**UE5** – Génétique médicale

Professeur Debette

Mardi **04/12/2012** de 8h30 à 10h30

Ronéotypeuse : Anne-Lorraine Choné

Ronéolecteur : Guillaume

**COURS 10 : GENETIQUE DES MALADIES MULTIFACTORIELLES**

*Ce cours fut synthétique car la prof est arrivée à 9h20 et a décidé d’aller vite pour finir son cours. Elle a précisé que ce n’est pas un cours de statistique et qu’elle ne posera aucune question sur les diapos qu’elle a zappé. L’essentiel selon elle est de comprendre les grands principes d’une recherche d’association génomique. Par ailleurs vous y verrez des répétitions avec le cours de master de santé publique dispensé par la même prof.*

**INTRODUCTION :**

1. **MALADIES COMPLEXES/MULTIFACTORIELLES VS MENDELIENNES :**
2. **ETUDES D’ASSOCIATION GENETIQUE VS ANALYSE DE LIAISON :**
3. **QUELQUES RAPPELS :**
4. **Les différents types de variations génétiques :**
5. **Equilibre de Hardy-Weinberg :**
6. **Déséquilibre de liaison :**
7. **ETUDES D’ASSOCIATION GENETIQUE (EAG) :**
8. **Etudes d’association sur « gène candidats » :**
9. **Etudes d’association génétique pangénomique :**
10. **ANANLYSE ET INTERPRETATION DES EAG :**
11. **Tests multiples :**
12. **Hétérogénéité de population :**
13. **Réplication :**
14. **CARACTERISATION DES SIGNAUX IDENTIFIES, PERSPECTIVES :**

**INTRODUCTION**:

**L’Epidémiologie génétique** est une branche de l’épidémiologie qui étudie le rôle des facteurs de risque génétiques et leurs interactions avec les facteurs de risques environnementaux sur la survenue de maladies.

On s’intéresse à la part génétique de maladies fréquentes multifactorielles pour 3 objectifs :

1. **Mieux comprendre la physiopathologie de ces maladies.**

Quand on identifie un facteur de risque génétique, c’est à dire un gène associé à une maladie, on peut observer une cascade physiopathologique. On identifie une protéine impliquée dans la survenue de la maladie.

* Ca nous permet de la tester comme **nouvelle cible thérapeutique**.
* Cette **protéine** peut servir comme **biomarqueur de la maladie**. La doser régulièrement permettra de monitorer la maladie dans son évolution (Exemple : marqueurs biologiques de suivi d’un cancer après chimio)
* En comprenant mieux l’interaction des FDR biologiques et leur interaction avec l’environnement, on peut aussi identifier de **nouvelles approches préventives**.

Exemples : La maladie de Crohn a permis de découvrir des gènes de susceptibilité qui ont mis en évidence le phénomène d’autophagie, avec le gène NOD2, fortement associé à la maladie, qui code pour un détecteur intra cellulaire de certaines composantes bactériennes, et le gène ATG16L1 qui dirige les microbes vers les lysosomes. *(Les noms des gènes de sont pas à retenir).* Ces découvertes ont lancé une voie de recherche thérapeutique de la maladie de Crohn qui vise à cibler le microbiome intestinal, avec des antibiotiques, des probiotiques, des prébiotiques. Mais pas d’applications concrètes pour l’instant.

-La polyarthrite rhumatoïde : L’interaction entre le FDR tabac et la présence de certains marqueurs génétiques (HLA et PTPN22), multiplie par 20 le risque de développer une polyarthrite rhumatoïde.

-Les Amish, qui vivent en communauté aux US, constituent un isolement génétique car ils se marient entre eux, donc les études sur leur génome sont intéressantes. Ils ont été étudiés pour la prédisposition à l’obésité, et l’étude a révélé que le gène FTO (normalement très associé à l’obésité), s’avère être protecteur contre l’obésité lorsqu’il est combiné à une activité physique.

1. **Mieux prédire le risque pour mieux cibler la prévention envers une population à risque :**

On espère pouvoir mieux prédire le risque des maladies en ajoutant la connaissance du terrain génétique.

Exemple : L’infarctus du myocarde est associé à une prédisposition génétique additionnée au fait que les gens fument. On voudrait donc augmenter la prévention tabagique chez les patients génétiquement prédisposés à l’infarctus (Cependant, il faudrait faire un screening ADN systématique chez chaque patient pour savoir s’il possède le gène de prédisposition… mais ce serait trop couteux !)

