



MINISTÈRE DE L'INTÉRIEUR

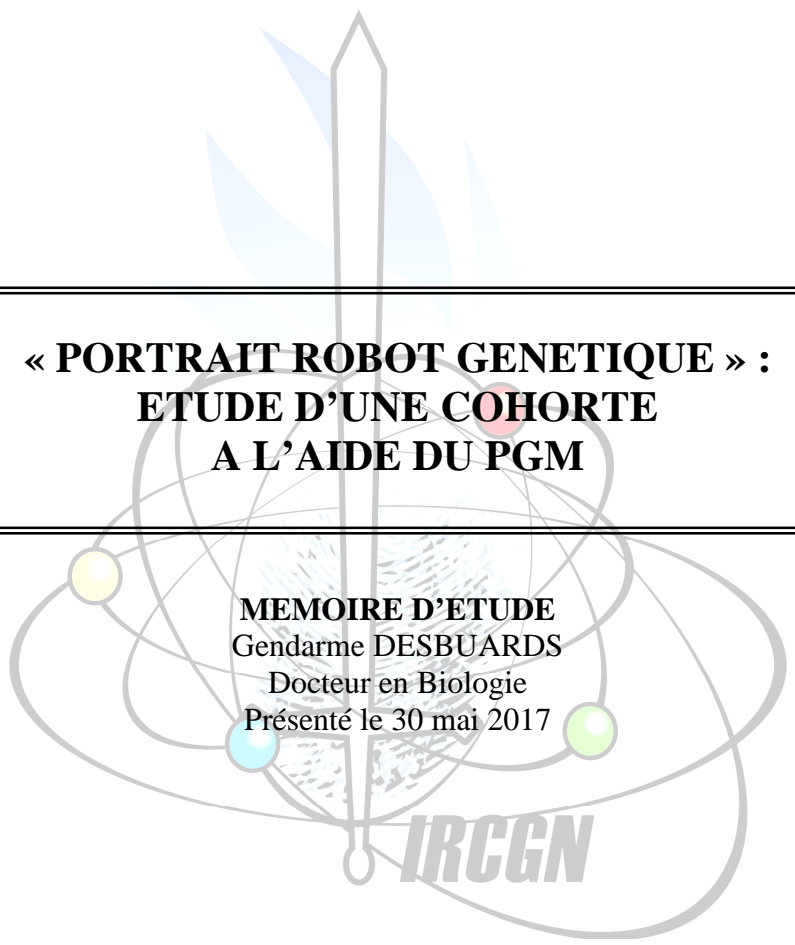


POLE JUDICIAIRE DE LA
GENDARMERIE NATIONALE

INSTITUT DE RECHERCHE CRIMINELLE
DE LA GENDARMERIE NATIONALE

DIVISION CRIMINALISTIQUE BIOLOGIE GENETIQUE

DEPARTEMENT BIOLOGIE



« PORTRAIT ROBOT GENETIQUE » :
ETUDE D'UNE COHORTE
A L'AIDE DU PGM

MEMOIRE D'ETUDE
Gendarme DESBUARDS
Docteur en Biologie
Présenté le 30 mai 2017

IRCGN

Travail réalisé sous la tutelle du capitaine PUSSIAU, expert de la DCBG

REMERCIEMENTS

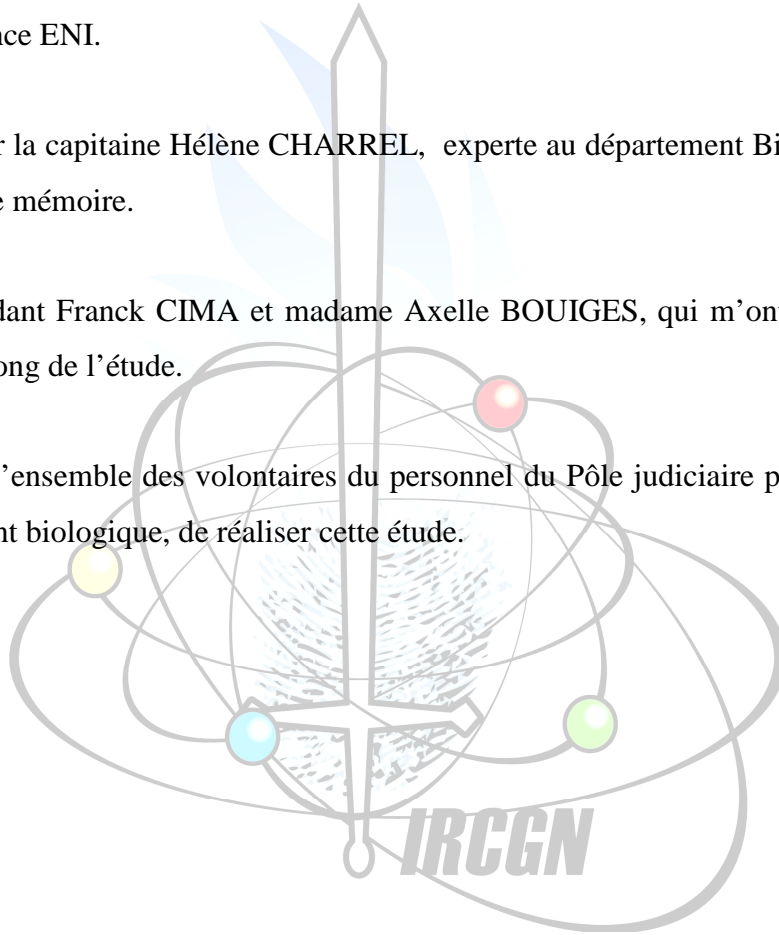
Je tiens à remercier le Capitaine Amaury PUSSIAU, expert judiciaire dans le domaine de l'identification humaine par empreintes génétiques de la « division criminalistique biologie génétique (DCBG) en sa qualité de tuteur pour cette étude.

J'adresse également mes remerciements au chef d'escadron PHAM-HOAI, chef du département biologie, pour m'avoir confié la responsabilité technique de cette étude et me permettre ainsi de passer ma soutenance ENI.

Je tiens à remercier la capitaine Hélène CHARREL, experte au département Biologie, en sa qualité de correctrice de ce mémoire.

Je remercie l'adjudant Franck CIMA et madame Axelle BOUIGES, qui m'ont apporté un soutien technique tout au long de l'étude.

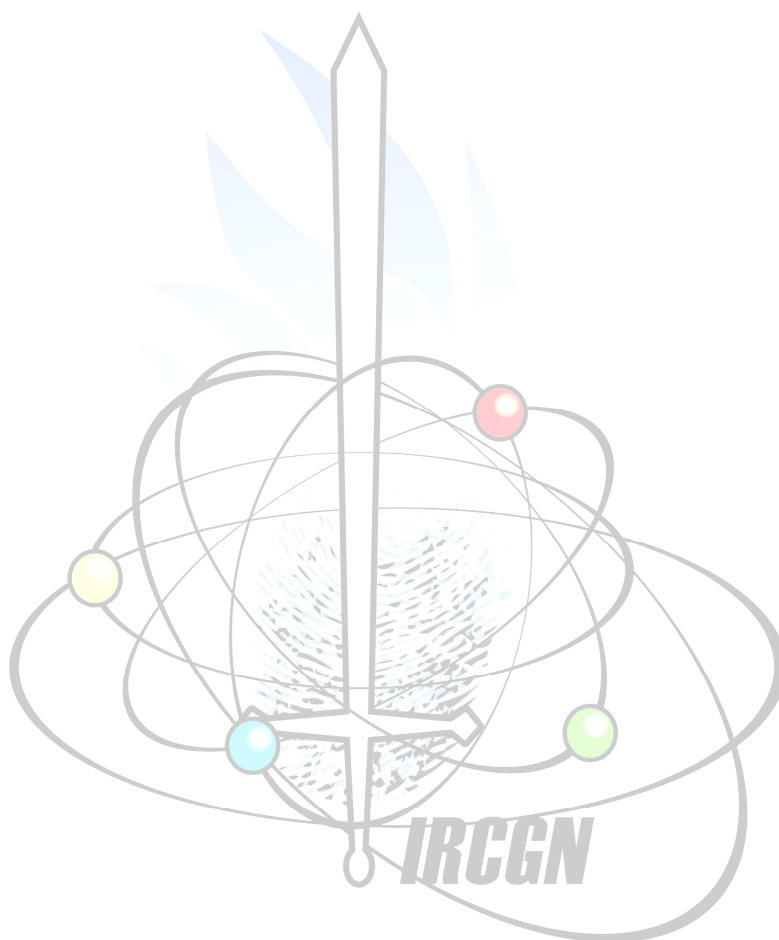
Enfin je remercie l'ensemble des volontaires du personnel du Pôle judiciaire pour m'avoir permis, via leur prélèvement biologique, de réaliser cette étude.



SOMMAIRE

SOMMAIRE	3
LISTE DES ABREVIATIONS	5
LISTE DES ILLUSTRATIONS	6
INTRODUCTION GENERALE	7
ETAT DE L'ART OU GENERALITES	9
I) L'ADN	9
1) <u>Généralités</u>	9
2) <u>Génotype et phénotype</u>	9
3) <u>Polymorphisme</u>	9
a) <i>Polymorphismes de longueur</i>	10
b) <i>Polymorphismes de séquence</i>	10
II) Le portrait robot génétique	11
1) <u>Intérêt du portrait robot génétique</u>	11
2) <u>Etude des SNP pour la détermination des traits phénotypiques</u>	11
a) <i>La couleur des yeux et des cheveux</i>	12
b) <i>Les origines biogéographiques</i>	13
c) <i>Autres caractéristiques morphologiques</i>	13
III) PGM	13
1) <u>Intérêt du PGM</u>	13
2) <u>Principe du PGM</u>	14
a) <i>Préparation de la librairie</i>	15
b) <i>Préparation de la matrice de séquençage</i>	16
c) <i>Etape de séquençage</i>	18
d) <i>Etape d'interprétation</i>	19
MATERIEL ET METHODES	21
1) <u>Cohorte de volontaires</u>	21
2) <u>Extraction quantification et dilution de l'ADN prélevé</u>	21
3) <u>Panel de SNP utilisés</u>	21
4) <u>Données brutes</u>	21
5) <u>Prédiction de l'origine biogéographique</u>	22
6) <u>Prédiction de la couleur des yeux et des cheveux</u>	22
7) <u>Etude de la reproductibilité</u>	22
8) <u>Etude de la sensibilité</u>	23
RESULTATS	24
I) Prédiction des origines biogéographiques	24
II) Prédiction des caractères phénotypiques	25
1) <u>La couleur des yeux</u>	25
2) <u>La couleur des cheveux</u>	31
III) Etude de la reproductibilité	34

IV) Etude de la sensibilité	34
DISCUSSION ET PERSPECTIVES	38
BIBLIOGRAPHIE	41
ANNEXES	42



LISTE DES ABREVIATIONS

ADN : Acide désoxyribonucléique

AIM : *Ancestry Informative Marker*

Amplicon : Fragment d'ADN amplifié par PCR

dNTP : Mélange des quatre désoxyribonucléotides

FNAEG : Fichier National Automatisé des Empreintes Génétiques

ISP : *Ion Sphere Particule.*

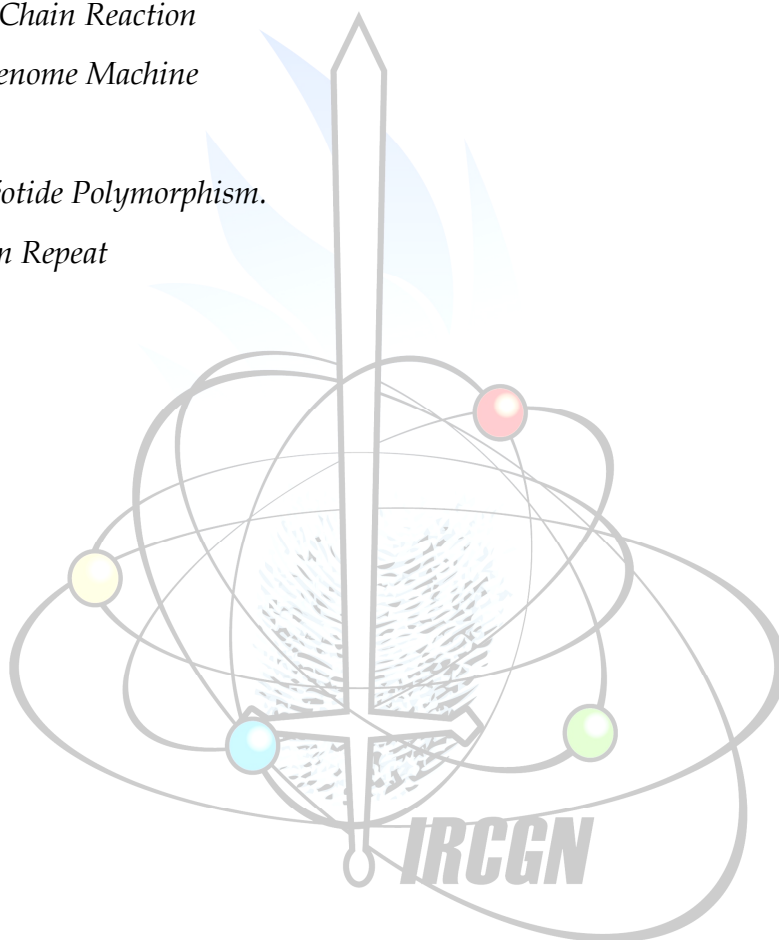
PCR : *Polymerase Chain Reaction*

PGM : *Personal Genome Machine*

Pb : *paire de base*

SNP: *Single Nucléotide Polymorphism.*

STR: *Short Tandem Repeat*



LISTE DES ILLUSTRATIONS

Figure n° 1 : PGM utilisé au sein du laboratoire

Figure n°2 : Schéma du principe général du séquençage par le PGM

Figure n° 3 : Schéma de la préparation de la librairie

Figure n° 4 : Schéma de l'amplification des banques

Figure n° 5 : Schéma de la phase d'enrichissement

Figure n° 6 : Une puce de séquençage

Figure n° 7 : Schéma de fonctionnement d'un puits de la puce

Figure n° 8 : Conversion par le serveur des données numériques en séquence nucléotidique

Tableau n°1 : Prédiction des origines biogéographiques de 20 échantillons

Tableau n°2 : Concordance entre la prédiction de la couleur des yeux et la couleur des yeux donnée

Tableau n°3 : Résultat des prédictions de la couleur des cheveux

Tableau n°4 : Concordance entre les prédictions et la couleur des cheveux

Tableau n°5 : Résultats de la prédiction de la couleur des yeux et des cheveux pour différentes quantités d'ADN

Graphique n°1 : Nombre de SNP séquencés en fonction de la quantité d'ADN pour l'individu n°1

Graphique n°2 : Nombre de SNP séquencés en fonction de la quantité d'ADN pour l'individu n°2

Annexe n°1 : Consentement éclairé

Annexe n°2 : Questionnaire phénotypique

Annexe n°3 : Cohorte

Annexe n°4 : Liste des SNPs du panel HIrisPlex

Annexe n°5 : Rapport de séquençage

INTRODUCTION GENERALE

« Identifier l'auteur de faits punissables par la loi est le but de toute enquête judiciaire, mais les moyens d'y parvenir ont considérablement évolué au cours du temps. Chaque personne se distingue de toute autre, par un ensemble de caractères physiques et morphologiques qui lui sont propres. Ainsi, l'identification consiste à rassembler les caractères spécifiques à chaque individu dans le but de déterminer d'une manière rigoureuse et objective son identité, et ce tout au long de son existence » (Buquet, Manuel de criminalistique moderne et de police scientifique 2006).

