040 flux réalisés (1 run de 300pb ou 2 run 200pb). 3. Si plus de 27h et moins de 48h se sont écoulés entre l'initialisation et le début du run. 4. Si un nettoyage au chlorite a été réalisé depuis une semaine et que l'appareil n'a pas été utilisé.
Solution Chlorite	1. Une fois par semaine, sauf si l'appareil n'a pas été utilisé depuis le dernier nettoyage au chlorite (dans ce cas un nettoyage à l'eau est réalisé). 2. Si les réactifs ont été laissés sur l'automate depuis plus de 48h.

3.3.2 Préparation

Enlever toutes les bouteilles encore attachées au séquenceur.

Laisser les *sipper* sur l'appareil.

Utiliser les bouteilles réservées au nettoyage de l'appareil (trois bouteilles de 250 mL et une bouteille de 2 L). Une bouteille de 250mL est réservée pour le nettoyage au chlorite.

IMPORTANT : S'assurer que le circuit d'azote est ouvert avant d'allumer le séquenceur.

3.3.3 Nettoyage à l'eau ultra-pure 18 MΩ

1. Vider et nettoyer les quatre bouteilles à l'eau ultra-pure 18 MΩ en les rinçant 3 fois avec de l'eau ultra-pure. Mettre 250mL d'eau 18 MΩ dans 1 bouteille de 250 mL annotée Wash-1.
2. Sélectionner Clean PGM sur l'écran tactile.
3. Rincer le *sipper* W1 avec de l'eau 18 MΩ et mettre en place la bouteille Wash-1 contenant l'eau.
4. Suivre les inscriptions affichées sur l'écran (mettre des réservoirs sous les *sipper* Nucléotides, la bouteille de 250 mL annotée Wash-3 et la bouteille de 2L annotée Wash-2 sans sceller les bouchons).
5. Une fois le nettoyage terminé, sélectionner Next pour retourner au menu principal.

3.3.4 Nettoyage au chlorite

1. Vider et nettoyer les 4 bouteilles à l'eau ultra-pure en les rinçant 3 fois avec de l'eau ultra-pure (1 bouteille de 250 mL est annoté Chlorite).
2. Remplir une bouteille en verre avec un litre d'eau 18 MΩ et mettre une PGM™ Cleaning Tablet (comprimé de chlorite). Laisser dissoudre.
3. Une fois le chlorite dissous, ajouter 1mL de NaOH à 1M et filtrer la solution (filtre 0,22 µm). La solution une fois préparée doit être utilisée dans les 3h.
4. Sélectionner Clean PGM sur l'écran tactile.
5. Ajouter 250 mL de chlorite dans la bouteille Chlorite.
6. Rincer le *sipper* W1 avec de l'eau 18 MΩ et mettre en place la première bouteille Chlorite.
7. Suivre les inscriptions affichées sur l'écran (mettre des réservoirs sous les *sipper* Nucléotides, la bouteille de 250 mL annotée Wash-3 et la bouteille de 2L annotée Wash-2 sans sceller les bouchons).
8. Une fois le premier cycle de nettoyage terminé, remplacer la bouteille Chlorite par une bouteille remplie d'eau 18 MΩ après avoir nettoyer le *sipper* W1 avec de l'eau 18 MΩ. Sélectionner Next.
9. Quand le deuxième cycle de nettoyage est terminé, sélectionner Next pour retourner au menu principal.

3.4 Initialisation du PGM

3.4.1 Avant initialisation

1. Récupérer les bouteilles Wash-1, 2 et 3 du kit et s'assurer que moins de quatre initialisations ont été réalisé. Les nettoyer en les rinçant 3 fois avec de l'eau 18 MΩ.
2. Mettre les tubes de dNTP à décongeler sur la glace.
3. Préparer 1 mL de NaOH à 100 mM.

3.4.2 Préparation du Wash 2

1. Remplir la bouteille avec 2L d'eau 18 MΩ.
2. Transvaser la totalité d'une bouteille Ion PGM Sequencing 300 W2 Solution dans la bouteille (mettre de côté la bouteille pour scanner plus tard le barcode).
3. Ajouter 150 µL de solution NaOH 100 mM préparé précédemment.
4. Homogénéiser la bouteille par plusieurs retournements.

3.4.3 Préparation du Wash 1 et 3

1. Ajouter 350 µL de 100 mM NaOH dans la bouteille Wash 1. Fermer la bouteille.
2. Ajouter 50 mL de Ion PGM Sequencing 1X W3 Solution dans la bouteille Wash 3 et fermer la bouteille.

3.4.4 Initialisation

1. S'assurer qu'une puce déjà utilisée est en place.
2. Dans la page d'accueil, sélectionner ***Initialize***.
3. Sur la page suivant presser sur ***Enter Barcode*** et scanner le barcode du ***Ion PGM Sequencing 300 W2 Solution*** présent sur la bouteille.
4. Presser sur ***Next*** et confirmer qu'une puce est en place ainsi que des sippers et des poubelles. Presser de nouveau sur Next.
5. L'appareil vérifie à cette étape la pression du gaz.
6. Enfiler de nouveaux gants et installer un nouveau *sipper* sur le bouchon Wash 2. Faire en sorte que le *sipper* ne touche aucune surface.
7. Installer la bouteille Wash 2.
8. Installer de nouveaux *sipper* sur les bouchons Wash 1 et 3, et y installer les bouteilles Wash 1 et 3.

9. Presser sur ***Next*** pour lancer l'initialisation (durée environ 30 min).
10. Le PGM teste ensuite l'étanchéité des différentes bouteilles, remplit la bouteille Wash 1 avec de la solution W2 et ajuste le pH de la solution W2. Compter environ 30min.

Troubleshooting

L'initialisation peut parfois échouer car le PGM n'a pas réussi à ajuster le pH de la solution W2 à 7,65. Selon le pH atteint procéder de la manière suivante :

1. *Cas n°1 : 7,65 < pH W2*
   Si le pH final est d'environ 7,80 ajouter 5 µL d'HCl à 10mM dans la bouteille Wash 2. Bien mélangé et redémarrer l'initialisation en appuyant sur Start. Après un premier cycle (ajout Wash 1, homogénéisation et mesure pH de W2) le pH sera descendu à environ 7,2.

