



HAL
open science

Évaluation de l'hétérogénéité mitochondriale par séquençage de l'ADN

Nicolas Pons, Pascal Barbry, Sophie Couvé, Sophie Gad-Lapiteau

► **To cite this version:**

Nicolas Pons, Pascal Barbry, Sophie Couvé, Sophie Gad-Lapiteau. Évaluation de l'hétérogénéité mitochondriale par séquençage de l'ADN. *Génomique, Transcriptomique et Protéomique [q-bio.GN]*. 2018. hal-01687510

HAL Id: hal-01687510

<https://ephe.hal.science/hal-01687510>

Submitted on 18 Jan 2018

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

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



École Pratique
des Hautes Études



MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

ÉCOLE PRATIQUE DES HAUTES ÉTUDES
Sciences de la Vie et de la Terre

MEMOIRE

Présenté
par
Pons Nicolas

pour l'obtention du Diplôme de l'École Pratique des Hautes Études

TITRE : Evaluation de l'hétérogénéité mitochondriale par séquençage de l'ADN

Soutenu le 11 Janvier 2018

devant le jury suivant :

Dr RENAUD-PAITRA Flore - **Président**
Dr BARBRY Pascal - **Tuteur scientifique**
Dr COUVE Sophie - **Tutrice pédagogique**
Dr PAQUIS Véronique - **Rapporteur**
Dr MARI Bernard - **Examineur**

Mémoire préparé sous la direction de :
BARBRY Pascal

UCA GenomiX et Laboratoire de Physiologie Génomique des Eucaryotes
Université Côte d'Azur et CNRS, IPMC,
660 route des lucioles, 06560 Valbonne

Et de

Couvé Sophie & Gad Sophie

Groupe Génétique Oncologique EPHE, UMR Inserm 1186
Institut Gustave Roussy, 114, rue Edouard Vaillant - 94800 Villejuif
EPHE (Science de la Vie et de la Terre)

Groupe de Recherche et d'Enseignement Thématiques de l'EPHE : Génomes, Cancers & Thérapies.

Remerciements

Mes premiers remerciements reviennent aux membres de mon jury Pascal Barbry, Sophie Gad-Lapiteau, Bernard Mari, Véronique Paquis et Flore Renaud d'avoir fait l'honneur d'accepter de juger mon travail pour le diplôme de l'Ecole Pratique des Hautes Etudes. Merci pour votre temps et votre implication.

Je remercie Sophie Couvé, d'avoir accepté de me suivre tout au long de ces deux années, de m'avoir assisté et orienté durant mon diplôme et je remercie Sophie Gad-Lapiteau d'avoir repris le flambeau et d'avoir tout mis en œuvre pour me permettre de présenter mon diplôme malgré l'urgence de la situation.

Un grand merci à mon tuteur scientifique et directeur Pascal Barbry pour m'avoir accueilli dans le laboratoire depuis mon BTS à maintenant, d'avoir cru en moi et m'avoir donné l'opportunité de continuer mes études en parallèle de mon travail sur la plateforme. Tu m'as beaucoup apporté aussi bien sur le plan professionnel que sur le plan humain. Pour avoir été aussi compréhensif lors des moments de stress, et pour m'avoir constamment poussé à m'améliorer.

J'aimerais remercier Bernard Mari, qui a été le premier à me donner une chance lors de ma première année de mes études supérieures ainsi que les suivantes alors que de rudes batailles font rage pour trouver un stage dans le milieu professionnel. Une personne avec qui j'ai eu la chance de partager de bons moments que se soit au labo comme autour d'une mousse.

Par la suite, je tiens à remercier la plateforme, Géraldine Rios, Agnès Paquet, Nicolas Nottet, Kévin Lebrigand et Marie-Jeanne Arguel qui en plus de m'avoir aidé d'un point de vue professionnel, m'a également fait découvrir le monde sous-marin. Je n'oublie surtout pas Virginie Magnone pour laquelle ma gratitude n'a pas de limite. Tu as été une figure maternelle tout au long de ces années, trahie parfois par le nom de ton fils lorsque tu souhaitais me parler. Je te remercie pour ta disponibilité, ton écoute ainsi que tes nombreux conseils.

Je remercie également les personnes de l'équipe, qui ont contribué à enrichir et égayer mes journées durant ces années. Merci à Laure-Emmanuelle Zaragosi pour son affabilité. Merci à

Roger Rezzonico, Rainer Waldmann, Carole Gaillard et Karine Robbe-Sermesant. Un grand merci à Gilles Ponzio, qui me faisait déjà écouter ses chants avant même ma naissance et qui selon les dires aurait pu faire de moi un chanteur. Merci pour ces bons moments passés, ces discussions du matin et ces instants de rire. Je remercie Charlotte Pandiani pour nos parties de tennis et les journées canyoning, Sandra Ruiz-Garcia avec laquelle j'ai pu améliorer mon anglais, je la remercie pour sa patience face à mes phrases parfois incohérentes, Marine Gautier, Marie Deprez, Amélie Cavard, Serena Diazzi et Julien Fassy pour nos soirées, notre dégustation de limocello, nos verres after-work et pour certaines, nos récurrentes et tant attendues discussions chats. Pour terminer l'équipe, je tiens à remercier les personnes qui ont fait partie de ma vie professionnelle et personnelle depuis le début de mon arrivée et pour lesquelles j'ai partagé de grandes conversations, des fous rires, des challenges sportifs et de nombreuses soirées. Je pense bien évidemment à Olivier Mercey, Richard Allan et Laura Moreno-Leon.

J'aimerais remercier les personnes extérieures à notre équipe, ceux qui ont embelli mes journées et mes soirées. Je pense en premier lieu Thomas, Miled et Hadi qui m'extirpaient aux aurores hors de mon lit et me motivaient pour aller au sport le matin. Je pense aussi à Sarah, Mélanie et Pierre lors de nos soirées et qui font partie de mes nombreux bons souvenirs. Je remercie également les étudiants de la Team Checler pour leurs soirées jeux.

Pour finir, je souhaite profondément remercier ma famille et surtout mes parents qui ont grandement contribué au parcours me permettant d'arriver à ce jour. Ils ont su me donner toutes les chances pour réussir. Je les remercie pour leur soutien tout au long de ces années, pour vos encouragements dont je vous en suis infiniment reconnaissant.

Sommaire

Abréviations	6
Liste des figures	8
Liste des tableaux.....	10
Liste des annexes	11
Introduction.....	12
I. Les cellules de l'épithélium respiratoire	12
a. Les cellules basales	13
b. Les cellules sécrétrices de mucus.....	14
c. Les cellules multiciliées	14
II. La mitochondrie	15
a. Généralité et structure	15
b. ADN mitochondrial.....	18
c. ADN mitochondrial et hétéroplasmie	19
d. Héritage mitochondrial.....	22
III. Séquençage et nouvelle technologie	25
a. Les différents types de séquençage	25
b. Cellule unique	27
Objectif	31

Matériels et méthodes	33
I. Culture cellulaire.....	33
II. Echantillons.....	33
a. RNAseq.....	33
b. Biopsie (10X Chromium Genomics)	33
III. Extraction	34
a. Blood and tissue Kit (Qiagen).....	34
b. Miniprep (Qiagen).....	34
IV. PCR (Polymerase Chain Reaction)	34
a. Short-PCR	34
b. Long-PCR	35
c. qPCR (quantitative PCR)	35
V. Purification des acides nucléiques	36
VI. Digestion enzymatique.....	37
VII. Sonication.....	37
VIII. Contrôles	37
a. Nanodrop.....	37
b. Bioanalyzer	38
IX. C1 Fluidigm	38
X. Chromium 10X Genomics	38

XI.	Séquençage.....	39
a.	RNA-seq.....	39
b.	DNA-seq	40
XII.	Bioinformatique	40
	Résultats.....	41
I.	Mise en place du séquençage d'ADNmt.....	41
II.	Comparaison des données de séquençage d'ADNmt.....	46
III.	Mise en place du séquençage d'ADNmt en cellule unique.....	51
IV.	Analyse mitochondriale des données RNAseq et 10X Genomics	56
a.	Détermination de l'haplogroupe mitochondrial.....	57
b.	Recherche de contamination	62
c.	Analyse de l'hétérogénéité.....	64
	Conclusion - Discussion	78
	Perspectives.....	81
	Bibliographies	82
	Annexes.....	87

Abréviations

ADN: Acide désoxyribonucléique

ADNc: Acide désoxyribonucléique complémentaire

ADNmt: Acide désoxyribonucléique mitochondrial

ADNn: Acide désoxyribonucléique nucléaire

ARN: Acide ribonucléique

ARNm: Acide ribonucléique messenger

ARNnc: Acide ribonucléique non codant

ARNt: Acide ribonucléique de transfert

ATP: Adénosine triphosphate

BPCO: Bronchopneumopathies chroniques obstructives

BT: Blood and Tissue

COX3: Cytochrome c oxydase 3

CRT: Terminaison réversible cyclique

CYTB: Cytochrome b

DMEM: Dulbecco's Modified Eagle's Medium

dNTP: Désoxynucléotide tri-phosphate

EIM: Espace inter-membranaire

HEK: Human embryonic kidney cell

HTS: High throughput sequencing

IFC: Integrated fluidic circuit

MAMs: Mitochondria-associated ER-membrane / point de contact entre la mitochondrie et le réticulum endoplasmique

MCC: Cellule multiciliée

MEM: Membrane externe mitochondriale

MIM: Membrane interne mitochondriale

ND1: NADH dehydrogenase 1

NGS: Next generation sequencing

Pb: Paire de base

PCR: Polymerase chain reaction

RT: Reverse transcription

SNA: addition simple de nucléotide

SNP: Single nucleotide polymorphism

SVF: Sérum de veau foetal

UCA: Université Côte d'Azur

Liste des figures

Figure 1 : Image de cellules basales au microscope confocal

Figure 2 : Image de cellules à mucus au microscope confocal

Figure 3 : Image d'une cellule multiciliée au microscope confocal

Figure 4 : Structure mitochondriale

Figure 5 : ADN mitochondrial

Figure 6 : Répartition aléatoire des mitochondries

Figure 7 : Migrations de la population humaine et les principaux haplogroupes d'ADNmt

Figure 8 : Principe de séquençage par terminaison réversible cyclique

Figure 9 : Principe de séquençage par addition simple de nucléotide

Figure 10 : Principe du C1 de Fluidigm

Figure 11 : Principe de la technique du Chromium de chez 10X Genomics

Figure 12: Migration des produits de PCR de trois régions codantes dur l'ADNmt en gel d'agarose 1%

Figure 13 : Migration du produit de « Long-PCR » sur gel d'agarose 0,7%

Figure 14 : Migration des produits de PCR sur gel d'agarose 1%

Figure 15 : Résultats de digestion de l'ADNmt avec EcoRI et HindIII

Figure 16 : Test de fragmentation par sonication

Figure 17 : Profil Bioanalyzer de la librairie d'ADNmt

Figure 18 : Fragmentation de l'ADNmt

Figure 19 : Librairies d'ADNmt à partir du nez et du sang du patient

Figure 20 : Comparaison des pourcentages d'hétéroplasmie entre l'ADN sanguin et l'ADN nasal

Figure 21 : Alignement sur IGV des séquences d'ADNmt des librairies de sang et de nez du patient PB

Figure 22 : Répartition nucléotidique pour le polymorphisme G1338

Figure 23 : Migration des produits de PCR à différentes quantités sur gel d'agarose 1%

Figure 24 : Migration des produits de PCR à 10pg sur gel d'agarose 1%

Figure 25 : Migration des produits de PCR des trois régions sur gel d'agarose 1%

Figure 26 : Profil BioAnalyzer de la librairie d'ADNmt obtenue à partir de 10pg d'ADN total

Figure 27 : Capture des cellules sur la plaque C1 à l'aide du LAS Imager

Figure 28 : Profil BioAnalyzer des librairies d'ADNmt d'une cellule

Figure 29 : Représentation t-SNE des cellules d'haplogroupe H et D4q dans l'échantillon mélanome

Figure 30 : Représentation graphique de la répartition cellulaire de l'échantillon PB

Figure 31 : Comparaison du nombre de polymorphismes entre le nez, le sang et le cluster MUC5AC

Figure 31 : Comparaison du nombre de polymorphismes entre le nez, le sang et le cluster MUC5AC

Figure 32 : Représentation nucléotidique avec IGV à la position 2617 de l'ADNmt

Figure 33 : Répartition des cellules avec une mutation dans l'échantillon D246 ali2

Figure 34 : Répartition des cellules avec la mutation A750G dans l'échantillon D246 ali22

Figure 35 : Répartition cellulaire de l'échantillon de biopsie de cornets nasaux

Figure 36 : Comparaison de polymorphismes entre les différents clusters

Figure 37 : Hétérogénéité de séquences sur IGV avec le patient D267

Figure 38 : Position des cellules mutées de l'échantillon D267

Figure 39 : Répartition cellulaire de l'échantillon de mélanome

Figure 40 : Position des cellules mutées dans l'échantillon

Liste des tableaux

Tableau 1 : Liste des amorces

Tableau 2 : Haplogroupe à partir de l'ADNmt provenant de données RNA-seq.

Tableau 3 : Comparaison de l'haplogroupe d'un patient provenant de deux tissus différents.

Tableau 4 : Haplogroupes des patients séquencés à partir du Chromium 10X Genomics.

Tableau 5 : Haplogroupes des cellules du patient PB

Tableau 6 : Haplogroupes des cellules du mélanome

Tableau 7 : Comparaison de six positions nucléotidiques issues du génome mitochondrial à partir de tissus nasaux et bronchiques chez un même patient.

Tableau 8 : Mutation retrouvée dans une seule condition

Liste des annexes

Annexe 1 : Arbre phylogénétique de l'ADNmt

Annexe 2 : Alignement des séquences des bibliothèques Nez et Sang sur l'ADNmt

Annexe 3 : Répartition cellulaire des cellules exprimant MUC5AC du patient PB

Annexe 4 : Répartition cellulaire du patient D246 Ali2

Annexe 5 : Répartition cellulaire du patient D246 Ali22

Annexe 3 : Répartition cellulaire des cellules exprimant KRT5 du patient 258

Annexe 7 : Position possédant au moins une cellule mutée

Annexe 8 : Répartition cellulaire du patient D267

Introduction

L'équipe dans laquelle je travaille, est située à l'Institut de Pharmacologie Moléculaire et Cellulaire (IPMC). Elle est dirigée par le docteur Pascal Barbry et organisée autour de deux activités : (1) L'équipe de recherche s'intéresse à la complexité des ARN non codants (ARNnc) dans les tissus respiratoires. Le tissu respiratoire est soumis en permanence à des stress physicochimiques et biologiques, qui peuvent altérer le bon fonctionnement de l'organe. Il doit disposer de programmes pour réparer le cas échéant ces lésions, par une combinaison de mécanismes de prolifération, migration et différenciation des cellules, afin d'assurer la bonne régénération d'un tissu fonctionnel. Toute perturbation de ces programmes peut provoquer des pathologies pulmonaires. Ces perturbations sont rencontrées dans des pathologies aussi diverses que les cancers pulmonaires, la fibrose pulmonaire, la mucoviscidose, les bronchopneumopathies chroniques obstructives (BPCO), ou l'asthme : (2) La plateforme de génomique fonctionnelle, intitulée UCA GenomiX, est un des 8 membres fondateurs de l'infrastructure nationale de France Génomique. La plateforme est certifiée ISO9001 depuis 2006, et son activité s'articule autour de développements en biologie moléculaire, génomique et bio-informatique. UCA GenomiX propose des solutions technologiques à la communauté académique, dans le cadre de collaborations locales, nationales et internationales. La plateforme de génomique fonctionnelle développe principalement son activité autour d'approches par séquençages à haut débit (HTS), puces à ADN ou PCR quantitatives (<http://www.genomique.info/>).

I. Les cellules de l'épithélium respiratoire

L'appareil respiratoire est recouvert par un épithélium cylindrique pseudo-stratifié. Cet épithélium est principalement constitué de cellules basales, de cellules multiciliées et de cellules sécrétrices de mucus, nonobstant quelques populations cellulaires accessoires (telles que les cellules neuro-endocrines, les cellules immunitaires résidentes, les ionocytes et les cellules tuft).

a. Les cellules basales

Les cellules basales sont en contact direct avec la lame basale (Figure 1). Leur quantité diminue en descendant le long de l'arbre bronchique et passe de 30% au niveau proximal à environ 6% de la population épithéliale dans les bronchioles (Boers, Ambergen, & Thunnissen, 1998). Ce sont des cellules de petite taille possédant un rapport noyau/cytoplasme très élevé. Ces cellules expriment des marqueurs intracellulaires qui leur sont propres et qui permettent de les identifier facilement. Ces marqueurs sont des cytokératines, comme la cytokératine-5 (KRT5) ainsi que des facteurs de transcription, comme TP63 (« Tumor Protein P63 »). Les cellules basales ont deux rôles essentiels au sein de l'épithélium. Ces cellules possèdent un rôle structural permettant la cohésion de l'épithélium sur la lame basale. Par ailleurs, ces cellules forment des liaisons avec les cellules voisines à l'aide de desmosomes. Dès lors, les cellules basales assurent le maintien de l'épithélium sur la lame basale (Evans, Cox, Shami, Wilson, & Plopper, 1989). Elles jouent également un rôle de cellules souches ou cellules progénitrices, capables de proliférer mais également de se différencier, notamment lors des mécanismes de réparation suite à une lésion, afin de reformer un épithélium fonctionnel. La capacité à former d'autres types cellulaires différenciés a été mise en évidence grâce à des souris transgéniques exprimant un gène rapporteur sous le contrôle d'un promoteur spécifique de cellules basales. En effet, Hong et collaborateurs ont pu montrer qu'après lésion de l'épithélium *in vivo*, les cellules multiciliées et sécrétrices de mucus néoformées dérivent toutes deux des cellules basales (Hong, Reynolds, Watkins, Fuchs, & Stripp, 2004).

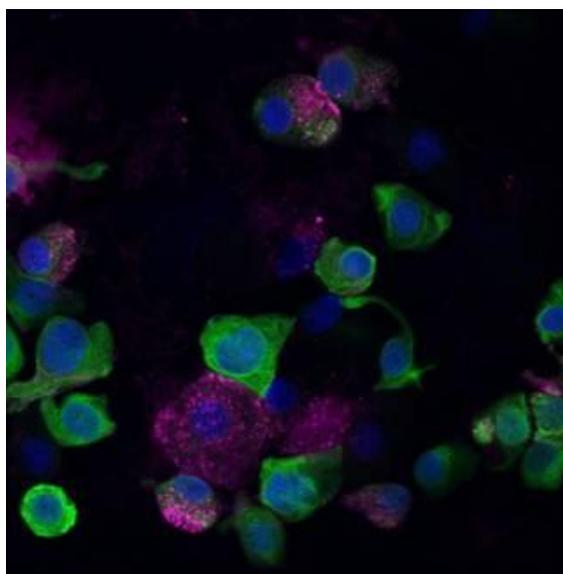


Figure 1: Image de cellules basales au microscope confocal

Les cellules ont été marquées avec du DAPI pour marquer le noyau (bleu) des cellules. Deux fluorochromes ont également été utilisés ciblant KRT5 (vert) qui se fixe sur la kératine des cellules basales et PLUNC (violet) qui se fixe sur les cellules basales prédifférenciées. L'image est grossie x60

b. Les cellules sécrétrices de mucus

Les cellules sécrétrices de mucus sont des cellules cylindriques également appelées cellules en gobelet (« goblet cell » en anglais) (Figure 2). Leur noyau est disposé dans la partie basale de la cellule. A l'inverse, toute la partie apicale est gorgée de vésicules de sécrétion, constituées principalement de glycoprotéines riches en acide sialique, les mucines (composant majeur du mucus) (Rogers, 1994). Ces cellules sont également capables de produire des molécules à forte activité anti-bactérienne (IgA sécrétoires, lactoferrine, lysozyme défensines) (Bals, Weiner, & Wilson, 1999).

Ces cellules sont disséminées au milieu de cellules multiciliées où elles représentent 15% des cellules cylindriques au sein de l'épithélium. Bien que les cellules sécrétrices de mucus soient largement minoritaires au sein de l'épithélium, leur proportion peut augmenter drastiquement en fonction des conditions de stress ou pathologiques (Park et al., 2006).

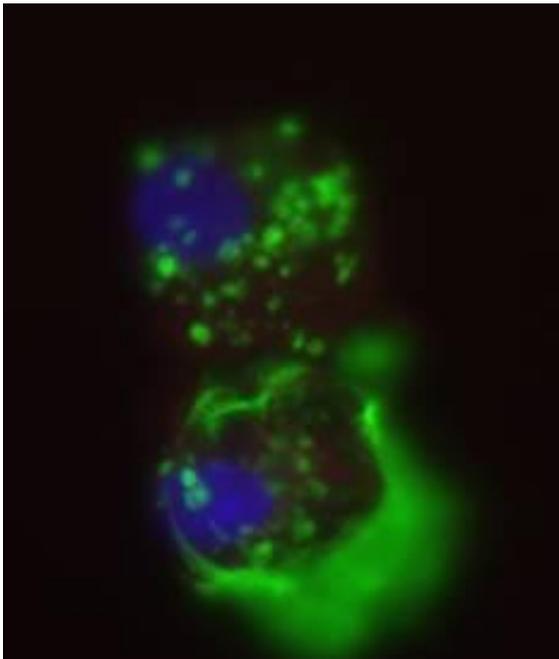


Figure 2 : Image de cellules à mucus au microscope confocal

Les cellules ont été marquées avec du DAPI pour marquer le noyau (bleu) des cellules. Un fluorochrome a été utilisé ciblant MUC5AC (vert) qui se fixe sur la mucine des cellules à mucus. L'image est grossie x60

c. Les cellules multiciliées

Les cellules multiciliées (MCC) de l'épithélium respiratoire sont de forme cylindrique ou pyramidale (Breeze & Wheeldon, 1977). Le battement coordonné de centaines de cils vibratiles à leur surface permet d'assurer la fonction de clairance mucociliaire. Chaque cellule multiciliée arbore à son pôle apical 200 à 300 cils motiles (Figure 3) (Rhodin, 1966).

Elles représentent au niveau de la trachée ou des bronches plus de 50% de la population cellulaire pour ne plus représenter que 15% des cellules épithéliales dans les voies respiratoires plus distales (Serafini & Michaelson, 1977). Les cellules multiciliées sont considérées comme différenciées de façon terminale, étant incapables de se diviser à nouveau pour proliférer. Cette règle est susceptible d’être remise en cause par la découverte du mécanisme de transdifférenciation des cellules multiciliées en cellules sécrétrices de mucus (Park et al., 2006).

Le noyau est situé à la base de la cellule alors que les mitochondries sont abondamment retrouvées au niveau apical de la cellule, fournissant ainsi de façon efficace l’énergie indispensable au battement ciliaire (Hansell & Moretti, 1969) (Blanquart, Giuliani, Jeulint, Guennou, & Marano, 1995).



Figure 3 : Image d’une cellule multiciliée au microscope confocal

Les cellules ont été marquées avec du DAPI pour marquer le noyau (bleu) des cellules. Un fluorochrome a également été utilisé ciblant AKAP14 (vert) qui se fixe sur les cils de la cellule multiciliée. L’image est grossie x60

II. La mitochondrie

a. Généralité et structure

Le terme « mitochondrie » est composé de deux mots, d’étymologie grecque, « mitos » signifiant « filament » et « chondros » signifiant « grain ». Il a été proposé en 1898 par le médecin allemand Carl Benda, pour décrire un organe de type filamenteux. La mitochondrie a été observée pour la première fois dans les années 1850, par le physiologiste Albert Von

Kölliker. Entre 1850 et 1880, plusieurs scientifiques ont observé la présence des mitochondries, de manière indépendante, au sein de différents types cellulaires.

Les mitochondries ont une structure en forme de bâtonnet ou de sphère de 0,5 à 1 μm de diamètre. Leur nombre est variable selon l'activité métabolique. Leur origine semble établie : elles dériveraient de l'endosymbiose d'une bactérie de la classe des α -protéobactéries (Lazcano & Peretó, 2017).

La mitochondrie est présente à des centaines de copies dans tous les types cellulaires, excepté dans les hématies. Son abondance varie en fonction du type cellulaire et des besoins énergétiques de la cellule. Par exemple, dans le foie et dans le muscle cardiaque, les mitochondries représentent respectivement 20% et 40% du volume cytosolique. Les mitochondries sont en mouvement continu dans la cellule grâce aux interactions avec le cytosquelette. Une mitochondrie est séparée du compartiment cytoplasmique par deux membranes distinctes qui délimitent cinq compartiments distincts : la matrice mitochondriale, l'espace inter-membranaire (EIM) qui sépare la membrane externe (MEM) de la membrane interne mitochondriale (MIM) et les crêtes qui représentent le compartiment généré par l'invagination de la membrane interne (Friedman & Nunnari, 2014) (Figure 4).

La matrice mitochondriale d'aspect granuleuse contient des mitoribosomes, de l'ADN circulaire, de l'ARNm et de l'ARNt. Elle est le siège de réactions métaboliques comme le cycle de Krebs qui permet la production de l'ATP cellulaire. Les mitochondries produisent près de 90% de l'énergie de la cellule.

L'espace inter-membranaire est un espace dense d'une épaisseur de 4 à 7 nm contenant des protons H^+ , dont les gradients transmembranaires jouent un rôle déterminant dans la phosphorylation oxydative, et des molécules de cytochrome c (cyt c) (impliqués dans les mécanismes d'apoptose).

La membrane externe (MEM) est une bicouche lipidique de 5 à 7 nm d'épaisseur, avec une composition proche de celle de la membrane plasmique. Elle contient 40% à 50% de lipides et 50% à 60% de protéines incluant les porines, protéines membranaires formant des canaux ioniques, qui permettent la libre diffusion de petits métabolites (ions et molécules de masse moléculaire inférieure à 10kDa) (Logan, 2006). La MEM sépare l'EIM du cytoplasme. Au départ considéré comme une simple enceinte de confinement de la mitochondrie, de récentes

études lui ont attribué des caractéristiques importantes liées à la physiologie et à la signalisation cellulaire, grâce à son dialogue avec les « mitochondria-associated ER-membranes » (MAMs) (Chami et al., 2008).

La membrane interne mitochondriale (MIM) est une bicouche lipidique de 5 à 6 nm avec une organisation très différente de celle de la membrane externe. Elle est composée de 80% de protéines et 20% de lipides. Elle est riche en cardiolipides, en transporteurs et complexes enzymatiques de la chaîne respiratoire, responsables de la production d'énergie. Cette membrane a une structure similaire à la membrane plasmique bactérienne. Contrairement à la MEM, elle ne possède pas de porines. La diffusion à travers la MIM ne peut se réaliser que via des transporteurs particuliers appelés translocases.

La MIM forme de nombreuses invaginations qui se replient dans des poches appelées crêtes. La conséquence majeure de ces invaginations est l'augmentation de la surface du complexe respiratoire afin d'amplifier de manière considérable, la capacité respiratoire de la mitochondrie et optimiser la production d'ATP (Patten et al., 2014). La taille et la forme des crêtes peuvent changer en fonction du signal intracellulaire. Ainsi, un remodelage des crêtes peut s'opérer suite à l'activation de la mort cellulaire programmée. La forme des crêtes demeure importante dans la formation et la stabilité de la chaîne du complexe respiratoire, en une structure quaternaire appelée « super complexe ». L'efficacité et le bon fonctionnement de la respiration mitochondriale, en réponse aux changements du métabolisme ou du stress cellulaire, dépendent directement de la forme des crêtes. De manière intéressante, les protéines qui contrôlent l'architecture des crêtes peuvent être localisées au niveau de la MIM ou des MAMs. Cette coopération inter-organelles permettrait de répondre aux changements du métabolisme mais reste encore peu caractérisée. L'interaction des compartiments intra- et inter-mitochondriaux demeure indispensable dans la modulation de l'activité mitochondriale.

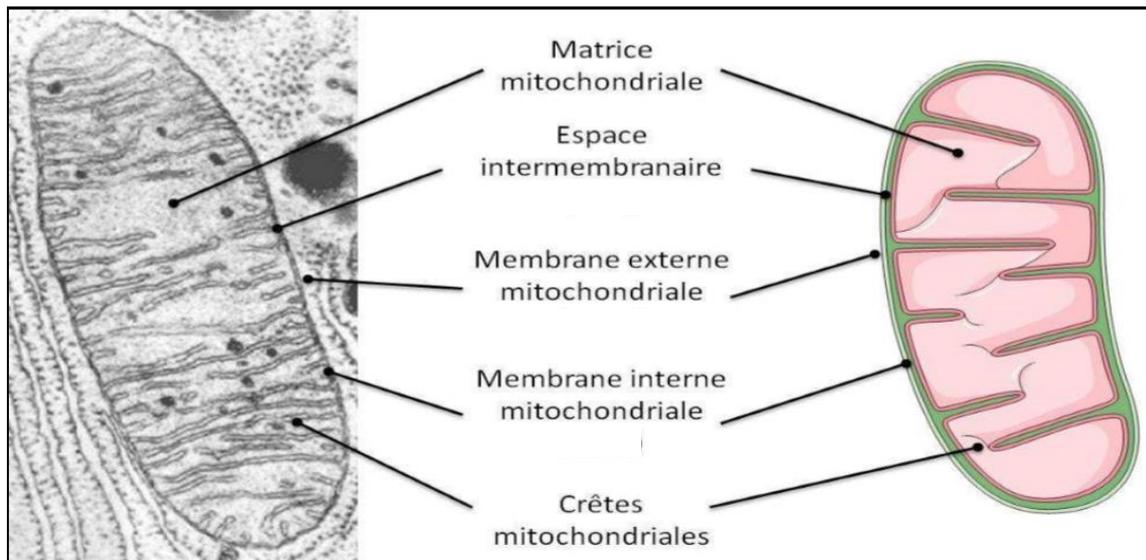


Figure 4 : Structure mitochondriale

La mitochondrie est composée d'une double membrane délimitant un espace intermembranaire et une matrice mitochondriale. La membrane interne forme de nombreux replis appelés crêtes mitochondriales (d'après Goiran 2016).

b. ADN mitochondrial

Les mitochondries possèdent leur propre information génétique sous forme d'ADN mitochondrial (ADNmt) (Figure 5). Chez l'Homme, le génome mitochondrial est une molécule circulaire double brin de 16569 paires de bases non associée à des histones et ne possédant aucun intron contrairement à l'ADN nucléaire (ADNn). Chaque cellule humaine possède plusieurs centaines à plusieurs milliers de mitochondries. Chaque mitochondrie contenant 2 à 10 copies d'ADNmt, la quantité totale d'ADN mitochondrial peut représenter jusqu'à 1% de la quantité totale d'ADN cellulaire, en dépit de sa faible taille ($1/200000^{\text{ième}}$ du génome nucléaire). Les deux brins du génome mitochondrial diffèrent par un brin lourd (H) riche en bases puriques (G et A) et un brin léger (L) riche en bases pyrimidiques (C et T). L'ADN est transcrit sur toute sa longueur à l'exception d'une région appelée boucle D (Displacement) contenant l'origine de réplication du brin lourd ainsi que les promoteurs de transcription des brins lourd et léger, et d'une région plus courte, présentant l'origine de réplication du brin léger. Le brin lourd code 14 ARN de transfert, les 2 ARN ribosomiaux et 12 protéines. Le brin léger code 1 protéine et les 8 ARN de transfert restant (Taanman, 1999).