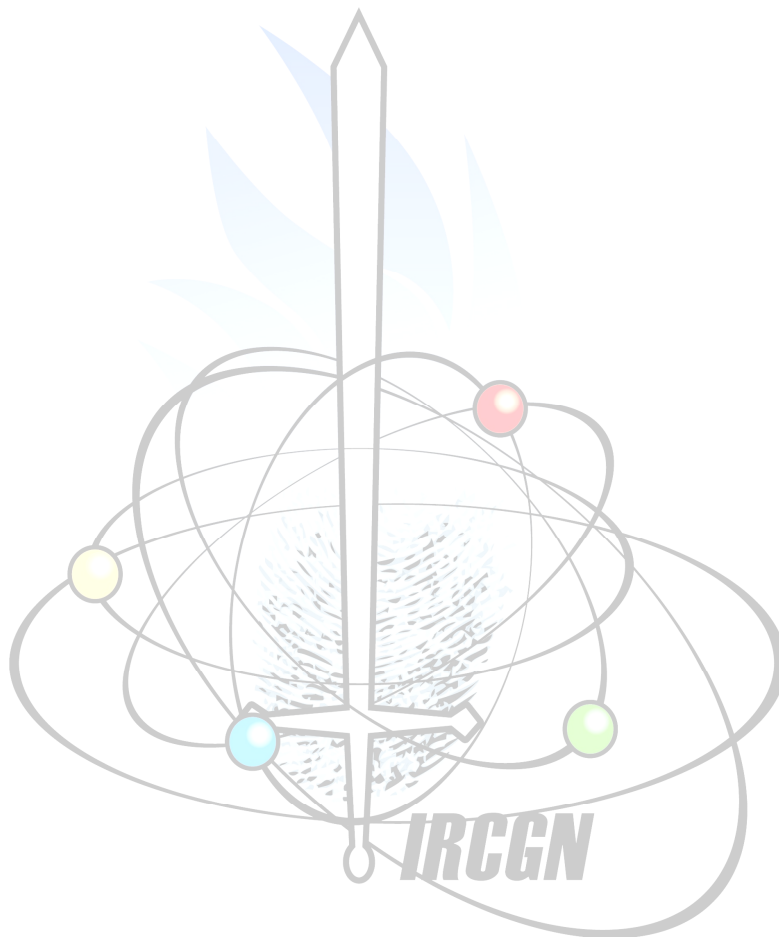
L'arrivée des analyses ADN en criminalistique a révolutionné l'identification des criminels. Basée sur des données scientifiques, l'ADN permet en effet d'identifier l'auteur d'un crime ou d'un délit à partir de petits échantillons recueillis sur la scène de crime. Son efficacité croissante en a fait rapidement un outil phare au sein de la police technique et scientifique.

Les profils non identifiés établis à partir des traces prélevées sur le terrain et les profils ADN de référence prélevés sur les auteurs ou les mis en cause dans le cas de crimes ou délits sont enregistrés dans la base de données génétique nationale : le FNAEG (Fichier National des Empreintes Génétiques). Créé en 1998, cette base de données contenant en 2016 plus de 3,42 millions de profils génétiques, effectue des rapprochements entre les profils génétiques des traces non identifiées et les profils génétiques des individus de référence afin d'identifier l'individu à l'origine de la trace.

Parfois les enquêteurs se trouvent dans l'impossibilité de tirer profit des échantillons d'ADN recueillis sur une scène de crime. En effet, la quantité d'ADN présente dans l'échantillon peut être insuffisante pour permettre d'obtenir un résultat interprétable, ou encore le profil génétique déterminé par l'analyse de l'ADN recueilli peut être inconnu de la base de données. Dans ces cas particuliers, l'information apportée par l'analyse ADN est inefficace pour identifier un individu.

A défaut de pouvoir identifier l'individu à l'origine de la trace, l'information génétique pourrait être exploitée pour fournir aux enquêteurs des orientations morphologiques et biogéographiques. Le but est ainsi de pouvoir exploiter au mieux l'ADN recueilli sur une scène de crime ou encore pour l'identification d'un cadavre. C'est dans cette optique que depuis quelques années, les laboratoires scientifiques et la police scientifique se sont intéressés à la possibilité d'établir des portraits-robots à partir de traces biologiques.

Le but de cette étude est de tester la possibilité de déterminer des caractères morphologiques apparents à partir de l'analyse d'ADN recueilli sur une scène d'infraction, et d'en connaître les limites et les perspectives.



ETAT DE L'ART OU GENERALITES

I) L'ADN

1) Généralités

L'ADN est le support de l'information génétique de tout être vivant. Les caractères morphologiques et physiologiques exprimés sont codés au niveau de séquences bien spécifiques que sont les gènes. Les gènes sont constitués de régions codantes ou exons et de régions non codantes ou introns. Pour autant, certains introns ont un rôle dans la régulation de l'expression des gènes.

Un gène peut exister en plusieurs versions dénommées *allèles*. Ainsi pour chaque caractère génétique, un individu va posséder deux allèles, un transmis par le père biologique et l'autre transmis par la mère biologique. Si ces deux allèles sont identiques, l'individu est dit homozygote ; à l'inverse si les deux allèles sont différents, l'individu est dit hétérozygote.

2) Génotype et phénotype

Le génotype est l'ensemble de l'information génétique d'un individu. C'est la composition allélique de tous les gènes.

Le phénotype est l'ensemble des traits et caractéristiques morphologiques et physiologiques observables chez un individu, résultat de l'expression et de la régulation du génotype. Ce phénotype est également influencé par d'autres gènes, ou par les conditions environnementales dans lequel évolue l'individu. Le phénotype pour un caractère donné peut résulter de l'expression de plusieurs gènes. Inversement, un même gène peut influencer l'expression de plusieurs traits ou caractéristiques à la fois.

3) Polymorphisme

Tous les individus d'une même espèce possèdent le même patrimoine génétique et sont par conséquent génétiquement très proches. Néanmoins, il existe des variations interindividuelles qui

sont à l'origine de l'unicité de chaque individu. Ces variations mineures constituent le polymorphisme.

Ce polymorphisme trouve son origine dans l'ADN et est de deux types : le polymorphisme dit de séquence et le polymorphisme dit de longueur.

a) *Polymorphismes de longueur*

Les polymorphismes de longueur sont des motifs nucléotidiques répétés, dont le nombre de répétitions est variable selon les individus. C'est le nombre de répétitions du motif qui constitue le polymorphisme. Ces polymorphismes sont localisés généralement dans les régions non codantes de l'ADN. Parmi eux, les microsatellites ou STR (*Short tandem repeat*) sont utilisés en criminalistique.

Les STR sont constitués de nucléotides de 2 à 6 pb répétés de 5 à 100 fois les uns à la suite des autres. Le nombre de répétitions des STR est très variable d'un individu à l'autre en raison du brassage génétique. Leur pouvoir discriminant élevé fait de ces polymorphismes des marqueurs de choix pour l'établissement de profils génétiques en criminalistique.

Bien que l'analyse des STR soit facile à mettre en œuvre, elle peut souvent se révéler infructueuse à partir d'extraits d'ADN dégradés. La longueur des séquences à analyser (environ 400 pb) limite fortement la taille des fragments d'ADN amplifiables et donc leurs détections.

b) *Polymorphismes de séquence*

Le polymorphisme de séquence correspond aux variations observées dans l'enchaînement des nucléotides pour une région d'ADN d'intérêt.

Les SNP (*Single Nucleotide Polymorphism*) sont la variation d'une seule paire de bases du génome entre individus d'une même espèce. Ils constituent les polymorphismes les plus répandues du génome humain et présentent une partition uniforme (en moyenne tous les 1000 pb).

Les SNP des chromosomes non sexuels ou autosomaux ont été les polymorphismes ponctuels les moins étudiés pendant de nombreuses années. En effet, la découverte de ces polymorphismes est très récente puisqu'elle correspond au séquençage complet du génome humain. De ce fait, tous les SNP autosomaux ne sont pas encore caractérisés. Aujourd'hui de nombreuses études visent à identifier les SNP impliqués dans la prédisposition à certaines maladies, dans la structure et la

diversité des populations humaines, définir des ensembles de SNP permettant d'identifier les individus et d'estimer leur origine biogéographique.

II) Le portrait robot génétique

1) Intérêt du portrait robot génétique

La prédiction des caractères phénotypiques d'un individu, à partir de l'ADN a été beaucoup étudiée depuis 2009, plus particulièrement en criminalistique. En effet, la capacité à prédire l'apparence physique d'un individu à partir d'un ADN retrouvé sur une scène de crime peut être d'une grande aide pour les enquêteurs, en limitant par exemple le nombre de suspects potentiels, ou donner une orientation d'enquête lorsque le profil génétique n'est pas connu dans une base de données, ou bien encore lorsque l'ADN est trop dégradé pour pouvoir établir un profil génétique conventionnel (à partir des STR) exploitable. Cela peut être d'autant plus intéressant dans le cadre de découverte de cadavre ou dans le cadre de victimes de désastre, lorsque les corps sont décomposés, carbonisés et qu'il n'est plus possible d'avoir une description précise des traits morphologiques.

2) Etude des SNP pour la détermination des traits phénotypiques

Afin d'établir un portrait robot génétique, il faut pouvoir prédire des caractéristiques phénotypiques d'un individu, caractéristiques issues de l'expression et de la régulation des gènes.

Les SNP sont des marqueurs de l'ADN, pour différents traits morphologiques d'un individu, tels que la couleur des yeux, la couleur et la forme des cheveux, la pigmentation de la peau, la calvitie, la taille, la présence de taches de rousseur, l'origine biogéographique, etc... Les marqueurs SNP sont d'autant plus pertinents, qu'ils peuvent être recherchés sur des fragments d'ADN très courts (quelques pb), ce qui n'est pas le cas pour l'étude nécessaire aux marqueurs STR. Il est donc possible d'obtenir une information à partir d'un ADN présent en faible quantité ou dégradé.

Les marqueurs SNP qui sont les plus étudiés ces dernières années, sont ceux de la couleur des yeux, de la couleur des cheveux et de l'origine biogéographique. C'est pourquoi dans le cadre de notre étude, nous nous sommes particulièrement intéressés à ces derniers.

a) *La couleur des yeux et des cheveux*

L'iris est le diaphragme de l'œil. Il donne sa couleur aux yeux : marrons, verts ou bleus. La couleur d'une lumière perçue par l'œil dépend de la ou les longueurs d'onde qui la composent. La couleur de l'œil dépend ainsi de la lumière qu'il émet ou réfléchit. Plus l'iris est riche en cellules mélanocytaires ou cellules pigmentées plus la lumière incidente est absorbée, et donc les yeux apparaissent de couleur marron. En revanche si l'iris est pauvre ou dépourvue en cellules pigmentées, la lumière incidente est réfléchiée et les yeux apparaissent de couleur bleu. Enfin, les yeux de couleur vert sont munis d'iris moyennement pigmentés qui absorbent une partie de la lumière incidente.

La couleur de l'iris est un polymorphisme phénotypique complexe et important dans les populations européennes, contrôlé très fortement génétiquement. En effet, de récentes et nombreuses études ont montré que le polymorphisme SNP étudié principalement au sein de certains gènes est associé à la variation de la couleur de l'iris dans la population européenne.

Ces mêmes cellules mélanocytaires ou pigmentées sont également responsables de la couleur des cheveux. La pigmentation mélanique du cheveu est déterminée par le patrimoine génétique de l'individu. Si les mécanismes aboutissant à la couleur des yeux sont de mieux en mieux identifiés, la prédiction de la couleur des cheveux est plus complexe et difficile. Néanmoins, des études récentes ont montré que des marqueurs d'ADN provenant de gènes différents étaient impliqués dans l'expression de la couleur des cheveux.

A partir de toutes ces données, une équipe néerlandaise a mis au point un modèle mathématique pour prédire la couleur des yeux et des cheveux. Il s'agit de l'« *Hiris-Plex* », comprenant 6 marqueurs d'ADN de type SNP pour prédire la couleur des yeux et 22 marqueurs d'ADN de type SNP (dont 4 marqueurs communs avec ceux des yeux) pour prédire la couleur des cheveux. Ce modèle permet d'obtenir une prédiction théorique de :

- 95% pour les yeux de couleur marron
- 94% pour les yeux de couleur bleu
- 74% pour les yeux de couleur intermédiaire, à savoir ni bleu ni marron
- 93% pour les cheveux de couleur roux
- 87% pour les cheveux de couleur noir
- 82% pour les cheveux de couleur brun
- 81% pour les cheveux blonds

b) Les origines biogéographiques

L'origine biogéographique peut être prédite en recherchant des différences de fréquence allélique entre les populations. C'est un système multiplex forensique constitué d'une combinaison de marqueurs AIM et de marqueurs SNP, aboutissant aux marqueurs AIM-SNP. Les marqueurs AIM (*Ancestry informative marker*) sont les marqueurs informatifs de l'ascendance, utilisés pour estimer la généalogie génomique individuelle. Afin d'améliorer la prédiction de l'origine biogéographique, ces marqueurs AIM ont été couplés à d'autres marqueurs SNP tels que les marqueurs correspondant à la pigmentation de la peau.

c) Autres caractéristiques morphologiques

Des études récentes ont été réalisées sur des marqueurs SNP associés à la calvitie, à la présence de tâche de rousseur, à la forme des cheveux, à la taille, aux traits physiques du visage. Bien que les premiers résultats soient prometteurs, le degré de prédiction n'est pas suffisant, à ce jour, pour pouvoir intégrer ces caractéristiques phénotypiques dans un portrait robot-génétique.

Notre étude porte donc sur la prédiction de la couleur des yeux, de la couleur des cheveux et des origines biogéographiques.

III) PGM

Afin de séquencer de manière fiable et rapide les différents marqueurs SNP utilisés dans notre étude, aux fins de prédire la couleur des yeux, la couleur des cheveux et l'origine biogéographique, nous avons utilisé le séquenceur « Ion Torrent PGM » (*Ion Torrent Personal Genome Machine*) commercialisé par la société THERMOFISHER.

1) Intérêt du PGM

Le PGM est un séquenceur haut débit de seconde génération. Ce type d'appareil a la capacité de séquencer d'importantes quantités d'ADN de plusieurs individus simultanément et en peu de temps, à l'inverse de la technique traditionnelle de Sanger ou du Snapshot. De plus, la sensibilité de ces appareils est plus importante et il faut peu de quantité d'ADN pour obtenir un profil génétique. Le PGM est donc l'outil idéal pour travailler sur les marqueurs génétiques que sont les SNPs.



Figure n°1 : PGM utilisé au sein du laboratoire

2) Principe du PGM

Le séquençage de SNP sur le PGM nécessite 4 phases opératoires (cf figure n° 2). La première phase (durée environ de 6 heures) consiste à préparer la librairie d'échantillons, la librairie étant la collection de fragments d'ADN, représentative de l'ADN à séquençer. La deuxième phase (durée environ de 6 à 7 heures) consiste à préparer la matrice de séquençage. La librairie va subir une amplification puis un processus d'enrichissement, dans le but d'obtenir de l'ADN d'intérêt monoclonal. La troisième phase est le séquençage réalisé sur l'appareil PGM (durée environ de 4,5 heures). La dernière phase est l'interprétation des données issues du séquençage (durée environ de 30 min par individu).

