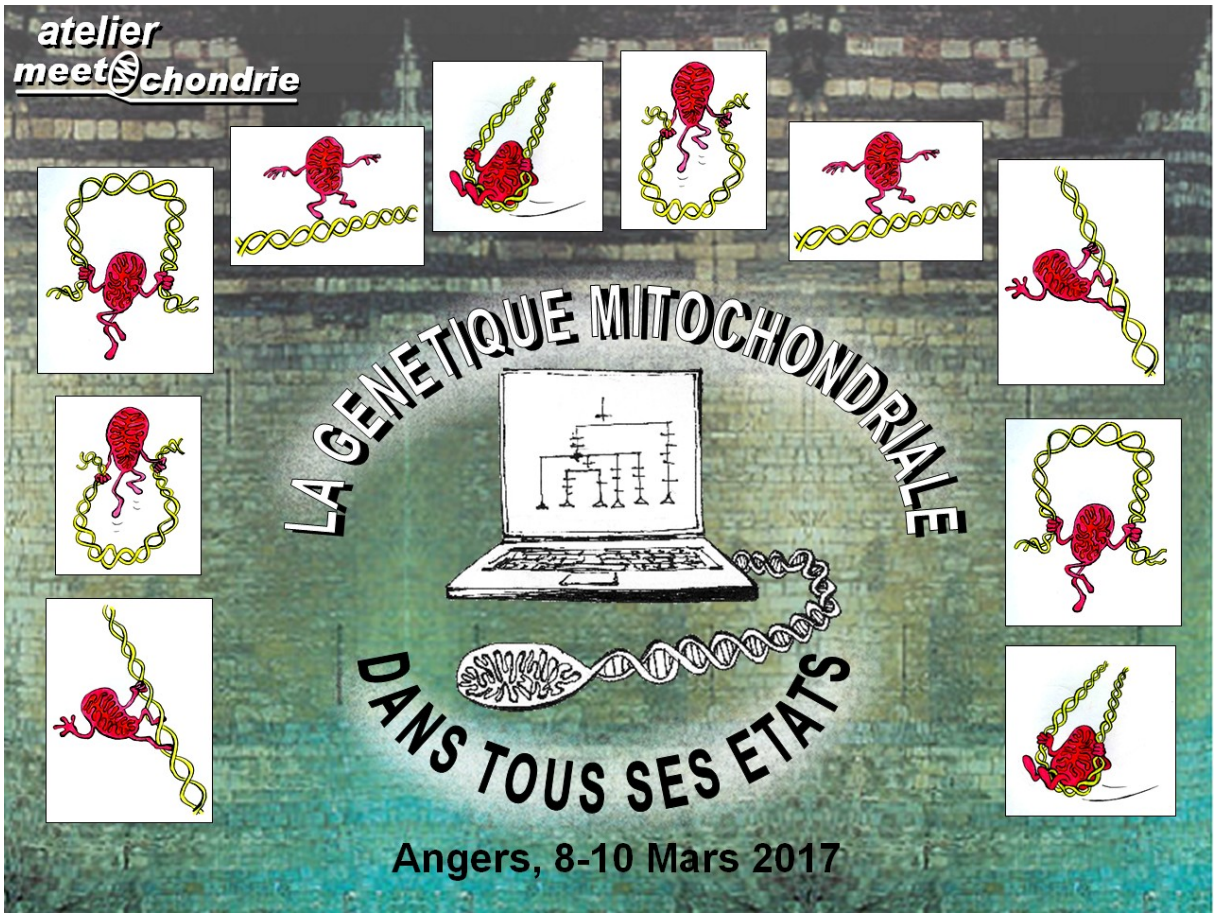
réseau meetochondrie

mitolab

ATELIER
8-10 mars 2017



Biologie moléculaire mitochondriale
Acquisition, Analyse de séquences d'**ADN mitochondrial**
Séquençage très haut débit.
Analyse bioinformatique
Exploration des variants pathogènes
Haplogroupes mitochondriaux

université angers

L'Institut de Biologie et Santé d'Angers et le laboratoire MitoVasc accueille l'atelier " la génétique mitochondriale dans tous ses états"!

L'atelier sera animé conjointement par des chercheurs de l'équipe MitoVasc (UMR 6015 U1083) d'Angers et de l'équipe Médecine Evolutive, laboratoire AMIS - UMR 5288 de Toulouse, avec la participation d'intervenants extérieurs.

Organisation : Arnaud Chevrollier, Vincent Procaccio
Contacts: Arnaud Chevrollier (MCF Biologie Moléculaire, Bioinformatique)
arnaud.chevrollier@univ-angers.fr
Vincent Procaccio, Pu-Ph Génétique ViProcaccio@chu-angers.fr

Encadrants :
Majida Charif (PhD, Post-Doc, Biologie Moléculaire, génétique, UMR6215, U1083, Mitolab) majida.charif@univ-angers.fr

Valérie Desquiret Dumas (PhD, Ingénieur CHU, Biologie Moléculaire, génétique, UMR6215, U1083, Mitolab) VaDesquiret@chu-angers.fr

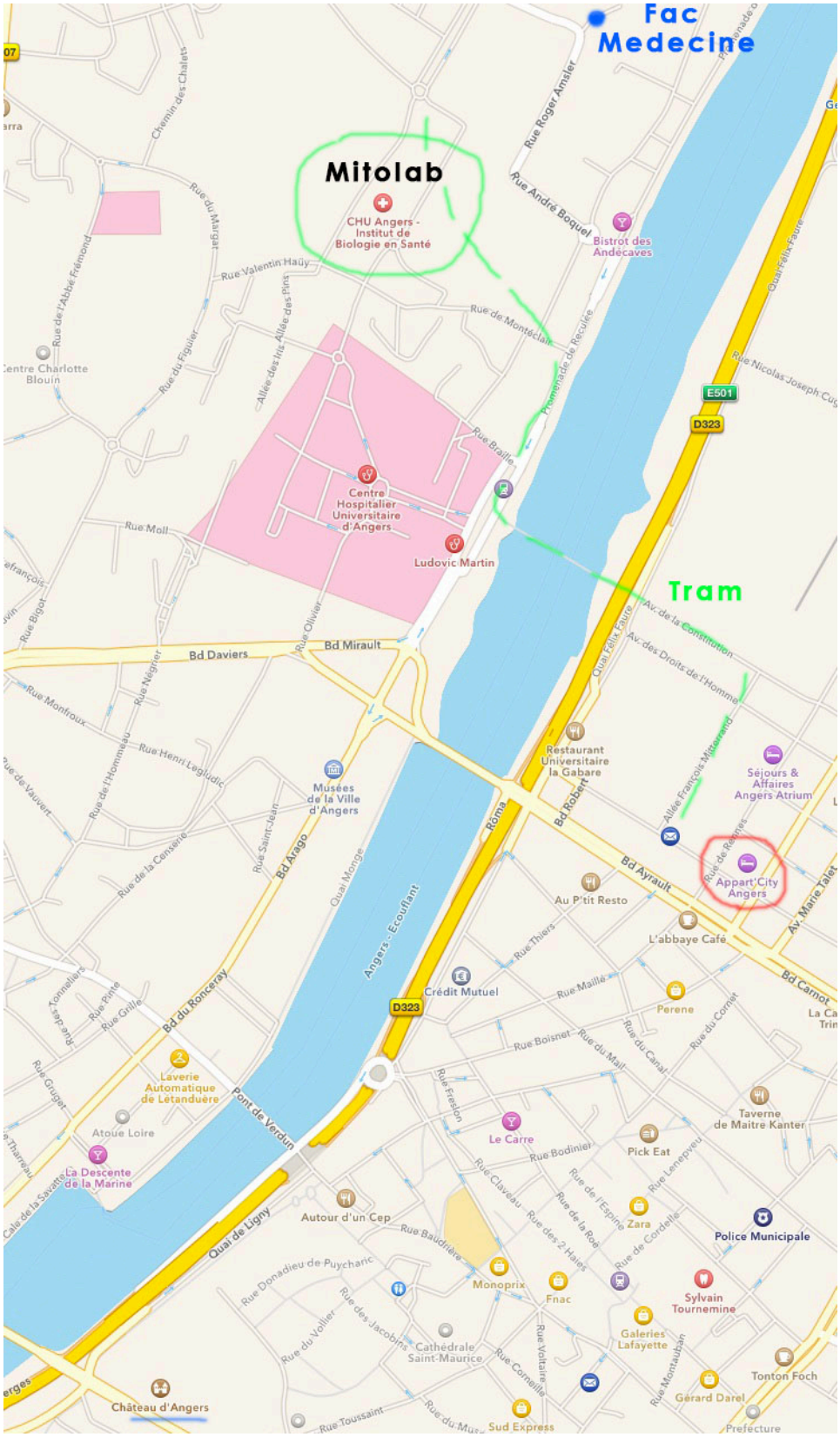
David Goudenège (PhD, Ingénieur CHU, Bioinformatique, UMR6215, U1083, Mitolab) David.Goudenege@chu-angers.fr
Pascal Reynier (Pu-Ph, Biochimie, CHU Angers)

Thierry Letellier (CR1 Inserm, UMR5288 Toulouse)

Les participants sont logés dans une résidence confortable en centre-ville, et deux soirées conviviales sont prévues.

réseau meetochondrie

mitolab



réseau meet chondrie

mitolab

Liste des participants

Amandine MORETTON	amandine.moretton@univ-bpclermont.fr	Laboratoire RGM, Université Blaise Pascal, Clermont-Ferrand
Aya, ABDUL-WAHED KAYALI	aya.kayali@inserm.fr	Inserm, Paris
Giulia BARCIA	giulia.barcia@aphp.fr	UF Génétique, Hôpital Necker Enfants Malades, Paris
Lisa BOUCRET	liboucret@chu-angers.fr	CHU ANGERS
Carine DELAUNA	carine.delaunay@univ-poitiers.fr	Université de Poitiers UMR7267 EES-EBI
Damien JEANDARD	damien.jeandard@etu.unistra.fr	UMR 7156 Strasbourg
Marine MALLETER	marine.malleter@univ-rouen.fr	Université de Rouen Normandie - laboratoire ABTE équipe Toxemac
Isabelle MARCADE	isabelle.marcade@univ-poitiers.fr	Université de Poitiers
Cécile PAGAN	cecile.pagan@chu-lyon.fr	Hospices Civils de Lyon
Cristina PANOZZO	cristina.panozzo@i2bc.paris-saclay.fr	I2BC Gif sur Yvette
Cécile SCHANEN	schanen-c@chu-caen.fr	CHRU CAEN
Jean Claude VIENNE	jcl.vienne@gmail.com	CHRU de Lille

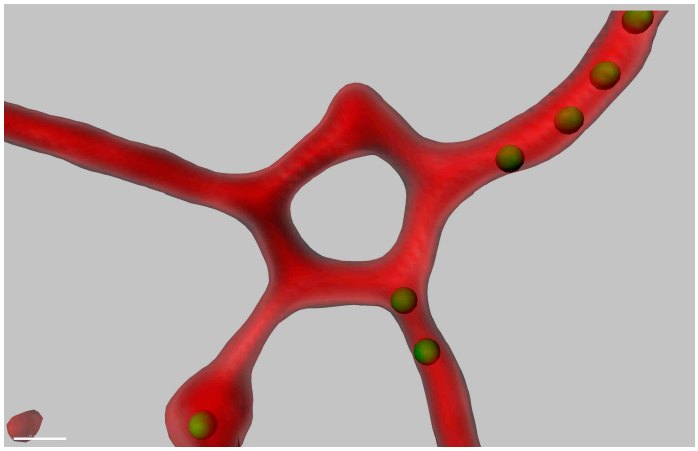
réseau

meetOchondrie

Chers participants, chers organisateurs,

Je suis très heureux de voir aboutir ce projet d'Atelier, qui est le premier que le **Réseau MeetOchondrie** organise depuis que nous avons pris la forme d'une Association.
Le Réseau organise des colloques et des journées thématiques qui sont des lieux d'échanges et d'interaction importants au sein de la communauté, mais les ateliers ont une importance pour l'Association. En effet, ce sont des évènements à la fois intenses et conviviaux ou le partage de l'expertise prend alors toute sa dimension, allant des aspects les plus théoriques aux plus concrets, et tant les participants que les encadrants en tirent grand bénéfice.
La génétique mitochondriale a une importance considérable pour le diagnostic, l'exploration et kle traitement de nombreuses pathologies liées à la mitochondrie, mais également pour la recherche fondamentale, notamment en évolution. L'atelier « ***La Génétique mitochondriale dans tous ses états*** » va ainsi permettre aux participants de découvrir et se familiariser avec les concepts et les outils les plus récents, en particulier avec les apports des technologies NGS et de la bioinformatique.
En remerciant chaleureusement les organisateurs et les encadrants pour leur dévouement, ainsi que les différents sponsors, je souhaite à tous un merveilleux atelier.

David Macherel
Président de l'Association Réseau MeetOchondrie



Préambule

Cet atelier est entièrement organisé sous l'égide de l'association Mitochondrie autour de la génétique mitochondriale qui depuis plusieurs années est devenue le centre d'intérêt de bon nombre de disciplines médicales et d'équipes de recherche. Plusieurs facettes de cette génétique si particulière seront abordées lors de cet atelier notamment par des conférences plénières. Influence sur évolution et migrations humaines ou utilisation des bases de données. Mais aussi sur les technologies récentes de prévention de ces maladies mitochondriales liées à des mutations de l'ADN mitochondrial avec les embryons à 3 parents qui posent un certain nombre de questions sur le plan technique mais aussi bioéthiques.

Différentes techniques récentes d'analyses à la fois qualitatives mais aussi quantitatives du génome mitochondrial seront développées lors ce de cet atelier. Avec la mise en place des nouvelles technologies de séquençage du génome, l'analyse du génome mitochondrial est en passe de devenir une analyse de routine. L'automatisation des techniques d'analyse du génome mitochondrial nous permet ainsi maintenant d'analyser de manière systématique cet ADN mitochondrial générant ainsi un accroissement considérable des données à analyser. De nouveaux outils d'analyse automatisée du génome mitochondrial ont été mis en place et des outils de prédiction de la pathogénicité des variants de l'ADN mitochondrial seront particulièrement développés au cours de plusieurs sessions bioinformatiques. Un important travail de formation et d'expertise à la génétique mitochondriale est ainsi nécessaire afin d'améliorer la compréhension des pathologies et la prise en charge de ces maladies particulièrement complexes et en pleine évolution.

Arnaud Chevrollier – Thierry Letellier – Vincent Procaccio

réseau meetochondrie

mitolab

Atelier : la génétique mitochondriale dans tous ses états

Programme

Mercredi 8 Mars

12h30-13h Accueil des participants, Café : IRIS Rue des capucins, CHU d'Angers

13h00-13h30 Présentation de l'atelier, David Macherel Association Meetochondrie

13h30-15h30

Atelier A Grp1

Quantité/qualité ADNmt /**PCR digitale,** Labo Mitolab Valérie Desquiret Dumas

Atelier B Grp2

Séquençage **NGS**, présentation technologique, visite plateau technique ; Préparation de la librairie. Labo Mitolab Majida Charif

15h30-17h30

Atelier B Grp1 Séquençage **NGS**, présentation technologique, visite plateau technique ; Préparation de la librairie. Labo Mitolab Majida Charif

Atelier A Grp2

Quantité/qualité ADNmt /**PCR digitale,** Labo Mitolab Valérie Desquiret Dumas

18h-19h **Conférence « Génétique mitochondriale, embryon à 3 parents, considérations médicales et bioéthiques** », Pr. **Julie Steffann**, Hopital Necker. Amphi 200, UFR Santé, Amphi 200

20h00 Diner en ville, Chez PontPont,

Jeudi 9 Mars

8h-12h00 introduction Haplogroupe Thierry

Atelier C Grp1 **Haplogroupage** par technique RFLP, polymorphismes de l'ADNmt. Labo Mitolab Majida Charif ; Valérie Desquiret Dumas

Atelier D Grp2

Analyse bioinformatique ADNmt total, salle info David Goudenege Med E201

12h30 Lunch CROUS

14h-18h00

Atelier C Grp2

Haplogroupage par technique RFLP, polymorphismes de l'ADNmt. Labo Mitolab Majida Charif ; Valérie Desquiret Dumas

Atelier D Grp1 **Analyse bioinformatique** ADNmt total, salle info David Goudenege Med E201

18h15-19h15 **Conférence « Haplogroupe, migration et pathologies »**
Dr. **Thierry Letellier**, Equipe de "Médecine Evolutive" Laboratoire AMIS - UMR 5288 Toulouse , Salle Conférence PBH

20h00 Diner en ville Chez Remi, 20h

Vendredi 10 Mars

8h-10h00

Atelier E Grp1 Analyse Bioinformatique NGS, « **mitome** » **salle info David Goudenege : Salle Med D201**

Atelier F Grp2 Analyse bioinformatique ADNmt **Haplogroupe** **salle Med Thierry Letellier Med G101**

10h00 -12h

Atelier F Grp1 Analyse bioinformatique ADNmt **Haplogroupe** **salle Med Thierry Letellier Med G101**

Atelier E Grp2 Analyse Bioinformatique NGS, « **mitome** » **salle Med D201 David**

12h-13h **Conférence « Evolution de la bioinformatique face au séquençage très haut débit, Big Databases » Pr. Christophe BEROUD,** UMR_S910, INSERM, GMGF, Marseille, France. Amphi ICO

13h15 Lunch IRIS

15h-16h00 **Table ronde : Validation fonctionnelle**
Pr. Pascal Reynier, Pr. Vincent Procaccio

16h00 Fin

université angers réseau meetochondrie

mitolab

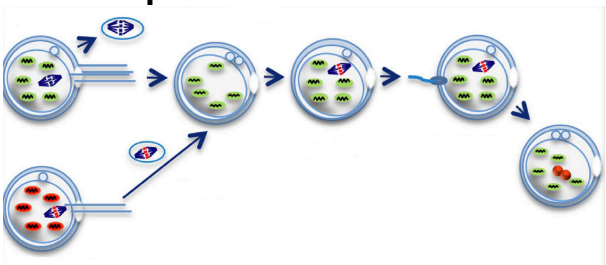
Mercredi 8 Mars à 18h, Amphi 200

UFR Santé, rue Haute de Reculée, Angers

Conférence « Génétique mitochondriale, embryon à 3 parents, considérations médicales et bioéthiques

Pr Julie Steffann

Université Paris-Descartes, Institut Imagine Unité UMR 1163
Hôpital Necker-Enfants Malades (AP-HP), Paris



La mitochondrie centrale énergétique de nos cellules a la particularité de posséder son propre génome. Les mutations de l'**ADN mitochondrial** (ADNmt) sont responsables de maladies sévères, dont le mode de transmission est exclusivement maternel. Elles ont la particularité d'être présentes le plus souvent à l'état hétéroplasmique (mélange de molécules d'**ADNmt mutées** et **sauvages**) définissant ainsi un taux de mutation ou hétéroplasmie, qui au-delà d'un certain seuil va entrainer l'apparition des symptômes.