2. *Cas n°2 : pH W2 < 7,65*
   Transvaser le volume restant dans la bouteille Wash 1 dans la bouteille Wash 2. Ajouter 350 µL d'NaOH à 100mM dans la bouteille Wash 1.

   Retourner à l'écran d'accueil pour lancer une nouvelle initialisation.

3.4.5 Préparation des tubes contenant les dNTPs

IMPORTANT : Manipuler les tubes de dNTPs soigneusement afin d'éviter toutes contaminations. **Changer de gants à chaque transfert**. Les laisser sur la glace durant le protocole. Utiliser des cônes à filtre.

1. Préparer quatre tubes 50 mL en collant les étiquettes fournit dans le kit permettant de les identifier selon le dNTP qui y sera ajouté (dGTP, dCTP, dATP et dTTP). Les laisser dans la glace.
2. Vortexer et centrifuger en mode spin les différents tubes une fois les dNTPs décongelés.

3. Transférer 20 µL de chaque dNTP dans leur tube attitré. **Changer de gants à chaque transfert pour éviter toute contamination**.

3.4.6 Installation des tubes dNTPs

1. Une fois la préparation du Wash 2 terminée par le PGM, jeter les *sipper* des ports dNTP.
2. Avec de nouveaux gants, attacher de nouveaux *sipper* sur chaque port dNTP.
3. Attacher chaque tube contenant les dNTPs en respectant l'ordre indiqué sur l'appareil. Il n'est pas nécessaire de changer de gants à chaque fois. Bien serré en faisant attention à ce que les tubes soient bien droit.
4. Suivre les indications affichées sur l'écran pour poursuivre l'initialisation (vérification de la pression dans tubes et bouteilles, remplissage des tubes dNTP avec du Wash 2, et mesure de pH de chaque solution)
5. Sélectionner ***Next*** pour finir l'initialisation et retourner dans la page d'accueil. L'appareil est prêt pour le run. Procéder au chargement de la puce.

3.5 Chargement de la puce (Ion 316™/Ion 318™)

3.5.1 Préparation de l'échantillon à séquencer

1. Homogénéiser l'échantillon à la pipette 100 fois puis vortexer 5 fois 1 s.
2. Vortexer le tube ***Control Ion Sphere™*** (stocké à -20°C) et centrifuger.
3. Ajouter 5 µL de ***Control Ion Sphere™*** dans l'échantillon à séquencer (Volume final = 55 µL).
4. Ajouter ensuite 150 µL d'***Annealing Buffer***. Homogénéiser à la pipette 10 fois puis vortexer 10 fois 1 s.
5. Centrifuger le tube 2 min à 15 500g.
6. Eliminer le surnageant en plaçant le cône à l'opposé du culot pour éviter de le disperser (noter le sens à laquelle le tube à été centrifugé). Laisser 15 µL dans le tube (comparer avec un tube témoin). Garder le cône sur la pipette.
7. Ajouter 12 µL de Sequencing Primer.
8. Homogénéiser en vortexant 10 fois 1s puis à la pipette 100 fois par aspiration/refoulement (avec le cône de l'étape 5).
9. Disposer le tube sur le thermocycleur et lancer le programme Annealing (2 min à 95°C ; 2 min à 37°C)
10. Ajouter, juste après la fin du programme, 4 µL de ***PGM 300 Sequencing Polymerase*** (Volume final = 31 µL).
11. Homogénéiser à la pipette par 100 fois aspiration/refoulement et centrifuger en mode *spin*. Replacer le tube sur le thermocycleur régler à 20°C et laisser incuber 5 min. Procéder à la vérification et au lavage de la puce pendant ce temps.

3.5.2 Vérification et lavage de la puce

Précautions quant à l'utilisation de la puce

Ne jamais placer la puce directement sur une surface quelconque, toujours la disposer sur son adaptateur quand elle n'est pas dans le PGM.

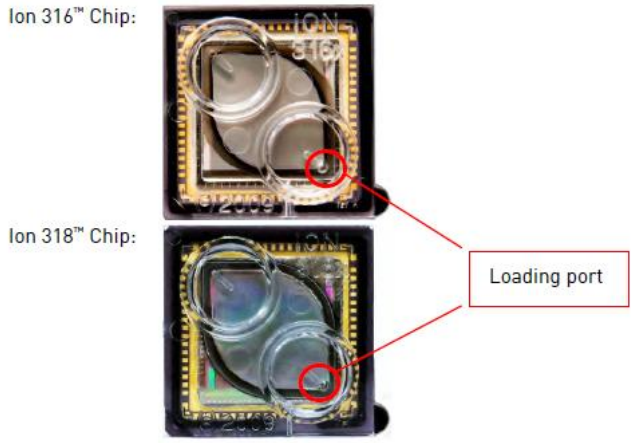
Ne jamais porter de gants lors de sa manipulation.

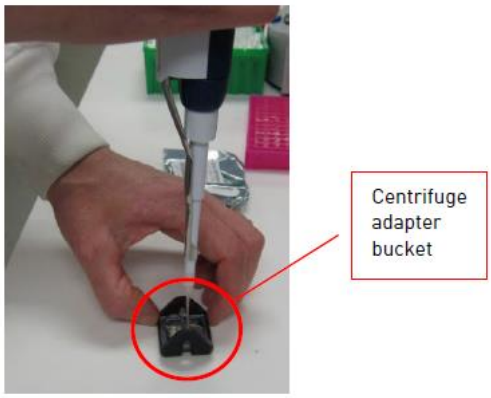
Poser son index sur la plaque situé à côté de l'emplacement pour la puce afin d'éliminer tout effet électrostatique à chaque fois que la puce doit être installé ou enlevé du PGM.

Vérification de la puce

Cette étape permet de vérifier que la puce ne présente aucun problème de fonctionnalité et que l'échantillon peut y être chargé.

