

UE11 – PARCOURS – Génétique - n° 2 20/02/2019 Alexandre Alcais alexandre.alcais@inserm.fr	RT : Joséphine Perreau & Jeanne-Eglantine Oms RL : Sixtine Olivier
---	--

## Introduction au Parcours Génétique

Plan:

- I. Introduction et conseils pour l'examen**
- II. Les quatre questions de la génétique**
  - A. Quand/Pourquoi la génétique ?**
  - B. Quelle question poser à la génétique ? / A quelle question répond la génétique ?**
    - 1. Deux phénotypes extrêmes
      - i. La BCGite disséminée
      - ii. La tuberculose pulmonaire
    - 2. Quel type de génétique?
      - i. Génétique moléculaire
      - ii. Génétique épidémiologique
  - C. Comment répondre à la question ?**
  - D. Que répondre à la génétique?**

Mot du RT : Cours à ne pas apprendre mais bien comprendre, c'est une introduction à la matière et une réflexion sur l'intérêt de la génétique et ses enjeux à notre époque. On a eu cours dans une salle d'ED, il n'a donc pas été enregistré. De plus, le professeur n'a pas utilisé de diaporamas. Ce n'est pas le cours sur ENCODE normalement prévu car le professeur ne pouvait pas venir. Il sera probablement reporté.  
 Le professeur va relire la ronéo mais il enverra la correction à la fin (quand il aura eu tous les cours), comme ça il pourra préciser des points s'il voit qu'ils n'étaient pas bien compris.

## I. Introduction et conseils pour l'examen

Le projet encode : le but est d'évaluer ce qui a un impact sur un phénotype dans le génome humain. 98% du génome a une fonction. Il n'y a pas que les parties codantes qui comptent et qui permettent de prédire. Il est accessible gratuitement.

**Examen :** Il sera sous forme de QCM, le prof s'en excuse et trouve ça « pauvre ». Les QCM sont souvent redondants. Les questions sont simples, il n'y a pas de piège. Pas possible d'avoir moins de 10 si on comprend « la question principale », et il est facile d'arriver à 10/12 points. Les points peuvent être pondérés.

On peut choisir d'être noté de deux façons et il faudra choisir à la première question de l'examen :

- « moins de gens répondent à la question plus elle vaut de point »
- ou « plus de gens répondent plus la question vaut de points »

C'est la deuxième option qui donne des bonnes notes. Il y a une deuxième session pour ceux qui ont raté l'examen. L'examen est fait pour être terminé en 1h mais il dure 2h.

Un cours est donné par le doyen à ne pas louper ! C'est le seul cours qu'il donne, et il y expose ses réflexions sur la génétique. Il est très conseillé d'y assister. *(C'est le 27/02/2019)*

Pour les autres cours, la présence n'est pas obligatoire mais l'idée est de travailler en petits groupes et de nourrir une discussion.

## II. Les quatre questions de la génétique

4 objectifs :

1. **Quand/Pourquoi la génétique ?**
2. **Quelle question poser à la génétique ? / A quelle question répond la génétique ?**
3. **Comment répondre à la question ?**
4. **Que répondre à la question ?**

Mot du RT: dans ce cours on va surtout aborder les deux premières questions. On survole la 3 à la fin. En fait les deux dernières seront traitées dans d'autres cours.

Actuellement, il y a énormément de données, c'est fascinant mais on s'y perd. L'important est de savoir interroger les données, qui sont déjà disponibles, contrairement au passé, où les données n'étaient pas disponibles. Cette quantité de données va continuer à augmenter, il faut donc garder en tête ses questions, ses idées, et générer des hypothèses de recherches. Aucune hypothèse n'est absurde ! Il faut être curieux.

En stage, l'important est la démarche plus que le résultat.

