

THESE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE PARIS XIII

Ecole doctorale Sciences, Technologies, Santé « GALILEE »

Présenté et soutenue par

Morad EL BOUCHTAOUI

**NOUVELLE APPROCHE EXPERIMENTALE
EN GENETIQUE CONSTITUTIONNELLE
DES MALADIES HUMAINES**

Directeur de thèse

Professeur Guilhem BOUSQUET

JURY

*Madame Laurence Verneuil
Madame Anne Janin
Madame Mélanie Di Benedetto
Monsieur Guilhem Bousquet
Monsieur Jean-Paul Feugeas
Monsieur Claude Delcayre*

*Présidente/ rapporteur
Examinatrice
Examinatrice
Directeur de thèse
Rapporteur
Examineur*

REMERCIEMENTS

Je tiens tout d'abord à remercier l'ensemble des membres du jury pour l'intérêt qu'ils ont porté à ce travail.

Je remercie le professeur Laurence Verneuil qui m'a permis de réaliser ce travail et me fait l'honneur d'être rapporteur et de présider mon jury.

Je remercie le professeur Jean-Paul Feugeas qui me fait l'honneur d'être rapporteur et de faire partie de mon jury de thèse.

Je remercie le Maître de Conférence Universitaire HDR Mélanie Di Benedetto d'avoir accepté d'examiner ce travail.

Je remercie le professeur Anne Janin et mon directeur de thèse, le professeur Guilhem Bousquet, qui me font l'honneur de faire partie de mon jury de thèse. Je les remercie chaleureusement pour leurs enseignements, leur patience et leur pédagogie. Qu'ils soient assurés de ma profonde gratitude.

Je remercie tout particulièrement Mme la directrice du laboratoire, le Pr Anne Janin pour m'avoir fait confiance pendant toutes ces années. Depuis décembre 2006, ma date d'arrivée dans le laboratoire, vous m'avez constamment encouragé à l'évolution professionnelle et personnelle, toujours avec bienveillance. Votre dévouement pour les autres n'a d'égale que mon affection.

Je terminerais par un proverbe : "***Choisis un travail que tu aimes, et tu n'auras pas à travailler un seul jour de ta vie***" (*Confucius*). Merci, car grâce à vous mon travail est une passion.

Je tiens à exprimer mes plus vifs remerciements au Pr Guilhem Bousquet qui fut pour moi un directeur de thèse attentif et disponible malgré ses nombreuses charges. Je suis extrêmement sensible à ses qualités humaines d'écoute et de compréhension. Sa compétence, sa rigueur scientifique et sa clairvoyance m'ont beaucoup appris. Ils ont été et resteront des moteurs de mon travail.

Je remercie bien entendu l'ensemble des membres de l'unité UMR_S1165 pour la bonne humeur et la convivialité qui règne dans l'équipe depuis toutes ces années.

Je remercie évidemment mes collègues de tous les jours. Tous particulièrement Guillaume, collègue de bureau, plus qu'un collègue, un ami.

Merci également à Christophe pour ces conseils, Toni pour son aide.

Je tenais à remercier également le reste de l'équipe, Annie, Mélanie, Luc, Philippe, Laurence, Charlotte.

Durant ces années, certains ont pris des chemins différents et j'ai une pensée pour eux : Stéphanie, Natalie, Allison et Rachida.

J'ai aussi croisé et travaillé avec de nombreuses autres personnes. Tout d'abord Nicolas et Sébastien du service de pathologie de l'hôpital Saint-Louis. Carèle et Dominique du l'équipe U944. Et Christelle de la plateforme de l'Institut Universitaire d'Hématologie. Enfin, un grand merci Chrystophe et Irmine Ferrera ainsi que Marcio de la plateforme ANIMA 5.

Enfin, un grand merci à ma famille pour leur soutien. Merci à ma petite femme Lucie pour son soutien indéfectible et pour m'avoir donné le meilleur des fils, Manil. Ce travail vous est dédié.

Table des matières

REMERCIEMENTS.....	2
INDEX DES TABLEAUX ET FIGURES.....	8
Introduction.....	10
I. Génétique des cancers.....	10
I 1. Bases moléculaires des cancers.....	11
I 1 1. Les cancers : des maladies génétiques somatiques.....	11
I 1 2. Proto-oncogènes, gènes suppresseurs de tumeur et gènes de réparation.....	12
I 2. Facteurs environnementaux.....	14
I 2 1. Consommation de tabac.....	15
I 2 2. Consommation d'alcool.....	15
I 2 3. Exposition aux radiations ionisantes.....	16
I 2 4. Indice de masse corporelle (IMC).....	16
I 3. Formes héréditaires de cancers.....	17
I 3 1. Génétique Mendélienne des cancers.....	17
I 3 2. Prédispositions aux cancers.....	19
I 3 2 1. Analyses de ségrégation.....	19
I 3 2 2. Prédispositions aux cancers identifiées.....	21
I 3 3. Liaison familiale et variant de forte pénétrance.....	22
I 3 4. Variant de moyenne et faible pénétrance.....	24
I 3 5. Les trois premières GWAS.....	24
I 3 6. Méta-analyse.....	25
I 4. Analyses génomiques haut débit.....	26
I 4 1. Séquençage nouvelle génération (NGS).....	26
I 4 2. Dans le cancer de l'ovaire.....	26
I 4 3. Dans le cancer gastro-intestinal.....	27
I 4 4. Dans les cancers pancréatiques.....	28
I 5. Mutation sporadique et forme familiale.....	28
II. Nouvelle approche en génétique constitutionnelle.....	30
II 1. Objectif du projet.....	31
II 2. Matériaux et méthodes.....	58
II 2 1. Echantillons de patients et de souris utilisés.....	58
II 2 1 1. Echantillons humains.....	58
II 2 1 2. Echantillons murins.....	58

II 2 2. Purification d'ADN analyse pangénomique des échantillons humains	59
II 2 3. Analyse pangénomique des échantillons humains	60
II 2 3 1. Défi technologique	60
II 2 3 2. La technologie OncoScan™ FFPE.....	60
II 2 4. Détection de la mutation de <i>MET</i> pR988C	61
II 2 4 1. PCR à haute résolution (PCR-HRM)	61
II 2 4 2. Séquençage Sanger.....	62
II 2 4 3. Digestion enzymatique	63
II 2 5. Création d'un modèle de souris transgéniques knock-in pour la mutation <i>MET</i> humain pR988C	64
II 2 5 1. Choix du fond génétique	64
II 2 5 2. Construction du plasmide.....	64
II 2 5 3. Transgénèse	67
II 2 5 4. Génotype des souris transgéniques <i>MET</i> pR988C.....	67
II 2 5 5. Identification de "Off-target"	68
II 2 5 6. Phénotype pR988C <i>MET</i> transgéniques.....	70
II 2 5 7. Inhibition fonctionnelle de la voie <i>MET</i>	72
II 3. Résultats	33
II 3 1. Cas index : une femme jeune avec un cancer du sein et d'autres maladies graves 33	
Figure : Généalogie de la patiente	Error! Bookmark not defined.
II 3 2. Identification d'une mutation constitutionnelle du gène <i>MET</i>	33
II 3 3. Identification de cette mutation <i>MET</i> pR988C, c2962C_T.....	41
II 3 3 1. Dans une séries de patients avec cancer du sein sporadique.....	41
II 3 3 2. Dans une séries de patients avec polyglobulie de Vaquez avec mutation JAK2 V617F	42
.....	43
II 3 4. Souris transgéniques <i>MET</i> pR988C, c2962C_T	44
II 3 5. Souris transgéniques <i>MET</i> pR988C, c2962C_T et maladies auto-immunes et cancers (Figure 2).....	50
II 3 6. Souris transgéniques <i>MET</i> pR988C, c2962C_T et syndromes myeloproliferatis (Figure 3)	52
II 3 7. Inhibition fonctionnelle de la voie <i>MET</i>	53
II 4. Discussion	58
II 5. Bibliographie	79

INDEX DES TABLEAUX ET FIGURES

Figure 1. Illustration du two-hit model de Knudson et Comings

Tableau 1 : Principales prédispositions génétiques et principaux gènes associés. Institut national du cancer 2017

Figure 2 : Généalogie de la patiente

Figure 3 : Méthodologie de la technologie oncoscan

Tableau 2 : Liste des mutations retrouvées par oncoscan sur le cancer du sein et les cellules myéloïdes CD34+

Figure 4 : Schéma des mutations retrouvées par oncoscan sur le cancer du sein et les cellules myéloïdes CD34+

Tableau 3 : Liste des LOH retrouvées par oncoscan à la fois sur le cancer du sein et les cellules myéloïdes CD34+

Figure 5 : Mutation de *MET* pR988C, c2962C_T constitutionnel et héréditaire

Figure 6: Etude de microsatellites autour du gène *VHL* chez le cas index et sa tante paternelle

Figure 7 : screening de la mutation *MET* pR988C, c2962C_T dans des cancer du sein sporadique et des polyglobulies de Vaquez

Figure 8 : Concentration de Cas9/sgRNA injecté et efficacité d'intégration

Figure 9 : Généalogie des souris transgéniques *MET* pR988C, c2962C_T obtenues

Figure 10 : Exemple de séquençage Sanger de souris transgéniques *MET* pR988C, c2962C_T

Figure 11 : Génotypage des souris transgéniques *MET* pR988C, c2962C_T

Figure 12 : Souris transgéniques *MET* pR988C, c2962C_T et maladies auto-immunes et cancers

Figure 13 : Souris transgéniques *MET* pR988C, c2962C_T et syndromes myeloproliferatis

Figure 14 : Localisation sur le récepteur des mutations connus de *MET*

Figure 15 : Voie de signalisation de Met

Figure 16 : Inhibition fonctionnelle de Met

Figure 17 : Cellules tumorales microdisséquées sur une coupe de cancer du sein

Figure 18 : Schéma de restriction de l'enzyme *XmnI*

Figure 19 : Séquence de sgRNA et de ssODN

Figure 20 : Méthode de la ddPCR Biorad

Figure 21 : Protocole d'analyse des souris transgéniques *MET* pR988C, c2962C_T

Figure 22 : Généalogie de la patiente avec la polyarthrite rhumatoïde de son père et de son grand-père paternel

Introduction

I. Génétique des cancers et facteurs de risque

I 1. Bases moléculaires des cancers

I 1 1. Les cancers : des maladies génétiques somatiques

Les cancers sont des proliférations clonales pathologiques causées par de multiples événements moléculaires (altérations génétiques et épigénétiques) d'une cellule de l'organisme. Le processus de transformation d'une cellule normale en une cellule tumorale consiste en une accumulation successive et séquentielle de mutations somatiques et permet de définir la majorité des cancers comme étant des maladies génétiques acquises. Il est possible de dénombrer au sein d'une cellule tumorale plus d'une centaine de mutations ¹.

De plus, il existe dans une même tumeur une hétérogénéité dans le nombre et le type de mutations somatiques retrouvées ².

On distingue trois grands types de mutations somatiques :

1. Les mutations des oncogènes. Ces mutations qui interviennent dans les différents processus de transformation maligne, de progression tumorale et de survie de la tumeur. Elles confèrent un avantage de croissance et ont été sélectionnées lors du développement des cellules tumorales. Ces mutations sont présentes en petit nombre, généralement moins de quinze et sont les mutations responsables de l'oncogenèse.

2. Les mutations des gènes suppresseurs de tumeur. Ils représentent le côté opposé du contrôle de la croissance cellulaire, agissant normalement pour inhiber la prolifération cellulaire et le développement de tumeur. Dans beaucoup de tumeurs, ces gènes sont perdus ou inactivés, enlevant ainsi les régulations négatives de prolifération cellulaire. Toutefois, ces mutations sont noyées dans la masse des autres mutations somatiques

3. Les "passenger mutations". Ces dernières ont un effet biologique neutre, ne confèrent pas d'avantage de croissance aux cellules tumorales et ne sont que le reflet de l'instabilité génétique de la cellule tumorale.

I 1 2. Proto-oncogènes, gènes suppresseurs de tumeur et gènes de réparation

En 2010, plus de 400 gènes impliqués dans l'oncogenèse ont été recensés par le Sanger Center dans le cadre du projet Cancer Genome Project ³. A ce jour, plus de 525 gènes impliqués dans le développement de cancers ont été identifiés ⁴ et ces gènes de cancer peuvent être schématiquement classés en trois grandes catégories : les proto-oncogènes, les gènes suppresseurs de tumeurs et les gènes de réparation. La catégorie des proto-oncogènes a été la première à être identifiée. Il s'agit de gènes impliqués dans la transduction des différents signaux reçus par la cellule et qui ont pour rôle de réguler positivement la prolifération et le cycle cellulaire. Leur altération dans les cellules tumorales active les proto-oncogènes en oncogènes, les rendant hyperactifs, favorisant ainsi la croissance cellulaire et diminuant la mort cellulaire programmée. Les proto-oncogènes ont été découverts initialement selon trois grandes voies historiques de recherche que sont l'analogie avec des oncogènes viraux, les études de transfection de ces gènes dans des modèles cellulaires comme les cellules *NIH3T3*, et l'étude des gènes impliqués dans les translocations observées dans les leucémies ou les lymphomes. L'activation des proto-oncogènes en oncogènes peut résulter d'une mutation ponctuelle avec un gain de fonction délétère de la protéine (mutations des gènes *RAS* dans de nombreux cancers), d'une amplification génique (amplification du gène *ERBB2* dans 25% des cancers du sein), de translocations chromosomiques dérégulant l'expression du gène (translocation juxtaposant le proto-oncogène *MYC* avec un gène codant une immunoglobuline) ou créant un gène de fusion codant pour une nouvelle protéine avec une

fonction aberrante (translocation récurrente t(9;22)(q34;q11) du chromosome Philadelphie de la leucémie myéloïde chronique, aboutissant à la création d'un gène de fusion entre les gènes BCR et ABL). Un seul événement moléculaire mutant un seul des deux allèles d'un proto-oncogène est nécessaire à leur activation, ce qui se traduit par un effet dominant des mutations activant les oncogènes.

La seconde catégorie correspond aux gènes suppresseurs de tumeurs qui possèdent des propriétés inverses aux proto-oncogènes puisqu'ils sont des régulateurs négatifs de la prolifération cellulaire et qu'ils favorisent l'apoptose. Ces gènes sont également appelés "gate-keepers" car ils sont responsables de la formation des points de contrôle ou checkpoints entre les différentes phases du cycle cellulaire. Ces "checkpoints" permettent aux cellules en division de prévenir et de réparer les anomalies de l'ADN survenues lors de la mitose ou lorsque ces anomalies sont trop importantes d'orienter la cellule vers l'apoptose. La troisième catégorie correspond aux gènes impliqués dans la réparation de l'ADN. Ces gènes ont pour rôle de maintenir l'intégrité de l'ADN et de réparer les anomalies génétiques susceptibles d'activer les proto-oncogènes et d'inactiver les gate-keepers. A l'inverse des proto-oncogènes, les gènes suppresseurs de tumeurs et les gènes de réparation sont inactivés dans les cellules tumorales. Cette perte de fonction peut également faire intervenir différents mécanismes. Il peut s'agir de délétions d'une région du chromosome contenant le gène suppresseur de tumeur, de mutations ponctuelles qui altèrent sa fonction, d'une interaction avec un oncogène viral ou encore d'altérations épigénétiques qui vont diminuer l'expression du gène (hyperméthylation du promoteur du gène *MLH1* dans les cancers du côlon). De plus, pour que la fonction d'un gène suppresseur soit perdue, il est nécessaire que les deux copies du gène soient inactivées ce qui explique le caractère récessif des anomalies moléculaires des suppresseurs de tumeurs au niveau cellulaire. Les altérations à la fois des proto-oncogènes et des gènes suppresseurs de tumeurs sont nécessaires pour la dérégulation de la croissance cellulaire. De plus,

l'emballlement du cycle cellulaire en l'absence de contrôle est responsable d'une instabilité génétique des cellules qui favorise la progression de la cellule vers le cancer et l'acquisition par la cellule tumorale de nouvelles propriétés comme la néo-angiogénèse, la capacité d'invasion tissulaire ou la résistance à une chimiothérapie.

I 2. Facteurs de risque environnementaux et comportementaux

En 2011, le nombre de nouveaux cas de cancer en France a été estimé à 365 500 pour les deux sexes confondus. Entre 1995 et 2005, les taux d'incidence tous cancers confondus ont augmenté de 14 % chez l'homme et de 17 % chez la femme. L'augmentation de l'espérance de vie et l'amélioration du dépistage ne suffisent pas à eux seuls à expliquer ces chiffres. Les modifications de l'environnement sont, entre autres, largement suspectées de contribuer à cette augmentation. À l'échelle mondiale et selon l'OMS, 19 % de tous les cancers peuvent être attribués à l'environnement, y compris au milieu professionnel, et sont à l'origine de 1,3 million de décès chaque année.

Sans qu'il soit possible, en l'état actuel des connaissances, d'estimer avec précision la part de l'augmentation liée aux expositions environnementales, le lien entre l'apparition de certains cancers et des expositions environnementales est suspecté voire même, dans certains cas, clairement établi. Pour l'Agence Française de Sécurité Sanitaire de l'Environnement et du Travail (AFSSET), l'environnement comprend les agents physiques, chimiques ou biologiques auxquels un individu peut être exposé dans les lieux de vie et de travail. Tandis que certains utilisent une définition plus restrictive de « l'environnement » en excluant les expositions professionnelles, d'autres, comme l'Organisation Mondiale de la Santé (OMS), intègrent également les facteurs relatifs aux modes de vie et les comportements individuels (tabagisme, consommation d'alcool, alimentation, activité physique...). Notons que le

développement d'un cancer résulte le plus souvent d'une association de différents facteurs et d'une accumulation de plusieurs altérations subies par les cellules tout au long de la vie et, probablement, d'une moindre efficacité des mécanismes de réparation des altérations au niveau de la cellule des personnes âgées. Ces mécanismes expliquent la fréquence accrue des cancers en vieillissant.

I 2 1. Consommation de tabac

Bien que la fumée du tabac soit une importante source de substances carcinogènes, pendant longtemps la cigarette n'a pas été considérée comme un facteur de risque établi de cancer du sein en population générale. Certaines études ont même rapporté un effet protecteur du tabac vis-à-vis du cancer du sein. En effet, fumer pourrait réduire le risque par un effet anti-œstrogénique ou une altération du métabolisme hormonal. Des études plus récentes ont suggéré une augmentation du risque chez les fumeuses, en particulier si l'exposition est longue ou débute avant la première grossesse, laissant penser que l'effet du tabac pourrait varier au cours de la vie et selon le début de l'exposition. En 2005, une analyse combinée de 20 études ayant mesuré à la fois l'exposition au tabac actif et au tabac passif a été réalisée. Cette étude a pu mettre en évidence une augmentation du risque de cancer du sein de 1.2 chez les femmes ayant été exposées directement ou indirectement à la fumée de cigarette par rapport aux femmes n'ayant jamais été exposées ⁵.

I 2 2. Consommation d'alcool

En population générale, l'alcool est le seul facteur de risque nutritionnel établi pour le cancer du sein. Ce risque relatif augmente d'environ 7% pour chaque 10 g additionnel d'alcool consommé par jour. En effet, comparativement aux femmes déclarant ne pas

consommer d'alcool, les consommatrices de 35 à 44 g d'alcool par jour ont un risque relatif de 1,32 et celles consommant plus de 45 g d'alcool par jour ont un risque relatif de 1,46. Par ailleurs, une étude montre une association linéaire du risque de cancer du sein avec une augmentation de la consommation d'alcool ⁶. Cette étude de la cohorte EPIC a trouvé une augmentation de 3% du risque de cancer du sein par verre d'alcool (10 g d'éthanol) par jour parmi 274 688 femmes dont 4 285 avaient développé un cancer du sein ⁶. L'alcool provoque une augmentation du niveau d'hormones dans le sérum et une production accrue de facteurs de croissance IGF (insulin-like growth factor). Les IGF agissent comme des mitogènes, inhibent l'apoptose et interagissent avec les œstrogènes. Une production accrue d'IGF augmente le risque de cancer du sein, surtout avant la ménopause ⁷.

I 2 3. Exposition aux radiations ionisantes

En population générale, les expositions aux radiations ionisantes ont été montrées comme étant associées à une augmentation du risque de cancer du sein ⁸. Par ailleurs, plusieurs études ont été menées sur les survivantes à la bombe atomique d'Hiroshima ou sur une population ayant subi des irradiations dans un cadre médical (irradiation par imagerie médicale ou à visée thérapeutique). Ces études montrent une augmentation du risque comprise entre 1,0 et 4,3 par Gray. Par ailleurs, plusieurs de ces études ont montré que le risque de cancer du sein associé aux irradiations était plus important pour les femmes ayant été exposées durant l'enfance ou l'adolescence ⁹⁻¹⁵.

I 2 4. Indice de masse corporelle (IMC)

Le surpoids est un facteur associé au risque de cancer du sein de façon différente, en fonction du statut ménopausique : le surpoids diminue le risque de cancer du sein avant la ménopause par unité d'IMC additionnelle), mais l'augmente en postménopause. En effet, avant la ménopause, les femmes en surpoids ont un plus faible nombre d'ovulations et une diminution du taux d'hormones sanguin. A la ménopause, la production d'œstrogènes est stoppée dans les ovaires mais elle se poursuit dans les tissus adipeux. Une femme ménopausée avec un IMC élevé a donc une production accrue d'œstrogènes par rapport aux femmes ayant un IMC « normal ».

I 3. Formes héréditaires de cancers et facteurs constitutionnels

I 3 1. Génétique Mendélienne des cancers

Si l'ensemble des cancers résultent d'une séquence d'anomalies génétiques et épigénétiques, on estime que moins de 5% des cancers correspondent à des formes héréditaires de cancers répondant à un déterminisme Mendélien. Ces cancers héréditaires sont liés à une anomalie constitutionnelle de l'ADN correspondant fréquemment à la mutation d'un gène suppresseur de tumeur ou de réparation de l'ADN, plus rarement d'un oncogène. En effet, chez les individus porteurs à l'état hétérozygote d'une mutation d'un gène suppresseur de tumeur, le premier événement moléculaire conduisant à la transformation maligne est déjà présent dans l'ensemble des cellules de l'organisme (sauf rares cas de mosaïcisme). Il suffit alors d'un seul événement moléculaire inactivant le second allèle pour perdre la fonction de la protéine, contre deux événements moléculaires dans une seule et même cellule dans le cas des cancers sporadiques. La présence chez un individu donné d'une mutation constitutionnelle d'un gène suppresseur de tumeur le prédispose donc à développer

des tumeurs plus fréquemment et plus précocement que dans la population générale. Cette prédisposition aux cancers liée à une mutation d'un gène suppresseur de tumeur (premier "événement génétique"), suivie d'une deuxième mutation somatique de l'allèle sain (second "événement génétique") correspond au "two-hit model" émis en 1971 par Knudson ¹⁶ (Figure 1).