Jusqu'à maintenant, **les résultats sont décevants** car :

* Les variants génétiques étudiés sont intéressants mais souvent associés à une **augmentation très faible du risque** de la maladie (la présence du gène multiplie par moins de 1.5 le risque de développer la maladie).
* Et dans les rares fois ou le risque est vraiment accru, par exemple concernant le gène Epsilon L de l’apoprotéine E, associé a la maladie d’Alzheimer (risque 2 ou 3 fois supérieur à celui de la population générale), on n’aboutit **pas toujours** sur **une solution préventive possible**…

Donc, encore une fois, pas d’application concrète au sujet de la prévention des facteurs de risque. Mais, en combinant en un score génétique tous les facteurs de risques identifiés dans une maladie (une 12aine de gènes impliqués dans Alzheimer par exemple), il serait possible de mieux prédire le risque, et ainsi mieux cibler les populations par la prévention.

1. **Mieux utiliser la pharmacogénétique, choisir un traitement personnalisé en fonction du profil génétique d’un patient donné :** (Savoir, en connaissant l’ADN de tel patient, s’il sera meilleur répondeur ou moins bon répondeur à tel traitement).

Ici encore, **pas d’application officielle concrète**, mais voici un exemple parlant : les personnes atteintes d’un risque coronarien aigu bénéficient d’une pose de stent et sont ensuite traitées par antiagrégant plaquettaire (Clopidogrel). Il se trouve que les patients répondent plus ou moins bien au Clopidogrel en raison d’un variant génétique du gène CYP 2C19 qui code pour une enzyme bio-activatrice du Clopidogrel : Les porteurs homozygotes pour l’allèle 2 métabolisent mal le Clopidogrel, donc il est préférable de donner à ces patients là un autre traitement.

Il n’y a **pas encore** eu **d’études thérapeutiques randomisées** qui prouvent que prendre en compte le génotype soit plus efficace que de l’ignorer pour choisir une approche thérapeutique. Mais la prof a beaucoup d’espoir qu’on y arrive dans quelques années.

1. **MALADIES COMPLEXES/MULTIFACTORIELLES VS MENDELIENNES :**

RAPPEL : Les **maladies multifactorielles** (polygéniques) sont **opposées aux** maladies à mode de transmission mendélien (**monogéniques**). Dans ce cours, seules les maladies complexes/polygéniques/multifactorielles nous intéressent. Elles ont de multiples facteurs de susceptibilité génétiques et de multiples FDR environnementaux. Les arbres généalogiques sont donc inutiles mais **on peut mesurer la contribution génétique** en calculant:

* **Le risque de récurrence** : proportion de personnes malades apparentées à un cas malade versus proportion d’individus malades dans la population générales. (Evalue le risque de présenter la maladie quand on a un apparenté malade)
* **L’héritabilité** : proportion de variance du phénotype due à des effets génétiques.

*(Désolée, la prof n’a pas été plus claire à ce sujet…)*

1. **ETUDES D’ASSOCIATION GENETIQUE VS ANALYSE DE LIAISON :**

**L’analyse de liaison** examine la co-transmission d’une génération à l’autre du phénotype et des allèles d’un marqueur génétique. Elle s’utilise **surtout dans les maladies mendéliennes**.

**Dans les maladies multifactorielles**, on utilise plutôt **les études d’association génétique** qui consistent à comparer la fréquence de variants génétiques entre un groupe de cas et un groupe de témoins. (C’est un test statistique, même principe que les études cas – témoins.)

Exemple :

A A T C G **A/C** T C T

A A T C G **A/C** T C T

1000 patients :

CC n=50

AC n=250

AA n=700

1000 témoins :

CC n=10

AC n=180

AA n=810

On compare la fréquence du génotype porteur de l’allèle rare (ici c’est C) chez les cas et chez les témoins. On dit que **l’allèle est associé** à un phénotype si la fréquence de l’allèle diffère plus entre les cas et les témoins que ne le voudrait le hasard.

**Attention**, quand on démontre une **association** entre un variant génétique et un phénotype, cela **n’implique pas** un lien de **causalité** ! Car il peut s’agir d’une région génomique qui co-ségrégue avec le gène de la maladie mais qui n’en est pas la cause.