## A. Quand la génétique ?

La réponse est facile. L'idée c'est qu'on veut expliquer les choses. Ex : Pasteur, les maladies infectieuses... On a découvert que les microbes pouvaient donner des maladies chez l'Homme, ça c'est LA découverte majeure. La deuxième question majeure n'est pas « comment traiter » mais du point de vue conceptuel la seule question qui se pose c'est : « pourquoi des gens qui sont exposés de la même façon à la même chose ne réagissent pas pareil ? ». Pourquoi l'un meurt sous 15 jours, et chez l'autre il ne se passe rien. La question de la génétique se pose quand il y a de **la variabilité phénotypique**.

S'il n'y a aucune différence, aucune variabilité, exemple : vous envoyez 500 personnes en zone d'endémicité du paludisme, si personne ne montre aucun symptôme en 3 mois, que voulez-vous expliquer ? Il n'y a rien à expliquer, donc il n'y a pas d'intérêt génétique.

La question de la génétique est intéressante lorsqu'il y a une variabilité phénotypique. ex : le cancer du poumon, certains fument beaucoup et ne développent rien, certains bronzent, d'autres prennent des coups de soleil.

Donc lorsque vous observez de la variabilité, vous cherchez à expliquer la variabilité.

Une fois qu'on a observé une variabilité il faut qu'il y ait une **variabilité résiduelle**. Ex : on connaît les facteurs de risque : tabac pour le cancer du poumon, soleil pour la peau etc... Il faut avoir éliminé les variations environnementales.

Variabilité résiduelle = après ajustement sur les facteurs « environnementaux » connus au sens très large du terme. L'étape de recherche bibliographique doit être exhaustive. Ensuite on ajuste sur les facteurs qu'on connaît et s'il reste qqc à expliquer on se penche sur la génétique. S'il reste beaucoup à expliquer, alors la question a de l'intérêt et on a de l'information. Plus ça varie, plus on a d'information.

C'est pour ça qu'on a tellement de maladies mendéliennes monogéniques car ça c'est le summum de la variabilité, une variabilité Oui/Non., ce n'est pas un continuum.

Donc c'est simple, mais c'est fondamental.

$Y$  (le phénotype) = Environnement + Génétique +  $E \times G$ , parfois l'environnement prend le dessus. Si vous expliquez tout avec l'environnement, il n'y a plus de place pour la génétique.

La génétique humaine est par nature individuelle. C'est-à-dire que : par exemple dans les maladies infectieuses : si on n'est pas exposé au microbe, on ne l'attrape pas.

Donc dans les maladies purement environnementales on trouve les maladies infectieuses, et dans les maladies purement génétiques il y a par exemple la phénylcétonurie.

La génétique humaine est une génétique acquise, conditionnelle. Elle est adaptée à l'environnement. Avant 1900, 50% des gens mouraient avant 30 ans (maladies infectieuses, violence, famine, guerre...)

Donc en conclusion il faut de la maladie et de la variabilité. Peut-être qu'il y a de la génétique dans tout mais il y a de l'environnement dans tout aussi, mais dans quelle proportion ? Il faut avoir et garder des convictions !

## B. Quelle question poser ?

En partant de l'hypothèse qu'il y a de la génétique, et qu'on travaille chez l'Homme, le phénotype on s'en fou.

Il faut décrire très précisément le phénotype, c'est pourquoi c'est difficile en psychiatrie car les phénotypes sont très diffus.

Donc si on travaille avec des maladies infectieuses, on a éliminé les facteurs de variabilité tels que la virulence, le temps d'exposition, les immuno supprimés, et malgré tout il y a une variabilité importante qui est observée. Si vous êtes dans un modèle expérimental chez la souris après vous pouvez vous amuser à tout contrôler.

Exemple des maladies infectieuses : On prend une population, on l'expose au microbe, certains vont développer une infection. Parmi ceux qui sont infectés, certains vont développer une forme clinique de la maladie. Comment on explique cette variabilité (car elle existe)? Et elle peut s'expliquer du côté du microbe : intensité d'exposition : facteur non génétiques ou alors par des facteurs génétiques spécifiques.