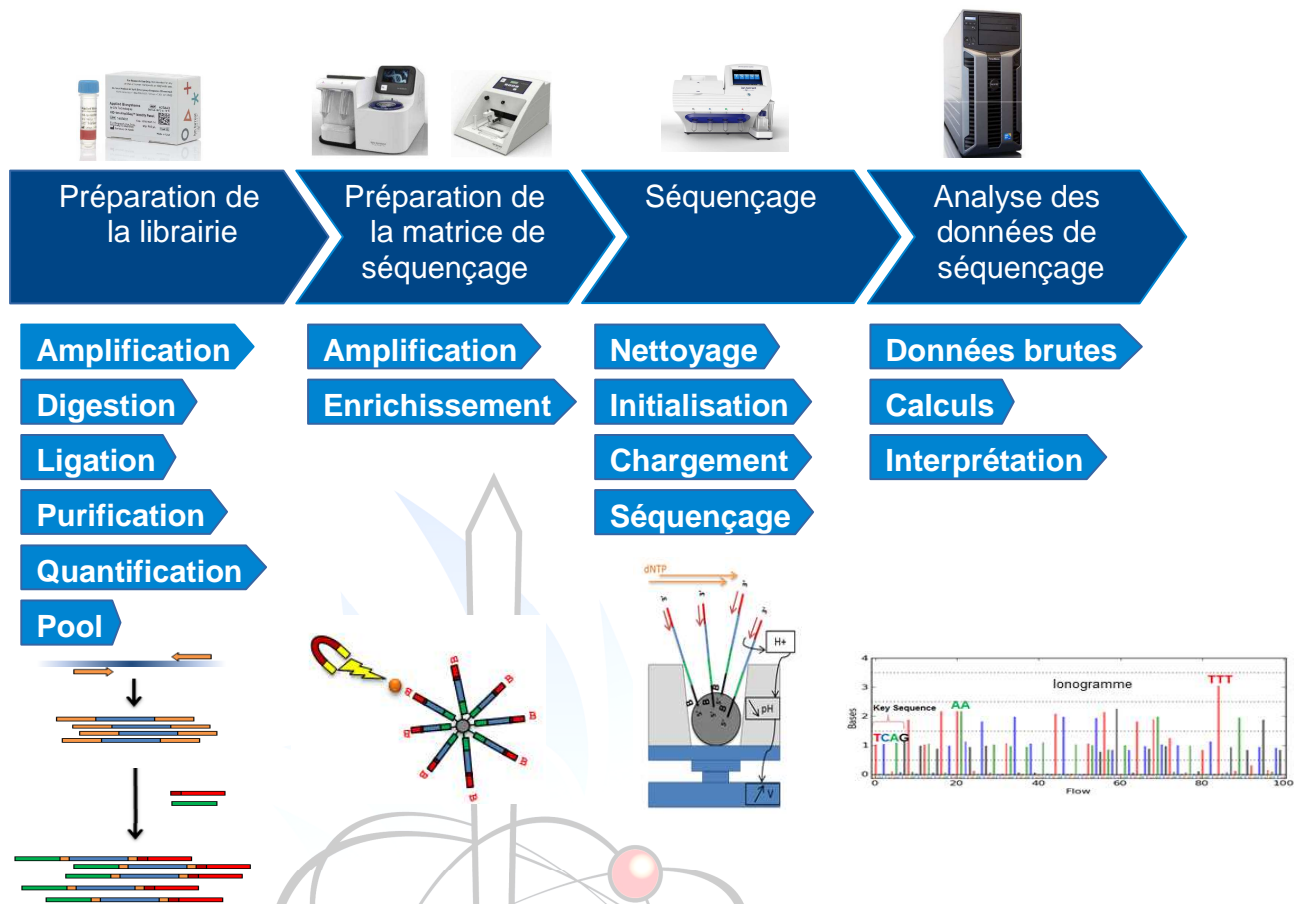


Figure n° 2 : Schéma du principe général du séquençage par le PGM

a) Préparation de la librairie

Il s'agit d'amplifier les régions cibles d'intérêt en utilisant des amorces spécifiques. Les amplicons obtenus sont traités avec le panel enzymatique « Fupa » pour partiellement digérer les amorces et phosphoryler les amplicons. Les amplicons sont alors liés avec des adaptateurs A et P1. A est une séquence nucléotidique incluant un « code-barres », permettant d'identifier les amplicons issus de chaque échantillon lors du séquençage. P1 est une séquence nucléotidique permettant la fixation des amplicons à une particule sphérique ISP (*Ion Sphere Particule*) nécessaire pour réaliser l'amplification monoclonale. Les amplicons sont ensuite purifiés.

Les librairies d'échantillons obtenues sont quantifiés et poolées.

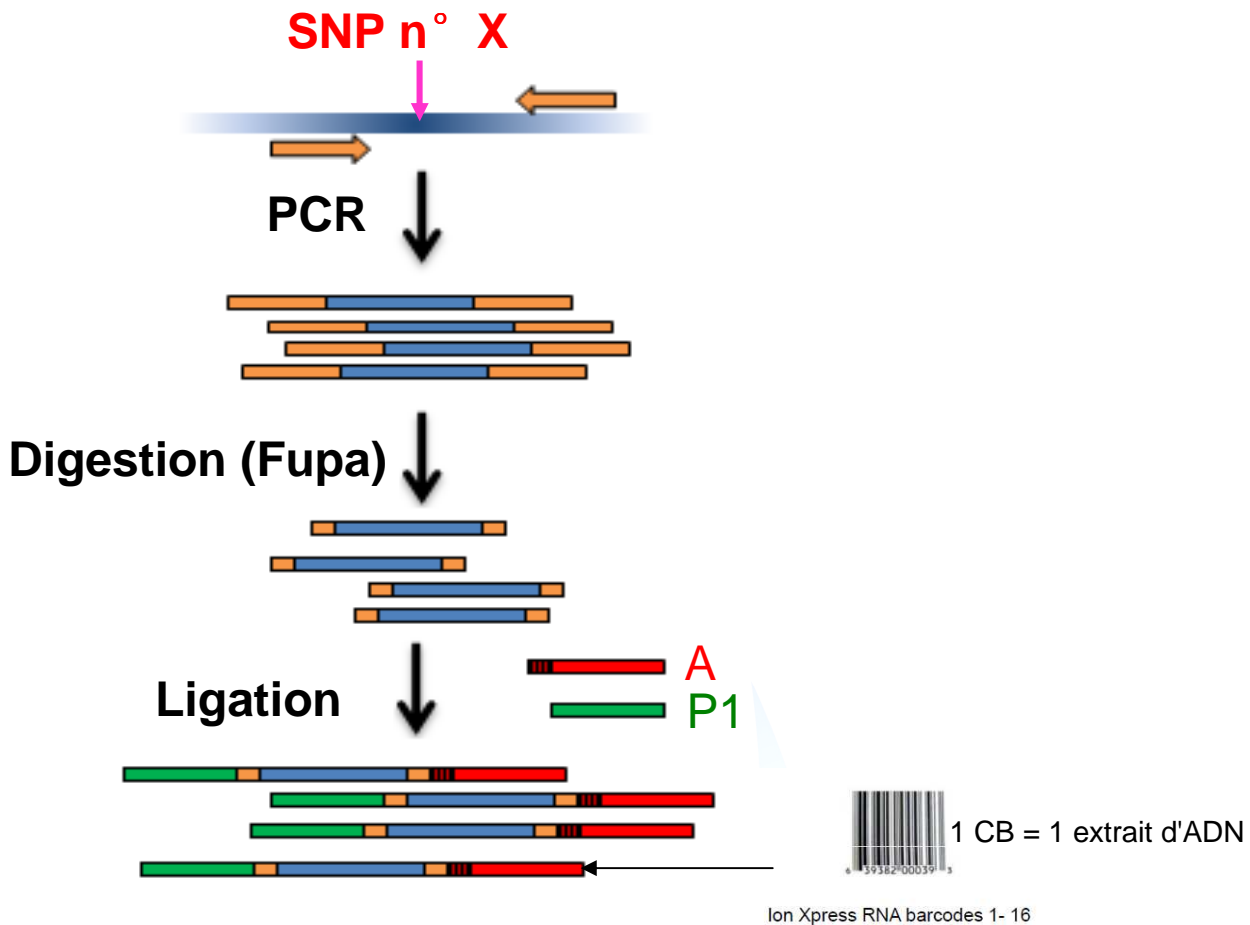


Figure n°3 : Schéma de la préparation de la librairie

b) Préparation de la matrice de séquençage

Les librairies vont subir une amplification clonale par PCR en émulsion. Ce procédé permet de générer des micro-réacteurs à PCR dans l'huile où les fragments d'ADN sont liés à une sphère et amplifiés afin d'obtenir sur chaque sphère une grande quantité monoclonale des régions d'intérêt de l'ADN. A la fin de cette PCR en émulsion, 3 types d'ISPs sont récupérés :

- les ISPs dont la surface est recouverte de brins d'ADN monoclonaux
- les ISPs dont la surface est recouverte de brins d'ADN polyclonaux
- les ISPs ne contenant aucun brin d'ADN

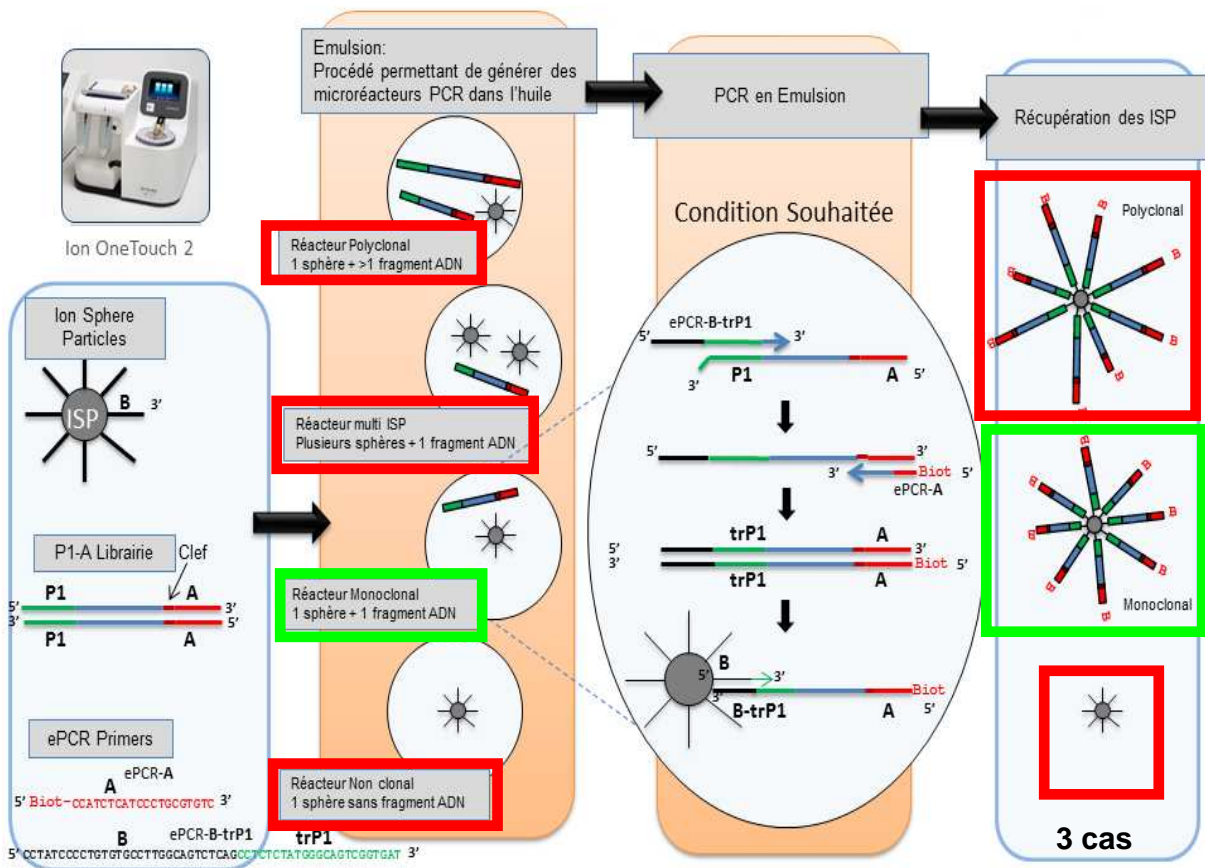


Figure n°4 : Schéma de l'amplification des bibliothèques

Les sphères contenant les fragments d'ADN (ISP positif) sont isolés des sphères ne contenant pas de fragment par ajout de billes magnétiques coâtées streptavidine.

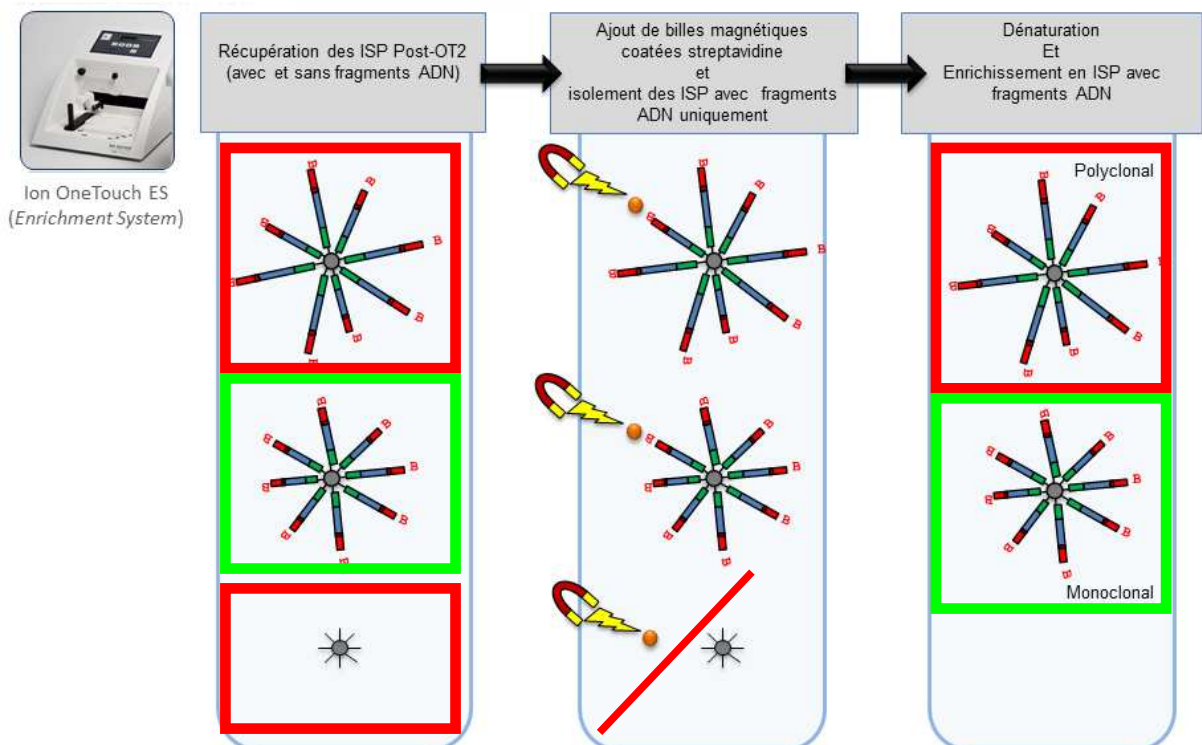


Figure n°5 : Schéma de la phase d'enrichissement

c) Etape de séquençage

Des amorces et de la polymérase sont ajoutées aux ISP positifs enrichis et chargés dans une puce pour être séquencés.