Quelques chiffres donnés par la prof pour se faire un ordre d’idées :

* Environ **3 milliards de nucléotides** dans la séquence de l’ADN humain.
* Environ **20 – 25 mille gènes**, de tailles variables : entre 100 et > 2 millions de nucléotides, constituant 1.5% du génome, le reste étant non codant ou régulateur.
* **99.9%** de la séquence d’ADN est **identique d’un individu à l’autre**, les 0.01% restants font la différence.

1. **QUELQUES RAPPELS :**
2. **Les différents types de variations génétiques :**

* Les **SNP** sont les plus étudiés à l’heure actuelle : c’est un polymorphisme mononucléotidique : au niveau d’un locus du génome, il y a 2 nucléotides différents possibles d’un individu à l’autre. Il y a des millions de SNP dans notre génome. On parle de **polymorphisme** si la fréquence de ce SNP est **> 1%** dans la population, et on parle de **variant ou de mutation** si la fréquence est **< 1%.**

Conséquences d’un SNP et signification fonctionnelle :

* **Neutre** si il est dans une séquence non codante, non régulatrice, ou si dans un exon il ne modifie pas l’acide aminé correspondant au triplet (acide aminé synonyme).
* **Modification du tau d’expression du gène** si le variant se trouve dans une séquence régulatrice (promoteur...)
* **Modification de la composition de la protéine** par remplacement d’un acide aminé par un autre (variants missense), par remplacement d’un acide aminé par un codon stop (variants nonsense), ou par épissage alternatif si le variant était dans un site d’épissage.
* Les **CNV**, qui sont des variations du nombre de copies d’une séquence d’ADN, comprenant 10 000 à 5 000 000 de bases. Les CNV sont présents en nombre moindre ou en excès, et sont situés en intergénique aussi bien qu’en intragénique.
* Les **polymorphismes de répétitions** sont des répétitions de séquences en tandem. (tels que les microsatellites) On ne s’y intéresse pas dans ce cours.

1. **Equilibre de Hardy-Weinberg :**

*(Cette notion a déjà été étudiée dans le cours 3 donc la prof a squizé cette diapo. Je vous en parle quand même.)*

L’équilibre de Hardy-Weinberg correspond à un équilibre de la fréquence allélique. C’est l’hypothèse selon laquelle dans une population donnée où **p est la fréquence de l’allèle 1** et **q est la fréquence de l’allèle 2**:

* La proportion des homozygotes pour l’allèle 1 est p2 ; la proportion des hétérozygotes est 2pq, et la proportion des homozygotes pour l’allèle 2 est q2.
* Donc la **fréquence génotypique** de ces allèles : **p2 + 2pq + q2 = 1**

Lors des études d’association génétique, on vérifie toujours dans le groupe des témoins que l’équilibre d’Hardy-Weinberg est respecté, et s’il ne l’est pas on suspecte une erreur technique dans l’étude.

1. **Déséquilibre de liaison :**

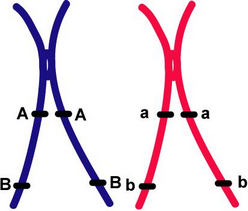
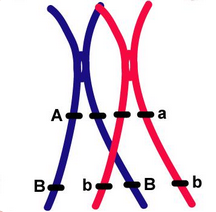
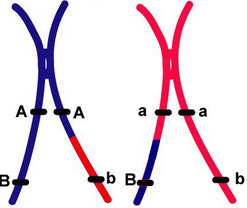
RAPPEL : Lorsqu’on est devant 2 variants génétiques (A ou a, et B ou b) situés sur le même chromosome, on sait qu’il y a 4 haplotypes possibles : AB, Ab, aB, ab.

* Si les 2 variants sont indépendants/en équilibre de liaison, la fréquence d’un haplotype sera égale a la probabilité du premier allèle multipliée par la probabilité du 2eme allèle.

Ex : P(Ab) = P(A) x P(b)

* Si 2 variants ne sont pas indépendants, cela signifie que la fréquence d’un haplotype ne dépend pas seulement de la probabilité de chaque allèle ; mais elle dépend aussi du lien de dépendance/ déséquilibre de liaison/ tau de recombinaison entre les 2.