Les femmes porteuses d'une mutation de l'ADNmt ont un risque élevé de transmettre une maladie grave à leur descendance, et les demandes de prise en charge en diagnostic prénatal (DPN) et/ou préimplantatoire (DPI) sont nombreuses. Des équipes ont récemment développé le don de cytoplasme, qui offre à ces femmes la possibilité de transmettre leurs caractères héréditaires portés par le génome nucléaire. Outre les problèmes éthiques avec la génération d'embryons à 3 parents, ces technologies posent un certain nombre de questions. Des interrogations subsistent quant à l'innocuité de cette approche du fait d'une possible perturbation du dialogue mito-nucléaire, de l'état hétéroplasmique induit par cette procédure, et de l'existence de possibles réversions génétiques.

Laboratoire MitoVasc
Arnaud Chevrollier – Vincent Procaccio

SFR 4208
Interactions cellulaires & Applications thérapeutiques

Séminaire SFR ICAT

Vendredi 10 Mars 2017

Horaire : 12H00

ICO - Site Paul Papin – Amphithéâtre

Pr Christophe Béroud

Université Aix-Marseille, INSERM, GMGF
APHM, Hôpital Timone Enfants,
Laboratoire de Génétique Moléculaire, Marseille

Evolution de la bioinformatique face au séquençage très haut débit, Big Databases

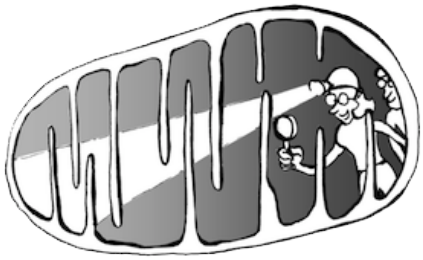
Les technologies de séquençage à haut débit sont maintenant fondamentales pour l'identification des mutations pathogènes dans les maladies génétiques humaines. Plus de 1000 gènes ont été identifiés entre 2010 et 2014 grâce au développement des technologies de séquençage complet. Cependant, malgré ce chiffre encourageant, le taux de détection des mutations pathogènes reste faible (entre 23% et 26%). Elle est due à plusieurs paramètres tels que les facteurs techniques, types de mutations, la suite d'outils bioinformatiques et les méthodes ou filtres utilisés.
Dans cette présentation, nous décrirons les étapes critiques des processus d'annotation et de filtration de variants pour identifier des mutations potentiellement pathogènes. Nous examinerons les éléments d'annotation clés et les systèmes conçus pour aider à recueillir ces informations critiques. Nous allons aussi décrire les options de filtration, leur efficacité et leurs limites, et fournir un flux de travail de filtration générique. Enfin, nous allons démontrer ce workflow en action et mettre en évidence les pièges potentiels à l'analyse. Nous examinerons également les applications du système UMD-Predictor, considéré comme l'outil de prédiction de pathogénicité le plus efficace pour les mutations faux-sens et les mutations synonymes (évaluation sur> 140 000 variations annotées), le plus rapide (3 à 20 fois plus rapide) et spécifique réduisant le nombre de mutations pathogènes candidates (25% à 50% des autres outils en moyenne) pour l'analyse de validation.

Laboratoire MitoVasc
Arnaud Chevrollier – Vincent Procaccio

réseau meetchondrie mitolab

Jeudi 9 Mars à 18h15, Salle Conférence IBS

CHU d'Angers



Conférence « La mitochondrie : une histoire dans notre Histoire »

Thierry Letellier
"Médecine Evolutive"
Laboratoire AMIS - UMR 5288
Toulouse

La mitochondrie occupe une place centrale dans notre métabolisme en produisant l'énergie indispensable à notre survie, l'ATP. Elle intervient aussi dans de nombreux processus cellulaires comme l'apoptose, la signalisation calcique ou la production de radicaux libres. Ainsi, la mitochondrie est impliquée dans de nombreuses pathologies (Cytopathies, Cancer, Alzheimer, …).

Notre histoire avec cet organite intracellulaire remonte au commencement de notre Histoire, il y a 1,5 milliard d'année, avec son implication dans la théorie endosymbiotique. L'ADN mitochondrial nous a permis de dater et de retracer les migrations de notre espèce dans sa conquête de nouveaux territoires.

Lors de cette conférence, je vous propose de remonter le passé en retraçant les principales étapes de la découverte des mitochondries et de ses fonctions mais surtout en montrant comment l'étude de l'ADN mitochondrial peut nous aider à (i) mieux comprendre nos origines et à (ii) considérer le fond génétique mitochondrial comme un facteur de risque ou un facteur protecteur dans certaines pathologies.

Laboratoire MitoVasc
Arnaud Chevrollier – Vincent Procaccio

université angers

réseau meetochondrie mitolab

Remerciements

Paul LAROCHE
Responsable technico-commercial - Service Commercial
+33 (0)6 49 99 27 11
paul.l@abioexpertise.com

ALLIANCE
BIO EXPERTISE
Z.A de Courbouton Le Tremplin - 35480 Guipry FRANCE
Tel : +33(0) 240 517 953
Fax : +33(0) 230 060 547
www.alliance-bio-expertise.com

Stéphane Colinet
Ingénieur Technico-Commercial Ouest
E-mail: scolinet@neb.com
Mobile: +33 (0)6 28 50 15 64
New England Biolabs France
5 Rue Henri Desbruères
Genopole Campus 1
Bâtiment 6
91030 EVRY CEDEX
N° vert Commandes: 0800 100 632 (orders.fr@neb.com)
Fax: 0800 100 610
Technique: 0800 100 633 (techsupport.fr@neb.com)

Atelier A

Quantification de l'ADN mitochondrial par PCR quantitative en temps réel

Contexte scientifique :

Le nombre de copies d'ADN mitochondrial est variable non seulement en fonction des tissus mais également en fonctions des besoins énergétiques et de diverses situations pathologiques. La maintenance de l'ADN mitochondrial est un domaine très étudié ces dernières années et joue un rôle prépondérant dans la physiologie et l'homéostasie cellulaire, des dysfonctionnement de ce système pouvant être impliquées dans le vieillissement ou le développement de pathologies neurodégénératives telles que la maladie de Parkinson dans laquelle une diminution du nombre de copies d'ADN mitochondrial a été observé (Pyle et al., 2016) mais également dans le développement tumoral.

D'un point de vue diagnostic, à l'heure actuelle des mutations dans 9 gènes de réplication de l'ADN mitochondrial et de formation du nucléoïde ont été identifiées et sont impliqués dans le syndrome de déplétion (MDS, Mitochondrial DNA Depletion Syndrome, Rooney et al., 2015). Dans ce cadre, une quantification précise du nombre de copies d'ADN mitochondrial dans les échantillons de patients se révèle être une aide précieuse pour le diagnostic. De même, l'utilisation croissante du séquençage à haut débit permet la découverte de nouveaux variants dans les gènes de maintenance et la quantification de l'ADN mitochondrial est là encore un élément clé permettant de confirmer la pathogénicité de ces variants.

De même, d'un point de vue recherche, de nombreuses conditions métaboliques et environnementales peuvent faire varier la balance entre la biogenèse mitochondriale et la mitophagie. Ainsi, La quantification du nombre de copies d'ADN mitochondrial est fréquemment utilisée dans la littérature scientifique mais se révèle souvent très variable selon les études. Les causes de cette hétérogénéité sont multiples et impliquent notamment de la technique d'extraction d'ADN, les amorces ainsi que des gènes nucléaires utilisés comme référence. De plus, l'ADN mitochondrial présente la particularité d'être partiellement intégré sous forme de séquences plus ou moins longues dans l'ADN nucléaire (NUMTs : Nuclear Mitochondrial DNA fragments), des pseudogènes non codants mais qui peuvent fausser la quantification. Une méthode robuste et quantitative est donc indispensable pour pouvoir quantifier de manière fiable et reproductible le nombre de copies d'ADN mitochondrial. La plupart des techniques de PCR quantitative s'appuient sur l'emploi d'un intercalant type SYBR Green qui présente l'avantage d'être peu onéreux et assez facile à employer. Cependant, sa fixation à l'ADN n'étant pas spécifique, il est nécessaire d'amplifier séparément chacun des gènes cibles (généralement deux gènes de l'ADN mitochondrial et deux gènes de référence dans l'ADN nucléaire) au moins en duplicat pour chaque. Outre la consommation répétée d'ADN (qui peut-être limitante pour les échantillons précieux), les imprécisions de pipetage et la variabilité interplaque peuvent rendre les résultats de cette quantification assez variables.

Une alternative peut-être l'utilisation de PCR quantitative avec hydrolyse de sonde augmentant la spécificité et permettant le multiplexage. Dans le laboratoire, nous avons développé une PCR quantitative en triplex de manière à quantifier dans le même puits le nombre de copies de deux gènes mitochondriaux ainsi que d'un gène nucléaire servant de référence. Cette technique permet de détecter seulement sur 10 ng d'ADN la quantité des trois gènes cibles dans le même puits diminuant par là même les incertitudes de

pipetage ainsi que la variabilité interplaque. Une des difficultés de cette méthode réside dans la forte différence de réprésentation entre l'ADN mitochondrial et nucléaire dans les échantillons. En effet, l'ADN mitochondrial est présent entre 2 à 10 copies par mitochondrie avec plusieurs centaines de mitochondries par cellule contrairement à l'ADN nucléaire présent seulement sous forme de deux copies. Il convient donc de se placer dans des conditions de PCR et à des concentrations de nucléotides permettant l'amplification des trois gènes à la fois sans que l'amplification des gènes mitochondriaux ne limite celle des gènes nucléaires. Si tel n'est pas le cas, l'aspect quantitatif de la méthode disparaît puisqu'en cas d'augmentation de quantité d'ADN mitochondrial, l'amplification du gène nucléaire va se trouver retardée et la quantité d'ADN mitochondrial par cellule va se trouver surestimée. Au contraire, en cas de dépletion, si l'amplification du gène nucléaire était limitée par celle des gènes mitochondriaux, elle va se trouver facilitée et la quantité d'ADN mitochondrial va donc être sous-estimée. De manière à réduire ces écarts, l'utilisation de séquences répétées du génome telles que les séquences Alu ont déjà proposées dans la littérature (Fragouli et al., 2015) et sont actuellement en cours de test au laboratoire.
Outre son aspect de quantification du nombre de copies d'ADN mitochondrial, le positionnement des amorces sur l'ADN mitochondrial (dans les gènes MT-CO1 et MT-ND4) permet également à cette méthode d'estimer le pourcentage d'ADN présentant la délétion commune de l'ADN mitochondrial emportant le gène MT-ND4 et non le gène MT-CO1. Ainsi, devant un écart de Ct entre les deux gènes mitochondriaux, la présence de la délétion commune peut être suspectée.
Des essais sont en cours au laboratoire pour déterminer le nombre de copies d'ADN mitochondrial par PCR digitale. Les premiers résultats affichent une sensibilité de l'ordre de 0.2 à 0.5 ng d'ADN pour la quantification de l'ADN mitochondrial et 5 ng pour celui de l'ADN nucléaire. L'utilisation des séquences Alu comme référence nucléaire pourrait permettre de réduire cet écart et autoriser le multiplexage en réduisant considérablement la quantité d'ADN utilisée.

Matériel :

- iQ Multiplex Mastermix (BioRad)
- dNTPS mix (20 mM, Eurogentec)
- Amorces et sondes (séquences en annexes)
- ADN à 10 ng/µl

Protocole expérimental :

Pour un puits, faire le mélange réactionnel suivant et conserver à l'abri de la lumière:

- iQ Multiplex Mastermix 15 µl
- Amorces et sondes : 0.5 µl de chaque (concentration finales : amorces MT-CO1 et MT-ND4 : 500 nM, amorces Actine : 900 nM, sonde MT-CO1 et MT-ND4 : 200 nM et sonde actine 500 nM)
- dNTPS (20 mM) : 0.5 µl
- H20 : 9 µl

Distribuer 29 µl de mélange réactionnel par puits et ajouter 1 µl d'ADN à 10 ng/µl

Centrifuger les barrettes et placer dans le thermocycleur CFX (BioRad) et exécuter le programme suivant :

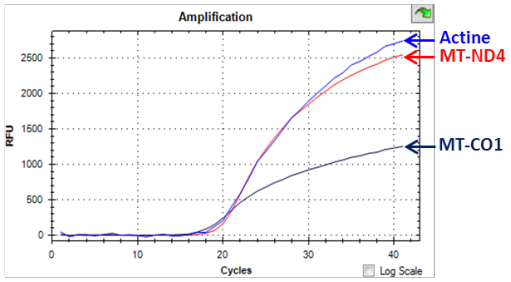
- 10 min à 95°C
- 40 cycles :
  - 15s à 95°C (dénaturation)
  - 30s à 54°C (hybridation)
  - Lecture de la fluorescence

Annexe 1 : séquences des amorces et des sondes Taqman

Nom du gène	Amorce sens (5'-3')	Amorce antisens (5'-3')	Sonde (5'-3')	Fluorochrome
Actine	CAGTGTGACATGGTGTATCT	AGAGGCGTACAGGGATAG	CCATGTACGTTGCTATCCAGGCTGT	6-FAM
MT-CO1	TCCACTATGTCCTATCAATA	GGTGTAGCCTGAGAATAG	CCATCATAGGAGGCTTCATTCACT	ATTO-700
MT-ND4	CGCACTAATTTACACTCA	GCTAGTCATATTAAGTTGTTG	ACATTCTACTACTCACTCTCACTGCC	Texas Red

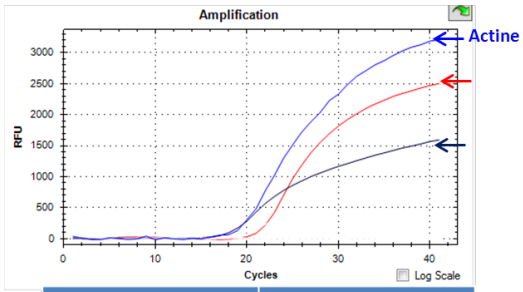
Annexe 2 : Exemple de résultats :

ADN de muscle de témoin



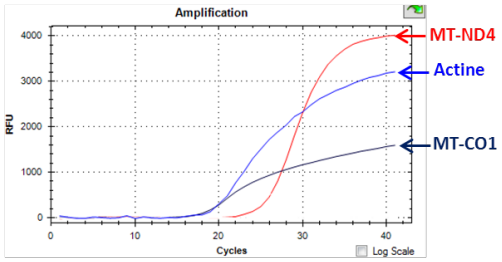
Gène	Ct
MT-CO1	16,15
MT-ND4	16,41
Actine	16,76

ADN de muscle présentant une délétion (1)



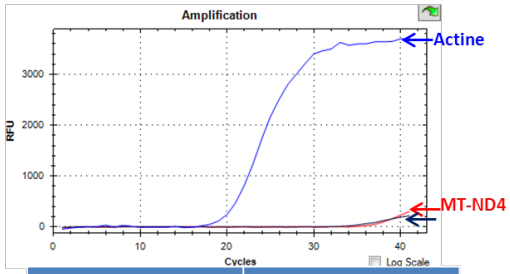
Gène	Ct
MT-CO1	16,54
MT-ND4	20,32
Actine	16,58

ADN de muscle présentant une délétion (2)



Gène	Ct
MT-CO1	16,68
MT-ND4	24,21
Actine	16,75

ADN de muscle présentant une déplétion



Gène	Ct
MT-CO1	39,49
MT-ND4	38,98
Actine	17,08

Délétion (1) : Délétion de 9650 à 14400 bornée par l'outil bioinformatique « Eklipse » à 78%
Délétion (2) : 8580-12908 bornée par l'outil bioinformatique « Eklipse » à 91%

Atelier B

Les méthodes de séquençage à haut débit

Au cours des dernières années est apparue une nouvelle génération de séquenceurs dits à haut débit. Ils opèrent en parallèle sur un très grand nombre de séquences courtes, reposant sur de nouvelles technologies physico-chimiques avec des débits jusqu'à 1000 fois supérieurs. Les capacités de ces nouveaux séquenceurs sont de plus en plus grandes pour un prix de revient de plus en plus faible.

De nombreuses applications médicales voient le jour, que ce soit par l'utilisation de panels de gènes choisis, ou le séquençage d'exome. Largement exploitée en recherche, cette technologie est aussi utilisée en diagnostic dans des laboratoires du monde entier, car c'est l'outil le plus performant dans le diagnostic de maladies rares avec anomalie du développement.

Préparation d'une librairie d'ADN mitochondrial pour NGS :

Principe:

La technique du NGS permet de détecter toute mutation de l'ADN mitochondrial (ADNmt) ainsi que les polymorphismes, c'est un séquençage entier de l'ADNmt,

Après amplification de l'intégralité de l'ADNmt par deux PCR longues chevauchantes de 8,5 kb chacune, on passe à l'étape de la fragmentation des produits amplifiés. L'étape suivante est la sélection sur gel des fragments de même taille et pour chaque patient on ajoute des codes barres (barcodes), de courtes séquences d'ADN qui permettent d'identifier ces patients et autorisent donc le multiplexage. Après des étapes d'amplification et de purification, les librairies de tous les patients sont regroupées dans un seul tube et chargées dans la puce.

Le principe général du séquençage Ion Torrent se base sur la différence du pH générée par la fixation d'une base, un ion H+ est libéré et la différence de pH ainsi engendrée est détectée et ainsi de suite jusqu'à la lecture de toutes les séquences. On obtient les résultats sous forme des fichiers VCF ou excel qui comprennent tous les variantes identifiés pour chaque patient, après on passe à l'étape d'analyse pour vérifier les mutations pathogènes.

Protocole expérimental :

1. Objet :

Ce mode opératoire explique comment réaliser une librairie barcodée d'ADN mitochondrial utilisable sur le séquenceur Ion Proton et le séquenceur S5 (Life Technologies) à partir de produits de long range PCR ou demi long range PCR.