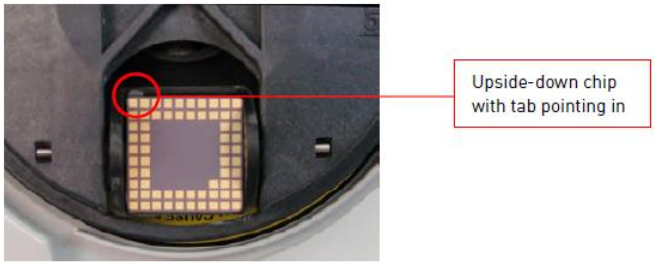
1. Sélectionner ***Run*** dans la page d'accueil et suivre les indications pour lancer la vérification.
2. Poser son index sur la plaque et installer la puce dans l'emplacement prévu à cet effet dans le PGM. Suivre les indications affichées sur l'écran.
3. Scanner le code barre situé sur le sachet de la puce.
4. Presser sur ***Chip Check***.
5. Vérifier l'absence de fuite autour de la puce.
6. Une fois la vérification terminée presser sur ***Next*** et suivre les indications pour enlever la puce et la remplacer par une puce usagée.





1. Incliner la puce à 45° de sorte que le loading port soit en bas
2. Aspirer le maximum d'*Annealing Buffer* restant à l'aide d'une pipette via le *loading port*. Eliminer le liquide.

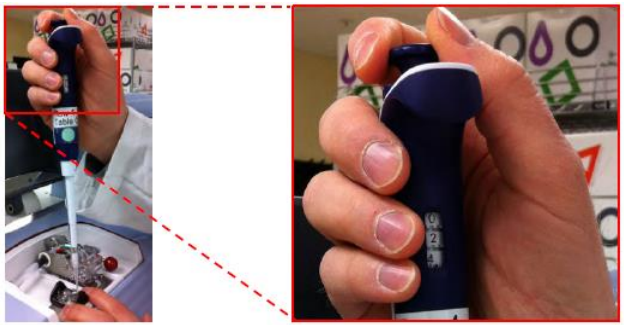
3. Placer la puce face vers le bas dans l'adaptateur et disposer l'ensemble dans la centrifugeuse avec la languette de la puce vers le centre.



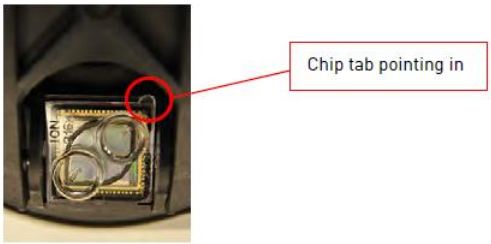
4. Centrifuger 5s pour vider complètement la puce. Disposer l'adaptateur sur la paillasse et enlever, à l'aide d'une pipette, le liquide restant.

Chargement de la puce

1. Récupérer le tube resté sur le thermocycleur et injecter, à l'aide d'une pipette Rainin SL-L100F, l'échantillon dans la puce :
    a. Régler la pipette à 30 µL. Laisser l'écrou de réglage volumétrique débloqué. Aspirer la totalité de l'échantillon.
    b. Insérer le cône dans le *loading port*. La pipette doit rester verticale tout au long du chargement.
    c. Injecter le liquide lentement et régulièrement en réglant le volume vers le bas à 1 µL par seconde jusqu'à 0,5 µL pour éviter l'introduction de bulles.



2. Centrifuger la puce 30s, la languette vers le centre.



3. Mélanger l'échantillon à l'intérieur de la puce :
    a. Régler la pipette à 20 µL.
    b. Incliner la puce à 45 degrés avec le loading port vers le bas. Descendre le piston de la pipette pour procéder à une aspiration et placer le cône à l'intérieur du *loading port*.
    c. Faire 3 aspirations/refoulements très lents pour éviter la formation de bulles.
4. Centrifuger la puce 30 s, la languette vers l'extérieur

5. Répéter de nouveaux les étapes 3. et 4. 3 fois en centrifugeant la puce languette vers l'intérieur, extérieur puis intérieur (la languette doit toujours être orientée vers l'intérieur lors de la dernière centrifugation).
6. Incliner la puce à 45 degrés et aspirer lentement le maximum de liquide via le *loading port*. Orienter la puce vers la gauche et la droite selon l'emplacement du liquide. Eliminer le liquide récupéré.
7. S'il reste encore de l'échantillon, réaliser une centrifugation de 5s en disposant la puce languette vers l'extérieur. Essayer de nouveau d'éliminer le liquide à l'aide de la pipette.
8. Procéder immédiatement au lancement du run.

Importation de la feuille de route et lancement du run

1. Presser sur ***Browse*** et sélectionner la feuille de route créée précédemment.
2. Vérifier que les différents paramètres soient corrects.
   Nombre de cycles : 160 cycles (640 flux)

   Planned Run : Nom du run donné

   Application : GENS

   Reference : hg19

   Barcode Kit : Ion Xpress

   Ion PGM Sequencing Kit: 300

3. Remplir les champs Project, Sample, Run Name et User Notes.

4. Après l'importation de la feuille de route, presser sur ***OK*** pour confirmer les paramètres.
5. Installer la puce chargée sur l'appareil puis presser sur ***Next***.
6. Vérifier au début du run d'éventuels fuites autour de la puce avec de fermer le couvercle.
7. Une fois le run terminé, l'écran retourne à la page d'accueil.

