

# Discovery of metabolic gene mutations causing intellectual delay

Wyeth W. Wasserman

David Aubert

Ahmed Rafik

University of Montpellier

Bio-informatique

05 jan 2015

## 1 Etude du génome

---

### 1.1 Pourquoi cette étude

Un laboratoire a réussi à mettre en évidence le lien entre des maladies entraînant un retard intellectuelle et des mutation genetique. Ce qui a poussé le docteur Wasserman et son équipe a se penché sur le sujet. Il commence par décrire les différents symptome que l'on peut observé chez une fraterie de nouveau né comme par exemple, à la naissance, des difficultés respiratoires qui réapparaissent à 2 ans et demi ainsi qu'à 3 ans et demi. Rapidement, ils ont pu écarté de nombreuses thèses.

### 1.2 WGS VS WES

**Définitions simplifiées** The Whole genome sequencing : c'est un procédé qui permet de détermine l'ensemble des séquences d'ADN du génome d'un organisme donné. The Whole exome sequencing : c'est un procédé qui va selectionner l'ensemble des séquences d'ADN qui encodent des protéines, et ensuite va séquencer celles-ci.

Le procédé WGS est donc plus lourd a utilisé que WES ( WGS = WESx6 bp)

### 1.3 le gene CA5A

D'abord, nous devons définir ce qu'est l'anhydrase carbonique en général : C'est une enzyme présente à la surface plasmique intracellulaire des globules rouges qui permet transforme le gaz carbonique  $CO_2$  en  $H_2CO_3$ . le CA5A est donc un gène de la famille des Anhydrases Carboniques. Il encode l'anhydrase carbonique dans les cellules mitochondrial.

Malgré le rôle de CA-VA dans le métabolisme intermédiaire extérieur des crises mortelles, on peut observer des phenotypes leger (leger retard dans le developement mental ou dans la croissance par exemple) c'est peut s'expliquer éventuellement de la manière suivante : Il est possible qu'il y ai un chevauchement fonctionnel de la production de bicarbonate avec l'enzyme mitochondriale CA5B. CA-VA étant une enzyme néonatale, elle devient moins important avec l'âge.

## 1.4 L'approche de l'équipe

L'équipe a eu une approche très classique pour aborder ce problème qui leur donna les résultats suivants : un premier séquençage avec WES sur 67 familles révéla que 3 n'étaient pas génétiques tandis que 64 l'étaient. Parmi ceux-ci, ils ont pu identifier des mutations chez 52 familles et pour les 12 restantes, il fallut utiliser WGS.

## 2 Medical Subject Heading Over-representation Profiles

---

### 2.1 Définition

MeSHOP est une application très complète. Celle-ci permet de centraliser les recherches et d'indexer les publications scientifiques. Celle-ci englobe des thématiques allant de la recherche génétique à la chimie en passant par les maladies, ce qui intéresse l'auteur.

### 2.2 fonctionnement

Le moteur de cette application repose sur une base de données MEDLINE et le moteur de recherche PubMed.

#### 2.2.1 Annotations

Chaque article devra être annoté avec un vocabulaire particulier. En effet, c'est cette annotation qui permettra une recherche efficace. Le système MeSH a une liste de vocabulaire (*PMIDs*) qui servira d'index pour le fichier.

### 2.3 Editer un article

Après l'écriture d'un article, il est nécessaire de le publier. Et c'est à ce moment là que va servir le système d'annotation. En effet une lecture par un robot de l'article sera faite et permettra, en fonction du vocabulaire utilisé, de classer l'article dans des catégories et de l'indexer pour des recherches.

Chaque mot sera classé en fonction de son nombre d'occurrence dans les *PMIDs*

### 2.4 Recherche

Un point important est la thématique recherchée. En effet, comme vu précédemment, les sujets d'articles englobent un grand nombre de thématiques (génétique, maladies, chimie etc...) et donc le contexte peut changer pour des mêmes mots ou combinaisons de mots.

## 3 Site de liaison d'allèle spécifique

---

### 3.1 Définition

### 3.2 allèles préférées

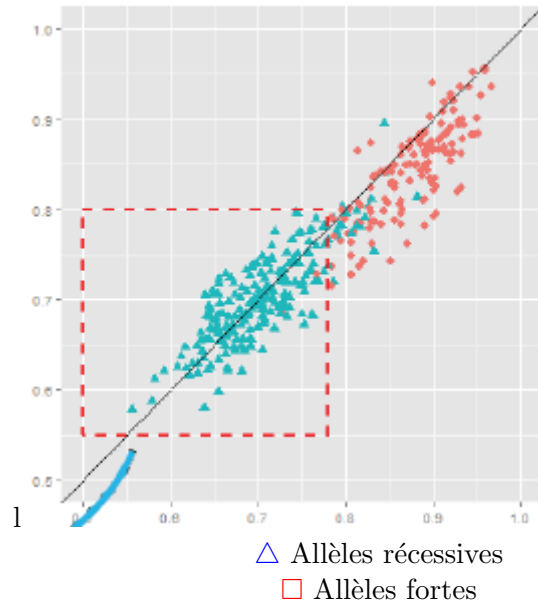
Mais comme les allèles fortes s'expriment plus que les allèles récessives cela peut être dû à plusieurs raisons mais la plus commune est que, par sélection naturelle, un gène subira une mutation, et aura un nombre supérieur de site de transcriptions que le précédent et ainsi pourra se transmettre plus facilement de génération en génération. De ce fait l'étude à de meilleurs chances de résultats si il est effectué sur des gènes forts.

Il s'agit d'une matrice où sont marquées les probabilités de chaque nucléotide d'apparaître dans le site de liaison pour la transcription, ainsi que sa position.

Mais ces travaux sont ralentis par le fait qu'il peut exister de multiples facteurs qui modifient le taux de transcription d'un site comme les mutations, l'hérédité et l'epigenetique.

### 3.3 Recherche de site de liaison pour la transcription pour les lymphomes

En effet, grâce aux outils présentés précédemment, beaucoup de données ont été recueillies,



**Figure 3.1:** *Allèle score*

notamment des échantillons d'ADN, d'ARN de malades atteints de lymphomes.

Ensuite certaines zones du génomes ont été ciblés, et une des premières remarques est que les sites de transcriptions ont eu des taux de mutations plus élevé comparés aux séquences saines.

La mort des cellules, à un rôle dans ce fonctionnement. En effet, la mort des cellules est régulé d'une manière naturel et saine mais si cette régulation est perturbée cela peut causer des problèmes, dans le cas d'une diminution et dans le cas d'une augmentation un lymphome, par exemple. Une augmentation des cellules (ou plutôt une non décroissance normale du nombre de cellules) augmentent le nombre de site de liaison augmente.