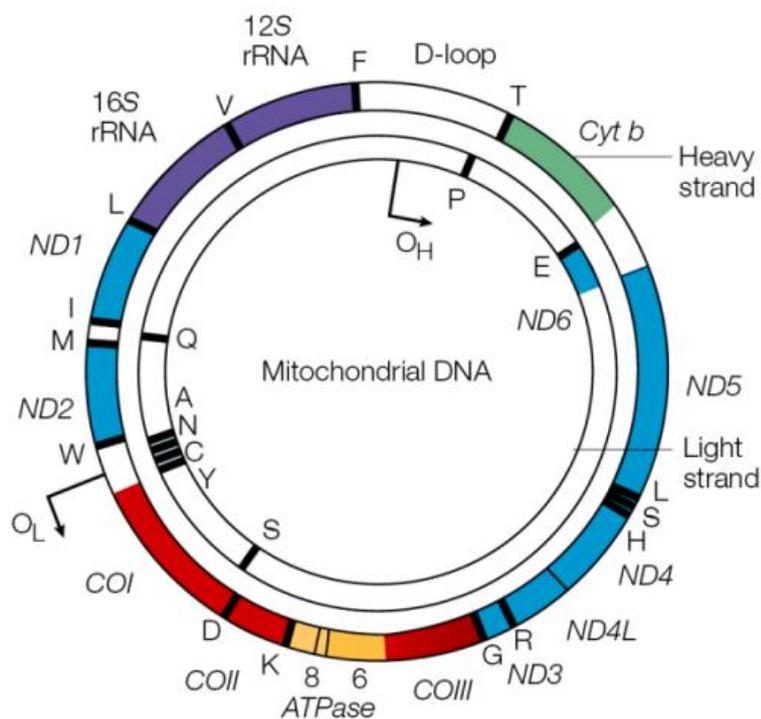


Figure 5 : ADN mitochondrial
 L'ADN mitochondrial possède 22 gènes d'ARN de transfert nécessaires à l'expression de l'ADNmt et 13 gènes codant des protéines de la chaîne respiratoire. La boucle D est une région de contrôle de la réplication et de la transcription qui comporte deux promoteurs de transcription et une origine de réplication (OH) (d'après Taylor and Turnbull 2005).

L'ADNmt présente une variabilité de séquence plus importante (environ dix fois) que l'ADN nucléaire. Une telle mutabilité de l'ADNmt s'explique par l'action des radicaux libres oxygénés générés en parallèle de la respiration sur un ADN non protégé par des histones, et par le faible nombre de mécanismes réparateurs efficaces (Martin, 2011). La réplication de l'ADN mitochondrial est dépendante de la seule ADN polymérase présente dans la mitochondrie : la polymérase γ ou polymérase G. La disponibilité et l'équilibre de dNTPs dans la mitochondrie sont les principaux déterminants qui définissent le taux de fidélité de la réplication de l'ADN mitochondrial (Gandhi & Samuels, 2011).

c. ADN mitochondrial et hétéroplasmie

Les cellules humaines possèdent plusieurs centaines à plusieurs milliers de mitochondries et chaque mitochondrie contient 2 à 10 copies d'ADNmt. Lors des divisions cellulaires, la répartition du capital mitochondrial d'une cellule mère s'effectue de manière aléatoire entre les cellules filles. Au sein d'une même cellule, plusieurs ADNmt différents peuvent coexister. Cette propriété connue sous le nom d'hétéroplasmie de l'ADNmt joue un rôle important dans l'expression des mutations mitochondriales (Wallace & Chalkia, 2013). L'expression phénotypique au niveau cellulaire va dépendre en grande partie des proportions d'ADN normal et d'ADN muté. Une augmentation du pourcentage d'ADN muté, peut conduire à une

diminution des capacités de la chaîne respiratoire à générer de l'ATP induisant une augmentation du stress oxydatif mitochondrial pouvant entraîner secondairement des phénomènes d'apoptose. La répartition aléatoire des mitochondries lors de la division cellulaire permet d'obtenir différents profils cellulaires à partir d'une cellule hétéroplasmique, via la répartition aléatoire de nouvelles mutations mitochondriales. On peut anticiper la présence de plusieurs populations de cellules apparemment identiques, mais différant en réalité par leurs populations mitochondriales : les cellules homoplasmiques ne présenteront qu'un ADNmt sauvage (non muté) et elles présenteront un fonctionnement normal, des cellules homoplasmiques ne présentant que des mitochondries portant une mutation particulière présenteront des déficits fonctionnels. Enfin, des cellules hétéroplasmiques possédant un nombre réduit de mitochondries portant cette même mutation présenteront un phénotype intermédiaire dépendant du taux de présence de la mutation (Figure 6) (Taylor & Turnbull, 2005).

Plus simplement, l'homoplasmie signifie que toutes les copies du génome mitochondrial sont identiques, l'hétéroplasmie qu'il existe un mélange de deux ou plusieurs génotypes mitochondriaux. C'est dans ce cadre que des mutations d'ADNmt peuvent mener à une maladie. Certaines mutations affectent toutes les copies du génome mitochondrial (mutation homoplasmique), tandis que d'autres ne sont présents que dans certaines copies du génome mitochondrial (mutation hétéroplasmique).

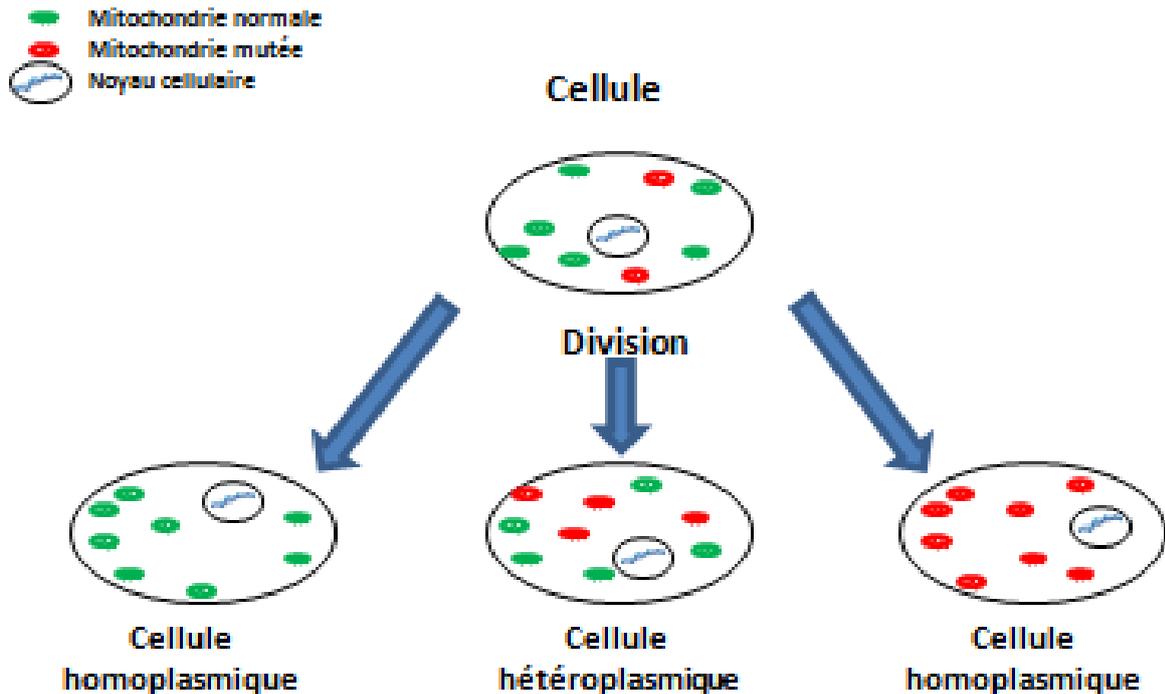


Figure 6 : Répartition aléatoire des mitochondries

La ségrégation mitochondriale lors de la division cellulaire peut donner trois types de cellules : une cellule homoplasmique normale, une cellule homoplasmique mutée et une cellule hétéroplasmique (d'après Sévin and Girard 1998).

Le concept d'homoplasmie est plus apparent que réel. Dans la plupart des individus, il n'y a aucune preuve d'hétéroplasmie, alors même que l'ADNmt subit constamment des mutations, en raison de la fidélité enzymatique de l'ADN polymérase, ou d'un environnement cytoplasmique génotoxique. Il peut en résulter une expansion clonale des mitochondries mutées, ou à l'inverse une perte de ces mutations ponctuelles ou de délétions. Parce que ces mutations acquises se produisent au hasard, toutes les mutations acquises sont présentes à un niveau bas et ne peuvent donc pas être détectées dans un homogénat tissulaire ou un échantillon de sang.

Les maladies mitochondriales ont été associées à un large spectre de symptômes appartenant aux maladies neurodégénératives incluant désordres du mouvement, cécité, surdité, épilepsie, accidents vasculaires, diabète et autres pathologies (Chaussonot & Paquis-Flucklinger, 2014). De plus, un dysfonctionnement mitochondrial représente un élément clé dans les phénomènes dus au vieillissement ou pour les maladies liées à l'âge. Chez l'homme, les maladies mitochondriales représentent le groupe le plus important des maladies

métaboliques. Plus de 200 mutations touchant les gènes de l'ADNmt ont été rapportées à ce jour (<http://www.mitomap.org/>).

d. Héritage mitochondrial

La transmission de l'ADN mitochondrial s'effectue uniquement par la lignée maternelle qui est résumée dans la théorie de l'Eve mitochondriale s'appuyant sur le fait que la transmission du patrimoine génétique mitochondrial ne s'effectue qu'à travers la mère (Giles, Blanc, Cann, & Wallace, 1980). L'Eve mitochondriale est le nom donné à une femme hypothétique considérée comme l'ancêtre commun de l'humanité par la lignée mitochondriale (Ayala, 1995).

Lors de la fécondation, le capital mitochondrial de l'ovocyte provient pour l'essentiel du gamète femelle, dont le cytoplasme plus important comporte bien davantage de mitochondries que le gamète mâle. De plus, lors de la fécondation, le spermatozoïde recrute en quelques minutes des autophagosomes qui dégradent les mitochondries présentes. La dégradation des mitochondries paternelles et de leur ADN mitochondrial nécessite une autophagie LC3-dépendante. De plus, des chercheurs ont identifié un gène, appelé cps-6, qui semble initier le processus de détérioration dans les mitochondries paternelles. Ils ont constaté que la suppression de cps-6 augmente le nombre des mitochondries paternelles dans l'embryon. Cela entraîne également davantage de décès embryonnaire (Zhou et al., 2016).

Le taux d'évolution de la séquence d'ADNmt permet dans un tel contexte le suivi de l'origine et du rayonnement des populations du point de vue de la propagation de l'ADN mitochondrial féminin. En conséquence, les polymorphismes d'ADNmt permettent le suivi des migrations des Hommes vers l'ouest de l'Afrique ou sur les différents continents. Ainsi, lorsque les Hommes, et dans ce contexte plus particulièrement les femmes, se sont déplacées vers les différents continents il y a environ 150.000 ans, des polymorphismes de séquence d'ADNmt ont été sélectionnés, dont la fréquence peut être plus élevée dans certaines populations (Stewart & Chinnery, 2015). Ces polymorphismes définissent de la sorte des haplogroupes spécifiques d'ADNmt propres à ces populations. Pour déterminer l'haplogroupe, on analyse des SNPs. Les SNPs (Single Nucleotide Polymorphism) sont des variations de couples de bases isolées dans une lignée d'ADN. Par exemple, une étude internationale menée par le professeur Jin li a révélé que le peuple chinois ne provient pas de

l'Homme de Pékin dans le nord de la Chine, mais de premiers humains en Afrique de l'Est qui se sont déplacés à travers l'Asie du Sud jusqu'à la Chine il y a environ 100.000 ans.

À l'ère de la PCR et du séquençage Sanger, le taux de mutation élevé a rendu possible la découverte de variations de séquence de l'ADNmt. Le manque de recombinaison a permis de regrouper les données des régions de l'ADNmt sous la forme d'un arbre phylogénétique (Annexe 1). La caractérisation des haplogroupes est de plus en plus fine, à mesure que de nouvelles données sont disponibles. En conséquence, les haplogroupes les plus courants ont reçu des étiquettes alphabétiques qui correspondent aux nœuds de l'arbre phylogénétique de l'ADNmt.

La nomenclature des haplogroupes d'ADNmt a été introduite au milieu des années 1990 avec des étiquettes A-G attribuées à la variation observée dans les lignées asiatiques et américaines (Torroni et al., 1993), H-K en Europe (Torroni et al., 1994), alors qu'une seule lettre, L, a été assignée pour décrire le plus haut niveau de variation observé en Afrique dans une étude utilisant un groupe asiatique (Chen et al., 1995). La nomenclature d'ADNmt actuellement utilisée (<http://www.phylotree.org/>) a une structure de branchement solide qui a été déterminée dans le cadre d'analyses détaillées de tous les génomes d'ADNmt (van Oven & Kayser, 2009). Ces détails topologiques de la phylogénie d'ADNmt ont été révélés étape par étape au cours des deux dernières décennies grâce aux contributions de nombreux groupes dans la couverture de données de populations de plus en plus nombreuses à travers le monde et grâce aux progrès de la technologie (Figure 7). Notamment le développement des techniques de séquençage de nouvelle génération offrant un excellent outil pour la détection de SNP dans le génome mitochondrial.

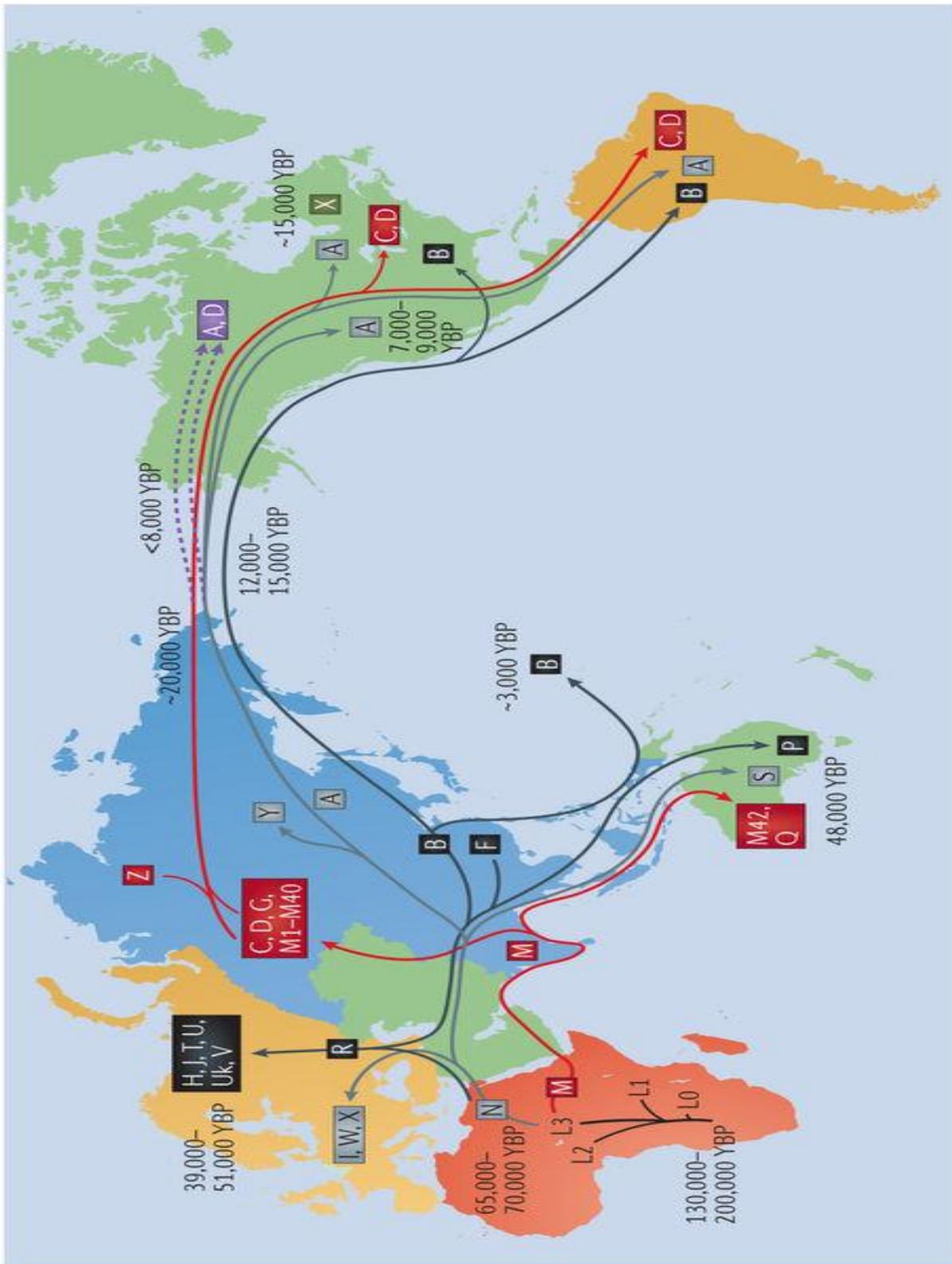


Figure 7 : Migrations de la population humaine et les principaux haplogroupes d'ADNmt.

Les mutations acquises au fil du temps ont divisé la population en plusieurs haplogroupes. Les principaux haplogroupes sont apparus entre 40 000 et 150 000 ans avant notre ère (YBP) et ont défini différentes populations humaines alors qu'elles migraient hors d'Afrique et peuplaient le globe (D'après Stewart & Chinnery, 2015).

III. Séquençage et nouvelle technologie

Depuis la découverte de la structure de l'ADN, de grands progrès ont été réalisés dans la compréhension de la complexité et de la diversité des génomes dans la santé et la maladie. Cependant, le débit limité et les coûts élevés du séquençage sont demeurés des barrières majeures. Au cours de la dernière décennie, les technologies de séquençage à haut débit (NGS) ont continué à évoluer en augmentant la capacité de séquençage d'un facteur de 100 à 1 000 - et ont incorporé des innovations révolutionnaires pour s'attaquer à la complexité des génomes (Zhang, Chiodini, Badr, & Zhang, 2011). Ces progrès ont permis de produire des longueurs de lecture aussi longues que certains génomes entiers, ils ont amené le coût du séquençage d'un génome humain en deçà des \$1 000 US (tel que rapporté par Veritas Genomics) avec un objectif à moins de \$100 d'ici quelques années. Cette situation positionne progressivement le séquençage comme un outil pertinent dans le domaine de la clinique. Les plateformes NGS fournissent de grandes quantités de données, avec des taux d'erreurs associés de l'ordre de 1% légèrement plus élevés et des longueurs de lecture généralement plus courtes (35-700 pb pour les approches de lecture courte) qu'avec des plateformes de séquençage Sanger, qui sont compensés par l'obtention d'un taux de couverture bien plus important (chaque nucléotide pouvant être lu un très grand nombre de fois).

a. Les différents types de séquençage

En permettant des approches de l'analyse des acides nucléiques qui soient à la fois qualitatives et quantitatives, le séquençage à haut débit permet de déterminer la succession de nucléotides d'une séquence d'acides nucléiques (analyse qualitative) mais aussi de dénombrer le nombre de ces séquences dans un échantillon donné (analyse quantitative). Il existe plusieurs types de séquenceurs à haut débit, utilisant chacun une technologie différente permettant la lecture des bases (Metzker, 2005). Le séquençage par terminaison réversible cyclique (CRT) développé par la société Illumina introduit des nucléotides fluorescents à chaque cycle d'extension d'une chaîne d'ADN par une polymérase (Figure 8). Après incorporation d'un seul dNTP à chaque brin complémentaire, les dNTPs non liés sont supprimés et la surface est imagée pour identifier quel dNTP a été incorporé à chaque grappe. Le fluorochrome peut alors être retiré et un nouveau cycle d'extension peut commencer (Goodwin, McPherson, & Richard McCombie, 2016). Le système Illumina qui utilise cette chimie occupe actuellement la plus grande part du marché des instruments de séquençage de

nouvelle génération (90% des lectures générées aujourd'hui). L'identification des dNTP est obtenue grâce à la microscopie de fluorescence à réflexion interne totale (TIRF) utilisant deux ou quatre canaux laser (Ju et al., 2006). Dans les premières plateformes Illumina, chaque dNTP était lié à un seul fluorochrome spécifique à un type de base et nécessitait quatre canaux d'imagerie différents. Les systèmes plus récents (NextSeq et Micro-Seq) utilisent un système à deux fluorochromes (Metzker, 2010).

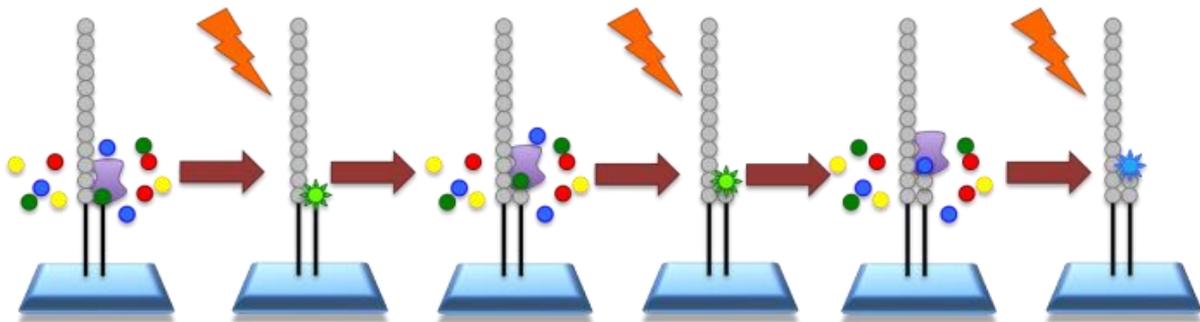


Figure 8 : Principe de séquençage par terminaison réversible cyclique

Le processus permet d'identifier simultanément les bases d'ADN lorsqu'elles sont incorporées dans la chaîne d'acide nucléique. Chaque base émet un signal de fluorescence unique lorsqu'elle est ajoutée au brin en cours de synthèse, ceci est utilisé pour déterminer la séquence d'ADN (<https://emea.illumina.com/>).

La seconde méthode de séquençage par synthèse est par addition simple de nucléotide (SNA, Figure 9). Contrairement à la CRT, les approches SNA s'appuient sur un seul signal pour marquer l'incorporation d'un dNTP sur un brin. En conséquence, chacun des quatre nucléotides doit être ajouté itérativement à une réaction de séquençage pour s'assurer qu'un et un seul dNTP est responsable du signal. En outre, cela ne nécessite pas que les dNTP soient bloqués, car l'absence du nucléotide suivant dans la réaction de séquençage empêche l'allongement. Ce type d'approche rencontre quelques limitations dans les régions homopolymères où des dNTP identiques doivent être ajoutés, car l'identification de séquence reposant sur une augmentation proportionnelle du signal, la détermination exacte du nombre de dNTPs incorporés peut alors être délicate. La précision est alors limitée dans les longues zones d'homopolymères. L'Ion Torrent de la société ThermoFisher utilise une cascade enzymatique pour générer un signal, le séquenceur détectant alors les ions H^+ qui sont relâchés lorsque chaque dNTP est incorporé. Le changement de pH détecté par un capteur est alors proportionnel au nombre de nucléotides détectés (Rothberg et al., 2011).

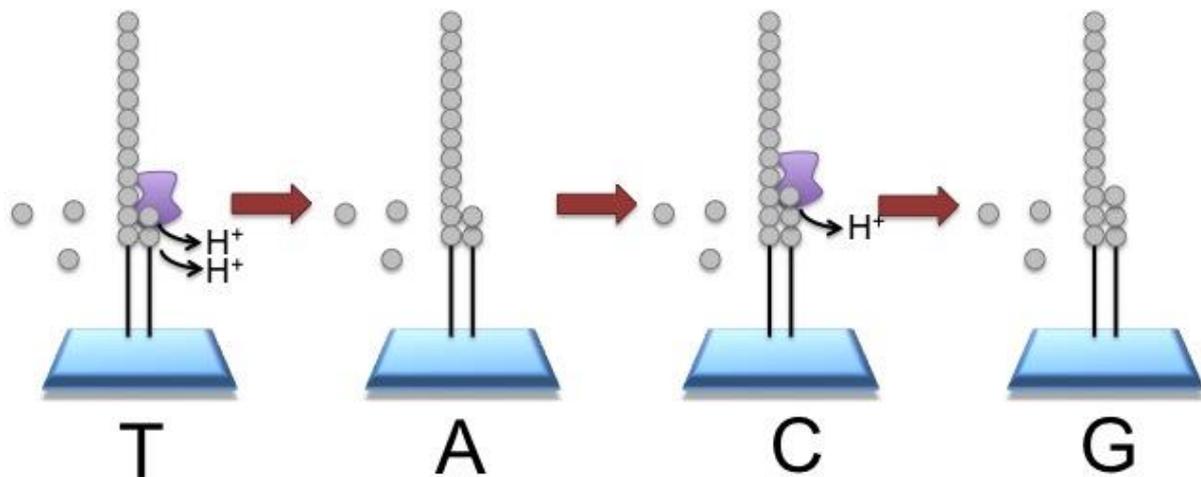


Figure 9 : Principe de séquençage par addition simple de nucléotide

Chaque base est ajoutée individuellement. Lorsqu'une base est appariée à la séquence, un H⁺ est relargué dans la solution et l'appareil détecte la variation de pH dans le milieu et indique la base intégrée (<https://www.thermofisher.com/>).

De multiples applications ont été développées à partir de ces nouvelles techniques : outre le DNA-seq qui permet le séquençage de génome *de novo*, le re-séquençage total ou partiel, le Chip-seq permet d'identifier des sites de liaison de protéines sur le génome, le RNA-seq de séquencer le transcriptome complet, le smallRNA-seq de séquencer de petits ARN comme les microARN (Ozsolak & Milos, 2011). Des améliorations continues depuis plus de 40 ans (1977 étant la date de publication du séquençage de l'ADN par Sanger, Maxam et Gilbert) permettent de réaliser désormais de telles expériences à des niveaux de sensibilité de plus en plus importants, avec l'objectif d'atteindre une sensibilité à l'échelle de la cellule unique « single-cell » (qui contient 10-20 pg d'acides nucléiques). Pour également améliorer le séquençage, la technologie « long-read » commence à faire son apparition permettant d'obtenir des informations de séquences sur de grandes longueurs (une à plusieurs centaines de Kb). En 2014, le MinION d'Oxford Nanopore Technologies est devenu disponible et permet le passage de l'ADN à travers un pore protéique et séquencer les bases nucléotidiques lorsque le courant passe à travers le pore (Clarke et al., 2009).

b. Cellule unique

La technologie de séquençage d'ADN a non seulement augmenté rapidité du débit, mais les progrès dans la préparation d'échantillons ont également diminué en termes de quantité d'ADN requise pour l'analyse, au point où il est désormais possible d'analyser le contenu en

ADN des cellules individuelles. De très nombreux travaux sont habituellement effectués dans un contexte biologique bien précis. Cependant, la diversité est bien supérieure à celle que nous pouvons mesurer dans un échantillon de tissu ou d'un prélèvement. Il existe en effet 10^{13} cellules dans le corps humain, qui correspondent à un très grand nombre de types cellulaires différents.

La génomique sur cellule unique vise à fournir de nouvelles perspectives à notre compréhension de la génétique en amenant l'étude des génomes au niveau d'une seule cellule. Ces outils ouvrent de nouvelles frontières en disséquant les contributions des cellules individuelles à la biologie des écosystèmes et des organismes. Par exemple, il est maintenant possible d'utiliser la génomique uni-cellulaire pour évaluer les rôles du mosaïcisme génétique dans la physiologie et déterminer la contribution de l'hétérogénéité génétique intra-tumorale dans des domaines comme le développement ou le traitement des cancers. Cette approche méthodologique repose sur la capacité d'étudier les molécules d'acides nucléiques à partir de cellules individuellement isolées (Gawad, Koh, & Quake, 2016).

Les approches sur cellule unique offrent des perspectives de résolution supérieure du contenu génomique des échantillons en réduisant la complexité du signal génomique à travers la séparation physique des cellules. Le séquençage d'une seule cellule issue d'un organisme multicellulaire peut révéler des variations génétiques rares et fournir des informations sur l'apparition de mutations et l'histoire naturelle (Shapiro, Biezuner, & Linnarsson, 2013) (Proserpio & Lönnberg, 2016).

L'acquisition de données de séquençage sur cellule unique de haute qualité comporte quatre défis techniques principaux: l'isolement physique efficace des cellules individuelles; l'amplification des acides nucléiques de cette cellule pour augmenter le signal; le séquençage de ces acides nucléiques avec suffisamment de profondeur pour identifier une variation; la capacité de discriminer parmi les données pertinentes au milieu des biais et des erreurs introduits au cours des différentes étapes. Pour maximiser la qualité des données obtenue d'une seule cellule et s'assurer que le signal est bien séparable du bruit technique, un examen minutieux de chacune de ces variables est indispensable lors de la conception des études (Wang & Navin, 2015).

Quelques sociétés ont mis en place des moyens permettant l'isolation de chaque cellule, et mon laboratoire s'est plus particulièrement intéressé aux technologies développées par la société Fluidigm et la société 10X Genomics. La société Fluidigm a mis au point une technique microfluidique pour séparer les cellules (Figure 10). Le principe technique est basé sur un système pneumatique permettant la capture cellulaire sur des plaques microfluidiques comportant 96 ou 800 « pièges » à cellules. Une fois la phase de capture accomplie, les divers pièges sont individualisés et chacune des cellules piégées est « poussée » dans une succession de microréacteurs isolés enchainant phases de lyse de la cellule puis la préparation des acides nucléiques visés amplifiés pour des applications en génomique (Kolodziejczyk, Kim, Svensson, Marioni, & Teichmann, 2015).