2. Long range PCR et demi-long range PCR :

L'amplification d'ADNmt peut se faire en une seule longue pcr ou deux demi longue PCR selon le protocole suivant :

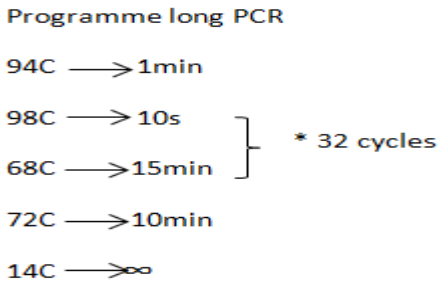
Mix LONG PCR	Pour 1 tube
H2O	29,5µl
Buffer 10X	5µl
dNTPs (2.5mM)	8µl
amorces F+R (20 µM)	5µl
Taq TAKARA	0.5µl

=>Distribution de 48µl par puits + ajout de 2µl d'ADN

Les amorces utilisées pour chaque réaction sont :

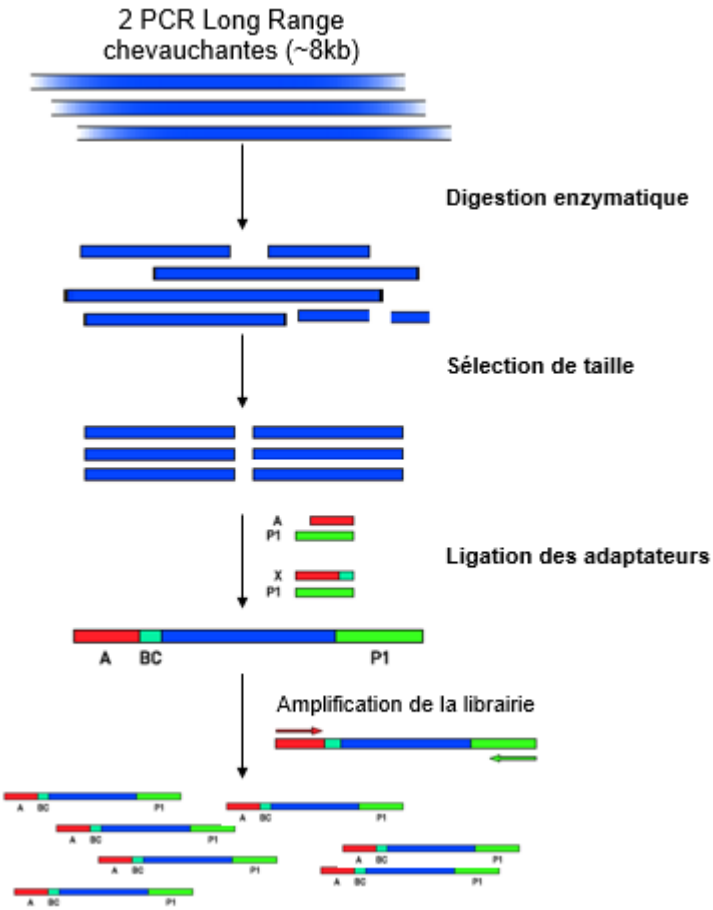
AMORCES	**amorce F 5'-3'**	**amorce R 5'-3'**
LONG PCR MITO	D24 : GGCACCCCTCTGACATCC	R32 : TAGGTTTGAGGGGGAATGCT
DEMI LONG 1	QCOX 1 FOR : TACGTTGTAGCCCACTTCCACT	QCYB REV : GCCCGATGTGTAGGAAGAG
DEMI LONG 2	QCYB FOR : AACTTCGGCTCACTCCTTGG	QCOX 1 REV : AGTAACGTCGGGGCATTCCG

Le programme de la PCR est le suivant :



Après amplification, les produits de la PCR sont contrôlés par électrophorèse sur gel d'agarose à 1%.

La préparation de la librairie est faite en utilisant le kit « **Ion Xpress™ Plus Fragment Library Kit » de Life Technologie**.



Principales étapes de la préparation de la librairie d'ADNmt

3. Purification des produits de PCR

3.1 Matériel requis

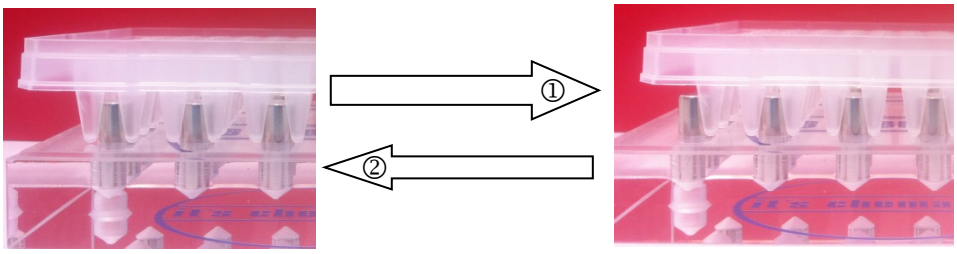
- Billes Agencourt® AMPure® XP reagent (Beckman Coulter).
- Ethanol 100%.
- Eau nucléase free.
- Tubes Eppendorf LoBind® 1,5 mL ou plaque 96 puits Applied Biosystem.
- Pipettes et cônes appropriés.
- Portoir magnétique pour tube ou pour plaque.

3.2 Mode opératoire

3.2.1 Préparation des réactifs

- Ramener les billes Agencourt® AMPure® XP reagent à température ambiante au moins 30 minutes avant l'utilisation. Les vortexer vigoureusement juste avant utilisation.
- Préparer extemporanément de l'éthanol à 70% (500µL/échantillon si travail en tube et 150µL/échantillon si travail en plaque).

3.2.2 Protocole de purification :

- Ajouter les billes Agencourt® AMPure® XP reagent (1,8x le volume de l'échantillon) à chaque échantillon. Mélanger par aspiration/refoulement et incuber la suspension à température ambiante pendant 5 minutes.
- Placer les tubes ou la plaque sur un portoir magnétique jusqu'à ce que la solution soit claire (au moins 3 minutes).
- Eliminer le surnageant sans toucher au culot.
- Rinçage à l'éthanol 70% : Ajouter 150 µL d'éthanol 70% par échantillon, incuber pendant 30 secondes, décaler la plaque d'une colonne sur le portoir magnétique de manière faire migrer le culot de billes, puis remettre la plaque à sa position initiale.
- 

- Eliminer le surnageant et répéter une seconde fois le lavage à l'éthanol 70%.
- Après le second lavage, éliminer le surnageant.
- Centrifuger rapidement la plaque et éliminer l'éthanol résiduel en laissant sécher le culot à l'air libre **au maximum 5 minutes.**
- Enlever les tubes du portoir magnétique et ajouter de l'eau nucléase (DL1 :100µl et DL2 : 125µl) free sur chaque culot de billes.
- Mélanger par aspiration/refoulement 5 fois.
- Placer les tubes ou la plaque sur un portoir magnétique jusqu'à ce que la solution soit claire (au moins 1 minutes).
- Transférer le surnageant contenant les amplicons purifiés dans un tube de 1,5 mL.
- Les amplicons purifiés peuvent être conservés à -20°C pendant 3 mois.

4. Dosage des amplicons purifiés :

Une quantification de l'ADN double brin est faite au Qubit (ThermoFischer) avec 1 µl de la librairie.

Effectuer le dosage à partir de 1µl du produit PCR purifié.

Réactif	Volume (µL) pour un échantillon
Qubit dsDNA HS buffer	199
Fluorochrome	1

5. Fragmentation enzymatique et purification :

5.1. Matériel requis

- Ion Xpress™ Plus Fragment Library Kit

Produit	Stockage
Ion Shear™ Plus 10X Reaction Buffer	Congélateur
Ion Shear™ Plus Enzyme Mix II	Congélateur
Ion Shear™ Plus Stop Buffer	Congélateur
Low TE	Congélateur avant ouverture Température ambiante après ouverture

- Autres réactifs :
  - Eau nucléase free ;
  - Agencourt® AMPure® XP reagent ;
  - Ethanol.
- Matériels et équipements :
  - Tubes Eppendorf® 1,5 mL LoBind ;
  - Tubes 0,2 mL ou barrettes ou plaque 96 puits ;
  - Thermocycleur ;
  - Pipettes et cônes appropriés ;
  - Portoir magnétique ;
  - Glace ;
  - Vortex ;
  - Centrifugeuse de paillasse.

5.2.Préparation des réactifs

- Ramener les billes Agencourt AMPure® XP à température ambiante au moins 30 min à l'avance et les vortexer vigoureusement juste avant l'utilisation.
- Préparer extemporanément de l'éthanol à 70% (300µL/échantillon si travail en plaque, 1,2 mL si travail en tube).

5.3.Fragmentation enzymatique

- Déterminer le volume à prélever pour obtenir 100ng de d'amplicons :
  Exemple :

N° échantillon	Concentration (ng/µL)	Quantité d'ADN (ng)	Produit de PCR (µL)	Eau nucléase free (µL)	Ion shear plus10X
1	61,2	100	1,63	32,13	5
	81,2	100	1,24		
2	26,2	100	3,817	29,01	5
	46	100	2,173		

- Vortexer Ion Shear™ Plus 10X Reaction Buffer et Ion Shear™ Plus Enzyme Mix II pendant 5 secondes, puis centrifuger brièvement et conserver sur la glace.
- Ajouter les réactifs dans l'ordre dans un tubes Eppendorf® 1,5 mL LoBind.

Ordre	Réactif	Volume (µL)
1	Produit de PCR	Y
2	Ion Shear™ Plus 10X Reaction Buffer	5
3	Eau nucléase free	35 –Y
	Total	40

- Vortexer vigoureusement, puis centrifuger brièvement.
- Ajouter à chaque échantillon 10 µL de Ion Shear™ Plus Enzyme Mix II, puis mélanger immédiatement le mélange par aspiration/refoulement 10 fois à l'aide d'une pipette réglée sur 40 µL
- Incuber à 37° pendant 15 minutes précisément.
- Après l'incubation, ajouter 5 µL de Ion Shear™ Plus Stop Buffer et vortexer vigoureusement.
- Conserver les tubes sur la glace.

5.4.Purification de l'ADN fragmenté

- Vortexer vigoureusement le réactif Agencourt® AMPure® XP.
- Ajouter 99µL de réactif Agencourt® AMPure® XP (1,8x le volume de l'échantillon) à chaque échantillon.
- Mélanger par aspiration/refoulement 5 fois, puis centrifuger brièvement.
- Incuber pendant environ 5 minutes à température ambiante.
- Centrifuger brièvement et mettre les tubes/la plaque sur un portoir magnétique pendant au moins 3 minutes, jusqu'à ce que la solution soit claire.
- Eliminer le surnageant sans toucher le culot.
- Laisser les tubes/la plaque sur le portoir magnétique et effectuer un rinçage à l'éthanol 70% (cf 3.2.2)
- Répéter le lavage à l'éthanol une seconde fois.
- Pour éliminer l'éthanol résiduel, centrifuger brièvement le tube ou la plaque, le placer sur un portoir magnétique et éliminer le reste de liquide à l'aide d'une pipette.
- Laisser sécher le culot pendant 3 - 5 minutes.
- Enlever les tubes du portoir magnétique et éluer les billes avec 25µL de low TE.
- Mélanger par aspiration refoulement 5 fois, puis vortexer pendant 10 secondes.
- Effectuer une centrifugation courte, puis placer les tubes ou la plaque sur un portoir magnétique pendant au moins 1 minute jusqu'à ce que la solution soit claire.
- Transférer le surnageant dans un tube 0,2 mL, une barrette ou une nouvelle plaque.

Le surnageant peut être conservé à -20°C jusqu'à la sélection de taille de la librairie non amplifiée à l'aide du système E-Gel® Sizeselect™.

6. Ligation des adaptateurs, réparation des coupures et purification

6.1. Matériel requis

- Réactifs fournis dans le kit Ion Plus Fragment Library :
    - 10X Ligase Buffer ;
    - DNA Ligase ;
    - Nick Repair Polymerase ;
    - dNTP Mix ;
    - Low TE.
- Réactifs fournis dans le kit Ion Xpress™ Barcode Adapters :
    - Ion Xpress™ P1 Adapter ;
    - Ion Xpress™ Barcode X (1 adaptateur barcode/librairie).
- Autres réactifs :
    - Eau nucléase free ;
    - Agencourt® AMPure® XP kit ;
    - Ethanol.
- Matériels et équipements :
    - Tubes Eppendorf® 1,5 mL LoBind ;
    - Tubes 0,2 mL, barrettes ou plaques 96 puits ;
    - Portoir magnétique ;
    - Pipettes et cônes appropriés ;
    - Vortex ;
    - Centrifugeuse de paillasse ;
    - Thermocycleur.

6.2.Préparation des réactifs

- Ramener les billes Agencourt AMPure® XP à température ambiante (au moins 30 min à l'avance) et vortexer juste avant l'utilisation.
- Préparer extemporanément de l'éthanol à 70% (300 µL/échantillon si travail en plaque, 1 mL si travail en tube).

6.3.Ligation des adaptateurs

- Ajouter les réactifs suivant dans un tube 0,2 mL, une barrette ou une plaque 96 puits :

Librairie barcodée	
Réactif	**Volume (µL)**
ADN fragmenté et purifié	≈ 25
10X Ligase Buffer	10
Ion P1 Adaptateur	2
Ion Xpress™ Barcode X	2
dNTP mix	2
Eau nucléase free	49
ADN ligase	2

Nick repair polymérase	8
Total	**100**

- Mélanger par aspiration/refoulement.
- Placer les tubes dans un thermocycleur et lancer le programme suivant :

Température	Temps
25°C	15 min
72°C	5 min
4°C	Hold (maximum 1h)

- Transférer tout le mélange réactionnel dans un tube tubes Eppendorf® 1,5 mL LoBind ou dans une plaque 96 puits pour réaliser l'étape de purification.

6.4.Purification

- Ajouter 120 µL de réactif Agencourt® AMPure® XP (1,2x le volume de l'échantillon) à chaque échantillon, et mélanger par aspiration/refoulement 5 fois.
- Incuber pendant environ 5 minutes à température ambiante.
- Centrifuger brièvement et placer le tube sur un portoir magnétique pendant au moins 3 minutes, jusqu'à ce que la solution soit claire.
- Eliminer le surnageant sans toucher le culot.
- Laisser le tube sur le portoir magnétique et effectuer un rinçage à l'éthanol 70% (cf. 3.2.2).
- Répéter le lavage à l'éthanol une seconde fois.
- Pour éliminer l'éthanol résiduel, centrifuger brièvement le tube ou la plaque, le/la placer sur un portoir magnétique et éliminer le reste de liquide à l'aide d'une pipette.
- Laisser sécher le culot pendant 3 - 5 minutes.
- Enlever les tubes/la plaque du portoir magnétique et ajouter 20µL de Low TE sur le culot pour le disperser. Mélanger par aspiration/refoulement 5 fois, puis vortexer pendant 10 secondes.
- Centrifuger brièvement et placer les tubes/la plaque sur un portoir magnétique pendant au moins 1 minute, jusqu'à ce que la solution soit claire.
- Transférer le surnageant dans un tube Eppendorf® 1,5 mL LoBind, une barrette ou une plaque 96 puits.
- Le surnageant peut être conservé à -20°C jusqu'à la sélection de taille de la librairie non amplifiée à l'aide du système E-Gel® Sizeselect™.

7. Sélection de la taille de la librairie avec le système E-Gel® Sizeselecttm

➢ Matériel requis

- <u>Réactif fournis dans le kit ion fragment plus :</u>
  - Low TE.
- <u>Autres réactifs, matériels et équipements :</u>
  - E-Gel® iBase et E-Gel Safe imager™ transilluminator ;

- E-Gel® SizeSelect™ 2% agarose gel ;
- PhiX174 DNA/HaeIII markers ;
- Eau nucléase free ;
- Pipettes et cônes appropriés.

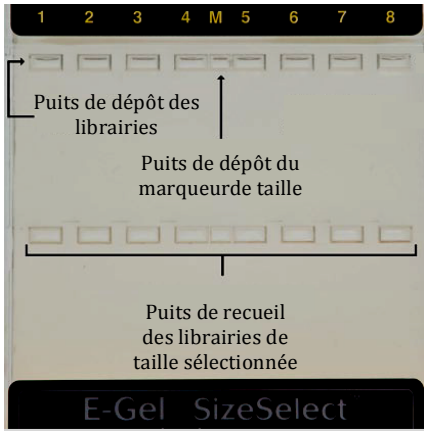
➢ **Préparation des réactifs**

Diluer le marqueur de taille au 1/40ème dans du Low TE (concentration finale 25ng/µL).

➢ **Sélection de la taille de la librairie**

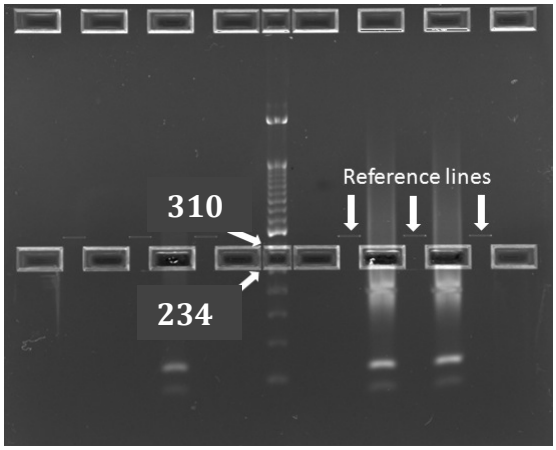
- Installer le E-Gel® iBase et le E-Gel Safe imager™ transilluminator, et placer le gel sur le système.

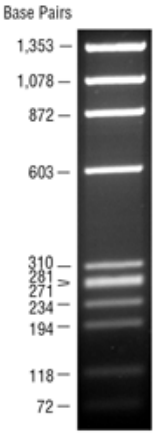
- Déposer l'intégralité de chaque librairie dans un puits de dépôt préalablement identifié. S'il reste des puits de dépôt vide, les remplir avec 25µL d'eau nucléase free.
- Déposer 10µL de marqueur de taille dans le puits de dépôt central du gel (voie M).
- Déposer 25µL d'eau nucléase free dans les puits de recueil 1 à 8, et 10µL dans le puits de recueil central (voie M).



- Placer le filtre ambré au-dessus du E-Gel® iBase, pour pouvoir observer la migration de la librairie.
- Sélectionner le programme "Run SizeSelect 2%" et programmer un temps de migration de 16 minutes.
- Appuyer sur "Go" sur l'E-Gel® iBase pour débuter l'électrophorèse.
- Au bout de 11 minutes, stopper la migration et ajouter environ 10 µL d'eau nucléase free dans tous les puits de recueil.
- Relancer la migration pendant environ 4 minutes.

- Pour les "200-base-read Ion Proton™ libraries" la taille cible des fragments est de 270pb, arrêter la migration quand le marqueur de 310 pb est à la limite supérieure du puits de recueil, et le marqueur de 234 pb à l'extrémité inférieure.





Marqueur de taille PhiX174 DNA/HaeII

- Récupération des librairies :
  - Transférer le contenu des puits de recueil à l'aide d'une pipette dans des barrettes de PCR ou une plaque 96 puits ;
  - Laver chacun des puits de recueil avec 10 µL d'eau nucléase free et pooler avec le reste de la librairie.
  - Au total, on récupère environ 30µL de librairie.
  -

8. Amplification de la librairie :

8.1.Matériel requis

- Réactifs fournis dans le kit Ion Plus Fragment Library :
  - Platinum® PCR SuperMix High Fidelity ;
  - Library Amplification Primer Mix.
- Autres réactifs :
  - Agencourt® AMPure® XP Kit ;
  - Ethanol.
  - Eau nucléase free.
- Matériels et équipements :
  - Thermocycleur
  - Tubes 0.2 mL ou plaque 96 puits ;
  - Tubes Eppendorf® 1,5 mL LoBind ;
  - Portoir magnétique ;
  - Pipettes et cônes appropriés ;
  - Vortex ;
  - Centrifugeuse de paillasse.

8.2.Préparation des réactifs

- Ramener les billes Agencourt AMPure® XP à température ambiante (au moins 30 min à l'avance) et vortexer juste avant l'utilisation.
- Préparer extemporanément de l'éthanol à 70% (300 µL/échantillon si travail en plaque, 1 mL si travail en tube).