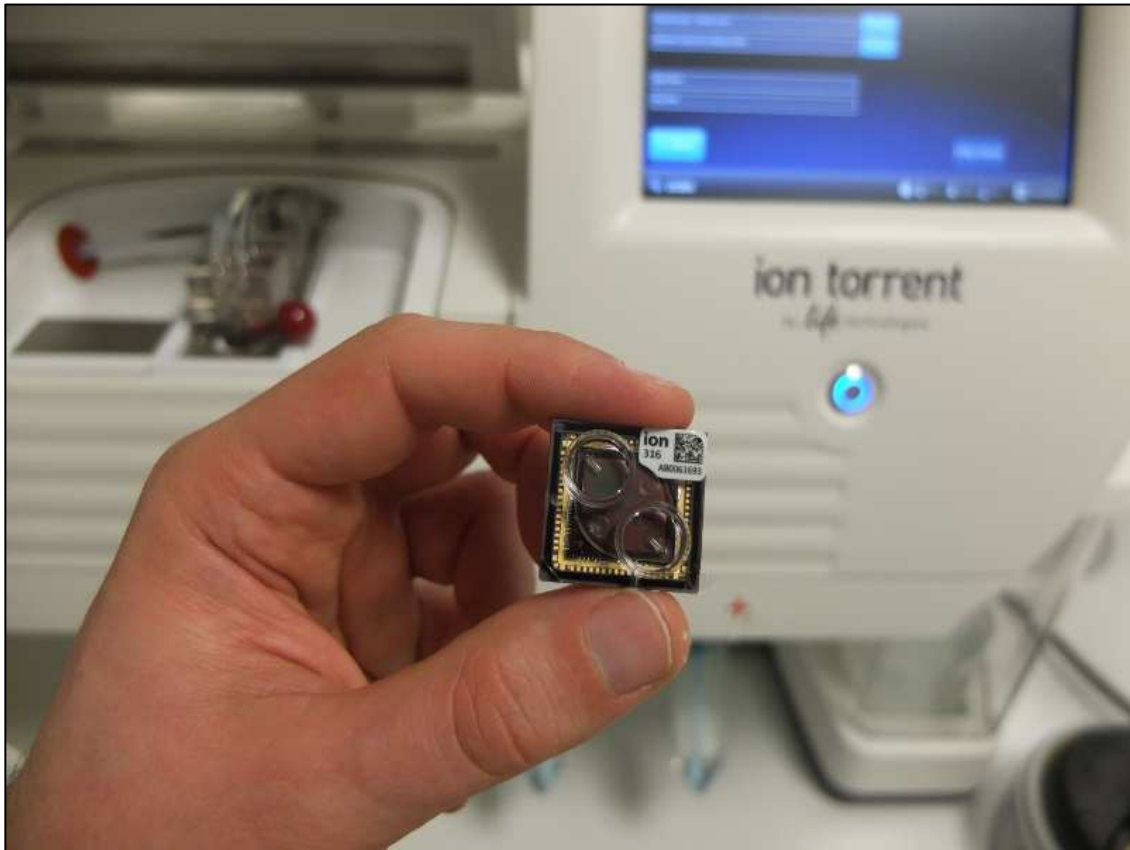


Figure n°6 : Une puce de séquençage

Cette méthode de séquençage est basée sur la détection d'ions hydrogène qui sont émis durant la polymérisation de l'ADN par introduction successive des différents dNTP (désoxyribonucléotides) L'introduction d'un dNTP complémentaire du brin d'ADN de l'échantillon, ce qui provoque un changement de pH du milieu. Ce changement est détecté par la puce qui transforme ce signal chimique en données numériques.

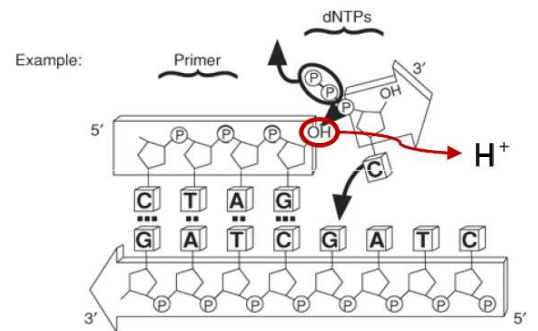
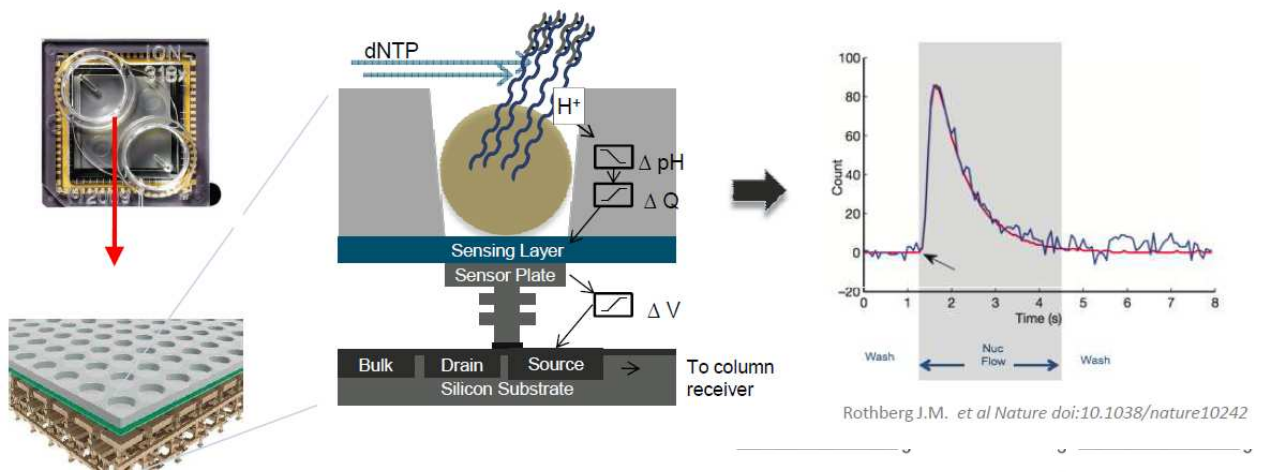
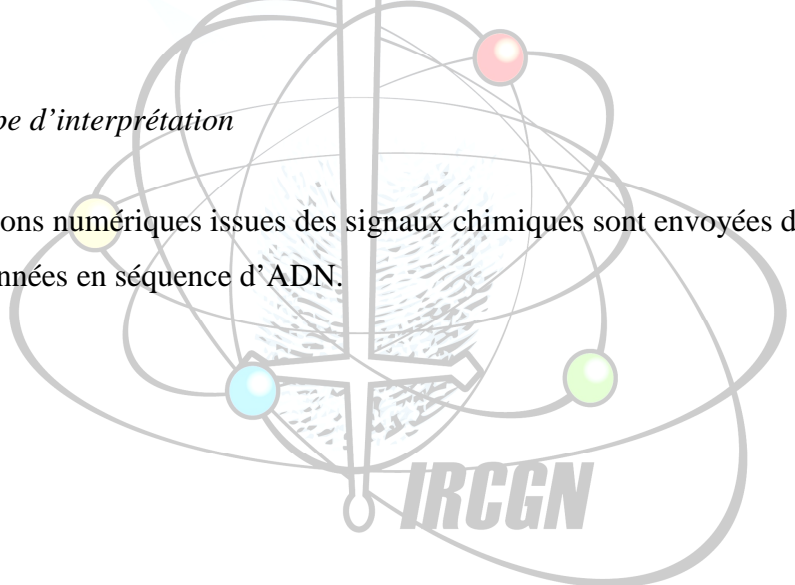


Figure n°7 : Schémas de fonctionnement d'un puits de la puce

d) Etape d'interprétation

Les informations numériques issues des signaux chimiques sont envoyées dans un serveur qui va convertir ses données en séquence d'ADN.



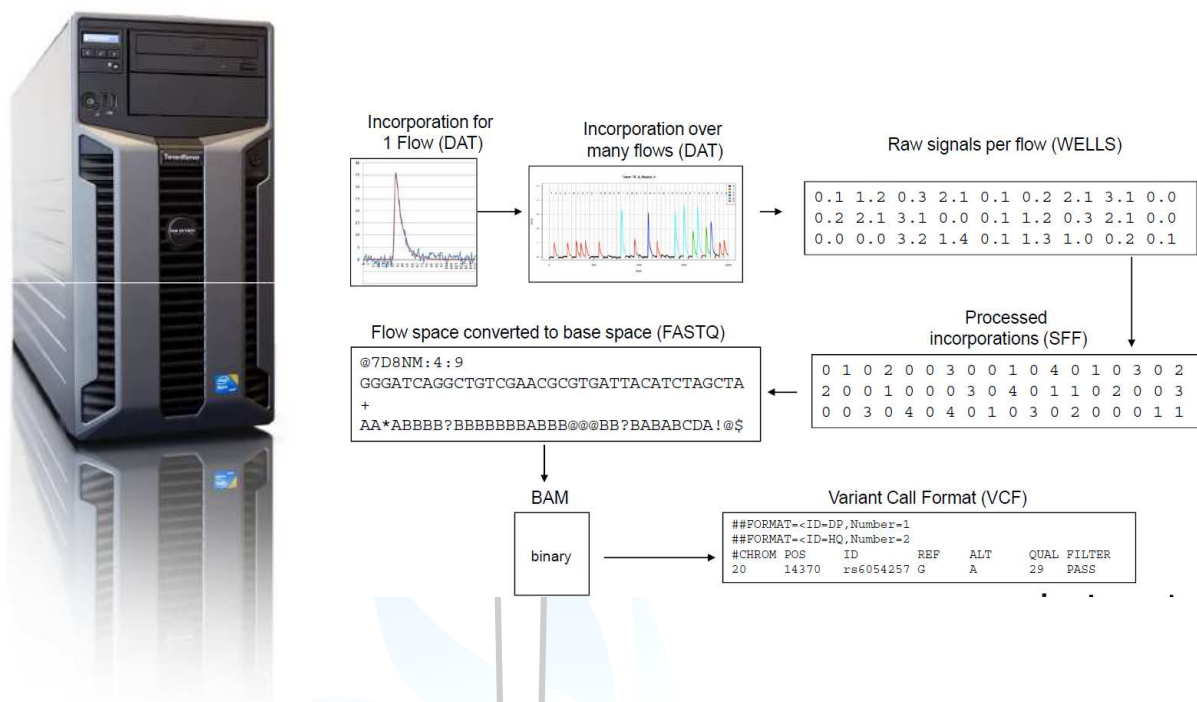


Figure n°8 : Conversion par le serveur des données numériques en séquence nucléotidique

BUT DE L'ETUDE : Est-il possible de prédire un portrait robot génétique d'une cohorte d'individus à l'aide du séquenceur PGM ?

Afin de répondre à cette question nous avons travaillé sur 3 axes, à savoir :

- la concordance entre les prédictions et les phénotypes étudiés (couleur des yeux, couleur des cheveux et origine biogéographique) des individus volontaires.
- la reproductibilité. Le test consiste à faire réaliser exactement la même expérience par des **opérateurs différents** en utilisant les mêmes échantillons, les mêmes types d'appareils et les mêmes méthodes d'analyses, afin d'obtenir dans notre étude le même génotype.
- la sensibilité. Il s'agit des tests de limite de détection permettant la détermination de la plus petite quantité d'ADN permettant d'obtenir, de façon reproductible un même génotype.

MATERIEL ET METHODES

1) Cohorte de volontaires.

L'ADN de 100 volontaires, identifié chacun par un code-barres de PGM00001 à PGM0000100, a été prélevé à l'aide d'écouvillons buccaux. Chaque volontaire de l'étude a complété et signé un consentement éclairé (cf annexe n°1), puis remplit un questionnaire concernant son morphotype (cf annexe n°2). Enfin pour chacun d'entre eux une capture photographique de l'iris et des cheveux est effectuée. Toutes les données obtenues sont consignées dans un tableau (cf. annexe n°3).

2) Extraction quantification et dilution de l'ADN prélevé

L'ADN des écouvillons buccaux est extrait sur la plate-forme robotisée du SCAGG-T. Les extraits d'ADN obtenus sont quantifiés puis sont dilués pour obtenir une concentration de 0,17 ng/μL. Le test est réalisé pour une quantité d'ADN de 1ng (soit 6 μl de librairie).

3) Panel de SNP utilisés

Le panel « Ancestry » commercialisé par la société Life Technologie est utilisé pour la prédiction des origines biogéographiques. Ce kit contient les amorces de 151 SNPs, marqueurs génétiques de l'origine des populations.

Le panel « *HirisPlex* » est utilisé pour la prédiction de la couleur des yeux et des cheveux. Ce panel contient les amorces de 24 SNPs (cf annexe n° 4), marqueurs génétiques de la couleur des yeux et des cheveux.

4) Données brutes

Les résultats de séquençage se présentent sous forme de rapport (cf annexe n° 5) comprenant les informations suivantes pour chaque séquençage :

- le % de remplissage de la puce

- le nombre total de lecture
- le % de lectures utilisable
- le % d'enrichissement
- les % d'ADN monoclonal et polyclonal
- la longueur des fragments d'ADN séquencés
- Pour chaque échantillon, le nombre de lectures

En fonction des séquences obtenues, un caractère phénotypique ou une origine biogéographique pourra être prédite.

5) Prédiction de l'origine biogéographique

20 échantillons de la cohorte sont utilisés pour l'étude de la prédiction de l'origine biogéographique (cf annexe n°3).

Les données brutes sont présentées sous forme de tableau indiquant en %, le degré d'appartenance aux 7 régions du monde suivantes : Amérique, Afrique, Caucase, Océanie, Asie de l'Est, Asie du Sud et Asie du Sud Ouest. Un degré de confiance est également indiqué par le logiciel, à savoir faible ou élevé. Ce degré de confiance est obtenu en fonction : du nombre total de SNP séquencé ; des fréquences alléliques de chaque population.

6) Prédiction de la couleur des yeux et des cheveux

60 échantillons de la cohorte sont utilisés pour la prédiction de deux caractères phénotypiques : la couleur des yeux et la couleur des cheveux.

Les données sont indiquées en possibilité d'appartenance aux 3 couleurs des yeux (bleu, intermédiaire, marron), à la couleur des cheveux (blond, roux, marron, noir) et à la teinte des cheveux (clair ou foncé).

Dans les deux cas, les données brutes sont comparées avec les informations recueillis des individus de la cohorte.

7) Etude de la reproductibilité

40 échantillons de la cohorte sont utilisés par deux opérateurs différents. Les génotypes, de la couleur des yeux et des cheveux, obtenus par chaque opérateur, sont comparés.

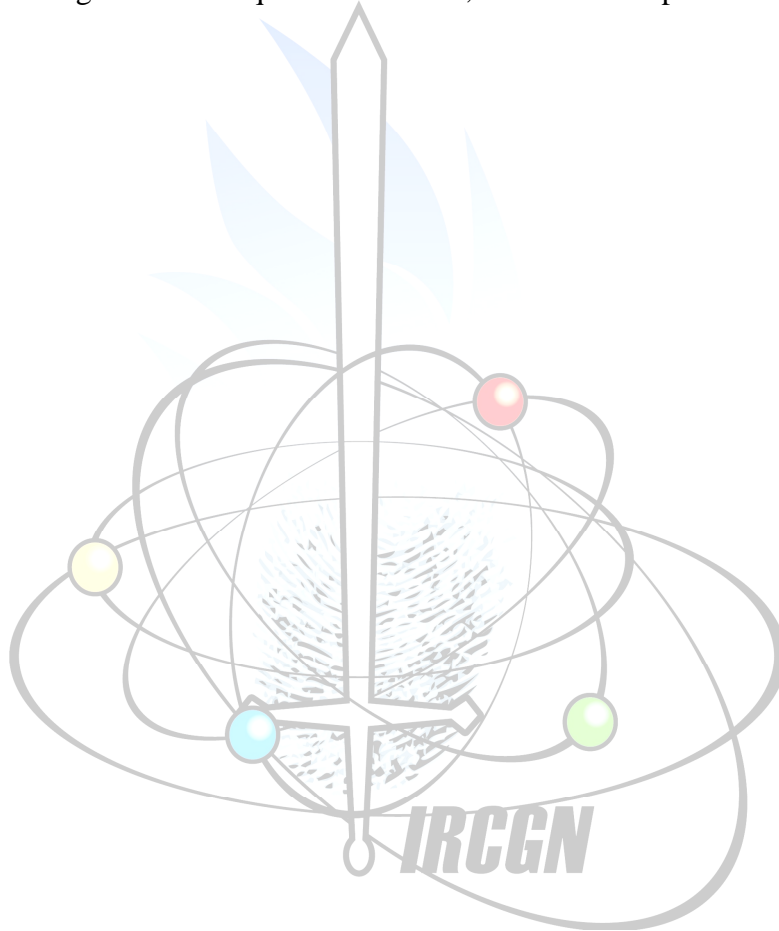
8) Etude de la sensibilité

Pour le test de sensibilité, les ADN de 2 individus différents identifiés n°1 et n°2, sont utilisés. Ces ADN sont dilués afin d'obtenir une gamme décroissante en quantité d'ADN.

Pour l'individu n°1, les quantités en picogrammes (pg) d'ADN sont: 329 ; 156, 75, 53, 19 et 11.