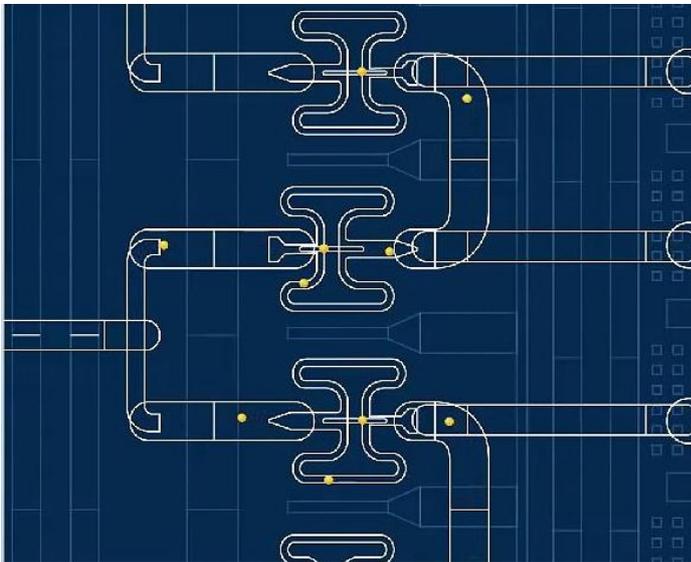


Figure 10 : Principe du C1 de Fluidigm

Grâce à un système de microfluidique permettant de faire circuler les cellules dans le réseau de canaux afin de les capturer individuellement (<https://www.fluidigm.com/>).

La société 10X Genomics utilise une technologie se basant sur la production d'émulsions « eau dans l'huile », sur le principe du « Drop-seq ». La technique consiste à utiliser de minuscules billes possédant à leur surface des oligonucléotides recouverts de séquences courtes d'oligonucléotides spécifiques de chaque bille, qui fonctionnent comme de véritables codes-barres nucléotidiques. L'appareil va former des gouttelettes contenant une et une seule cellule, une seule bille recouverte d'un code-barre unique, ainsi que l'ensemble des réactifs nécessaires à la lyse des cellules et à la transcription inverse des ARN (Figure 11). Dans cette émulsion, la lyse et la transcription inverse des ARN polyadénylés propres à une et une seule cellule sont donc effectuées à l'intérieur d'une et une seule gouttelette (Zheng et al., 2017).

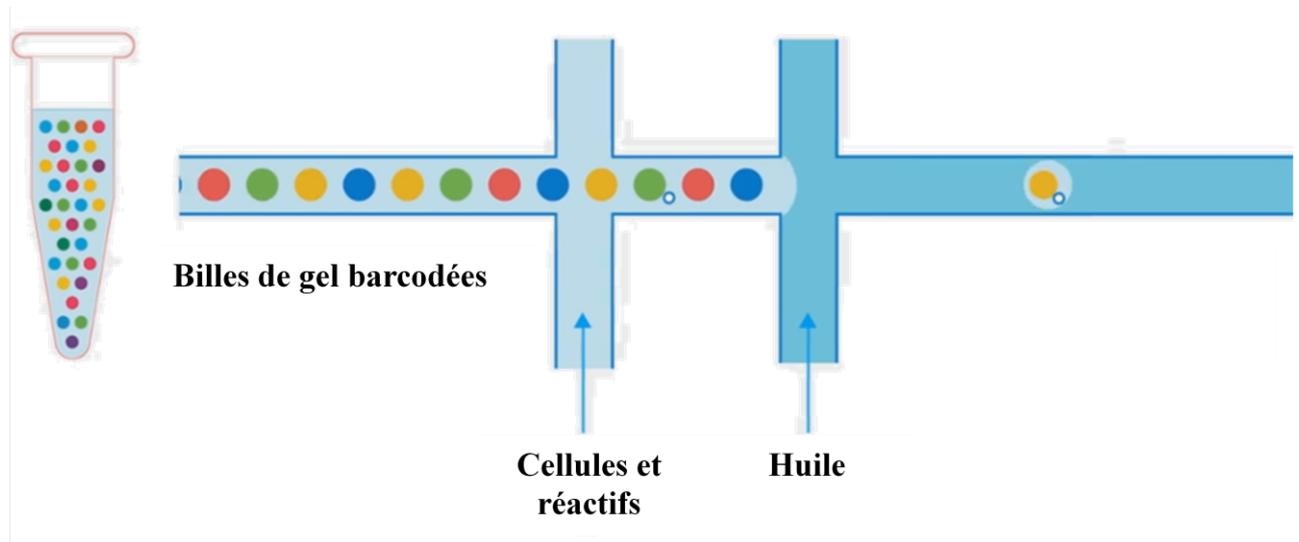


Figure 11 : Principe de la technique du Chromium de chez 10X Genomics

Une gouttelette va être formée grâce à l'huile encapsulant une cellule, les réactifs et une bille possédant un code-barre (<https://www.10xgenomics.com/>).

Objectif

Les mitochondries sont capitales au fonctionnement cellulaire, elles apportent à la cellule 90% de son énergie. Le nombre de mitochondries présentes au sein de la cellule est dépendant de la fonction ainsi que l'activité de la cellule. Par exemple, les cellules cardiaques possèdent un nombre de mitochondrie équivalent à 30% de leurs volumes cellulaires (Piquereau et al., 2013). L'ADN mitochondrial est aussi présent en plusieurs exemplaires au sein d'une mitochondrie. Lors de la réplication de cet ADN mitochondrial, des mutations peuvent être introduites. Au sein d'une même cellule, plusieurs molécules d'ADNmt portant des séquences différentes peuvent ainsi coexister (hétéroplasmie). Cette hétéroplasmie de l'ADNmt peut être à l'origine d'une pathologie lors le taux de l'ADNmt muté dépasse un certain seuil. La présence de nombreuses copies d'ADNmt par mitochondrie et par cellule, rend possible une variabilité entre tissus d'un même individu, voire même au sein de chacune des cellules de ce tissu (Perales-Clemente et al., 2016).

Les technologies d'aujourd'hui nous permettant de séquencer l'ADN mitochondrial, j'ai eu l'opportunité de travailler sur une plateforme de séquençage affiliée à une équipe de recherche s'intéressant aux voies aériennes, où j'ai eu à comparer les séquences mitochondriales de cellules d'origines embryonnaires différentes mais issues d'un même individu. Certaines mutations ou SNP (Single nucléotide Polymorphism) permettent d'établir une phylogénie mitochondriale au sein de l'espèce humaine, offrant un moyen d'étudier certaines propriétés tissulaires et/ou cellulaire propre à un individu.

Dans le cadre du Diplôme de l'EPHE, j'ai été en charge de mettre en place, au sein de la plateforme, une analyse par séquençage de l'ADN mitochondrial à l'échelle de la cellule unique que j'ai appliqué à l'analyse de différents résultats en cellule unique à partir de cellules isolées de cornets nasaux issus de différents patients. L'idée initiale était de profiter de la redondance d'informations sur la séquence mitochondriale pour suivre les comportements de clones cellulaires particuliers associés à des variations de séquences, que nous pourrions suivre.

La combinaison du séquençage à haut débit avec les approches en cellule unique mises en œuvre par mon équipe, m'a permis dans un premier temps l'identification des haplogroupes pour toutes les cellules issues d'un même patient. J'ai pu ainsi vérifier que le même

haplogroupe était bien identifié lorsque l'ADN mitochondrial était produit et séquencé à partir d'une grande quantité de cellules. Ce faisant, cette démarche expérimentale m'a permis de mettre en place un crible rapide permettant d'écartier les risques d'une potentielle contamination externe par un expérimentateur lors du prélèvement du tissu ou lors de son séquençage en cellule unique. Dans un deuxième temps, j'ai tenté d'identifier des signatures moléculaires mitochondriales qui s'accumuleraient dans un tissu, un type cellulaire ou une situation physiopathologique, pour nous permettre d'étudier son évolution dans un environnement cellulaire complexe (lignage cellulaire).

Matériels et méthodes

I. Culture cellulaire

Les cellules HEK ont été cultivées dans des flasques dans un milieu de culture DMEM avec l'ajout de 10% SVF et 1% Pénicilline/Streptomycine. Les cellules sont incubées à 37°C avec 5% de CO₂.

II. Echantillons

a. RNAseq

Les échantillons utilisés pour le séquençage du transcriptome proviennent de patients atteints de maladies respiratoires allergiques. L'étude GENEPRO a été conçue pour comparer les profils d'expression des brossages nasaux et bronchiques chez 3 groupes de patients: rhinite allergique et asthme, rhinite allergique isolée et témoins sains.

b. Biopsie (10X Chromium Genomics)

A la réception de la biopsie, le prélèvement est incubé à 4°C toute la nuit en présence de 0,1% de Pronase permettant la dissociation des cellules. Le jour suivant, les cellules dissociées sont mises en culture dans un flask durant 4-6 jours selon leur croissance. Lorsque les cellules sont à confluence, les cellules sont mises en culture sur un filtre semi perméable préalablement traité au collagène. Dans un premier temps, ces cellules sont immergées dans un milieu de culture riche au niveau apical et basolatéral du filtre afin de permettre la prolifération des cellules basales. Au moment où les cellules arrivent à confluence, le milieu sur la face apicale est retiré afin de former l'interface air/liquide, et de mimer ainsi les conditions physiologiques des voies respiratoires permettant de induire la différenciation des cellules basales. Ces cellules sont ensuite individualisées à l'aide de l'appareil Chromium (10X Genomics) et séquencées sur l'appareil NextSeq500 (Illumina).

III. Extraction

a. Blood and tissue Kit (Qiagen)

L'ADN total est extrait des cellules à l'aide du kit Blood and Tissue (B-T). Le principe de cette technique est basé dans un premier temps sur l'utilisation de Protéinase K et un tampon de lyse pour dénaturer les membranes et protéines. L'éthanol est ajouté pour précipiter l'ADN puis transférer dans une colonne avec une membrane de silice afin de ne capturer que l'ADN précipité. On ajoute 20 µl de protéinase K pour 5×10^6 cellules HEK resuspendues dans 200µl de PBS. On ajoute 200 µl de tampon AL puis l'échantillon est incubé à 56°C durant 10 min. On ajoute 200 µl d'éthanol 100%. Le volume total est transféré dans une colonne DNeasy Mini Spin. La colonne est lavé avec du tampon AW1. L'ADN total purifié est élué dans 500µl d'H₂O.

b. Miniprep (Qiagen)

L'ADN plasmique supérieur à 10 kb est extrait des cellules à l'aide du kit Miniprep (Mini). Le principe de cette technique est basé dans un premier temps sur une lyse alcaline des cellules, bactéries et mitochondries. L'ADN génomique est précipité et le tube est centrifugé afin séparé le surnageant contenant l'ADN plasmidique du culot contenant l'ADN génomique. Le surnageant est passé à travers un gel de silice pour capturer l'ADN plasmidique pour le laver puis l'éluer. Les cellules sont centrifugées et le culot est resuspendu dans 200 µl de tampon P1. On ajoute 250 µl de tampon P2 et 350 µl de tampon N3. Le surnageant est transféré dans les colonnes Qiaprep 2.0 Spin. Les colonnes sont lavées avec du tampon PB et les extraits d'ADN plasmidiques sont éluées dans 50 µl d'H₂O.

IV. PCR (Polymerase Chain Reaction)

a. Short-PCR

Pour effectuer la PCR de courts fragments, 2 couples d'amorces par gène (Tableau 1) ont été choisis et synthétisés (IDT). Le milieu réactionnel est constitué de 1 µl d'ADN total (200ng) + 11,5µl de mix (9,5µl H₂O + 1µl d'amorce sens 10µM + 1µl d'amorce antisens 10µM) et 12,5µl de GoTaq G2 2X (Promega).

Le programme utilisé est :

- 94°C 4 min
- 94°C 40 sec
- 55°C 40 sec
- 72°C 1 min
- 72°C 5 min

} 30 cycles

b. Long-PCR

Le principe de la Long-PCR est d'amplifier à l'aide d'une polymérase haute fidélité de long fragment d'ADN, 16kb dans notre cas. Le kit de PCR utilisé est la LongAmp Taq DNA Polymerase (NEB). Le milieu réactionnel est constitué de 12 µL de Premix (Taq, Tampon), 1,5 µl de dNTPs 10mM, 2 µL d'amorce sens 10 µM, 2 µL d'amorce antisens 10 µM (Tableau 1), 25 µL H₂O, auquel sont ajoutés 2 µL d'ADN total (200ng).

Le programme utilisé est :

- 94°C 40 sec
- 94°C 30 sec
- 57°C 40 sec
- 65°C 14 min
- 65°C 10 min

} 30 cycles

c. qPCR (quantitative PCR)

La qPCR est une technique d'amplification dans le but de quantifier le nombre de molécule. Un marqueur fluorescent est intégré lors de chaque cycle d'amplification et cette fluorescence est dosée à chaque cycle. Ces données de fluorescence représentent la quantité de produits amplifiés à cet instant. Ainsi plus l'échantillon est concentré moins il faudra de cycle pour atteindre le signal fluorescent significativement supérieur au bruit de fond défini par le Ct (Cycle Treshold). Ce Ct permettra de déterminer la quantité de molécule d'ADN dans l'échantillon. Les librairies de séquençage sont dosées à l'aide du Kit qPCR KAPA Library Quantification de la société KAPABiosystems. Des échantillons standards sont également utilisés à plusieurs concentrations de 20pM à 0,0002pM. Nous distribuons dans les puits de la plaque de qPCR 4µl de standard ou de notre librairie préalablement diluée puis un volume de 6µl de KAPA SYBR FAST qPCR Master Mix, contenant les deux amorces. La plaque est ensuite mise dans l'appareil LightCycler de la société Roche.

Le programme utilisé est : - 95°C 5 min
 - 95°C 30 sec }
 - 60°C 45 sec } 35 cycles
 - 65 - 95°C (Melt curv)

Gènes	Séquences des amorces
COX3 sens (Long PCR)	5' CACGAGAAAGCACATACCAAG 3'
COX3 reverse (Long PCR)	5' GCCTGCAGTAATGTTAGCGG 3'
MT-ND1 F1	5' ATGGCCAACCTCCTACTCCT 3'
MT-ND1 R1	5' GCGGTGATGTAGAGGGTGAT 3'
MT-ND1 R2	5' AGAAGAGCGATGGTGAGAGC 3'
MT-CYTB F1	5' TATCCGCCATCCCATACATT 3'
MT-CYTB R1	5' GGTGATTCCTAGGGGGTGT 3'
MT-CO3 R1	5' GGAAGCCTGTGGCTACAAAA 3'
MT-CO3 F1	5' CCCGCTAAATCCCCTAGAAG 3'
Amorce 1 qPCR	5' AATGATACGGCGACCACCGA 3'
Amorce 2 qPCR	5' CAAGCAGAAGACGGCATAACGA 3'

Tableau 1 : Liste des amorces

Le tableau liste les différentes amorces utilisées et leurs séquences nucléotidiques correspondantes. Les amorces COX3 sens et antisens sont utilisées pour l'amplification « Long-PCR » de l'ADN mitochondrial. Les amorces MT sont utilisées pour les PCR de courte amplification. Les amorces 1 et 2 sont utilisées pour la qPCR afin de quantifier la librairie.

V. Purification des acides nucléiques

La purification a été effectuée à l'aide des billes magnétiques Ampure XP (Beckman Coulter). Cette technique est basée sur la sélection de taille des fragments d'ADN. Nous avons utilisé une proportion à volume/volume de billes Ampure XP, c'est-à-dire 50 µl de billes Ampure XP pour 50 µl de volume PCR permettant de conserver la totalité des fragments supérieurs à 100pb.

VI. Digestion enzymatique

Pour effectuer la digestion enzymatique, nous avons utilisé deux enzymes de restriction EcoRI et HindIII. Le principe de cette technique est basé sur la reconnaissance et la coupure d'une séquence spécifique par une enzyme de restriction. Dans deux tubes séparés, nous avons ajouté 1 µl d'ADN à 1 µg/µl, 1 µl de 10X Buffer (Buffer H pour EcoRI et Buffer E pour HindIII), 1 µl d'enzyme de restriction et 16 µl d'eau. Le mélange est ensuite incubé 1 heure à 37°C puis afin d'inactiver l'enzyme la température est augmentée à 65°C durant 15 min. Le produit de digestion est ensuite purifié avec les billes AmpureXP.

VII. Sonication

La fragmentation des échantillons s'effectue grâce à l'appareil S2 Focused-ultrasonicators de la société Covaris. Le principe de cette méthode est d'utiliser des ondes sonores qui agitent les particules et créent une cassure de l'ADN. L'échantillon est transféré dans un micro-tube de verre (AFA fiber snap) et le volume est complété à 130 µl afin de totalement remplir le tube. La particularité de celui-ci est qu'il possède une tige de plastique à l'intérieur augmentant l'efficacité de fragmentation lors de l'ultrasonication.

Programme de fragmentation d'ADNmt: - Peak Intensity: 5
 - Duty Cycle: 10%
 - Cycles Burst: 200
 - Time: 7 cycles de 60sec

L'ADNmt fragmenté est alors purifié à l'aide des billes AmpureXP.

VIII. Contrôles

a. Nanodrop

Le contrôle quantitatif des différents échantillons est déterminé par lecture au spectrophotomètre Nanodrop (NanoDrop ND-1000). L'absorbance est mesurée à 260 nm afin de quantifier les acides nucléiques. La pureté de l'échantillon est visualisée avec le rapport d'absorbance A260/A280 et 260/230. Les rapports calculés doivent être autour de 1,8. Un rapport 260/280 inférieur à 1,8 signifie que des protéines et des restes de phénol contaminent

l'échantillon et si le rapport 260/230 est inférieur à 1,8, l'échantillon est contaminé en solvant et sels.

b. Bioanalyzer

La qualité des ARN ou de l'ADN est déterminée par électrophorèse capillaire microfluidique sur un Bioanalyzer (Agilent 2100). Pour les ARN, le RIN (RNA Integrity Number) est une valeur sur 10, déterminant si l'ARN est intact (RIN=10) ou s'il est dégradé (RIN<10).

IX. C1 Fluidigm

L'appareil C1 de chez Fluidigm permet d'isoler 96 cellules à l'aide d'une plaque IFC. On ajoute le réactif C1 Harvest à cette plaque pour préparer les micro-réseaux. Les cellules sont dissociées puis diluées dans un réactif de suspension avant d'être déposées dans la plaque IFC. Un marqueur de mort cellulaire, un tampon de lavage et un réactif de préchargement sont ensuite ajoutés et la plaque peut être chargée. Durant cette étape, l'appareil va capturer individuellement 96 cellules. Une fois la phase de capture accomplie, nous observons au microscope LAS Imager les sites de capture des cellules individuelles afin de vérifier si celles-ci sont vivantes. On ajoute ensuite le tampon de lyse et le mix de « Long-PCR ». Les cellules sont transférées dans des chambres microfluidiques successives dans lesquelles vont s'exécuter séparément la phase de lyse puis la PCR de l'ADNmt. A la fin de ces étapes, le contenu est alors acheminé dans un puits, nous permettant la collecte des acides nucléiques amplifiés.

X. Chromium 10X Genomics

Pour débiter la préparation, un mélange contenant du RT Reagent Mix, RT Primer, de l'Additive A et un mix de RT Enzyme est préparé. Un mélange cellulaire contenant 1 500 cellules dissociées est ajouté au mélange de RT et cette suspension est transférée dans un puits. Un volume de 40µl de billes vortexées préalablement est transféré dans un autre puits et 270µl d'huile est ajouté dans un puits séparé. La puce est ensuite insérée dans l'appareil Chromium. Durant cette étape, l'appareil va former une émulsion de 100.000 gouttelettes contenant zéro ou une cellule et une seule bille avec un code-barre, et l'ensemble des réactifs nécessaires à la lyse des cellules et à la transcription inverse des ARN (Figure 11). A la sortie

de l'appareil, il faut absolument vérifier que l'émulsion est homogène. L'émulsion est délicatement transférée dans une plaque et va être incubée à 53°C durant 45 min afin d'effectuer la transcription inverse. Une fois la réaction effectuée, l'émulsion est ensuite purifiée à l'aide des billes DynaBeads MyOne Silane puis les fragments d'ADNc sont amplifiés par PCR à l'aide d'un mix contenant un Amplification Master Mix, ADNc Additive et un mix d'ADNc primer. A partir de l'amplification, on construit la librairie en fragmentant l'ADNc, les extrémités des fragments sont réparées et une base A est ajoutée à l'extrémité 3'. Une purification est ensuite effectuée avec les billes SPRI Select (Beckman Coulter) puis l'adaptateur P7 est ligué à une extrémité des fragments. Cette construction est amplifiée par PCR, quantifiée à l'aide du kit de qPCR KAPA Library Quantification puis diluée à 4nM et séquencée sur le NextSeq500.

XI. Séquençage

L'appareil utilisé sur la plateforme est le NextSeq500 de la société Illumina. Cette société utilise une technologie basée sur de l'amplification à partir d'un support solide possédant des brins complémentaires aux adaptateurs utilisés pour la librairie. Ces adaptateurs sont liés à chaque extrémité de la séquence cible. Son amplification va former des colonies d'un seul et même fragment. Les quatre types de bases sont ajoutés successivement et celles non-hybridées sont éliminées. Ce type de séquençage permet d'incorporer un nucléotide à la fois et le cycle est ainsi répété jusqu'à la taille attendue du fragment. Les échantillons sont regroupés dans un tube après les avoir dilués à une même concentration (4nM). Ce tube est dénaturé, dilué à 1,8pM puis chargé sur une cellule de séquençage Nextseq 500 et séquencé sur la plateforme NextSeq 500 (Illumina) avec la chimie paired-end 75pb x 75pb ou 26pb x 98pb. Nous pouvons ainsi obtenir plus de 400 millions de séquences en 24 heures.

a. RNA-seq

Les librairies sont obtenues à partir de 500ng d'ARN total grâce au kit Truseq Stranded Total RNA (Illumina). Les librairies sont fabriquées en premier lieu par la transcription inverse en obtenant un ADN complémentaire. Un double brin est obtenu puis les extrémités sont réparées et des nucléotides A sont ajoutés. L'ajout de ce nucléotide va permettre l'utilisation des adaptateurs avec des codes-barres (permet le multiplexage des échantillons) et ainsi effectuer la ligation. Les librairies sont amplifiées durant 15 cycles.

b. DNA-seq

Les librairies sont obtenues grâce au kit Truseq Nano DNA Library Preparation Kit (Illumina) à partir de l'ADN mitochondrial préalablement amplifié et fragmenté avec le sonicateur. Les librairies sont fabriquées en premier lieu par réparation des extrémités des fragments et des nucléotides A sont ajoutés. On ligue des adaptateurs et des codes-barres (permet le multiplexage des échantillons) puis les librairies sont amplifiées durant 15 cycles.

XII. Bioinformatique

Après séquençage de la librairie, les données de sortie du NextSeq500 sont ensuite exportées. Les fichiers obtenus sont traités par les bio-informaticiens du laboratoire afin de convertir le signal en base sous forme de fichier .fastq avec le script bcl2fastq. Ces séquences brutes sont ensuite alignées sur le génome de référence avec le logiciel de mapping STAR permettant d'attribuer à chaque lecture sa position chromosomique. Nous pouvons ensuite identifier et extraire les séquences spécifiques de la mitochondrie sous forme d'un fichier .bam.

Concernant les résultats de Chromium de chez 10X Genomics, une étape supplémentaire de tri, à l'aide de l'étiquette de cellule, est effectuée afin d'attribuer chacune des séquences à la cellule qui lui est associée.

Une recherche de polymorphisme sur un seul nucléotide (SNP) est faite à partir des fichiers .bam pour déterminer les mutations vis-à-vis du génome mitochondrial de référence (NC_012920) et va ainsi recréer la séquence entière en indiquant pour chaque position la base majoritaire. Ce fichier final est au format .fasta.

Résultats

I. Mise en place du séquençage d'ADNmt

Afin d'obtenir des réponses sur l'hétérogénéité mitochondriale, j'ai dû dans un premier temps mettre en place le séquençage de l'ADNmt sur la plateforme de génomique de l'IPMC.

La première étape a consisté à extraire la totalité du matériel génétique d'un culot cellulaire de cellules HEK. J'ai testé deux kits d'extraction, le premier permettant l'extraction totale du matériel génétique des cellules (Blood and Tissue), le second étant un kit permettant l'extraction d'ADN plasmidique, mieux ciblé pour l'ADN mitochondrial (Miniprep).

Suite à l'extraction, j'ai vérifié la présence de l'ADNmt dans notre échantillon. Pour cela, j'ai amplifié 3 régions génomiques mitochondriales réparties sur la totalité des 16kb de l'ADNmt. J'ai utilisé une amorce sens (F1) et une amorce antisens (R1 et/ou R2) spécifiques de 3 régions codantes différentes (gène MT-ND1, MT-COX3 et MT-CYTB) sur le génome mitochondrial.

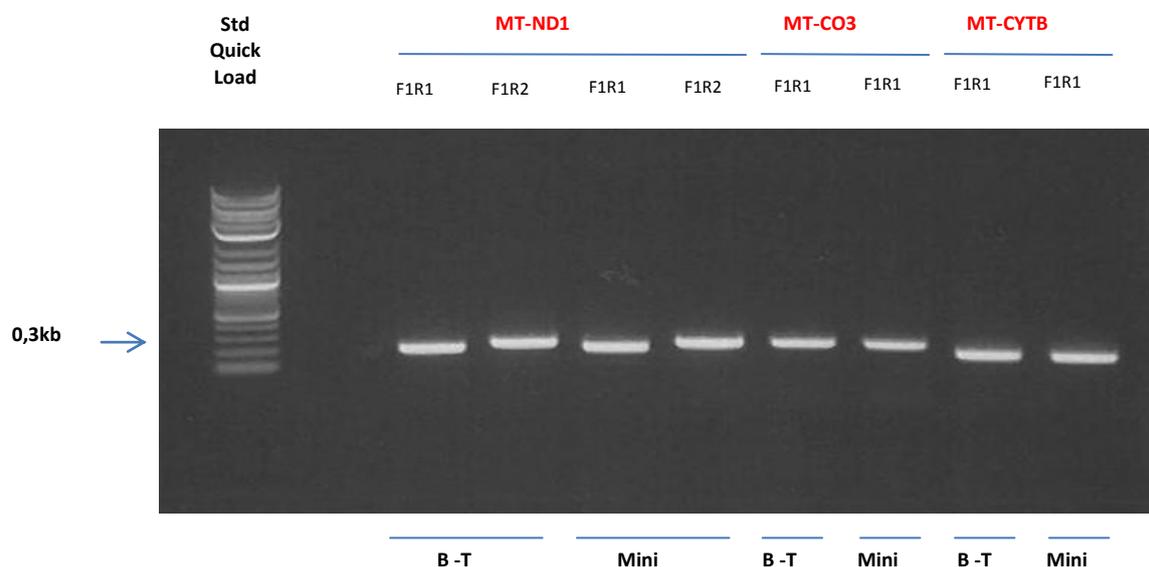


Figure 12: Migration des produits de PCR de trois régions codantes dur l'ADNmt en gel d'agarose 1%.

B-T correspond à la technique d'extraction effectuée avec le kit Blood and Tissue (Qiagen) et Mini correspond à la technique d'extraction effectuée avec le kit Miniprep (Qiagen). Des combinaisons d'amorces ont été effectuées sur trois gènes différents ND1, CO3 et CYT-B. L'ADN mitochondrial a été amplifié à partir de 200ng d'ADN total.

La figure 12 indique la présence d'une bande d'environ 250 pb, correspondant à la taille attendue des fragments mitochondriaux amplifiés pour chacun de nos gènes.

A partir de cette première vérification, il s'agissait ensuite de trier en cellule unique sur le C1 Fluidigm et il convenait alors de préparer l'extrait par lyse cellulaire. Le matériel génétique obtenu à partir de la méthode d'extraction de l'ADN total étant le plus similaire à la lyse cellulaire, c'est donc sur la technique d'extraction avec le kit Blood and Tissue que j'ai poursuivi mon projet.

L'amplification des petits fragments de PCR démontrant la présence de l'ADN mitochondrial, je me suis alors attaché à réaliser une amplification « Long-PCR » sur la totalité du génome mitochondrial à partir de 1ng d'ADN total.

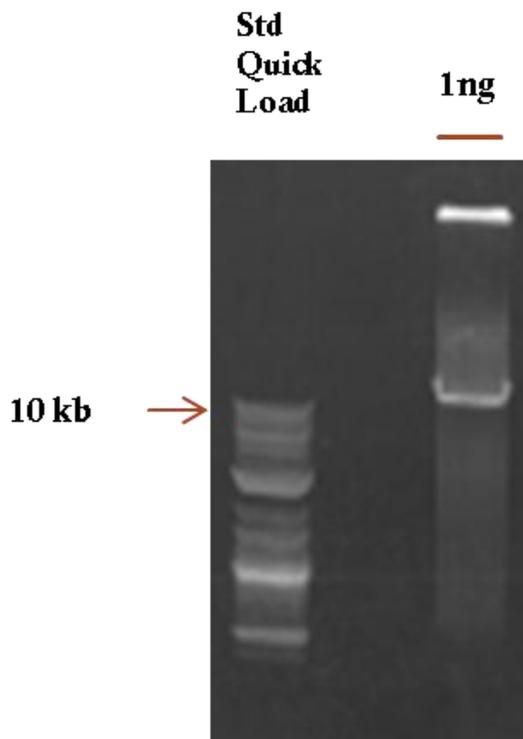


Figure 13 : Migration du produit de « Long-PCR » sur gel d'agarose 0,7%.

Les amorces s'hybrident sur une partie du gène CO3 ont été utilisées à partir de 1ng d'ADN total de départ.

Les résultats de la « long-PCR », démontrent l'amplification correcte de l'ADN mitochondrial à partir de 1ng de matériel génétique de départ.

Pour confirmer que la potentielle bande vue sur le gel correspondait bien à une amplification de la séquence attendue, j'ai découpé à partir du gel la bande observée à 15-16kb de la figure

13. Deux types de vérification ont été effectués. Une purification puis une amplification avec les 3 couples d'amorces ND1, CO3 et CYT précédemment décrits, ont été effectués.

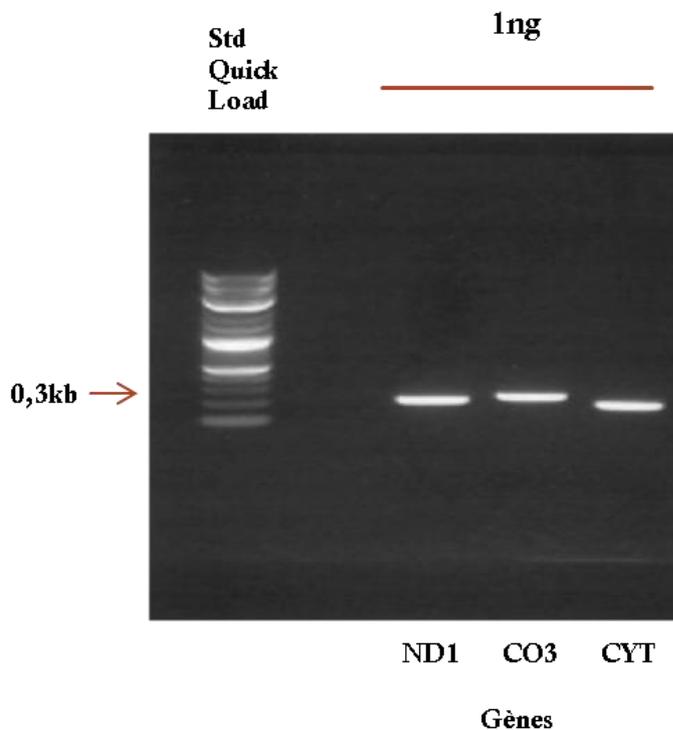


Figure 14 : Migration des produits de PCR sur gel d'agarose 1%.