8.3.Amplification de la librairie

- Pour chaque échantillon réaliser le mélange réactionnel suivant :

Réactif	**Volume (µL)**
Platinum® PCR SuperMix High Fidelity	100
Library Amplification Primer Mix	5
Librairie non amplifiée sélectionné par taille	25
Total	**130**

- Placer les tubes dans un thermocycleur et suivant le programme suivant :

Répétition	**Objectif**	**Température**	**Temps**
1 cycle	Dénaturation	95°C	5 min
8 cycles	Dénaturation Hybridation Elongation	95°C 58°C 70°C	15 sec 15 sec 1 min
—	Maintien à T°	4°C	∞

8.4.Purification de la librairie amplifiée

- Ajouter 195µL de Agencourt® AMPure® XP Kit à chaque échantillon, mélanger par aspiration/refoulement 5 fois, puis centrifuger brièvement.
- Incuber pendant 5 minutes à température ambiante.
- Centrifuger brièvement et placer le tube sur un portoir magnétique pendant au moins 3 minutes, ou jusqu'à ce que la solution soit claire.
- Eliminer le surnageant sans toucher le culot.
- Laisser le tube sur le portoir magnétique et effectuer un rinçage à l'éthanol 70% (cf. 3.2.2).
- Répéter le lavage à l'éthanol une seconde fois.
- Pour éliminer l'éthanol résiduel, centrifuger brièvement le tube/la plaque, puis le/la placer sur un portoir magnétique et éliminer le reste de liquide à l'aide d'une pipette.
- Laisser sécher le culot pendant maximum 5 minutes.
- Enlever les tubes du portoir magnétique et ajouter 20µL d'eau nucléase free ou de low TE sur le culot pour le disperser. Mélanger par aspiration/refoulement 5 fois. Et vortexer pendant 10 secondes.
- Centrifuger brièvement et placer les tubes/la plaque sur un portoir magnétique pendant au moins 1 minute, jusqu'à ce que la solution soit claire.
- Transférer le surnageant dans un tube Eppendorf® 1,5 mL LoBind identifié.
- Les librairies peuvent être conservé à -20°C pendant 6 mois.

5- Qualification et dilution des librairies :

➢ **Quantification de la librairie :**

- Effectuer le dosage de l'ADN au Qubit à partir sur 3 µL de librairie.

➢ **Dilution de la librairie :**

- Diluer les librairies à 70 pM (pour Ion Proton) et 55pM (pour leS5) dans de l'eau nucléase free. Les librairies diluées peuvent être conservées 48h entre 2 et 8°C.

LANCEMENT ION CHEF™

1. Objet

Ce mode opératoire explique quelles sont les étapes à réaliser pour lancer l'appareil Ion Chef™. L'Ion Chef™ est l'appareil qui effectue sur 2 puces simultanément la PCR en émulsion, l'enrichissement des sphères et le chargement de la puce en 15 heures 30 minimum.

2. Dilution des librairies

- Diluer les librairies dans de l'eau nuclease-free (à préparer au maximum 48h avant de lancer l'Ion Chef™) à 70pM (pour Ion Proton) et 55pM (pour leS5).

Library	Recommended concentration [1]
RNA-Seq	50 pM
TargetSeq™ Exome	50 pM
Ion AmpliSeq™ Exome	50 pM
Ion AmpliSeq™ Exome RDY	50–100 pM
Ion AmpliSeq™ Comprehensive Cancer Panel	50 pM
Human CEPH Control 200 library[2]	Dilute 1 µL into 24 µL Nuclease-free Water

[1] Recommendations are based on qPCR quantification. If libraries are quantified with a 2100 Bioanalyzer® instrument, a higher calculated concentration may need to be used for equivalent input.
[2] Obtained from the Ion PI™ Controls 200 Kit (Cat. no. 4488985)

- Préparer au moins 30µL de mélange des différentes librairies.
- Conserver les librairies prêtes à +4°C ou dans la glace jusqu'au chargement de l'Ion Chef™.

3. Préparation et chargement de l'Ion Chef™

3.1. Matériels et réactifs

- Conserver les réactifs et consommables dans de bonnes conditions et dans leur position droite.
- Tous les réactifs et consommables sont à usage unique.

• Réactifs

o Consommables fournis dans 1 boîte Ion PI™ Hi-Q™ Chef **Supplies** (température ambiante) en 4 exemplaires :
  - 2 Chip Adapters
  - 1 kit pour l'étape d'enrichissement : Enrichment Cartridge
  - 1 kit de cônes : Tip Cartridge
  - 1 plaque de PCR
  - 1 couvercle de plaque de PCR en aluminium
  - 2 couvercles de centrifugeuses : Recovery Station Lid
  - 12 tubes de centrifugeuses : Recovery Tubes

o Réactifs fournis dans la boîte Ion PI™ Hi-Q™ Chef **Solutions** (température ambiante 15-30°C) :
  - 4 boîtes de Ion PI™ Hi-Q™ Chef Solutions

o Réactifs fournis dans la boîte Ion PI™ Hi-Q™ Chef **Reagents** (-30°C à -10°C):
  - 4 boîtes de Ion PI™ Hi-Q™ Chef Reagents

o Autres réactifs
  - 2 librairies (25µL par librairie) : Ion PI™ IPSs amplifiées à 70pmol/L
  - Eau nucléase free

• Matériels :
  - Pipettes et cônes appropriés
  - Tubes Eppendorf® 1,5 mL LoBind
  - 2 Ion PI™ Chip v3

3.2. Préparation des réactifs

• A faire avant de commencer :
  - Avoir effectué son plan de run sur le Torrent Server.
  - Sortir au moins 45minutes avant la boîte Ion PI™ Hi-Q™ Chef Reagents du congélateur.
  - Allumer l'Ion Chef™ avant de le lancer grâce à l'interrupteur situé à l'avant en bas à droite de l'appareil (cf §5.2).

3.3. Présentation de l'appareil

Ion Chef™ vue de face

Ecra	Bouton d'allumage

Bouton d'ouvert ure de	Erreurs et

Ecran tactile de l'appareil

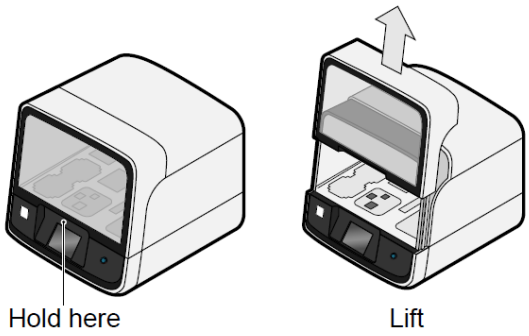
3.4. Allumage de l'appareil

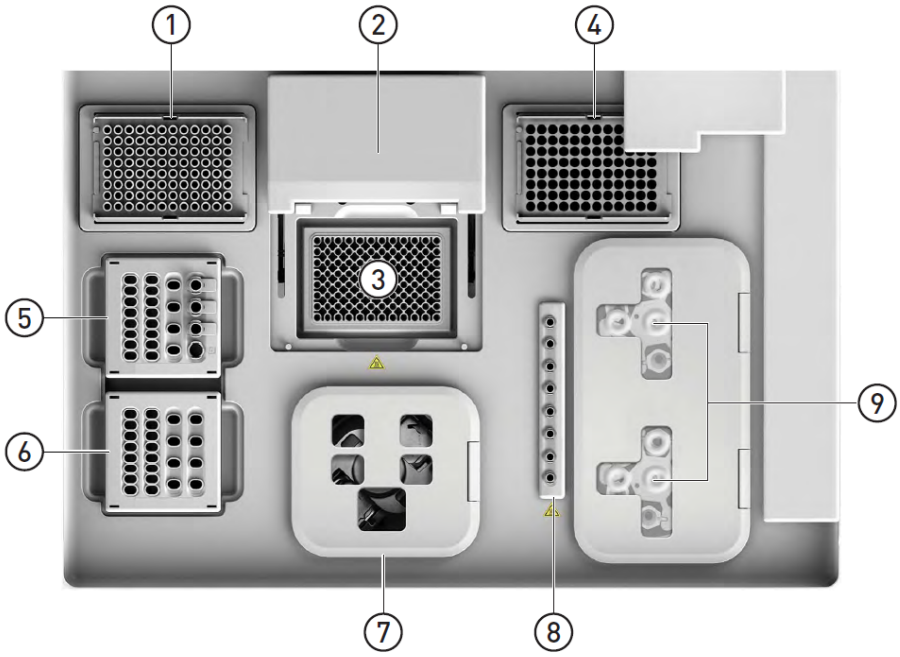
- Allumer l'appareil grâce à l'interrupteur situé à l'avant en bas à droite.

3.5. Installation des consommables et réactifs

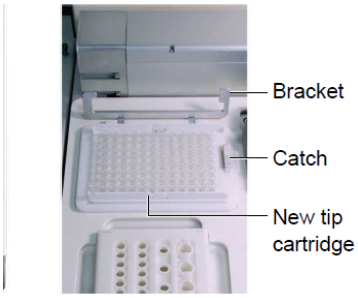
- *S'assurer que le nettoyage de l'appareil a bien été réalisé sinon l'effectuer avant le chargement des réactifs, nettoyer les compartiments de l'appareil avant le chargement si nécessaire.*

- Ouvrir la porte de l'appareil en appuyant sur l'icône de l'écran tactile puis attendre le déclic avant de faire glisser la porte vers le haut et de la bloquer (butée).



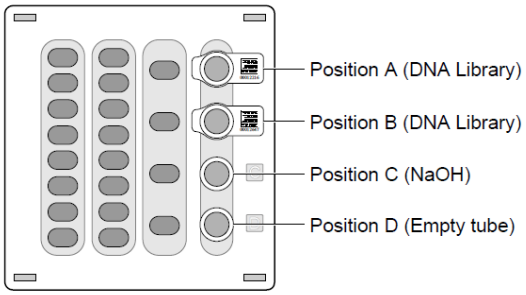


- Prendre la boîte de cônes **vide** du run précédent laissée en position 1 et la placer en position 4 (poubelle), si cela n'a pas été fait. Changer de gants.
- Placer une nouvelle boîte de cônes en position 1 (Ion PI™ Hi-Q™ Chef Tip Cartridge)
    - Enlever l'opercule de la nouvelle boîte de cônes.

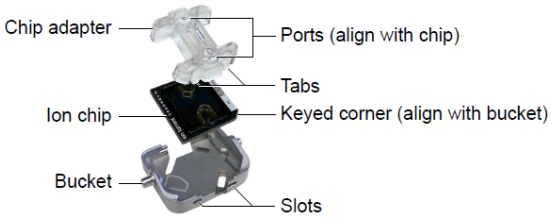


    - Fermer le clapet pour bloquer la boîte en le verrouillant dans le loquet.
- Disposer la plaque de PCR en position 3 et faire glisser sous le couvercle du thermocyleur (position 2) le couvercle en aluminium de la plaque de PCR.

- Charger le kit de réactifs Ion PI™ Hi-Q™ Chef **Reagents** en position 5 (remise à température ambiante depuis 45minutes).
    - Tapoter le kit de réactifs afin de placer les réactifs dans la partie inférieure des compartiments.
    - Déboucher les 4 tubes du kit (les 2 tubes pour les librairies, le tube NaOH et le tube vide) et les positionner dans l'appareil.



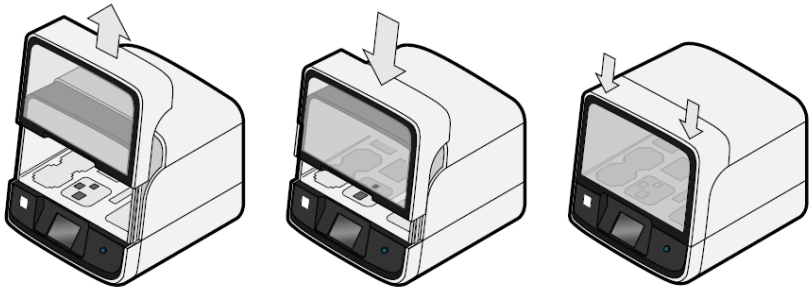
    - Mettre 25µL de chaque librairie dans les tubes A et B (attention de bien respecter les numéros des références saisies dans le plan de run) – bien enfoncer les tubes dans leur emplacement.
- Charger le kit de solution Ion PI™ Hi-Q™ Chef **Solutions** en position 6 (tapoter le kit afin de placer les réactifs dans la partie inférieure des compartiments).
- Placer les consommables des centrifugeuses (position 9).
    - Charger 6 tubes dans les 6 plots des 2 centrifugeuses.
    - Disposer un couvercle transparent en plastique sur chaque centrifugeuse.
    - Fermer le couvercle commun des 2 centrifugeuses.
- Charger le kit d'enrichissement en position 8.
- Charger les puces dans la centrifugeuse qui leur est dédiée en position 7.
    - Placer chaque puce dans un des supports de la centrifugeuse et la bloquer en clippant la Chip Adapter au dessus.



    - Replacer les supports dans la centrifugeuse.
    - Fermer le couvercle de la centrifugeuse.
- Vérifier que tout soit bien en place et que les kits soient bien positionnés dans leurs emplacements respectifs avant de lancer l'appareil.

4. Lancement de l'Ion Chef™

- Sur l'écran tactile de l'appareil, appuyer sur "Set up Run".
- "Step by step" : fermer la porte de l'appareil en la levant avant de l'abaisser et de la verrouiller des 2 côtés.



- Appuyer sur « Start Check», pour débuter la vérification de chargement puis attendre que l'instrument scanne les barre codes des consommables et réactifs.
- Si l'instrument détecte des anomalies (consommables manquants), il affiche des avertissements qu'il faut corriger avant de poursuivre.
- Quand « Deck Scan Complete » appuyer sur « Next ».
- Vérifier les informations du « Data Destination » : le nom du kit de l'IC, le type de puce et le barre codes des puces, sélectionner, si besoin, avec le menu déroulant le plan de run pour chaque puce et si besoin les modifier, puis "Next".
- Sélectionner la proposition du « Run Options » : « Timer », et rentrer l'heure à laquelle on veut récupérer nos puces le lendemain et appuyer sur « NEXT ».

Run Options	**Récupération des puces/échantillons**
Time	Immédiatement après le run (12.75h après le lancement du run)
Pause	Run plus long (jusqu'à 10.8h de plus)

Rq: Calculer pour avoir le temps d'effectuer l'initialisation du séquenceur avant de récupérer les puces.

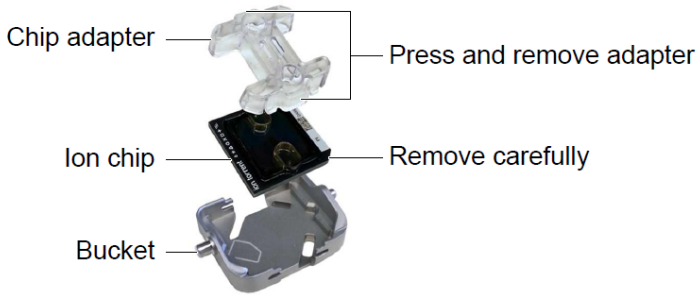
- puis sélectionner « yes ».
- Appuyer sur "Start run" pour que le run débute.

Rq: Possibilité d'interrompre le run à n'importe quel moment en appuyant sur "Cancel" puis "Yes" mais le run sera définitivement arrêté.

- Lorsque le run est terminé (« Run Complete »), appuyer sur « NEXT » et passer au déchargement de l'Ion Chef™ et au séquençage des puces immédiatement.

5. Déchargement de l'Ion Chef™

- Lorsque le run est terminé, appuyer sur l'image d'ouverture de porte et attendre le déclic pour glisser la porte vers le haut et la bloquer (buttée).
- Ouvrir le couvercle des centrifugeuses des puces et prendre l'ensemble portoirs-puces-adaptateurs.
- Récupérer les puces en les dissociant des portoirs en appuyant sur les 2 extrémités de l'adaptateur, replacer les portoirs dans la centrifugeuse et jeter les adaptateurs.



- Mettre une des 2 puces dans la boîte en plastique de conservation des puces et placer un morceau de parafilm au dessus de la puce avant de fermer la boîte et de la mettre à +4°C (max 6h pour lancer la puce sur le séquenceur – dans ce cas, sortir la puce 20 minutes avant du réfrigérateur et la placer à l'abri de la lumière pour qu'elle revienne à température ambiante).
- Procéder immédiatement au séquençage de la deuxième.
- Enlever et jeter tous les consommables et réactifs de l'Ion Chef™ à l'exception de la boîte de cônes vide en position 1 et de la boîte de réactifs en position 5
    - Enlever et jeter la plaque de PCR
    - Enlever et jeter la boîte de cônes usagés en position 4
    - Enlever et jeter la boîte de solutions Ion PI™ IC Solution 200
    - Enlever et jeter la boîte d'enrichissement
    - Enlever et jeter tous les consommables des centrifugeuses (couvercles, tubes)

Rq : Vérifier qu'il n'y ait pas d'excès de liquide présent dans la centrifugeuse, sinon procéder au nettoyage des centrifugeuses à l'isopropanol ou éthanol.

    - Refermer le couvercle de la centrifugeuse de puces.
- Changer de gants et placer la boîte de cônes vide en position 4 pour le prochain run.
- Enlever la boîte de réactifs Ion PI™ IC Reagents 200 de la position 5, fermer les tubes contenant les librairies et les conserver à +4°C en attendant d'effectuer le dosage Qubit reflétant le taux de multiclonalité si nécessaire, puis jeter la boîte de réactifs.
- Effectuer un cycle de décontamination UV de l'appareil.
- Eteindre l'appareil grâce à l'interrupteur situé à l'avant en bas à droite.

LANCEMENT S5 SYSTEM

Ce mode opératoire explique quelles sont les étapes à réaliser pour lancer le séquenceu S5.

1. Préparation et initialisation du système S5 :

Rq : pensez à sortir la cartouche de réactifs gardés à -20C, 45min avant l'utilisation.

✓ *Réactifs*

- Conserver les réactifs dans de bonnes conditions et dans leur position droite.
- Tous les réactifs et consommables sont à usage unique.

• Réactifs
- 4 bouteilles de solution de lavage
- Une bouteille de solution de nettoyage
- 4 cartouches de réactifs

✓ *Initialiser le séquenceur S5 :*

1. Dans le menu principal de l'écran de l'instrument, appuyez sur Initialiser. La porte se déverrouille.
2. Retirez la bouteille de solution de lavage Ion S5 ™ pour accéder au réservoir de déchets, puis retirez et videz le réservoir.
3. Réinstallez le réservoir de déchets vides.
4. Remplacez la cartouche de réactifs Ion S5 ™ par une nouvelle cartouche équilibrée à température ambiante.
5. Retirez le capuchon rouge de flacon de la solution de lavage Ion S5 ™ et installez-le.
6. Assurez-vous que la puce de séquençage utilisée de la séquence précédente est correctement placée dans la pince à puce et que la pince à copeaux est enfoncée complètement.
7. Si nécessaire, installez un nouveau flacon de solution de nettoyage Ion S5 ™.
8. Fermez la porte, puis touchez Suivant.
9. Lorsque l'initialisation est terminée (~ 30-40 minutes), touchez Accueil.

L'instrument est maintenant prêt pour le séquençage.

2. Séquençage :

Nous vous recommandons de mettre en ordre les puces chargées sur le Séquenceur Ion S5 ™ ou Ion S5 ™ XL aussitôt que possible après le chargement des puces et l'initialisation de l'instrument.
1. Après la fin de l'initialisation, appuyez sur Exécuter dans l'écran de l'instrument. La porte déverrouille.
2. Retirez la puce de séquençage utilisée, puis insérer la 1ère puce chargée.
3. Fermer la porte de l'instrument, puis appuyer sur suivant.
4. Assurez-vous que le plan remplis correspond au code de la puce insérée, puis appuyez sur suivant.
5. Si cette séquence doit être la première de deux séquences de cette initialisation, désactivez la case à cocher « Activer le nettoyage post-exécution ».
6. Confirmer que la porte de l'instrument est fermée, puis appuyer sur Démarrer pour lancer le séquençage.
7. à la fin du séquençage de la 1eère puce, insérer la deuxième en suivant les mêmes étapes.