Pour l'individu n° 2, les quantités en picogrammes (pg) d'ADN sont: 323 ; 181 ; 81 ; 54 ; 22 et 9.

Chaque point de la gamme est séquencé deux fois, en utilisant le panel « *HirisPlex* ».



RESULTATS

I) Prédiction des origines biogéographiques

Nous avons fait le choix d'analyser en partie des échantillons d'individus issus de métissage, afin de connaître les limites de la prédiction des origines biogéographiques.

Echantillon	PREDICTION POPULATION (%)							Degré de Confiance	Origine donnée par les volontaires
	Amérique	Asie Est	Océanie	Afrique	Europe	Asie Sud	Asie Sud Ouest		
PGM00004	10		5		85			Elevé	Europe
PGM00007				50	35		15	<i>Faible</i>	Europe Réunion
PGM000010					100			<i>Faible</i>	Europe Kazakhstan
PGM000011		15	15	15	55			Elevé	Europe Réunion, Inde
PGM000043					50	5	45	Elevé	Europe Algérie (Pieds noir)
PGM000044					100			Elevé	Europe Liban
PGM000046	15			25	60			Elevé	Europe Mexique
PGM000052					100			Elevé	Europe
PGM000067	5	35	5	55				Elevé	Mayotte
PGM000071			5		60	35		Elevé	Europe Ile Maurice
PGM000073	5			15	10	70		Elevé	Asie du Sud
PGM000074			5		95			Elevé	Europe
PGM000086			5		90	5		Elevé	Europe
PGM000088					100			Elevé	Europe
PGM000089			10	10	40	40		Elevé	Europe Asie du Sud
PGM000090			10	35	55			<i>Faible</i>	Europe Antilles
PGM000091		25	10	20			45	Elevé	Asie Ouest Afrique
PGM000098		100						Elevé	Asie
PGM000099				20	5	55	20	Elevé	Réunion
PGM0000100					60		40	Elevé	Europe

Tableau n°1 : Prédiction des origines biogéographiques de 20 échantillons

Sur les 20 prédictions, 3 obtiennent un degré de confiance faible, soit 15% de l'ensemble des échantillons étudiés.

Sur les individus se disant strictement d'origine européenne, nous obtenons une prédiction concordante à 86%. Pour un seul échantillon le n°PGM0000100, il n'y a pas de concordance entre la prédiction et les informations fournies. En effet l'individu s'est déclaré d'origine strictement européenne or il est obtenu un résultat de prédiction de 60% européen et 40% d'Asie du Sud Ouest.

Concernant les individus se disant strictement d'origine asiatique (2 échantillons), les prédictions de l'origine biogéographique concordent avec les données recueillies.

8 individus ont un parent d'origine européenne et un autre parent d'une autre origine. Sur ces 8 échantillons, 6 d'entre eux présentent une prédiction de l'origine européenne aux environs de 50%. Seuls les échantillons PGM00007 et PGM000044 n'obtiennent pas une prédiction de l'origine européenne aux environs de 50%. En effet, il est de 35% pour l'échantillon PGM00007 et de 100% pour l'échantillon PGM000044. Concernant la deuxième origine de ces 8 individus la concordance entre les prédictions et les informations fournies est de 75%.

Pour les 3 derniers échantillons, n° PGM000067, PGM000091 et PGM000099, il y a une correspondance entre les informations fournies et la prédiction des origines biogéographiques.

II) Prédiction des caractères phénotypiques

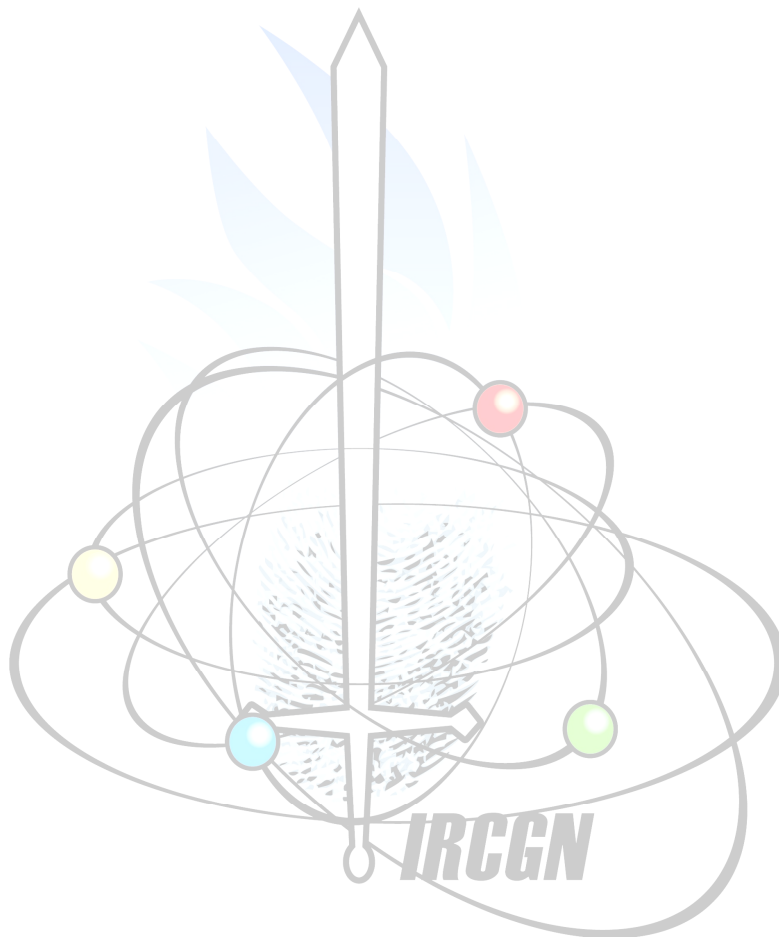
1) La couleur des yeux

Les résultats de la prédiction de la couleur sont présentés de la manière suivante :

- numéro de l'échantillon
- photographie de l'iris de l'individu correspondant à l'échantillon
- le score obtenu pour chaque couleur de l'iris (Bleu, Inter pour Intermédiaire et Marron)
- la prédiction de la couleur de l'œil de l'échantillon. Cette dernière est déterminée en fonction du score de la couleur le plus élevé entre la couleur bleu et la couleur marron. Si la couleur bleu ou la couleur marron n'atteint pas un score supérieur ou égal à 0,7 la couleur prédite est la couleur intermédiaire.
- un tableau de concordance entre les prédictions et les données fournies par les volontaires.

Remarque :

- La photographie de l'iris, correspondant à l'individu dont l'échantillon est le n° PGM000064, n'est pas présentée, car non fournie.
- les résultats sont triés dans l'ordre décroissant du score obtenu pour la couleur bleu



PGM000017			PGM000014			PGM000052			PGM000074			PGM000027			PGM000031			PGM00003			PGM00004		
Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron
0.968	0.027	0.005	0.953	0.032	0.015	0.953	0.032	0.015	0.946	0.043	0.011	0.931	0.053	0.016	0.929	0.054	0.017	0.917	0.052	0.031	0.917	0.052	0.031

BLEU

BLEU

BLEU

BLEU

BLEU

BLEU

BLEU

BLEU

PGM000094			PGM00002			PGM000015			PGM000062			PGM000024			PGM000060 : YEUX VAIRON			PGM000021		
Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron
0.917	0.052	0.031	0.891	0.068	0.041	0.891	0.068	0.041	0.891	0.068	0.041	0.858	0.081	0.061	0.806	0.077	0.117	0.681	0.109	0.21

BLEU

BLEU

BLEU

BLEU

BLEU

BLEU

INTER

PGM00005			PGM000063			PGM000034			PGM000026			PGM000097			PGM000016			PGM000070			PGM000096		
Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron
0.623	0.275	0.102	0.35	0.228	0.422	0.274	0.22	0.506	0.259	0.235	0.506	0.259	0.235	0.506	0.252	0.165	0.583	0.215	0.233	0.552	0.196	0.173	0.631

INTER

INTER

INTER

INTER

INTER

INTER

INTER

INTER

PGM000032			PGM000010			PGM000036			PGM000044			PGM000009			PGM000028			PGM000058			PGM000059					
Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron			
0.196	0.173	0.631	0.16	0.213	0.627	0.16	0.213	0.627	0.15	0.227	0.623	0.143	0.155	0.702	0.143	0.155	0.702	0.143	0.155	0.702	0.143	0.155	0.702	0.143	0.155	0.702

INTER

INTER

INTER

INTER

MARRON

MARRON

MARRON

MARRON

PGM000049			PGM000012			PGM000029			PGM000078			PGM000047			PGM000050			PGM000046			PGM000011		
Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron
0.121	0.218	0.66	0.101	0.136	0.763	0.101	0.136	0.763	0.101	0.136	0.763	0.087	0.192	0.721	0.087	0.192	0.721	0.083	0.092	0.826	0.052	0.116	0.832

INTER

MARRON

MARRON

MARRON

MARRON

MARRON

MARRON

MARRON

PGM000019			PGM000022			PGM000083			PGM000030			PGM000007			PGM000089			PGM000061			PGM000001		
Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron
0.052	0.116	0.832	0.052	0.116	0.832	0.052	0.116	0.832	0.041	0.077	0.882	0.028	0.064	0.909	0.028	0.064	0.909	0.014	0.051	0.935	0.005	0.059	0.937

MARRON

MARRON

MARRON

MARRON

MARRON

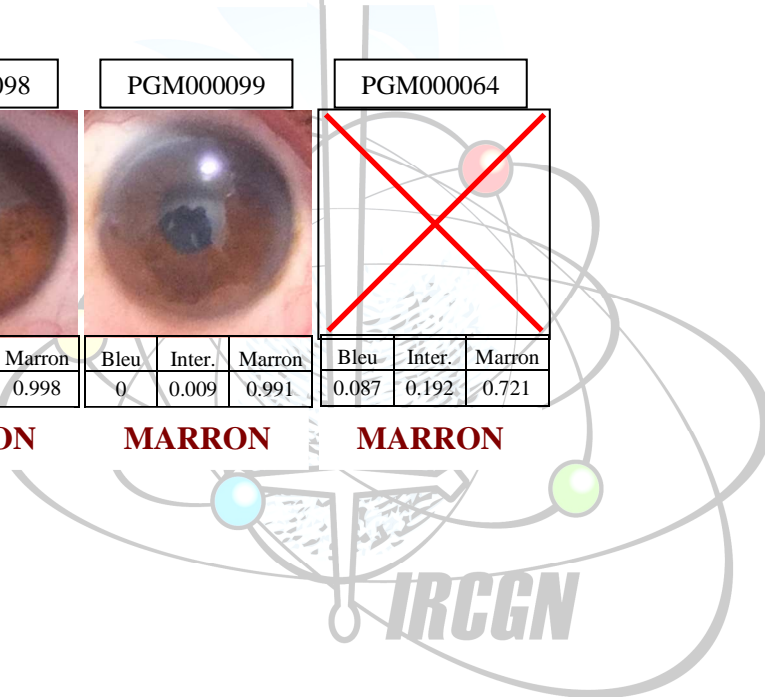
MARRON

MARRON

MARRON

PGM00008			PGM000067			PGM000020			PGM000081			PGM000018			PGM000040			PGM000043			PGM000090					
Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron
0.004	0.085	0.911	0.003	0.021	0.976	0.002	0.043	0.956	0.001	0.019	0.981	0	0.003	0.997	0	0.002	0.998	0	0.009	0.99	0	0.007	0.993	0	0.007	0.993
MARRON			MARRON			MARRON			MARRON			MARRON			MARRON			MARRON			MARRON			MARRON		

PGM000091			PGM000095			PGM000098			PGM000099			PGM000064		
Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron	Bleu	Inter.	Marron
0	0.009	0.991	0	0.018	0.982	0	0.002	0.998	0	0.009	0.991	0.087	0.192	0.721
MARRON			MARRON			MARRON			MARRON			MARRON		



Echantillon	Prédiction	Photo de la couleur des yeux	Concordance	Echantillon	Prédiction	Photo de la couleur des yeux	Concordance
PGM000017	BLEU	BLEU	OUI	PGM000058	MARRON	INTER	NON
PGM000014	BLEU	BLEU	OUI	PGM000059	MARRON	MARRON	OUI
PGM000052	BLEU	BLEU	OUI	PGM000012	MARRON	INTER	NON
PGM000074	BLEU	BLEU	OUI	PGM000029	MARRON	MARRON	OUI
PGM000027	BLEU	BLEU	OUI	PGM000078	MARRON	MARRON	OUI
PGM000031	BLEU	BLEU	OUI	PGM000047	MARRON	INTER	NON
PGM000003	BLEU	BLEU	OUI	PGM000050	MARRON	INTER	NON
PGM000004	BLEU	BLEU	OUI	PGM000046	MARRON	MARRON	OUI
PGM000094	BLEU	BLEU	OUI	PGM000011	MARRON	MARRON	OUI
PGM000002	BLEU	BLEU	OUI	PGM000019	MARRON	MARRON	OUI
PGM000015	BLEU	BLEU	OUI	PGM000022	MARRON	MARRON	OUI
PGM000062	BLEU	BLEU	OUI	PGM000083	MARRON	MARRON	OUI
PGM000024	BLEU	BLEU	OUI	PGM000030	MARRON	MARRON	OUI
PGM000060	BLEU	VAIRON (BLEU)	OUI	PGM000007	MARRON	MARRON	OUI
PGM000021	INTER	INTER	OUI	PGM000089	MARRON	MARRON	OUI
PGM000005	INTER	INTER	OUI	PGM000061	MARRON	MARRON	OUI
PGM000063	INTER	MARRON	NON	PGM000001	MARRON	INTER	NON
PGM000034	INTER	INTER	OUI	PGM000008	MARRON	MARRON	OUI
PGM000026	INTER	INTER	OUI	PGM000067	MARRON	MARRON	OUI
PGM000097	INTER	INTER	OUI	PGM000020	MARRON	MARRON	OUI
PGM000016	INTER	INTER	OUI	PGM000081	MARRON	MARRON	OUI
PGM000070	INTER	INTER	OUI	PGM000018	MARRON	MARRON	OUI
PGM000096	INTER	INTER	OUI	PGM000040	MARRON	MARRON	OUI
PGM000032	INTER	MARRON	NON	PGM000043	MARRON	MARRON	OUI
PGM000010	INTER	INTER	OUI	PGM000090	MARRON	MARRON	OUI
PGM000036	INTER	INTER	OUI	PGM000091	MARRON	MARRON	OUI
PGM000044	INTER	INTER	OUI	PGM000095	MARRON	MARRON	OUI
PGM000049	INTER	INTER	OUI	PGM000098	MARRON	MARRON	OUI
PGM000009	MARRON	MARRON	OUI	PGM000099	MARRON	MARRON	OUI
PGM000028	MARRON	MARRON	OUI	PGM000064	MARRON	INTER	NON

Tableau n°2 : Concordance entre la prédiction de la couleur des yeux et la couleur des yeux donnée

Pour l'ensemble des 60 échantillons, il y a 87% de concordance entre la prédiction de la couleur des yeux et les données fournies par les volontaires. En revanche, les résultats sont très différents en fonction de la couleur des yeux à prédire. En effet, la concordance entre la prédiction et les données fournies est de :

- **100% pour les individus ayant les yeux de couleur bleu**
- **93% pour les individus ayant les yeux de couleur marron**
- **67% pour les individus ayant les yeux de couleur intermédiaire**

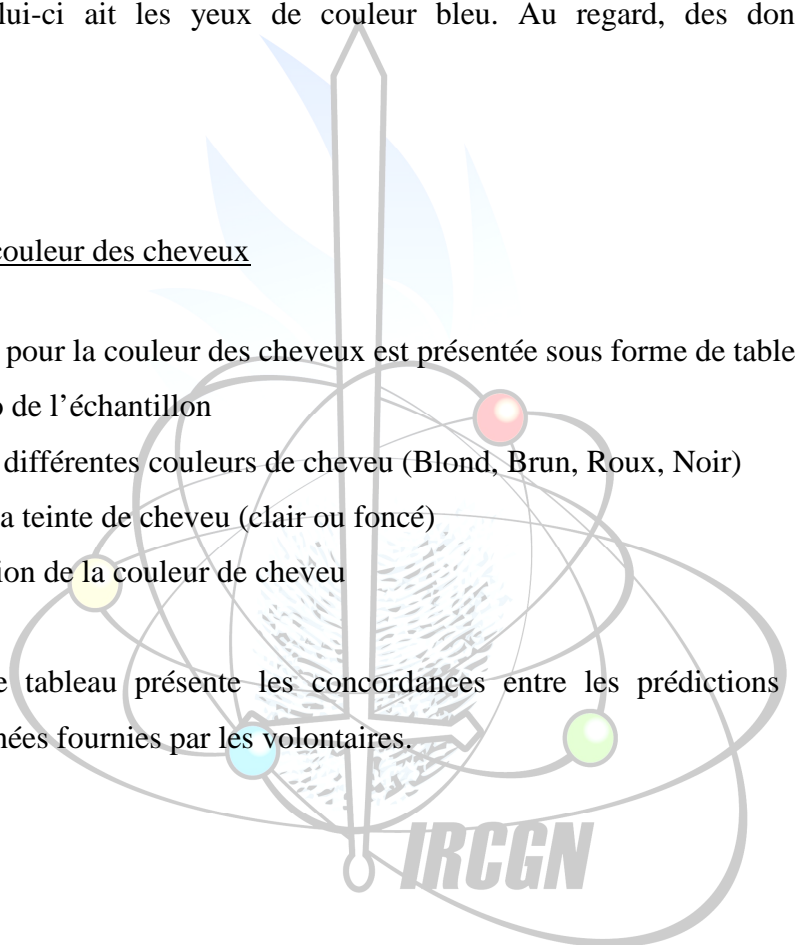
Mentionnons que l'échantillon correspondant à l'individu aux yeux vairon donne une prédiction que celui-ci ait les yeux de couleur bleu. Au regard, des données fournis, cela correspond.