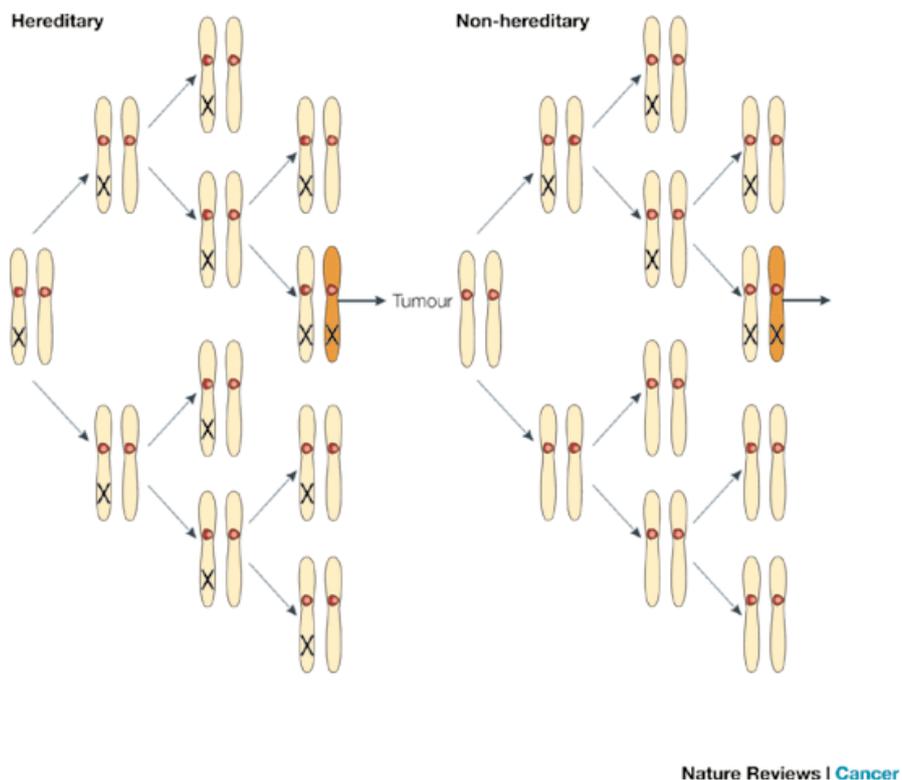


Figure 1. Illustration du two-hit model de Knudson et Comings

La prédisposition au rétinoblastome, où l'inactivation des deux allèles du gène *RB1* est nécessaire (mais peut-être pas suffisante) à la transformation tumorale des neuroblastes de la rétine du jeune enfant. Dans plus de 60 % des cas, l'atteinte est unilatérale et survient après l'âge de 12 mois. Dans la plupart de ces cas, les deux mutations ont été acquises. En revanche, dans près de 40 % des cas, l'atteinte est bilatérale et survient avant l'âge de 12 mois et, dans

près de 10 % des cas (25 % des cas bilatéraux), l'un des deux parents a été atteint dans l'enfance. Chez ces enfants, l'une des deux mutations du gène *RBI* est constitutionnelle, l'autre est somatique. Le gène *RBI*, cloné en 1986, est le premier gène de prédisposition aux cancers ayant été identifié. Son identification a ouvert un nouveau champ pour les études génétiques.¹⁷⁻¹⁹

La transmission à la descendance de ces mutations conduit à des tableaux familiaux avec de nombreux cancers dans la famille définissant ainsi les syndromes de prédisposition héréditaires aux cancers.

I 3 2. Prédisposition aux cancers

I 3 2 1. Analyses de ségrégation

Les analyses de ségrégation permettent en partie de répondre à ces interrogations. Ces analyses étudient la distribution familiale des cancers dans des familles et cherchent à déterminer le modèle qui explique le mieux les données observées, notamment à mettre en évidence un gène majeur parmi l'ensemble des facteurs intervenant dans le déterminisme d'une maladie. Ces analyses ont été faites soit dans des familles clairement identifiées comme appartenant à des syndromes familiaux, soit dans des familles recensées par l'intermédiaire d'un cas non sélectionné pour ses antécédents familiaux. Dans le premier cas, ces analyses cherchent surtout à confirmer l'existence d'un gène dominant rare responsable de cette catégorie de familles et essaient d'estimer les risques des sujets susceptibles. Dans le second cas, elles visent à proposer un modèle général expliquant l'excès de risque familial. Dans le cancer du sein, la plupart des études ont conclu à l'existence d'un gène majeur dominant rare

expliquant 5 à 10 % de tous les cas ²⁰. L'analyse de ségrégation basée sur l'étude CASH (Claus et coll., 1991) a proposé l'existence d'un gène dominant ayant une fréquence de 0,003 dans la population et associé à un risque chez les susceptibles de 14% avant 40 ans, 38% avant 50 ans et 67% avant 70 ans. L'évolution du risque en fonction de l'âge des sujets susceptibles et l'excès de risque chez les sujets jeunes, sont compatibles avec la théorie multi-étape des cancers puisque les sujets susceptibles ont déjà, par l'existence d'une mutation constitutionnelle, « franchi » une étape (ou au moins une étape) dans le processus de la tumorigenèse. De plus cette courbe est totalement compatible avec l'excès de cas précoces observé dans les familles « historiques ». Les méthodes utilisées présentent en général des limites majeures. En particulier, ces méthodes comparent des modèles entre eux, de sorte que le modèle génétique retenu est le plus vraisemblable parmi ceux qui sont comparés, ce qui ne veut pas dire que les distributions familiales de cas soient effectivement expliquées par ce modèle. Ainsi les études qui ne comparent que des modèles monogéniques ne peuvent conclure à l'existence d'autres formes d'hérédité. Dans cet ordre d'idée, Dizier et coll. (1993) ont montré que l'on pouvait conclure à l'existence d'un gène dominant en présence de données générées dans un modèle supposant l'existence de deux gènes interagissant entre eux (phénomène d'épistasie). Par ailleurs, comme cela est désormais possible grâce aux développements des modèles dits régressifs (Bonney, 1986 ; Abel et Bonney, 1990; Demenais, 1991), il est important de prendre en compte l'effet de covariables environnementales qui peuvent expliquer une ressemblance familiale non génétique due aux expositions communes dans la famille. Ceci sera largement commenté dans le chapitre 6 "Intéactions entre facteurs génétiques et facteurs de la vie reproductive dans l'étiologie du cancer du sein". Enfin, il faut remarquer que les estimations des fréquences géniques et des pénétrances sont biaisées si l'on ne prend pas correctement en compte le recensement des familles.

I 3 2 2. Gènes connus de prédisposition aux cancers

En moins de 20 ans, une quarantaine de gènes de prédisposition aux cancers ont été identifiés (*Tableau 1*). Ils correspondent aux situations de prédisposition les plus simples à mettre en évidence, la partie émergée d'un iceberg que représenterait l'ensemble des prédispositions aux cancers. En effet, il s'agit soit de prédispositions transmises selon le mode dominant et associées à un risque tumoral élevé, conduisant alors souvent à une concentration familiale de cancers, soit de syndromes dont les manifestations désignent la prédisposition.

Prédispositions génétiques	Principaux gènes associés
Syndrome seins-ovaires	BRCA1, BRCA2, PALB2, RAD51
Syndrome de lynch	MLH1, MSH2, MSH6, PMS2, EPCAM
Adénomes hypophysaires familiaux	AIP
Ataxie-télangiectasie	ATM, MRE11A
Cancer gastrique diffus familial	CDH1
Carcinome papillaire rénal héréditaire	FH, MET
Hyperparathyroïdisme	CDC73, CASR
Maladie de cowden	PTEN, PIK3CA
Maladie de fanconi	FANC
Maladie de von hippel-lindau	VHL
Mélanome malin familial	CDKN2A, MITF, BAP1, CDK4
Néoplasies endocriniennes	MEN1, RET, CDKN1B
Neurofibromatoses	NF1, NF2, LZTR1, SMARCB1, SPRED1
Phéochromocytome-paragangliome héréditaire	SDH, TMEM127, MAX, EPAS1
Polyposes adénomateuses familiales	APC, MUTYH
Rétinoblastome	RB1
Syndrome de birt-hogg-dubé	FLCN
Syndrome de bloom	BLM
Syndrome de carney	PRKAR1A
Syndrome de gorlin	PTCH1
Syndrome de li-fraumeni	TP53, CHEK2
Syndrome de nijmegen	NBN
Syndrome de peutz-jeghers	STK11
Syndrome de polypose juvénile	BMPR1A, SMAD4
Xeroderma pigmentosum	XP

Tableau 1 : Principales prédispositions génétiques et principaux gènes associés. Institut national du cancer ; 2017

I 3 3. Liaison familiale et variants de forte pénétrance

L'un des grands succès des approches de liaison est l'identification des gènes de prédisposition aux cancers du sein et de l'ovaire, *BRCA1* et *BRCA2*. Après des études de ségrégation ayant retenu un modèle dominant avec ses paramètres de fréquence génique et de pénétrance, la réunion d'une trentaine de familles sélectionnées a permis à Mary-Claire King, en 1990, de localiser *BRCA1* en 17q2²¹. Puis, dans le cadre d'un consortium de laboratoires, la position du gène recherché a pu être finement précisée, et a conduit à son identification^{22,23}. Quant à l'approche gène-candidat, l'identification des gènes *MLH1*, *MSH2* et *MSH6*, dont les altérations sont associées à un risque élevé de cancer du côlon et de l'endomètre : alors que des régions du génome perdues de façon récurrente dans des adénocarcinomes du côlon de personnes s'inscrivant dans un syndrome HNPCC (hereditary non-polyposis colorectal cancer) étaient recherchées par une approche de perte d'hétérozygotie de régions microsatellites, des pertes d'allèle, et non des gains d'allèle, ont été en réalité mises en évidence dans ce type de tumeurs, évoquant un défaut de réparation d'erreurs de réplication de l'ADN (phénotype MSI, *microsatellites instable*). Ces gains d'allèles étaient très semblables à ce qui était observé chez certaines bactéries ; les homologues humains des gènes altérés chez ces bactéries, en l'occurrence *hMLH1*, *hMSH2* et *hMSH6*, ont été identifiés, et des mutations constitutionnelles de ces gènes ont été mises en évidence²⁴.

Ces dernières années ont montré que les gènes de prédisposition aux cancers n'appartenaient pas à des voies cellulaires isolées, dont les altérations ne conduiraient qu'à des cancers. Ainsi, des anomalies du développement empruntent des voies communes à certaines prédispositions aux cancers : voies Sonic-Hedgehog-Gli, avec holoprosencéphalie et syndrome de Greig d'une part, et syndrome de Gorlin, d'autre part, syndrome associé à un risque majeur de carcinomes basocellulaires multiples et à un risque plus faible de

médulloblastome²⁵. De la même façon, les gènes de la cascade des TGF- β sont impliqués à la fois dans des anomalies vasculaires et des tumeurs digestives²⁶. Une autre voie passionnante mérite d'être mentionnée, celle du facteur HIF-1 : il s'agit d'un gène clé de l'induction de l'angiogenèse, dont l'expression est finement régulée par différents gènes, dont le gène VHL à l'origine de la maladie de von Hippel Lindau. D'autres gènes, codant pour des enzymes du cycle de Krebs, ont été récemment impliqués dans certaines prédispositions aux cancers : gènes des sous-unités de la succinate déshydrogénase (SDH) et gène de la fumarate hydratase (FH), dont les altérations à l'état hétérozygote sont, respectivement, associées à un risque de paragangliomes et de cancer du rein. À l'état homozygote ou hétérozygote composite, ces altérations sont responsables d'anomalies graves du métabolisme et de troubles neurologiques²⁷.

Il est difficile de faire une présentation synthétique et exhaustive de l'ensemble des prédispositions aux cancers identifiées aujourd'hui : faut-il retenir une présentation par gène, par type de mutation (mutations activatrices, inactivatrices), par identité de la voie cellulaire dans laquelle il est impliqué, par mode de transmission de la prédisposition (dominant, récessif), par maladie prédisposante, par localisation tumorale ? Les connaissances récentes viennent encore compliquer toute tentative de présentation simple : en effet, les mêmes gènes peuvent conduire à des pathologies bien différentes selon que les altérations constitutionnelles sont mono- ou bi-alléliques. Il en est ainsi, notamment, des altérations de *BRCA2*, à l'origine d'un risque élevé de cancer du sein en cas d'altération mono-allélique, mais responsable d'une maladie de Fanconi en cas d'altération bi-allélique²⁸. De plus, l'appartenance à la catégorie des gènes suppresseur de tumeur ou à celle des gènes de stabilité du génome n'est plus exclusive : à titre d'exemple, le gène *APC* est impliqué dans la voie Wnt de transduction du signal mitogène, mais aussi dans la ségrégation des chromosomes lors de la mitose²⁹.

Les variantes de forte pénétrance de ces gènes expliquent environ 20 % du risque de cancer familial ³⁰.

I 3 4. Variants de moyenne et faible pénétrance

La plupart des chercheurs pensent que l'architecture génétique de la maladie sporadique est polygénique, dans laquelle la sensibilité résulte de l'effet agrégé de nombreuses variantes à faible pénétration. Ainsi, les GWAS (genome-wide association study) sont utilisés pour rechercher ces variantes en testant les polymorphismes nucléotidiques ou SNP (single-nucleotide polymorphism) dans des anémies de Fanconi génotypés dans un grand nombre de cas et de témoins sains.

I 3 5. Les trois premières GWAS de cancer du sein

Les trois premières GWAS de cancer du sein ont été publiées simultanément en 2007. Dans une de ces études, Stacey et al. ³¹ ont utilisé 4554 cas et 17 577 contrôles pour identifier deux SNP communs, rs13387042 et rs3803662. Dans le deuxième de ces GWAS, Easton et al. ³² ont identifié cinq loci de susceptibilité en utilisant 4398 cas de cancer du sein et 4316 contrôles pour l'étape de découverte, et plus de 20 000 cas et 20 000 témoins pour l'étape de confirmation. Ces loci contenaient des SNP dans ou près de *FGFR2*, *TNRC9*, *MAP3K1*, *LSP* et *H19*. Enfin, Hunter et al. ³³ ont mené une étude d'association à deux étapes à l'échelle du génome en utilisant 2921 cas de cancer du sein postménopausique et 3214 témoins, et a identifié quatre SNP *FGFR2* introniques, reproduisant ainsi de manière indépendante la découverte d'Easton et al.

À ce jour, plus de 60 GWAS de cancer du sein ont été réalisées. À mesure que ce nombre grandit, l'avantage de la méta-analyse - la combinaison de la preuve à travers plusieurs études - devient évident.

I 3 6. Méta-analyse

La première méta-analyse à grande échelle, menée par Michailidou et al. ³⁴ en 2013, employaient 55 342 cas et 54 455 témoins de neuf GWAS et identifiaient 41 nouveaux loci de sensibilité. Deux ans plus tard, une méta-analyse encore plus grande, comprenant plus de 120 000 individus de 52 études, a permis d'identifier 15 autres loci de susceptibilité ³⁵, ce qui porte le nombre actuel de loci de sensibilité identifiés par GWAS à 84.

Le cancer colorectal (CC) peut également être considéré comme une maladie complexe en tenant compte de l'hypothèse commune d'une maladie avec un modèle polygénique héréditaire où les composants génétiques des maladies complexes communes correspondent principalement à des variantes d'effet faible ou modéré. A ce jour, 30 variantes de susceptibilité à faible pénétration ont été identifiées pour le CC ³⁶.

Ainsi, grâce à ces GWAS plusieurs variantes de susceptibilité à faible pénétration ont été identifiées. Bien que les variantes causales et les mécanismes prédisposants pour les associations de risque soient actuellement inconnus, ils ont ouvert des champs inexplorés dans la tumorigénèse mammaire. Par exemple, le facteur de croissance des fibroblastes 2 (*FGFR2*), un oncogène qui est surexprimé dans un sous-ensemble de cancer du sein ³⁷, contient un SNP intronique qui peut affecter un site de liaison du facteur de transcription pour le gene *FGFR2*. Un SNP dans la région 8q24, dépourvue de gènes ou d'activité transcriptionnelle, permet de réguler la transcription de l'oncogène *MYC* situé à plus de 300 kb ³⁸.

Ces études suggèrent que les gènes de faible pénétrance peuvent expliquer la prédisposition génétique sous-jacente au risque de cancer ³⁹ bien que leur valeur fonctionnelle reste à ce jour peu connue.

I 4. Analyses génomiques à haut débit

I 4 1. Séquençage nouvelle génération (NGS)

Avec l'avènement et l'amélioration de la nouvelle technologie de séquençage, le NGS est de plus en plus appliqué en recherche sur la génomique du cancer. Plus récemment, le NGS a été adoptée en oncologie clinique pour favoriser un traitement personnalisé du cancer. Le NGS est utilisé pour la détection de translocations, les inversions, les insertions et les suppressions, la détection de variations de numéros de copies, la détection de mutation dans des cancers familiaux. Le NGS présente de nombreux avantages, tels que la capacité de séquencer complètement tous les types de mutations pour un grand nombre de gènes.

Il y a eu plusieurs articles récents portant sur l'utilisation du NGS dans le dépistage génétique des cancers familiaux ⁴⁰⁻⁴³.

En outre, les gènes de faible pénétrance qui ne sont pas fréquemment analysés mais impliqués dans des syndromes de cancer familial pourraient être analysés si des risques de maladie peuvent être attribués à des mutations de ces gènes.

I 4 2. Dans le cancer de l'ovaire

Environ 15% des femmes atteintes d'un cancer de l'ovaire invasif portent une mutation détectable dans les gènes *BRCA1* ou *BRCA2* gènes ^{44,45} et le test des gènes *BRCA1* et *BRCA2* est recommandé pour toutes les femmes atteintes d'un cancer de l'ovaire

⁴⁶. Plus récemment, un nombre croissant de cas de cancer de l'ovaire sont attribués à des mutations dans les gènes du syndrome de Lynch, notamment *MSH6* ^{47,48}. Une étude récente a révélé que plus d'un cancer de l'ovaire sur 5 était associé à une mutation germinale du gène et plus de 30% des femmes étudiées n'avaient pas d'antécédent familial de cancer du sein ou de l'ovaire ⁴⁴. Des mutations dans un nombre croissant d'autres gènes ont été trouvées chez les patientes atteintes d'un cancer de l'ovaire, y compris: *ATM*, *BARD1*, *BRIP1*, *CHEK2*, *MLH1*, *MRE11A*, *MSH6*, *NBN1*, *PALB2*, *PMS2*, *PTEN*, *RAD50*, *RAD51C* et *TP53* ^{44,47,49}. Cette augmentation du nombre de gènes impliqués dans le cancer de l'ovaire suggère l'utilisation de tests basées sur un panel de gènes.

I 4 3. Dans le cancer gastro-intestinal

Jusqu'à 6% des cancers colorectaux peuvent être dus à un syndrome dit à haut risque ⁵⁰. Le syndrome de Lynch nécessite par exemple l'analyse de 5 gènes (*EPCAM*, *MLH1*, *MSH2*, *MSH6* et *PMS2*). Les analyses en immunohistochimie (IHC) des tumeurs peuvent identifier les gènes à cibler, mais ne sont pas suffisantes. Le NGS peut maintenant être utilisé comme évaluation de première intention pour le diagnostic génétique du syndrome Lynch. Beaucoup de familles avec une polypose colique et les caractéristiques du syndrome de Lynch ne se sont pas révélées avoir des mutations dans les gènes *APC*, *MUTYH* ou les 5 gènes associés au syndrome de Lynch. Des études ont identifié des mutations dans les gènes *POLE*, *POLD1* et d'autres polymérases dans de telles familles conduisant à un diagnostic de polypose associée à la polymérisation (PPAP) ^{51,52}. Avec cette nouvelle preuve de gènes supplémentaires avec un chevauchement phénotypique de syndromes hémorragiques de cancer colorectal ou de polypose communs, les tests NGS peuvent être une approche efficace et rentable de l'évaluation génétique. Les tests de panel avec une attention particulière aux gènes inclus devraient être pris en considération dans le cas où un test initial spécifique de

gène pour un individu répondant aux critères de diagnostic de la polypose adénomateuse familiale (FAP), la FAP atténué, polypose à association MUTYH ou syndrome Lynch est négatif.

I 4 4. Dans les cancers pancréatiques

Environ 5% à 10% de l'adénocarcinome pancréatique sont familiaux ⁵³. Plusieurs gènes ont été associés à un risque accru de cancer du pancréas incluant *BRCA1*, *BRCA2*, *CDKN2A*, *PALB2*, *PRSS1*, *STK11* et les gènes du syndrome de Lynch ^{54,55}. Des études NGS utilisant des panneaux de cancer multigéniques pour dépister les patients atteints de cancer du pancréas ont identifié des mutations dans les gènes *ATM*, *BARD1*, *BRCA1*, *BRCA2*, *CHEK2*, *FANCM*, *NBN*, *MLH1*, *MSH2*, *MSH6*, et *TP53* ^{56,57}. *L'American College of Gastroenterology* recommande actuellement de tester *BRCA1*, *BRCA2*, *PALB2*, *CDKN2A* et *ATM* pour tous les cas de cancer du pancréas familiaux, en ajoutant les gènes du syndrome de Peutz-Jeghers, du syndrome de Lynch et de la pancréatite héréditaire si l'histoire clinique suggère ces conditions ⁵⁸.