Ce **déséquilibre de liaison** est **lié au phénomène de recombinaison** des chromosomes homologues au moment de la méiose : des échanges de segments s’opèrent entre les 2 chromosomes homologues. Il y a plein de recombinaisons lors d’une méiose, c’est une vraie loterie.

 =>  => 

* Si les variants **A et B** sont **très proches** géographiquement, alors il y a de fortes chances pour qu’ils restent ensemble malgré la recombinaison. Ils sont donc **dépendants**.
* Mais si **A et B** sont **très éloignés** (comme sur le schéma ci-dessus), la probabilité qu’ils soient séparés au moment de la recombinaison est plus élevée, donc ils sont considérés comme **indépendants**.
* En réalité, la dépendance n’est pas seulement liée à la distance entres 2 gènes sur le chromosome, mais aussi à des causes structurales qui font que certains points sont plus propices à la segmentation au moment de la recombinaison. On les appelle des **hotspots**. On a donc parfois 2 variants qui sont très proches mais néanmoins indépendants car séparés par un hotspot.

Aujourd’hui, avec la connaissance du génome humain, on peut à peu près prédire, avec un logiciel, que tel variant et tel variant vont toujours se recombiner ensemble ; dans ce cas on dit que le déséquilibre de liaison est complet. En cas d’indépendance on dit qu’il n’y a pas de déséquilibre de liaison. Entre les deux, le déséquilibre de liaison est chiffré.

1. **ETUDES D’ASSOCIATION GENETIQUE (EAG) :**

Il y a 2 types d’études d’association génétique :

1. **Etudes d’association sur « gène candidat » :**

On teste l’association d’une maladie avec un polymorphisme candidat, c'est-à-dire dont on pense (arbitrairement) que ce soit une mutation potentiellement associée à la maladie. C’est l’approche qu’on utilisait dans les années 90 – 2000. Il se trouve qu’elle est **très biaisée** car :

* **Si l’hypothèse de départ est fausse**, on n’a aucune chance de démontrer quoi que ce soit.
* Et si par hasard on avait trouvé une association, il était **impossible de la reproduire** sur un autre échantillon de sujets, donc elle n’était pas validable. Il y a donc eu beaucoup d’études lancées au hasard pendants les débuts de la génétique, mais un très petit nombre d’entre elles se sont avérées utiles et ont pu être publiées.

Ce que les généticiens reprochent aujourd’hui aux études d’association sur gène candidat, c’est qu’elles n’ont pas été conçues pour être reproduites sur un autre groupe indépendant. Or c’est aujourd’hui un pré-requis indispensable si on veut vérifier une différence significative, sinon on est « envahis » de faux positifs, d’après la prof.

Exemples : - L’association entre l’inactivation du gène LRP1 et l’anévrisme de l’aorte, constatée chez la souris avant d’être vérifiée chez l’homme, était une approche raisonnable et plausible.

-Constater que la maladie d’Alzheimer est associée très fortement avec l’AVC, car les patients qui sont atteints de l’un des 2 ont de fortes chance d’avoir un jour le 2 ème, et se dire «Tiens, on va tester les variants génétiques qu’on connait déjà pour la maladie d’Alzheimer dans les AVC pour voir si c’est vraiment lié.» était encore un raisonnement qui tenait plutôt debout.

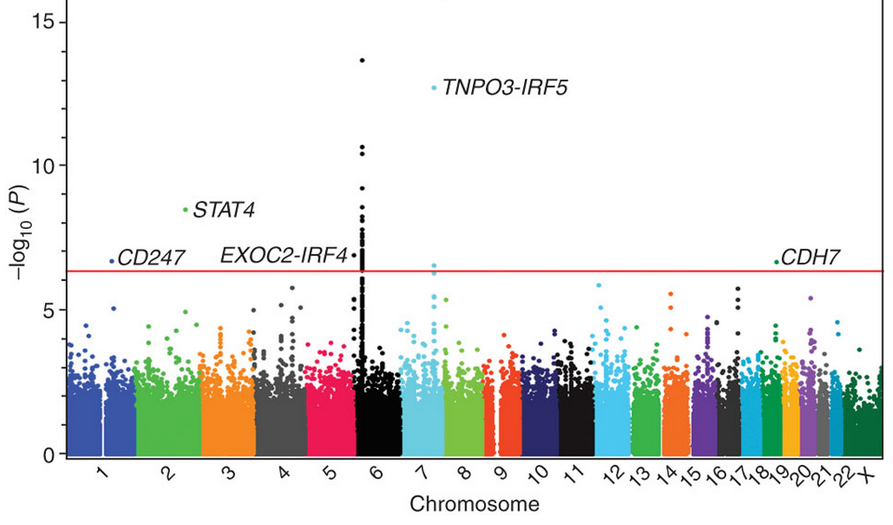
-Mais remarquer que les AVC sont dus à une thrombose dans les artères ; en déduire qu’il faut tester un gène (au pif parmi les N variants de chacun des X gènes de la coagulation et de l’hémostase) pour peut être trouver la mutation en cause de l’AVC ; c’était une approche fumeuse et très incomplète. Il faut donc bien choisir son gène candidat en s’appuyant sur des arguments solides dans ce genre d’étude, et en essayant de prendre 2 variants indépendants pour avoir à faire le moins de tests possible par économie.

