

Discovery of metabolic gene mutations causing intellectual delay

Wyeth W. Wasserman

David Aubert

Ahmed Rafik

University of Montpellier

Bio-informatique

05 jan 2015

Aujourd'hui il existe quelques domaines qui occupent une place importante du monde de la recherche, à cause de leurs importances et de leurs complexités et qui font régulièrement appel aux centres de calculs haute performances. L'étude du génome en fait partie.

Aujourd'hui beaucoup des maladies difficilement curables sont dûes à des problèmes de mutations et de génétique (Trisomie, cancer etc...). Ces problèmes trouvent leurs sources dans l'ADN et/ou son expression.

A l'heure actuelle, le séquençage du génome pose des problèmes à la recherche et ralenti son avancée.

Le Dr. Wasserman est à la tête d'un groupe travaillant sur la partie calcul et analyse du génome. Il travaille sur des nouvelles méthodes de calculs et des outils pour identifier et implémenter des bases de données pour améliorer la recherche. Le Dr. Wasserman essaie d'avancer la compréhension sur les facteurs de transcriptions qui jouent un rôle important dans l'expression du génome.

Il sera présenté dans cet article l'outil MeSH, qui est une base de données spécialisée dans le domaine de la recherche médical, ensuite la détection des transcriptions si importantes dans le fonctionnement. Ensuite une fois ces transcriptions détectées comment elles sont utilisés, et comment elles s'expriment mais dans un premier temps l'intérêt de l'étude du génome, la portée ainsi que les méthodes utilisées.

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 déterminer l'ensemble des séquences d'ADN du génome d'un organisme donné. The Whole exome sequencing : c'est un procédé qui va sélectionner l'ensemble des séquences d'ADN qui encodent des protéines, et ensuite va séquencer celles-ci.

Le procédé WGS est donc plus lourd à utiliser 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 transformer 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'était pas génétique tandis que 64 l'était. Parmi ceux-ci, ils ont pu identifier des mutations chez 52 familles et pour les 12 restantes, il fallu 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

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

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

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

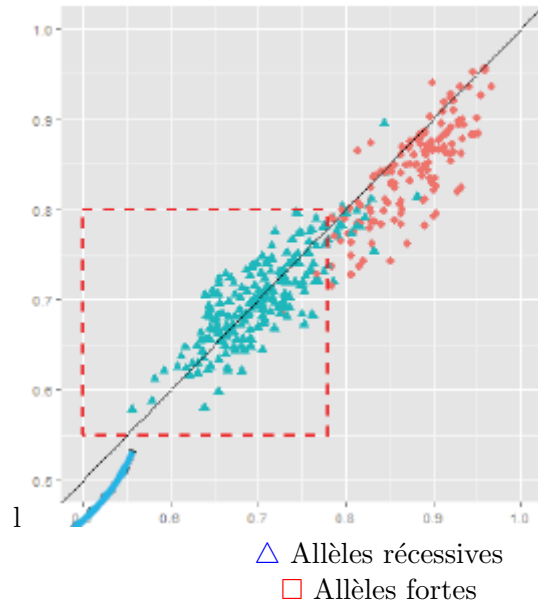


Figure 3.1: *Allèle score*

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 meilleures chances de résultats si il est effectué sur des gènes forts.

Pour trouver les places de liaisons de transcription, une unité a été créée: **PWM**: La matrice de position de poids.

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.

Ceci a une importance particulière car grâce à ces matrices, il a été possible de remarquer que les liaisons se faisaient plus sur les allèles récessives que les autres.

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'épigénétique.

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

C'est ici que la présentation recoupe avec les travaux du Dr. Wasserman.

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

lules 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.