Les amorces des 3 différents gènes (ND1, CO3 et CYT-B) ont été utilisées à partir de la purification extraite du gel «Long-PCR».

Nous pouvons observer sur la figure 14, la présence d'une bande pour chacun des couples d'amorces utilisés. Cela confirme que c'est bien l'ADN mitochondrial qui a été amplifié.

La seconde vérification effectuée a été de digérer avec des enzymes de restrictions l'extrait purifié du gel. Après digestion par EcoRI et HindIII, j'ai comparé les bandes obtenues (Figure 15.B) avec leurs tailles théoriques attendues (Figure 15.A).

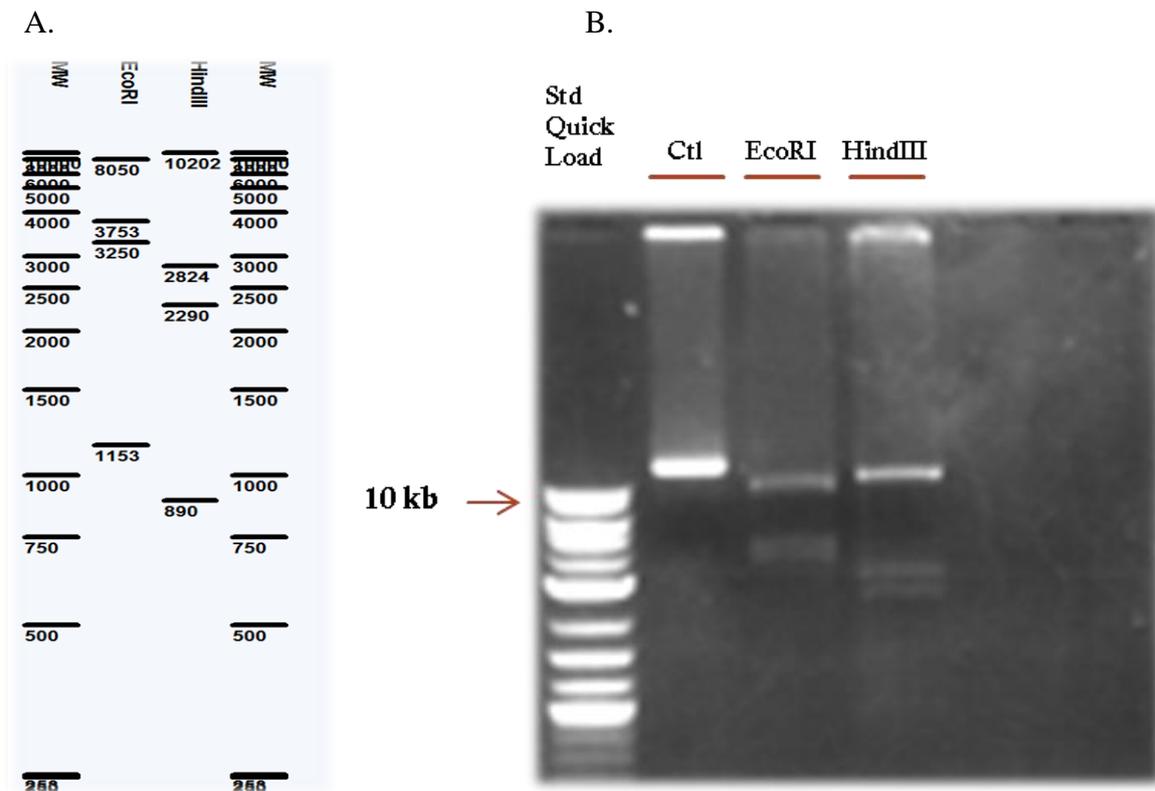


Figure 15 : Résultats de digestion de l'ADNmt avec EcoRI et HindIII.

La figure A est la représentation théorique des coupures effectuées par les enzymes de restriction EcoRI et HindIII sur l'ADN mitochondrial amplifié. La figure B est la migration des produits de digestion enzymatique à partir d'ADNmt amplifié sur gel d'agarose 1%.

En comparant la figure 15.A avec la figure 15.B, on peut observer pour la digestion EcoRI que les bandes attendues à 8050pb, 3753pb et 3250pb sont bien retrouvées dans le produit de digestion. Pour la digestion HindIII, on retrouve également sur la figure 15.B, les bandes attendues à 10202pb, 2824pb et 2290pb. Les bandes à 1150pb pour EcoRI et à 890pb pour HindIII sont probablement en très faible quantité et ne sont pas visibles sur le gel d'agarose.

A partir de ces deux vérifications, nous pouvons affirmer que nous amplifions bien la totalité du génome mitochondrial grâce à la « Long-PCR ».

Maintenant que nous avons la confirmation de pouvoir amplifier l'ADN mitochondrial, il nous faut, pour construire une librairie de séquençage, des fragments d'ADN mitochondriaux de taille d'environ 180pb. Cette taille est nécessaire afin de pouvoir ajouter les adaptateurs et permettre le séquençage. Pour cela, j'ai utilisé une fragmentation par sonication pour casser les longs fragments d'ADNmt. J'ai testé la fragmentation à différents temps sur l'amplification d'ADNmt à partir de 200ng d'ADN total.

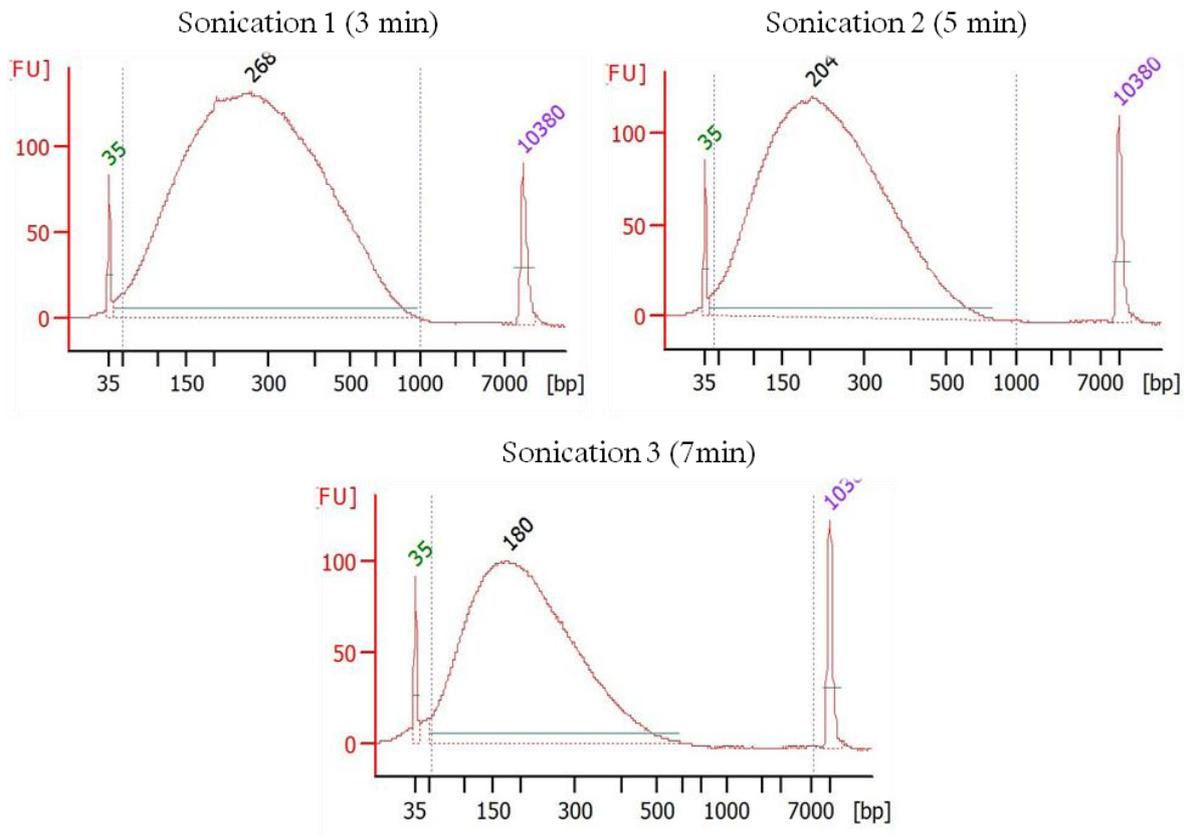


Figure 16 : Test de fragmentation par sonication.

Trois temps différents (2min, 5min et 7min) de fragmentation ont été effectués sur de l'ADNmt obtenu par « Long-PCR ».

On peut observer dans la figure 16, que le temps optimal pour obtenir la taille de fragments adéquate est de 7min.

Suite à la fragmentation, j'ai effectué la librairie de séquençage à partir de ces fragments d'ADNmt. Pour cela, j'ai réparé les extrémités de ces fragments car lors de la cassure, la séparation entre les deux fragments n'est pas nette. Ensuite une base A est ajoutée sur l'extrémité 3' pour permettre la ligation des adaptateurs. La construction est ensuite purifiée et analysée sur le Bioanalyzer.

Library mtDNA HEK

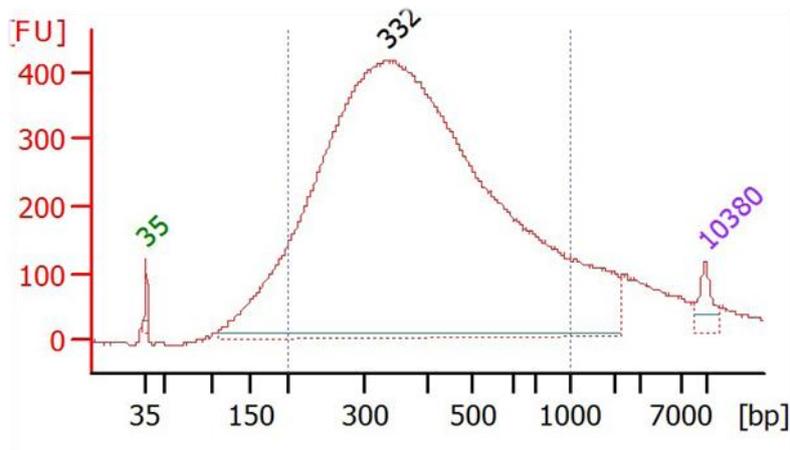


Figure 17 : Profil Bioanalyzer de la librairie d'ADNmt

Le pic à 35pb est le Lower marker et le pic à 10380pb est le upper marker, ils permettent de définir la taille de l'échantillon. La « cloche » observée représente l'ensemble des fragments de la librairie.

Nous pouvons observer sur la figure 17, un profil en forme de cloche dont le pic se situe à 332pb. Cela correspond à la taille attendue pour un séquençage sur Nextseq500. J'ai également effectué un contrôle de la librairie en effectuant une qPCR utilisant des amorces spécifiques des adaptateurs utilisés pour la construction de la banque. Le profil observé sur la figure 17 est bien celui désiré, et cette séquence possède bien à ses extrémités les adaptateurs attendus.

II. Comparaison des données de séquençage d'ADNmt

Après avoir mis en place la construction d'une librairie d'ADNmt sur notre plateforme, l'étape suivante a été de faire une librairie d'ADNmt sur des échantillons provenant de tissus d'origines embryonnaires différentes d'un même patient, afin que le plus grand nombre de divisions cellulaires possibles permette de mesurer le nombre de différences de séquences pouvant s'accumuler chez un même individu. J'ai prélevé les cellules sanguines (d'origine mésodermale) et les cellules nasales (d'origine endodermale) d'un même patient sur lesquelles j'ai extrait l'ADN total et amplifié l'ADNmt. J'ai ensuite effectué une fragmentation avec l'appareil Covaris.

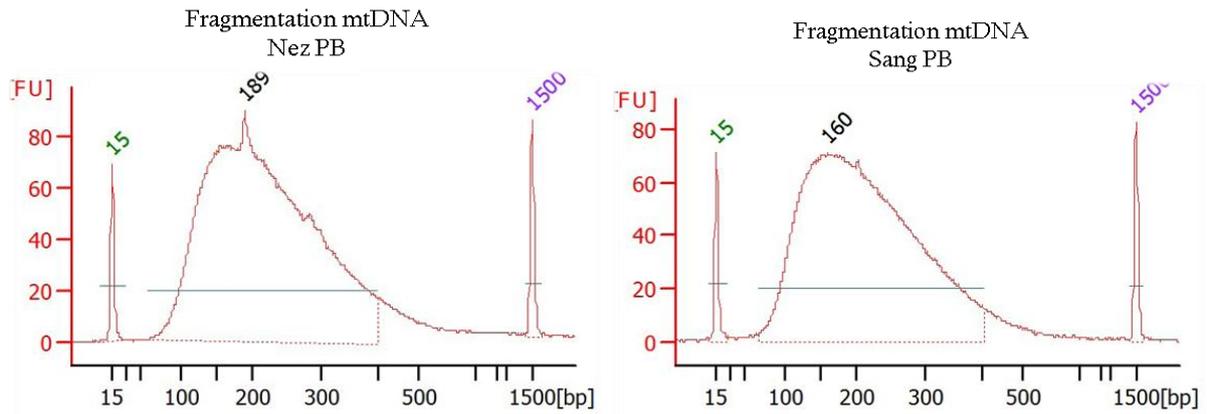


Figure 18 : Fragmentation de l'ADNmt.

L'ADNmt amplifié à partir du sang et nez d'un patient a été fragmenté à l'aide du Covaris durant 7min.

Après m'être assuré de la taille des fragments d'ADNmt (Figure 18), j'ai effectué la construction de la librairie.

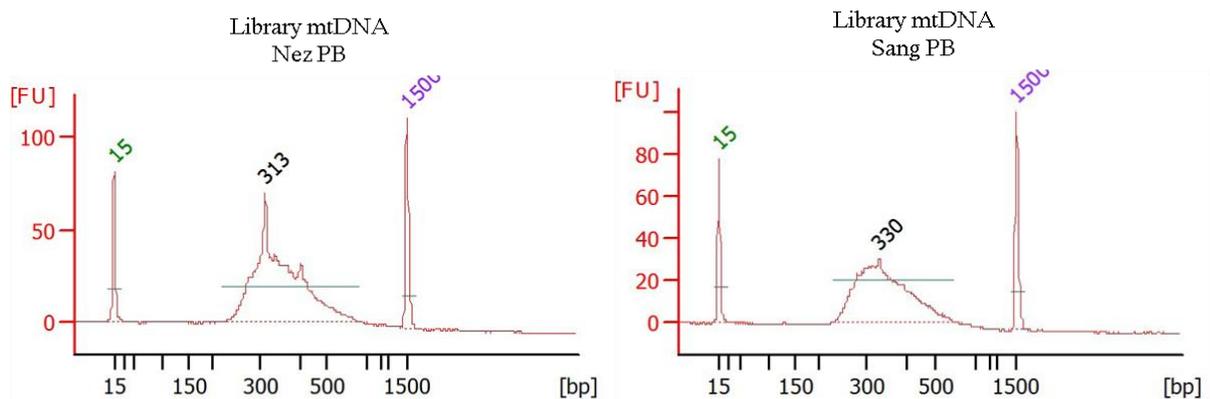


Figure 19 : Bibliothèques d'ADNmt à partir du nez et du sang du patient.

Les deux profils BioAnalyzer ci-dessus sont les contrôles des bibliothèques d'ADNmt des échantillons de nez et de sang du patient PB. La taille obtenue des fragments pour chacune des bibliothèques est 313pb pour le nez et 330pb pour le sang.

Les bibliothèques sont alors dosées par qPCR et je les mélange ensemble à 4nM. Les échantillons sont mélangés pour pouvoir les séquencer en une seule fois. Le mélange est alors dénaturé par NaOH pendant 5 min puis stoppé par du Tris-HCl. Les bibliothèques sont alors diluées puis séquencées dans l'appareil Illumina, le Nextseq500.

Une fois le séquençage effectué, les fichiers obtenus sont analysés et les séquences obtenues sont attribuées pour chaque bibliothèque à l'aide de l'index ajouté avec l'adaptateur. Ces séquences sont ensuite alignées sur l'ADNmt de référence et sont regroupées dans un fichier .bam. Ce fichier est visualisable dans un Genome Browser comme IGV (Integrative Genome

Viewer) et nous permet de visualiser les mutations des échantillons par rapport à la référence et entre eux directement sur le génome mitochondrial complet.

Lors de l'analyse de ces deux échantillons, j'ai identifié plusieurs mutations spécifiques communes aux deux échantillons qui permettent de définir son haplogroupe. L'haplotype du patient PB est ainsi W3a1c. W3a1c correspond à une branche de l'arbre phylogénétique mitochondrial. Il a été déterminé à l'aide Haplogrep (<http://haplogrep.uibk.ac.at/>) (Weissensteiner Hansi et al. 2016). Chacune des mutations appartenant à son haplogroupe a été écartées afin de mettre en lumière les néo-mutations (Annexe 2). Aucune mutation totale, c'est-à-dire une substitution à 100% d'une base à la place de la base de référence n'a été trouvée. En revanche, des taux différents de substitutions de bases à différentes positions ont été observés.

A partir de ces deux séquençages, effectués sur deux types cellulaires distincts, j'ai effectué une étude sur l'hétéroplasmie. Pour ce faire, j'ai sélectionné la seconde base la plus abondante sur chacune des positions. En effet, la présence de séquence en grande quantité pour une deuxième base correspondrait à une hétéroplasmie et il serait ainsi plus facile de visualiser l'écart du nombre de séquences lues pour la deuxième base présente pour chaque position entre les deux échantillons. J'ai alors calculé le pourcentage d'abondance de cette base par rapport au nombre total de séquences détectées. La figure 20 montre la relation entre ces ratios, dans le Sang (abscisse) versus Nez (ordonnée).

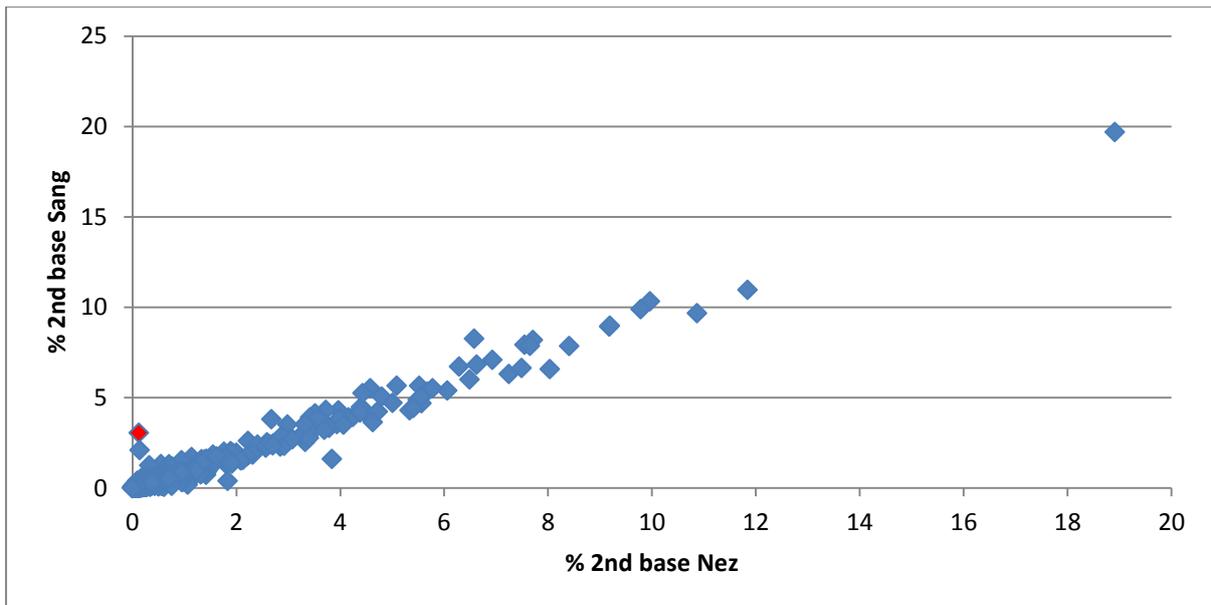


Figure 20 : Comparaison des pourcentages d'hétéroplasmie entre l'ADN sanguin et l'ADN nasal.

Le pourcentage de la seconde base la plus abondante de chacune des positions de l'ADNmt a été répertorié. Le graphique ci-dessus est la comparaison de ces pourcentages entre l'échantillon de Nez et de Sang du patient PB. Le point rouge correspond à la position 1338.

Nous observons sur la figure 20 une bonne concordance globale entre les deux échantillons, ce qui démontre la bonne stabilité de l'hétéroplasmie, c'est-à-dire que le pourcentage d'hétéroplasmie de la seconde base est identique entre les deux échantillons. Il est toutefois à noter que quelques positions nucléotidiques présentent des pourcentages différents, telle la position 1338.

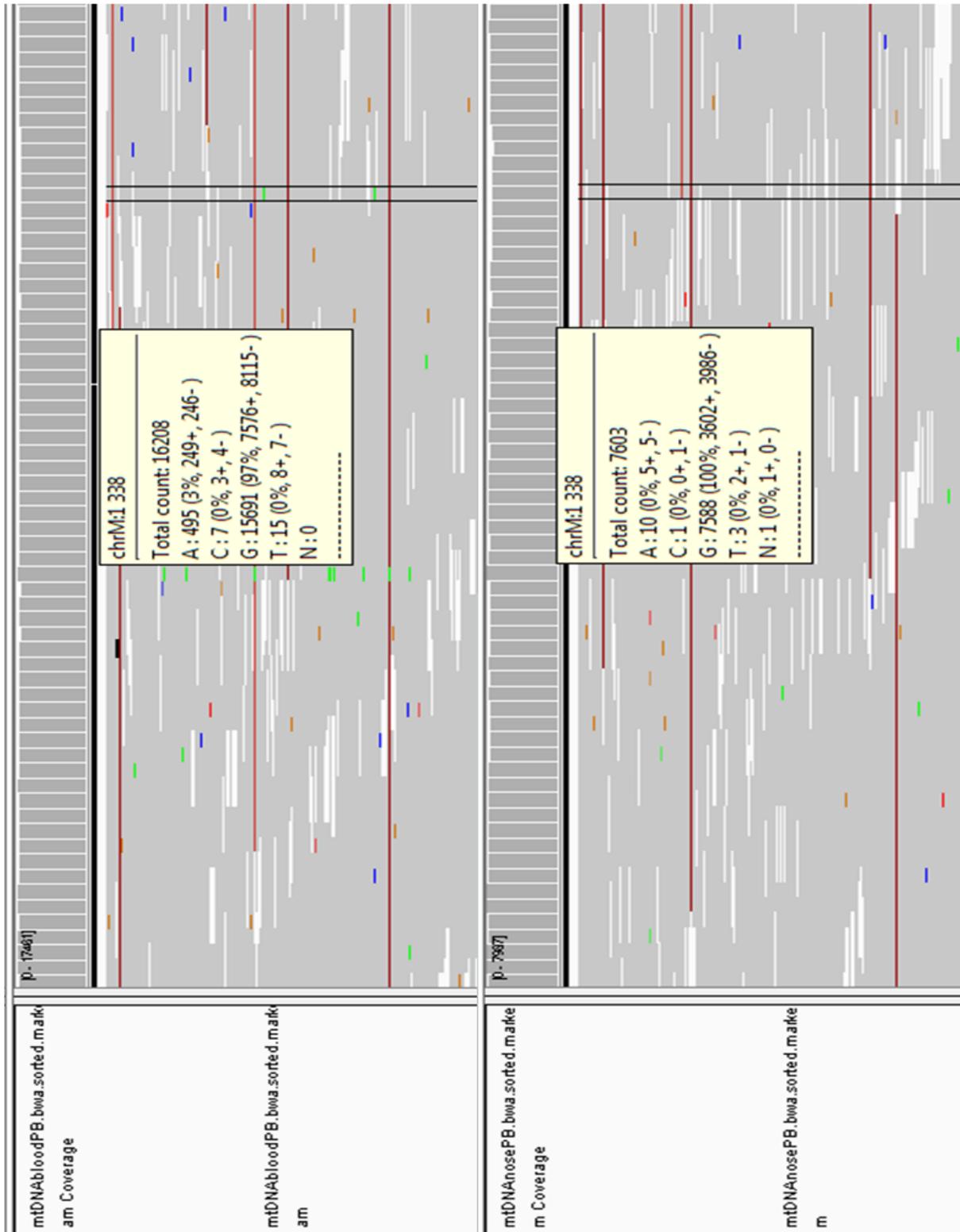


Figure 21 : Alignement sur IGV des séquences d'ADNmt des bibliothèques de sang et de nez du patient PB

Les séquences lues et traitées sont alignées sur le génome mitochondrial de référence. La position 1338 semble montrer une différence dans la répartition des différentes bases.

La figure 21, représente une capture d'écran sous IGV permettant de visualiser les alentours de la position 1338 sur l'ADNmt, on observe respectivement un pourcentage de 97% et 100% de la base G pour le sang et le nez. La base A est détectée dans 3% seulement pour l'échantillon de sang, j'ai détaillé le nombre de séquences lues pour chacune des bases à la position 1338.

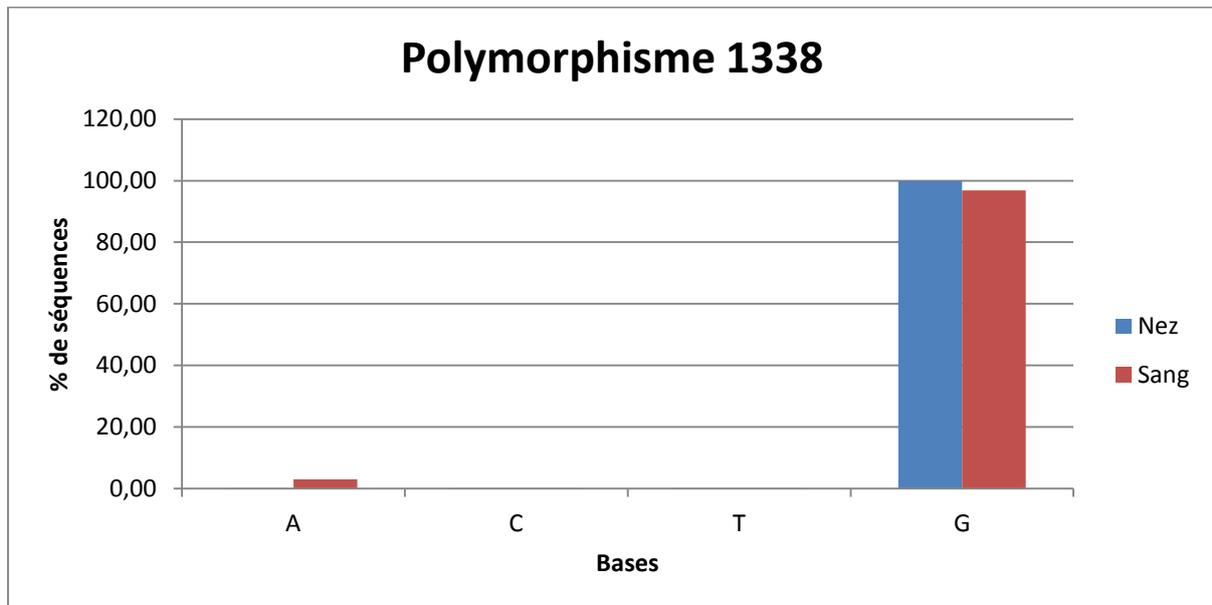


Figure 22 : Répartition nucléotidique pour le polymorphisme G1338

Chaque base à la position 1338 est comparée entre les échantillons de Nez et de Sang du patient PB.

Ces pourcentages ne nous indiquent pas si ces variations de séquences sont issues de quelques cellules portant une substitution totale, ou si toutes les cellules présentent un niveau constant d'hétéroplasmie.

Pour approfondir cela, j'ai souhaité mettre en place le séquençage de l'ADNmt à l'échelle de la cellule unique.

III. Mise en place du séquençage d'ADNmt en cellule unique

Afin de mettre au point le séquençage de l'ADNmt en cellule unique, j'ai à nouveau utilisé les cellules HEK et j'ai répété le travail en diminuant la quantité d'ADN total de départ jusqu'à pouvoir construire une librairie d'ADNmt à partir de seulement 10pg d'ADN total, soit la quantité d'ADN présent dans une seule cellule.

J'ai tenté dans un premier temps d'amplifier l'ADNmt à l'aide de la PCR courte avec les amorces s'hybridant sur le gène ND1, utilisé lors des précédentes mises au point.

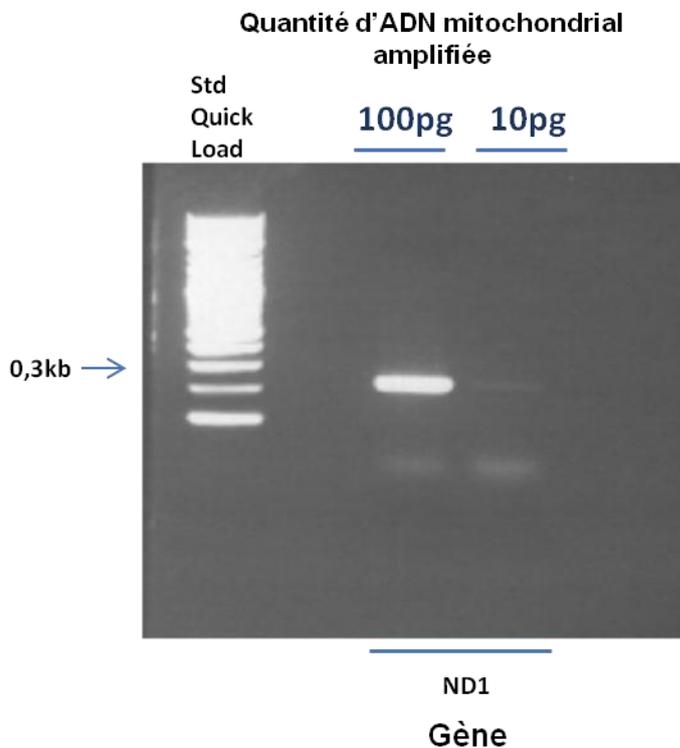


Figure 23 : Migration des produits de PCR à différentes quantités sur gel d'agarose 1%. La totalité de l'amplification à 10pg de l'ADN mitochondrial avec les amorces ND1 a été déposée sur le gel. Le dépôt de l'amplification à 100pg d'ADN mitochondrial avec les amorces ND1 sert de contrôle positif.