I 5. Mutations sporadiques et formes familiales

Le carcinome rénal (RCC) familial est dû à des mutations dans plusieurs gènes, y compris le gène *VHL*, qui est également muté dans les formes sporadiques. Cependant, un pourcentage important de RCC familial reste inexplicé. Récemment, il a été découvert que le gène *BAP1* est muté dans le RCC sporadique. *BAP1*, qui code pour une ubiquitine ligase nucléaire, est un gène suppresseur de tumeur à double impact. Les mutations *BAP1* somatiques sont associées aux carcinomes rénaux à cellules claires de haut grade (ccRCC). Pour déterminer si *BAP1* prédispose au RCC familial, le gène *BAP1* a été séquencé

chez 83 sujets non apparentés avec une forme familiale de RCC familial sans mutation de *VHL*. Un nouveau variant a été identifié : c.41T> A; p.L14H. Le variant p.L14H perturbe un résidu fortement conservé dans le domaine catalytique, un domaine fréquemment ciblé par des mutations non-sens. La famille avec le variant *BAP1* a été caractérisée par un RCC à cellules claires précoce, de haut grade de Fuhrman, sans maladie de von Hippel-Lindau. Ces résultats suggèrent que *BAP1* est un gène prédisposant au RCC familial ⁵⁹.

Environ 5% de tous les carcinomes rénaux (RCC) sont familiaux ⁶⁰. Plusieurs gènes, *VHL*, *MET*, *FLCN*, *FH* ont été identifiés comme responsables de la carcinogénèse rénale ⁶¹⁻⁶³. Cependant, la base génétique d'un pourcentage significatif de RCC familial reste inconnue. Il existe un précédent pour les gènes *VHL* mutés dans la lignée qui sont également mutés dans le cadre sporadique, et ainsi, les gènes mutés de manière somatique peuvent expliquer le RCC familial s'il est muté dans la lignée germinale.

Ainsi, les panels peuvent également inclure d'autres gènes de pénétrance modérée à faible, par exemple *BARD1* et *NBN*, dont il existe peu de données concernant le risque de cancer du sein et même moins d'informations sur les options de gestion. ^{6, 13, 16} Les progrès extraordinaires ont été réalisés dans la génétique du cancer au cours des 20 dernières années. La diffusion rapide du NGS et l'identification des gènes HBOC (Hereditary Breast and Ovarian Cancer) de faible et moyenne pénétrance ont profondément modifié la connaissance de la génétique du cancer du sein.

Reste un écueil, la plupart des approches d'identification sont familiales ou sur de large population. Surtout, la plupart des anomalies identifiées n'ont pas de valeur fonctionnelle connue.

II. Nouvelle approche en génétique constitutionnelle

II 1. Objectif du projet

Dans mon projet de thèse, je propose une approche individuelle innovante combinant plusieurs innovations technologiques pour identifier et étudier la valeur fonctionnelle de gènes de pénétrance faible à modérée.

Nous sommes partis d'un cas d'une patiente de 36 ans ayant un cancer du sein, une maladie de Vaquez, et une polyarthrite rhumatoïde sans contexte familial ni syndrome génétique familial évident. (*Figure 2*)

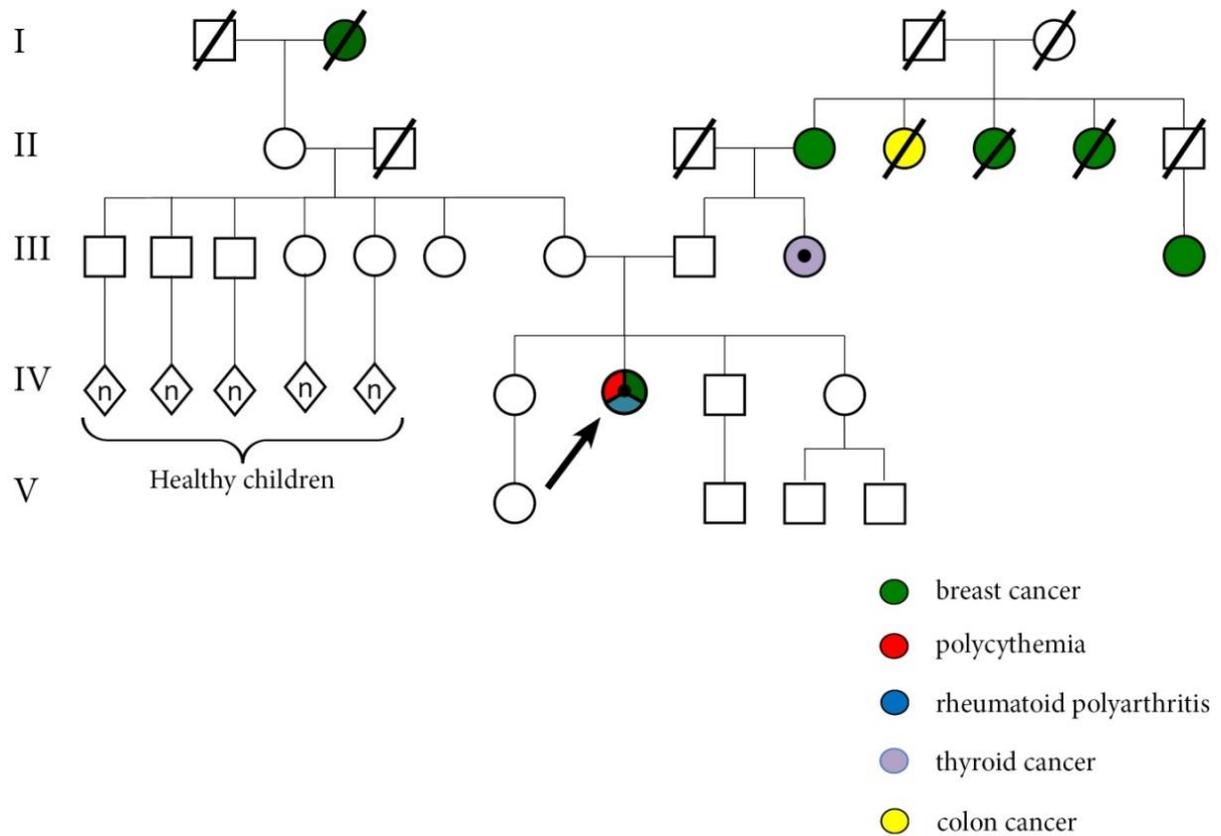


Figure 2 : Généalogie de la patiente

Dans notre étude, nous proposons une approche individuelle innovante en étudiant des échantillons tissulaires d'une femme jeune avec une association inhabituelle de cancers et d'autres maladies graves, afin d'identifier des mutations constitutionnelles de faible et moyenne pénétrance et démontrer leur valeur fonctionnelle.

II 3. Résultats

II 3 1. Cas index : une femme jeune avec un cancer du sein et d'autres maladies graves

Une femme de 36 ans a été diagnostiquée avec un cancer du sein localisé surexprimant *HER2* et une polyglobulie de Vaquez *JAK2* V617F mutée, concomitants d'une polyarthrite rhumatoïde ⁶⁴.

Une analyse minutieuse de l'histoire familiale de la patiente a retrouvé un cancer papillaire de la thyroïde à la troisième génération chez une tante paternelle (III, 9), quatre cancers du sein patrilinéaires et un cancer du côlon dans la deuxième génération (*Figure 2*). Soulignons que notre patiente ne portait aucune mutation dans les gènes de prédisposition au cancer du sein, de forte pénétrance (*BRCA1*, *BRCA2*, *TP53*, *PTEN*) ou de moyenne pénétrance (*CHEK2*, *PALB2* et *ATM*).

II 3 2. Identification d'une mutation constitutionnelle du gène *MET*

Afin de rechercher des mutations dans d'autres gènes candidats de faible et moyenne pénétrance chez notre patiente (cas index), nous avons choisi une approche originale et avons effectué une analyse **génomique à haut débit portant sur une étude de variation du nombre de copie** et de mutations ponctuelles au niveau des cellules de tumorales du cancer du sein microdisséquées et des cellules CD34+ triées à partir d'une aspiration de moelle osseuse.

La fixation en alcool formol acide acétique (AFA) de la plupart des échantillons tumoraux de patients dont on dispose représente l'un des défis majeurs de l'étude pangénomique comme c'est le cas pour le cancer du sein de notre cas index. Ces échantillons

d'ADN sont fréquemment endommagés par l'exposition au formaldéhyde et à un environnement très acide du fixateur. Notre intérêt était donc d'utiliser une technique d'analyse moléculaire très sensible pour détecter des anomalies moléculaires sur des petites quantités d'ADN dégradé et fragmenté. Des études sur la technologie OncoScan ont comparé la concordance des résultats de l'analyse des variations du nombre de copies d'un gène sur tissu congelé ou sur tissu formolé. Ces études rapportent un taux de concordance compris entre 92 et 99,99%⁶⁵⁻⁶⁸.

En utilisant la technologie OncoScan dédiée aux échantillons fixés en formol (*Figure 3*) avec un seuil de probabilité >9 pour filtrer les données, nous avons identifié 17 mutations potentielles (*Tableau 2*). Parmi les 541 mutations testées, seule la mutation pR988C, c2962C_T de *MET* était commune aux deux populations de cellules tumorales, avec un score de probabilité de 48,5 et 34,1 respectivement pour le cancer du sein et les cellules CD34 (*en gras dans la Tableau 2 et Figure 4*).

Quand nous nous sommes focalisés sur les variations de nombre de copie, les événements communs étaient des pertes d'hétérozygotie (LOH) essentiellement dans des régions non-codantes (*Tableau 3*). Une LOH située sur le chromosome 3p25.3 et comprenant les gènes Von Hippel Lindau (*VHL*) et *FANCD2* était commune au cancer du sein et aux cellules CD34. Aucune mutation « hotspot » commune de *VHL* n'a été identifiée dans ces prélèvements.

En utilisant une technique de PCR-HRM et un séquençage Sanger, nous avons confirmé la présence de la mutation hétérozygote pR988C, c2962C_T de *MET* dans les cellules de cancer du sein microdisséquées et les cellules CD34 de la moelle osseuse. Nous avons également retrouvé la même mutation au niveau de l'ADN constitutionnel extrait à partir de bulbes de cheveux (*Figure 5A*). Nous avons eu l'opportunité de tester le cancer de la

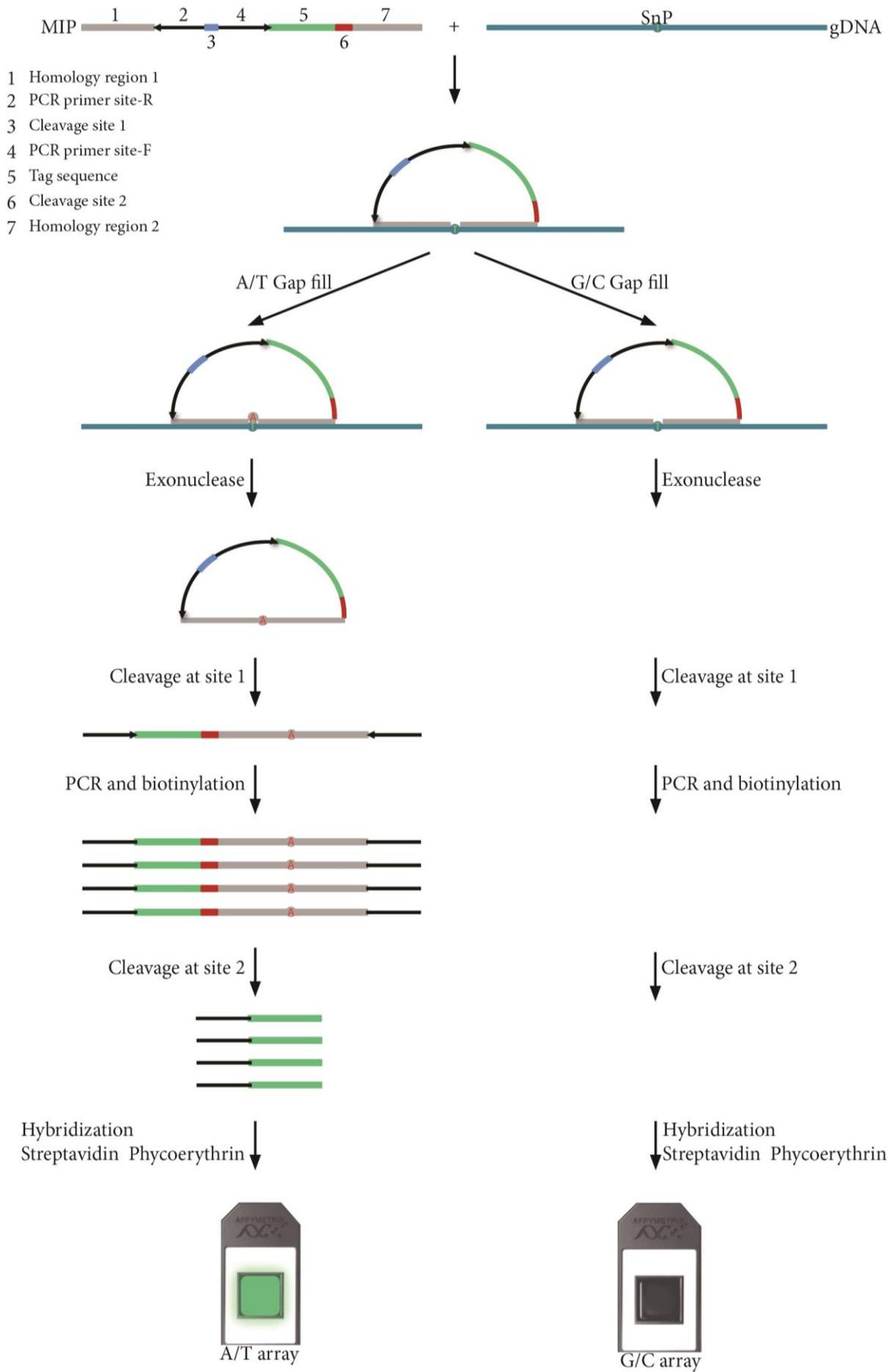


Figure 3: Méthodologie de la technologie Oncoscan

	Probability score	Gene	Mutation	Chromosome
Breast cancer	48.5	<i>MET</i>	pR988C, c2962C_T	7
	9.8	<i>HRAS</i>	pG12S, c34G_A	11
CD34-positive bone marrow cells	34.1	<i>MET</i>	pR988C, c2962C_T	7
	18.1	<i>PTEN</i>	pI101T, c302T_C	10
	13	<i>NF2</i>	pQ362X, c1084C_T	22
	12.6	<i>FGFR3</i>	pA369A, c1107G_T	4
	12.3	<i>HRAS</i>	pG12D, c35G_A	11
	12.1	<i>SMARCB1</i>	pR201X, c601C_T	22
	11.9	<i>RB1</i>	pR455X, c1363C_T	13
	11.8	<i>RB1</i>	pR787X, c2359C_T	13
	11.3	<i>VHL</i>	pE160K, c478G_A	3
	10.4	<i>NF2</i>	pR196X, c586C_T	22
	10.2	<i>RB1</i>	pQ702X, c2104C_T	13
	10	<i>CTNNB1</i>	pS45P, c133T_C	3
	9.5	<i>APC</i>	pR1450X, c4348C_T	5
	9.5	<i>NF2</i>	pQ456X, c1366C_T	22
	9.4	<i>MEN1</i>	p_c654_plus_3A_G	11
9	<i>ABL1</i>	pF359V, c1075T_G	9	

Tableau 2 : Liste des mutations retrouvées par oncoscan sur le cancer du sein et les cellules myéloïdes CD34+

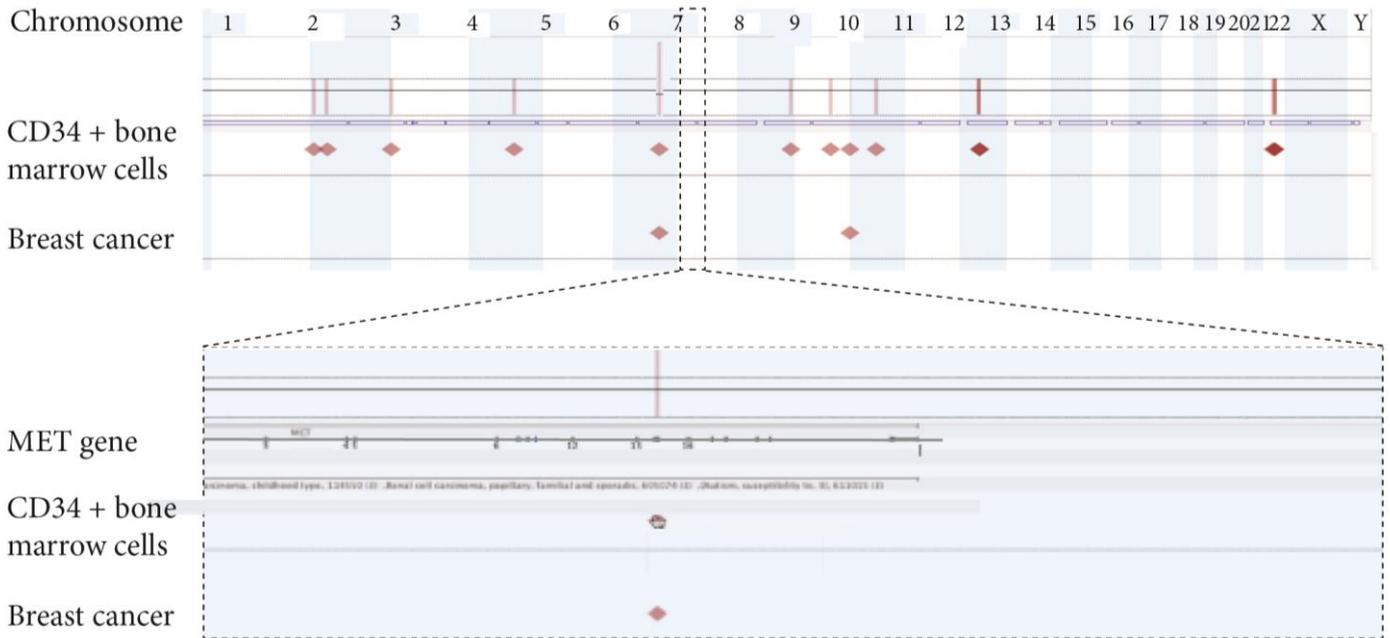


Figure 4 : Schéma des mutations retrouvées par oncoscan sur le cancer du sein et les cellules myéloïdes CD34+

Chromosome	Localisation	Cytoband	Event	Gene
1	35,380,596-37,125,966	p34.3	LOH	none
1	46,930,969-47,650,912	p33	LOH	none
1	144,009,053-147,099,711	q21.1 - q21.2	LOH	none
2	80,799,534-81,458,366	p12	LOH	none
2	107,783,376-108,411,960	q12.3	LOH	none
2	138,077,793-138,719,412	q22.1	LOH	none
3	9,909,419-10,606,834	p25.3	LOH	VHL, FANCD2
3	44,122,466-45,130,076	p21.32 - p21.31	LOH	none
3	104,935,498-105,659,604	q13.11	LOH	none
5	30,093,131-30,957,792	p13.3	LOH	none
5	50,677,388-51,431,355	q11.1 - q11.2	LOH	none
7	19,136,273-20,106,990	p21.1	LOH	none
7	142,241,083-143,017,287	q34	LOH	none
8	5,813,252-6,355,849	p23.2 - p23.1	LOH	none
8	46,950,145-49,032,046	q11.1 - q11.21	LOH	none
8	104,432,606-105,889,089	q22.3	LOH	none
9	44,870,087-45,412,818	p11.2	LOH	none
10	32,499,853-33,442,298	p11.22	LOH	none
11	9,605,225-10,371,796	p15.4	LOH	none
11	30,786,463-32,168,216	p14.1 - p13	LOH	none
12	96,820,650-97,722,341	q23.1	LOH	none
15	20,161,372-20,977,119	q11.1 - q11.2	LOH	none
17	0-1,444,026	p13.3	LOH	none
X	49,847,131-58,545,809	p11.22 - p11.1	LOH	none

Tableau 3 : Liste des LOH retrouvées par oncoscan à la fois sur le cancer du sein et les cellules myéloïdes CD34+

thyroïde de la tante paternelle de notre patiente (III; 9 sur la Figure 2), et avons retrouvé la même mutation *MET* pR988C, c2962C_T, nous permettant de confirmer qu'il s'agit d'une mutation héréditaire du côté paternel. En outre, nous avons montré qu'elle était associée à une surexpression de la protéine MET par les cellules tumorales à la fois dans le cancer du sein et le cancer de la thyroïde (Figure 5B).

Pour le locus de *VHL*, utilisant des marqueurs microsatellites (Figure 6), nous avons retrouvé une LOH au niveau des cellules de cancer du sein, confirmant les résultats d'ONCOSCAN. Les cellules CD34-positives et l'ADN constitutionnel de notre patiente avaient un profil de déséquilibre allélique, suggérant un mosaïcisme de cette LOH de *VHL*. Au contraire, le profil microsatellite du cancer de thyroïde de la tante (III; 9) était normal, suggérant que les anomalies *VHL* sont *de novo* chez notre patiente.