Lorsque le séquençage est terminé, l'instrument effectue automatiquement la procédure de nettoyage, sauf si la case Activer le nettoyage post-exécution a été désélectionnée. Après le nettoyage, l'écran revient au menu principal.

Atelier C

Extraction ADN et haplogroupage par technique RFLP

Partie I : Extraction ADN :

L'extraction d'ADN à partir de prélèvement de cellules buccales est faite en utilisant le kit « High Pure PCR Template Preparation Kit », Roche Diagnostics :

- Après prélèvement, il faut tremper la cytobrosse pendant une nuit dans 800µl de PBS 1X puis bien l'égoutter dans le PBS le matin.

Remarque: Avant de commencer la purification, chauffer le tampon d'élution à + 70 °C.

- **Protocole expérimental :**

1. Ajouter 600µl de tampon de liaison.
   • Ajouter 120µl de protéinase K (reconstitué).
   • Mélanger immédiatement et incuber à + 70 ° C pendant 15 min.
2. Ajouter 300µl d'isopropanol et bien vortexer.
3. • Insérez la colonne dans un tube collecteur.
   • Ajouter 600µl de l'échantillon sur la colonne et centrifuger pendant 1 min à 11000×g.
   • Jeter l'éluât puis redéposer 600µl et refaire la centrifugation.
   • Jeter l'éluât puis redéposer le reste et centrifuger pendant 1 min à 11000×g.
4. Après centrifugation:
   •Jeter l'éluât et le tube collecteur.
   • Combinez la colonne avec un nouveau tube collecteur.
   • Ajouter 500 µl de tampon d'élimination des inhibiteurs sur la colonne.
   • Centrifuger pendant 1 min à 8 000 × g.
5. Jeter l'éluât et le tube collecteur.
   • Combinez la colonne avec un nouveau tube collecteur.
   • Ajouter 500µl de tampon de lavage sur la colonne.
   • Centrifuger pendant 1 min à 8 000 × g et jeter l'éluat.
6. Répéter l'étape 5.
7. Après avoir éliminé l'éluat et changé de tube collecteur
   • Centrifuger tout l'ensemble pendant 10s supplémentaires à une vitesse maximum.
   Le temps de centrifugation supplémentaire assure l'élimination du tampon de lavage résiduel.
   • Jeter le tube collecteur.
8. Pour éluer l'ADN:
   • Insérez la colonne dans un tube de microcentrifugeuse stérile de 1,5 ml.
   • Ajouter 100 µl de tampon d'élution, préchauffé à 70°C sur la colonne, attendre 1 minute
   • Centrifuger le tube pendant 1 min à 8 000 × g.
   Recommencer cette étape en ajoutant 100 µl de tampon d'élution sur la colonne et centrifuger 1 min à 8 000 xg.
9. Le tube contient l'ADN purifié qui peut être utilisé directement ou stocké entre +2 et + 8°C ou -15 à -25 ° C pour une analyse ultérieure.

10. Après l'extraction, la concentration ainsi que la pureté de l'ADN est déterminée grâce au nanodrop.

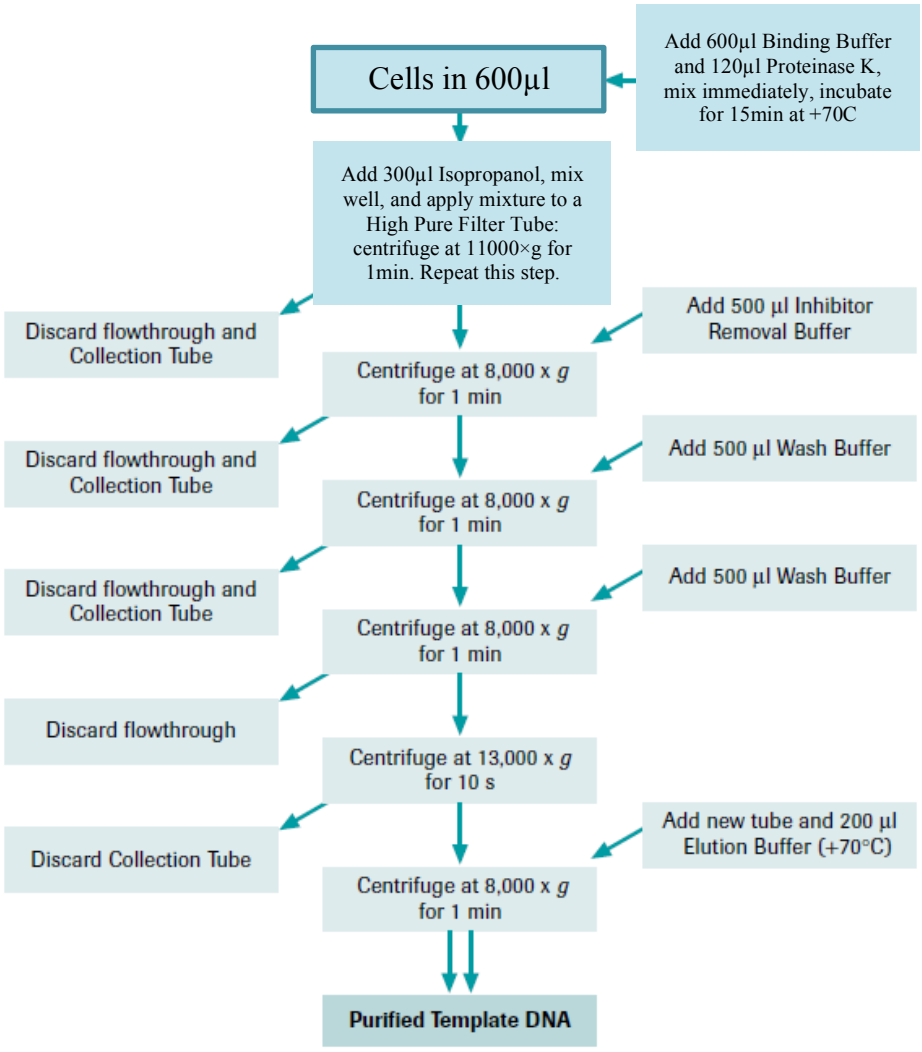


Schéma du protocole d'extraction d'ADN par le kit « High Pure PCR Template Preparation Kit »

Partie II : Détermination des Haplogroupes par technique RFLP :

La détermination des haplogroupes se fait par une technique simple appelée RFLP (Restriction Fragment Length Polymorphism) ou Polymorphisme de longueur de fragment de restriction. Le principe de cette technique se base sur une amplification génique de l'ADN extrait par des amorces spécifiques, suivie d'une digestion enzymatique par des enzymes de restrictions.

1- Amplification par PCR (Polymerase Chain Reaction) :

a- Principe :

L'une des propriétés de toutes les polymérases est de synthétiser le brin complémentaire à partir d'une amorce. Cette propriété est indispensable à la stabilité de l'information cellulaire, et elle est mise en profit dans la technique PCR décrite par Mullis en 1985, pour amplifier par réplication successive d'une séquence désirée.

La technique PCR permet l'amplification exponentielle d'une région spécifique d'un acide nucléique donné par l'utilisation d'une polymérase thermostable, d'amorces et des quatre désoxyribonucléotides (dNTP). L'ensemble est soumis à une série de cycles de température afin d'en obtenir une quantité suffisante de la séquence d'ADN désirée.

L'amplification se déroule en plusieurs cycles comprenant les étapes suivantes :

- **Une dénaturation** de l'ADN double brin par chauffage entre 90-95°C ;
- **Une hybridation** des amorces qui s'effectue à une température qui correspond à la température de fusion Tm du couple d'amorces (40-70°C) ;
- **Une élongation** de l'ADN qui se fait à partir de l'extrémité 3'OH libre de l'amorce dans le sens 5' 3'. La durée de polymérisation dépend de la taille du fragment d'ADN à amplifier.

Le nombre de cycle est variable, généralement situé entre 25 et 35 cycles. Ainsi que la durée et la température des différentes étapes sont des paramètres spécifiques qui sont définis pour chaque expérience.

Les composantes d'une réaction PCR :

L'ADN matrice : Les meilleurs résultats sont obtenus avec de l'ADN parfaitement purifié.

Le Tampon de la Taq polymérase : il sert à maintenir la stabilité et le pH du milieu réactionnel au niveau optimal pour la Taq polymérase.

Les amorces sens et anti-sens : ce sont des petites séquences d'ADN d'environ 20 bases. Elles sont capables de s'hybrider de façon spécifique grâce à la complémentarité des bases, ainsi que leur extrémité 3'OH libre vont servir d'amorces pour l'ADN polymérase Les concentrations d'amorces habituellement utilisées sont comprises entre 0.2 et 0.8µM.

Le chlorure de magnésium (MgCl2): c'est l'élément clef de l'hybridation des amorces qui dépend de la température et de la concentration ionique. En plus, c'est un cofacteur qui catalyse le fonctionnement de l'ADN polymérase. Les concentrations optimales sont en général comprises entre 0,5 et 2,5mM.

Les dNTPs : Ce sont les monomères de base utilisés par la Taq Polymérase pour synthétiser les brins d'ADN complémentaires.

La Taq polymérase ou ADN polymérase : c'est une enzyme thermorésistante extraite de la bactérie *Thermus aquaticus*. Elle est capable de résister à la température (+ 95°C) sans qu'elle perde son activité. Les quantités optimales d'enzyme sont comprises entre 1 et 2,5 unités.

b- Protocole expérimental :

Les séquences cibles sera amplifiée en utilisant les amorces F6809 et R7301 spécifiques pour la détermination de l'haplogroupes H et les amorces L13398 et H14559 pour la détermination de l'haplogroupe J.

		X Nb. Echt	
Eau	4,8		
Master Mix	7,5		
Primer-F(10µM)	0,35		
Primer-R(10µM)	0,35		
ADN (10ng/µl)	2		
Volume final	15	**V à dis**	**13**

La réaction PCR est effectuée dans les conditions suivantes : d'abord une dénaturation initiale à 94°C pendant 4 min ensuite, 35 cycles comprenant chacun une dénaturation à 94°C pendant 35s, une hybridation à 60°C pendant 35s et une élongation à 72°C pendant 40s et finalement une extension à 72°C pendant 7 min.

Les séquences des amorces utilisées sont les suivants :

Haplogroupe	Amorces sens	Amorces anti-sens	Taille
H	AGCATATTTCACCTCCG	TATTACTGCTGTTAGAGA	493bp
J	AAATAGGAGGACTACTCAAAA	GATTGTTAGCGGTGTGGTCG	1170bp

Les séquences amplifiées sont les suivants :

- **Hplogroupe H :**

Séquence normale :

AGCATATTTCACCTCCGCTACCATAATCATCGCTATCCCCACCGGCGTCAAAGTATTTAGCTGACTC
GCCACACTCCACGGAAGCAATATGAAATGATCTGCTGCAGTGCTCTGAGCCCTAGGATTCATCTTT
CTTTTCACCGTAGGTGGCCTGACTGGCATTGTATTAGCAAACTCATCACTAGACATCGTACTACACG
ACACGTACTACGTTGTAGCCCACTTCCACTATGTCCTATCAATAGGAGCTGTATTTGCCATCATAGG
AGGCTTCATTCACTGATTTCCCCTATTCTCAGGCTACACCCTAGACCAAACCTACGCCAAAATCCAT
TTCACTATCATATTCATCGGCGTAAATCTAACTTTCTTCCCACAACACTTTCTCGGCCTATCCGGAA
TGCCCCGACGTTACTCGGACTACCCCGATGCATACACCACATGAAACATCCTATCATCTGTAGGCT
CATTCATTTCTCTAACAGCAGTAATA

Séquence mutée :

AGCATATTTCACCTCCGCTACCATAATCATCGCTATCCCCACCGGCGTCAAAGTATTTAGCTGACTC
GCCACACTCCACGGAAGCAATATGAAATGATCTGCTGCAGTGCTCTGAGCCCTAGGATTCATCTTT
CTTTTCACCGTAGGTGGCCTGACTGGCATTGTATTAGCAAACTCATCACTAGACATCGTACTACACG
ACACGTACTACGTTGTAGCTCACTTCCACTATGTCCTATCAATAGGAGCTGTATTTGCCATCATAGG
AGGCTTCATTCACTGATTTCCCCTATTCTCAGGCTACACCCTAGACCAAACCTACGCCAAAATCCAT
TTCACTATCATATTCATCGGCGTAAATCTAACTTTCTTCCCACAACACTTTCTCGGCCTATCCGGAA
TGCCCCGACGTTACTCGGACTACCCCGATGCATACACCACATGAAACATCCTATCATCTGTAGGCT
CATTCATTTCTCTAACAGCAGTAATA

- **Hplogroupe J :**

Séquence normale :

AAATAGGAGGACTACTCAAAACCATACCTCTCACTTCAACCTCCCTCACCATTGGCAGCCTAGCAT
TAGCAGGAATACCTTTCCTCACAGGTTTCTACTCCAAAGACCACATCATCGAAACCGCAAACATAT

CATACACAAACGCCTGAGCCCTATCTATTACTCTCATCGCTACCTCCCTGACAAGCGCCTATAGCAC
TCGAATAATTCTTCTCACCCTAACAGGTCAACCTCGCTTCCCCACCCTTACTAACATTAACGAAAAT
AACCCCACCCTACTAAACCCCATTAAACGCCTGGCAGCCGGAAGCCTATTCGCAGGATTTCTCATT
ACTAACAACATTTCCCCCGCATCCCCCTTCCAAACAACAATCCCCCTCTACCTAAAACTCACAGCCC
TCGCTGTCACTTTCCTAGGACTTCTAACAGCCCTAGACCTCAACTACCTAACCAACAAACTTAAAAT
AAAATCCCCACTATGCACATTTTATTTCTCCAACATACTCGGATTCTACCCTAGCATCACACACCGC
ACAATCCCCTATCTAGGCCTTCTTACGAGCCAAAACCTGCCCCTACTCCTCCTAGACCTAACCTGAC
TAGAAAAGCTATTACCTAAAACAATTTCACAGCACCAAATCTCCACCTCCATCATCACCTCAACCC
AAAAAGGCATAATTAAACTTTACTTCCTCTCTTTCTTCTTCCCACTCATCCTAACCCTACTCCTAATC
ACATAACCTATTCCCCCGAGCAATCTCAATTACAATATATACACCAACAAACAATGTTCAACCAGT
AACTACTACTAATCAACGCCCATAATCATACAAAGCCCCCGCACCAATAGGATCCTCCCGAATCAA
CCCTGACCCCTCTCCTTCATAAATTATTCAGCTTCCTACACTATTAAAGTTTACCACAACCACCACC
CCATCATACTCTTTCACCCACAGCACCAATCCTACCTCCATCGCTAACCCCACTAAAACACTCACCA
AGACCTCAACCCCTGACCCCCATGCCTCAGGATACTCCTCAATAGCCATCGCTGTAGTATATCCAA
AGACAACCATCATTCCCCCTAAATAAATTAAAAAAACTATTAAACCCATATAACCTCCCCCAAAAT
TCAGAATAATAACACACCCGACCACACCGCTAACAATC

Séquence mutée :

AAATAGGAGGACTACTCAAAACCATACCTCTCACTTCAACCTCCCTCACCATTGGCAGCCTAGCAT
TAGCAGGAATACCTTTCCTCACAGGTTTCTACTCCAAAGACCACATCATCGAAACCGCAAACATAT
CATACACAAACGCCTGAGCCCTATCTATTACTCTCATCGCTACCTCCCTGACAAGCGCCTATAGCAC
TCGAATAATTCTTCTCACCCTAACAGGTCAACCTCGCTTCCCCACCCTTACTAACATTAACGAAAAT
AACCCCACCCTACTAAACCCCATTAAACGTCTGGCAGCCGGAAGCCTATTCGCAGGATTTCTCATT
ACTAACAACATTTCCCCCGCATCCCCCTTCCAAACAACAATCCCCCTCTACCTAAAACTCACAGCCC
TCGCTGTCACTTTCCTAGGACTTCTAACAGCCCTAGACCTCAACTACCTAACCAACAAACTTAAAAT
AAAATCCCCACTATGCACATTTTATTTCTCCAACATACTCGGATTCTACCCTAGCATCACACACCGC
ACAATCCCCTATCTAGGCCTTCTTACGAGCCAAAACCTGCCCCTACTCCTCCTAGACCTAACCTGAC
TAGAAAAGCTATTACCTAAAACAATTTCACAGCACCAAATCTCCACCTCCATCATCACCTCAACCC
AAAAAGGCATAATTAAACTTTACTTCCTCTCTTTCTTCTTCCCACTCATCCTAACCCTACTCCTAATC
ACATAACCTATTCCCCCGAGCAATCTCAATTACAATATATACACCAACAAACAATGTTCAACCAGT
AACTACTACTAATCAACGCCCATAATCATACAAAGCCCCCGCACCAATAGGATCCTCCCGAATCAA
CCCTGACCCCTCTCCTTCATAAATTATTCAGCTTCCTACACTATTAAAGTTTACCACAACCACCACC
CCATCATACTCTTTCACCCACAGCACCAATCCTACCTCCATCGCTAACCCCACTAAAACACTCACCA
AGACCTCAACCCCTGACCCCCATGCCTCAGGATACTCCTCAATAGCCATCGCTGTAGTATATCCAA
AGACAACCATCATTCCCCCTAAATAAATTAAAAAAACTATTAAACCCATATAACCTCCCCCAAAAT
TCAGAATAATAACACACCCGACCACACCGCTAACAATC

2- L'électrophorèse sur gel d'agarose :

L'électrophorèse est une méthode permettant la séparation de particule chargée sous l'action d'un champ électrique uniforme.

A pH neutre, les molécules d'ADN sont chargées négativement du faite de la présence du phosphate, et migrent vers l'anode quand elles sont soumises à un champ électrique, leur rapport charge/taille étant constant, ces molécules se séparent en fonction de la facilité avec laquelle elles se progressent à travers les mailles constituées par le gel. La séparation est ainsi assurer par l'effet de la filtration du gel. La vitesse de la migration d'une molécule d'ADN dépend de deux paramètres : sa taille et la concentration en agarose du gel, mais le voltage et la force ionique du tampon interviennent aussi.

L'ADN n'est pas visible à l'œil nu, pour cela on mélange la solution d'ADN avec le tampon de charge contenant :

♦ Du glycérol qui permet d'augmenter la densité de la solution d'ADN à déposer dans le gel afin de pouvoir l'entraîner vers le fond du puits, cela évite à l'échantillon de remonter à la surface du tampon et donc contaminer les autres puits.

♦ Marqueur de mobilité : c'est le bleu de bromophénol qui permet de visualiser la migration.

Enfin, la visualisation de l'ADN sur gel d'agarose est réalisée grâce à un colorant fluorescent, GelRed. Ce composé possède la propriété de s'intercaler entre les paires de bases des acides nucléiques. La révélation se fait par exposition du gel d'agarose aux radiations UV. Les fragments d'ADN se présentent sous forme de bandes fluorescentes.

b- Préparation du gel d'agarose 2% :

Pour 100ml du gel on ajoute 100 ml de tampon TBE 1X et 2g d'agarose

1- Faire fondre le gel sur la plaque chauffante, une fois fondue, l'agarose est laissé refroidir.

2- Couler le gel dans un support de gel portant les peignes bien installées ;

3- Enlever les peignes et placer le gel sur la cuve remplie par le TBE 1X ;

4- Les électrodes sont branchées en respectant la polarité, le voltage est de 100V ;

5- Arrêter la migration dès que le colorant de la migration de la solution du dépôt a parcouru le ¾ du gel.