Analyses diagnostiques BRCA avec le PGM

Description

Plan

Etape 1 : Récupération des données

1) Récupération des données brutes
2) Récupération des données de qualité
   - QC reports
   - RGT graphs

Etape 2 : Analyse NextGene

1) Conversion des données
2) Analyse des données

Note sur la conservation des données

Note sur la 2ème analyse

Etape 3 : Import des données dans FileMaker

1) Création d'une série
2) Import des données

   Note sur le cas où 2 runs ont été passés pour une même série

Etape 4 : Validation des résultats

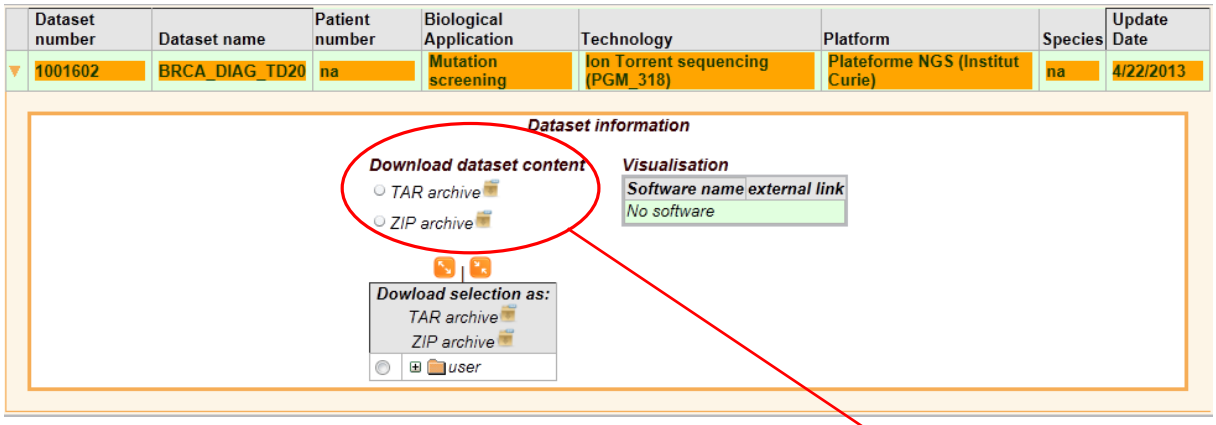
1) Vérification des données
2) Validation en Sanger

Annexes

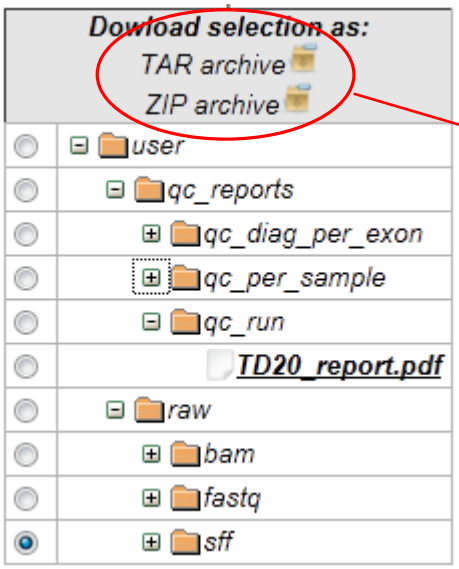
1) NextGene viewer
2) Paramètres d'alignement (SNV et Indel)
3) Logigramme d'analyse Indel récurrents
4) Logigramme de validation des résultats dans FMP

V. Etape 1 : Récupération des données

- Se connecter à http://bioinfo-portal.curie.fr/KDI
- Entrer son identifiant et son mot de passe
- valider en cliquant sur login
- Cliquer sur le projet « Diagnostic sur BRCA1 et BRCA2 »
- Dans « Dataset filters », sélectionner votre run : Dataset Name = n° de Run
- Cliquer sur Apply filters, une ligne comportant votre run doit apparaitre :
- Cliquer sur la flèche orange :



- Cliquer sur "+" pour dérouler l'arborescence :



Attention : totalité des fichiers téléchargés en

Pour télécharger seulement les fichiers sélectionnés,

C. 1) Récupération des données brutes

- Sélectionner les données sous le format **.sff**
- Télécharger les données en cliquant sur « *TAR archive* » (~8Gb soit ~10min de téléchargement)
- Sauvegarder sur le bureau du PC NextGene, dans un nouveau dossier au nom du run, puis dans un sous-dossier nommé « Données brutes »
- Extraire les données (clic droit sur le fichier « tar » > extraire ici)

D. 2) Récupération des données de qualité

➢ QC reports

- Sélectionner le rapport de run (ex : qc_reports > qc_run > TD20_report.pdf)
- Sélectionner les rapports « diagnostics » par échantillon (ex : qc_reports > qc_per_sample > TD20R1 > qc_diag > TD20R1-NGS-QcDiagnocticReport.pdf)
- Sauvegarder dans Synapse : P:\Analyses NGS\QC Reports\ « dossier au nom du run »

➢ RGT graphs

- Sélectionner les fichiers d'analyses RGT (ex : analyses_target-pgm_U900 > TD20.RGT_analysis > TD20.RGT.foldchange_graphs.pdf + fichier « csv »)
- Sauvegarder dans Synapse : P:\Analyses NGS\RGT graphs\ « dossier au nom du run »

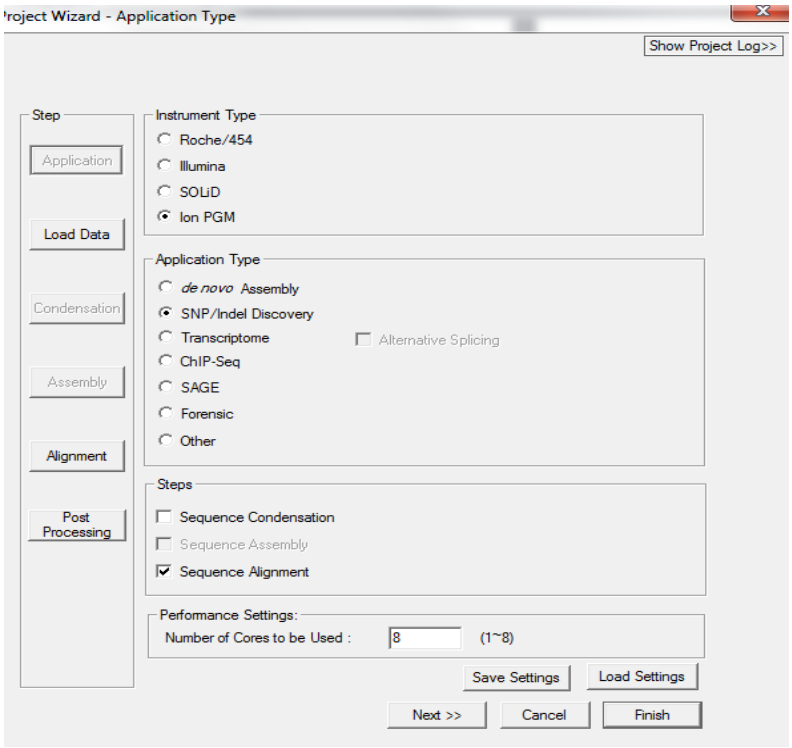
VI. Etape 2 : Analyse NextGene

Note : Attention à la version NextGene utilisée. Selon la version des paramètres peuvent varier.