La figure 23 montre qu'il est possible d'amplifier une courte séquence d'ADNmt à partir de 100pg et 10pg d'ADN total. J'ai ensuite essayé d'amplifier la totalité de l'ADNmt grâce à la « Long-PCR » en réduisant les quantités d'ADN total de départ.

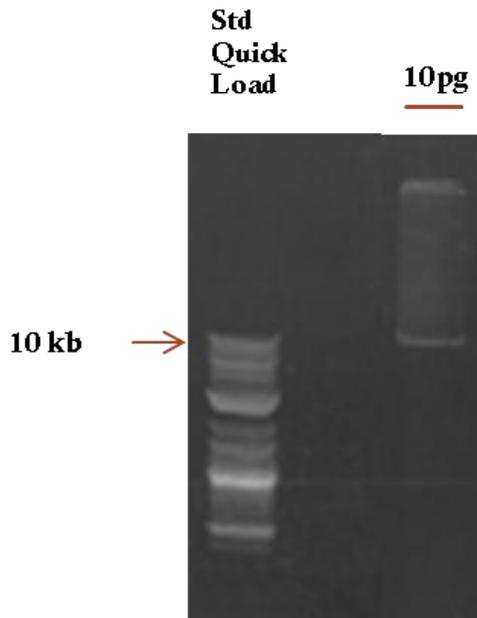


Figure 24 : Migration des produits de PCR à 10pg sur gel d'agarose 1%.
La totalité de l'amplification à 10pg de l'ADN mitochondrial avec les amorces COX3 « Long-PCR » a été déposée sur le gel.

Comme montré sur la figure 24, la bande que nous observons légèrement au-dessus de 10kb suggère l'amplification correcte des 16,5kb de l'ADNmt. Pour contrôler que cette bande correspond bien à l'ADNmt, j'ai effectué, à partir du produit de la « Long-PCR », un courte PCR avec les amorces des trois différents gènes ND1, CO3 et CYT.

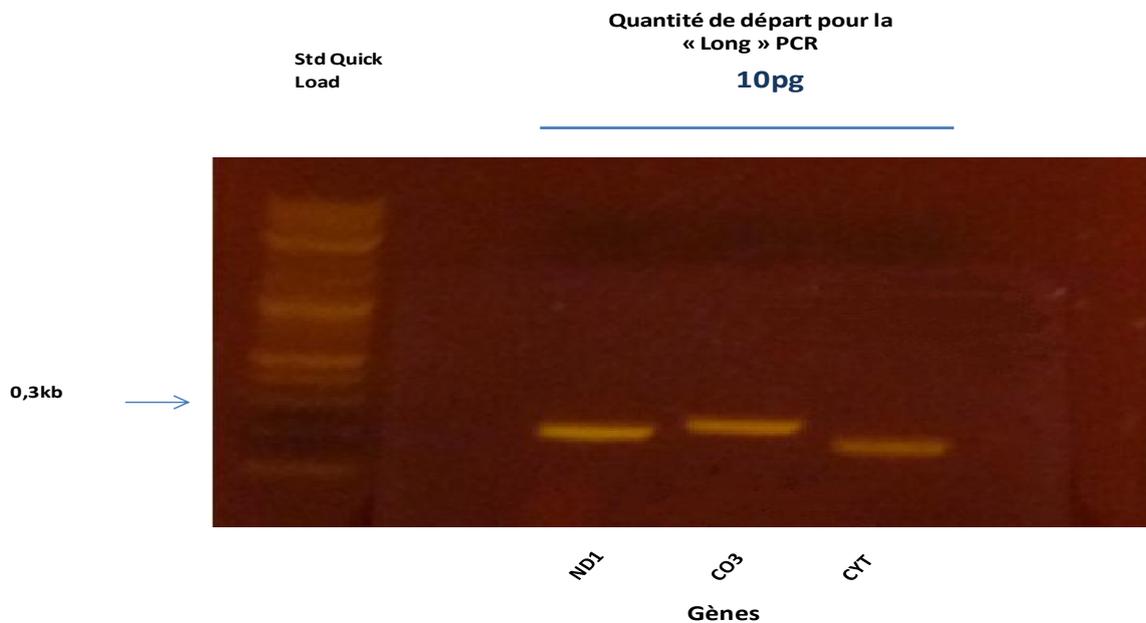


Figure 25 : Migration des produits de PCR des trois régions sur gel d'agarose 1%.
La bande obtenue pour la « long » PCR à partir de 10pg a été découpée du gel d'agarose et purifiée. Une nouvelle amplification à partir de cette purification a été faite avec les amorces des trois régions codantes ND1, CO3 et CYT-B.

On peut voir sur la figure 25, l'amplification des 3 courtes séquences à partir de l'amplification complète de l'ADNmt, ce qui montre qu'il s'agit bien de la bonne séquence.

La fragmentation de l'ADNmt est effectuée par sonication suivie par la construction de la librairie que j'ai dosée au BioAnalyzer.

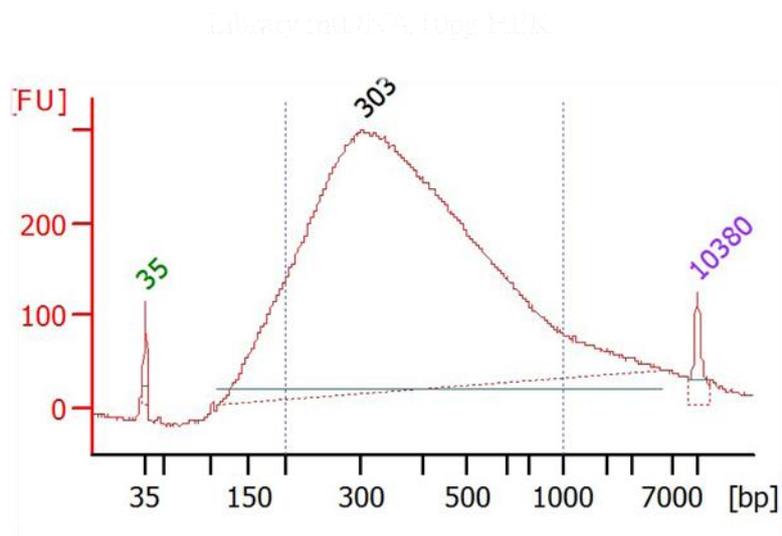


Figure 26 : Profil BioAnalyzer de la librairie d'ADNmt obtenue à partir de 10pg d'ADN total.

Le pic à 35pb est le Lower marker et le pic à 10380pb est le upper marker, ils permettent de définir la taille de l'échantillon. La « cloche » observée représente l'ensemble des fragments de la librairie obtenus à partir de 10pg d'ADN total.

Suite à l'obtention d'une librairie d'ADNmt à partir de 10pg d'ADN total (correspondant à la quantité équivalente de matériel génétique présente dans une et une seule cellule), j'ai alors tenté de reproduire la production de la librairie d'ADNmt en cellule unique en utilisant des cellules isolées sur l'appareil C1 (Fluidigm). Pour cela, j'ai créé un protocole spécifique pour le C1 permettant d'effectuer dans l'appareil, la lyse de chaque cellule puis d'amplifier l'ADNmt dans les mêmes conditions que la « Long-PCR ».

Pour commencer, le chargement de la plaque C1 avec les différents tampons est effectué. Une quantité de 15000 cellules HEK préalablement dissociées ont été introduites dans la plaque ainsi que les réactifs. La plaque est ensuite chargée dans l'appareil et l'appareil fait ensuite circuler les cellules afin de les capturer individuellement (Figure 27)

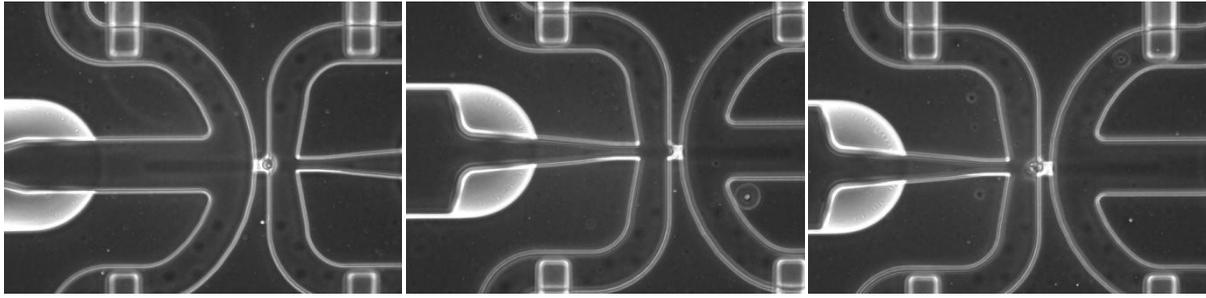


Figure 27 : Capture des cellules sur la plaque C1 à l'aide du LAS Imager.

Les cellules ont circulé dans ce réseau microfluidique. On observe à l'aide du LAS Imager, la capture d'une seule cellule.

Après le flux cellulaire distribué dans la plaque, j'ai vérifié à l'aide du LAS Imager, un microscope me permettant de photographier les 96 puits, les puits avec une seule et unique cellule capturée comme montré dans la figure 27. Lors du chargement de la plaque, j'ai mis également un réactif permettant de visualiser la mort cellulaire afin d'exclure les puits avec des cellules capturées mortes. Tous ces contrôles en milieu de parcours vont me permettre de pouvoir prélever uniquement les acides nucléiques des puits ayant une seule cellule vivante.

La plaque est ensuite remise dans l'appareil C1 pour effectuer la lyse et l'amplification de chaque cellule.

Sur cette expérience, le nombre de cellules isolées est plutôt faible (40 cellules isolées et vivantes / 96 chambres réactionnelles). Après amplification, j'ai récupéré dans chacun des puits un volume de 8 μ l contenant le lysat cellulaire ainsi que l'ADNmt amplifié.

J'ai ainsi effectué pour chacune des cellules la fragmentation par sonication puis la construction de la librairie.

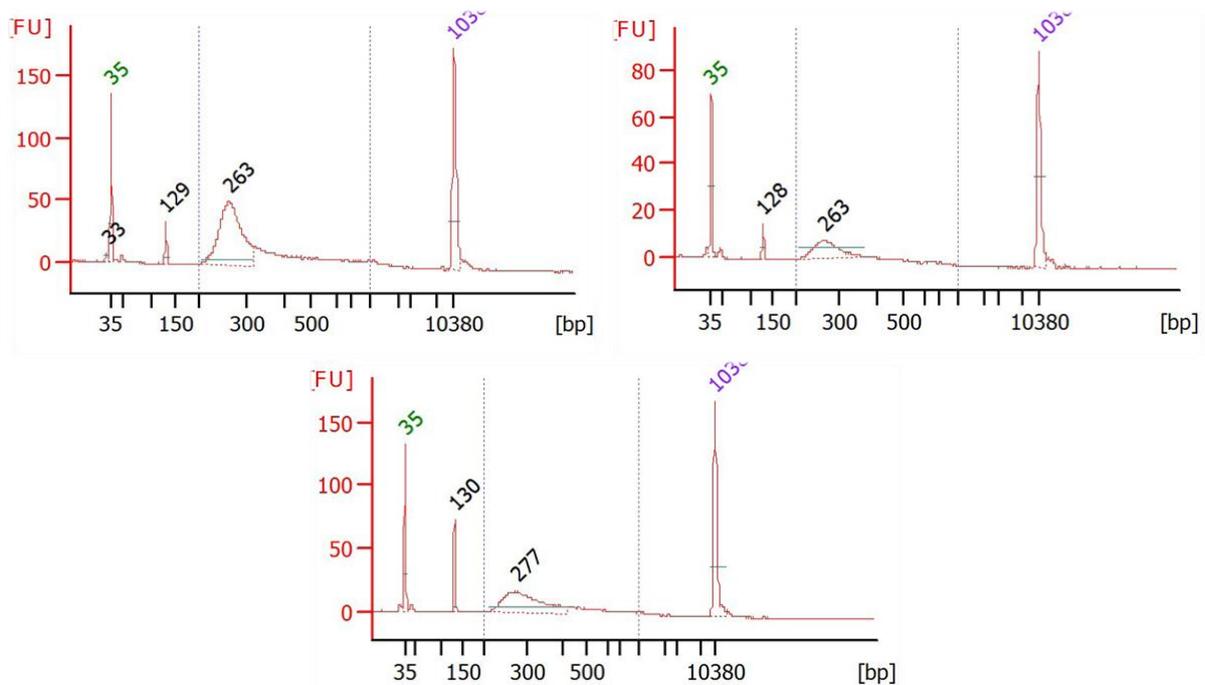


Figure 28 : Profil BioAnalyzer des bibliothèques d'ADNmt d'une cellule.

Le pic à 35pb est le Lower marker et le pic à 10380pb est le upper marker, ils permettent de définir la taille de l'échantillon. La « cloche » observée représente l'ensemble des fragments de la bibliothèque obtenus à partir d'une seule cellule. Le pic vers 130pb correspond au dimère d'adaptateur.

Après contrôle des bibliothèques au BioAnalyzer, on peut observer le pic attendu correspondant au fragment d'ADNmt avec les adaptateurs à chaque extrémité. La taille de la bibliothèque à 260pb est légèrement plus petite qu'attendue mais la fragmentation a probablement été plus efficace sur la faible quantité d'ADNmt d'une cellule.

Sur l'ensemble de ces cellules récoltées, seules 3 cellules m'ont permis d'obtenir une banque séquençable.

IV. Analyse mitochondriale des données RNAseq et 10X Genomics

Du fait du faible rendement (seulement 3 bibliothèques sur 96 puits) lors de l'expérience effectuée sur l'appareil C1, nous avons réorienté notre étude sur l'analyse du génome mitochondrial à partir de données de séquençage d'ARN, disponible par ailleurs. En effet, l'ADN mitochondrial est transcrit dans sa quasi-totalité en ARNm, ARNt et ARNr. Dans un premier temps, afin de vérifier la faisabilité de l'approche, j'ai utilisé des résultats d'échantillons RNAseq analysés dans le cadre du projet GENEPRO auquel je suis associé. L'étude

GENEBRO compare la composition cellulaire et les profils d'expression des brossages nasaux et bronchiques chez des patients asthmatiques, ou rhinitiques à des contrôles sains. Cette première approche avait pour but d'effectuer une évaluation de faisabilité quant aux déterminations d'haplotypes à partir de données RNAseq « réelles ». L'arrivée de l'appareil Chromium de chez 10X Genomics dans notre laboratoire nous permettant un séquençage du transcriptome à l'échelle de la cellule unique sur un grand nombre de cellules m'a ensuite permis d'analyser de nouvelles données, obtenues cette fois-ci au niveau de la cellule unique, et d'effectuer une analyse fine des transcrits mitochondriaux et de leurs séquences associées.

Mon objectif a ensuite été de savoir si les données RNAseq couvraient suffisamment le génome mitochondrial et si ces résultats étaient exploitables pour (1) une détermination de l'haplogroupe des patients et (2) une détermination en cellule unique sur Chromium 10X Genomics. Ce type d'analyse pouvant permettre de détecter facilement d'éventuelles contaminations par des cellules étrangères, en recherchant la présence de plusieurs haplogroupes différents, j'ai testé cette application sur différentes expériences. Mes analyses ont donc été effectuées sur deux tissus (nez/bronche en bulk ARN, sang/nez en bulk ADNmt), que j'ai ensuite comparé avec des données de Chromium obtenues sur des cellules épithéliales respiratoires différenciées in vitro (culture 3D en interface air/liquide). J'ai enfin effectué une analyse de données de Chromium correspondant à : (1) une biopsie de cornet nasale et (2) un mélanome oculaire.

a. Détermination de l'haplogroupe mitochondrial

A partir des données de RNA-seq du projet GENEPRO, j'ai établi l'haplogroupe de chacun des échantillons à l'aide du logiciel Haplogrep (<http://haplogrep.uibk.ac.at/>) (Weissensteiner Hansi et al. 2016). Pour cela, j'ai extrait à partir des séquences nucléotidiques consensus les mutations caractéristiques des différents haplogroupes. Après production de fichiers .fasta, comportant la séquence consensus de chaque échantillon, définie comme la suite nucléotidique prépondérante à partir de l'ensemble des données de séquençage. Haplogrep détermine l'haplogroupe après analyse des SNP (Single Nucleotide Polymorphism) en fournissant également un pourcentage de fiabilité. Ce pourcentage de fiabilité est obtenu à partir du nombre de SNP trouvés par rapport aux SNP spécifiques de chaque haplogroupe. La séquence de référence Cambridge correspond à l'haplogroupe H2a2a. Elle correspond à la

séquence mitochondriale humaine de référence à laquelle toutes les autres séquences sont comparées lors de l'étape d'alignement des lectures.

Echantillons	Haplogroupes	Fiabilité	Origine
PB	W3a1c	98%	Nord est Europe, nord ouest Asie
D163	H2a2a	100%	Proche orient, sud de l'Europe
D159	H2a2a	100%	Proche orient, sud de l'Europe
D139	J1c8	79%	Caucase
D138	H2a2a	63%	Proche orient, sud de l'Europe
D137	H2a2a	100%	Proche orient, sud de l'Europe
D128	H2a2a	73%	Proche orient, sud de l'Europe
D103	H2a2a/J1c8	68%/79%	Caucase
D87	A2r	91%	Asie
D59	H2a2a	83%	Proche orient, sud de l'Europe
D46	H2a2a	100%	Proche orient, sud de l'Europe
B-VIL JU	H3g1	92%	Europe occidental
B-SIC JU	J1c4	98%	Caucase
B-SEV CA	I1a	97%	Nord est de l'Europe
B-ROU VI	T2b	98%	Mésopotamie
B-RIP TH	H1h1	89%	Proche orient, sud de l'Europe
B-RIC RO	H1e1a	100%	Proche orient, sud de l'Europe
B-NEG MO	U6a1b2	92%	Afrique du nord
B-MAR NI	U5b1+16189	95%	Asie occidentale
B-LIV DE	H2b	92%	Proche orient, sud de l'Europe
B-LET LI	T2b	98%	Mésopotamie
B-LEO VA	H8b	94%	Proche orient, sud de l'Europe
B-LEH CH	H1a3c1	95%	Proche orient, sud de l'Europe
B-ISO SA	U5a2b	97%	Asie occidentale
B-DES JU	HV6	96%	Proche orient
B-DAS AL	H5c	94%	Proche orient, sud de l'Europe
B-CHE MA	H1+16311	82%	Proche orient, sud de l'Europe
B-CAS LA	J1c1	100%	Caucase
B-CAR XA	U5b1+16189	97%	Asie occidentale

B-BRE PA	T1a1k	100%	Mésopotamie
B-BOU WA	H1q	94%	Proche orient, sud de l'Europe
B-BOU AD	H7	90%	Proche orient, sud de l'Europe
B-BIA BL	U2e1	93%	Nord est Afrique, moyen orient
B-BEN MI	X2d2	97%	Nord est Europe
B-YAQ YO	H5c	87%	Proche orient, sud de l'Europe
B-GUI AI	H44b	100%	Proche orient, sud de l'Europe

Tableau 2 : Détermination de l'haplogroupe à partir de l'ADNmt provenant de données RNA-seq.

L'haplogroupe de chacun des échantillons a été établi à partir des données de RNAseq à l'aide du logiciel Haplogrep 2. La fiabilité est un pourcentage correspondant au nombre de polymorphismes trouvés sur le nombre de polymorphismes attendues pour un haplogroupe donné. A droite correspond la provenance géographique d'origine de l'haplogroupe.

Ces analyses montrent qu'il est possible de déterminer l'haplogroupe des différents échantillons à partir de données d'échantillons RNAseq.

Pour confirmer l'haplogroupe des patients, j'ai comparé des résultats à partir de deux RNAseq de deux tissus différents provenant d'un même patient. C'est à partir de prélèvements nasaux et de prélèvements bronchiques que j'ai effectué cette comparaison.

Echantillons	Haplogroupes Bronche	Fiabilité	Haplogroupes Nez	Fiabilité
BEN MI	X2d2	97%	X2d2	98%
BIA BL	U2e1	93%	U2e1	94%
BOU AD	H7	90%	H7	90%
BOU WA	H1q	94%	H1q	94%
BRE PA	T1a1k	100%	T1a1k	100%
CAR XA	U5b1+16189	97%	U5b1+16189	97%
CAS LA	J1c1	100%	J1c1	100%
CHE MA	H1+16311	82%	H1+16311	84%
DAS AL	H5c	94%	H5c	94%
DES JU	HV6	96%	HV6	96%
GUI AI	H44b	100%	H44b	100%
ISO SA	U5a2b	97%	U5a2b	97%
LEH CH	H1a3c1	95%	H1a3c1	97%

LEO VA	H8b	94%	H8b	94%
LET LI	T2b	98%	T2b	98%
LIV DE	H2b	92%	H2b	92%
MAR NI	U5b1+16189	95%	U5b1+16189	95%
NEG MO	U6a1b2	92%	U6a1b2	92%
RIC RO	H1e1a	100%	H1e1a	100%
RIP TH	H1h1	89%	H1h1	89%
ROU VI	T2b	98%	T2b	98%
SEV CA	I1a	97%	I1a	97%
SIC JU	J1c4	98%	J1c4	99%
VIL JU	H3g1	92%	H3g1	92%
YAQ YO	H5	87%	H5	87%

Tableau 3 : Comparaison de l'haplogroupe d'un patient provenant de deux tissus différents

J'ai comparé l'haplogroupe des échantillons nez et bronche à partir des données RNAseq du même patient à l'aide du logiciel Haplogrep 2. La fiabilité est un pourcentage correspondant au nombre de polymorphismes trouvés sur le nombre de polymorphismes attendus pour un haplogroupe donné.

On note la correspondance parfaite entre les différents haplogroupes ainsi définis, ce qui établit que l'haplogroupe peut bien être déterminé dans chacun des tissus à partir des données des échantillons RNAseq.

J'ai ensuite analysé les haplogroupes des cellules des voies aériennes en culture 3D, d'un échantillon de mélanome et d'un brossage nasal, à partir de leurs données Chromium respectives :

Echantillons	Haplogroupe
D231	T1a1
D237	L2a1c6
D246	H1e1a
D258	H80
D266	T2
D267	U5b2c
Mélanome 3	J1d6
Brossage PB	W3a1c

Tableau 4 : Haplogroupes des patients séquencés à partir du Chromium 10X Genomics.

J'ai établi l'haplogroupe de chacun des échantillons à partir des données de Chromium à l'aide du logiciel Haplogrep 2.

Après avoir déterminé l'haplogroupe de chacun des patients, à partir d'un regroupement de toutes les séquences disponibles, j'ai voulu savoir s'il était également possible de déterminer l'haplogroupe de chaque cellule isolée. A partir de toutes les séquences mitochondriales détectées de chaque cellule, nous pouvons reconstruire la séquence consensus et ainsi déterminer leurs haplogroupes à l'aide du programme Haplogrep 2.

Identification de la cellule	Haplogroupe	Fiabilité
AAACCTGCACCAACCG_1	W3a1c	80%
AAACGGGCAAAGTCAA_1	W3a1c	74%
AAATGCCACATCCAA_1	W3a1c	83%
AACACGTGTATGAAAC_1	W3a1c	78%
AACCATGGTCGACTGC_1	W3a1c	72%
ACACGGGAGAATCTCC_1	W3a1c	74%
ACACGGGCATGTTGAC_1	W3a1c	77%
ACAGCTAGTGCACGAA_1	W3a1c	74%
ACATACGCATCGATGT_1	W3a1c	74%
ACAGTAAGCTGGAAC_1	W3a1c	83%
ACCAGTAGTGAACCTT_1	W3a1c	84%
ACCTTTATCTCGCTTG_1	W3a	73%
ACGAGGAGTGGTACAG_1	W3a1c	80%
ACGTCAAAGATAGCAT_1	W3a1c	72%
ACTTACTCAGCGTCCA_1	W3a1c	74%
ACTTTCATCGCGGATC_1	W3a1c	78%
AGAATAGCAAGTAGTA_1	W3a1c	73%
AGACGTTACGCGTAAG_1	W3a1c	61%
AGAGCGACACAAGCCC_1	W3a1c	74%

Tableau 5 : Haplogroupes des cellules du patient PB

Ceci est l'haplogroupe à partir de la séquence mitochondrial de chacune des cellules du patient PB à l'aide du logiciel Haplogrep 2. La fiabilité est un pourcentage correspondant au nombre de polymorphismes trouvés sur le nombre de polymorphisme attendu pour un haplogroupe donné.

Les traitements présentés concernent uniquement le patient PB pour lequel je dispose également des données de séquençage de l'ADNmt à partir des échantillons Nez et Sang. La séquence consensus a été obtenue pour chacune des 290 cellules disponibles. La détermination de l'haplogroupe W3a1c pour la plupart des cellules montre qu'il est possible d'obtenir un haplogroupe correct à partir d'une seule cellule. Le pourcentage de fiabilité est déterminé en fonction du nombre de mutations correctement identifiées par rapport au nombre de mutations spécifiques de l'haplogroupe.

b. Recherche de contamination

Après avoir criblé l'haplogroupe des cellules, dans différents projets de cellule unique avec le Chromium effectués sur la plateforme de séquençage et dans l'équipe de recherche, je me suis aperçu que certaines cellules ne possédaient pas le même haplogroupe que celui désigné pour le patient. Dans le cas de 8 échantillons, l'haplogroupe H2a2a1 apparaissait dans une cinquantaine de cellules avec un pourcentage de fiabilité proche de 50%. Cet haplogroupe H2a2a1 était particulièrement surprenant dans le cas de l'échantillon de mélanome oculaire d'un projet effectué sur la plateforme puisque l'haplogroupe de l'individu correspondait à J1d6.

Voici les résultats obtenues :

Code à barre cellulaire	Haplogroupe	Fiabilité
ACATACGAGGGCTCTC-1	J1d6	69%
ACATCAGCAAGGTTCT-1	J1d6	67%
ACATCAGGTTTAGCTG-1	J1d6	58%
ACATGGTAGTGGCACA-1	H2a2a1	50%
ACATGGTCACAAGACG-1	J1d6	69%
ACATGGTGTAGCGATG-1	H1	54%
ACATGGTGTCCAACA-1	J1d6	66%
ACCAGTACACGGCTAC-1	J1d6	60%
ACCAGTAGTTAAGAAC-1	J1d6	77%
ACCACTTCAAACAAG-1	J1d6	61%
CGTTAGATCCTTTCTC-1	D4q	53%
ACCGTAACAAGTTGTC-1	J1d6	75%
ACCGTAATCACTATTC-1	J1d6	86%
ACCGTAATCTAACTGG-1	H2a2a1	50%
ACGAGCCCAGACAAGC-1	J1d6	82%
ACGAGGACACTTGGAT-1	J1d6	61%
ACGAGGAGTAAGTAGT-1	H2a2a1	50%

Tableau 6 : Haplogroupes des cellules provenant de l'échantillon de mélanome oculaire
Les séquences mitochondriales consensus sont utilisées pour déterminer l'haplogroupe de chacune des cellules du mélanome avec l'outil Haplogrep 2. La fiabilité est un pourcentage correspondant au nombre de polymorphismes trouvés sur le nombre de polymorphismes attendus pour un haplogroupe donné.

Nous constatons donc la présence de cellules ayant l'haplogroupe J1d6 et de cellules avec les haplogroupes H2a2a1 et D4q. J'ai initialement pensé à une contamination cellulaire lors de la manipulation. J'ai alors représenté les cellules en fonction de leur profil d'expression dans un graphe t-SNE. Ce type de graphe permet de représenter l'ensemble des cellules en fonction

de leur profil d'expression, en utilisant une projection sur un plan. En effet, une valeur va être attribuée à chaque cellule correspondant à leur profil d'expression et positionnera la cellule dans ce plan permettant une meilleure visualisation. La proximité entre cellules indique une ressemblance des profils d'expressions. Sur la figure 29, on voit ainsi apparaître plusieurs groupes de cellules associées principalement à l'haplogroupe J1d6. On note toutefois également la présence de cellules d'haplogroupes H et D4q.