6- Tremper le gel dans le Gelred pendant 10min et passer à l'observation.

3- Digestion enzymatique :

Les enzymes de restriction sont capables de digérer l'ADN au niveau de séquences spécifiques appelées sites de restriction. Il existe, pour chaque enzyme, un tampon spécifique et une température pour lesquels l'activité est optimale. La quantité d'enzyme utilisée pour une digestion est en fonction de la quantité d'ADN traité, de sa taille, du nombre de sites de restriction et du volume réactionnel.

Les produits de digestion de l'ADN par de telles enzymes forment une collection de fragments de taille variable en fonction du nombre et de l'emplacement des sites reflétant la présence et la fréquence des sites de restriction qui leur sont spécifique, c'est cette spécificité qui est exploité pour la mise en évidence d'une mutation au niveau du gène : une présence / absence de site de restriction entraîne une variation de longueur des fragments. Cette variation de longueur de fragments peut être mise en évidence par une digestion enzymatique de l'ADN génomique suivie d'une électrophorèse sur gel de polyacrylamide ou gel d'agarose.

a- Protocole expérimental :

La digestion du produit PCR est réalisée dans des tubes contenant 10µl du produit PCR amplifié de chaque échantillon, 1µl de l'enzyme ALuI et BstnI, 3µl du tampon de l'enzyme, 16µl de l'eau distillée stérile dans un volume final de 30µl. le mélange réactionnel est incubé à 37°C pendant 1 heure.

Le contrôle de la digestion se fait par électrophorèse sur gel d'agarose à 3%. En présence de l'haplogroupe H, nous obtiendrons 3 bandes : 244pb, 188bp et 61pb. A l'inverse, en absence d'haplogroupe H, on obtient 4 bandes : 244pb, 158bp, 61pb et 30bp.

En présence de l'haplogroupe J, nous obtiendrons une bande de 1170pb. A l'inverse, en absence d'haplogroupe J, on obtient 2 bandes : 873pb et 297bp.

Le tableau ci-dessus présente, les amorces et les enzymes utilisés pour la détermination des autres haplogroupes.

Hg	SNP/Indel	RFLP	Primers (5'-3')	Size
A	663	+III 663	534-553 / 725-706	191
B	9-bp deletion	N/A	8188-8207 / 8366-8345	178/169
C	13263	-II 13259	13001-13020 / 13403-13384	402

C	13263	+I 13262	13001-13020 / 13403-13384	402
C4	15204	-15204	15161-15180 / 15676-15658	515
D	5178A	-I 5176	5151-5170 / 5481-5464	330
D5	10397	+I 10396	10279-10296 / 10569-10550	290
E	7598	-I 7598	7367-7384 / 7628-7610	261
F1	12406	-I 12406	12385-12405 / 12576-12595	191
F1	12406	-II 12406	12385-12405 / 12576-12595	191
F2	7828	+I 7828	6890-6909 / 7131-7115	241
G	4883	+II 4830	4651-4670/ 4952-4934	301
H	7028	-I 7025	6890/ 7301	493
H2	4769	+I 4769	4651-4670 / 4952-4934	303
H8	13101	+II 13101	13001-13020 / 13403-13384	402
HV	14766	-II 14766	14407-14424 / 14810-14791	403
V	4577	-I 4577	4500-4519 / 4678-4659	178
I, X	1719	-I 1715	1615-1643 / 1899-1879	284
J	13708	-OI 13704	13537-13556 / 13851-13832	314
K	9055	-II 9052	8925-8953 / 9100-9081	175
M/N	10398	+I 10394	10279-10296 / 10569-10550	290
M	10400	+I 10398	10279-10296 / 10569-10550	290
N9	5417	+I 5417	5151-5170 / 5481-5464	330
R	12705	+II12705	12599-12618 / 12785-12766	186
R2	14305	-I 14304	13940-13959/ 14385-14366	445
R6	12285	-I 12285	12104-12124 / 12338-12309	234
T	15607	+I 15606	15409-15428 / 15728-15709	319
UK	12308	+I 12308	12104/ 12338	628
U1	13104	+I 13104	13001-13020 / 13403-13384	402
U4	4646	+RsaI 4646	4500-4519 / 4678-4659	178
U5	13617	-II 13617	13537-13556 / 13851-13832	314
W	8994	-III 8994	8925-8953 / 9100-9081	175
X	14470	+I 14465	14407-14424 / 14810-14791	403
X2e	15310	+DI 15310	15161-15180 / 15676-15658	515
Y	7933	+I 7933	7871-7890 / 8020-8001	149

Atelier D- E

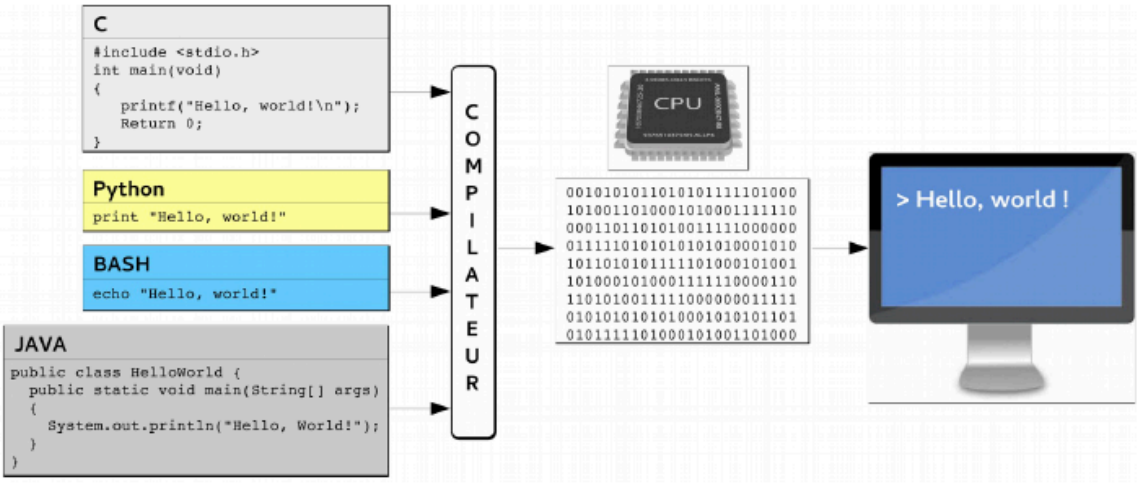
Analyse bioinformatique
ADMmt total / mitome

<u>Encadrant</u> : David Goudenège (david.goudenege@chu-angers.fr)

I – Notions et concepts informatiques

A) Une histoire de langages

Contrairement aux humains qui utilisent majoritairement un **système décimal**, les ordinateurs eux utilisent un **système binaire** basé sur 2 chiffres : le 0 et le 1. Ces deux chiffres correspondant aux deux états éléctriques possibles. Toutes les communications et instructions aux processeurs se feront donc par ce codage binaire. Ne parlant pas le binaire, si on veux « discuter » avec un processeur, on va devoir utiliser des langages de plus haut niveau, plus compréhensibles pour les humains mais convertibles en binaire par le processeur : on parlera de langages de programmation.
Il existe une multitude de langage de programmation car la seule contrainte est de disposer d'un « traducteur » (ou compilateur) faisant correspondre une syntaxe lisible en un code machine. Pour chacun des langages il existe des fonctions déjà codées sous forme de modules, librairies ou packages. Par exemple les modules BioPython, BioPerl et BioJava implémentent des fonctions dédiées à la bioinformatique (manipulation FASTA, lecture BAM ...).



Principes des langages de programmation. Un code source écrit dans un langage de haut niveau est traduit en binaire par un compilateur. Les instructions en binaire peuvent alors être interprétées par le CPU.

B) Une histoire d'algorithmique

Un **algorithme** peut être vu comme une « recette de cuisine » qui indique à un ordinateur les tâches à effectuer. Il précise une méthode pour procéder à des actions (ouvrir un fichier, trier une liste ...). Il nécessite de définir des **variables** et des **conditions logiques**. Une variable est un objet qui possède un nom qui fait référence a un espace de la mémoire vive dans lequel est stockée une donnée : la valeur de la variable. Les variables ont différents types : integer, float, string, booléen, liste, tableau, objet... Les conditions logiques vont elles permettre de structurer les algortithmes, elles peuvent être conditionnelles (si, sinon) ou répétitives (tant que, pour ...). Les algorithmes vont être codés suivant un langage de programmation dans des scripts, un programme peut comporter plusieurs scripts. Un pipeline (ou workflow) peut contenir plusieurs programmes qui s'enchainent suivant le principe que l'entrée de l'un et la sortie de l'autre.

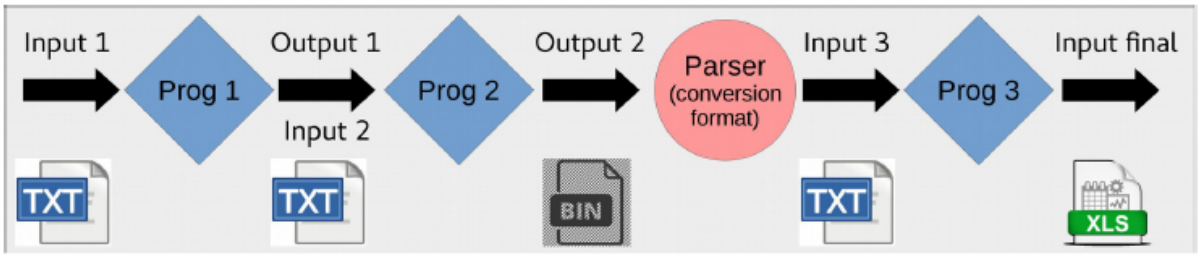


Schéma simplifié d'un pipeline informatique.

C) L interpréteur de commande (CLI)

Pour qu'un ordinateur soit capable de faire fonctionner un programme informatique, la machine doit être en mesure d'effectuer un certain nombre d'opérations préparatoires afin d'assurer les échanges entre le processeur, la mémoire, et les ressources physiques (périphériques). C'est le rôle du **système d'exploitation**, il peut être décomposé en 3 éléments principaux :

- Le **noyau** (kernel) représentant les fonctions fondamentales du système d'exploitation telles que la gestion de la mémoire, des processus, des fichiers, des entrées-sorties principales, et des fonctionnalités de communication.

- L'**interpréteur de commande** (**shell**) permettant la communication avec le système d'exploitation par l'intermédiaire d'un langage de commandes.
- Le **système de fichiers** (file system), permettant d'enregistrer les fichiers dans une arborescence..

La plupart des outils bioinformatiques ne disposent pas d'interface graphique et ne peuvent donc être lancés que via le shell. De plus ces outils sont souvent dédiés à des environnements unix (Mac ou Linux), nous allons donc nous concentrer sur le shell Linux. Pour accéder au shell, il faut ouvrir un **terminal** (console virtuelle) qui permet d'entrer ces instructions en ligne de commande. La navigation dans l'arborescence des fichiers et l'exécution de programmes se fera dans ce terminal en entrant des commandes définies. Les espaces permettent de séparer les arguments d'une instruction, comme dans la commande :

```
$ echo "Bonjour"
```

« echo » indique qu'il faut utiliser le programme d'affichage « echo » et l'argument (le mot à afficher est séparé par un espace). C'est pour cette raison que les informaticiens détestent les espaces dans les noms de fichiers car en ligne de commande ils sonts interprétés comme des séparateurs d'arguments. Dans l'exemple ci-dessus ainsi que dans le reste de cet atelier le caractère "$" indique le début de la ligne de commande (comme dans un vrai terminal), il n'est donc pas à recopier. Il existe un très grand nombre de commandes dans le shell, pour connaitre leur fonctionnement, il faut taper la commande suivie de "--help" (ex : « echo --help »).
Liste des commandes disponibles : https://ss64.com/bash/, sinon les plus utilisées :

NAVIGATION ARBORESCENCE
pwd Affiche le chemin absolu du répertoire courant (Print Working Directory).
cd [répertoire] Change de répertoire (Change Directory). Va dans répertoire ou dans le répertoire de l'utilisateur s'il n'y a pas d'argument. Si ".." est indiqué en argument, déplace dans le répertoire supérieur. Si "-" est indiqué en argument, déplace dans le répertoire précédent.
ls [répertoire] Liste le contenu des répertoires ou le nom des fichiers passés en arguments (liste le répertoire courant si pas d'argument).
mkdir [répertoire] Créé les répertoires (MaKe DIRectory) passés en arguments.
sudo [command] L'accomplissement de tâches privilégiées (ou tâches d'administration) s'effectue à travers un « filtre » puissant appelé **sudo**. Son principe est le suivant: - Toutes les tâches administratives ne peuvent être exécutées qu'à travers l'utilitaire d'administration sudo. - Lorsqu'un utilisateur tente d'exécuter une tâche administrative à travers le filtre sudo, cet utilitaire vérifie que cet utilisateur a le droit d'effectuer cette tâche. Dans le cas contraire, il bloque la tâche.
FICHIERS
cp [source] [destination]

Copie (CoPy) les fichiers source vers destination.
-i : demande confirmation avant écrasement (interactive)
-f : écrase sans demander confirmation (force)
-R ou -r : copie aussi les répertoires (recursive)
Attention : sans l'option -R (ou -r), la commande cp ne pourra pas copier les répertoires.

mv [source] [destination]
Renomme/déplace (MoVe) les fichiers source vers destination.
-i : demande confirmation avant écrasement (interactive)
-f : écrase sans demander confirmation (force)

rm [fichier]
Supprime (ReMove) les fichiers passés en arguments.
-i : demande confirmation avant suppression (interactive)
-f : supprime sans demander confirmation (force)
-R : supprime aussi les répertoires (recursive)

TRAITEMENT DE FICHIERS ET FILTRES

cat [fichier]
Affiche le contenu des fichiers texte passés en arguments.

wc [fichier]
Affiche le nombre de lignes, de mots et de caractères (Word Count) contenus dans les fichiers passés en arguments.
-l : affiche uniquement le nombre de lignes (line)
-w : affiche uniquement le nombre de mots (word)
-c : affiche uniquement le nombre de caractères (character)

grep [truc] [fichier]
Affiche uniquement les lignes, des fichiers passés en argument, correspondantes à l'expression rationnelle (ou expression régulière) regexp.
-v : inverse le résultat de la commande (affiche seulement les lignes ne correspondant pas à regexp)
-c : retourne le nombre de correspondances
-n : affiche les numéros des lignes correspondantes
-l : affiche les noms des fichiers contenant des lignes correpondant à regexp
-i : ne tient pas compte de la casse des caractères

head -<n> [fichier]
Affiche les n premières lignes (ou les 10 premières si n n'est pas spécifié).

tail <n> [fichier]
Affiche les n dernières lignes (ou les 10 dernières si n n'est pas spécifié).

sort [fichier]
Trie un fichier passé en paramètres.

LA REDIRECTION DE SORTIE STANDARD
Le caractère **">"** après une commande indique qu'il faut rediriger la sortie standard (l'écran) vers un fichier. Si on reprends notre exemple :

```
$ echo "Bonjour" > output.txt
```

Cette commande n'affichera rien dans le terminal mais écrira "Bonjour" dans le fichier "output.txt"

II – Manipulation et Formats de fichiers

A) Le plus ancien : FASTA (1988)

Les fichiers FASTA sont utilisés pour écrire les séquences nucléiques et protéiques. Ils doivent obligatoirement comporter une entête précédée du symbole ">" indiquant le titre de la séquence puis la séquence en passant à la ligne.

```
>sequence1_exemple
ATGCTCGCTGATGATAGATAGATAGATAGATAGTTTTGGTGATAGCCGCGCGC
CAAAACCCTTGGGATCGATGAGCTGGCCAAAACCCTTGGGATCGAT
>sequence2_exemple
AGCTCATCGATCCCAAGGGTTTTGGCCAGCTCATCGATCCCAAGGGTTTTGG
CGCGCGGCTATCACCAAAACTATCTATCTATCTATCTATCATCAGCGAGCAT
```

Astuce :

Un « grep » sur les signes supérieurs d'un fichier FASTA permet de lister les séquences présentes :

```
$ grep ">" hg19.fasta
>chr1
>chr2
...
>chrX
>chrM
```

Astuce :

Le programme **samtools** (dédiés à la manipulation des fichiers BAMs et SAMs) permet extraire des régions précises dans un fichier FASTA (à condition que ce fichier soit préalablement indexé) :

- Indexation

```
$ samtools faidx hg19.fasta (indexation)
```

- Extraction

```
$ samtools faidx hg19.fasta chrX:120181441-120181541
>chrX:120181441-120181541
AGGGCAACCCGCGCCGGACCCTTCCTTCCTAGTCGCGGGGAGTCTGAGAAAGCGCACCTG
TTCCGCGACCGTCACGCACCCCTCCTCCGCCTGCCGCGATG
```

B) Les incontournables du NGS : FASTQ & BAM

1) FASTQ

C'est le format qui permet de stocker les séquences des reads et leur qualité (phred) en sortie du séquençage (extension ".fastq" ou ".fq"). Il ne contient aucune information d'alignement. Pour les données pairées, on aura un donc 2 fichiers FASTQ. Les qualités des bases sont codées en ASCII.

```
@EAS54_6_R1_2_1_413_324 => la 1ère ligne doit commencer par "@" puis le nom du read
CCCTTCTTGTCTTCAGCGTTTCTCC          => la 2ème ligne contient la séquence
+                                  => la 3ème ligne commence par +
;;3;;;;;;;;;;;;7;;;;;;;88      => la 4ème ligne contient les qualités pour chaque base du read
@EAS54_6_R1_2_1_540_792 => nouveau read
```

```
TTGGCAGGCCAAGGCCGATGGATCA
+
;;;;;;;;;;;7;;;;;-;;;3;83
```

Astuce :
Pour convertir une qualité codée en ASCII en entier (ici ";"), il faut lancer ce script PERL :

```
$ perl -e 'print ord(";") - 33;'
26
```

2) BAM (Binary Alignment/Map)

C'est le format le plus utilisé pour stocker les informations d'alignement des reads. Il est en fait la version binaire du format texte SAM (Sequence Alignment/Map), il est prend ainsi moins d'espace de stockage.
Pour rappel, si l'on dispose d'un génome de référence on va aligner les reads (au format FASTQ) sur ce dernier : on parlera de **« mapping »**. Si l'on ne dispose d'aucune référence on utilisera des méthodes d'**« alignement *de novo* »**. Les outils de mapping les plus connus sont BWA, BOWTIE et TMAP (Ion Torrent). Des informations détaillés sur les spécifications du format BAM sont disponibles sur https://samtools.github.io/hts-specs/SAMv1.pdf.

Un fichier SAM/BAM peut être décomposé en 2 grandes sections :

- **L'entête (header)** : Chaque ligne commence par un "@", les champs sont séparés par des tabulations et chaque champs est sous format "TAG:VALUE".

```
@HD   VN:1.0 SO:coordinate      => VN=Format version / SO=Alignement triès ou non
@PG   ID:CASAVA   VN:CASAVA-1.8.2        => Informations sur les programmes utilisés
@SQ   SN:chr1     LN:249250621    => Informations sur la référence de mapping utilisée
@SQ   SN:chr2     LN:243199373
...
@SQ   SN:chr21    LN:48129895
@SQ   SN:chrM     LN:16571
@RG   ID:Sample2812 LB:1   ...          => Groupes de reads (même lane, même échantillons ...)
```

- L'alignement : Chaque ligne d'alignement doit comporter 11 champs obligatoires.