Version actuelle : v2.3.1

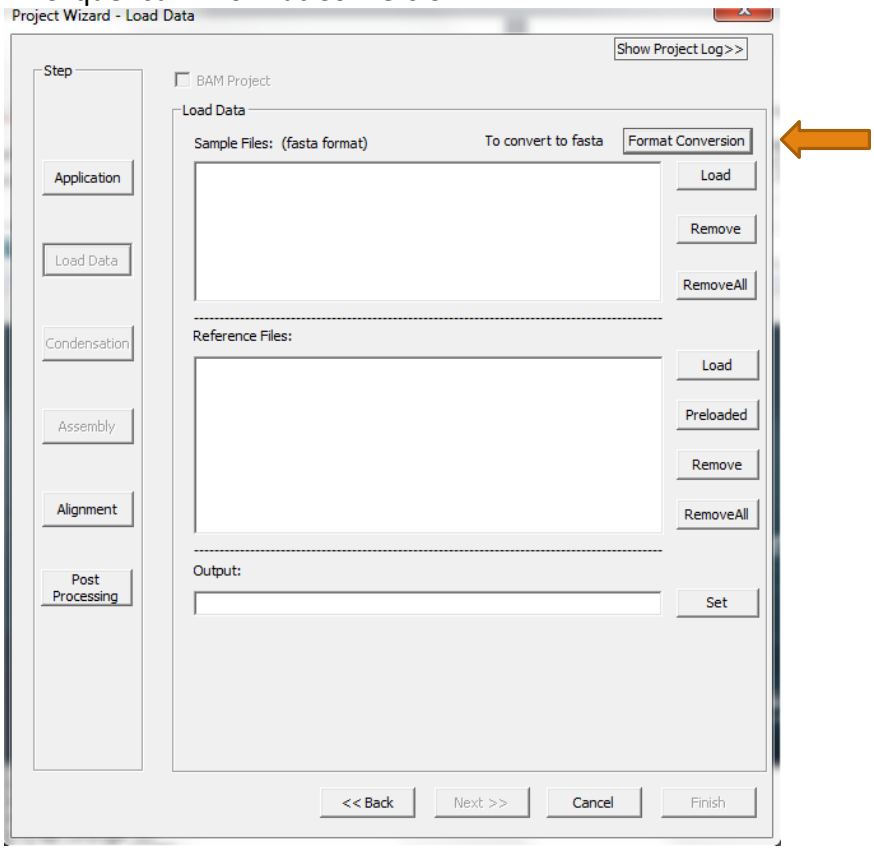
1) Conversion des données

- Ouvrir le logiciel NextGene. La fenêtre « Project Wizard – Application Type » s'ouvre. (si non : dans « File », cliquer sur « Open Project Wizard »)
- Cocher les paramètres suivants (Reséquencage sur PGM)

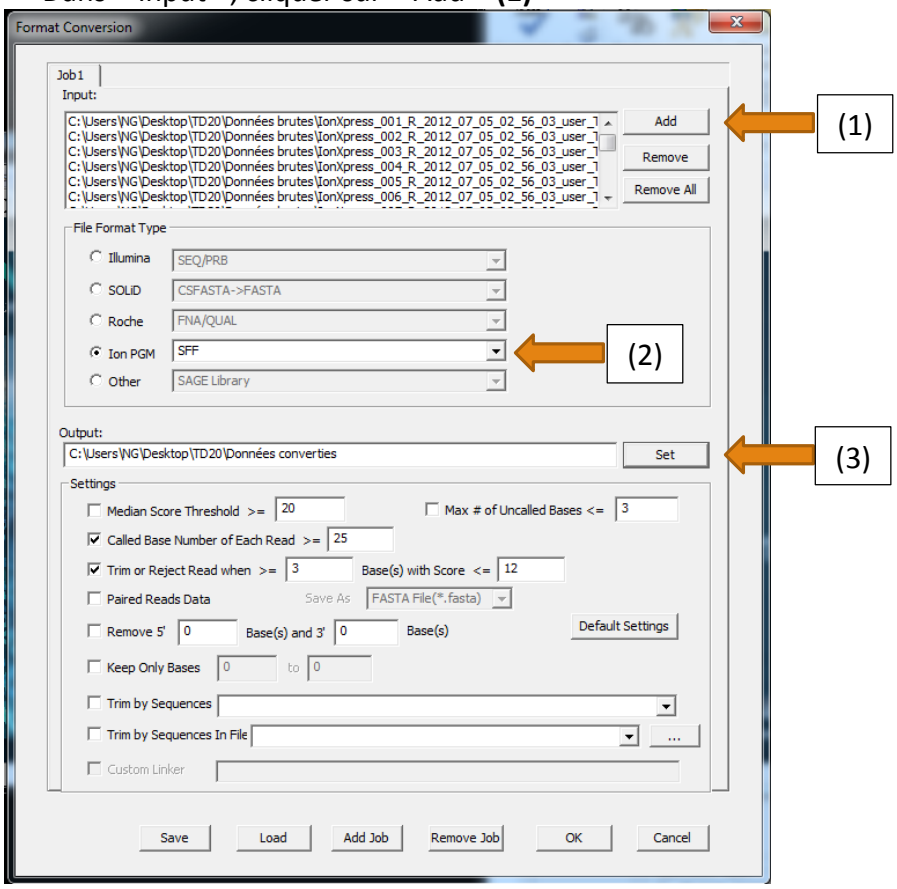


- Cliquer sur « Next »