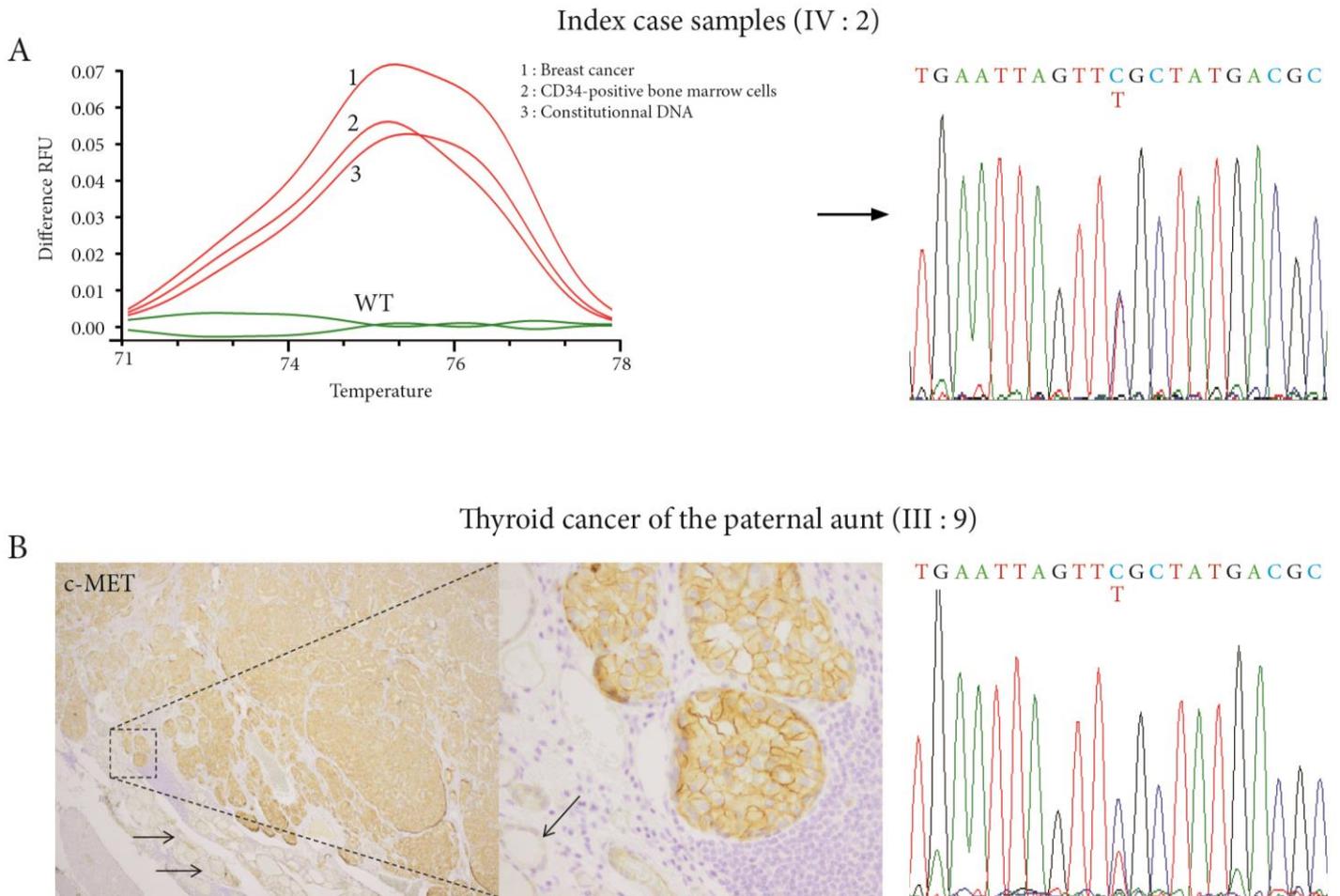


Figure 5 : Mutation *MET* pR988C, c2962C_T constitutionnel et héréditaire

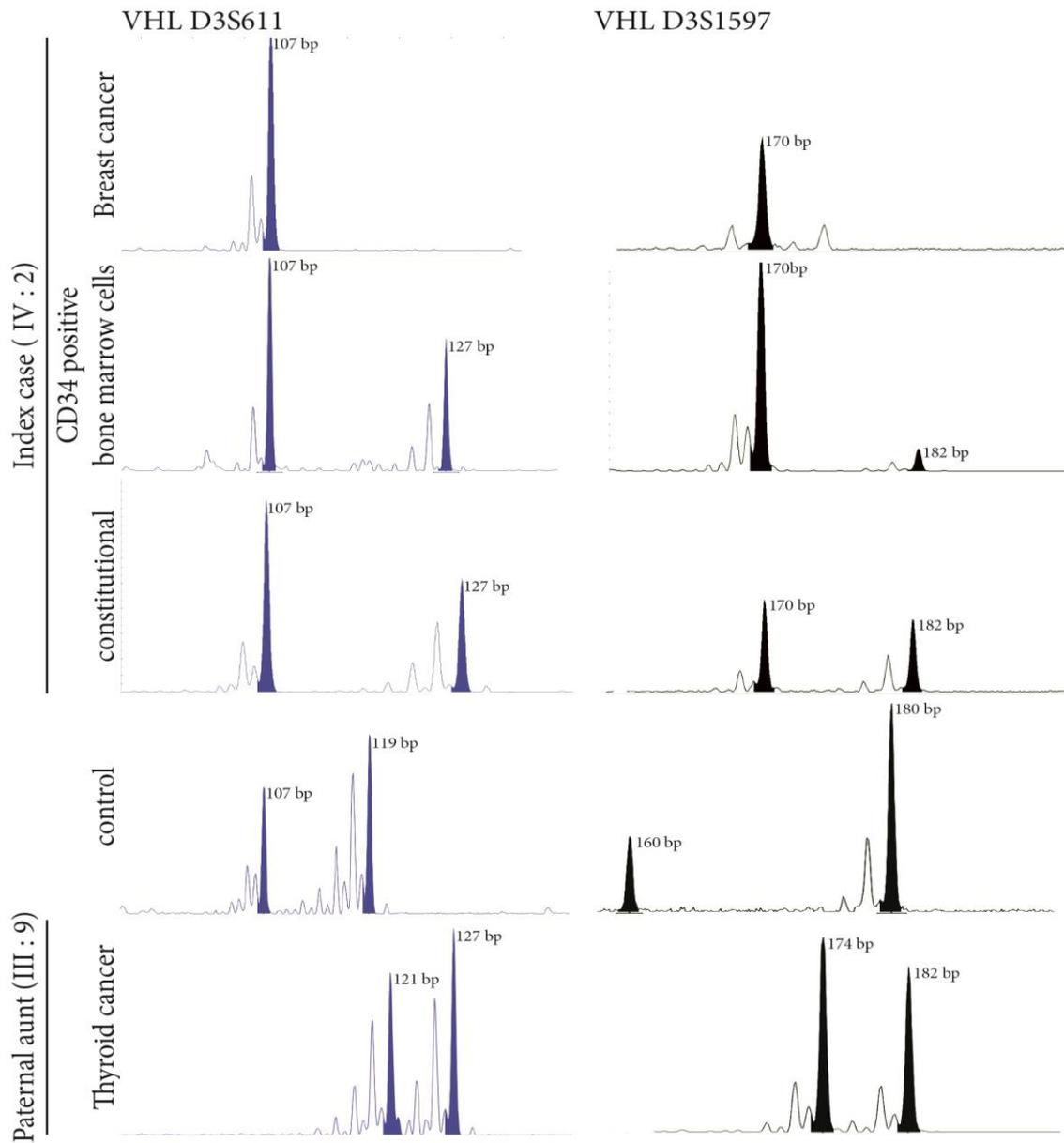


Figure 6 : Etude de microsatellites autour du gène *VHL* chez le cas index et sa tante paternelle

II 3 3. Identification de cette mutation *MET* pR988C, c2962C_T

II 3 3 1. Dans une séries de patients avec cancer du sein sporadique

Dans une série de 105 cancers du sein sporadiques de l'Hôpital Saint Louis (Paris, France), nous avons utilisé la PCR-HRM pour détecter la mutation *MET* pR988C, c2962C_T. Elle a été identifiée dans un patient avec un cancer du sein intracanalair triple négatif (Figure 7A).

II 3 3 2. Dans une séries de patients avec polyglobulie de Vaquez avec mutation *JAK2* V617F

Nous avons aussi testé l'ADN de sang total de 96 patients avec une polyglobulie de Vaquez avec la mutation *JAK2* V617F. En utilisant l'ADN-HRM pcr, la mutation *MET* pR988C, c2962C_T a été identifié dans un patient de 87 ans (Figure 7B). Étonnamment, elle avait aussi une histoire personnelle de polyarthrite rhumatoïde chronique évolutive et une cryoglobulinémie.

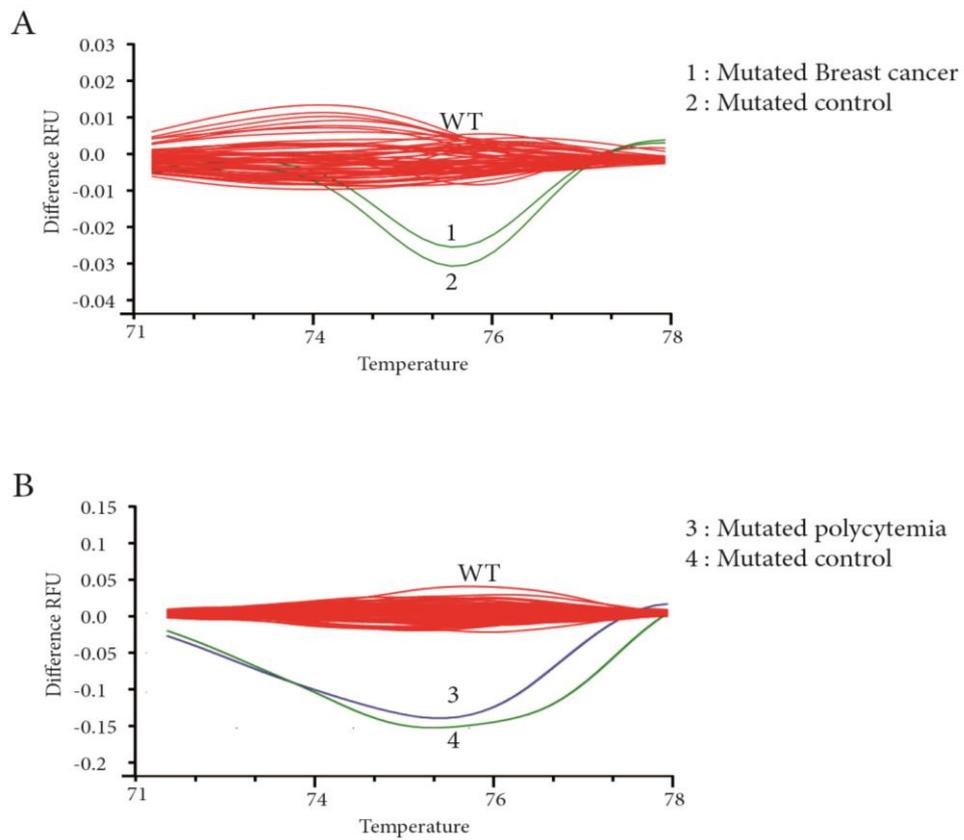


Figure 7 : Screening de la mutation *MET* pR988C_T, c2962C_T dans des cancers du sein sporadique (A) et dans des polyglobulies de Vaquez (B)

II 3 4. Souris transgéniques *MET* pR988C, c2962C_T

Puisque la mutation *MET* pR988C, c2962C_T est localisé sur un codon identique chez l'homme et la souris, nous avons décidé de développer un model transgénique de souris pour montrer la valeur fonctionnelle de cette mutation localisée :

Séquence humaine

```
3'TCTTTAACAAGCTCTCTCTTTCTCTCTGTTTTAAGATCTGGGCAGTGAATTAGTTCGCTATGACGC  
AAGAGTACACACTCCTCATTGGATAGGCTTGTAAGTGCCCGAAGTGTAAGTCCAACACTACAGAGAT  
GGTTTCAAATGAGT5'
```

Séquence murine

```
3'TCTTTAACAAGCTCTTTCTTTCTCTCTGTTTTAAGATCTGGGCAGTGAATTAGTTCGCTACGATGC  
AAGAGTA  
GCAGAAATTGTTCGAGAAAGAAAGAGAGACAAAATTCTAGACCCGTCACTTAATCAAACGATGCT  
AGCTTCTG5'
```

La mutation *MET* pR988C, c2962C_T est en surbrillance jaune et les différences entre les 2 séquences en rouge.

Des modifications précises et efficaces du génome présentent une grande valeur dans les tentatives de compréhension des rôles de gènes particuliers et d'autres éléments génétiques dans les processus biologiques ainsi que dans diverses pathologies. Au cours des dernières années, les nouvelles méthodes de modification du génome connues sous le nom d'édition du génome, qui utilisent les nucléases dites "programmables", sont entrées en vigueur. Une véritable révolution dans l'édition du génome a été provoquée par l'introduction du système CRISPR-Cas (clusterisé régulièrement intercalé short palindromic repeats-CRISPR), dans lequel une de ces nucléases, c'est-à-dire Cas9, joue un rôle majeur. Ce système est basé

sur les éléments du mécanisme bactérien et archaïque responsable de l'immunité acquise contre les infections aux phages et du transfert de matériel génétique étranger. Les microorganismes incorporent des fragments d'ADN étranger dans les loci CRISPR présents dans leurs génomes, ce qui permet une reconnaissance et une élimination rapides des infections futures. Il existe plusieurs types de systèmes CRISPR-Cas parmi les procaryotes, mais seuls les éléments de CRISPR type II sont employés dans l'ingénierie génomique. Le CRISPR-Cas de type II utilise de petites molécules d'sgRNA (CRRNA et TraCRRNA) pour diriger précisément la nuclease effectrice - Cas9 - vers un site spécifique du génome, c'est-à-dire la séquence complémentaire de l'ARNc. Cas9 peut être utilisé pour:

- introduire des changements stables dans les génomes, par exemple dans le processus de génération d'animaux knock-out et knock-in et de lignées cellulaires
- activer ou calmer l'expression d'un gène d'intérêt
- visualiser des sites spécifiques dans les génomes des cellules vivantes.

Les outils CRISPR-Cas ont été utilisés avec succès pour la génération de modèles animaux et cellulaires d'un certain nombre de maladies, par exemple des types spécifiques de cancer. À l'avenir, l'édition du génome par des nucléases programmables peut trouver une large application en médecine, par exemple dans les thérapies de certaines maladies d'origine génétique et dans la thérapie de patients infectés par le VIH.

En utilisant la technologie CRISPR-Cas9, les cellules d'une femelle C57Bl6 ont été micro-injectées avec deux types de sgRNA à deux concentrations d'enzyme Cas9 et sgRNA, dans le cytoplasme ou dans le noyau. Après l'incubation et la division cellulaire jusqu'au stade blastocyste (autour de 100 cellules), ce dernier a été testé pour l'intégration de la mutation *MET* pR988C, c2962C_T. En utilisant la méthode de restriction enzymatique et l'ADN-HRM pcr, nous avons pu vérifier le taux de réussite de l'intégration de la mutation avant la

réimplantation des embryons. Nous pouvons ainsi déterminer les conditions optimales, avec un ratio de 50ng/20ng de Cas9/sgRNA microinjecté directement dans le noyau donnant 19-25 % des embryons portant la mutation (Figure 8).

Concentration of Cas9/sgRNA	Type of SgRNA	Sub-cellular site of injection	Number of mutated blastocysts (%)
25ng/10ng	Sg5	Nucleus	0/20 (0%)
		Cytoplasm	1/16 (6%)
	Sg8	Nucleus	0/16 (0%)
		Cytoplasm	0/15 (0%)
50ng/20ng	Sg5	Nucleus	3/16 (19%)
		Cytoplasm	NA
	Sg8	Nucleus	6/24 (25%)
		Cytoplasm	4/17 (23.5%) + 1 indel/17

Figure 8 : Concentrations de Cas9/sgRNA injecté et leur efficacité d'intégration

Vingt et un jours après la re-implantation, une première génération de 17 souris a été obtenue, avec 4 animaux avec la mutation (23.5 %), 2 avec la mutation *MET* pR988C, c2962C_T et avec une indel (*souris 8, 10, 12 et 19, Figures 9 et 11*). Les deuxièmes et troisièmes générations, obtenues par l'accouplement des fondateurs 10 et 12, ont mené à l'élimination de l'indel et des potentiels « off-targets » pour 2 souris (*Figures 10 et 11B*). Un total de 23 souris portant la mutation a été ainsi obtenue, génotypées par restriction enzymatique, ADN-HRM pcr et le séquençage Sanger. La variation du nombre de copie a été réalisée pour toutes les souris avec **un ratio muté/normal allant** des 1 à 12 (*Figure 11*). Malheureusement, nous n'avons pas pu obtenir d'accouplement fructueux après la troisième génération, **par l'accouplement heterozygous ou homozygous**.

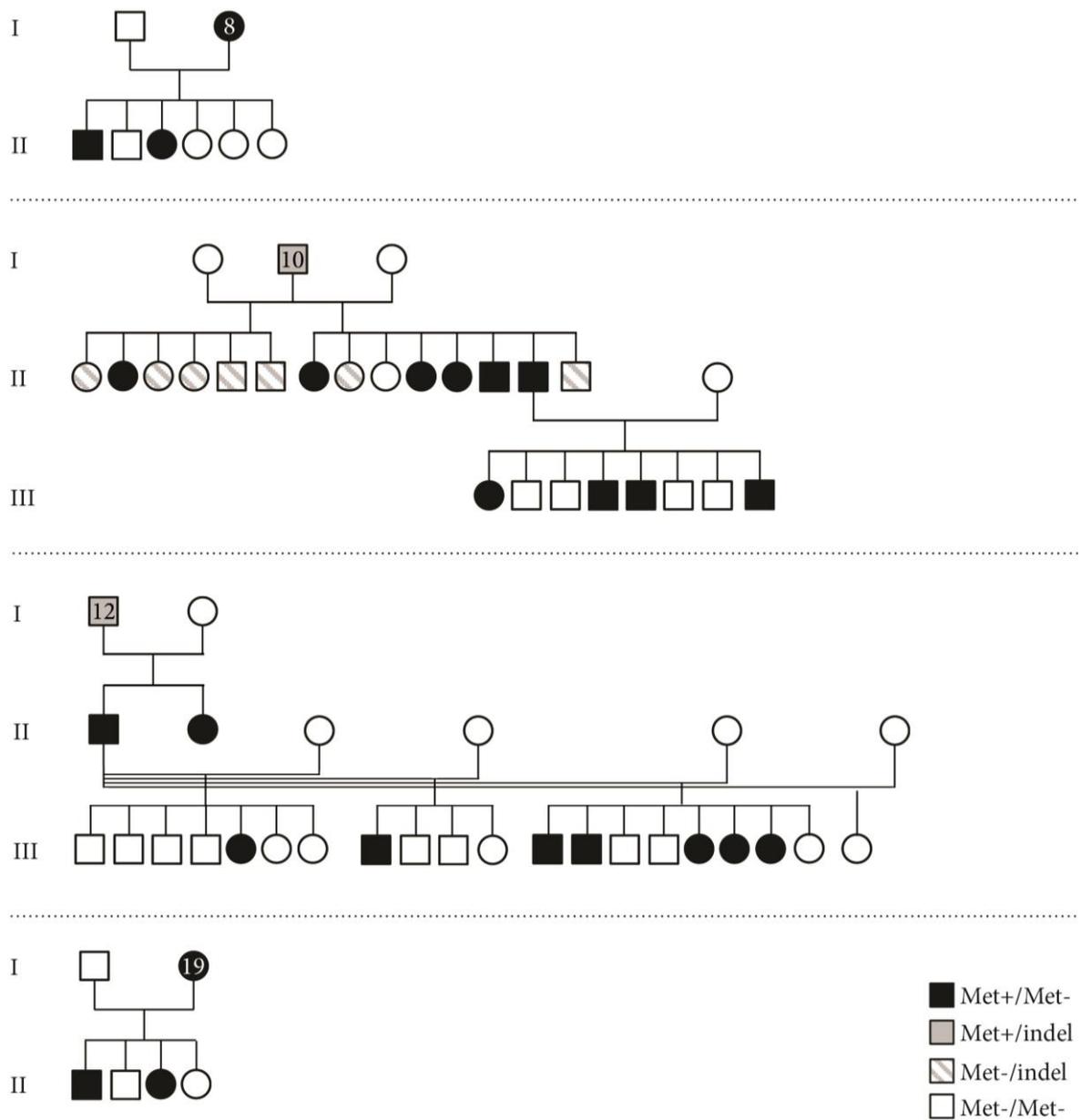
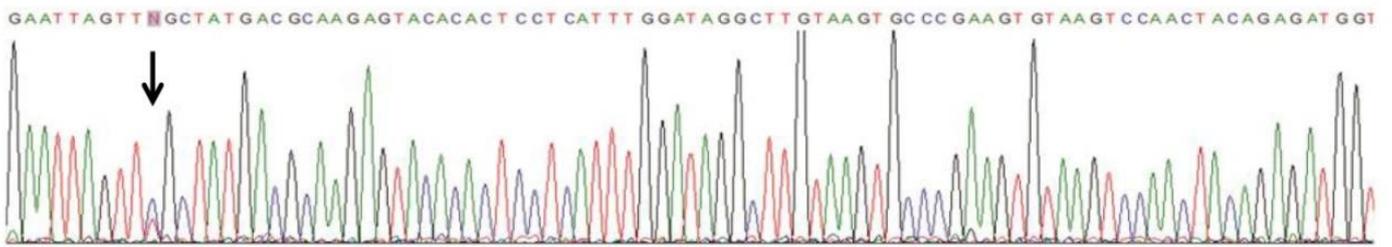


Figure 9 : Généalogie des souris transgéniques *MET* pR988C, c2962C_T obtenues

II:7, from the lineage of funder 10, *MET*^{+/MET-}



I:2, Funder 10, *MET*^{+/Indel}

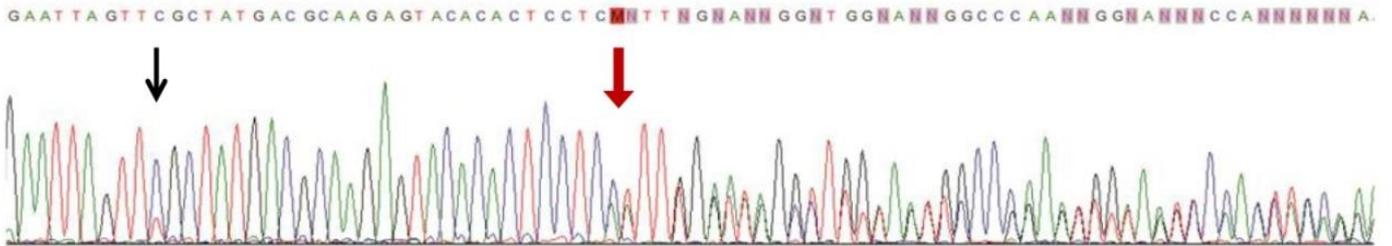
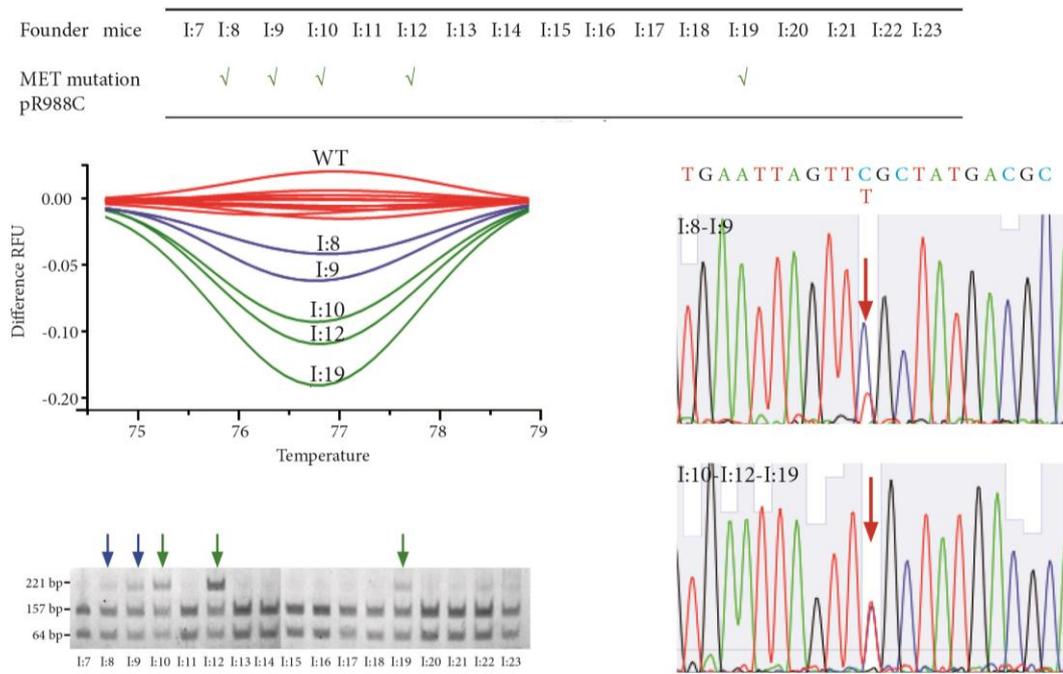