2) La couleur des cheveux

La prédiction pour la couleur des cheveux est présentée sous forme de tableau, incluant :

- le numéro de l'échantillon
- les % des différentes couleurs de cheveu (Blond, Brun, Roux, Noir)
- les % de la teinte de cheveu (clair ou foncé)
- la prédiction de la couleur de cheveu

Un deuxième tableau présente les concordances entre les prédictions de la couleur des cheveux et les données fournies par les volontaires.



Echantillon	blond	brun	roux	noir	clair	fonce	Prédiction	Echantillon	blond	brun	roux	noir	clair	fonce	Prédiction
PGM00001	0.168	0.64	0.002	0.19	0.485	0.515	Brun foncé - Noir	PGM000040	0.003	0.183	0	0.814	0.003	0.997	Noir
PGM00002	0.654	0.309	0.018	0.019	0.969	0.031	Blond - Blond foncé	PGM000043	0.028	0.627	0.025	0.32	0.09	0.91	Brun foncé - Noir
PGM00003	0.687	0.261	0.005	0.047	0.927	0.073	Blond - Blond foncé	PGM000044	0.458	0.503	0.003	0.036	0.886	0.114	Brun - Brun foncé
PGM00004	0.762	0.222	0.002	0.014	0.976	0.024	Blond	PGM000046	0.077	0.31	0	0.613	0.145	0.855	Noir
PGM00005	0.04	0.799	0.003	0.158	0.242	0.758	Brun - Brun foncé	PGM000047	0.012	0.139	0.847	0.002	0.895	0.105	Roux
PGM00007	0.092	0.54	0	0.368	0.16	0.84	Brun foncé - Noir	PGM000049	0.18	0.629	0.044	0.147	0.549	0.451	Brun foncé - Noir
PGM00008	0.035	0.629	0.002	0.333	0.082	0.918	Brun - Brun foncé	PGM000050	0.048	0.812	0.002	0.138	0.214	0.786	Brun - Brun foncé
PGM00009	0.383	0.318	0.295	0.005	0.989	0.011	Blond - Blond foncé	PGM000052	0.686	0.182	0.126	0.006	0.994	0.006	Blond - Blond foncé
PGM000010	0.161	0.77	0.004	0.065	0.597	0.403	Brun - Brun foncé	PGM000058	0.386	0.351	0.256	0.007	0.977	0.023	Blond - Blond foncé
PGM000011	0.364	0.447	0.001	0.187	0.664	0.336	Brun foncé - Noir	PGM000059	0.193	0.157	0.645	0.004	0.99	0.01	Roux
PGM000012	0.434	0.382	0.165	0.019	0.963	0.037	Blond - Blond foncé	PGM000060	0.343	0.41	0.002	0.246	0.645	0.355	Brun foncé - Noir
PGM000014	0.639	0.18	0.171	0.01	0.987	0.013	Blond - Blond foncé	PGM000061	0.101	0.334	0.001	0.564	0.17	0.83	Noir
PGM000015	0.354	0.553	0.079	0.014	0.963	0.037	Brun - Brun foncé	PGM000062	0.645	0.298	0.002	0.054	0.909	0.091	Blond - Blond foncé
PGM000016	0.584	0.354	0.002	0.06	0.905	0.095	Blond - Blond foncé	PGM000063	0.093	0.74	0.003	0.164	0.387	0.613	Brun - Brun foncé
PGM000017	0.363	0.549	0.006	0.082	0.82	0.18	Brun - Brun foncé	PGM000064	0.09	0.758	0.003	0.149	0.375	0.625	Brun - Brun foncé
PGM000018	0.004	0.18	0	0.816	0.003	0.997	Noir	PGM000067	0.007	0.163	0	0.829	0.009	0.991	Noir
PGM000019	0.279	0.602	0.002	0.116	0.627	0.373	Brun foncé - Noir	PGM000070	0.121	0.677	0.003	0.199	0.435	0.565	Brun foncé - Noir
PGM000020	0.023	0.489	0	0.488	0.04	0.96	Brun - Brun foncé	PGM000074	0.34	0.53	0.107	0.023	0.953	0.047	Brun - Brun foncé
PGM000021	0.371	0.501	0.003	0.125	0.679	0.321	Brun foncé - Noir	PGM000078	0.474	0.473	0.004	0.048	0.863	0.137	Blond foncé-brun
PGM000022	0.274	0.592	0.003	0.132	0.555	0.445	Brun foncé - Noir	PGM000081	0.065	0.478	0	0.456	0.19	0.81	Brun - Brun foncé
PGM000024	0.447	0.471	0.04	0.042	0.838	0.162	Brun - Brun foncé	PGM000083	0.155	0.098	0.745	0.001	0.996	0.004	Roux
PGM000026	0.566	0.352	0.005	0.077	0.872	0.128	Blond foncé-brun	PGM000089	0.102	0.44	0.001	0.457	0.216	0.784	Noir
PGM000027	0.411	0.512	0.007	0.071	0.858	0.142	Brun - Brun foncé	PGM000090	0.001	0.176	0	0.824	0.001	0.999	Noir
PGM000028	0.271	0.553	0.003	0.173	0.656	0.344	Brun foncé - Noir	PGM000091	0.011	0.176	0	0.813	0.017	0.983	Noir
PGM000029	0.4	0.419	0.005	0.176	0.72	0.28	Brun foncé - Noir	PGM000094	0.54	0.166	0.292	0.003	0.995	0.005	Blond - Blond foncé
PGM000030	0.154	0.543	0.002	0.302	0.318	0.682	Brun foncé - Noir	PGM000095	0.078	0.412	0	0.509	0.128	0.872	Noir
PGM000031	0.045	0.778	0.002	0.175	0.197	0.803	Brun	PGM000096	0.099	0.063	0.837	0.001	0.994	0.006	Roux
PGM000032	0.397	0.445	0.003	0.155	0.718	0.282	Brun foncé - Noir	PGM000097	0.006	0.023	0.97	0.001	0.917	0.083	Roux
PGM000034	0.038	0.8	0.077	0.085	0.293	0.707	Brun - Brun foncé	PGM000098	0.001	0.056	0	0.943	0.003	0.997	Noir
PGM000036	0.119	0.697	0.003	0.181	0.338	0.662	Brun foncé - Noir	PGM000099	0.018	0.44	0	0.542	0.022	0.978	Noir

Tableau n°3 : Résultat des prédictions de la couleur des cheveux

Echantillon	Prédiction	Données fournies sur la couleur des cheveux	Concordance	Echantillon	Prédiction	Couleur des cheveux Couleur des cheveux	Concordance
PGM00001	Brun foncé - Noir	Brun foncé	OUI	PGM000040	Noir	Noir	OUI
PGM00002	Blond - Blond foncé	Blond foncé	OUI	PGM000043	Brun foncé - Noir	Brun foncé	OUI
PGM00003	Blond - Blond foncé	Brun	NON	PGM000044	Brun - Brun foncé	Brun foncé	OUI
PGM00004	Blond	Noir	NON	PGM000046	Noir	Noir	OUI
PGM00005	Brun - Brun foncé	Brun	OUI	PGM000047	Roux	Roux	OUI
PGM00007	Brun foncé - Noir	Noir	OUI	PGM000049	Brun foncé - Noir	Noir	OUI
PGM00008	Brun - Brun foncé	Brun foncé	OUI	PGM000050	Brun - Brun foncé	Noir	NON
PGM00009	Blond - Blond foncé	Brun	NON	PGM000052	Blond - Blond foncé	Blond foncé	OUI
PGM000010	Brun - Brun foncé	Noir	NON	PGM000058	Blond - Blond foncé	Blond foncé	OUI
PGM000011	Brun foncé - Noir	Noir	OUI	PGM000059	Roux	Roux	OUI
PGM000012	Blond - Blond foncé	Brun	NON	PGM000060	Brun foncé - Noir	Brun foncé	OUI
PGM000014	Blond - Blond foncé	Blond	OUI	PGM000061	Noir	Noir	OUI
PGM000015	Brun - Brun foncé	Blond foncé	NON	PGM000062	Blond - Blond foncé	Brun	NON
PGM000016	Blond - Blond foncé	Brun	NON	PGM000063	Brun - Brun foncé	Brun foncé	OUI
PGM000017	Brun - Brun foncé	Brun	OUI	PGM000064	Brun - Brun foncé	Brun foncé	OUI
PGM000018	Noir	Noir	OUI	PGM000067	Noir	Noir	OUI
PGM000019	Brun foncé - Noir	Brun	OUI	PGM000070	Brun foncé - Noir	Brun	OUI
PGM000020	Brun - Brun foncé	Brun foncé	OUI	PGM000074	Brun - Brun foncé	Blond foncé	NON
PGM000021	Brun foncé - Noir	Brun foncé	OUI	PGM000078	Blond foncé-brun	Blond foncé	OUI
PGM000022	Brun foncé - Noir	Noir	OUI	PGM000081	Brun - Brun foncé	Brun foncé	OUI
PGM000024	Brun - Brun foncé	Brun foncé	OUI	PGM000083	Roux	Roux	OUI
PGM000026	Blond foncé-brun	Brun	OUI	PGM000089	Noir	Noir	OUI
PGM000027	Brun - Brun foncé	Noir	NON	PGM000090	Noir	Noir	OUI
PGM000028	Brun foncé - Noir	Brun foncé	OUI	PGM000091	Noir	Brun foncé	NON
PGM000029	Brun foncé - Noir	Brun foncé	OUI	PGM000094	Blond - Blond foncé	Blond foncé	OUI
PGM000030	Brun foncé - Noir	Brun foncé	OUI	PGM000095	Noir	Noir	OUI
PGM000031	Brun	Brun foncé	OUI	PGM000096	Roux	Roux	OUI
PGM000032	Brun foncé - Noir	Brun foncé	OUI	PGM000097	Roux	Roux	OUI
PGM000034	Brun - Brun foncé	Brun foncé	OUI	PGM000098	Noir	Noir	OUI
PGM000036	Brun foncé - Noir	Noir	OUI	PGM000099	Noir	Noir	OUI

Tableau n°4 : Concordance entre les prédictions et la couleur des cheveux

Pour l'ensemble des 60 échantillons, il y a 80% de concordance entre les prédictions de la couleur de cheveu et les données de la cohorte. Néanmoins, les résultats sont différents en fonction de la couleur de cheveu. En effet, la concordance entre la prédiction et la couleur de cheveu « actuelle » de l'individu est de :

- 100% pour les individus ayant la couleur de cheveu roux (5/5)
- 79% pour les individus ayant la couleur de cheveu noir (15/19)
- 78,5% pour les individus ayant la couleur de cheveu brun (22/28)
- 75% pour les individus ayant la couleur de cheveu blond (6/8)

Parmi les discordances observées, quasiment toutes sont minimales : une prédiction de coloration brun foncé alors que la couleur réelle est noire et inversement; une prédiction de couleur de couleur blond foncé alors que la coloration réelle est brun clair et inversement. Un seul échantillon, le numéro PGM00004 a un résultat totalement discordant, à savoir une prédiction de couleur de cheveux blond alors que l'individu a les cheveux de couleur noire. Nous n'avons aucune explication pour justifier ce résultat.

III) Etude de la reproductibilité

Deux opérateurs différents, le capitaine PUSSIAU et le gendarme DESBUARDS, ont séquencé et analysé, chacun de leur côté, les 40 mêmes échantillons en utilisant le panel « *HirisPlex* » afin de prédire la couleur des yeux et des cheveux. Les résultats des génotypages obtenus sont strictement identiques pour les deux opérateurs.

IV) Etude de la sensibilité

Les données obtenues sont présentées sous forme de tableau incluant :

- le numéro de l'individu
- les différentes quantités d'ADN
- les scores des différentes couleurs des yeux
- les scores des différentes couleurs des cheveux
- les prédictions de la couleur des yeux et de la couleur des cheveux

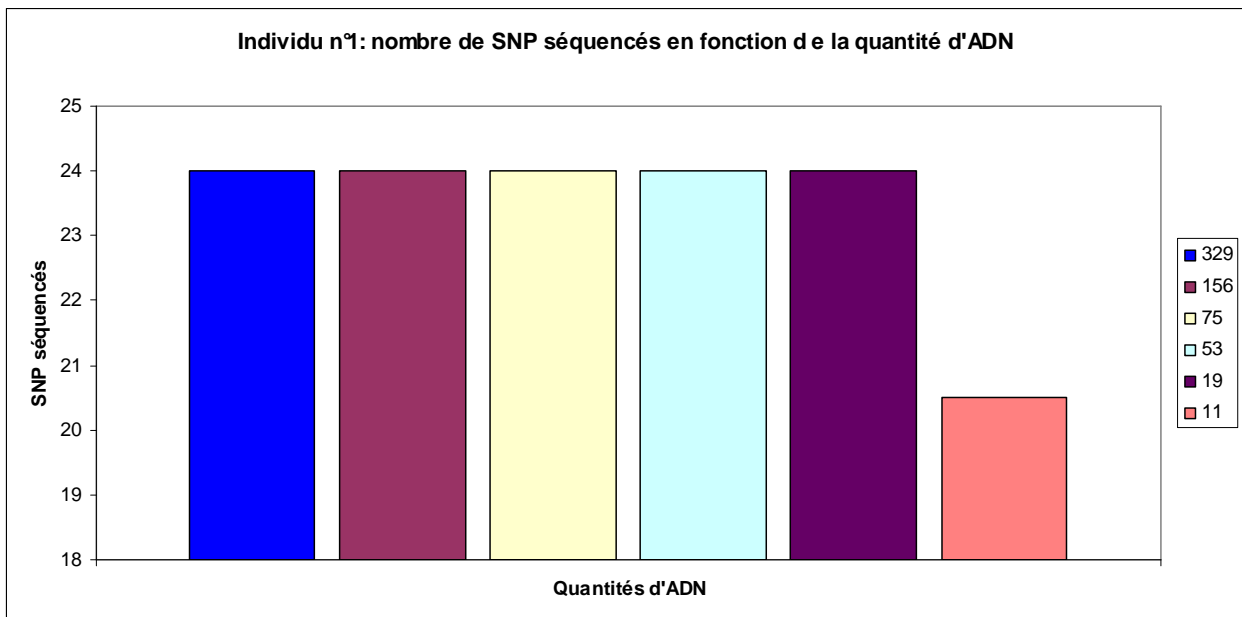
Le nombre de SNP séquencés en fonction de la quantité d'ADN est également présenté.