1. **Etudes d’association génétique pangénomique :** (GWAS en anglais : genome wild association study)

Cette fois-ci on part **sans hypothèse préalable**. Le **GWAS** consiste à séquencer tout le génome d’une personne. On parle de screening, de génotypage à haut débit. Les variants génétiques sont au nombre de 5 000 000 dans l’ADN humain, et on va étudier les associations entre chacun d’entre eux individuellement et la maladie qui nous intéresse. Ce qui correspond à 5 000 000 de tests d’association génétique.

Cette technique n’est possible que depuis l’avènement révolutionnaire des technologies à haut débit (depuis 2007), qui permettent de séquencer le génome de 1000 personnes en une semaine, et aussi grâce à un projet international, un consortium, qui s’appelle **Hap Map**, qui catalogue les variations génétiques existant dans différentes populations ethniques. C’est un projet collaboratif à très grande échelle.

Le **génotypage** se fait **sur des micropuces**, après hybridation de l’ADN marqué, puis il y a une détection de fixation  « séquence-spécifique », puis l’ADN est séquencé, mis en parallèle avec celui de centaines ou milliers d’autres personnes, et cela permet d’exploiter une interprétation informatisée. Cela demeure cependant une technique très couteuse.

**PRESENTATION DES RESULTATS**

*(Schéma trouvé sur internet car mes photos prises en cours étaient mauvaises, mais si vous comparez avec les diapos, vous verrez que c’est très ressemblant.)*

Il s’agit ici d’une recherche d’association génétique avec les AVC sur des dizaines de milliers de personnes. Chaque petit point du graphique correspond à une valeur p d’association entre un variant génétique et les AVC, (l’étude concernait 2 500 000 variants, donc en réalité il y a 2 500 000 points sur le graphique).

En ordonnées : Les valeurs de p en -log10. Donc par exemple la valeur 8 correspond à p = 10-8.

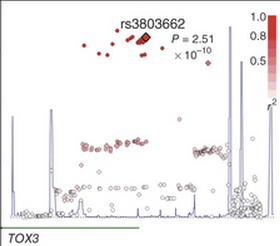
En abscisses : La position du variant sur les chromosomes.

NB : **p est le seuil de significativité**, dont nous parlions déjà en P1 (*en P1 on choisissait souvent comme seuil p = 0.05*). Plus on fait de tests, plus on est exigeant et plus p doit être petit. **En génétique**, c’est à partir du seuil **p = 5 x 10-8** que l’étude est significative. Ce seuil est matérialisé par la ligne horizontale qui coupe en 2 le graphique, donc seuls les variants correspondant aux points situés au dessus de la ligne remplissent ce critère.

**INTERPRETATION :**

On conclut de ce graphe que les valeurs de p les plus faibles (jusqu’à p = 10-14) sont au niveau du chromosome 6. Donc il y a une association significative entre un variant génétique dans le chromosome 6 et les AVC.

Si on zoome ce type de graphique, (mêmes échelles d’ordonnées et d’abscisses), on voit une ligne continue qui traduit le tau de recombinaison de l’ADN. Plus le tau de recombinaison est élevé, moins les variants de part et d’autre d’un pic sont susceptibles d’être en équilibre de liaison, et vice versa. (Autrement dit, les pics qui apparaissent correspondent aux hotspots de recombinaison : 2 points de part et d’autre d’un pic correspondent à 2 gènes indépendants sur un même chromosome, alors que 2 points qui sont proches et qui ne sont pas séparés par un pic sont 2 gènes dépendants.)