- Cliquer sur « Format Conversion »



- Dans « Input », cliquer sur « Add » **(1)**



- Charger les données brutes précedemmment téléchargées au format .sff (tous les fichiers peuvent être traités en bloc)
- Cliquer sur Ion PGM, sélectionner SFF **(2)**
- Cliquer sur « Set » pour l'output (=dossier de destination des données converties) **(3)**
- Dans « Settings », utiliser les paramètres de conversion par défaut (indiqués ci-dessus)
- Dans le dossier du run créer un dossier « Données converties »
- Cliquer sur « OK » (durée de conversion env 1 min par Barcode)

2) Analyse des données

L'analyse des données de SNV et d'Indel se font séparément (car les paramètres de filtrage des données de séquences sont différentes).

- Dans le dossier de run (sur le bureau du PC NextGene), créer 2 sous-dossiers d'analyses pour les SNV et les Indel.
  Exemple :

Attention : La structure et le nom des dossiers doivent être strictement respectés pour la prise en charge automatique d'import dans Filemaker.

L'analyse de l'ensemble des données de séquences d'une série peut se faire automatiquement grâce à un fichier « .ngjob ». Celui-ci contient les informations des fichiers d'échantillons, des séquences de référence et des configurations d'analyse. Il est généré grâce à un script créé sous FileMaker Pro.

5. - - Analyse des SNV

> Création du fichier .ngjob

- Dans Synapse : P:\Analyses NGS, ouvrir le fichier FileMaker « creation_NGJOB.fmp12_v2 »

Note : chemin de connection à Synapse : \\Synapse\analyses_NGS

Login : ID@curie.net

- Sélectionner (voir exemple page suivante) :

  > le nom du projet (= nom du run) **(1)**

  > le chemin du 1er échantillon (dans « Données converties » **(2)**

  Bien sélectionner le fichier « converted.fna » (et pas le fichier « removed.fna »)

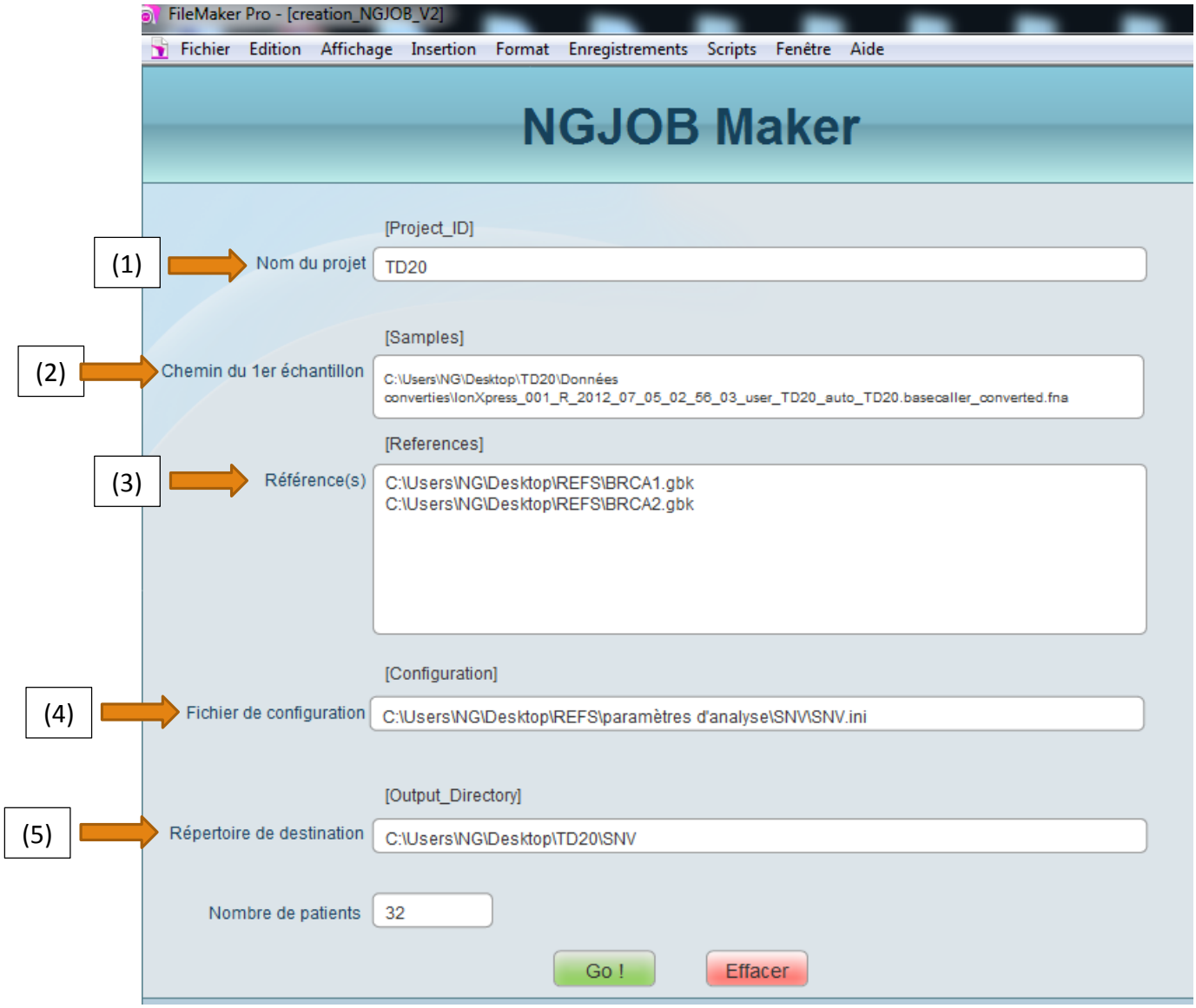
  > les séquences de référence « BRCA1.gbk » et « BRCA2.gbk » (dossier REFS sur le bureau) **(3)**