Individu	Quantité (pg)	YEUX				CHEVEUX						Prédiction
		Bleus	Inter	Marron	Prédiction	Blond	Brun	Roux	Noir	Clair	Foncé	
1	329	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	329	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	156	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	156	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	75	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	75	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	53	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	53	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	19	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	19	0.972	0.022	0.007	BLEU	0.702	0.22	0.002	0.076	0.93	0.07	Blond-Blond foncé
	11	0.972	0.022	0.007	BLEU	0.734	0.223	0.003	0.04	0.95	0.05	Blond
11	0.937	0.036	0.027	BLEU	0.345	0.259	0.001	0.395	0.719	0.281	Brun foncé - Noir	
2	323	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	323	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	181	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	181	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	81	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	81	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	54	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	54	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	22	0.936	0.043	0.021	BLEU	0.767	0.198	0.001	0.034	0.963	0.037	Blond
	22	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
	9	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond
9	0.936	0.043	0.021	BLEU	0.801	0.17	0.001	0.027	0.972	0.028	Blond	

Tableau n°5 : Résultats de la prédiction de la couleur des yeux et des cheveux pour différentes quantités d'ADN

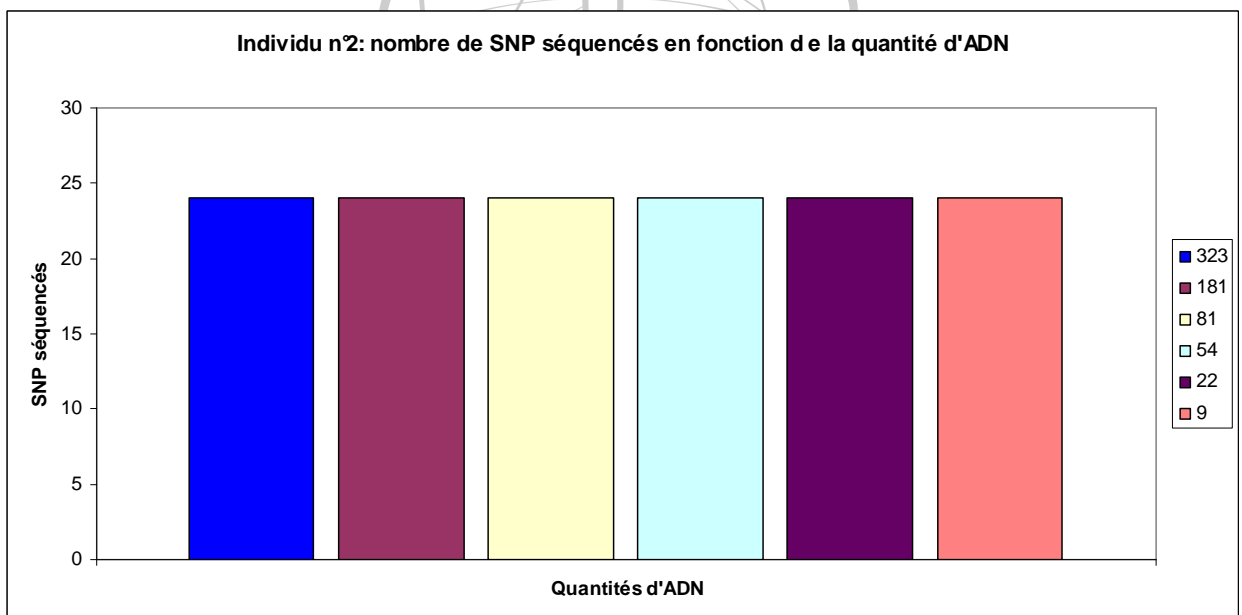
Pour l'individu n°1, les mêmes résultats de prédiction de la couleur des yeux et de la couleur des cheveux sont obtenus pour les quantités d'ADN de 329 pg à 19 pg. Il y a une légère différence de prédiction pour le premier séquençage de l'ADN à une quantité de 11 pg. En effet, l'individu n°1 est prédit blond alors que les autres résultats le donnent blond-blond foncé. Le second séquençage de l'ADN à 11 pg donne lui des résultats de prédiction totalement différents. Les cheveux de l'individu n°1 sont prédits de couleur brun foncé ce qui est totalement à l'opposé des résultats obtenus pour les autres quantités d'ADN. Concernant la prédiction des yeux celle-ci est identique quelque soit la quantité d'ADN avec cependant, une différence au niveau des scores obtenus pour chaque couleur des yeux, à la quantité de 11 ng.

Pour l'individu n° 2, les mêmes résultats de prédiction de la couleur des yeux et de la couleur des cheveux sont obtenus pour les quantités d'ADN de 323 pg à 54 pg puis pour 9 pg. Pour la quantité d'ADN à 22 pg, nous observons une différence des scores concernant les différentes couleurs de cheveux bien que la prédiction obtenue soit identique à celle des autres quantités d'ADN.



Graphique n°1 : Nombre de SNP séquencés en fonction de la quantité d'ADN pour l'individu n°1

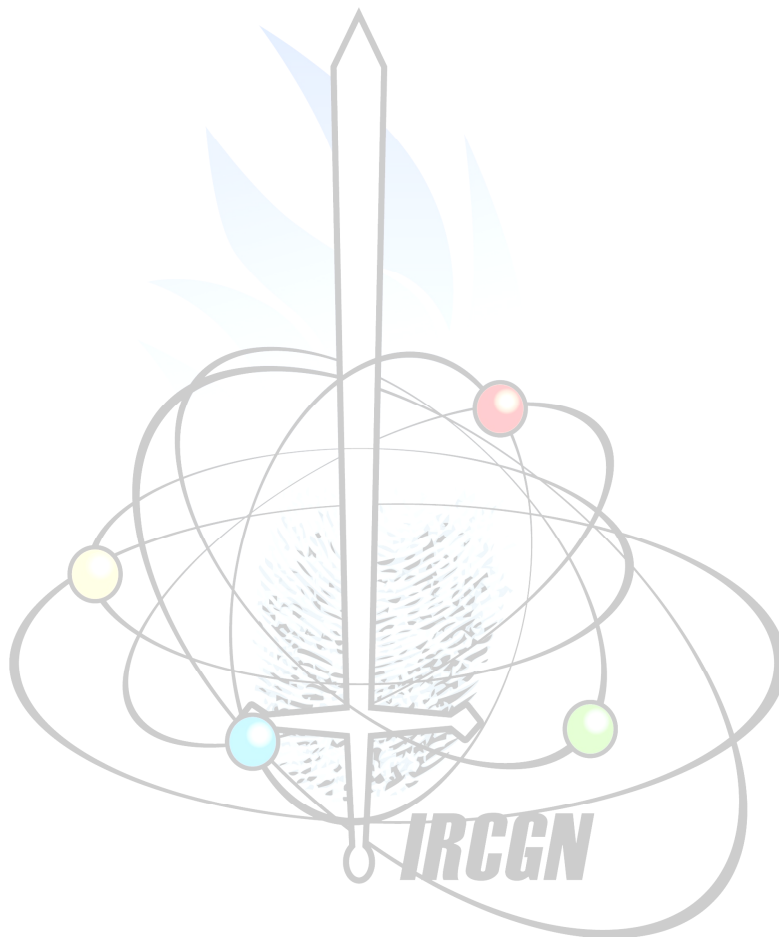
Pour l'individu n°1, la totalité des 24 SNPs du panel « *HirisPlex* » sont séquencés pour les quantités d'ADN de 329 à 19 pg. Pour la quantité de 11 pg, le nombre de SNP séquencé diminue pour une moyenne de 20,5 SNP séquencés.



Graphique n° 2 : Nombre de SNP séquencés en fonction de la quantité d'ADN pour l'individu n°2

Pour l'individu n°2, la totalité des 24 SNPs du panel « *HirisPlex* » sont séquencés pour toutes les quantités d'ADN. La différence des résultats obtenus entre la quantité d'ADN à 22 pg et les autres quantités s'explique par le fait d'une différence de génotypage pour un des SNP.

Au vu des résultats, le séquençage de l'ADN est fiable et reproductible pour une quantité d'ADN minimale d'environ 50 pg.



DISCUSSION ET PERSPECTIVES

Dans cette étude, nous avons séquencé différents SNP d'intérêts à l'aide d'un séquenceur haut débit de seconde génération, le PGM afin de tester ; d'une part la sensibilité de l'appareil ; d'autre part de tester la reproductibilité, et enfin d'évaluer les limites d'interprétation d'un portrait-robot génétique. Nous avons utilisé l'ADN de volontaires pour prédire leur origine biogéographique et deux caractères phénotypiques (la couleur des yeux et des cheveux), et de comparer les résultats obtenus avec les données recueillis auprès des volontaires.

Dans le cadre de notre étude, nous avons testé la reproductibilité et la sensibilité d'un séquenceur haut débit de seconde génération, le PGM. Pour tester la reproductibilité, deux opérateurs ont séquencé des SNP d'intérêt chacun de leur côté, en utilisant le panel « *HirisPlex* ». Les résultats des génotypages obtenus sont strictement identiques pour les deux opérateurs, montrant ainsi la reproductibilité du PGM.

Pour tester la sensibilité de l'appareil, des SNP issus d'ADN à différentes quantités sont séquencés en utilisant le panel « *HirisPlex* ». Nous obtenons des résultats fiables et reproductibles jusqu'à une quantité d'ADN d'environ 50 pg. En dessous de cette valeur, il est difficile de se prononcer pour une interprétation des résultats obtenus.

Au vu des résultats de reproductibilité et de sensibilité, le PGM est un outil indispensable pour l'établissement d'un portrait robot génétique.

En vue d'évaluer la possibilité de réaliser un portrait-robot-génétique et d'en connaître les limites, nous nous sommes intéressés à la prédiction de l'origine biogéographique et de 2 caractères morphologiques, la couleur des yeux et la couleur des cheveux.

Au niveau de l'origine biogéographique, nous avons fait le choix d'analyser préférentiellement l'ADN d'individus issus de métissage, afin de connaître les limites de prédiction. Il en ressort que dans la majorité des cas, l'origine biogéographique réelle concorde avec la prédiction établie. En effet, nous obtenons une concordance de 100% pour les origines strictement asiatiques et de l'ordre de 87% pour les origines strictement européennes. Néanmoins, ce résultat est à relativiser, vu le nombre relativement restreint d'échantillons concernés pour ces deux prédictions. Pour la prédiction de métissage, comprenant une origine européenne et une autre origine, l'origine européenne est retrouvée dans 75% des cas. Les autres origines issues des mélanges correspondent aux prédictions même s'il est difficile à partir d'un résultat de prédiction

de définir exactement l'origine des différents brassages génétiques. Par exemple, si une prédiction donne 15% d'origine asie du sud ouest, 35% d'origine africaine et 50% européenne, il est difficile d'établir que l'individu est en réalité originaire de l'île de la Réunion, même si cela correspond.

Un certain nombre de discordance peut être expliqué par le fait que de certains individus, ayant participé à l'étude, ne connaissent pas exactement leurs origines. C'est pourquoi, il est utile de poursuivre l'étude en incluant un plus grand nombre d'individus. Cependant, au vu des résultats obtenus, la prédiction de l'origine biogéographique est un outil très intéressant pour la réalisation d'un portrait robot génétique.

Afin d'affiner la qualité d'un portrait-robot génétique, il est nécessaire de coupler les prédictions de l'origine biogéographique avec la prédiction de caractères phénotypiques. Dans notre étude, nous nous sommes intéressés à la couleur des yeux et des cheveux, du fait de l'existence d'un modèle de prédiction, concernant ces 2 caractères phénotypiques, fiable et rodé.

Pour la couleur des yeux, le pourcentage de concordance entre les prédictions et la couleur réelle de chaque individu est assez élevé, à savoir de l'ordre de 87%. Néanmoins, ce résultat est à nuancer, en fonction de la couleur de l'œil à prédire. En effet, nous obtenons une concordance : de 100% pour les yeux bleus ; 93% pour les yeux marrons ; seulement de 69% pour les yeux intermédiaires. Ce dernier résultat peut être expliqué par le fait que tous les gènes impliqués dans la synthèse ou la régulation de la coloration dite « intermédiaire » de l'iris n'ont pas encore été identifiés. De plus, il est nécessaire à moyen terme, d'améliorer la qualité de distinction entre les différentes couleurs de l'iris.

La concordance entre la prédiction de la couleur des cheveux et la couleur de cheveux réelle de l'individu est assez élevée, de l'ordre de 80%. Toutefois, ces résultats restent inférieurs à ceux obtenus pour la prédiction de la couleur des yeux. Comme pour la prédiction des yeux, les résultats sont différents en fonction de la couleur de cheveux prédite. En effet le pourcentage de concordance entre la prédiction et la couleur de cheveux réelle est de 100% pour les cheveux roux, 79% pour les cheveux noirs, 78,5% pour les cheveux bruns et de 75% pour les cheveux blonds. Néanmoins il est à souligner que de nombreuses discordances sont minimales à savoir : une prédiction de coloration brun foncé alors que la couleur réelle est noire et inversement; une prédiction de couleur de couleur blond foncé alors que la coloration réelle est brun clair et inversement. Un seul échantillon, le numéro PGM00004 a un résultat totalement discordant, à savoir une prédiction de couleur de cheveux blond alors que l'individu a les cheveux de couleur noire.

Les discordances minimales observées peuvent être dues d'une part à l'environnement et d'autre part à la génétique. En effet, sous l'effet d'une exposition prolongé au soleil, la couleur des

cheveux a tendance à s'éclaircir. De plus, surtout chez les hommes, les cheveux avec l'âge, grisonnent et provoquent un assombrissement de la couleur des cheveux. Enfin, la couleur des cheveux peut changer entre l'enfance et l'adolescence de l'individu. Nous avons remarqué que dans certains cas de discordance, la prédiction de la couleur des cheveux correspondait à la couleur de cheveux qu'avait l'individu lors de son enfance. Ce changement de couleur serait dû à des variations hormonales régulées génétiquement. Des études sont menées par différents laboratoires en vue d'identifier les gènes responsables de ce changement.

L'établissement d'un portrait-robot génétique à partir des prédictions de l'origine biogéographique et des caractères phénotypiques tels que la couleur des yeux et des cheveux, peut être d'une aide précieuse dans l'identification à partir d'ossements ou de corps très dégradés. En effet, le portrait robot génétique peut compléter voire conforter, au sein de l'institut, les examens réalisés dans le service anthropologie pour la reconstitution faciale ou déterminer l'origine ethnique du cadavre.

Le portrait robot génétique peut également être utilisé dans le cadre d'orientation d'enquête lorsque les enquêteurs disposent d'un profil génétique exploitable mais non connu FNAEG. Cela peut permettre de discriminer certains types d'individus ou suspects. Néanmoins, il faudrait plus de caractères phénotypiques à prédire, comme la calvitie, l'âge, la taille, les traits du visage (distance entre le nez et la bouche). Les SNPs marqueurs de ces caractéristiques phénotypiques sont très étudiés par différents laboratoires. Au sein de la DCBG, nous travaillons actuellement sur les SNPs marqueurs de la calvitie et de la pigmentation de la peau.

Il est possible également d'utiliser le PGM pour séquencer les STRs ou l'ADN mitochondrial, avec une plus grande sensibilité que par électrophorèse capillaire. En effet, le séquençage à haut débit permet de mieux exploiter des ADN dégradés voire très dégradés. Cependant la mise en œuvre du PGM est complexe, (multiples étapes pour aboutir au séquençage et la durée de l'expérience) et nécessite donc une pratique régulière. De plus, il faut noter que via son mode de séquençage par détection de la variation du pH, il est difficile d'avoir un résultat fiable lorsqu'un SNP se trouve dans une séquence nucléotidique comprenant 5 à 6 mêmes nucléotides à la suite.