Les GWAS nécessitent de **très grands effectifs** (centaines de milliers de patients) parce que :

* Il est nécessaire de corriger les résultats au fil des tests multiples et d’**affiner cette valeur p**, qui doit être très stricte. Pour certaines maladies (Ex : l’obésité), on a des centaines de milliers de patients sous la main donc on utilise facilement le GWAS, mais pour d’autres pathologies, il est difficile de trouver 100 000 patients (Ex : Alzheimer).
* L**’effectif** doit être **d’autant plus important que les variants** qu’on étudie **sont rares**.
* L’**effectif** doit être **d’autant plus grand que le risque relatif** qu’on observe **est faible** (faible = 1.2 à 1.5 fois le risque de la population générale).
* Donc obligation de travailler ensemble à l’**échelle mondiale**, dans des consortia internationaux, d’autant que le matériel utilisé est couteux ce qui implique que la mise en commun des données est précieuse.

**Bilan**: Avantages et inconvénients des études d’associations sur gène candidat et des GWAS :

|  |  |  |
| --- | --- | --- |
|  | **GENES CANDIDATS** | **GWAS** |
| **AVANTAGES :** | * - Coute moins cher * - Nécessite effectifs moindres | * - Permet de découvrir de nouveaux gènes car approche agnosique (100 gènes découverts) * - Couvre mieux variation génétique |
| **INCONVENIENTS :** | * - Ne permet pas de découvrir de nouveaux gène ; on travaille seulement avec les gènes suspectés * - Résultats très décevants en moyenne | * - Nécessite très grands effectifs (collaboration...) * - Cout élevé * - Nécessite infrastructures adéquates |

1. **ANANLYSE ET INTERPRETATION DES EAG :**
2. **Tests multiples :**

Comme on étudie un très grand nombre de variants génétiques **on est obligé de corriger p** en fonction du nombre de tests effectués. Parfois on étudie même plusieurs maladies à la fois (Exemple : études des variants des triglicérides, du HDL et du LDL dans le cadre d’une recherche sur les composantes génétiques du cholestérol), et donc il y a des corrections sur le nombre de phénotypes étudiés.

Pour un seuil p = 0.05, cela signifie qu’on a 5 résultats sur 100 qui sont positifs par simple hasard, ce qui est considérable. On a donc besoin de **plus de tests pour améliorer la significativité et la puissance du résultat**.

*Ensuite, la prof a sauté une diapo intitulée ‘Comment interpréter un test statistique ?’ Ce sont les bases statistiques (différences entre sur lesquels la prof ne revient pas parce que ce n’est pas l’objet de son cours. Si ca vous amuse de replonger dans ces notions, vous pouvez toujours vous référer aux diapos… ou à vos cours de P1 !)*

Par convention, en génétique, le **seuil de significativité** se calcule ainsi :

**0,05 / nombre de tests effectuée**

Le nombre de **tests significatifs** sera le nombre de tests **pour lesquels**

**p expérimental < seuil de significativité calculé.**

Exemple : On effectue 5 tests, donc notre seuil de significativité sera

p = 0.05 / 5

= 0.01

Or les p expérimentaux obtenus après chacun des 5 tests sont les suivants :

p = 0.02, p = 0.049, p = 0.035, p = 0.009 et p = 0.18

Seul 0.009 < 0.01 donc seul le test pour lequel p = 0.09 est significatif et peut être exploité.

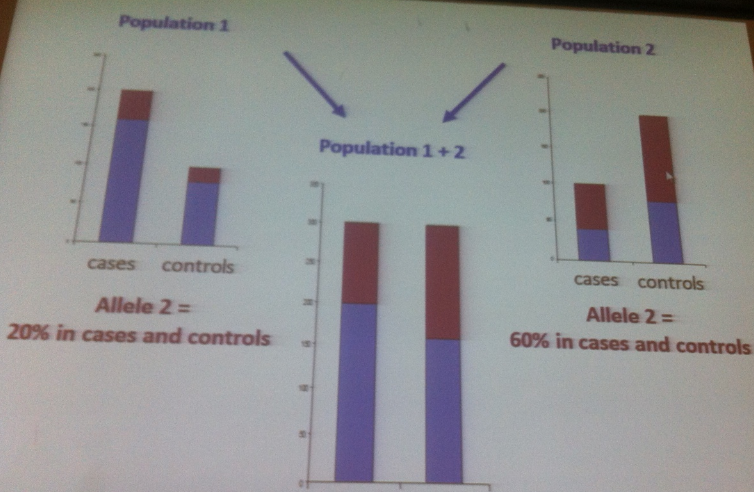
Mais en pratique, comme dit précédemment, la plupart du temps on prend plutôt **p =** **5 x 10-8** comme seuil, car cela exige des résultats encore plus précis. C’est un seuil admis de façon internationale.