  > le fichier de configuration « SNV.ini » **(4)**

  > le répertoire de destination (SNV) **(5)**

  > le nombre de patients

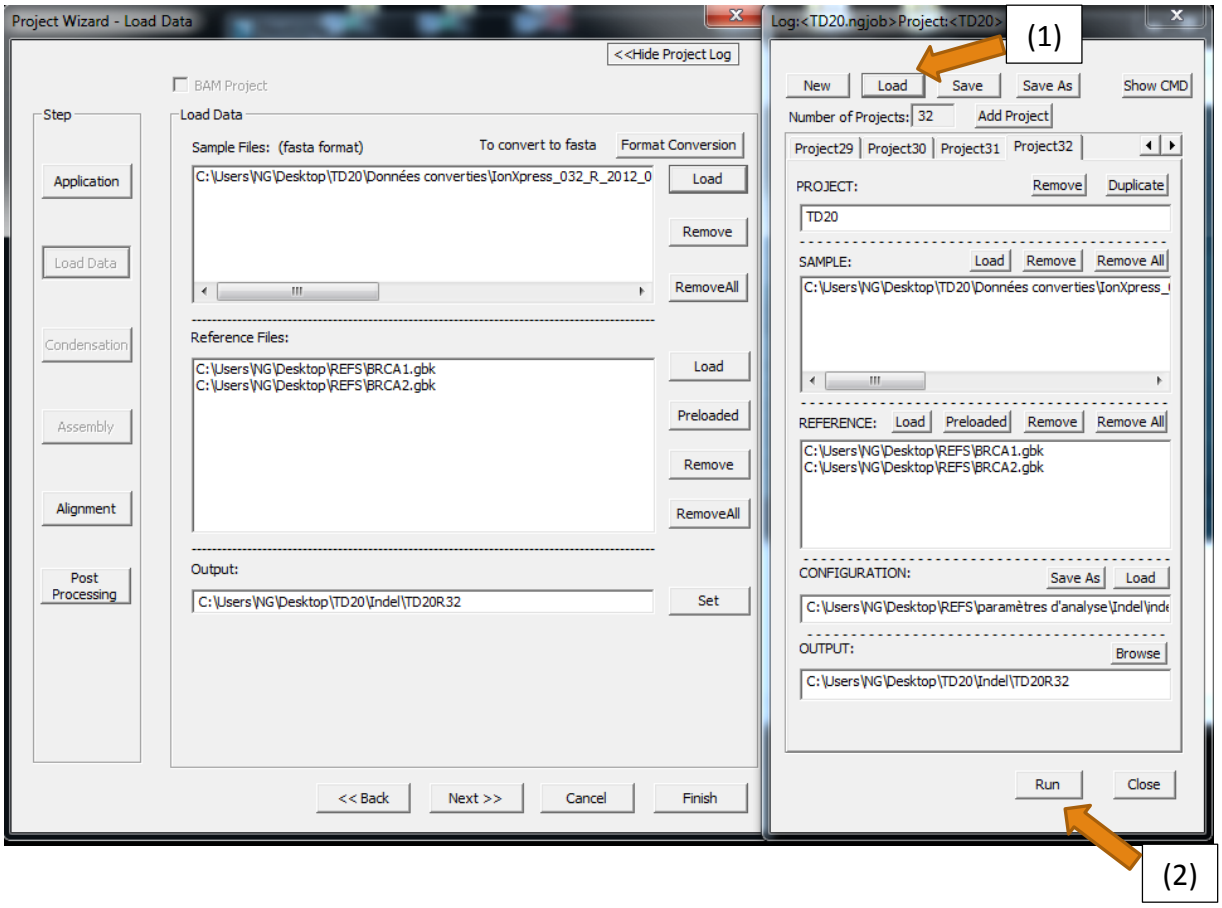
- Cliquer sur « Go ! »



- Un fichier .ngjob portant le nom du run est créé sur le bureau du PC NextGene.

> Analyse NextGene (voir exemple page suivante)

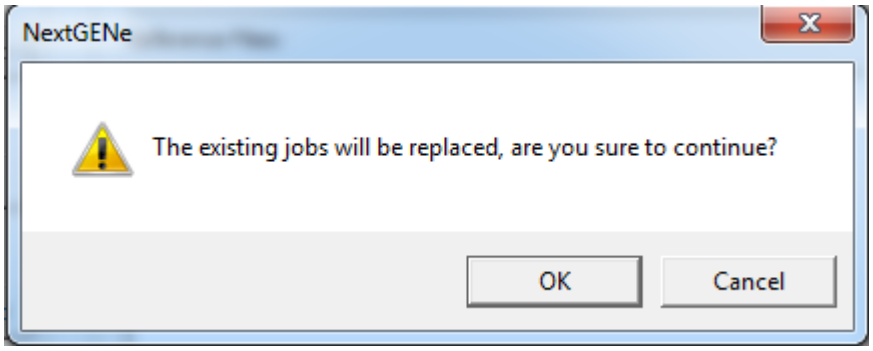
- Cliquer sur « File » > « Open Project Wizard »
- Cliquer sur « Show Project Log », une nouvelle fenêtre s'ouvre > cliquer sur « Load » **(1)**
- Sélectionner le fichier .ngjob du run à analyser
- Vérifier que les différents paramètres importés soient corrects
- Cliquer sur « run » **(2)** (durée d'analyse env 1 min par Barcode)



- Vérifier pour le 1er échantillon que le tableau de « Mutation Report » soit correct :

Ouvrir sous excel le fichier correspondant, et vérifier que dans la colonne « Mutation call » correspondent bien des SNV.