Figure 10 : Exemple de séquence Sanger de souris transgéniques *MET* pR988C, c2962C_T

A



B

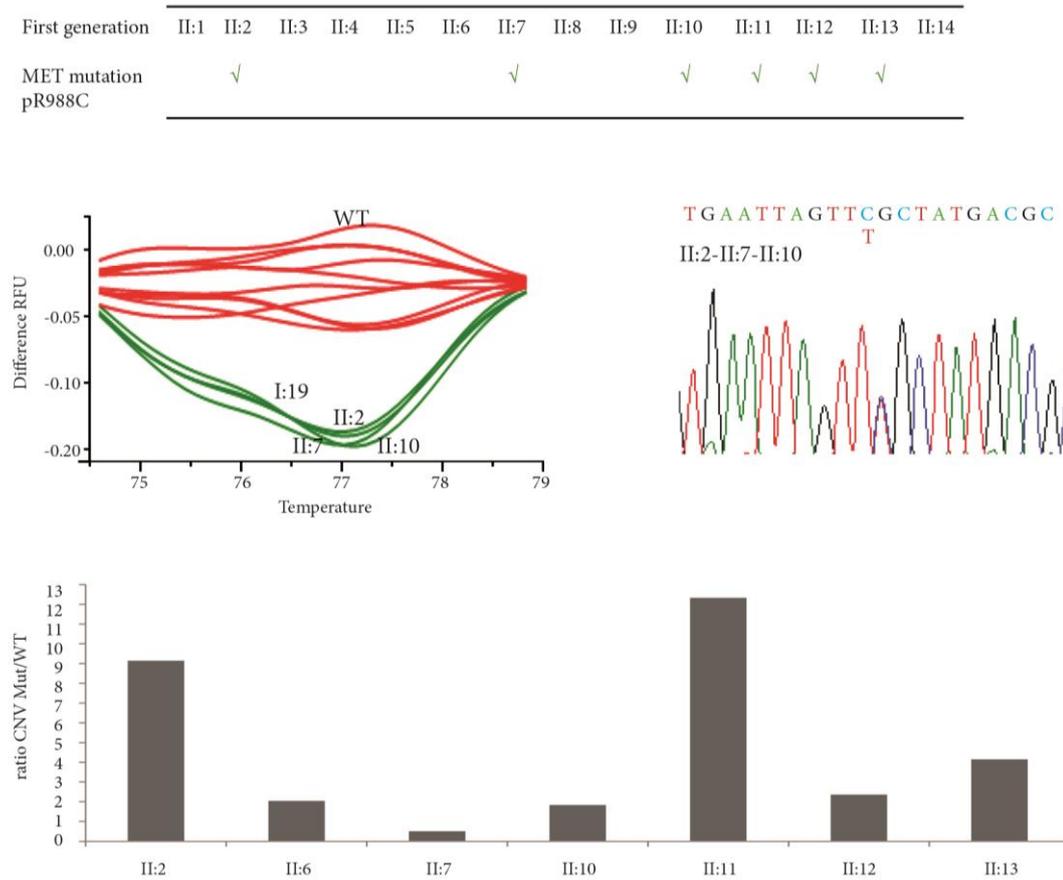


Figure 11 : Génotypage des souris transgéniques *MET* pR988C, c2962C_T

II 3 5. Souris transgéniques *MET* pR988C, c2962C_T et maladies auto-immunes et cancers (Figure 12)

Dès l'âge de 9 mois, toutes les souris mutées avaient des signes histologiques de syndrome de Sjögren, caractérisé par un infiltrat lymphocytaire **peri-canalair**e dense (score de Chisholm de III ou IV) dans les glandes salivaires et lacrymales. L'ELISA a montré un niveau des anticorps anti-SSA et anti-SSB murine sérique significativement plus élevé dans les souris mutées par rapport aux souris normales ($p < 0.05$). Dans 22 des 23 souris mutées, nous avons aussi constaté des signes histologiques de synovial hyperplasique et un infiltrat inflammatoire au niveau dans le distal, mais pas proximal des articulations. Les niveaux sérique des anticorps anti-peptides cycliques citrullinés (anti-CPP aussi appelé ACPA) et d'anticorps anti-dsDNA étaient significativement plus hauts dans les souris mutées que dans les souris normales ($p < 0.01$).

Nous avons trouvé les signes de thyroïdite dans 14 des 23 souris. Une souris mutée de 10 mois (II:10), a aussi développé un carcinome cutané sur l'oreille. Les souris normales n'avaient aucun signe de Sjögren, de polyarthrite rhumatoïde ou de cancer.

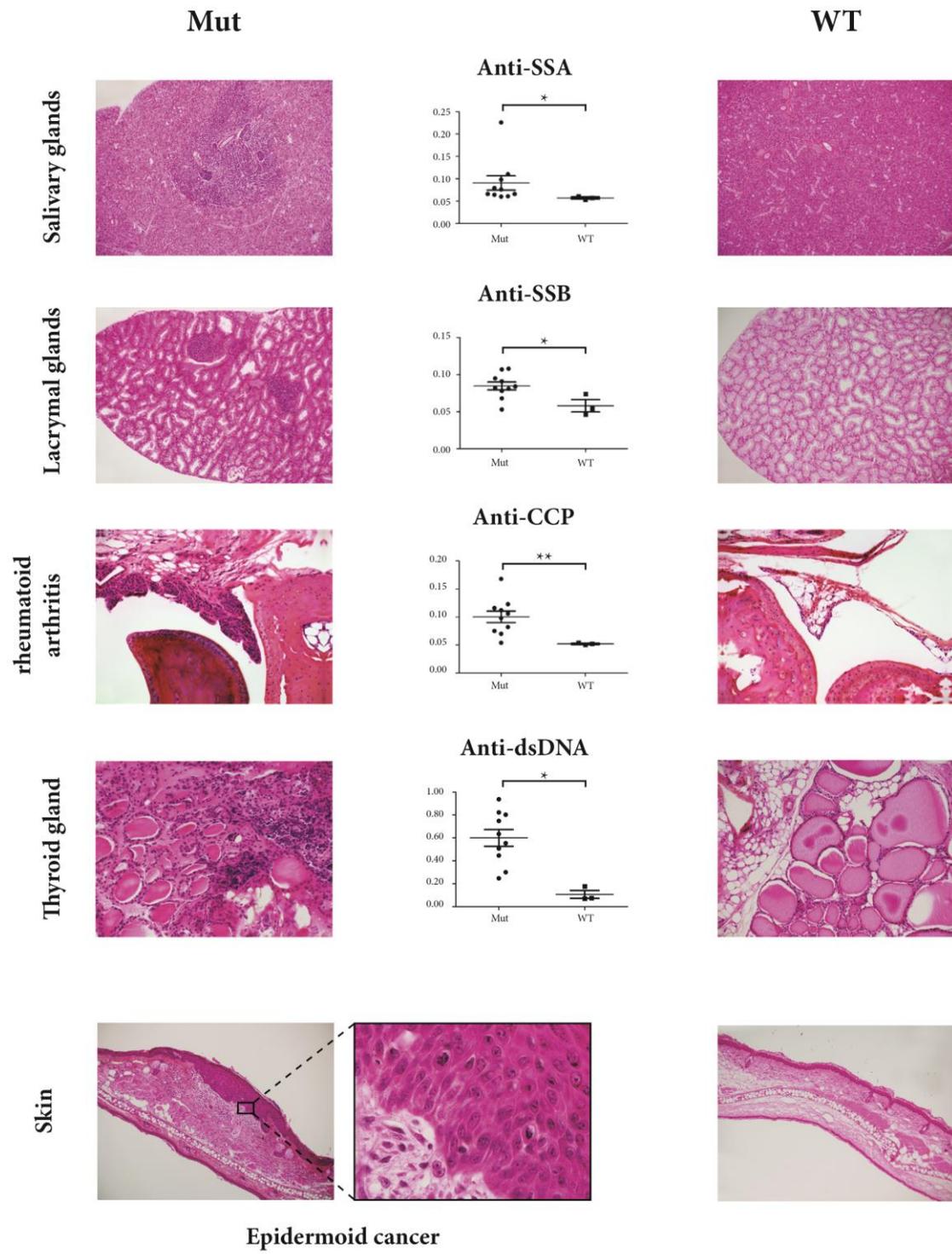


Figure 12 : Souris transgéniques MET pR988C, c2962C_T et maladies auto-immunes et cancers

II 3 6. Souris transgéniques *MET pR988C, c2962C_T* et syndromes myeloproliferatis

(Figure 13)

Les numérations obtenues à 1, 6 et 9 mois pour 12 souris mutées, ont montré une augmentation significative de nombres de plaquettaires comparées aux souris normales et ce, dès le premier mois (moyennne à 9 mois de $1078.103/mm^3 (\pm 137)$ contre $775.103/mm^3 (\pm 282)$, $p0.05$) (*Figure 13A*)

Les analyses de la moelle osseuse en cryométrie de flux ont été obtenues pour 8 souris mutées et 8 souris normales pour évaluer les pourcentages des différentes lignées (*Figure 13B et 20C*). Pour des populations LK, nous avons constaté que les souris mutées avaient une densité médullaire plus élevée que les souris normales, avec les valeurs cellulaires absolues de $24.1 \times 10^6 (\pm 2.6)$ contre $9.4 \times 10^6 (\pm 1.1)$, $p0.01$, (*Figure 13D*). Cette densité cellulaire plus haute a été trouvée dans toutes les trois lignée progénitrices myéloïdes (*Figure 13E*). Les plus grands nombres ont été trouvés pour les GMP (*Figure 13F*).

Les analyses des coupes tissulaires ont confirmé cette densité myéloïde plus élevée dans des souris mutées par rapport aux souris normales. De plus, des ovaires volumineux ont été trouvés dans deux souris mutées. L'examen microscopique a montré une prolifération hematopoietic extra-medullaire avec une différenciation osseuse dans ces deux ovaires volumineux (*Figure 13G*).

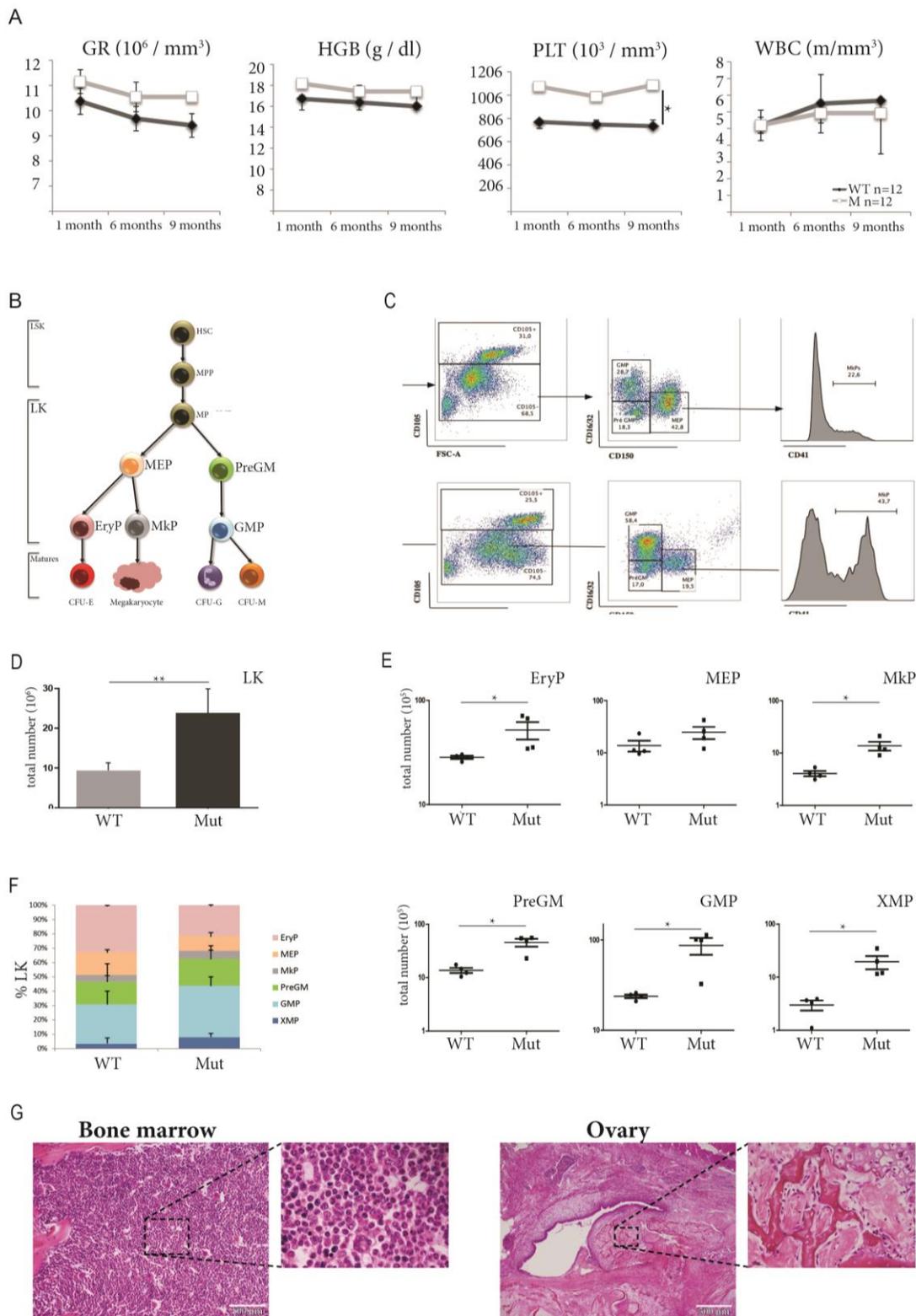


Figure 13 : Souris transgéniques *MET* pR988C, c2962C_T et syndrome myéloprolifératifs

II 3 7. Inhibition fonctionnelle de la voie *MET*

L'activation du récepteur transmembranaire MET/HGFR (Hepatocyte Growth Factor Receptor) induit sa dimérisation, la phosphorylation des tyrosines Y1234 et Y1235 au niveau du domaine tyrosine kinase (TK) et de Y1349 et Y1356 dans la région carboxy-terminale, menant aux signaux en aval (*Figures 14 et 15*)⁶⁹. Le domaine juxtamembranaire est fortement conservé (JM) et agit comme un régulateur négatif de transduction du signal, permettant l'ubiquitination de MET et sa dégradation⁷⁰. Les mutations

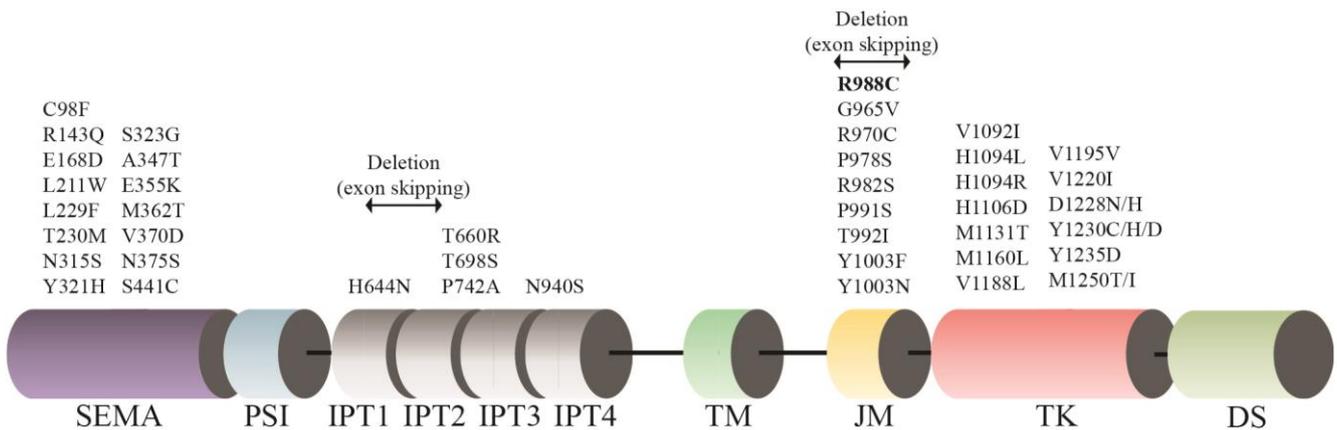


Figure 14 : Les mutations connues de *MET*

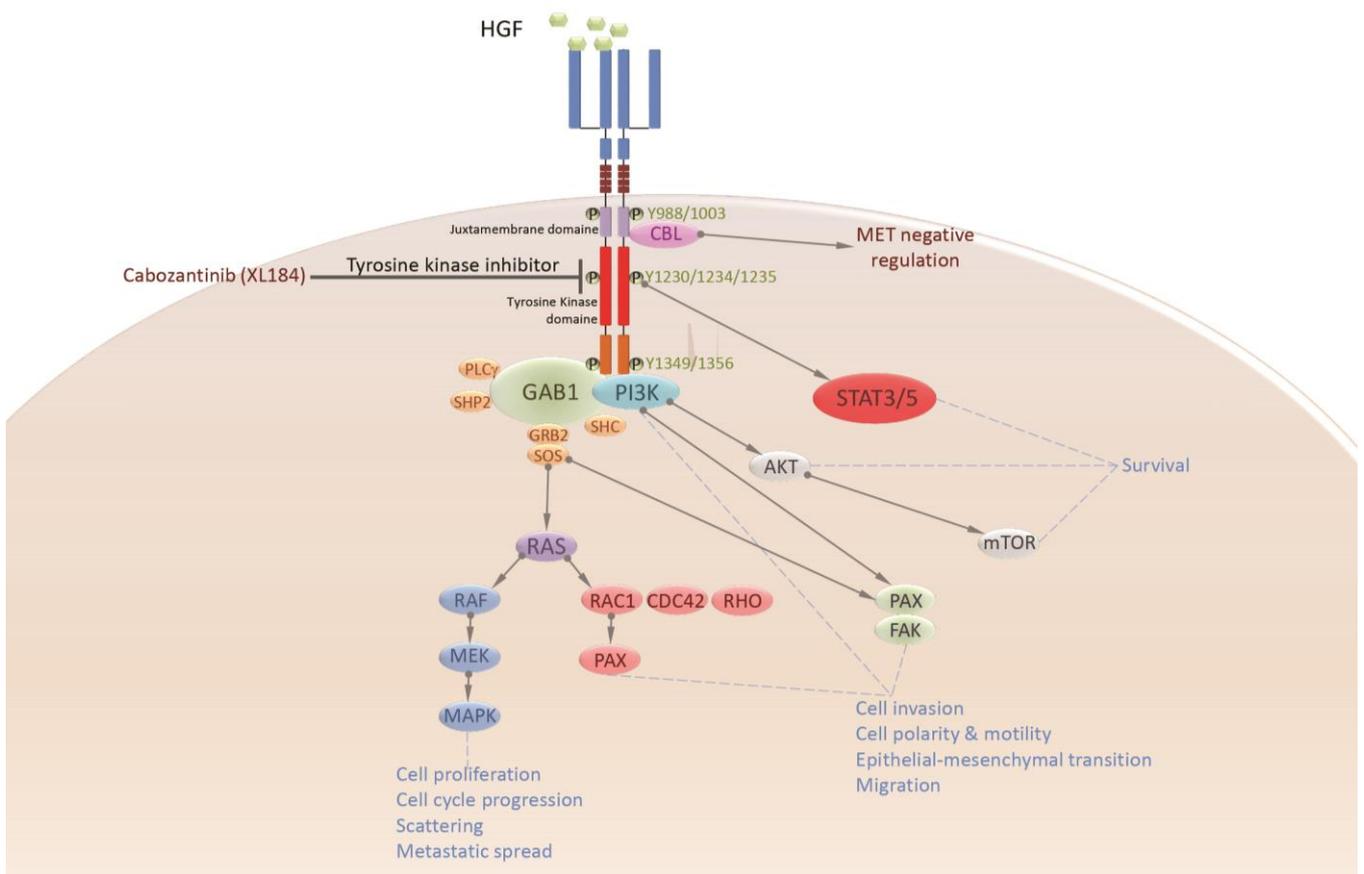


Figure 15 : Voie de signalisation de Met

dans le domaine JM pourraient supprimer cette régulation négatif⁷¹ et mener ainsi à l'activation constitutive de MET.

La mutation *MET* pR988C, c2962C_T est localisé dans le domaine JM (Figure 14). Pour définir sa valeur fonctionnelle, nous avons administré un inhibiteur de MET, le cabozantinib, aux souris mutées pendant 28 jours. Sur des échantillons d'aspiration de moelle osseuse séquentiels et des analyses de sang (jour 0, jour 28 et jour 42), cette inhibition de la voie MET a conduit à une diminution de GMP dans la moelle osseuse des souris mutées et

une baisse du nombre de globules rouges et du niveau d'hémoglobine et à une normalisation de concentration sérique des anticorps anti-CCP et des anticorps anti-SSB au jour 28 dans les souris mutées. Ces effets ont été suivis par un retour aux niveaux d'origines après une période sans traitement de 14 jours.

Pour résumer, la technologie CRISPR-Cas9 nous a permis d'obtenir des souris transgéniques *MET* pR988C mutées qui ont reproduit les maladies auto-immunes et la myéloprolifération trouvée chez notre cas d'index. L'utilisation d'un médicament anti-MET a validé le rôle fonctionnel de cette mutation et de l'activation de la voie *MET* dans le maintien de la myéloprolifération, la polyarthrite rhumatoïde et le syndrome de Sjögren (*Figure 16*).