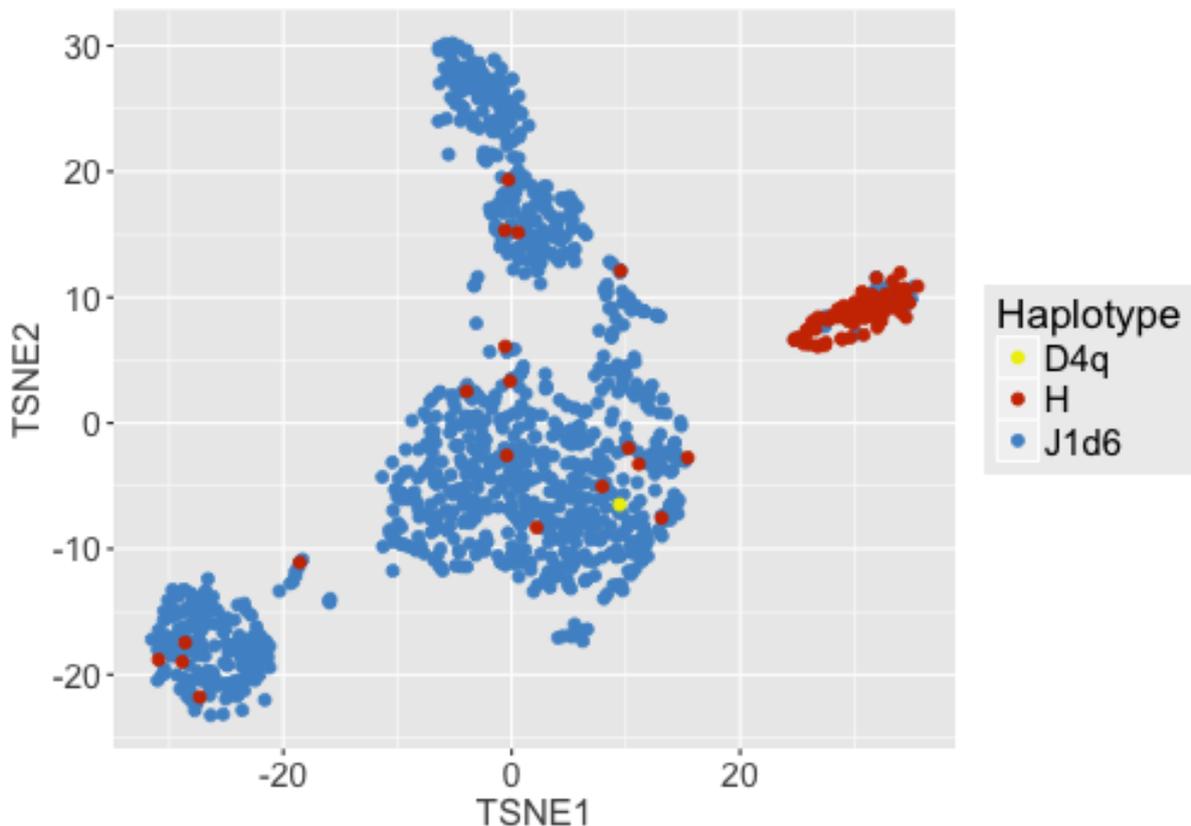


Figure 29 : Représentation t-SNE des cellules d'haplogroupe H et D4q dans l'échantillon mélanome

Ce mode de représentation permet de comparer les différentes cellules présentes (représentées par des points) en fonction de leur profil d'expression. Deux cellules adjacentes présentent des profils d'expression proches. Les points bleus correspondent aux cellules ayant l'haplogroupe J1d6, les points rouges aux cellules avec l'haplogroupe H et le point vert est la cellule avec l'haplogroupe D4q

Nous pouvons remarquer que la grande majorité des cellules H se regroupe dans un cluster à droite du graphique. En analysant plus précisément les cellules ayant l'haplogroupe H, j'ai remarqué que ce sont celles ayant le moins de couverture, c'est-à-dire que peu de séquences ont été obtenues pour ces cellules. La couverture de la séquence est insuffisante pour

déterminer correctement les 16569 paires de bases de l'ADN mitochondrial. Lors de la détermination de l'haplogroupe, le logiciel HaploGrep2 prévoit des mutations lui permettant d'orienter le résultat et lorsque l'ensemble des mutations attendues fait défaut, il affecte par défaut l'haplogroupe H2a2a1 qui correspond à la séquence de référence utilisée par le logiciel. Ceci explique la présence de deux haplogroupes extrêmement différents, qui ne font pas partie du même sous-groupe dans l'arbre phylogénétique (www.phylotree.org). Des mutations supplémentaires auraient donc conduit à attribuer par erreur l'haplogroupe D4q que ne possèdent pas les autres cellules. Deux explications sont susceptibles d'expliquer les résultats : (1) il y a eu une contamination par des cellules appartenant à une personne d'haplogroupe D4q, ce qui semble impossible au vu du nombre de cellules D4q (1 cellule / 1067 cellules); (2) il s'agit de cellules (cancéreuses ou pas) ayant accumulé de nombreuses néo-mutations qui aurait modifié l'haplogroupe. L'hypothèse la plus parcimonieuse étant celle se référant à une lecture d'insuffisante qualité (53% de fiabilité), c'est celle-ci que nous invoquons. Ce résultat est intéressant puisqu'il suggère de mettre en place ce contrôle de façon systématique pour vérifier a posteriori les mesures, et s'assurer des bonnes pratiques de laboratoire. Dans ce cas, on peut s'assurer à la fois de l'absence de contamination tout en visualisant les groupes de cellules ayant une faible couverture nucléotidique.

c. Analyse de l'hétérogénéité

Nos premières comparaisons ont montré (Figure 20) que les niveaux d'hétéroplasmie restent relativement constants entre tissus d'origine embryonnaire différente. J'ai toutefois voulu préciser ces niveaux, en me positionnant désormais au niveau de cellules uniques. J'ai toutefois commencé mon analyse par celle des résultats obtenus à partir des échantillons RNAseq dans le cadre du projet GENEPRO, en poursuivant la comparaison des deux prélèvements effectués à partir d'un même patient, dans le nez et la bronche. Le tableau 7 récapitule les 6 bases présentant la plus forte hétérogénéité de séquençage. Ces 6 bases différentes sont situées en positions 310, 405, 432, 439, 440 et 442 dans le génome mitochondrial

Position	N1-ESTMA				B1-ESTMA			
	%A	%T	%C	%G	%A	%T	%C	%G
310	0,3	46,4	47,9	5,4	0,6	42,3	52	5,1
405	0	100	0	0	5,5	94,5	0	0

432	89,2	0	8,4	2,4	94,8	3,5	0	1,7
439	83,3	0	16,7	0	89,8	0	10,2	0
440	90	0	7,8	2,2	96,5	0	3,5	0
442	7,6	79,3	0	13,1	0	93,2	1,7	5,1

Tableau 7 : Comparaison du génome mitochondrial à partir de tissus nasaux et bronchiques.

Les données de séquençage de tissus nasaux et bronchiques issues du même patient ont été analysées et comparées. Parmi les polymorphismes, six positions nucléotidiques se sont démarquées pour lesquelles on observe une différence dans la répartition des bases. Nous présentons dans ce tableau le pourcentage de chacune des bases pour chaque position.

Dans le tableau 7, nous observons que plusieurs nucléotides distincts peuvent être associés à une même position. Par exemple, en position 405, il y a environ 5% de A et 95% de T dans l'échantillon B1-ESTMA, mais 100% de T dans l'échantillon N1-ESTMA. De façon générale, les cases jaunes indiquent des pourcentages variables d'une base, indicatifs d'une différence entre les deux tissus. Sur ces échantillons, nous observons donc une répartition différente des bases entre deux tissus d'un même individu, qui reste toutefois marginale. Les séquences obtenues à partir des données transcriptomiques couvrant l'essentiel du génome mitochondrial, le nombre de bases chez lesquelles ce phénomène est observé reste faible. Nous constatons donc une très faible variabilité mitochondriale entre ces deux tissus distincts issus d'un même individu : les seules variations observées ne concernent jamais la base la plus abondante, qui reste inchangée, et affecte de façon marginale les deuxièmes bases les plus abondantes (hétéroplasmie).

Afin de préciser l'ampleur de ce phénomène, j'ai compilé les pourcentages des différentes bases de l'ensemble des patients du projet GENEPRO analysés pour ces mêmes bases. On observe des différences entre les patients et parfois quelques différences entre les deux tissus mais dans la totalité des cas, le nombre de séquences lues est alors trop faible. Ces résultats suggèrent davantage des erreurs de séquençage qu'une réelle hétéroplasmie. Les résultats précédents réfutent l'hypothèse selon laquelle des différences d'hétéroplasmie seraient observées entre types cellulaires différents, en tout cas au niveau global.

Une situation différente pouvant apparaître au niveau de la cellule unique, je me suis intéressé aux données transcriptomiques du patient PB pour lequel des informations sur la séquence de l'ADNmt et sur le transcriptome mitochondrial avec le Chromium de 10X Genomics étaient disponibles.

Ces données ont été représentées après projection dans un graphe t-SNE, qui permet de visualiser dans une représentation en deux dimensions les cellules en fonction de leur profil d'expression. Ce profil « t-SNE » utilise un mode de visualisation dans lequel les cellules ayant la plus grande similarité en termes de profils d'expression seront positionnées à proximité l'une de l'autre.

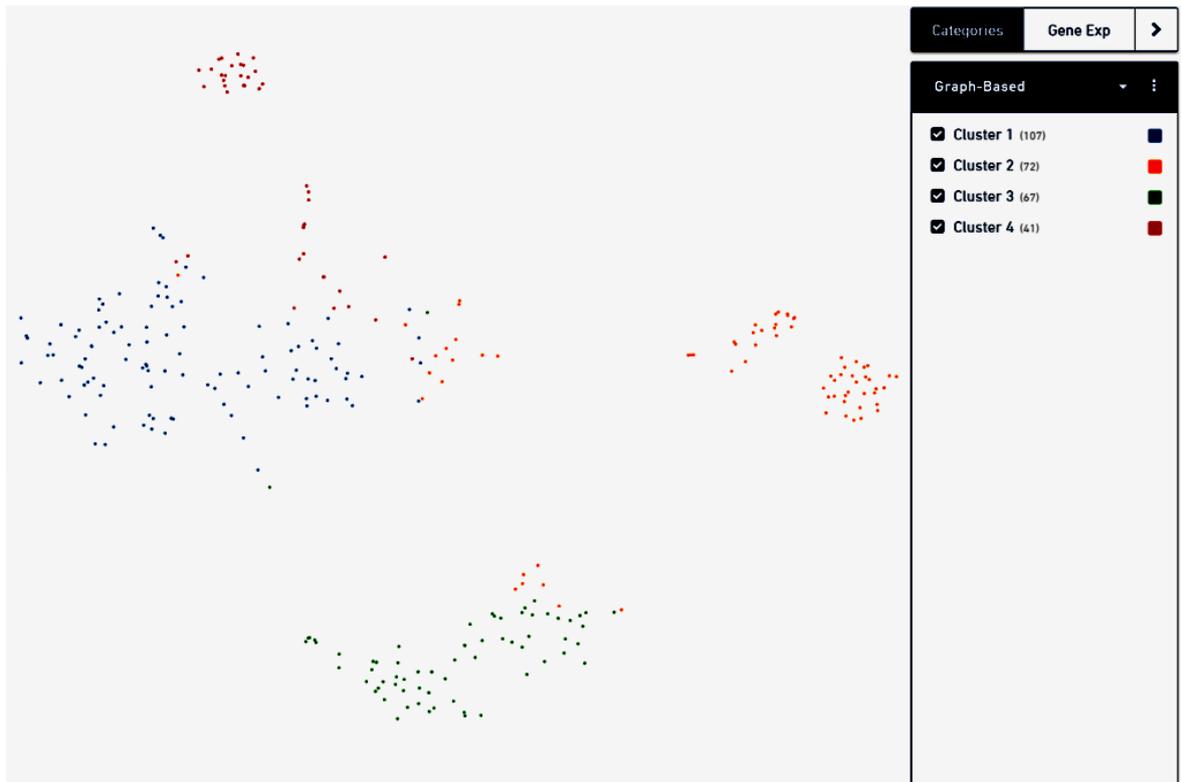


Figure 30 : Représentation graphique de la répartition cellulaire de l'échantillon PB

Les cellules sont regroupées entre elles par leur similitude d'expression génique et forment des clusters. Le logiciel d'analyse définit par défaut 4 clusters de cellules pour le brosseur nasal du patient PB, qui correspond aux cellules à mucus (MUC5AC) pour le cluster 1 en bleu, le cluster 2 en orange correspond aux hématies, le cluster 3 en vert correspond aux cellules multiciliées et le cluster 4 en rouge, correspond aux cellules lymphocytaires (CCL5)

Cette représentation fait apparaître des groupes distincts de cellules, qui peuvent correspondre à des types cellulaires différents. De façon intéressante, l'un des deux groupes identifiés présente de hauts niveaux d'expression pour le gène Muc5AC et correspond à des cellules épithéliales respiratoires (cellules à mucus) (Annexe 3). A contrario, un deuxième groupe exprime des niveaux élevés de marqueurs lymphocytaires. Il devenait donc théoriquement possible à partir de cet échantillon de disposer de mesures à l'échelle de la cellule unique de cellules de même origine embryonnaire que les échantillons de nez et de sang séquencés ci-

dessus. A partir de cette répartition cellulaire, nous avons extrait les fichiers .fasta et .bam de chacun de ces clusters pour comparer leur séquence mitochondriale consensus.

Le faible nombre de séquences lues pour le groupe cellulaire lymphocytaire CCL5 rend difficiles les interprétations des données associées. Deux des trois autres clusters étant également mal couverts au niveau du séquençage, je n'ai gardé que le groupe des cellules à mucus. J'ai répertorié toutes les positions des variants présents dans le nez, le sang et le groupe cellulaire MUC5AC puis je les ai comparés.

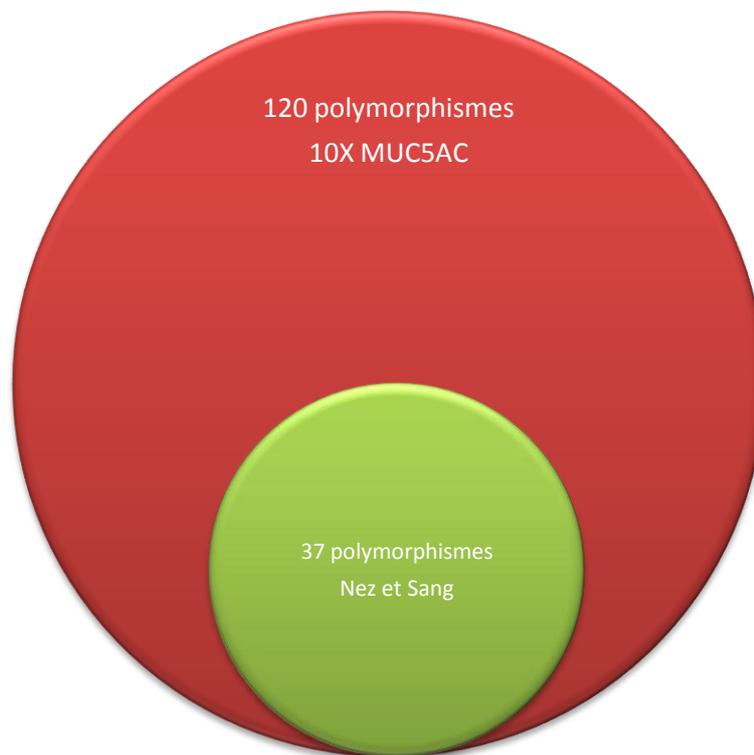


Figure 31 : Comparaison du nombre de polymorphismes entre le nez, le sang et le cluster MUC5AC

Ce tableau reprend le nombre de mutations trouvées dans le séquençage d'ADNmt du Nez et du Sang et les données Chromium sur le cluster MUC5AC. Nous observons 37 polymorphismes dans le nez et le sang et 157 polymorphismes dont 120 ne sont retrouvés que dans les données Chromium.

J'identifie 37 polymorphismes communs entre le sang et le nez. Pour le groupe MUC5AC, ces 37 mutations ont été également retrouvées et correspondent en réalité aux polymorphismes de l'haplogroupe W3a1c. Par contre les autres mutations ont été triées afin de ne conserver que les polymorphismes ayant suffisamment de séquences lues pour

discriminer ceux qui pourraient être dus à une erreur de lecture lors du séquençage. Sur les 120 polymorphismes observés, seules 18 mutations ont été retenues.

Une variation fréquemment observée correspond à la position 2617 qui se trouve sur le gène codant l'ARNr 16S. Cette position est également détectée dans d'autres expériences avec le Chromium, effectuées à partir d'autres patients. Les prélèvements étant des brossages nasaux, la question était de savoir si la mutation pouvait être associée à l'un des types cellulaires détectés, tel que les cellules positives pour Muc5ac (cellule sécrétrice de mucus). Aucune variation au niveau de cette position n'est détectable dans les deux séquençages d'ADNmt (Figure 32).

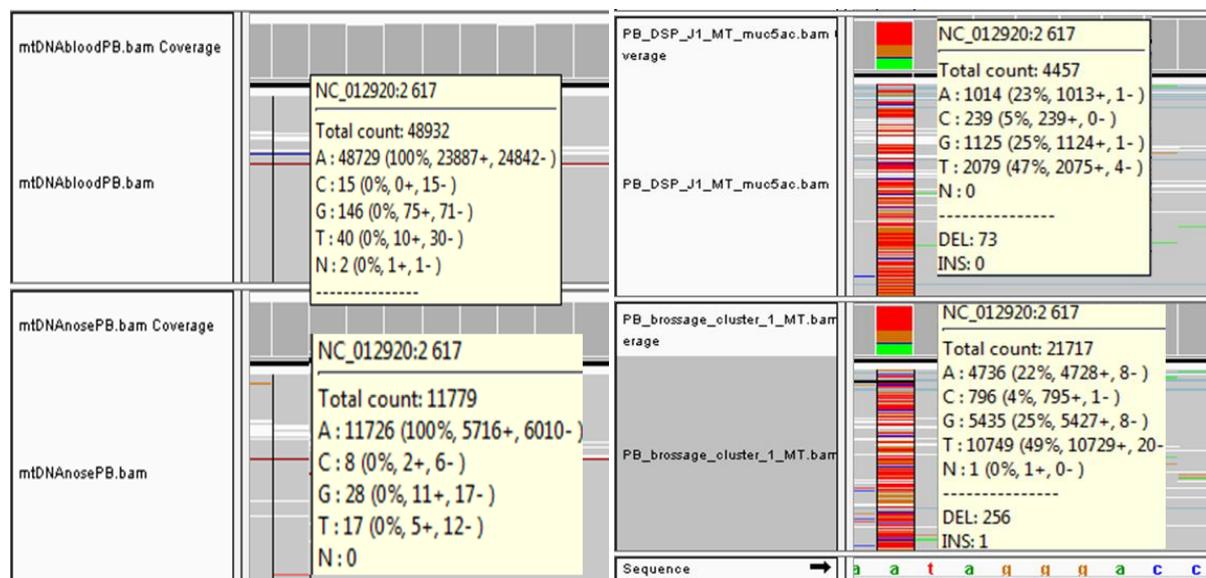


Figure 32 : Représentation nucléotidique avec IGV à la position 2617 de l'ADNmt
Les séquences lues et traitées sont alignées sur le génome mitochondrial de référence. La position 2617 semble montrer une différence dans la répartition des différentes bases pour les clusters provenant de données de Chromium mais aucune mutation n'est présente dans les données RNAseq.

Le même polymorphisme, observé à cette même position, a été trouvé dans plusieurs autres projets de Chromium appartenant à des individus différents, dans des types cellulaires également différents. L'hypothèse la plus probable est donc que cette variation soit en réalité un artefact lié à la technique plutôt qu'un véritable polymorphisme.

Je me suis également rendu compte que toutes les autres variations fréquemment identifiées sur cellule unique étaient situées au début ou en fin de brin séquencé. Le polymorphisme de la position 1338 (Figure 21) n'est retrouvé ni dans les résultats Chromium du patient PB, ni dans la biopsie nasale, ni dans les résultats RNAseq. Il est à noter que cette hétéroplasmie

apparaissait davantage dans les lignées sanguines, et qu'une analyse en cellule unique sur ces cellules permettrait sans doute de préciser plus exactement cette variation.

Certains cornets nasaux ont été traités au laboratoire en interface Air-Liquide afin de favoriser la différenciation des cellules basales en cellules multiciliées. Les mêmes cellules sont restées dans cette interface air/liquide pendant 2 temps différents (2 jours et 22 jours) J'ai donc comparé les données issues d'une même culture après différents temps de différenciation en interface air-liquide pour voir si cette situation permettait de suivre l'apparition de certaines variations (Taylor et al., 2003).

Pour cela, j'ai trié pour chacun des clusters, (Annexe 4 et 5) pour toutes les cellules, par ordre croissant toutes les positions qui présentaient une mutation. J'ai comptabilisé pour chaque position de chaque cellule le nombre de fois où une base était identique à celle de la base de référence. J'ai déterminé un nombre conséquent de positions qui possédaient une mutation dans au moins une cellule. Les positions mutées associées au plus grand nombre de cellules correspondaient principalement aux polymorphismes de l'haplogroupe.

Mutation D246 Ali2	Mutation D246 Ali22
16093	750
4272	
2832	
3092	
8273	
13041	

Tableau 8 : Mutation retrouvée dans une seule condition

Dans la colonne de gauche, ce sont les mutations présentes dans au moins une cellule uniquement de la condition ali2 du patient D246. Dans la condition ali22 du patient D246, seulement une mutation est retrouvée à la position 750 et n'est pas présente dans la condition ali2 du même patient.

Après avoir mis de côté les mutations de l'haplogroupe, un nombre restreint de positions est apparue. Ces mutations correspondent principalement à l'échantillon ayant passé deux jours en condition air-liquide mais n'apparaissent plus dans la même culture après 22 jours de différenciation. J'ai alors cherché à savoir si ces mutations étaient localisées dans un groupe cellulaire ou si elles étaient réparties de manière aléatoire.

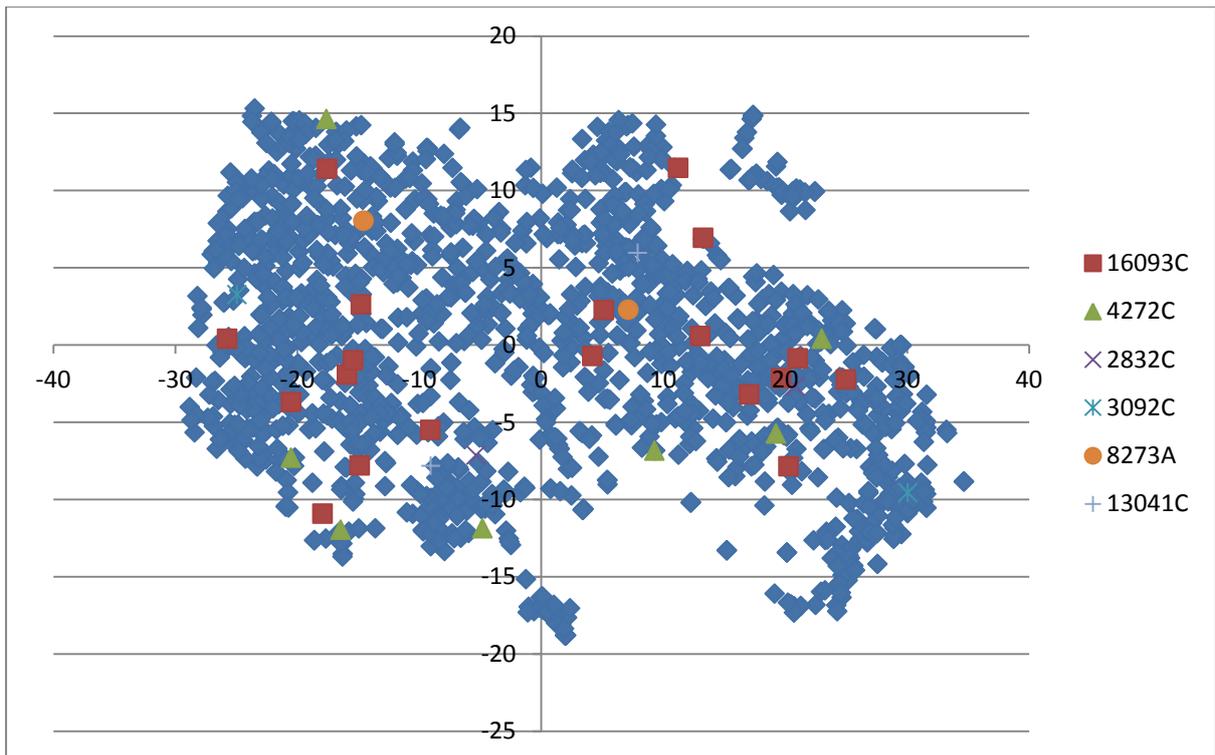


Figure 33 : Répartition des cellules avec une mutation dans l'échantillon D246 ali2

Les positions t-SNE ont été utilisées pour obtenir la même représentation que sur l'outil cloupe. Les cellules sans les mutations sont représentées en bleu et celles ayant une des mutations sont représentées d'une autre couleur.

On peut observer dans la figure 33 que les cellules avec la même mutation ne sont pas regroupées dans un seul et même cluster. Cela pourrait provenir de néo-mutations apparues de manière aléatoire (hétérogénéité inter-cellulaire) mais sans réel lien entre-elle. Plus probablement, cela pourrait correspondre aussi à des erreurs de séquençage.

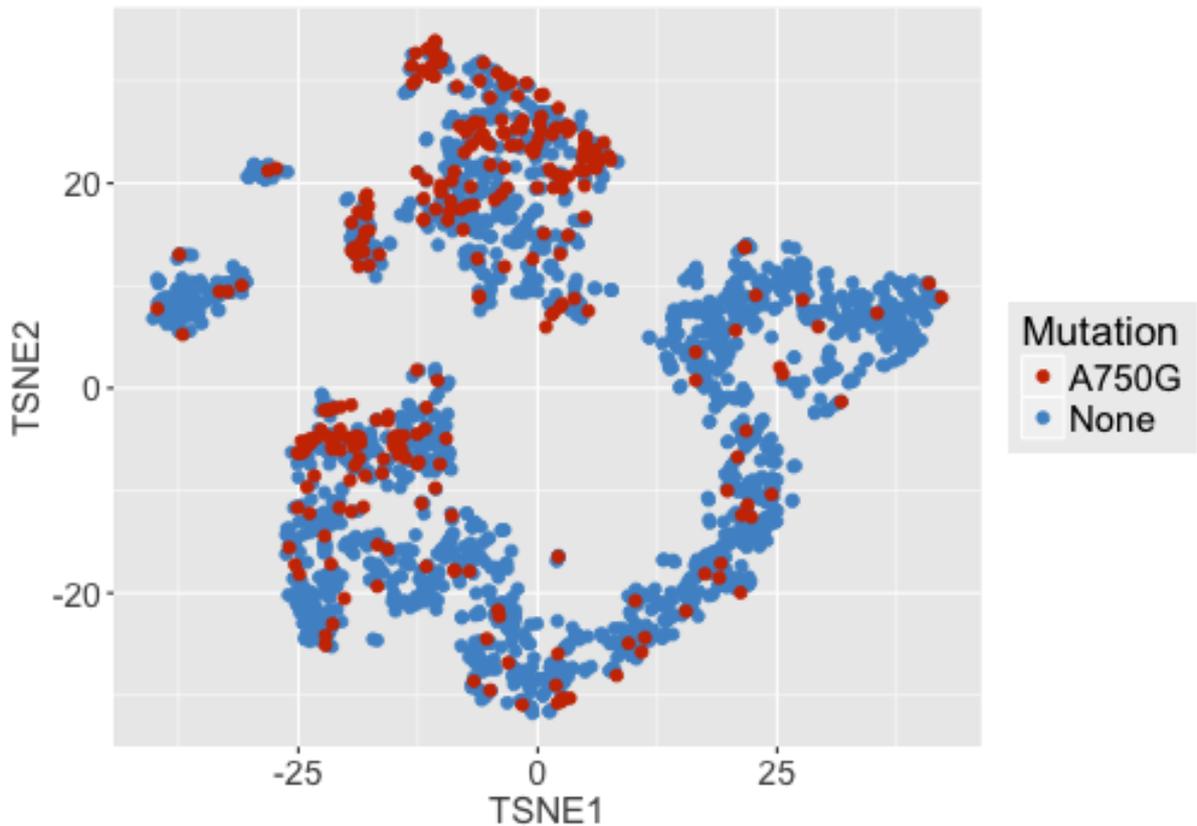


Figure 34 : Répartition des cellules avec la mutation A750G dans l'échantillon D246 ali22

Les positions t-SNE ont été utilisées pour obtenir la même représentation que sur l'outil cloupe. Les cellules possédant la mutation A750G sont en rouge et les cellules sans la mutation sont en bleu. La partie à droite correspond aux cellules basales, au centre ce sont les cellules prédifférenciées et le cluster à gauche correspond aux cellules multiciliées.

Concernant l'échantillon D246 ali22, on retrouve la mutation A750G principalement dans les cellules prédifférenciées. Ce schéma laisse penser que l'apparition de la mutation A750G serait induite au cours de la différenciation.

Afin de disposer d'une meilleure couverture de la séquence mitochondriale, j'ai ensuite regroupé les cellules en cluster de 200 à 500 cellules, en fonction de leurs profils d'expressions. Voici la représentation graphique du patient D258 (Figure 35).

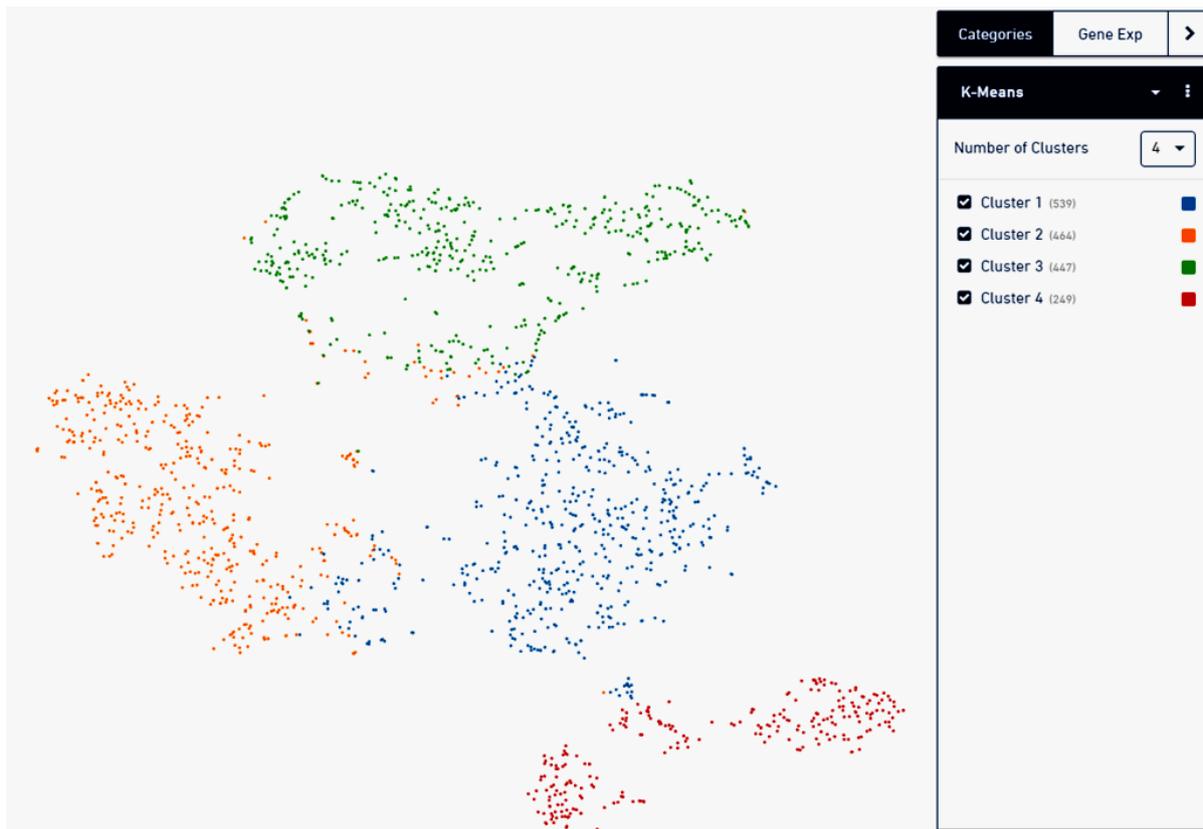


Figure 35 : Répartition cellulaire de l'échantillon de biopsie de cornets nasaux

Ce mode de représentation permet de comparer les différentes cellules présentes (représentées par des points) en fonction de leur profil d'expression. Les cellules sont regroupées entre elles par leur similitude d'expression génique et forment des clusters. On observe 4 clusters de cellules pour l'échantillon de biopsie de cornets nasaux. Les clusters 3 et 2 correspondent aux cellules basales, le cluster 1 aux cellules prédifférenciées et le cluster 4 correspond aux cellules multiciliées.