Colonne	Champ	Type	Description
1	QNAME	Chaine de caractère	Nom du read
2	FLAG	Entier	Code correspondant à des infos d'alignment
3	RNAME	Chaine de caractère	Nom de la reference ou s'aligne le read (ex : chr1)
4	POS	Entier	1ère position du mapping
5	MAPQ	Entier	Qualité de mapping
6	CIGAR	Chaine de caractère	Codification représentant l'alignment avec les matchs, les mismatches, les indels, le soft-clipping
7	RNEXT	Chaine de caractère	Nom de la référence mappant le read suivant
8	PNEXT	Entier	Position du prochain mapping le read suivant
9	TLEN	Entier	Longueur du read
10	SEQ	Chaine de caractère	Séquence du read
11	QUAL	Chaine de caractère	Qualité des bases (idem FASTQ)

Les fichiers BAMs/SAMs peuvent être **visualiser** grâce à des programmes comme Integrative Genome Viewer (IGV, http://software.broadinstitute.org/software/igv/) et GenomeBrowse (http://goldenhelix.com/products/GenomeBrowse/index.html).

Astuce :
Un fichier BAM peut être désaligner en le convertissant en format FASTQ, grâce à l'outil **"bamtofastq"** fourni avec le programme **"bedtools"**.

C) SAMTOOLS : le couteau suisse pour la manipulation des fichiers SAMs et BAMs

Ce programme est fournis avec de nombreux modules dédiés à la manipulation des fichiers SAMs et BAMs. Nous l'avons déjà vu pour son module **"faidx"** (cf partie II-A).

1) L'option "view"

C'est ce module qui permet de « débinariser » un BAM en SAM :

```
$ samtools view [options] <in.bam> [region...]
```

On peut lui ajouter de nombreuses options, voici quelques exemples :

- Afficher que le header du BAM.

```
$ samtools view -H test.bam
@HD     VN:1.0  SO:coordinate
@PG     ID:CASAVA       VN:CASAVA-1.8.2
@SQ     SN:chr1 LN:249250621
@SQ     SN:chr2 LN:243199373
...
@SQ     SN:chr21        LN:48129895
@SQ     SN:chrM         LN:16571
```

- Compter le nombre de reads total.

```
$ samtools view -c test.bam
1457505
```

- Créer un nouveau fichier BAM (option -b) en ne gardant que les reads inclus ou chevauchant la région de 120000 à 120500 du chromosome 2.

```
$ samtools view -b test.bam chr2:120000-120500 > test_reduce.bam
```

- Compter le nombre de reads sur cette même région

```
$ samtools view -c test.bam chr2:120000-120500
986
```

- Créer un fichier SAM en ne gardant que les reads ayant une qualité d'alignement >=20.

```
$ samtools view -q 20 test.bam > test_q20.sam
```

- Compter le nombre de reads totals avec une qualité de mapping >=10

```
$ samtools view -c -q 10 test.bam
1096447
```

- Créer un nouveau fichier BAM en ne gardant que les reads inclus ou chevauchant les régions définis dans un fichier .bed (cf partie II-D).

```
$ samtools view -b -L regions.bed test.bam > test_reduce_bed.bam
```

2) L'option "sort" et "index"

Pour être manipulé par samtools et par la plupart des programmes bioinformatiques, **un fichier BAM doit obligatoirement être trié et indexé.** Cela permet d'accélérer la récupération des données d'alignement qu'il contient. Le tri permet d'ordonner les reads par rapport à leur position d'alignement sur la référence (chr1, chr2 ..... chrM). L'index schématise les régions d'alignement pour optimiser la lecture des BAMs.

- Trier un fichier BAM.

```
$ samtools sort test.bam > test_sorted.bam
```

- Indexer le fichier BAM trier.

```
$ samtools index test_sorted.bam
```

Cela va générer un fichier test_sorted.bam.bai dans le répertoire courant.

Astuce :
L'outil « sambamba » permet de paralléliser l'exécution de samtools et donc d'augmenter la vitesse de ses différents modules.

D) L'indispensable : BED (Browser Extensible Data)

1) Le format

C'est le format de fichier qui permet d'indiquer des régions génomiques, c'est à dire des positions sur le génome, et d'y associer ou non des informations (annotations).
Par exemple, lors de séquençage de panel de gènes, on va définir un fichier BED listant toutes les régions amplifiées en y intégrant le nom des gènes et des exons. Comme tous fichiers tabulés, il peut s'ouvrir avec Excel. Son format est très simple, c'est un fichier tabulé dont seul les 3 premières colonnes sont obligatoires à savoir : **chromosome<tab>start<tab>end**
Il peut y avoir ou non une ligne d'entête qui commence par le symbole "#".

Ci dessous un exemple pour les positions des amplicons du gène OPA1.

chr3	193310840	193311068	AMPL7155534561	0	+	.	OPA1_Exon1
chr3	193310973	193311198	AMPL7158140951	0	+	.	OPA1_Exon1
chr3	193332442	193332641	AMPL7154847563	0	+	.	OPA1_Exon2
chr3	193332630	193332830	AMPL7158140948	0	+	.	OPA1_Exon2
...							
chr3	193415194	193415419	AMPL7155534582	0	+	.	OPA1_Exon29
chr3	193415408	193415610	AMPL7153773601	0	+	.	OPA1_Exon29

2) BEDTOOLS

A la manière de samtools, c'est le « couteau suisse » pour la manipulation des fichiers BED. Il va aussi manipuler des BAMs, des FASTA et des VCFs. Voici les commandes les plus utilisées :

- Trier un fichier BED et enregistrer dans un nouveau fichier.

```
$ bedtools sort -i panel_design.bed > panel_design_sorted.bed
```

- Calculer une couverture (=nombre de reads) d'un BAM sur différents intervalles d'un BED.

```
$ bedtools coverage -a panel_design.bed -b sample.bam
```

chr14	24563339	24563594	PCK2	utr_E1	**89**	**255**	**255**	**1.0000000**
chr14	24563614	24563643	PCK2	cds_E1	**91**	**29**	**29**	**1.0000000**
chr14	24566100	24566346	PCK2	cds_E2	**892**	**246**	**246**	**1.0000000**
chr14	24567411	24567596	PCK2	cds_E3	**990**	**185**	**185**	**1.0000000**
chr14	24567683	24567887	PCK2	cds_E4	**721**	**204**	**204**	**1.0000000**
chr14	24568257	24568445	PCK2	cds_E5	**994**	**188**	**188**	**1.0000000**
chr14	24568766	24568929	PCK2	cds_E6	**1159**	**163**	**163**	**1.0000000**
chr14	24569203	24569514	PCK2	cds_E7	**1577**	**311**	**311**	**1.0000000**
chr14	24569534	24569962	PCK2	utr_E7	**101**	**120**	**428**	**0.2803738**

La sortie est un fichier tabulé avec 4 colonnes en plus (en gras) : nombre de reads<tab>nombre de bases dans l'intervalle avec une couverture>0<tab>taille de l'intervalle<tab>% de bases dans l'intervalle avec une couverture>0

- Calculer une couverture (=nombre de reads) d'un BAM sur différents intervalles d'un BED mais position par position.

```
$ bedtools coverage -a panel_design.bed -b sample.bam -d
chr14   24563319        24569982        PCK2    utr_E1  1       1
chr14   24563319        24569982        PCK2    utr_E1  2       1
chr14   24563319        24569982        PCK2    utr_E1  3       1
chr14   24563319        24569982        PCK2    utr_E1  4       6
chr14   24563319        24569982        PCK2    utr_E1  5       6
chr14   24563319        24569982        PCK2    utr_E1  6       10
chr14   24563319        24569982        PCK2    utr_E1  7       10
chr14   24563319        24569982        PCK2    utr_E1  8       12
chr14   24563319        24569982        PCK2    utr_E1  9       12
```

La sortie est un fichier tabulé avec 2 colonnes en plus (en gras) : position dans l'intervalle<tab>nombre de reads

E) Le final : VCF

C'est le format qui permet de stocker les variants identifiés lors de l'étape de « calling ». Tout comme le fichier bed c'est un fichier tabulé comportant des lignes d'entête commencant par des "#" et une ligne par variant. Il peut donc être ouvert via Excel. Les lignes d'entêtes permettent de préciser les champs disponibles dans les lignes des variants. Ils peuvent contenir les variants pour un ou plusieurs échantillons.

Chacune des lignes variants est construite de la manière suivant :
- le chromosome
- la position du variant
- son identifiant, par exemple dans dbSNP
- la réference observée dans le génome utilisé
- la variation observée dans l'échantillon (dans le fichier BAM). Si plusieurs variations pour une même référence, les alternatives seront séparées par des ","
- la qualité du calling (score phred indiquant la fiabilité de ce variant)
- le filtre qui indique si le variant passé (PASS) ou non (q10) les filtres de l'outils de calling
- le champ "INFO" avec différentes informations sur la couverture, la fréquence ...
  (le détail des informations et leur acronyme seront dans le header)
- le champ "FORMAT" avec les informations sur le génotype d'un ou plusieurs échantillons
  (le détail des informations et leur acronyme seront dans le header)

Ci dessous, un exemple pour 1 échantillon (NA00001) :

```
##fileformat=VCFv4.2
##fileDate=20090805
##source=myImputationProgramV3.1
##reference=file:///seq/references/1000GenomesPilot-NCBI36.fasta
##contig=<ID=20,length=62435964,assembly=B36,species="Homo sapiens",taxonomy=x>
##phasing=partial
##INFO=<ID=NS,Number=1,Type=Integer,Description="Number of Samples With Data">
##INFO=<ID=DP,Number=1,Type=Integer,Description="Total Depth">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency">
##INFO=<ID=AA,Number=1,Type=String,Description="Ancestral Allele">
##INFO=<ID=DB,Number=0,Type=Flag,Description="dbSNP membership, build 129">
```

```
##INFO=<ID=H2,Number=0,Type=Flag,Description="HapMap2 membership">
##FILTER=<ID=q10,Description="Quality below 10">
##FILTER=<ID=s50,Description="Less than 50% of samples have data">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=GQ,Number=1,Type=Integer,Description="Genotype Quality">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth">
##FORMAT=<ID=HQ,Number=2,Type=Integer,Description="Haplotype Quality">
#CHROM POS ID        REF ALT  QUAL FILTER INFO                              FORMAT      NA00001
20 14370   rs6054257 G   A      29 PASS NS=3;DP=14;AF=0.5;DB;H2           GT:GQ:DP:HQ 0|0:48:1:51,51
20 17330   .         T   A      3  q10  NS=3;DP=11;AF=0.017              GT:GQ:DP:HQ 0|0:49:3:58,50
20 1110696 rs6040355 A   G,T    67 PASS NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:GQ:DP:HQ 1|2:21:6:23,27
20 1230237 .         T   .      47 PASS NS=3;DP=13;AA=T                   GT:GQ:DP:HQ 0|0:54:7:56,60
20 1234567 microsat1 GTC G,GTCT 50 PASS NS=3;DP=9;AA=G                    GT:GQ:DP 0/1:35:4
```

Rq : Il existe aussi un nouveau format basé sur le VCF mais dédié au génome complet : le "GVCF". Il comporte des informations supplémentaires.
(http://gatkforums.broadinstitute.org/gatk/discussion/4017/what-is-a-gvcf-and-how-is-it-different-from-a-regular-vcf)

Rq : Comme samtools et bedtools, il existe un vcftools permettant de manipuler les fichiers VCFs (https://vcftools.github.io/). Je ne le conseille qu'au bioinformaticien dans le sens ou les VCFs peuvent être analysés comme des tableaux excel classique.

III – Bioinformatique et diagnostic NGS
A) Le principe global

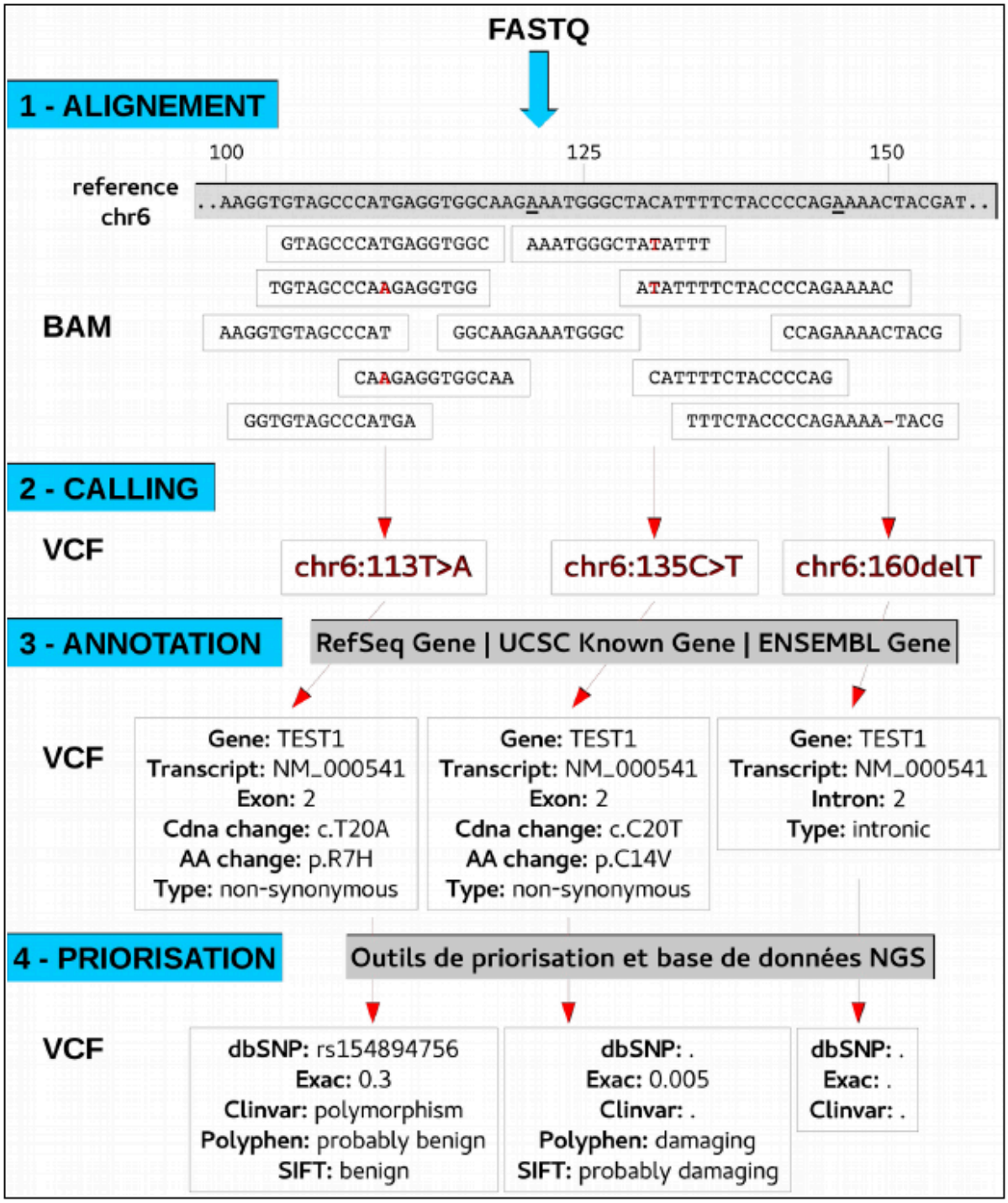


Schéma des 4 étapes principales utilisées dans un diagnostic NGS. Le schéma ne prends pas en compte les étapes de contrôle qualité, contrôle de couverture et du rendu final des résultats.

B) L'alignement

Dans le cas du diagnostic, l'alignement des reads obtenus se fait sur un génome de référence humain (hg19, hg20) ou mitochondrial (rCRS,CRS). Cette étape est cruciale, car toutes les autres étapes du diagnostic NGS en découleront. Elle est souvent réalisée par des outils fournis avec les séquenceurs, comme :

- CASAVA pour les technologies ILLUMINA.
- TMAP pour les technologies LIFE.

Je ne recommande pas de modifier les paramètres de ces outils ni de les changer car ils ont été optimisés pour leur technologie et sont donc ceux qui donnent les meilleurs résultats.

Il peut néanmoins être intéressant de les relancer dans le cas de pseudogènes et/ou de zones de duplication qui perturbent l'alignement.

De plus, un étape d'optimisation de l'alignement peut également être envisagée.

1) La gestion des pseudogènes et des duplications

Que ce soit pour le mitome ou pour d'autres panels, il peut arriver qu'une région amplifiée présente de forte similarité avec une autre région du gènome. Dans ce cas, au moment de l'alignement, l'outil ne saura pas où placer un read présentant un même score d'alignement pour 2 régions. Suivant les paramètres de l'outils, soit ce read sera rejeté, soit il sera distribué aléatoirement entre ces 2 régions mais avec un score d'alignement (mapQ) de zéro. Il en résultera donc un « trou » dans notre couverture.

Une solution peut être de masquer sur notre référence, les régions comportant des similarités avec notre design, et de faire un alignement sur cette référence « avec masque ».

- Récupérer les séquences amplifiées

```
$ bedtool getfasta -fi hg19.fasta -bed design_amplicon.bed
```

- Utiliser BLAT ou BLAST pour déterminer s'il existe des régions ayant de fortes similarités (>95%). Récupérer les positions de ces régions et les formater en BED (=pseudo_dupl.bed).
- Masquer ces régions dans une nouvelle référence

```
$ bedtool maskfasta -fi hg19.fasta -fo hg19_masked.fasta -bed pseudo_dupl.bed
```

- Si on ne dispose pas des fichiers FASTQ, on peut désaligner un BAM

```
$ bamToFastq -i sample.bam -fq sample.fq (Life)
```

ou

```
$ bamToFastq -i sample.bam -fq sample_R1.fq -fq2 sample_R2.fq (Illumina)
```

- Relancer un alignement

```
$ tmap mapall -n 8 -f hg19_masked.fasta -r sample.bam -v -Y -u --prefix-exclude 5 -o 2 stage1 map4 > sample_masked.bam (Life)
```

ou

```
$ bowtie2 -x hg19 -1 sample_R1_fq -2 sample_R2_fq -S sample.sam -p 8 -t (Illumina attention la sortie est un fichier SAM, il faudra utiliser samtools pour le convertir en BAM, le trier et l'indexer)
```

Rq : Pour aligner sur une référence il faut qu'au préalable le fichier d'index ait été créé :

```
$ tmap index -f hg19.fasta
$ bowtie2-build hg19.fasta hg19
```

Cette technique va forcer des reads normalement sur des régions hors-design à s'aligner sur notre design. On générera donc des faux-variants, mais au moins on est sur de ne pas avoir de faux-négatifs.

2) La suite logicielle GATK pour le réalignement et la recalibration

Elle permet d'améliorer un alignement autour de SNPs et d'indels connus. L'étape de recalibration permet de corriger des scores de qualité estimés par les séquenceurs.