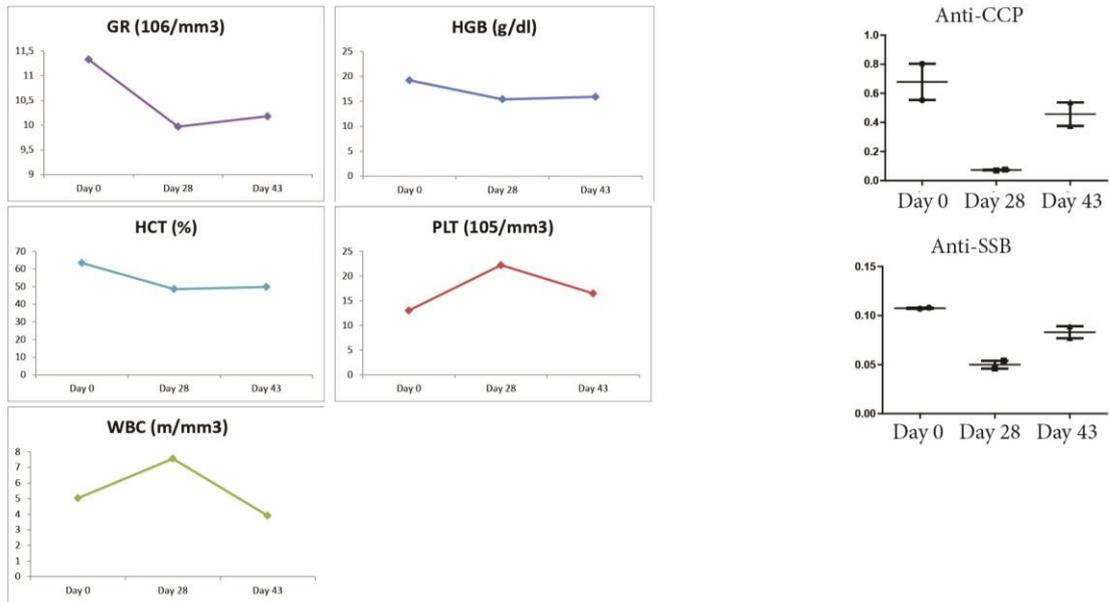
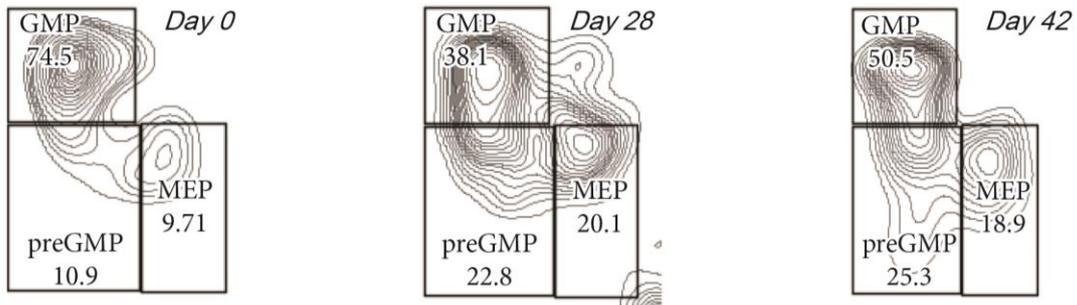
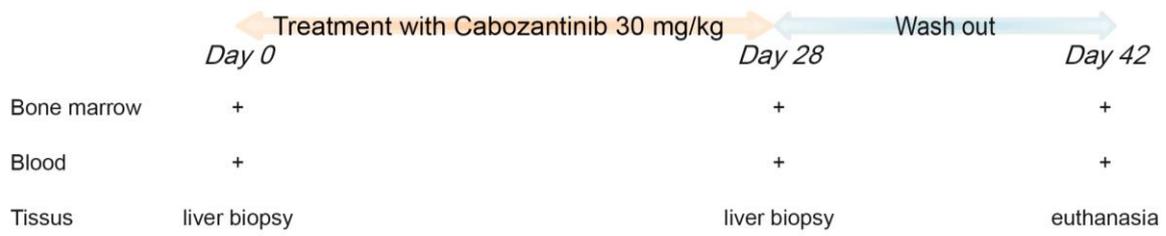


Figure 16 : Inhibition fonctionnelle de Met

II 2. Matériaux et méthodes

II 2 1. Echantillons de patients et de souris utilisés

II 2 1 1. Echantillons humains

Nous avons étudié dans un premier temps les échantillons de notre cas index. Un prélèvement fixé et inclus en paraffine d'un cancer du sein ainsi qu'une aspiration médullaire pour laquelle nous avons trié les cellules progénitrices CD34 positive. Nous avons également étudié l'ADN extrait à partir de follicule pileux de cette patiente.

Un prélèvement fixé et inclus en paraffine du cancer de la thyroïde de la tante paternel de notre cas index a également été étudié.

Une série de 105 prélèvements fixés et inclus en paraffine de cancers du sein sporadique a été analysé ainsi 96 prélèvements sanguins de patients pour lesquels une polyarthrite rhumatoïde a été diagnostiquée.

II 2 1 2. Echantillons murins

Pour le suivi dans le temps de certaines souris, des prélèvements ont été réalisés de façon séquentielle : des prélèvements sanguins ainsi que des aspirations de moelle osseuse fémorale.

Pour l'analyse, les souris sont mises à mort. Et une dissection minutieuse et complète est réalisée. Chaque prélèvement est fixé et inclus en paraffine avec un deuxième prélèvement réalisé en miroir qui est congelé.

II 2 2. Purification d'ADN analyse pangénomique des échantillons humains

L'analyse pangénomique a été réalisée sur des cellules de tumeurs du cancer du sein et sur les cellules progénitrices CD34 positives triées à partir d'une aspiration médullaire. Selon le respect des lois françaises de bioéthiques, la patiente avait été informée de l'utilisation de ses échantillons pour des travaux de recherche après que le diagnostic soit posé et ne s'y ai pas opposé. Un consentement éclairé a été obtenu conformément à la Loi française de bioéthique (2004-800, 08/06/2004). La sélection des cellules tumorales du cancer a été réalisée sur coupes de tissus fixées et inclus en paraffine par microdissection laser (*Figure 17*). Les cellules CD34 + ont été obtenus par tri magnétique à l'aide de billes couplées à un d'anticorps anti-CD34.

L'ADN a été purifié avec le Kit DNA mini kit de Qiagen, quantifié sur Nanodrop et qualifiés par électrophorèse.

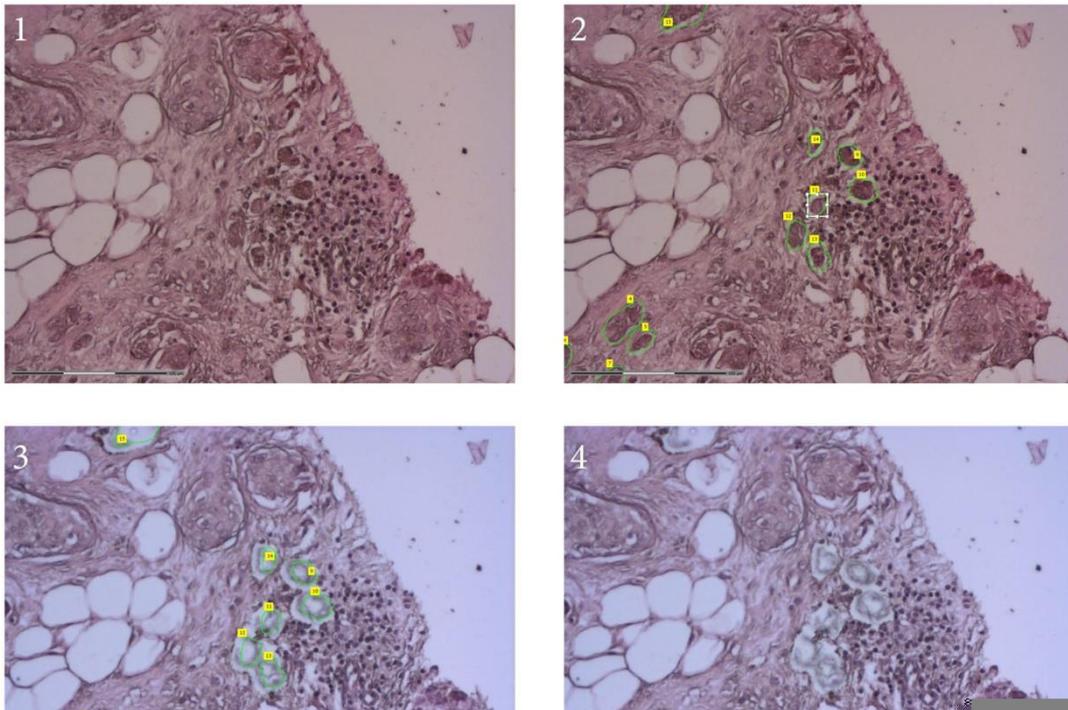


Figure 17 : Cellules tumorales microdisséquées sur une coupe de cancer du sein

II 2 3. Analyse pangénomique des échantillons humains

II 2 3 1. Défi technologique

II 2 3 2. La technologie OncoScan™ FFPE

L'analyse pangénomique a été effectuée à l'aide d'OncoScan™ FFPE qui permet d'analyser des faibles quantités d'ADN (75 ng recommandés), ce qui permet de réaliser la technique sur des prélèvements pauvres en cellules d'intérêt comme des prélèvements issus de biopsies ou de lignées cellulaires. Elle permet de génotyper 330 000 SNPs, des altérations du

nombre de copie, des pertes d'hétérozygotie (LOH) et des mutations somatiques.

Schématiquement, cette technique utilise une sonde inverse (MIP) constituée d'acides nucléiques, dont les 2 extrémités vont s'hybrider sur une cible située sur l'ADN génomique, à 2 endroits distants d'environ 40 paires de base, séparé par un nucléotide (le variant à géotyper). Une fois hybridées, les deux extrémités forment une sonde circulaire, appelée aussi « sonde cadenas », résistante aux exonucléases qui serviront à dégrader les autres nucléotides au cours de la technique. La sonde ainsi fermée va pouvoir être amplifiée par PCR, puis taguée et Détectée par hybridation sur une puce à ADN composée de sondes complémentaires (*Figure 3*). Oncoscan est capable d'interroger 541 mutations somatiques spécifiques du cancer, avec une couverture de plus de 200 gènes suppresseurs de tumeur et oncogènes. L'analyse des données moléculaires issues de la technique OncoScan® FFPE était réalisée grâce aux logiciels « Nexus Express Software for OncoScan FFPE ».

II 2 4. Détection de la mutation de *MET* pR988C

La mutation qui a été détectée par Oncoscan a été vérifiée à l'aide de trois méthodes différentes : PCR à haute résolution (PCR-HRM), séquençage Sanger et digestion par restriction enzymatique.

II 2 4 1. PCR à haute résolution (PCR-HRM)

Pour les échantillons humain, l'analyse de la PCR-HRM de la mutation *MET* permet une identification rapide et à haut débit des variantes dans les régions d'intérêt sans séquençage. La PCR est réalisée en présence d'un fluorochrome intercalant à l'ADN

bicaténaire. Ce fluorophore présente de faibles niveaux de fluorescence lorsqu'il est non lié, mais il est très fluorescent à l'état lié. Les amorces utilisées pour l'amplification par pcr sont :

primer <i>MET</i> sens	5'-GCCTATCCAAATGAGGAGTGTGT-3'
primer <i>MET</i> antisens	5'-TCTGTTTTAAGATCTGGGCAGTGA-3'

La PCR a été réalisées en utilisant le système CFX96TM en temps réel de BioRad dans un total de 20µL contenant 5 µl de ADN génomique (20ng), 15 µl de supermix SsoFastTMAlimor 1 X (Bio-Rad) et 0.4µM de chaque amorce. Le programme utilisé est : une étape initiale de dénaturation à 94° C pendant 2 mn, suivi de 45 cycles de dénaturation (95° C pendant 5 secondes) et d'hybridation (60° C pendant 10 s). Une fois la PCR terminée, une courbe de fusion est réalisée. L'amplicon (typiquement de 50 à 500 pb de long) est progressivement dénaturé en augmentant la température par pas de 0,2 ° C pour finalement atteindre 95 ° C. Chaque série de PCR inclus un contrôle négatif et chaque échantillon a été testé en trois exemplaires. Au cours de cette étape, la courbe de fusion peut être tracée à partir des valeurs d'émission de fluorescence qui diminuent en fonction de la température croissante. La forme unique de la courbe de fusion dépend de la longueur de l'amplicon, de sa séquence, de sa teneur en GC et de la complémentarité des brins d'ADN. Les courbes de fusions après amplification ont été analysées avec le logiciel d'analyse "precision Melt Analysis software".

II 2 4 2. Sequencage Sanger

Les amplicons obtenus par PCR-HRM ont été séquencé par la méthode Sanger. Elle repose sur la propriété qu'ont les ADN polymérase de synthétiser un brin complémentaire

d'un brin matrice en présence de dNTP (désoxyribonucléotide triphosphate) et de ddNTP (didésoxyribonucléotide triphosphate) marqués par fluorescence. Ces derniers servent de terminateurs d'élongation de façon aléatoire dans la réaction. Ainsi, 20 µl de produits PCR ont été purifiés à l'aide de colonnes d'affinité "QIAquick PCR Purification Kit" Qiagen. Le marquage a été réalisée à l'aide de "BigDye®-Terminator sequencing Kit" dans les deux sens selon le programme suivant : une dénaturation initiale à 94° C pendant 3 mn, 25 cycles à 94° C pendant 10 secondes et une élongation à 60° C pendant 20 s. Après une 2ieme étape de purification, les produits ont été migrés sur un séquenceur automatique 16 capillaires, ABI-PRISM® 3130xl GeneticAnalyzer. L'analyse a été réalisé garce au logiciel SeqScape v 2.5.

II 2 4 3. Digestion enzymatique

Cette étape fait intervenir une endonucléase, ici *XmnI*, qui possèdent un site actif d'hydrolyse de l'ADN, mais aussi un site de reconnaissance de la séquence cible, aussi appelée site de restriction. Environ 10 à 15 µl produit de PCR a été digéré avec cinq unités d'enzyme de restriction *XmnI* et l'électrophorèse a été effectuée sur gel d'agarose 2 % (Figure 18)

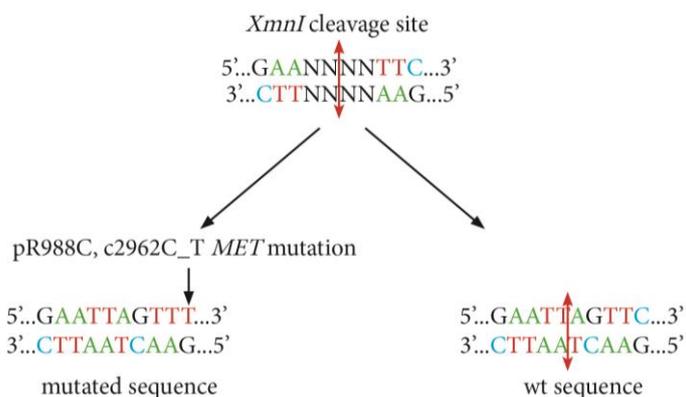


Figure 18 : Schéma de restriction de l'enzyme *XmnI*

II 2 5. Création d'un modèle de souris transgéniques knock-in pour la mutation *MET* humain pR988C

II 2 5 1. Choix du fond génétique

Les souris C57BL/6 ont été choisis pour leurs polyvalence, idéal pour créer un modèle transgénique, et sont de très bon reproducteurs. Elles ont été achetées chez Charles River à 3 à 4 semaines. Toutes les expériences ont été effectuées conformément aux directives du "National Institutes of Health" et les recommandations européennes (2010/63/UE). Le Comité d'éthique du conseil de l'institut universitaire a approuvé cette étude pour expérimentales animales (N ° 2012-15/728-0115).

II 2 5 2. Construction du plasmide

Le système "clustered regularly interspaced short palindromic repeats"/"CRISPR-associated protein 9" (CRISPR/Cas9) a été décrit grâce à l'étude des mécanismes immunes adaptatives de bactéries et d'archées contre des acides nucléiques de pathogènes tels que des virus^{72,73}. Cela met en jeu des petits ARN interférents capables de reconnaître de courtes séquences de pathogènes.

Les gènes *Cas* contiennent des opérons et des régions CRISPR correspondantes à des séquences nucléotidiques de pathogènes, ou espaceurs, entrecoupées de séquences répétées. Dans une première phase, appelée phase adaptative, la réponse se met en place en intégrant des séquences nucléiques de pathogènes appelées proto-espaceurs en 3' de la région CRISPR. Au cours de l'expression génique de la région CRISPR et de l'interférence, la formation d'ARN précurseur, pre-crRNA, est suivie de leur clivage. Cela donne lieu à des crRNA interférents capables de se fixer sur les séquences de pathogènes correspondants à celles intégrées en fin de région CRISPR. Le crRNA interagit avec un crRNA transactivateur (tracrRNA). Cet ensemble oriente ensuite l'action des protéines Cas.

Les études de ces systèmes ont été menées majoritairement dans espèce bactérienne *Streptococcus pyogenes*. Des études du système CRISPR/Cas9 d'autres espèces bactériennes devraient permettre d'élargir les possibilités d'édition du génome à partir de cette famille d'endonucléases⁷⁴.

Le système CRISPR / Cas9 permet des altérations génétiques efficaces et précises en induisant des coupures ciblées d'ADN double brin (DSB) qui stimulent les mécanismes de réparation de l'ADN cellulaire, tels que la jonction d'extrémités non homologues (Non-Homologous End-Joining (NHEJ)) et la réparation dirigée par homologie sans erreur (Homology directed repair (HDR))⁷⁵. Le NHEJ se caractérise par l'introduction aléatoire d'insertion ou de deletion (indels) au site ciblé de l'ADN, permettant ainsi au gène de supprimer un ou deux allèles à travers l'induction de mutations⁷⁶. En revanche, le HDR

provoque des événements knock-in qui entraînent des mutations ponctuelles précises ou l'insertion d'un fragment de séquence souhaitée au niveau du locus ciblé, tel que des remplacements de codons ou des insertions de gènes rapporteurs, par recombinaison avec une séquence homologue exogène ⁷⁶. Dans les cellules des mammifères, le NHEJ est plus efficace que le HDR. Dans les cellules humaines proliférantes, il a été montré que le NHEJ répare 75% des DSB, tandis que HDR répare les 25% restants ⁷⁷, et les cellules souches embryonnaires de souris ont montré un ratio similaire ⁷⁸. L'efficacité du HDR suite à une DSB induite par CRISPR / Cas9 est également assez peu efficace (0,5 à 20%) par rapport à l'efficacité élevée du NHEJ (qui peut atteindre jusqu'à 100%) dans les cellules de mammifères et les embryons de souris ^{79,80}. Par conséquent, la faible efficacité de l'édition de gènes précis médiée par CRISPR / Cas9 reste un défi majeur dans la génération de lignées cellulaires ou de modèles animaux avec la mutation souhaitée ou la correction de la mutation génétique pour la thérapie génique de maladies humains héréditaires.

La conception du plasmide CRISPR/Cas9 a été réalisée avec deux sgRNAs (sgRNA 5 et sgRNA 8) (*Figure 19*). Le plasmide de co-expression des sgRNA-Cas9 (CMV-T7-hspCas9-T2A-GFP-H1-gRNA linéarisé SmartNuclease vector) a été obtenu en

	sequences (5'-3')
Fwsg5RNA	5'AGGGGAAGAGTACACACTCCTCATT3'
Rvsg5RNA	5'AAACAATGAGGAGTGTGACTCTTC 3'
Fwsg8RNA	5'AGGGGGTTAAAGACTTTGCTGTCCAC 3'
Rvsg8RNA	5'AAACGTGACAGCAAAGTCTTTAACC 3'
ssODN	5'CTGTGACAGCAAAGTCTTTAACAAGCTCTCTCTTTCTCTCTGTTTTAAGATCTGGGCAGT GAATTAGTTtGCTATGACGCAAGAGTACACACTCCTCATTTGGATAGGCTTGTAAGTGCCC GAAGTGTAAGTCCAAC3'

Figure 19 : Séquence de sgRNA et de ssODN

utilisant le système de SmartNuclease de Cas9 précision de SBI. Le plasmide ayant intégré le sgRNA a été transfecté dans des bactéries *e. coli* à l'aide du kit One Shot™ TOP10

Chemically Competent *E. coli*

Afin de purifier ces plasmides amplifiés, une sélection à la kanamycine des bactéries transfectées a été réalisée. L'utilisation du kit PureLink HiPure plasmide Midiprep a permis l'extraction de l'ADN plasmidique.

II 2 5 3. Transgénèse

Des souris C57BL/6J femelles ont reçu une injection de gonadotrophine (eCG ou Pregnant Mare Serum Gonadotropin (PMSG)) et de gonadotrophine chorionique humaine (hCG) avec un intervalle de 48 h et accouplés avec des souris mâles C57BL/6J. Après fécondation, l'enzyme Cas9 ARNm (2ng/μl) et les sgRNAs sgRNA5 et sgRNA8 (5ng/μl pour chaque sgRNA) ont été mélangées et injectés dans les pronucléus des embryons unicellulaire. Après la transfection, les zygotes ont été cultivées dans un milieu M16 à 37° C, jusqu'au stade blastocyste (environ 100 cellules) et transférées dans des souris pseudopregnant.

Cette étape a été réalisée à l'animalerie (ANIMA 5) de l'Institut Cochin. Lorsque les fondateurs ont été identifiés, ils ont été transférés à l'animal de l'Institut universitaire d'hématologie de l'hôpital Saint-Louis pour la reproduction. Pour chaque génération, un génotypage a été effectuée sur des souris de 3 semaines.

II 2 5 4. Génotype des souris transgéniques MET pR988C

Comme pour les échantillons de notre cas-index, le génotypage des souris a été réalisé par PCR-HRM et séquençage Sanger en utilisant les amorces suivantes :

primer souris <i>MET</i> sens	5'- GTTTTGTTATTATCCGGGCTCTTCCTGT -3'
primer souris <i>MET</i> antisens	5'- GATATTTCTCAGGATAGTAAACTGAATT -3'

Une digestion par enzyme de restriction a également été réalisée à l'aide de l'enzyme *XmnI* comme décrit ci-dessus.