Note : le message suivant apparait après l'import d'un 2è fichier .ngjob > cliquer sur ok :



6. - - Analyse des Indel

Procéder de la même façon que pour l'analyse des SNV, en modifiant les paramètres de configuration et le fichier de destination :

- Créer un fichier .ngjob, avec les mêmes données, sauf pour :

  > le fichier de configuration « Indel.ini »

  > le répertoire de destination (Indel)

- Cliquer sur « Go ! »

- Dans NextGene, importer le fichier .ngjob du run à analyser

- Vérifier que les différents paramètres importés soient corrects

- Cliquer sur « run » (durée d'analyse env 1 min par Barcode)

- Vérifier pour le 1er échantillon que le tableau de « Mutation Report » soit correct :

Ouvrir sous excel le fichier correspondant, et vérifier que dans la colonne « Mutation call » correspondent bien des Indel.

Transférer les dossiers « SNV » et « Indel » générés sur Synapse

P:\ Analyses NGS\BRCA_multiplicom\analyse\« dossier au nom du run »

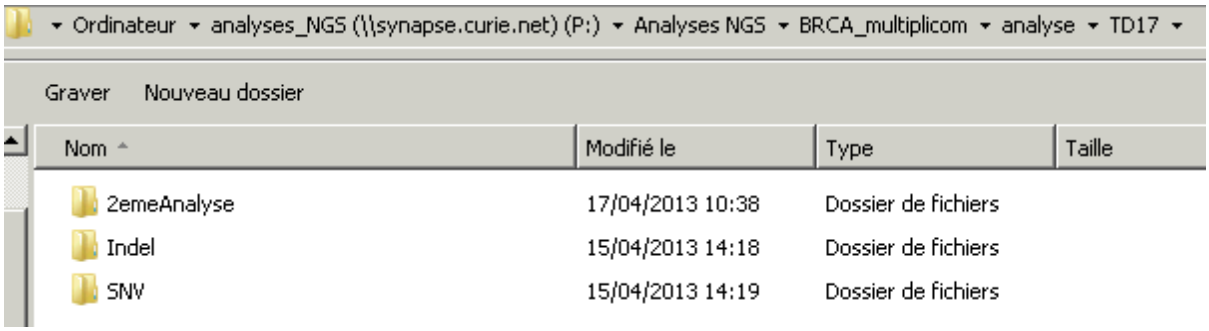
E. Note sur la conservation des données :

- Les données brutes et converties (stockées sur le PC NextGene) sont supprimées au fur et à mesure de la validation des séries.

- Les données analysées (dossiers « SNV » et « Indel » sont conservées sur Synapse.

F. Note sur la 2ème analyse :

Création sur le "PC NextGene", dans le dossier de run, d'un sous-dossier noté "2emeAnalyse". Ce dossier est ensuite copié sur Synapse, comme dans l'exemple suivant:



Questions sur la 2ème analyse:
- faut-il repartir du départ, càd des données brutes extraites de KDI, ou seulement vérifier la cohérence des noms de fichiers des données brutes et converties?
- puisque l'import des données et l'analyse NextGene se font automatiquement grâce au fichier ".ngjob", faut-il refaire la manip ou seulement vérifier la cohérence des no

VU

NANCY, le **16 septembre 2013**

Le Président de Thèse

Professeur P. JONVEAUX

NANCY, le **17 septembre 2013**

Le Doyen de la Faculté de Médecine

Professeur H. COUDANE

AUTORISE À SOUTENIR ET À IMPRIMER LA THÈSE N°6617

NANCY, le 26/09/2013

LE PRÉSIDENT DE L'UNIVERSITÉ DE LORRAINE

Professeur P. MUTZENHARDT

RÉSUMÉ DE LA THÈSE

Le travail de thèse proposé a pour sujet l'utilisation d'une plateforme de séquençage à haut-débit dans l'étude des gènes *BRCA1* et *BRCA2*. Ce projet a été réalisé en collaboration avec le laboratoire de Génétique de l'Institut Curie de Paris. La première partie d'introduction présente l'intérêt de l'étude des gènes *BRCA1* et *BRCA2* dans les prédispositions génétiques aux cancers du sein et de l'ovaire. Elle rappelle également les différents mécanismes de mutationnels (incluant l'impact sur l'épissage) et les principes généraux du séquençage à haut-débit.

La seconde partie concerne le champ d'application de la technique de séquençage à haut-débit par le séquenceur Personal Genome Machine ® (PGM) dans le cadre de la prise en charge diagnostique des prédispositions génétiques aux cancers du sein (et de l'ovaire). Elle présente le protocole de cette technique utilisé ainsi que les difficultés d'analyse et de traitements des données rencontrées dans le cadre d'une utilisation diagnostique.

La troisième partie concerne l'utilisation du PGM au sein d'un projet de recherche visant à utiliser la puissance des outils de séquençage à haut-débit afin d'évaluer l'impact des variants d'épissage pour les gènes *BRCA1* et *BRCA2*. Une approche par séquençage des produits de RT-PCR obtenus pour plusieurs patients porteurs de mutation affectant l'épissage a permis d'obtenir plusieurs données intéressantes et prometteuses. L'analyse bio-informatique des résultats reste cependant très délicate et laisse présager la nécessité de tester d'autres approches expérimentales afin d'affiner l'utilisation de cette technique dans l'évaluation de l'impact des variants d'épissage.

Les plateformes de séquençage à haut-débit sont en réalité des plateformes de biologie moléculaire aux multiples champs d'application. L'intérêt de ce sujet de thèse réside donc dans la description d'une approche maitrisée du séquençage à haut-débit (diagnostic *BRCA1* et *BRCA2*) et le développement d'une activité de transfert (impact des variants d'épissage).

TITRE EN ANGLAIS

BRCA1 and BRCA2 analysis by using next-generation sequencing platforms

---

THÈSE : MÉDECINE SPÉCIALISÉE – ANNÉE 2013

MOTS CLEFS : BRCA1, BRCA2, Séquençage haut-débit, diagnostic, variants, épissage

---

INTITULÉ ET ADRESSE :

UNIVERSITÉ DE LORRAINE

Faculté de Médecine de Nancy

9, avenue de la Forêt de Haye

54505 VANDOEUVRE LES NANCY Cedex