1. **Hétérogénéité de population :**

**Les fréquences alléliques varient d’une région à l’autre**. Donc si on a :

- Une population 1 avec 2 fois plus de cas que de témoins, et pour laquelle la fréquence de l’allèle étudié est de 20% aussi bien chez les cas que chez les témoins.

- Une population 2 avec 2 fois plus de témoins que de cas, et pour laquelle la fréquence de l’allèle étudié est de 60% chez les cas et chez les témoins (car origine ethnique différente de la population 1).



Si on mélange les cas de la population 1 et 2, et si on mélange les témoins de la population 1 et 2, on aura l’impression que l’allèle est moins fréquent chez les cas que chez les témoins (33% vs 45% si on effectue le calcul) ; alors que ce n’est pas le cas, ni dans la population 1, ni dans la population 2 ! L’étude est biaisée.

**Il faut donc faire nos études sur des populations qui ont la même origine ethnique**, et il existe même des méthodes pour corriger par stratification les potentiels mélanges ethniques qui auraient pu être faits par mégarde dans un test. *(La prof n’entre pas dans les détails et précise que ce n’est pas important ; que la principale chose à retenir est qu’il faut juste faire attention aux origines ethniques quand on fait des tests génétiques.)*

D’autre part, ces mêmes méthodes permettent de situer géographiquement et ethniquement un individu grâce à son génotype, en comparant son profil génétique avec ce qu’on connait du profil génétique des différentes populations sur terre.

1. **Réplication :**

La réplication génétique est essentielle, c'est-à-dire qu’il faut absolument que toute association soit confirmée dans une étude indépendante, idéalement menée par d’autres investigateurs.

1. **CARACTERISATION DES SIGNAUX IDENTIFIES, PERSPECTIVES :**

- Quand on trouve une **association** entre un variant et une maladie, ca **ne veut pas dire** qu’on a trouvé le variant **causal**. Parfois on tombe directement sur la mutation en cause, mais parfois on trouve un variant qui est simplement en déséquilibre de liaison avec le variant causal. Pour essayer de trouver le variant causal, (étape après le GWAS) **il faut** :

* soit **génotyper** plus finement **plus de marqueurs** dans la région
* **ou** carrément **séquencer la région**.

-Quand on trouve une association avec un variant codant dans un gène, il est quand même très probable que ce soit le gène en question qui soit impliqué dans la maladie.

-Par contre, quand on a un variant génétique qui est dans une région régulatrice proche d’un gène, ca ne veut pas forcément dire que c’est ce gène le plus proche qui est associé à la maladie. Car des variants génétique peuvent moduler l’expression d’un gène qui est à distance, parfois même sur un chromosome. Dans un tel cas, on a du mal à trouver le gène impliqué.

-**Le GWAS est difficilement exploitable** pour des pathologies qui impliquent de très nombreux gènes (Exemple : 38 gènes pour le diabète de type 2), ou **pour des pathologies hétérogènes**, (Exemple : pour les AVC, il s’est avéré qu’il faut séparer les AVC d’origine cardiologique, athéromateuse…) d’où l’importance de la description du phénotype avant de commencer une étude.

- **Le GWAS n’explique que 10 à 20% de l’héritabilité** d’une maladie. Voila où on en est aujourd’hui.

- En post-GWAS, il existe aussi des puces qui ont été conçues pour rechercher les variants rares chez un patient sans avoir à faire un séquençage pangénomique.

- Les perspectives sont les suivantes : s’intéresser aussi aux CNV, aux modifications épigénétiques, à l’ADN mitochondrial, etc… qui sont mis de coté pour l’instant, mais qui feront un jour surement l’objet de recherches intéressantes.

**FIN**