- Premiere etape du realignement au niveau des INDELS avec GATK RealignerTargetCreator

```
$ java -jar GenomeAnalysisTK.jar -T RealignerTargetCreator -I sample.bam -R hg19.fasta -L design.bed -o design.intervals -known Mills_and_1000G_gold_standard.indels.hg19.sites.vcf
```

- Seconde etape du realignement au niveau des INDELS avec GATK IndelRealigner

```
$ java -jar GenomeAnalysisTK.jar -T IndelRealigner -I sample.bam -R hg19.fasta -L design.bed -targetIntervals design.intervals -o sample_realigned.bam
```

- Premiere etape du BQSR (Base Quality Score Recalibration) avec GATK BaseRecalibrator

```
$ java -jar GenomeAnalysisTK.jar -T BaseRecalibrator -I sample_realigned.bam -R hg19.fasta -L design.bed -knownSites dbSNP_b146_GRCh37p13.vcf -o sample.bqsr
```

- Deuxieme etape du BQSR (Base Quality Score Recalibration) avec GATK PrintReads

```
$ java -jar GenomeAnalysisTK.jar -T PrintReads -I sample_realigned.bam -R hg19.fasta -L design.bed -BQSR sample.bqsr -o sample_realigned_recalibrated.bam
```

Rq : Pour utiliser les outils GATK, il faut qu'au préalable le fichier .**dict** ait été créé. Pour cela, il faut utiliser un module du programme PICARD (https://broadinstitute.github.io/picard/):

```
$ java -jar picard.jar CreateSequenceDictionary R=hg19.fasta O=hg19.dict
```

C) Le calling

1) Généralités

Cette étape consiste à rechercher les variations vis à vis d'une référence d'un fichier BAM.

Il existe une pléthore d'outils de calling, les plus connus pour le constitutionnel étant :

- ***GATK UnifiedGenotyper*** (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_genotyper_UnifiedGenotyper.php)
- ***GATK HaplotypeCaller*** (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_haplotypecaller_HaplotypeCaller.php)
- ***Platypus*** (http://www.well.ox.ac.uk/platypus)
- ***LoFreq*** (https://sourceforge.net/projects/lofreq/)
- ***VarScan*** (http://dkoboldt.github.io/varscan/)
- ***Torrent Variant Caller*** (plugin de le Torrent Suite de LIFE)
- **et les autres…** (https://omictools.com/germline-snp-detection-category)

Comme souvent en bioinformatique, on utilise le principe « ***Unite and conquer*** », c'est à dire qu'on utilise plusieurs approches (outils) pour en faire un consensus. Il est donc plus que recommandé d'utiliser plusieurs outils de calling. Tous ces outils vont en effet avoir leur propre algorithme de détection avec des paramètres ajustables en entrée.

2) Exemples

- Lancer GATK UnifiedGenotyper

```
$ java -jar GenomeAnalysisTK.jar -T UnifiedGenotyper -I sample.bam -L design.bed -R hg19.fasta -o sample_GATKu.vcf -mbq 10 -glm BOTH -dt BY_SAMPLE -dcov 2000 -gt_mode DISCOVERY -rf BadCigar
```

-mbq : qualité de base minimale pour un variant
-glm : calling des SNPs, des INDELs ou des deux (BOTH)
-dt : type de downsampling pour un locus (NONE|ALL_READS|BY_SAMPLE)
-dcov : nombre de reads maximum pour le downsampling de couverture
-gt_mode : En mode DISCOVERY et non GENOTYPE_GIVEN_ALLELES
-rf : filtrer les reads ayant des erreurs dans les notations CIGAR

- Lancer Platypus

```
$ python Platypus.py callVariants --bamFiles=sample.bam --refFile=hg19.fasta --filterDuplicates=0 --output=sample_platypus.vcf --minReads=10 --minMapQual=20 --minBaseQual=10 –minVarFreq=0.1
```
--filterDuplicates : ne pas filtrer les reads dupliqués
--minReads : couverture minimale pour caller un variant
--minMapQual : qualité d'alignement minimale pour un read
--minBaseQual : qualité de base minimale pour un variant
--minVarFreq : fréquence minimale pour un variant

- Lancer LoFreq

```
$ ./lofreq call -f hg19.fasta -l design.bed -a 1 -q 10 -Q 10 -m 20 -o sample_lofreq.vcf --min-cov 25 --call-indels --no-default-filter sample.bam
```
-a : P-Value cutoff
-q : qualité de base minimale générale
-Q : qualité de base minimale pour un variant
-m : qualité d'alignement minimale pour un read
--min-cov : couverture minimale pour caller un variant
--call-indels : inclure la recherche d'INDELs
--no-default-filter : n'appliquer aucun filtres

Rq : Les fichiers en sortie seront tous au format VCF mais pas forcément en même version de VCF et certains outils normalisent à gauche la notation des variants ou non. Certains acceptent, des variants multiples pour une même position et une même ligne et pas d'autres. Une étape de normalisation/décomposition est donc souvent nécessaire.

D) L'annotation

1) Généralités

Cette étape permet de déterminer où se produit la mutation au niveau génomique, dans quelles régions (intergénes, introns, exons), dans quels gènes, quel type de mutation (missense, stop, frameshift) et quelle est la nomenclature HGVS du variant.
Cette étape est plus complexe qu'il y parait. En effet, les annotations génomiques ne sont pas figées dans le temps et les bases de données utilisées (RefSeq, Ensembl, UCSC) subissent des mises à jour fréquentes aussi bien pour les gènes que pour les transcrits. Par exemple, certains transcrits peuvent être rallongés au niveau de leur UTR de plus de 2000bp entre 2 versions, ou être décalés de positions au niveau de leurs exons. Il est donc difficile de garantir aucune erreur d'annotation, surtout si l'on souhaite se passer des solutions payantes du type ALAMUT. Les autres peuvent être rangées en 2 grands groupes, les solutions en lignes (qui implique des problèmes de sécurité et de confidentialité) et les outils en local (qui sont souvent moins à jour).

Les outils en ligne :

- VEP (http://www.ensembl.org/info/docs/tools/vep/script/index.htmlATELIER P2.docx)
C'est l'outil dédié d'Ensembl, il permet d'obtenir également des résultats d'outils de priorisation.
Il peut être intérroger via une API.

- NCBI Variant Reporter
(https://www.ncbi.nlm.nih.gov/variation/tools/reporterATELIER P2.docx)
C'est l'outil du NCBI, son principal intérêt est qu'il est du coup toujours à jour avec la base RefSeq. Il permet d'avoir en plus les annotation de dbSNP et de ClinVar. Il existe aussi une API mais qui peut s'avérer très longue pour un grand nombre de variants.

- wANNOVAR (http://wannovar.wglab.org/ATELIER P2.docx)
C'est la version en ligne du célébre ANNOVAR. Il fait donc l'annotation et la priorisation en même temps.

Les outils en local :

- ANNOVAR (http://annovar.openbioinformatics.org/en/latest/)
C'est l'outil en local le plus connu et le plus utilisé malgré des erreurs d'annotation connues. Il permet d'annoter suivant différentes Dbs et permet également d'obtenir les résultats d'outils de priorisation.

- TRANSVAR (http://transvar.readthedocs.io/en/latest/index.html)
C'est le petit dernier qui corrige beaucoup des erreurs d'ANNOVAR même s'il ne corrige pas les erreurs dues aux hétérogénéités des versions des gènes et transcrits. Il est vraiment dédié à l'annotation et ne gère donc pas la priorisation.

2) Exemples

Après avoir installé les outils et téléchargé les bases de donnée nécessaires, comme indiqué dans leur documentation :

- Convertir le fichier VCF au format d'entrée d'ANNOVAR

```
perl convert2annovar.pl -format vcf4 sample_GATKu.vcf > sample_GATKu.avinput
```

- Lancer l'annotation sur refGene avec ANNOVAR

```
perl table_annovar.pl sample_GATKu.avinput path_db_annovar -buildver hg19 -protocol refGene -operation g -nastring . -outfile sample_GATKu_annovar.out
```

Le fichier de sortie (sample_GATKu_annovar.out) sera un fichier tabulé ouvrable avec excel.

- Lancer l'annotation avec TransVar d'un seul variant

```
transvar ganno -i 'chr22:g.31322870C>CG' --refversion hg19 --refseq
```

- Lancer l'annotation avec TransVar sur un fichier vcf

```
transvar ganno --vcf sample_GATKu.vcf --refversion hg19 --refseq --mem > sample_GATKu_transvar.vcf
```

Le fichier de sortie (sample_GATKu_transvar.vcf) sera un fichier VCF annoté.

E) La priorisation

C'est l'étape de jonction entre la bioinformatique et la bioanalyse. Une fois, les variants déterminés, il faut trier pour ne garder que ceux pouvant expliquer le phénotype clinique observé.

C'est sur cette partie que se fait actuellement le plus gros travail de développement d'algorithmes et de bases de données dédiées. On peut diviser la priorisation en 4 groupes :

- les outils *in silico* prédisant un impact fonctionnel et/ou un score de pathogénicité
- les bases de données populationnelles intégrant des données de NGS et nous permettant donc d'obtenir une fréquence d'occurence de notre variant
- les bases de données cliniques regroupant des informations sur les variants impliqués dans une ou plusieurs pathologies
- les meta-outils, combinant plusieurs approches souvent dans une interface de bioanalyse dédiée.

1) Les outils in silico de priorisation

Ils vont attribuer un score de pathogénicité pour chacun des variants. Pour ce faire, ils vont utiliser une ou plusieurs approches dont :

- la conservation interspécifique
- la modification de structure 2D/3D
- la présence dans un domaine fonctionnel
- la tolérance du gène aux mutations (RVIS pour Residual Variation Intolerance Score)

Certains sont très orientés sur la prédiction fonctionnel, il en existe une petite centaine sur le site OMICtools (https://omictools.com/functional-predictions-category), les plus utilisés étant : **Polyphen2**, **SIFT**, **MutationTaster**, **PhyloP**, **dbNSFP**...
D'autres sont des vrais outils de priorisation (https://omictools.com/variant-prioritization-category) comme : **UMD-Predictor, VaRank, VAAST, SPRING** ...

2) Les bases de données NGS
Ces bases vont regouper de nombreuses données d'exomes et/ou génomes et vont ainsi permettre de connaitre la fréquence d'un variant dans un grand nombre de séquencage NGS. Il faudra cependant faire attention aux types de données intégrées car elles peuvent bien entendu contenir des projets NGS liée à une pathologie (par exemple EXAC contient énormément de schyzophrène) et donc biaiser la fréquence observée. Les plus connus sont :
- **1000Genomes** (http://www.internationalgenome.org/ATELIER P2.docx)
- **ExAC** avec 60706 WES (http://exac.broadinstitute.org/)
- **ESP** (NHLBI Exome Sequencing Project) avec 6503 WES (http://evs.gs.washington.edu/EVS/)
- **CG69** avec 69 WGS (http://www.completegenomics.com/ATELIER P2.docx)
- **HRC** (Haplotype Reference Consortium) avec 64976 haplotypes et 39M de SNPs (http://www.haplotype-reference-consortium.org/)
- **dbSNP** dont la version 149 contient actuellement 154M de SNPs
- **Kaviar** avec 162M de SNPs, 13200 WGS et 64600 WES (http://db.systemsbiology.net/kaviar/)
- **gnomAD** (genome Aggregation Database) encore en beta mais qui intégre déjà 123136 WES et 15496 WGS (http://gnomad.broadinstitute.org/ATELIER P2.docx)
- **MITOMAP** qui contient actuellement 30589 mtDNAs humains (http://mitomap.org/MITOMAP)

3) Les bases de données cliniques
(https://omictools.com/variant-disease-association-databases-category)
Elles vont regroupées des informations sur une ou plusieurs pathologies avec une notion de mutation reportée/confirmée. En voici quelques unes :
- **ClinVar** qui est très généraliste (https://www.ncbi.nlm.nih.gov/clinvar/ATELIER P2.docx)
- **OMIM** qui est plus une DB liant gènes et génotypes (https://www.ncbi.nlm.nih.gov/omimATELIER P2.docx)
- **LOVD** qui regroupe diverses DBs spécialisées de pathologies (http://www.lovd.nl/ATELIER P2.docx)
- **DECIPHER** dédié aux maladies rares (https://decipher.sanger.ac.uk/)
- **MITOMAP** qui regroupe les variants mtDNA

4) Les meta-outils
Ils vont combiner des résultats d'outils de priorisation et des informations des bases de données. Ils sont souvent associés à une interface graphique facilitant l'interrogation. Il sont plutôt à utiliser pour confirmer ou infirmer des variants candidats préalablement filtrés. Voici quelques outils d'aide à l'interprétation :
- **Alamut Visual** (http://www.interactive-biosoftware.com/alamut-visual/ATELIER P2.docx)
- **VarSome** (https://omictools.com/variant-prioritization-category)

F) La spécificité du diagnostic NGS sur le génome mitochondrial

La spécificité du génome mitochondrial va se faire sur les différentes étapes de l'analyse bioinformatique en diagnostic NGS.

- Au niveau de l'alignement, si on est sur un séquençage d'ADN total, on peut être géné par les pseudogènes. La circularité du gènome va également perturber l'alignement en coupant les reads qui couvrent l'origine (ce qui n'est pas forcément limitant pour une analyse en diagnotic). L'alignement va aussi être biaisé par la présence d'homopolymères et de zones répétées.

- Au niveau du calling, il faut être souple sur les paramètres limitant la fréquence des variants. En effet, si l'on recherche des mutations hétéroplasmiques il faudra baisser ce seuil. Bien entendu, combiner avec les problèmes d'alignement, cela va générer un grand nombre de variants artéfactuels qu'il faudra donc filtrer au mieux par la suite.

- Au niveau de la priorisation, les outils classiques vont être très mauvais car pas du tout adapté aux spécificités du génome mitochondrial. De même, les bases de données classiques ne contiennent souvent pas ou peu de données sur les variants mtDNA.

La banque de données **MITOMAP** et son outil d'interrogation **MITOMASTER** restent donc pour le moment incontournables pour un rendu de résultats sur des variants mtDNA connus. Pour les nouveaux variants ou ceux pas encore confirmés, quelques outils peuvent quand même aider au diagnostic :
- **MitImpact2** qui intégre les résultats précalculés d' outils de priorisation (http://mitimpact.css-mendel.it/). Attention, il n'intègre que les gènes codant pour une protéine.
- **Mamit-tRNA** pour les variants ARNt (http://mamit-trna.u-strasbg.fr/human.asp)
- **eKLIPse** qui est un outil développé au CHU d'Angers et qui permet de détecter des délétions sporadiques et multiples sur des données NGS mtDNA (actuellement en cours de finalisation).

Des pipelines dédiés à l'analyse NGS mtDNA existe aussi :
- **MToolBox** qui est actuellement le plus abouti (https://github.com/mitoNGS/MToolBox)
- **MtDNA-Server** (https://mtdna-server.uibk.ac.at/)
- **MitoSeek** (https://github.com/riverlee/MitoSeek)

Atelier F

TP analyse Bioinfo de l'ADNmt.

A - Matériel :

Séquences ADNmt au format fasta : Séquence 1, Séquence 2, séquence 3, séquence 4

B - Analyse d'ADNmt :

1 – Analyse des séquences 1, 2, 3 et 4 :

a - Détermination des haplogroupes mitochondriaux à des séquences :

Utilisation d'HaploGrep 2 :

Site : https://haplogrep.uibk.ac.at/
Manuel : http://haplogrep.uibk.ac.at/help.html
Article de Référence : https://doi.org/10.1093/nar/gkw233

Quels sont les haplogroupes des séquences 1, 2, 3 et 4 ?

b - Trouver les haplogroupes dans l'arbre mondial et Trouver l'Histoire de ces haplogroupes et Trouver la Fréquence mondial de ces haplogroupes :

Site : http://www.phylotree.org/index.htm; wikipedia, pubmed, blog (http://blogs.discovermagazine.com/gnxp/2011/06/what-if-youre-wrong-haplogroup-j/)

Quelles sont les Histoires des haplogroupes des séquences 1, 2, 3 et 4 ?

c - Trouver la présence des polymorphismes de l'haplogroupe dans la population mondial et sa localisation géographique :

Localisation géographique du polymorphisme : http://www.mtdb.igp.uu.se/

MitoWheel : Trouver sur quel gène se trouve le polymorphisme :
http://www.mitowheel.org/mitowheel.html

Existe-t-il un polymorphisme d'intérêt particulier et dans quelle séquence ?

2 – Analyse des polymorphismes pathogènes :

Avez-vous détecté un polymorphisme potentiellement pathogène ?

Trouver sur quel gène se trouve le polymorphisme ?

Trouver l'association du polymorphisme avec les pathologies
Utilisation de MitoMap : http://www.mitomap.org/MITOMAP

Verifier haplogroupe avec Mitomaster :
http://www.mitomap.org/foswiki/bin/view/MITOMASTER/WebHome

3 – Analyse des relations haplogroupes/pathologies :

Trouver l'association de l'haplogroupe J avec les pathologies mitochondriale en utilisant Pubmed

C – Analyse des Gènes Nucléaires :

1 - Rechercher le schéma des oxydations phosphorylantes dans KEGG (Base de données Biochimique) :

Site : http://www.kegg.jp/kegg/

2 - Rechercher sur le schéma des oxydations phosphorylantes les gènes nucléaires de la cytochrome c oxydase et notamment le sous unité 4 et cliquer sur la case

3 - Trouver l'isoforme 2 humain de la sous-unité 4 (Gène : COX4I2).

4 - Trouver l'indentification Ensembl et NCBI du gène COX4I2 sur Kegg et cliquer directement sur le lien.

Pour info :
NCBI : https://www.ncbi.nlm.nih.gov/

Ensembl : http://www.ensembl.org/index.html

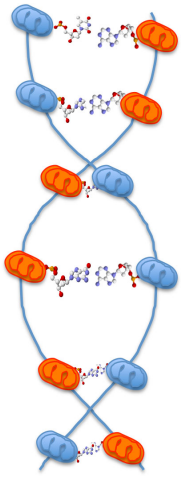
Sur Ensembl :

- Noter la localisation de COX4i2
- Regarder la table des variant pour connaitre les polymorphismes présents sur ce gène.
- Tirer les variants par fréquence (Global MAF) et au niveau des exons (consequences :All)

Quelles sont les informations disponibles pour le variant rs540773473 ?

- Sélectionner le variant de plus fréquent et noter son variant Id (Rs…..) et cliquer
- Regarder sa répartition mondiale.

Quelle est la fréquence de ce variant en Finlande ?



15h-16h00 **Table ronde : Validation fonctionnelle**

Pr. Pascal Reynier, Pr. Vincent Procaccio

Les variants pathogènes et non pathogènes de l'ADNmt seront sans doute précisément répertoriés dans quelques années avec probablement une catégorie de variants exprimant leur pathogénicité dans des contextes particuliers (fond génétique mitochondrial, di-génisme, médicament,…). La période actuelle voit l'ADNmt systématiquement séquencé chez un grand nombre de patients suspects de maladies mitochondriales et dans de nombreuses populations saines. Il en découle une avalanche de variants à pathogénicité incertaine sur lesquels on doit statuer. Une analyse globale des ressources disponibles est nécessaire pour établir la pathogénicité de ces variants et leur implication dans la maladie incriminée : Clinique, étude familiale, bibliographie (de nombreux variants initialement décrits comme pathogènes s'avèrent être des polymorphismes), bases de données, taux de mutant, biochimie, tests fonctionnels (cybrides), prédiction de structure pour les ARNt,… Notre table ronde consistera à discuter de ces approches, de leur intérêt et de leurs limites.

16h00 Fin