Les différents clusters déterminent le schéma de différenciation à partir des cellules basales en cellules multiciliées. En utilisant les marqueurs génétiques, on observe un parcours de différenciation. On retrouve principalement le gène KRT5 dans les clusters 2 et 3 qui un marqueur établi des cellules basales (Annexe 6). Dans le cluster 4, DTHD1 s'exprime principalement dans les cellules multiciliées. Le cluster 1 serait quant à lui une étape de prédifférenciation entre les cellules basales et les cellules multiciliées. J'ai recherché parmi chacun des clusters des variations de séquences un petit nombre de cellules qui seraient présentes dans chacun des groupes cellulaires et qui permettraient un lignage cellulaire depuis les cellules basales jusqu'aux cellules les plus différenciées.

Pour cela, j'ai étudié, sur le logiciel IGV chacune des positions présentant une variation du pourcentage de substitution pour chacun des clusters à partir des fichiers .bam. Après avoir exclu les mutations de l'haplogroupe, quelques mutations avec un pourcentage différents et

présentant parfois une variation entre clusters ont été représentés sur les graphiques ci-dessous.

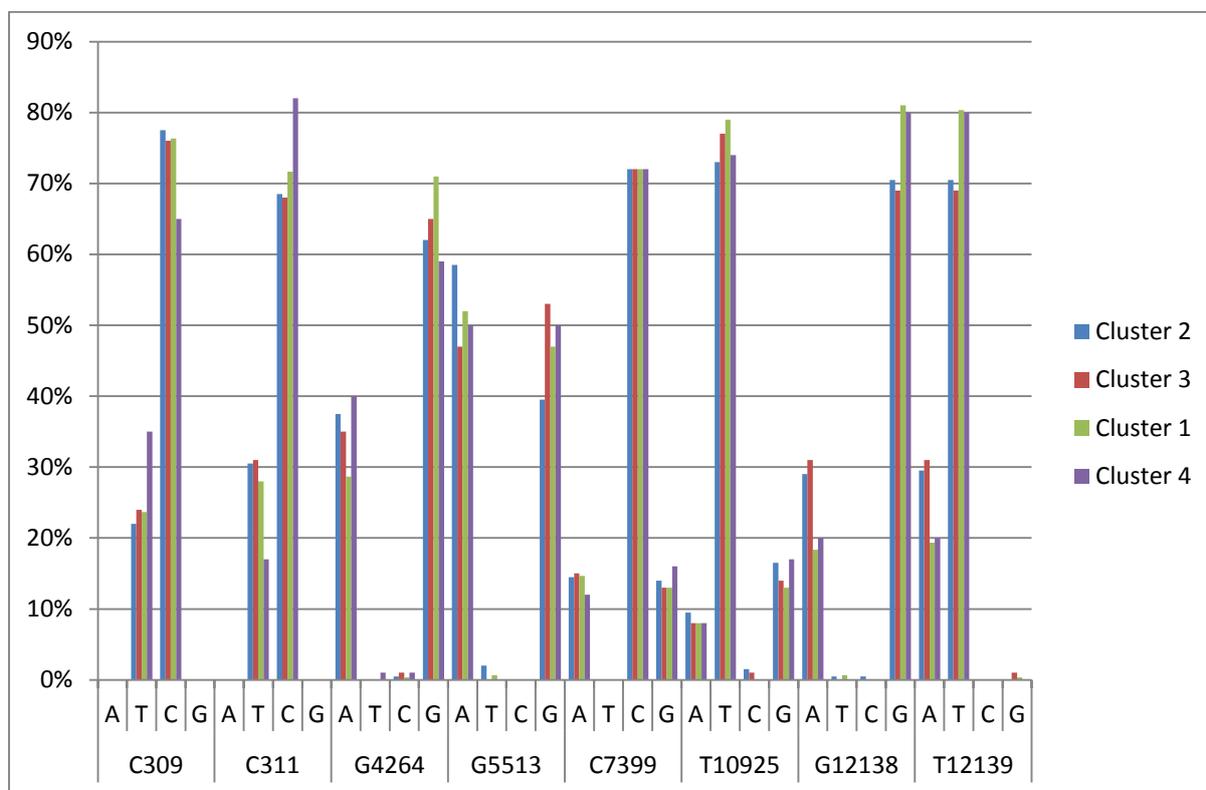


Figure 36 : Comparaison de polymorphismes entre les différents clusters

Dans cette figure, la répartition nucléotidique à différentes positions est comparée entre les 4 clusters. Les clusters 2 et 3 de couleurs bleu et rouge correspondent aux cellules basales, le cluster 1 de couleur verte aux cellules prédifférenciées et en violet le cluster 3 aux cellules multiciliées. Le pourcentage des 4 bases de chacun des clusters sont représentées pour 8 positions différentes (309, 311, 4264, 7399, 10925, 10925 et 12139).

On peut observer des variations sur la figure 36 comme la position 311 où il y a une diminution du pourcentage de la base T. Ces variants laissent supposer qu'il y a bien une hétéroplasmie associée à certains clusters mais la faible différence entre chacun d'eux ne me permet pas de conclure que cet écart de pourcentage des bases est caractéristique de l'état de différenciation.

Enfin une analyse des mutations, pour toutes les cellules d'une même condition, a été effectuée. J'ai comptabilisé pour chaque position de chaque cellule le nombre de fois où la base de la cellule était identique à celle de la base de référence. Il m'est apparu un nombre conséquent de positions qui possédaient une mutation sur au moins une cellule. Les positions mutées possédant un grand nombre de cellules étaient principalement les polymorphismes

correspondant à l'haplogroupe. Après avoir mis de côté les mutations de l'haplogroupe, quelques positions sont apparues.

A partir de ces résultats, nous pouvons observer une trentaine de mutation présentant au moins une seule cellule ayant une mutation (Annexe 7). Chaque position a été vérifiée sur IGV afin de contrôler la qualité de la séquence. Par exemple, en position 295, j'ai détecté plusieurs séquences portant un T au lieu d'un C avec plus de 16000 séquences lues.

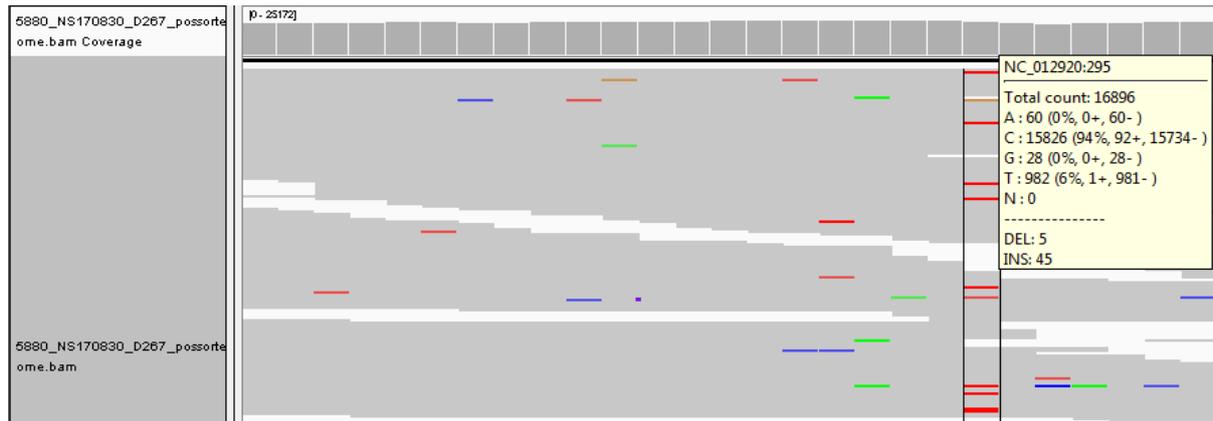


Figure 37 : Hétérogénéité de séquences sur IGV avec le patient D267

Les séquences lues et traitées sont alignées sur le génome mitochondrial de référence. La position 295 semble montrer une différence dans la répartition des différentes bases.

Comme on peut le voir sur la figure ci-dessus, il y a 6 % de séquences lues avec une substitution de la base T à la place de la base A de référence. A partir de cela, j'ai sélectionné les cellules portant la mutation et j'ai essayé de voir où elles se positionnaient lors de la répartition cellulaire. Seulement deux cellules sur 1600 cellules sont ressorties avec un T à la place d'un C.

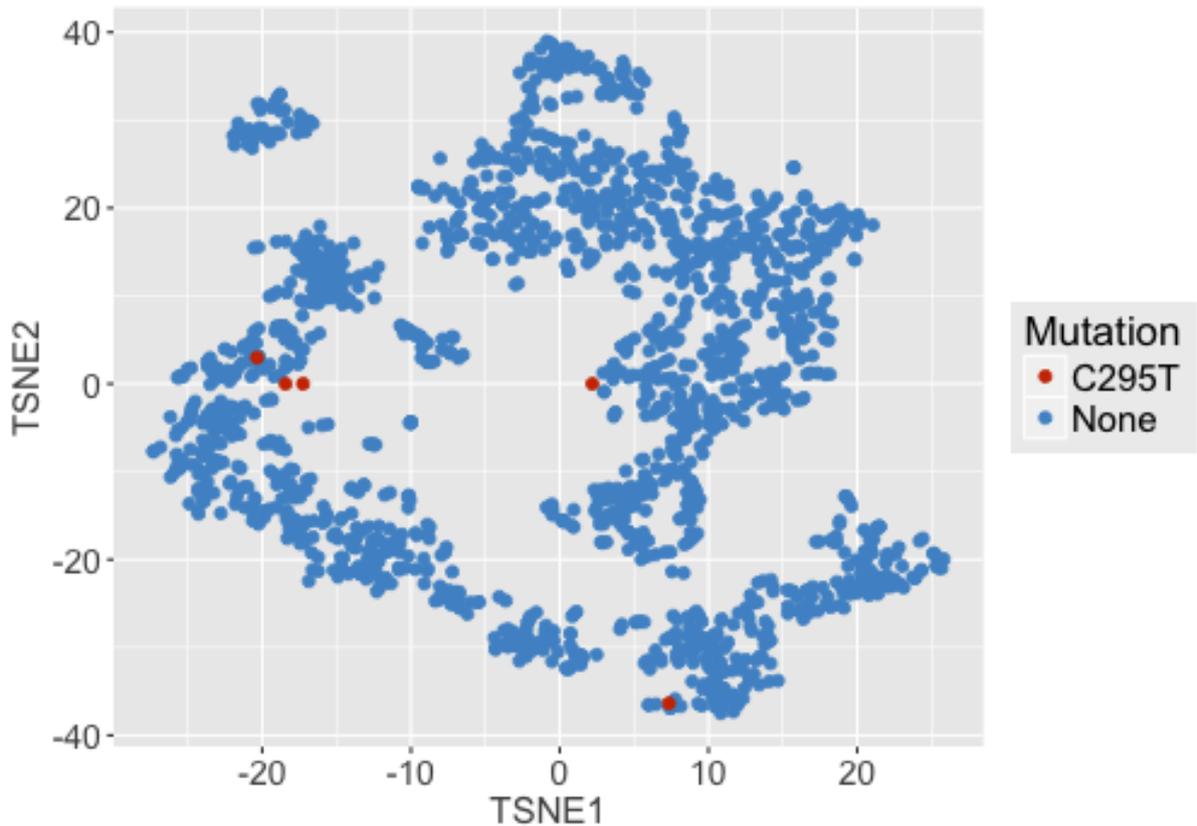


Figure 38 : Position des cellules mutées de l'échantillon D267

Les positions t-SNE ont été utilisées pour obtenir la même représentation que sur l'outil cloupe (Annexe 8). Cinq cellules apparaissent avec la mutation C295T mais dans des clusters différents.

Nous pouvons observer que les cinq cellules n'apparaissent pas dans le même cluster. La raison pour laquelle seulement cinq cellules sont apparues avec la mutation alors que 6% des séquences lues ont un T à la place d'un C, est probablement due à la création de la séquence consensus de chacune des cellules. Si pour une seule et même cellule, le nombre de séquences lues avec la mutation est moins important que le nombre de séquences lues sans la mutation, le programme informatique va mettre la base présente majoritairement. Cela ne veut néanmoins pas signifier qu'il n'y a pas du tout la mutation dans cette cellule.

Au vu des résultats obtenus, l'hypothèse initiale qui était que le suivi de l'hétéroplasmie permettrait de suivre le lignage cellulaire semble devoir être abandonné dans le cas d'un tissu normal, en raison de sa rareté. Une dernière exploitation possible résidait toutefois dans l'analyse de lignées cancéreuses, chez lesquelles des taux de mutations plus importants seraient observés. Pour finir, j'ai donc analysé des données sur cellule unique de mélanome oculaire. L'objectif a été de chercher des mutations spécifiques à une lignée ou à un groupe de cellules cancéreuses permettant une discrimination mutationnelle entre chaque lignée.

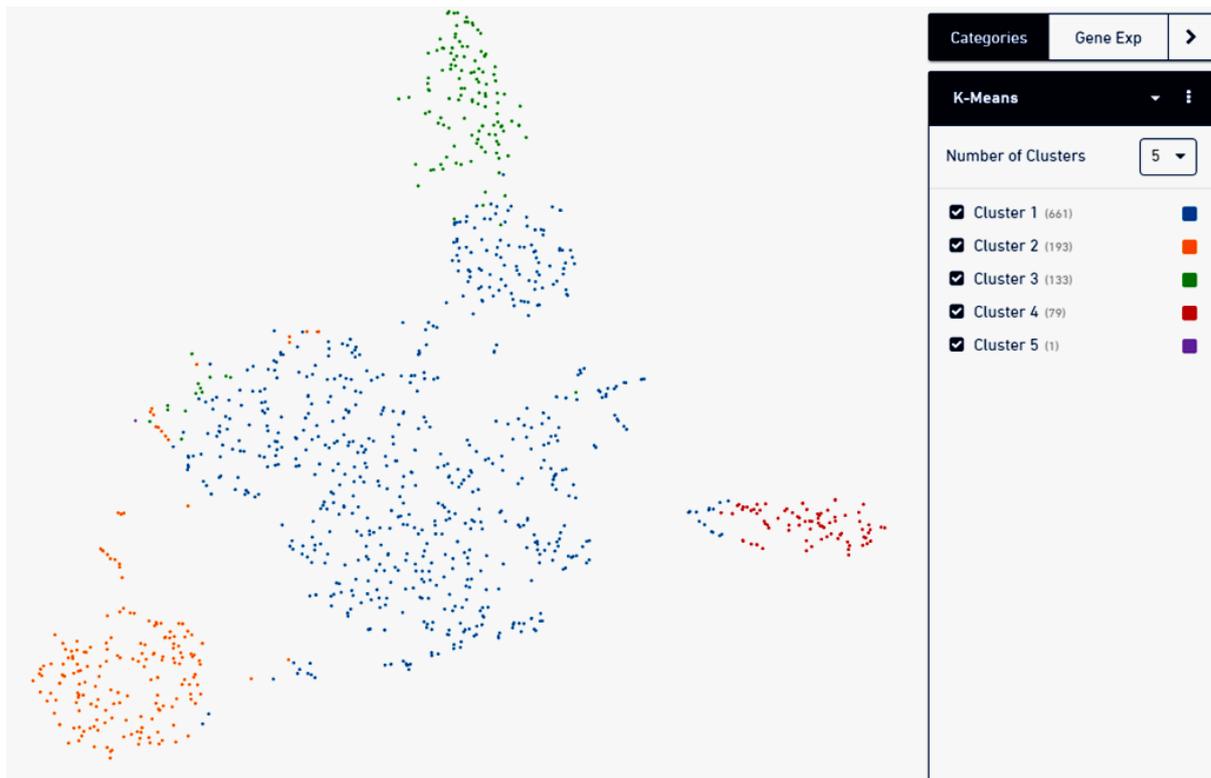


Figure 39 : Répartition cellulaire de l'échantillon de mélanome

Les cellules sont regroupées entre elles par leur similitude d'expression et forment des clusters. On observe 5 clusters de cellules pour le mélanome.

J'ai exclu les cellules ayant l'haplogroupe H et D4q et j'ai recherché des variations possible de séquences sur chacune des 1000 cellules restantes. Après avoir mis de côté les mutations de l'haplogroupe, certaines bases supplémentaires sont apparues. De façon notable, cinq cellules présentent les mêmes variations sur 17 positions différentes (263, 8364, 10398, 4769, 9168, 8860, 8360, 8362, 13708, 11719, 7789, 11251, 12612, 15452, 15326, 7028, 8361, 1476). Je les ai donc extraits pour observer leur disposition sur la représentation graphique des cellules.

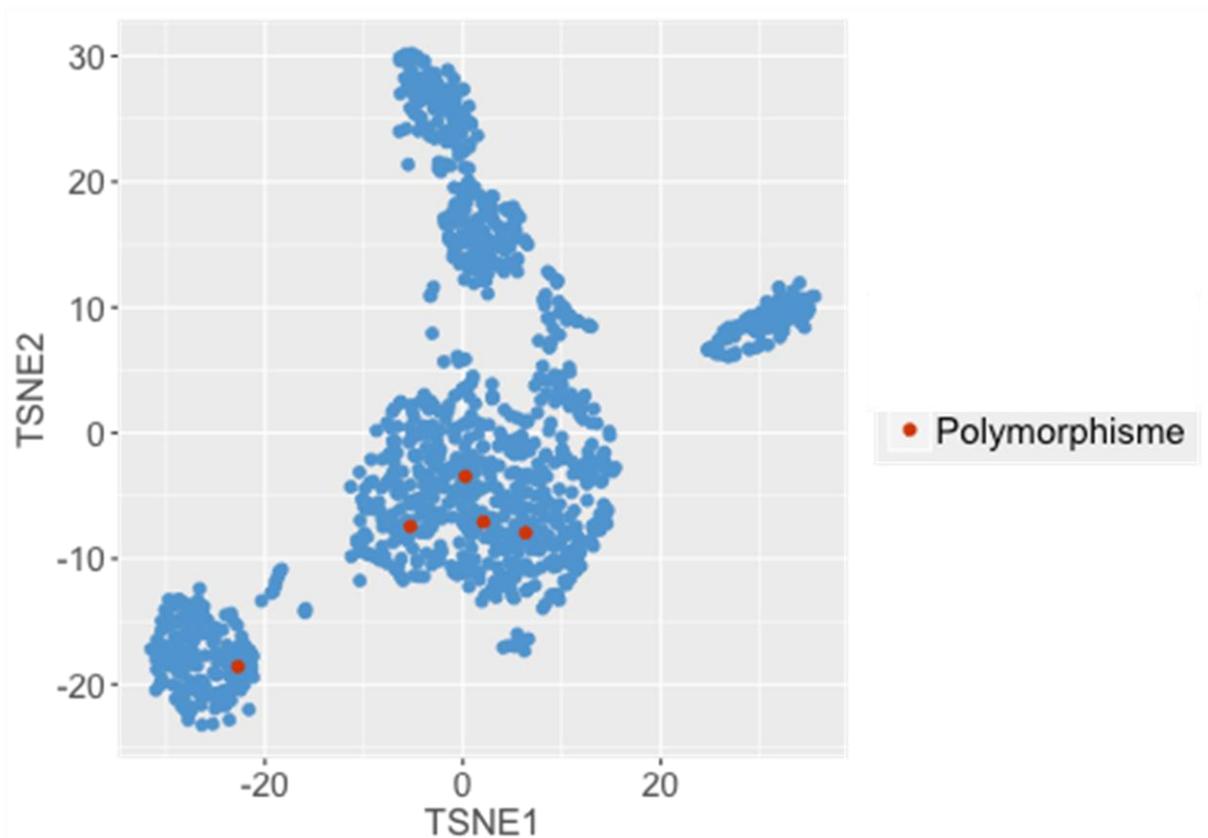


Figure 40 : Position des cellules mutées dans l'échantillon

Les points bleus représentent l'ensemble des cellules unique et les points rouges sont les cellules ayant le même polymorphisme sur 17 positions différentes. Ces résultats proviennent de données RNAseq sur cellule unique du mélanome.

On peut observer que ces cellules mutées ne se regroupent pas toutes ensemble dans un seul et même cluster et qu'une cellule se retrouve dans un autre cluster. La détection de 5 cellules partageant 17 mutations identiques suggère une certaine causalité.

Conclusion - Discussion

Plusieurs questions biologiques peuvent se résumer par des questions de lignages cellulaires. Etablir un lignage précis peut permettre de rendre compte de propriétés aussi importantes que la structure, la dynamique ou le développement d'une cellule particulière, ou éclairer son futur destin face au vieillissement, à la progression d'une maladie ou en réponse à un traitement. L'évolution d'une tumeur ou de métastases, la maturation du système immunitaire et la dynamique des cellules souches sont autant d'illustrations de domaines biologiques différents dans lesquels une connaissance approfondie des arbres de lignages cellulaires peut permettre d'obtenir des résultats essentiels. Pour y parvenir, la détection de néo-mutations somatiques attachées spécifiquement à certaines cellules normales au sein d'organismes multicellulaires a été envisagée par plusieurs auteurs (Behjati et al., 2014). Avec l'hypothèse que des mutations aient lieu lors de la réplication mitotique, le nombre de mutations s'accumulant au sein d'une cellule serait directement lié au nombre de divisions présentes, en lien avec des processus de dommages et/ou de réparation de l'ADN qui pourraient être différents en fonction des types cellulaires. De telles cellules porteraient donc des mutations distinctes. Cependant suivre de telles modifications nécessite des informations génomiques fiables pour chaque cellule, et les technologies nécessaires pour y parvenir ne sont pas encore nécessairement à maturité. Pour l'heure, les approches disponibles ne couvrent qu'incomplètement les génomes, et les erreurs systématiques liées aux technologies d'amplification et de séquençage peuvent fausser la reconstruction des lignages. Cette problématique a été abordée par l'équipe d'Ehud Shapiro, qui combine la technologie de la cellule unique identifiée grâce à un code-barre cellulaire et des PCR multiplex sur de multiples microsatellites afin de retracer le lignage cellulaire (Biezuner et al., 2016). Alors que Biezuner et ses collègues se sont intéressés aux séquences micro-satellites du génome nucléaire, l'objectif initial de mon travail a plutôt été d'essayer de suivre ce lignage à l'échelle de la cellule unique à partir des informations de séquences des mitochondries, bien plus abondantes. Plusieurs niveaux de variations peuvent ainsi y être étudiés, puisque chaque cellule comprend théoriquement de nombreuses copies d'un génome parfaitement identique, et qu'il est donc possible de détecter des variations simples de séquences nucléotidiques, affectant une position particulière sur le génome. De telles différences permettent de définir à partir de toutes les différences observées un haplogroupe, regroupant des individus possédant les mêmes mutations sur le génome mitochondrial permettant ainsi la caractérisation de

populations spécifiques. Mon travail a montré qu'il était possible de mettre en évidence un haplogroupe mitochondrial non seulement en analysant des données de RNAseq globales, mais aussi à partir de résultats de séquençage réalisés sur cellule unique préalablement individualisées utilisant la technologie Chromium. J'ai toutefois observé que la détermination de l'haplogroupe était très dépendante de la couverture du séquençage des génomes mitochondriaux. Ces premiers résultats suggèrent qu'il est donc indispensable d'effectuer d'abord une recherche systématique l'haplotypage sur chaque cellule de chaque prélèvement afin de définir des critères de qualité de nos expériences et de ne sélectionner pour les analyses statistiques ultérieures que les seules cellules ayant une couverture suffisante du transcriptome mitochondrial (figure 29). Ces résultats sont encore préliminaires et il est indispensable dans nos prochaines expériences sur cellules uniques de vérifier que dans le cas de cellules présentant un haplogroupe différent de celui de toutes les cellules de l'échantillon que cette différence n'est pas le résultat artefactuel résultant d'une trop faible couverture de séquençage ou d'une contamination par les différents intermédiaires qui manipulent les échantillons.. Il nous est toutefois impossible actuellement de mettre en évidence ce type de contamination dans un échantillon, si l'haplogroupe du manipulateur est le même que l'haplogroupe de référence international H2a2a défini par l'Université de Cambridge, ou s'il est identique à celui du patient auquel appartient l'échantillon. En effet, la présence de contamination ne pouvant jamais être exclue dans ce type d'expériences, l'identification de ces problèmes potentiels nous permettra lors de chacune des étapes du protocole, d'anticiper d'éventuelles contaminations et de définir ainsi un cahier des charges pour les utilisateurs de ces protocoles.

Les travaux réalisés dans l'équipe de Janes ont permis de caractériser des cellules progénitrices multipotentes de l'appareil respiratoire supérieur en suivant la descendance d'une cellule grâce aux accumulations de mutations (Teixeira et al., 2013). Cependant mes analyses (échantillons RNAseq) ne m'ont pas permis de détecter de différence entre le nez et les bronches chez les patients du projet GENEPRO. J'ai alors réalisé le même type d'analyse en comparant 2 tissus ayant une origine embryonnaire plus éloignées, à savoir des prélèvements nasaux et sanguins, en partant du principe qu'ils pouvaient potentiellement avoir accumulé davantage de variants au cours de leur différenciation, acquérant ainsi une signature génétique qui leur soit propre. Malgré la différence d'origine embryonnaire de ces 2 tissus, aucune différence nette n'a pu être observée. Toutefois de manière intéressante, j'ai pu

identifier un variant en position 1338 de l'ADN mitochondrial présent dans les cellules sanguines, mais pas dans le tissu nasal. Dans ce variant, la Guanine (G) en position 1338 est substitué par une Adénine (A). Cependant ce variant ne représente qu'un faible pourcentage (3%) des séquences analysées. Il est donc désormais important de savoir si ce pourcentage est un paramètre assez robuste pour différencier significativement les deux tissus. Si c'est le cas, cette variation pourrait être liée à une population de cellules spécifiques qu'une mesure globale sur un grand nombre de cellules ne permettrait pas d'évaluer correctement. Une expérience sur cellule unique issue d'un brassage nasal du même patient a ensuite été réalisée. J'ai pu m'apercevoir de la présence dans cet échantillon de cellules sanguines comme l'atteste la présence de marqueurs tels que CD4, CD8, CCL5 (i.e. cellules immunitaires). J'ai donc souhaité exploiter cet échantillon mixte pour vérifier ici la présence de la substitution G>A que l'on observe dans les tissus analysé indépendamment. Cependant la trop faible couverture de séquençage dans cette expérience ne m'a pas permis de répondre à cette question. A ce jour l'unique expérience que nous avons réalisée en comparant le sang et le nez ne nous permet donc pas d'affirmer qu'il existe réellement un variant lignée dépendant. Il serait donc intéressant de recruter d'autres donneurs sains pour lesquelles il serait possible d'obtenir concomitamment un prélèvement sanguin et un brossage nasal. Il est peu probable que nous retrouvions ce variant 1338 chez d'autres mais il est possible que nous identifions d'autres types de variants potentiellement associés à la lignée lymphocytaire. Concernant le patient PB auquel nous avons associé le polymorphisme 1338, pour palier la trop faible couverture des cellules CCL5+, on pourrait concevoir une expérience sur le Chromium en utilisant un prélèvement sanguin.

C'est finalement les cellules dissociées à partir d'un mélanome oculaire qui présentent le plus clairement une hétéroplasmie dans certaines cellules (figure 40). En groupant les cellules ayant un profil d'expression homogène, j'ai pu analyser certains groupes de cellules. Les premières analyses suggèrent quelques variations entre cellules et une analyse plus systématique doit maintenant être entreprise. Pour approfondir ces analyses, les données issues des cellules de biopsies de cornets nasaux et de mélanomes oculaires individualisées sur Chromium seront plus exploitées de manière plus approfondies.

Perspectives

L'hypothèse selon laquelle nous pourrions suivre une signature cellulaire à partir de variations mitochondriales n'a pas pu être complètement validée dans l'analyse initiale qui a porté essentiellement sur un seul patient, même si à ce stade une hétéroplasmie semble bien pouvoir être détectée entre les deux échantillons de sang et de nez. J'ai ainsi détecté un écart de 3% pour un variant à la position 1338 dans l'échantillon de sang qui se trouve être absent dans l'échantillon nasal. Le problème actuellement posé reste à savoir si une telle différence peut être considérée comme suffisamment significative.

Un problème majeur qui reste posé dans ces conditions est de pouvoir quantifier les plus exactement possible les niveaux relatifs entre les différentes séquences, ce que ne permet pas nécessairement le protocole de séquençage que j'ai utilisé. Il serait sans doute intéressant pour ce faire d'utiliser une technologie en cours de développement au sein de mon laboratoire, qui permet de corriger les biais d'amplification lors des PCR. Le principe général consiste à introduire de courtes amorces nucléotidiques aléatoires lors de la construction des banques d'ADN complémentaire. Chacune de ces séquences, appelées UMI (Unique Molecular Identifiers), est associée à une et une seule molécule d'ADN complémentaire. Il est alors possible de ne décompter qu'une seule fois une molécule amplifiée en multiples copies lors de la PCR. Ces séquences UMIs joueront un rôle central pour améliorer la fiabilité des résultats. Des travaux seront donc poursuivis, dans le cadre des recherches du laboratoire afin de mettre en place un « pipeline » d'analyses de données sur cellule unique qui soit parfaitement fiable. Il pourrait contenir certaines propositions résultant de mon travail, comme l'utilisation de l'haplogroupe cellulaire pour tracer d'éventuelles contaminations.

A plus long terme, l'analyse des nombreuses mitochondries présentes au sein d'une seule et même cellule permettra certainement d'effectuer des analyses plus fines. En effet, des variations entre différentes copies d'ADN mitochondrial présentes au sein même d'une seule cellule pourraient être accessibles à travers des techniques de séquençage en « long-read ». Ces techniques combinées avec des UMIs permettraient d'identifier une et une seule molécule d'ADN mitochondrial et de distinguer de la sorte toutes les différentes copies présentes au sein d'une seule cellule. Nous pourrions alors confirmer les niveaux d'hétéroplasmie au sein d'une même cellule.