Au vu de notre étude, il est donc possible, à l'aide du PGM, d'établir un portrait robot génétique, certes sommaire au jour d'aujourd'hui, pour compléter les examens scientifiques d'autres services, ou apporter une orientation d'enquête aux unités de terrain.

BIBLIOGRAPHIE

- *Manuel de criminalistique moderne et de police scientifique*, Buquet (2006)
- *Utilisation des SNP pour l'identification humaine*, Keyser (2006)
- *Advantages and limitations of next generation sequencing technologies: A comparison of electrophoresis and non-electrophoresis methods*, Hert et al (2008)
- *Eye color and the prediction of complex phenotypes from genotypes*, Liu et al (2009)
- *IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information*, Walsh et al (2011)
- *The HIrisPlex system for simultaneous prediction of hair and eye colour from DN,, Walsh et al (2013)*
- *Evaluation of the Ion Torrent™ HID SNP 169-plex*, Borsting et al (2014)
- *Progress toward an efficient panel of SNPs for ancestry inference*, Kidd et al (2014)
- *Forensic DNA Phenotyping: Predicting human appearance from crime scene material for investigative purposes*, Kayser (2015)
- *Evaluation of the Precision ID Ancestry Panel for crime case work: A SNP typing assay developed for typing of 165 ancestral informative markers*, Pereira et al (2016)

ANNEXES



Emplacement du code barre

POLE JUDICIAIRE DE LA
GENDARMERIE NATIONALE

INSTITUT DE RECHERCHE CRIMINELLE
DE LA GENDARMERIE NATIONALE

CONSENTEMENT ECLAIRE

Je soussigné(e), nom, prénom, grade ou qualité :

..... Année de naissance :

Déclare être volontaire et accepte que soit effectué sur ma personne un prélèvement de cellules buccales pour contribuer à une étude scientifique portant sur l'analyse des variations ponctuelles de l'ADN génomique.

J'ai bien été informé(e) que l'objectif de cette étude consiste à analyser uniquement le polymorphisme de séquence de l'ADN de type SNP à des fins de prédiction exclusivement limitée aux seuls **caractères morphologiques apparents**, tels que la couleur des yeux, le type et la couleur des cheveux, la couleur de la peau, les traits morphologiques liés à l'origine bio-géographique, la prédisposition à la calvitie ou aux taches de rousseur.

Ces recherches seront pratiquées conformément à la législation en vigueur et notamment en vertu de la loi n°94-653 du 29 juillet 1994 et des articles 226-25 à 226-30 du Code Pénal et des articles 16-10 et 16-11 du Code Civil.

Je déclare avoir déjà fait l'objet d'un prélèvement de cellules buccales conformément au décret n° 2013-406 du 16 mai 2013 relatif à la mise en œuvre de traitements de données à caractère personnel dénommés « outils de recherche de contamination ADN » (ORCA) de l'IRCGN.

J'atteste avoir parfaitement compris la finalité de l'étude et avoir compris les explications qui m'ont été données de façon claire et intelligible.

Fait à Cergy-Pontoise, Le :	Signature de la personne prélevée :
--------------------------------	-------------------------------------

Annexe n°1

QUESTIONNAIRE PHENOTYPIQUE

Identifiant CB du prélèvement :

Sexe : M F

Âge : ans

Couleur naturelle des yeux : bleu clair / foncé intermédiaire gris / vert brun clair / foncé

Présence de spots : NON OUI , combien :



Présence d'anneaux : NON OUI petit / moyen / gros



Couleur naturelle des cheveux :



blond clair blond foncé roux brun clair brun foncé noir

Couleur des cheveux dans l'enfance : identique autre :

Cheveux grisonnants NON OUI, âge d'apparition :

Forme naturelle des cheveux : lisse ondulé bouclé



Présence de taches de rousseur : NON OUI (un peu / beaucoup)

Prédisposition à la calvitie :



Existe-t-il des antécédents de calvitie dans la famille : NON OUI : Mère / Père / les deux

Âge d'apparition de la calvitie :

Couleur de la peau : clair (type 1 - 2) intermédiaire (type 3 - 4) foncé (type 5 - 6)

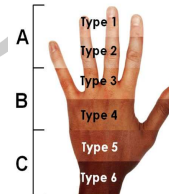
tendance à bronzer rapidement : NON OUI

Origine biogéographique :

Précision :

- Europe N S E O
- Afrique N C S E O
- Asie N C S E O
- Amérique N C S
- Océanie

- métissage : NON OUI :
- père :
 - mère :
 - grand-père :
 - grand-mère :



CB	origine biogéographique	Age	Sexe		yeux			cheveux						Enfance	
			Masculin	Féminin	bleu	inter	marron	blond clair	blond foncé	roux	brun	brun foncé	noir		grisonnant
PGM00001	Europe Sud	43		1		1						1			
PGM00002	Europe Ouest	39	1		1				1						
PGM00003	Europe Ouest	42		1	1						1				Blond foncé
PGM00004	Europe	36	1		1								1		
PGM00005	Europe Ouest	43		1		1					1				
PGM00006	Europe Ouest	43		1			1				1				
PGM00007	Europe - Martinique	39	1				1						1		
PGM00008	Europe Ouest	28		1			1					1			
PGM00009	Europe Ouest	31	1				1				1				
PGM00010	Europe-kazakhstan	56	1			1							1		
PGM00011	Inde-Afrique	35		1			1						1		
PGM00012	Europe Ouest	46	1			1					1				
PGM00013	Europe Ouest	36	1		1				1						
PGM00014	Europe Ouest	36	1		1			1							
PGM00015	Europe Nord	31	1		1				1						
PGM00016	Europe Ouest	33		1		1					1				
PGM00017	Europe Ouest	30		1	1						1				
PGM00018	Afrique Nord	34	1				1						1		
PGM00019	Europe Ouest	37	1				1				1				
PGM00020	Europe Ouest	32		1			1					1			
PGM00021	Europe Ouest	44	1			1						1			
PGM00022	Europe O et S	38	1				1						1		
PGM00023	Europe Ouest	29		1			1				1				
PGM00024	Europe Ouest	44	1		1							1			
PGM00025	Europe Sud	32		1			1					1			
PGM00026	Europe Ouest	41		1		1					1				
PGM00027	Europe E et S	31	1		1								1		Brun foncé
PGM00028	Europe Sud	55	1				1					1			
PGM00029	Europe Sud	29		1			1					1			
PGM00030	Europe Ouest	36	1				1					1			
PGM00031	Europe O et S	25		1	1							1			
PGM00032	Europe Sud	30		1			1					1			
PGM00033	Europe Ouest	32	1				1				1				

Annexe n°3 – partie 1

En jaune, les échantillons choisis pour l'étude de la prédiction des origines biogéographiques

CB	origine biogéographique	age	Sexe		yeux			cheveux							
			Masculin	Féminin	bleu	inter	marron	blond clair	blond foncé	roux	brun	brun foncé	noir	grisomant	Enfance
PGM000034	Europe ouest	50	1			1						1			
PGM000035	Europe Ouest	42	1		1						1				
PGM000036	Europe Ouest	43	1			1							1		
PGM000037	Europe Sud	36	1			1					1				
PGM000038	Europe O et S	23		1			1					1			
PGM000039	Europe Ouest	47	1				1					1			
PGM000040	Afrique	48	1				1						1		
PGM000041	Europe O	29	1				1						1		
PGM000042	Europe Ouest	36		1			1				1				
PGM000043	Afr N - Eur S	43	1				1					1			
PGM000044	Europe O Asie O	29	1			1						1			
PGM000045	Europe Ouest	29		1			1				1				
PGM000046	Ame C Eur O	36	1				1						1		
PGM000047	Europe O et E	29	1			1				1					
PGM000048	Europe O	44	1				1				1			1	
PGM000049	Europe O	39	1			1							1		
PGM000050	Europe O	31	1			1							1	1	
PGM000051	Europe O	29		1		1					1				
PGM000052	Europe E et O	36		1	1				1						
PGM000053	Europe E et O	40	1		1						1				
PGM000054	Europe S et O	34	1		1							1			
PGM000055	Eur O Asie SE	36	1				1						1		
PGM000056	Europe O	28	1				1				1				
PGM000057	Europe O	29	1				1					1			
PGM000058	Europe S et O	36	1			1			1						
PGM000059	Europe S	34		1			1			1					
PGM000060	Europe S	46	1		1V		1V					1		1	
PGM000061	Eur O Asie SE	38	1				1						1		
PGM000062	Europe O	45	1		1						1			1	
PGM000063	Europe O	39	1				1					1			
PGM000064	Europe O	54		1		1						1			
PGM000065	Europe Sud	33	1			1					1				
PGM000066	Europe S	45	1			1								1	

Annexe n°3 – partie2

V= yeux vairon

En jaune, les échantillons choisis pour l'étude de la prédiction des origines biogéographiques

CB	origine biogéographique	age	Sexe		yeux			cheveux								
			Masculin	Féminin	bleu	inter	marron	blond clair	blond foncé	roux	brun	brun foncé	noir	grisonnant	Enfance	
PGM000067	Afrique SE	27	1				1							1		
PGM000068	Europe S	25		1			1							1	1	
PGM000069	Europe S	33		1			1					1				
PGM000070	Europe E et O	29	1				1					1			1	
PGM000071	Eur O et Ile mau	33	1				1							1		
PGM000072	Europe O	47	1				1						1		1	
PGM000073	Asie S	46	1				1						1			
PGM000074	Lituanie	39		1	1				1							
PGM000075	Europe O	45	1		1							1				
PGM000076	Europe O	33	1				1		1							
PGM000077	Europe S	37	1				1						1			
PGM000078	Euro S et O	32		1			1		1							
PGM000079	Europe S	29	1				1						1			
PGM000080	Europe O	46	1				1					1				
PGM000081	Europe S	43	1				1						1			
PGM000082	Europe O	36	1				1									
PGM000083	Europe O	31	1				1			1				1		
PGM000084	Europe S	39	1				1									
PGM000085	indeterminé	42	1		1				1					1		
PGM000086	Eur O	38	1				1						1			
PGM000087	Europe O	41		1			1					1				
PGM000088	Eur E et O	43	1				1					1				
PGM000089	Asie S eur O	34	1				1							1		
PGM000090	Amér C Eur O	24		1			1							1		
PGM000091	Asie O Af SE	27	1				1						1			
PGM000092	Europe O	41	1				1					1				
PGM000093	Europe O	44	1				1					1			1	
PGM000094	Europe O	49	1		1				1							
PGM000095	Europe O	16	1				1							1		
PGM000096	Europe O	49	1				1			1						
PGM000097	Europe O	43		1			1			1						
PGM000098	Asie S	21	1				1							1		
PGM000099	Reunion	36	1				1							1		
PGM000100	Eur S	41	1				1							1		

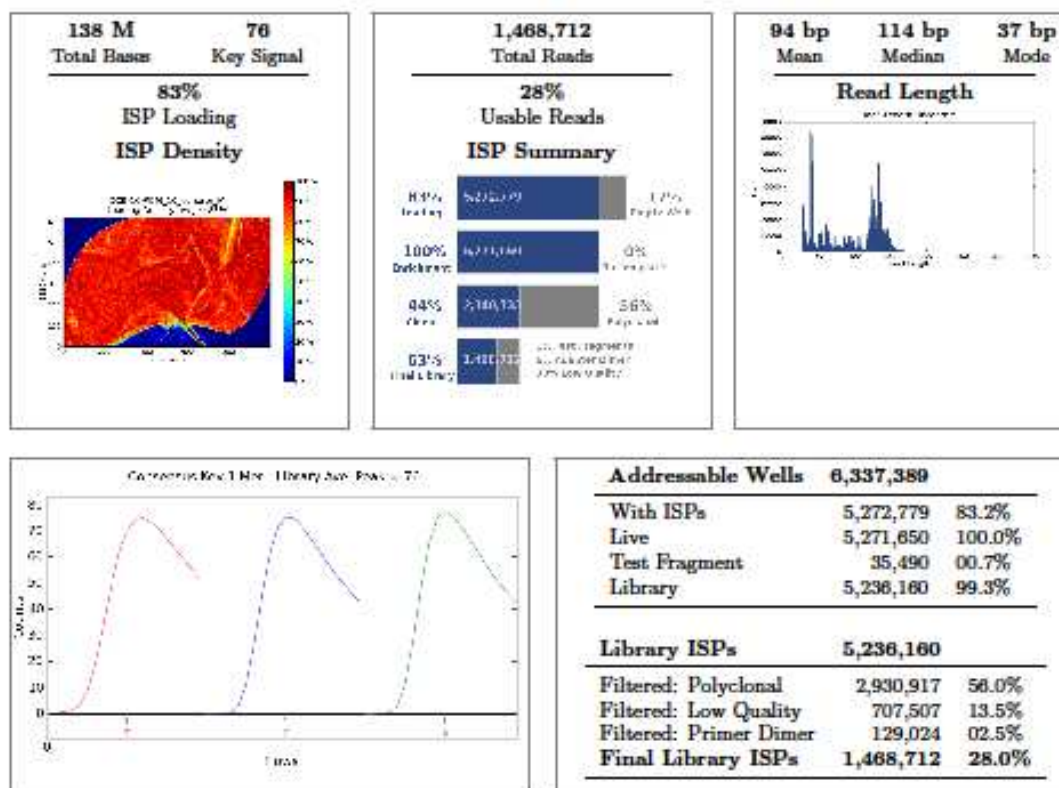
Annexe n°3 – Partie 3

En jaune, les échantillons choisis pour l'étude de la prédiction des origines biogéographiques

SNP	CHROMOSOME	GENE
<i>N29insA</i>	16	MC1R
<i>rs11547464</i>	16	MC1R
<i>rs885479</i>	16	MC1R
<i>rs1805008</i>	16	MC1R
<i>rs1805005</i>	16	MC1R
<i>rs1805006</i>	16	MC1R
<i>rs1805007</i>	16	MC1R
<i>rs1805009</i>	16	MC1R
<i>Y1520CH</i>	16	MC1R
<i>rs2228479</i>	16	MC1R
<i>rs1110400</i>	16	MC1R
<i>rs28777</i>	5	SLC45A2
<i>rs16891982</i>	5	SLC45A3
<i>rs12821256</i>	12	KITLG
<i>rs4959270</i>	6	EXOC2
<i>rs12203592</i>	6	IRF4
<i>rs1042602</i>	11	TYR
<i>rs1800407</i>	15	OAC2
<i>rs2402130</i>	14	SLC24A4
<i>rs12913832</i>	15	HERC2
<i>rs2378249</i>	20	ASIP/PIGU
<i>rs12896399</i>	14	SLC24A4
<i>rs1393350</i>	11	TYR
<i>rs683</i>	9	TYRP1

Annexe n°4 : Liste des 24 SNP du panel *HirisPlex*

Run Summary



Barcode Name	Sample	Bases	$\geq Q20$	Reads	Mean Read Length
No barcode	none	7,520,891	5,546,325	108,351	69 bp
IonXpress_001	M1	5,455,679	4,711,283	59,682	91 bp
IonXpress_002	M2	4,199,871	3,636,108	47,064	89 bp
IonXpress_003	M3	4,317,036	3,829,501	45,673	95 bp
IonXpress_004	M4	3,838,910	3,359,813	41,199	93 bp
IonXpress_005	M5	2,543,717	2,220,918	25,350	100 bp
IonXpress_006	M6	3,570,449	3,136,222	38,827	92 bp
IonXpress_007	M7	4,512,235	3,971,470	47,157	96 bp
IonXpress_008	M8	4,729,037	4,082,560	48,868	97 bp
IonXpress_009	M9	4,535,247	3,975,978	45,128	100 bp
IonXpress_010	M10	5,122,971	4,450,492	52,261	98 bp
IonXpress_011	M11	4,220,298	3,726,494	45,821	92 bp