II 2 5 5. Identification de "Off-target"

Le risque de off-target par CRISPR est associé à la spécificité de la séquence cible. Par exemple, dans les 23 pb du sgRNA, on peut tolérer des divergences allant jusqu'à 5 pb, ce qui peut conduire à une off target ⁸¹. Le nombre d'off-target a été évalué à l'aide de la variation du nombre de copie par droplet digital Polymerase Chain Reaction (ddPCR) en utilisant le système de flux de travail ddPCR de BioRad. La ddPCR) offre une amélioration profonde par rapport à la PCR en temps réel pour l'analyse du CNV. Le milieu réactionnel de ddPCR contient des amorces et des sondes fluorescentes pour chaque séquence d'intérêt (par exemple un CNV et un locus de contrôle à deux copies) et est répartie en milliers de gouttelettes de quelques nanolitres à une dilution suffisante que la plupart des gouttelettes ne contiennent que 0 ou 1 copie d'intérêt. Après PCR, le nombre de gouttelettes fluorescentes est compté (*Figure 20*).

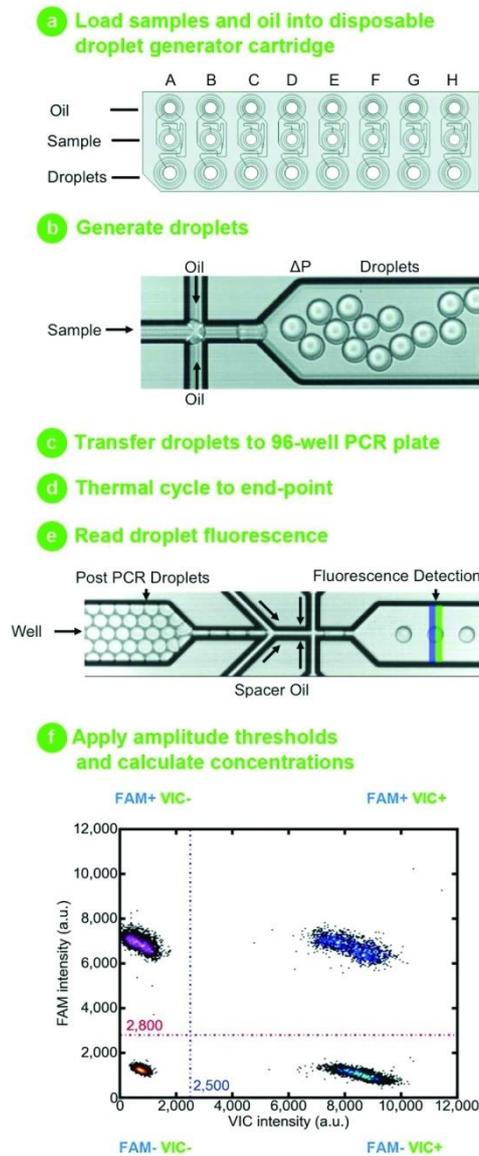


Figure 20 : Méthode de la ddPCR Biorad

(a) Les échantillons et l'huile permettant de générer les gouttelettes sont chargés dans une cartouche à huit canaux. (b) Un vide est appliqué au puits, qui aspire l'échantillon et l'huile à travers une buse de focalisation de flux où sont formées des gouttelettes monodisperses de 1 nL. En moins de 2 min, huit échantillons sont transformés en huit jeux de 20 000 gouttelettes. (c) Les gouttelettes stabilisées par le tensioactif sont transférées par pipette dans une plaque de PCR à 96 puits. (d) L'amplification des gouttelettes par PCR en point final (35-45 cycles) est effectuée dans un thermocycleur conventionnel. (e) La plaque

est chargée sur un lecteur qui compte les gouttelettes de chaque puits et les transmet à un seul fichier après une détection bicolore de ~ 1 000 par seconde. (f) Les gouttelettes sont assignées comme positives ou négatives en fonction de leur amplitude de fluorescence ⁸².

Le mélange réactionnel contient 20ng d'ADN génomique, 10µl de "ddPCR Supermix for probes no dUTP", 1µl de sonde *Met* (Mm00192978_cn, Life Technologies) et 1µl de TFrc sonde de référence (4458366, Life Technologies) par puits, avec un volume final de 20 µl. Le mix de PCR a été ensuite placé dans des DG8 cartouches de génération de gouttelettes QX100 avec 70 µL d'huile de génération de gouttelettes. Après fermeture, les cartouches ont été placés dans le système de génération de gouttelettes Bio-Rad QX200. Le volume final de gouttelettes en suspension dans l'huile était d'environ 40 µL. La PCR a été réalisée dans le système de PCR en temps réel CFX96 (Bio-Rad), avec une première étape de dénaturation à 95° C pendant 10 mn, suivi de 40 cycles de dénaturation (95° C pendant 15 s) et d'hybridation (60° C pendant 1 mn). Chaque série de PCR inclus un contrôle négatif. Les résultats de ddPCR ont été générés en utilisant le lecteur de gouttelettes QX100 et analysés à l'aide du logiciel QuantaSoft. Le ratio de gouttelettes positives pR988C *Met* sur les gouttelettes positive TFrc_a été calculée. Un ratio de 0,8 à 1,2 a été considéré comme un nombre de copie normale pour la mutation *MET* pR988C .

II 2 5 6. Phénotype pR988C *MET* transgéniques

Le sang et la moelle osseuse a été prélevé et analyses à trois moments différents (3 mois, 6 mois et 9 mois) pour chaque souris. À l'euthanasie, tous les organes, y compris la moelle osseuse, ont été systématiquement disséqués pour des analyses tissulaires. Tous les échantillons de tissus ont été traités et couper en deux parties : une fixée en AFA et incluse en paraffine, l'autre congelés dans l'azote liquide et stocké dans la Tumorbank de l'Hôpital

Saint-Louis (Figure 21). Toutes les coupes de tissu ont été effectuées avec un microtome automatique HM-340 à 5 µm et colorées à l'hématoxyline-éosine (HE). Deux pathologistes ont effectué les analyses de tissus.

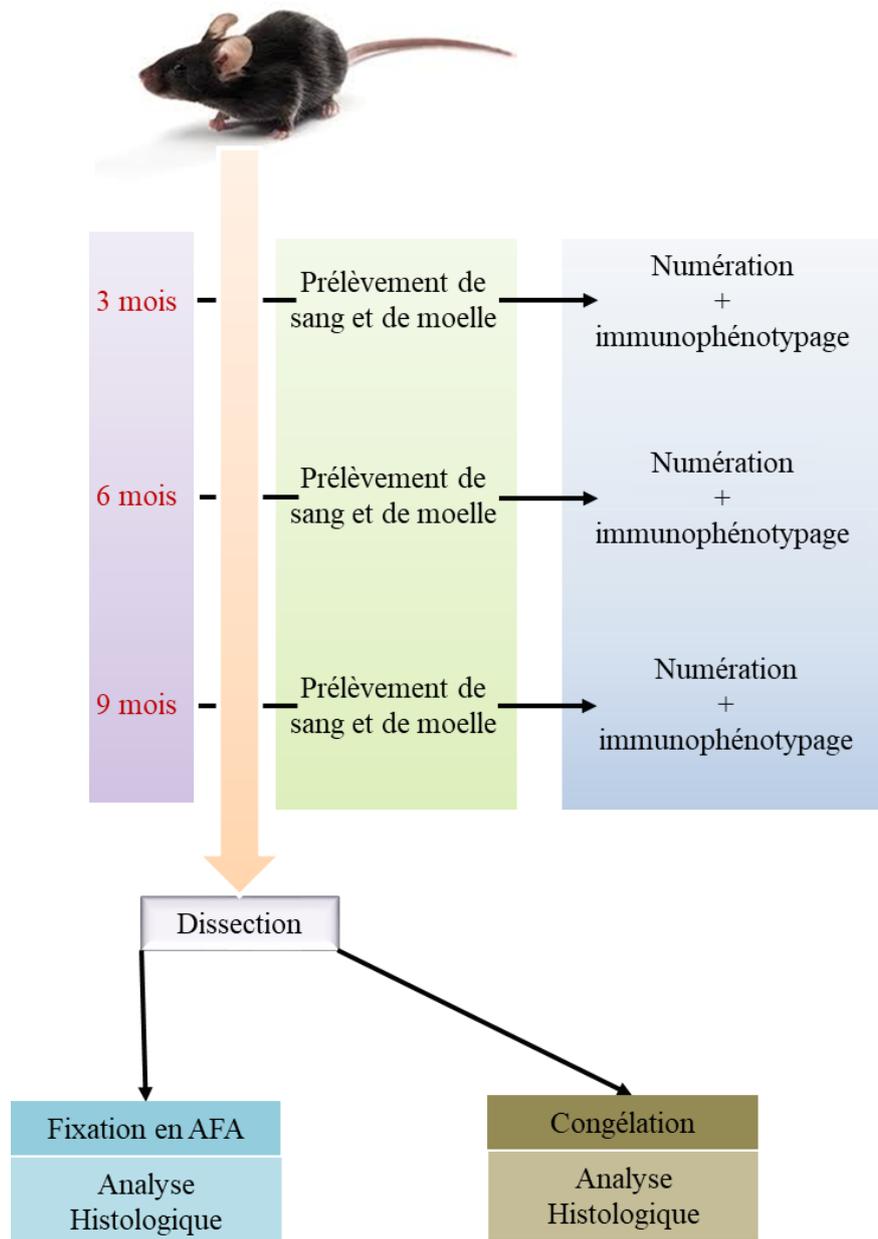


Figure 21 : Protocole d'analyse des souris transgéniques *MET* pR988C, c2962C_T

Pour l'évaluation de l'auto-immunité, le niveau d'auto-anticorps dans le sérum de chaque souris a été mesurée par ELISA (anti-SSA, anti-SSB, anti-ADN double brin, anti-CENT et anti-CCP) selon les protocoles des fabricants.

Des hémogrammes ont été obtenus avec un analyseur MS9-5.

Pour l'analyse de la moelle osseuse, les cellules ont été aspirés à partir de fémurs de chacune des souris et ont été resuspendu dans du PBS/SVF 2 %. Tampon de lyse des globules rouges a été ensuite ajouté. Les Granulocyte Monocyte précurseur (PreGM), les progéniteurs des granulocyte-monocyte (GMP), les progéniteurs des mégacaryocyte et des érythrocytes (MEP), les progéniteurs des érythrocytes (EryP) et les progéniteurs des mégacaryocyte (MkP) ont été marqué avec un panel d'anticorps de rat anti-souris dont le PacBlue cocktail (B220, Ter-119, CD11b CD3e, Gr-1), PercpCy5.5 CD117 (Clone 2 b 8), PE/Cy7 Sca-1 (Clone D7), BUV737 CD16/32 (Clone 2.4G2), Brillant Violet 650 CD150 (Clone TC15-12F12.2), Alexa Fluor 647 CD105 (Clone MJ7/18), Brillant Violet 605 CD41 (Clone MWReg30). Toutes les expériences ont été effectuées à l'aide d'un Fortessa de LSR FACs (BD Biosciences) selon la stratégie de fenêtrage dans les "Extended data"(Behrens et al., 2016) et analysées avec le logiciel FlowJo® (FlowJo, LLC, USA).

II 2 5 7. Inhibition fonctionnelle de la voie *MET*

Pour inhiber la voie *MET* chez les souris nous avons traités avec du cabozantinib (Sellekchem, USA), un inhibiteur *MET* qui agit au niveau du domaine tyrosine kinase, à la dose de 30 mg/kg, par gavage quotidien pendant un mois. Le médicament a été dilué dans 100ml de NaCl à 2 % de DMSO.

Afin d'évaluer l'efficacité du traitement, nous avons effectué des biopsies du foie guidée par échographie, des aspirations de la moelle osseuse et des analyses de sang à trois moments : avant le traitement, après un mois de traitement, et après une période de deux semaines qui ont suivies la fin du traitement.

L'auto-immunité et analyses de la moelle osseuse ont été réalisées comme décrit ci-dessus.

II 4. Discussion

Ici, nous avons démontré la faisabilité d'une approche individuelle en génétique constitutionnelle: i) en étudiant plusieurs échantillons de lésions différentes chez une même jeune patiente, (ii) en utilisant des technologies haut débit d'analyse génomique, (iii) en utilisant la technologie CRISPR-Cas9 pour obtenir des souris transgéniques pour un variant d'intérêt, en quelques mois. Nous avons pu ainsi démontrer la valeur fonctionnelle de la mutation *MET* pR988C, c2962C_T et sa participation tant dans la genèse des néoplasmes myéloprolifératifs que des maladies auto-immunes, un lien important jamais identifié. La polyarthrite rhumatoïde et le syndrome Sjögren sont fréquemment associés⁸³⁻⁸⁵. Pour ces deux maladies auto-immunes, les accumulations familiales ont été rapportées^{86,87}. Les études d'associations génomiques ont identifié plus de 10 allèles dont la valeur fonctionnelle est inconnue, comme HLA-DRB1, IRF5, STAT4, ou BLK⁸⁸⁻⁹². Ici, nous avons démontré le rôle constitutionnelle de la mutation *MET* pR988C, c2962C_T dans le développement de polyarthrite rhumatoïde et le syndrome Sjögren, avec une forte pénétrance phénotypique dans notre modèle murin. Sur la base des résultats de ces données expérimentales, nous nous sommes renseigné sur la présence de polyarthrite rhumatoïde et le syndrome Sjögren dans la famille de notre cas index et avons constaté que son père (III, 8) et son grand-père paternel (II, 3) avait la polyarthrite rhumatoïde (*Figure 22*).

Pour confirmer le rôle de la voie MET dans le développement de polyarthrite rhumatoïde et le syndrome Sjögren, nous avons utilisé le médicament cabozantinib pour inhiber cette voie dans notre modèle de souris. Après 4 semaines d'exposition au médicament, les concentrations sériques d'auto-anticorps spécifiques, anti-SSB et anti-CCP, ont été normalisées. Dans des études antérieures, le rôle de HGF, seul ligand de MET⁹³, a été proposé dans la polyarthrite rhumatoïde. Dans 136 patients, les niveaux élevés de HGF

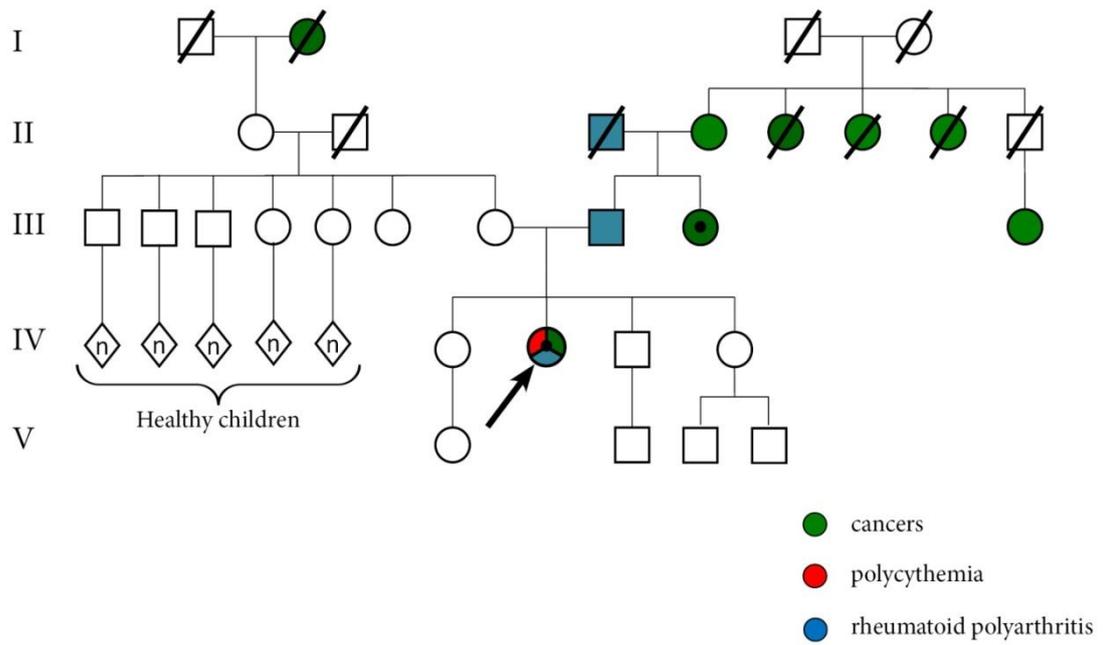


Figure 22 : Généalogie de la patiente avec la polyarthrite rhumatoïde de son père et de son grand-père paternel

plasmatique ont été associés à des destructions articulaires sévères ⁹⁴ et le blocage expérimental de HGF utilisant NK4, un inhibiteur de HGF, diminuées les lésions arthritique chez la souris ⁹⁵. Ici nous avons démontré le rôle d'activation du récepteur MET dans le syndrome de Sjögren et la polyarthrite rhumatoïde.

Dans l'ensemble, ces résultats cliniques et expérimentaux ouvrent le champ à une nouvelle thérapie ciblée en testant des inhibiteurs de MET, déjà disponibles dans la pratique clinique, chez des patients avec une polyarthrite rhumatoïde sévère, un syndrome de Sjögren et l'activation de la voie MET.

Les patients avec des maladies auto-immunes ont 20 % plus de risque de développer des néoplasmes myéloprolifératifs ⁹⁶, y compris la polyglobulie de Vaquez comme pour notre cas index. L'inflammation chronique des maladies auto-immunes pourrait produire des mutations dans des cellules souches hématopoïétiques ⁹⁷, en particulier la mutation *JAK2* V617F ⁹⁸. Cette mutation *JAK2* a été retrouvée parmi 16 autres mutations dans les cellules progénitrices myéloïdes CD34 positives chez notre cas index (voir la Table 1). Quand nous avons testé une série de 96 autres patients avec une polyglobulie de Vaquez avec la mutation *JAK2*, nous avons identifié la mutation *MET* pR988C dans l'ADN d'un cas et cette femme avait aussi une cryoglobulinémie et une polyarthrite rhumatoïde.

L'origine génétique des néoplasmes myéloprolifératifs est établie pour des mutations somatiques dans les gènes *JAK2*, *CALR* ou *MPL* qui conduisent à une prolifération clonale, mais seul des polymorphismes ont été identifiés par des associations d'études génomiques haut-débits comme facteurs génétiques constitutionnels ⁹⁹.

Ici, nous avons démontré que la mutation *MET* pR988C, c2962C_T était suffisant pour induire une myéloprolifération dans toutes les souris mutées. De plus, la participation des trois lignées myéloïdes est un argument pour un effet fonctionnel précoce de cette mutation au niveau des progéniteurs myéloïdes.

Comme dans la polyarthrite rhumatoïde ou le syndrome Sjögren^{94,100}, le niveau de HGF sérique est augmenté chez les patients avec une polyglobulie de Vaquez, indépendamment de la mutation *JAK2* V617F¹⁰¹.

Notre cas index a développé une polyarthrite rhumatoïde et une polyglobulie de Vaquez dans un contexte de mutation *MET* pR988C, c2962C_T constitutionnel. Nous avons pu reproduire ce phénotype chez des souris transgéniques pour la mutation *MET*. Mais notre cas index a aussi développé un cancer du sein et sa tante paternelle avait un cancer de la thyroïde avec la mutation *MET* pR988C.

MET est un proto-oncogène¹⁰² et des mutations nonsense activatrices ont été trouvées dans le domaine tyrosine kinase (TK) de patients avec des cancers du rein papillaire héréditaires^{103,104}, un syndrome autosomale dominant rare avec une forte pénétrance¹⁰⁵. Des mutations somatiques ont principalement été identifiées dans les domaines TK et SEMA de *MET* dans des carcinomes hépatocellulaire de l'enfance et dans des cancers de la tête et le cou, gastrique et du poumon⁹³ (*Figure 21*). La mutation *MET* pR988C, c2962C_T n'est pas localisée dans le TK ou SEMA, mais dans le domaine JM^{106,107}. Jusqu'à présent, cette mutation a été rapportée comme étant sporadique dans des cancers de la thyroïde et du poumon^{108,109}. Dans notre étude, l'analyse de plusieurs lésions d'une même patiente, combiné avec l'enquête de sa famille, nous a permis de démontrer que, dans ce cas, la mutation de *MET* pR988C était constitutionnelle. Quand nous avons testé une série de 105 cancers du sein sans histoire de cancer du sein familial, nous avons retrouvé cette mutation *MET* dans un cas; Malheureusement, nous n'avons pas pu réaliser une enquête familiale puisque le patient était mort.

Ces enquêtes familiales, principalement basées sur des analyses de données rétrospectives, sont parfois difficiles à réaliser. Cependant, notre étude soulève la question des mutations constitutionnelles induisant des cancers jusqu'ici considérés comme sporadique.

Nous n'avons pas reproduit de cancers du sein ou de la thyroïde, mais l'étude systématique des pathologies nous a permis de montrer des thyroïdites lymphocytaires dans 14 des 23 souris transgéniques et, chez les patients, la thyroïdite lymphocytaire est associée à un risque 2 fois accru au cancer papillaire de la thyroïde.

De plus, une des souris transgéniques a développé un carcinome cellulaire cutané sur l'oreille, **une zone exposée**. Le risque de carcinomes cutané est significativement augmenté avec l'application d'un cancérogènes chimique est combinée avec une activation forte de la voie *MET* chez des souris transgéniques pour HGF¹¹⁰. L'activation de cette voie est peut-être trop faible dans nos souris transgéniques avec la mutation *MET* pR988C. Les mutations dans le domaine JM suppriment l'ubiquitination de MET et sa dégradation dans des lignées cellulaire humaines de cancer du poumon^{71,93}, mais le niveau d'activation de la voie MET n'est pas connu.

Dans un autre modèle transgénique de mutation *MET* activatrice dans le domaine TK, une considérable amplification du gène *MET* est requise pour le développement de cancer du sein^{93,111}.

Dans des souris transgéniques pour HGF, les hauts niveaux de HGF sérique ont été associés au développement de cancer du sein¹¹². Chez notre cas index, en plus de la mutation *MET* pR988C, nous avons trouvé une LOH constitutionnel incluant FANCD2, un gène impliqué dans la réparation de l'ADN¹¹³, comme des gènes BRCA impliqué dans le cancers du sein héréditaires. Ceci renforce le rôle de facteurs génétiques constitutionnels dans le développement de cancers d'habitude considérés comme sporadique.