Bibliographies

- Ayala, F. J. (1995). The Myth of Eve : Molecular Biology and human Origins. *Science*, 270, 1930–1936.
- Bals, R., Weiner, D. J., & Wilson, J. M. (1999). The innate immune system in cystic fibrosis lung disease. *Journal of Clinical Investigation*, 103(3), 303–7. <https://doi.org/10.1172/JCI6277>
- Behjati, S., Huch, M., van Boxtel, R., Karthaus, W., Wedge, D. C., Tamuri, A. U., ... Stratton, M. R. (2014). Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature*, 513(7518), 422–425. <https://doi.org/10.1038/nature13448>
- Biezuner, T., Spiro, A., Raz, O., Amir, S., Milo, L., Adar, R., ... Shapiro, E. (2016). A generic, cost-effective, and scalable cell lineage analysis platform. *Genome Research*, 26, 1588–1599. <https://doi.org/10.1101/gr.202903.115>
- Blanquart, C., Giuliani, I., Jeulint, C., Guennou, C., & Marano, F. (1995). in Vitro Exposure of Rabbit Tracheal to SO₂: Effects on Morphology and Ciliary Beating. *Toxicology In Vitro*, 9(2), 123–132.
- Boers, J. E., Ambergen, A. W., & Thunnissen, F. B. J. M. (1998). Number and Proliferation of Basal and Parabasal Cells in Normal Human Airway Epithelium. *American Journal of Respiratory and Critical Care Medicine*, 157(6), 2000–2006. <https://doi.org/10.1164/ajrccm.157.6.9707011>
- Breeze, R. G., & Wheeldon, E. B. (1977). The Cells of the Pulmonary Airways. *American Review of Respiratory Disease*, 116(4), 705–777. <https://doi.org/10.1164/arrd.1977.116.4.705>
- Chami, M., Oulè, B., Szabadkai, G., Tacine, R., Rizzuto, R., & Paterlini-Bréchet, P. (2008). Role of SERCA1 Truncated Isoform in the Proapoptotic Calcium Transfer from ER to Mitochondria during ER Stress. *Molecular Cell*, 32, 641–651. <https://doi.org/10.1016/j.molcel.2008.11.014>
- Chaussonot, A., & Paquis-Flucklinger, V. (2014). An overview of neurological and neuromuscular signs in mitochondrial diseases. *Revue Neurologique*, 170(5), 323–338. <https://doi.org/10.1016/j.neurol.2014.03.007>
- Chen, Y.-S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A. S., Wallace , D. C., & Wallace, D. C. (1995). Analysis of mtDNA Variation in African Populations Reveals the Most Ancient of All Human Continent-Specific Haplogroups. *Am. J. Hum. Genet*, 57, 133–149.
- Clarke, J., Wu, H.-C., Jayasinghe, L., Patel, A., Reid, S., & Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, 4, 265–270. <https://doi.org/10.1038/NNANO.2009.12>

- Evans, M. J., Cox, R. A., Shami, S. G., Wilson, B., & Plopper, C. G. (1989). The Role of Basal Cells in Attachment of Columnar Cells to the Basal Lamina of the Trachea. *American Journal of Respiratory Cell and Molecular Biology*, *1*(6), 463–470. <https://doi.org/10.1165/ajrcmb/1.6.463>
- Friedman, J. R., & Nunnari, J. (2014). Mitochondrial form and function. *Nature*, *505*(7483), 335–343. <https://doi.org/10.1007/978-1-62703-673-3>
- Gandhi, V. V., & Samuels, D. C. (2011). A Review Comparing Deoxyribonucleoside Triphosphate (dNTP) Concentrations in The Mitochondrial and Cytoplasmic Compartments of Normal and Transformed Cells. *Nucleosides, Nucleotides and Nucleic Acids*, *30*, 317–339. <https://doi.org/10.1080/15257770.2011.586955>
- Gawad, C., Koh, W., & Quake, S. R. (2016). Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*, *17*(3), 175–188. <https://doi.org/10.1038/nrg.2015.16>
- Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Genetics*, *77*(11), 6715–6719.
- Goiran, T. (2016). *Etude de la protéine PINK1 dans la maladie d'Alzheimer et le Cancer Cérébral*. Université Cote d'Azur.
- Goodwin, S., McPherson, J. D., & Richard McCombie, W. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*, 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Hansell, M. M., & Moretti, R. L. (1969). Ultrastructure of the mouse tracheal epithelium. *Journal of Morphology*, *128*(2), 159–169. <https://doi.org/10.1002/jmor.1051280203>
- Hong, K. U., Reynolds, S. D., Watkins, S., Fuchs, E., & Stripp, B. R. (2004). Basal Cells Are a Multipotent Progenitor Capable of Renewing the Bronchial Epithelium. *American Journal of Pathology*, *164*(2), 577–588.
- Ju, J., Kim, D. H., Bi, L., Meng, Q., Bai, X., Li, Z., ... Turro, N. J. (2006). Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(52), 19635–40. <https://doi.org/10.1073/pnas.0609513103>
- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C., & Teichmann, S. A. (2015). The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell*, *58*, 610–620. <https://doi.org/10.1016/j.molcel.2015.04.005>
- Lazcano, A., & Peretó, J. (2017). On the origin of mitosing cells: A historical appraisal of Lynn Margulis endosymbiotic theory. *Journal of Theoretical Biology*, *434*, 80–87. <https://doi.org/10.1016/j.jtbi.2017.06.036>
- Logan, D. C. (2006). The mitochondrial compartment. *Journal of Experimental Botany*, *57*(6), 12225–1243. <https://doi.org/10.1093/jxb/erj151>

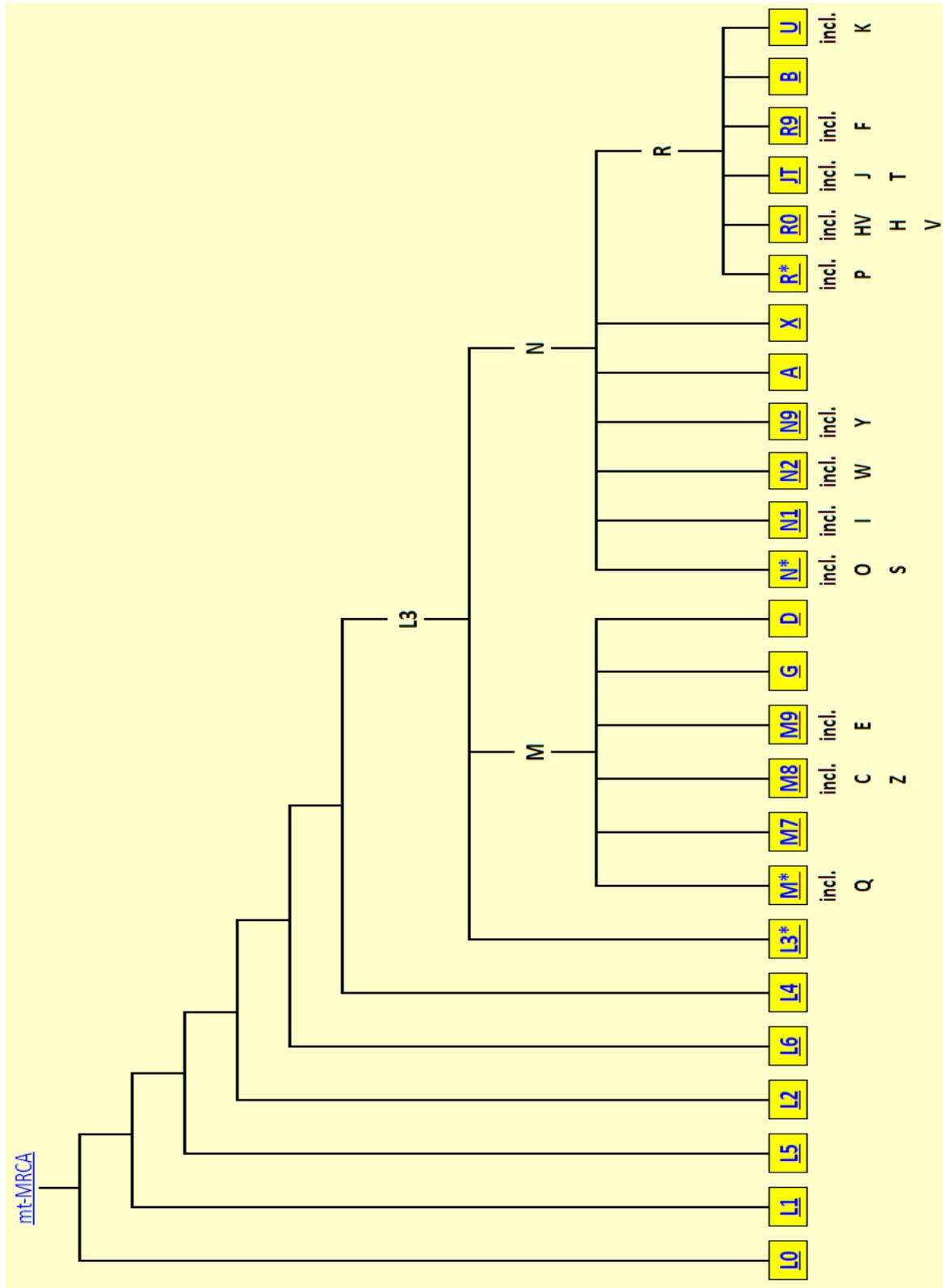
- Martin, S. A. (2011). Mitochondrial DNA repair. In F. Storici (Ed.), *DNA repair - On pathway to fixing DNA damage and errors* (Vol. 3, pp. 313–338). InTech.
- Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Res.*, *15*(12), 1767–76. <https://doi.org/10.1101/gr.3770505.with>
- Metzker, M. L. (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics*, *11*, 31–46. <https://doi.org/10.1038/nrg2626>
- Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*, *12*(2), 87–98. <https://doi.org/10.1038/nrg2934.RNA>
- Park, K.-S., Wells, J. M., Zorn, A. M., Wert, S. E., Laubach, V. E., Fernandez, L. G., & Whitsett, J. A. (2006). Transdifferentiation of Ciliated Cells during Repair of the Respiratory Epithelium. *American Journal of Respiratory Cell and Molecular Biology*, *34*, 151–157. <https://doi.org/10.1165/rcmb.2005-0332OC>
- Patten, D. A., Wong, J., Khacho, M., Soubannier, V., Mailloux, R. J., Pilon-Larose, K., ... Slack, R. S. (2014). OPA1-dependent cristae modulation is essential for cellular adaptation to metabolic demand. *The EMBO Journal*, *33*, 2676–2691. <https://doi.org/10.15252/emj>
- Perales-Clemente, E., Cook, A. N., Evans, J. M., Roellinger, S., Secreto, F., Emmanuele, V., ... Nelson, T. J. (2016). Natural underlying mtDNA heteroplasmy as a potential source of intra-person hiPSC variability. *The EMBO Journal*, *35*, 1979–1990. <https://doi.org/10.15252/emj>
- Piquereau, J., Caffin, F., Novotova, M., Lemaire, C., Veksler, V., Garnier, A., ... Joubert, F. (2013). Mitochondrial dynamics in the adult cardiomyocytes: which roles for a highly specialized cell? *Frontiers in Physiology*, *4*, 102. <https://doi.org/10.3389/fphys.2013.00102>
- Proserpio, V., & Lönnberg, T. (2016). Single-cell technologies are revolutionizing the approach to rare cells. *Immunology and Cell Biology*, *94*(3), 225–229. <https://doi.org/10.1038/icb.2015.106>
- Rhodin, J. A. (1966). The ciliated cell. Ultrastructure and function of the human tracheal mucosa. *The American Review of Respiratory Disease*, *93*(3), Suppl:1-15. <https://doi.org/10.1164/arrd.1966.93.3P2.1>
- Rogers, D. F. (1994). Airway goblet cells: responsive and adaptable front-line defenders. *Eur Respir J*, *7*(7), 1690–1706. <https://doi.org/10.1183/09031936.94.07091678>
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., ... Bustillo, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, *475*, 348–352. <https://doi.org/10.1038/nature10242>

- Serafini, S. M., & Michaelson, E. D. (1977). Length and distribution of cilia in human and canine airways. *Bulletin Europeen de Physiopathologie Respiratoire*, 13(4), 551–559.
- Sévin, C., & Girard, J. (1998). La chaîne respiratoire et sa génétique. In *Médecine thérapeutique / Pédiatrie* (Vol. 2, pp. 227–34). John Libbey Eurotext.
- Shapiro, E., Biezuner, T., & Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics*, 14, 618–630. <https://doi.org/10.1038/nrg3542>
- Stewart, J. B., & Chinnery, P. F. (2015). The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nature Reviews Genetics*, 16, 530–542. <https://doi.org/10.1038/nrg3966>
- Taanman, J.-W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta*, 1410(2), 103–123.
- Taylor, R. W., Barron, M. J., Borthwick, G. M., Gospel, A., Chinnery, P. F., Samuels, D. C., ... Turnbull, D. M. (2003). Mitochondrial DNA mutations in human colonic crypt stem cells. *The Journal of Clinical Investigation*, 112(9), 1351–60. <https://doi.org/10.1172/JCI19435>
- Taylor, R. W., & Turnbull, D. M. (2005). Mitochondrial DNA Mutations in Human Disease. *Nature Reviews Genetics*, 6(5), 389–402. <https://doi.org/10.1038/nrg1606>
- Teixeira, V. H., Nadarajan, P., Graham, T. A., Pipinikas, C. P., Brown, J. M., Falzon, M., ... Janes, S. M. (2013). Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. *eLife*, 2. <https://doi.org/10.7554/eLife.00966.001>
- Torrioni, A., Lott, M. T., Cabell, M. F., Chen, Y.-S., Lavergne, L., & Wallace, D. C. (1994). mtDNA and the Origin of Caucasians: Identification of Ancient Caucasian-specific Haplogroups, One of Which Is Prone to a Recurrent Somatic Duplication in the D-Loop Region. *Am. J. Hum. Genet*, 55, 760–776.
- Torrioni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V, Larsen, M., ... Wallace, D. C. (1993). Asian Affinities and Continental Radiation of the Four Founding Native American mtDNAs. *Am. J. Hum. Genet*, 53, 563–590.
- van Oven, M., & Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, 30(2). <https://doi.org/10.1002/humu.20921>
- Wallace, D. C., & Chalkia, D. (2013). Mitochondrial DNA Genetics and the Heteroplasmy Conundrum in Evolution and Disease. *Cold Spring Harbor Perspectives in Biology*, 5(11), a021220. <https://doi.org/10.1101/cshperspect.a021220>

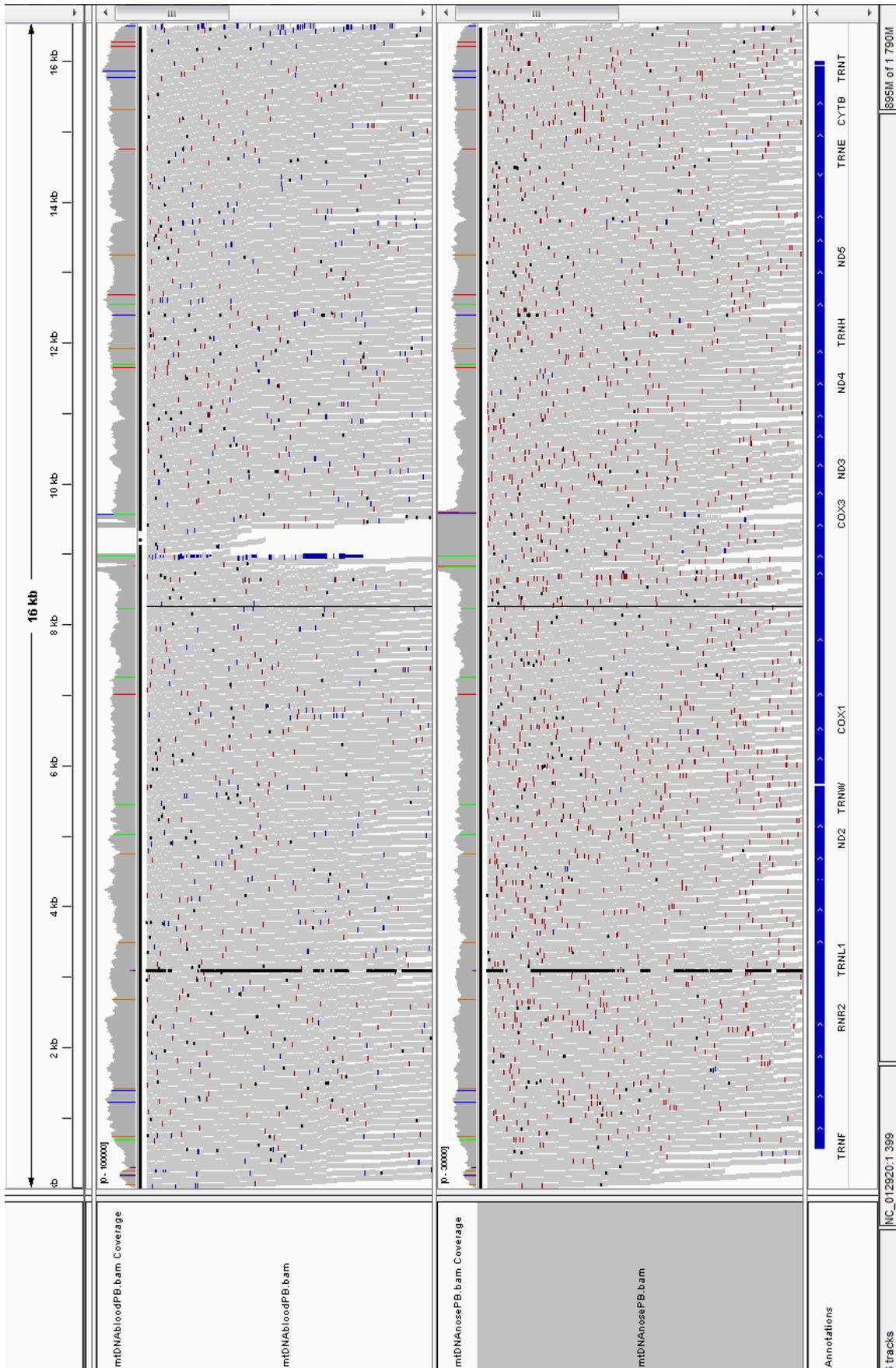
- Wang, Y., & Navin, N. E. (2015). Advances and Applications of Single-Cell Sequencing Technologies. *Molecular Cell*, 58(4), 598–609.
<https://doi.org/10.1016/j.molcel.2015.05.005>
- Zhang, J., Chiodini, R., Badr, A., & Zhang, G. (2011). The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38(3), 95–109.
<https://doi.org/10.1016/j.jgg.2011.02.003>.The
- Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., ... Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, 8, 14049. <https://doi.org/10.1038/ncomms14049>
- Zhou, Q., Li, H., Li, H., Nakagawa, A., Lin, J. L. J., Lee, S., ... Xue, D. (2016). Mitochondrial endonuclease G mediates breakdown of paternal mitochondria upon fertilization. *Science*, 353(6297), 394–399.
<https://doi.org/10.1126/science.aaf4777>.Mitochondrial

Annexes

Annexe 1 : Arbre phylogénétique de l'ADNmt.



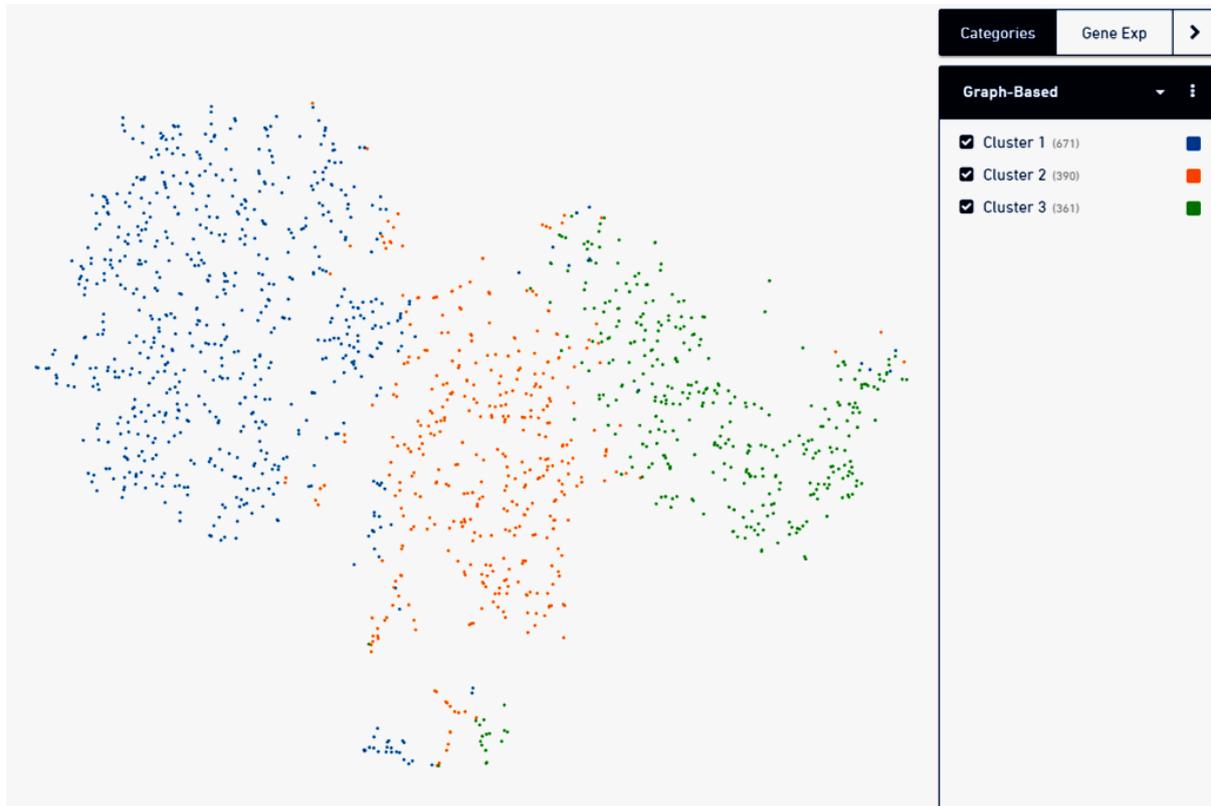
Annexe 2 : Alignement des séquences des bibliothèques Nez et Sang sur l'ADNmt.



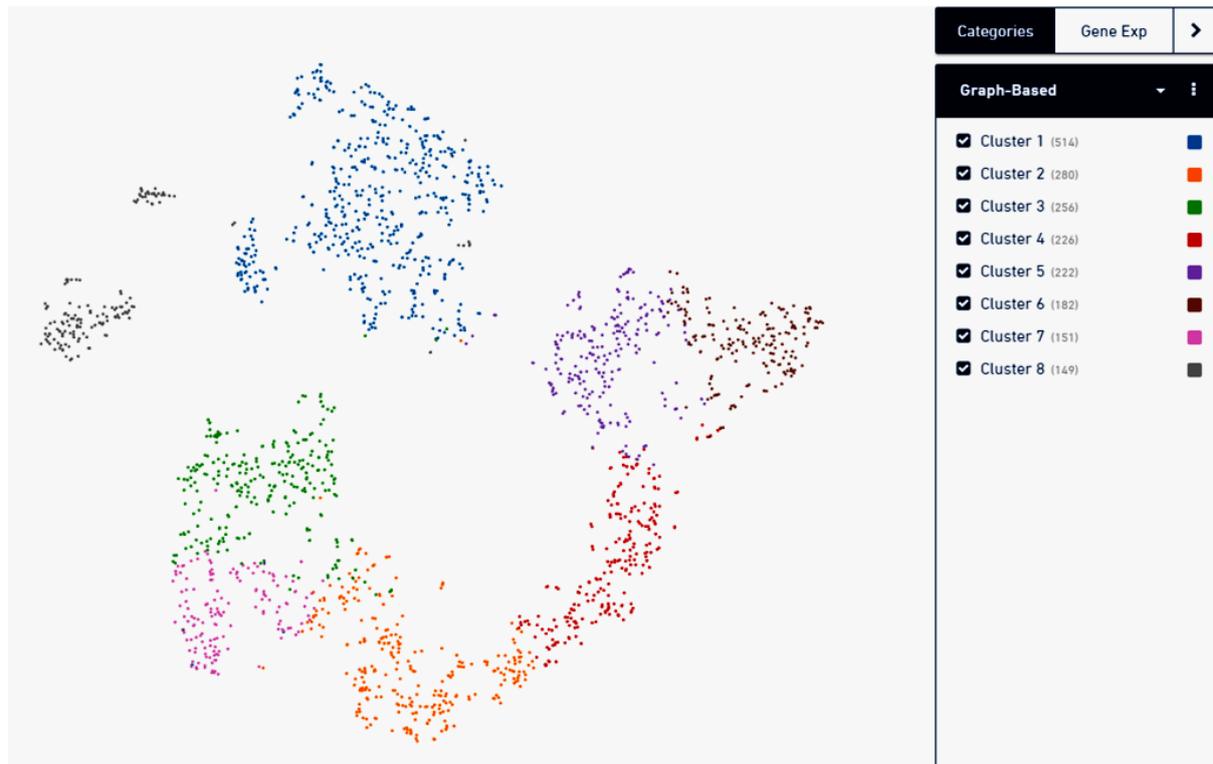
Annexe 3 : Répartition cellulaire des cellules exprimant MUC5AC du patient PB



Annexe 4 : Répartition cellulaire du patient D246 Ali2



Annexe 5 : Répartition cellulaire du patient D246 Ali22



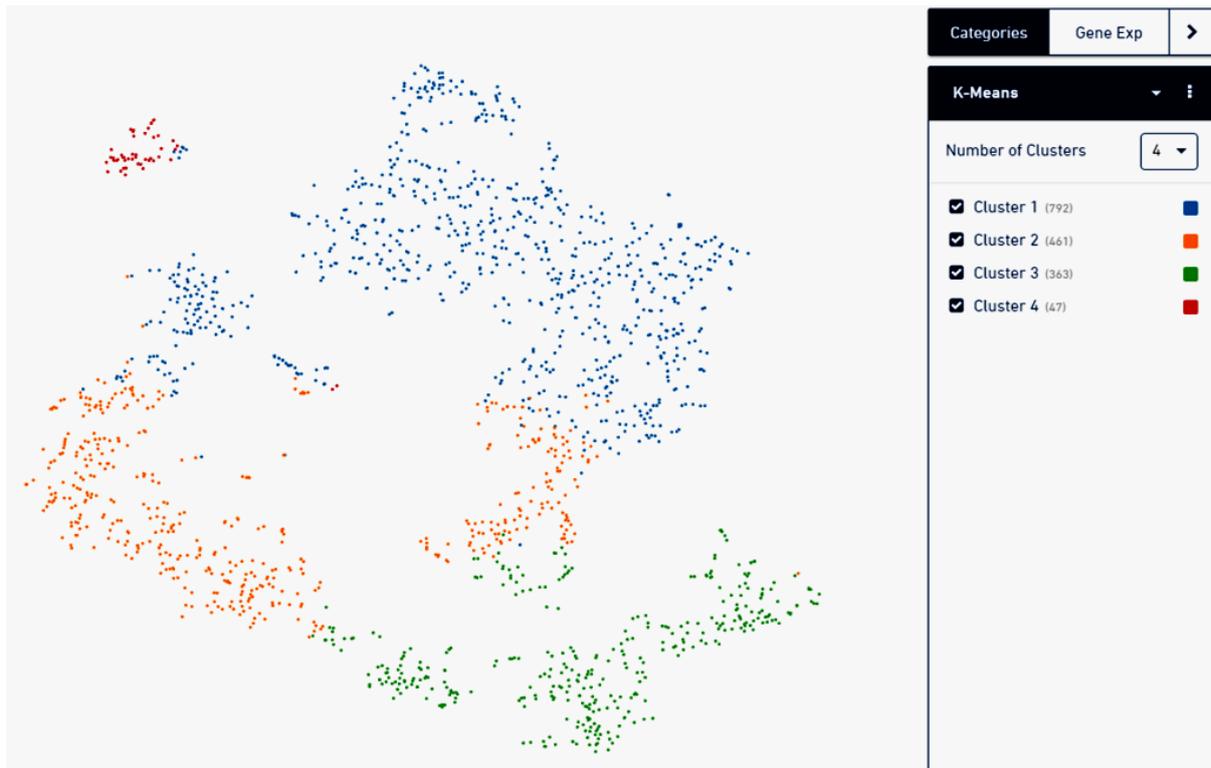
Annexe 6 : Répartition cellulaire des cellules exprimant KRT5 du patient D258.



Annexe 7 : Position possédant au moins une cellule mutée

Pos	Ref	Nb ref	Nb n	Nb W	Nb R	Nb total
295	C	1633	27	0	0	1660
183	A	1309	352	0	0	1661
214	A	1403	258	0	0	1661
989	T	1562	99	0	0	1661
3412	G	1019	642	0	0	1661
4769	A	309	1352	0	0	1661
14766	C	124	1537	0	0	1661
42	T	1300	362	0	0	1662
178	A	1307	355	0	0	1662
747	A	1480	182	0	0	1662
859	T	1496	166	0	0	1662
1273	G	1634	28	0	0	1662
1301	G	1645	17	0	0	1662
1371	T	1658	4	0	0	1662
1413	T	1633	29	0	0	1662
2737	T	1645	17	0	0	1662
3465	A	1171	491	0	0	1662
3467	A	1171	491	0	0	1662
3508	A	992	670	0	0	1662
3971	T	1659	3	0	0	1662
4064	A	1661	1	0	0	1662
4309	G	292	1370	0	0	1662
4412	G	401	1261	0	0	1662
4446	G	528	1134	0	0	1662
4450	G	527	1135	0	0	1662
5133	A	1459	203	0	0	1662
7203	T	1661	1	0	0	1662
7277	A	1662	0	0	0	1662
7298	A	1662	0	0	0	1662
9418	A	1435	227	0	0	1662
9723	T	1662	0	0	0	1662
10197	G	1659	2	0	1	1662

Annexe 8 : Répartition cellulaire du patient D267



ECOLE PRATIQUE DES HAUTES ETUDES

SCIENCE DE LA VIE ET DE LA TERRE

Evaluation de l'hétérogénéité mitochondriale par séquençage de l'ADN

Pons Nicolas

Soutenu le 11 Janvier 2018

Résumé

Les mitochondries sont présentes en grand nombre dans la plupart des cellules de l'organisme. D'origine exclusivement maternelle, les mitochondries présentent un taux de mutation 10 fois plus élevé que l'ADN nucléaire. Les différences présentées par le génome mitochondrial entre individus et la présence de nombreuses copies de l'ADNmt par mitochondrie et par cellule, peut offrir un moyen d'étudier certaines propriétés tissulaires et/ou cellulaire au sein même d'un individu. Pour répondre à cette question, j'ai mis en place au sein du laboratoire l'amplification de l'ADNmt et j'ai optimisé le protocole afin de l'adapter à un système microfluidique (C1, Fluidigm) afin de pouvoir effectuer le séquençage de l'ADNmt d'une et une seule cellule. Dans un second temps, j'ai étudié, à partir de données RNAseq bulk et RNAseq Chromium disponibles, les variations détectables au niveau des séquences mitochondriales. Grâce à ces analyses, il m'a été possible de définir directement l'haplogroupe mitochondrial de cellules uniques à partir de données de RNAseq. J'ai pu également identifier quelques variations intercellulaires au sein même d'échantillons de patients. Par contre, le nombre de cellules détectées dans l'analyse n'a pas été suffisant pour me permettre d'utiliser ces variants pour effectuer des expériences de lignage cellulaire. Ces travaux montrent qu'il est possible d'utiliser l'haplogroupe pour tracer d'éventuelles contaminations et comme contrôle qualité des expérimentations. Une étude plus approfondie sera nécessaire pour documenter plus exactement les niveaux d'hétéroplasmie dans les différents modèles biologiques étudiés au sein de mon équipe.

MOTS-CLES : Mitochondrie, haplogroupe, variant, hétéroplasmie, NGS, Single-cell.