Notre approche combinant l'étude des échantillons cellulaires et tissulaires de plusieurs lésions d'une même patiente, l'utilisation d'analyses génomiques haut débit et de technologie CRISPR-Cas9 nous a permis de démontrer pour la première fois un lien génétique fonctionnel

entre la polyarthrite rhumatoïde, syndrome de Sjögren et polyglobulie de Vaquez. Ce qui ouvre le champ d'application pour cette approche individuelle dans le domaine de facteurs génétiques constitutionnels.

II 5. Bibliographie

- 1 Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153-158, doi:10.1038/nature05610 (2007).
- 2 Shi, J. *et al.* Somatic Genomics and Clinical Features of Lung Adenocarcinoma: A Retrospective Study. *PLoS medicine* **13**, e1002162, doi:10.1371/journal.pmed.1002162 (2016).
- 3 Futreal, P. A. *et al.* A census of human cancer genes. *Nat Rev Cancer* **4**, 177-183, doi:10.1038/nrc1299 (2004).
- 4 Uhlen, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419, doi:10.1126/science.1260419 (2015).
- 5 Johnson, K. C. Accumulating evidence on passive and active smoking and breast cancer risk. *Int J Cancer* **117**, 619-628, doi:10.1002/ijc.21150 (2005).
- 6 Tjonneland, A. *et al.* Alcohol intake and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition (EPIC). *Cancer Causes Control* **18**, 361-373, doi:10.1007/s10552-006-0112-9 (2007).
- 7 Yu, H. Alcohol consumption and breast cancer risk. *JAMA* **280**, 1138-1139 (1998).
- 8 Ronckers, C. M., Erdmann, C. A. & Land, C. E. Radiation and breast cancer: a review of current evidence. *Breast Cancer Res* **7**, 21-32, doi:10.1186/bcr970 (2005).
- 9 Preston, D. L. *et al.* Radiation effects on breast cancer risk: a pooled analysis of eight cohorts. *Radiat Res* **158**, 220-235 (2002).
- 10 van Leeuwen, F. E. *et al.* Roles of radiation dose, chemotherapy, and hormonal factors in breast cancer following Hodgkin's disease. *J Natl Cancer Inst* **95**, 971-980 (2003).
- 11 Tokunaga, M. *et al.* Incidence of female breast cancer among atomic bomb survivors, 1950-1985. *Radiat Res* **138**, 209-223 (1994).
- 12 Land, C. E. *et al.* A case-control interview study of breast cancer among Japanese A-bomb survivors. I. Main effects. *Cancer Causes Control* **5**, 157-165 (1994).
- 13 Land, C. E. *et al.* Incidence of female breast cancer among atomic bomb survivors, Hiroshima and Nagasaki, 1950-1990. *Radiat Res* **160**, 707-717 (2003).
- 14 Hoffman, D. A. *et al.* Breast cancer in women with scoliosis exposed to multiple diagnostic x rays. *J Natl Cancer Inst* **81**, 1307-1312 (1989).
- 15 Boice, J. D., Jr., Preston, D., Davis, F. G. & Monson, R. R. Frequent chest X-ray fluoroscopy and breast cancer incidence among tuberculosis patients in Massachusetts. *Radiat Res* **125**, 214-222 (1991).
- 16 Knudson, A. G. Two genetic hits (more or less) to cancer. *Nat Rev Cancer* **1**, 157-162, doi:10.1038/35101031 (2001).
- 17 Cavenee, W. K. *et al.* Genetic origin of mutations predisposing to retinoblastoma. *Science* **228**, 501-503 (1985).
- 18 Friend, S. H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643-646, doi:10.1038/323643a0 (1986).
- 19 Lee, W. H. *et al.* Human retinoblastoma susceptibility gene: cloning, identification, and sequence. *Science* **235**, 1394-1399 (1987).
- 20 Williams, W. R. & Anderson, D. E. Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. *Genet Epidemiol* **1**, 7-20, doi:10.1002/gepi.1370010104 (1984).
- 21 Hall, J. M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684-1689 (1990).

- 22 Easton, D. F., Bishop, D. T., Ford, D. & Crockford, G. P. Genetic linkage analysis in
familial breast and ovarian cancer: results from 214 families. *The Breast Cancer
Linkage Consortium. American journal of human genetics* **52**, 678-701 (1993).
- 23 Miki, Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene
BRCA1. *Science* **266**, 66-71 (1994).
- 24 de la Chapelle, A. Genetic predisposition to colorectal cancer. *Nat Rev Cancer* **4**, 769-
780, doi:10.1038/nrc1453 (2004).
- 25 Pasca di Magliano, M. & Hebrok, M. Hedgehog signalling in cancer formation and
maintenance. *Nat Rev Cancer* **3**, 903-911, doi:10.1038/nrc1229 (2003).
- 26 Waite, K. A. & Eng, C. From developmental disorder to heritable cancer: it's all in the
BMP/TGF-beta family. *Nature reviews. Genetics* **4**, 763-773, doi:10.1038/nrg1178
(2003).
- 27 Eng, C., Kiuru, M., Fernandez, M. J. & Aaltonen, L. A. A role for mitochondrial
enzymes in inherited neoplasia and beyond. *Nat Rev Cancer* **3**, 193-202,
doi:10.1038/nrc1013 (2003).
- 28 Howlett, N. G. *et al.* Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* **297**,
606-609, doi:10.1126/science.1073834 (2002).
- 29 Fodde, R., Smits, R. & Clevers, H. APC, signal transduction and genetic instability in
colorectal cancer. *Nat Rev Cancer* **1**, 55-67, doi:10.1038/35094067 (2001).
- 30 Stratton, M. R. & Rahman, N. The emerging landscape of breast cancer susceptibility.
Nat Genet **40**, 17-22, doi:10.1038/ng.2007.53 (2008).
- 31 Stacey, S. N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer
susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-869,
doi:10.1038/ng2064 (2007).
- 32 Easton, D. F. *et al.* Genome-wide association study identifies novel breast cancer
susceptibility loci. *Nature* **447**, 1087-1093, doi:10.1038/nature05887 (2007).
- 33 Hunter, D. J. *et al.* A genome-wide association study identifies alleles in FGFR2
associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-874,
doi:10.1038/ng2075 (2007).
- 34 Michailidou, K. *et al.* Large-scale genotyping identifies 41 new loci associated with
breast cancer risk. *Nat Genet* **45**, 353-361, 361e351-352, doi:10.1038/ng.2563 (2013).
- 35 Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000
individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet* **47**, 373-
380, doi:10.1038/ng.3242 (2015).
- 36 Esteban-Jurado, C. *et al.* New genes emerging for colorectal cancer predisposition.
World J Gastroenterol **20**, 1961-1971, doi:10.3748/wjg.v20.i8.1961 (2014).
- 37 Moffa, A. B. & Ethier, S. P. Differential signal transduction of alternatively spliced
FGFR2 variants expressed in human mammary epithelial cells. *J Cell Physiol* **210**,
720-731, doi:10.1002/jcp.20880 (2007).
- 38 Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-
specific long-range interaction with MYC. *Proc Natl Acad Sci U S A* **107**, 9742-9746,
doi:10.1073/pnas.0910668107 (2010).
- 39 Antoniou, A. C. *et al.* Evidence for further breast cancer susceptibility genes in
addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol* **21**, 1-
18, doi:10.1002/gepi.1014 (2001).
- 40 Morgan, J. E. *et al.* Genetic diagnosis of familial breast cancer using clonal
sequencing. *Hum Mutat* **31**, 484-491, doi:10.1002/humu.21216 (2010).
- 41 Schroeder, C., Stutzmann, F., Weber, B. H., Riess, O. & Bonin, M. High-throughput
resequencing in the diagnosis of BRCA1/2 mutations using oligonucleotide

- resequencing microarrays. *Breast Cancer Res Treat* **122**, 287-297, doi:10.1007/s10549-009-0639-z (2010).
- 42 Summerer, D. *et al.* Microarray-based multicycle-enrichment of genomic subsets for targeted next-generation sequencing. *Genome Res* **19**, 1616-1621, doi:10.1101/gr.091942.109 (2009).
- 43 Walsh, T. *et al.* Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A* **107**, 12629-12633, doi:10.1073/pnas.1007983107 (2010).
- 44 Walsh, T. *et al.* Mutations in 12 genes for inherited ovarian, fallopian tube, and peritoneal carcinoma identified by massively parallel sequencing. *Proc Natl Acad Sci U S A* **108**, 18032-18037, doi:10.1073/pnas.1115052108 (2011).
- 45 Pal, T. *et al.* BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. *Cancer* **104**, 2807-2816, doi:10.1002/cncr.21536 (2005).
- 46 Daly, M. B. *et al.* Genetic/familial high-risk assessment: breast and ovarian, version 1.2014. *J Natl Compr Canc Netw* **12**, 1326-1338 (2014).
- 47 Minion, L. E. *et al.* Hereditary predisposition to ovarian cancer, looking beyond BRCA1/BRCA2. *Gynecol Oncol* **137**, 86-92, doi:10.1016/j.ygyno.2015.01.537 (2015).
- 48 Pal, T. *et al.* Frequency of mutations in mismatch repair genes in a population-based study of women with ovarian cancer. *Br J Cancer* **107**, 1783-1790, doi:10.1038/bjc.2012.452 (2012).
- 49 LaDuca, H. *et al.* Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med* **16**, 830-837, doi:10.1038/gim.2014.40 (2014).
- 50 Valle, L. Genetic predisposition to colorectal cancer: where we stand and future perspectives. *World J Gastroenterol* **20**, 9828-9849, doi:10.3748/wjg.v20.i29.9828 (2014).
- 51 Spier, I. *et al.* Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. *Int J Cancer* **137**, 320-331, doi:10.1002/ijc.29396 (2015).
- 52 Bellido, F. *et al.* POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. *Genet Med* **18**, 325-332, doi:10.1038/gim.2015.75 (2016).
- 53 Lynch, H. T. *et al.* Familial pancreatic cancer: a review. *Semin Oncol* **23**, 251-275 (1996).
- 54 Slater, E. P. *et al.* PALB2 mutations in European familial pancreatic cancer families. *Clin Genet* **78**, 490-494, doi:10.1111/j.1399-0004.2010.01425.x (2010).
- 55 Axilbund, J. E. & Wiley, E. A. Genetic testing by cancer site: pancreas. *Cancer J* **18**, 350-354, doi:10.1097/PPO.0b013e3182624694 (2012).
- 56 Hu, C. *et al.* Prevalence of Pathogenic Mutations in Cancer Predisposition Genes among Pancreatic Cancer Patients. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **25**, 207-211, doi:10.1158/1055-9965.EPI-15-0455 (2016).
- 57 Grant, R. C. *et al.* Prevalence of germline mutations in cancer predisposition genes in patients with pancreatic cancer. *Gastroenterology* **148**, 556-564, doi:10.1053/j.gastro.2014.11.042 (2015).
- 58 Syngal, S. *et al.* ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol* **110**, 223-262; quiz 263, doi:10.1038/ajg.2014.435 (2015).

- 59 Farley, M. N. *et al.* A novel germline mutation in BAP1 predisposes to familial clear-cell renal cell carcinoma. *Mol Cancer Res* **11**, 1061-1071, doi:10.1158/1541-7786.MCR-13-0111 (2013).
- 60 Pavlovich, C. P. & Schmidt, L. S. Searching for the hereditary causes of renal-cell carcinoma. *Nat Rev Cancer* **4**, 381-393, doi:10.1038/nrc1364 (2004).
- 61 Linehan, W. M. Genetic basis of kidney cancer: role of genomics for the development of disease-based therapeutics. *Genome Res* **22**, 2089-2100, doi:10.1101/gr.131110.111 (2012).
- 62 Malinoc, A. *et al.* Biallelic inactivation of the SDHC gene in renal carcinoma associated with paraganglioma syndrome type 3. *Endocr Relat Cancer* **19**, 283-290, doi:10.1530/ERC-11-0324 (2012).
- 63 Ricketts, C. J. *et al.* Succinate dehydrogenase kidney cancer: an aggressive example of the Warburg effect in cancer. *J Urol* **188**, 2063-2071, doi:10.1016/j.juro.2012.08.030 (2012).
- 64 Aletaha, D. *et al.* 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum* **62**, 2569-2581, doi:10.1002/art.27584 (2010).
- 65 Tuefferd, M. *et al.* Genome-wide copy number alterations detection in fresh frozen and matched FFPE samples using SNP 6.0 arrays. *Genes Chromosomes Cancer* **47**, 957-964, doi:10.1002/gcc.20599 (2008).
- 66 Lyons-Weiler, M., Hagenkord, J., Sciulli, C., Dhir, R. & Monzon, F. A. Optimization of the Affymetrix GeneChip Mapping 10K 2.0 Assay for routine clinical use on formalin-fixed paraffin-embedded tissues. *Diagn Mol Pathol* **17**, 3-13, doi:10.1097/PDM.0b013e31815aca30 (2008).
- 67 Lips, E. H. *et al.* Reliable high-throughput genotyping and loss-of-heterozygosity detection in formalin-fixed, paraffin-embedded tumors using single nucleotide polymorphism arrays. *Cancer Res* **65**, 10188-10191, doi:10.1158/0008-5472.CAN-05-2486 (2005).
- 68 Jacobs, S. *et al.* Genome-wide, high-resolution detection of copy number, loss of heterozygosity, and genotypes from formalin-fixed, paraffin-embedded tumor tissue using microarrays. *Cancer Res* **67**, 2544-2551, doi:10.1158/0008-5472.CAN-06-3597 (2007).
- 69 Matsumoto, K., Umitsu, M., De Silva, D. M., Roy, A. & Bottaro, D. P. Hepatocyte growth factor/MET in cancer progression and biomarker discovery. *Cancer Sci* **108**, 296-307, doi:10.1111/cas.13156 (2017).
- 70 Peschard, P. *et al.* Mutation of the c-Cbl TKB domain binding site on the Met receptor tyrosine kinase converts it into a transforming protein. *Mol Cell* **8**, 995-1004 (2001).
- 71 Kong-Beltran, M. *et al.* Somatic mutations lead to an oncogenic deletion of met in lung cancer. *Cancer Res* **66**, 283-289, doi:10.1158/0008-5472.CAN-05-2749 (2006).
- 72 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607, doi:10.1038/nature09886 (2011).
- 73 Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331-338, doi:10.1038/nature10886 (2012).
- 74 Fonfara, I. *et al.* Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic acids research* **42**, 2577-2590, doi:10.1093/nar/gkt1074 (2014).
- 75 Li, G. *et al.* Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. *Sci Rep* **7**, 8943, doi:10.1038/s41598-017-09306-x (2017).

- 76 Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol* **26**, 52-64, doi:10.1016/j.tcb.2015.07.009 (2016).
- 77 Mao, Z., Bozzella, M., Seluanov, A. & Gorbunova, V. Comparison of nonhomologous end joining and homologous recombination in human cells. *DNA Repair (Amst)* **7**, 1765-1771, doi:10.1016/j.dnarep.2008.06.018 (2008).
- 78 Johnson, R. D. & Jasin, M. Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. *EMBO J* **19**, 3398-3407, doi:10.1093/emboj/19.13.3398 (2000).
- 79 Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370-1379, doi:10.1016/j.cell.2013.08.022 (2013).
- 80 Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-918, doi:10.1016/j.cell.2013.04.025 (2013).
- 81 Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-826, doi:10.1038/nbt.2623 (2013).
- 82 Hindson, B. J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* **83**, 8604-8610, doi:10.1021/ac202028g (2011).
- 83 Fallahi, P. *et al.* The association of other autoimmune diseases in patients with autoimmune thyroiditis: Review of the literature and report of a large series of patients. *Autoimmun Rev* **15**, 1125-1128, doi:10.1016/j.autrev.2016.09.009 (2016).
- 84 Moutsopoulos, H. M., Webber, B. L., Vlagopoulos, T. P., Chused, T. M. & Decker, J. L. Differences in the clinical manifestations of sicca syndrome in the presence and absence of rheumatoid arthritis. *Am J Med* **66**, 733-736 (1979).
- 85 Gilboe, I. M., Kvien, T. K., Uhlig, T. & Husby, G. Sicca symptoms and secondary Sjogren's syndrome in systemic lupus erythematosus: comparison with rheumatoid arthritis and correlation with disease variables. *Ann Rheum Dis* **60**, 1103-1109 (2001).
- 86 Deighton, C. M., Walker, D. J., Griffiths, I. D. & Roberts, D. F. The contribution of HLA to rheumatoid arthritis. *Clin Genet* **36**, 178-182 (1989).
- 87 Sabio, J. M., Milla, E. & Jimenez-Alonso, J. A multicase family with primary Sjogren's syndrome. *J Rheumatol* **28**, 1932-1934 (2001).
- 88 Lessard, C. J. *et al.* Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. *Nat Genet* **45**, 1284-1292, doi:10.1038/ng.2792 (2013).
- 89 Li, Y. *et al.* A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjogren's syndrome at 7q11.23. *Nat Genet* **45**, 1361-1365, doi:10.1038/ng.2779 (2013).
- 90 Gregersen, P. K., Silver, J. & Winchester, R. J. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* **30**, 1205-1213 (1987).
- 91 Gourraud, P. A. *et al.* A new classification of HLA-DRB1 alleles differentiates predisposing and protective alleles for autoantibody production in rheumatoid arthritis. *Arthritis Res Ther* **9**, R27, doi:10.1186/ar2131 (2007).
- 92 Korman, B. D., Kastner, D. L., Gregersen, P. K. & Remmers, E. F. STAT4: genetics, mechanisms, and implications for autoimmunity. *Curr Allergy Asthma Rep* **8**, 398-403 (2008).

- 93 Graveel, C. R., Tolbert, D. & Vande Woude, G. F. MET: a critical player in tumorigenesis and therapeutic target. *Cold Spring Harb Perspect Biol* **5**, doi:10.1101/cshperspect.a009209 (2013).
- 94 Grandaunet, B. *et al.* Association between high plasma levels of hepatocyte growth factor and progression of radiographic damage in the joints of patients with rheumatoid arthritis. *Arthritis Rheum* **63**, 662-669, doi:10.1002/art.30163 (2011).
- 95 Tsunemi, S. *et al.* Molecular targeting of hepatocyte growth factor by an antagonist, NK4, in the treatment of rheumatoid arthritis. *Arthritis Res Ther* **15**, R75, doi:10.1186/ar4252 (2013).
- 96 Kristinsson, S. Y., Landgren, O., Samuelsson, J., Bjorkholm, M. & Goldin, L. R. Autoimmunity and the risk of myeloproliferative neoplasms. *Haematologica* **95**, 1216-1220, doi:10.3324/haematol.2009.020412 (2010).
- 97 Sorensen, A. L. & Hasselbalch, H. C. Antecedent cardiovascular disease and autoimmunity in Philadelphia-negative chronic myeloproliferative neoplasms. *Leuk Res* **41**, 27-35, doi:10.1016/j.leukres.2015.11.017 (2016).
- 98 Barbui, T. *et al.* Inflammation and thrombosis in essential thrombocythemia and polycythemia vera: different role of C-reactive protein and pentraxin 3. *Haematologica* **96**, 315-318, doi:10.3324/haematol.2010.031070 (2011).
- 99 Tapper, W. *et al.* Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nat Commun* **6**, 6691, doi:10.1038/ncomms7691 (2015).
- 100 Balarini, G. M. *et al.* Serum calprotectin is a biomarker of carotid atherosclerosis in patients with primary Sjogren's syndrome. *Clin Exp Rheumatol* **34**, 1006-1012 (2016).
- 101 Boissinot, M. *et al.* Anti-inflammatory cytokines hepatocyte growth factor and interleukin-11 are over-expressed in Polycythemia vera and contribute to the growth of clonal erythroblasts independently of JAK2V617F. *Oncogene* **30**, 990-1001, doi:10.1038/onc.2010.479 (2011).
- 102 Cooper, C. S. *et al.* Molecular cloning of a new transforming gene from a chemically transformed human cell line. *Nature* **311**, 29-33 (1984).
- 103 Schmidt, L. *et al.* Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. *Nat Genet* **16**, 68-73, doi:10.1038/ng0597-68 (1997).
- 104 Schmidt, L. *et al.* Two North American families with hereditary papillary renal carcinoma and identical novel mutations in the MET proto-oncogene. *Cancer Res* **58**, 1719-1722 (1998).
- 105 Verine, J. *et al.* Hereditary renal cancer syndromes: an update of a systematic review. *Eur Urol* **58**, 701-710, doi:10.1016/j.eururo.2010.08.031 (2010).
- 106 Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543-550, doi:10.1038/nature13385 (2014).
- 107 Reungwetwattana, T. & Ou, S. H. MET exon 14 deletion (METex14): finally, a frequent-enough actionable oncogenic driver mutation in non-small cell lung cancer to lead MET inhibitors out of "40 years of wilderness" and into a clear path of regulatory approval. *Transl Lung Cancer Res* **4**, 820-824, doi:10.3978/j.issn.2218-6751.2015.12.03 (2015).
- 108 Tyner, J. W. *et al.* MET receptor sequence variants R970C and T992I lack transforming capacity. *Cancer Res* **70**, 6233-6237, doi:10.1158/0008-5472.CAN-10-0429 (2010).
- 109 Boland, J. M. *et al.* MET and EGFR mutations identified in ALK-rearranged pulmonary adenocarcinoma: molecular analysis of 25 ALK-positive cases. *J Thorac Oncol* **8**, 574-581, doi:10.1097/JTO.0b013e318287c395 (2013).

- 110 Cataisson, C. *et al.* MET signaling in keratinocytes activates EGFR and initiates squamous carcinogenesis. *Sci Signal* **9**, ra62, doi:10.1126/scisignal.aaf5106 (2016).
- 111 Graveel, C. R. *et al.* Met induces diverse mammary carcinomas in mice and is associated with human basal breast cancer. *Proc Natl Acad Sci U S A* **106**, 12909-12914, doi:10.1073/pnas.0810403106 (2009).
- 112 Ma, J. *et al.* Somatic mutation and functional polymorphism of a novel regulatory element in the HGF gene promoter causes its aberrant expression in human breast cancer. *J Clin Invest* **119**, 478-491, doi:10.1172/JCI36640 (2009).
- 113 Krzystolik, K. *et al.* Large deletion causing von Hippel-Lindau disease and hereditary breast cancer syndrome. *Hered Cancer Clin Pract* **12**, 16, doi:10.1186/1897-4287-12-16 (2014).