Si vous travaillez sur le cancer du poumon et que vous ne prenez pas en compte le tabac vous allez pouvoir tomber sur une composante génétique forte alors que finalement ça se rapporte au tabac. Il y a une variabilité importante résiduelle. Si on travaille sur des souris on peut expérimenter à l'infini, c'est « génétique 2.0 » car on contrôle tout ce qu'on peut contrôler, on peut même contrôler les gènes. (ex en invalidant des gènes, en les dupliquant etc)

Exposition au microbe à Infection à Clinique : le passage d'une phase à l'autre peut s'expliquer par la virulence du microbe, le temps d'exposition... (coté microbe) ou alors par des facteurs non génétiques ou génétiques (coté infecté).

En revanche, on est chez la souris... c'est un modèle.

Chez l'Homme, on ne peut pas faire ça. On ne peut pas mettre les gens sous un régime particulier, un mode de vie particulier, invalider un gène etc... on ne peut pas. Mais il faut avoir les yeux ouverts car finalement la nature le fait pour vous. Parce que à chaque méiose, sur 10 millions de pb, il y a une mutation de novo (1/10 millions). Donc on fait du « in natura ». C'est à dire qu'on étudie l'impact des mutations créent par la nature *via la méiose*.

### 1. Deux phénotypes extrêmes

- On est vacciné par le BCG, c'est très bien pour protéger ses enfants d'une infection pulmonaire. Mais une fois de temps en temps (1/100 000, 1/200 000), l'enfant meurt après le vaccin. On se demande alors « J'ai fait une erreur? ». C'est inexplicable ce qu'il se passe. À l'autopsie on se rend compte que l'enfant est envahis par une tuberculosis atténuée, c'est un cas de **BCGite disséminé**. En effet, même s'il est affaibli ++, le BCG reste un vaccin vivant. Il y a des enfants tellement susceptibles aux immunobactéries que les vacciner revient à les tuer. C'est très rare mais quand ça arrive cela doit nous alerter. C'est un premier phénotype, ayant une fréquence de 1/500 000.
- Ensuite on travaille sur quelqu'un infecté d'un tuberculosis depuis qu'il a 40 ans. C'est un phénotype beaucoup plus fréquent (1/100 voir 1/1000 en France).

D'un côté on a donc quelque chose de **très rare** et de **très sévère** et de l'autre quelque chose de finalement **moins sévère** mais de **moins rare**.

i) Cas de la BCGite disséminé

Le phénotype de cette maladie est très sévère, notre esprit a donc tendance à penser que la **cause génétique est très sévère**. On a l'impression que c'est une **lésion monogénique rare**.

Ce phénotype très sévère va avoir un impact **catastrophique** au niveau de l'**individu**. En effet si on a ce variant on va développer la maladie et mourir. Cependant ce variant est très très rare, peu de personnes seront touchés par ce dernier. Ainsi la **population** ne sera que **faiblement touchée**.

ii) Cas de la tuberculose pulmonaire

Cette pathologie est assez **fréquente**. On va donc avoir tendance à penser qu'un unique variant ne peut pas être responsable. Ainsi, quand un phénotype est moins sévère et moins rare on a tendance à dire que la **contribution génétique est complexe**.

*On n'est pas si absurde du point de vue de la génétique de l'évolution. En effet, la tuberculose existait avant même l'Homo Sapiens. Ça serait donc surprenant que l'Homo Sapiens ait conservé dans son génome une séquence qui le rende très vulnérable à la pathologie. En principe c'est éliminé par l'évolution, c'est la sélection naturelle.*

L'impact au niveau de l'**individu** va être relativement **faible**. En effet, si un individu possède les multiples variants, ses chances de contracter la tuberculose pulmonaire sont multipliées par 1,5 ou 2. L'effet de ces variants est modéré, l'individu ne va pas forcément développer la tuberculose. En revanche, l'impact au niveau de la **population** est **important** car le phénotype est fréquent, un grand nombre de la population peut donc être touché.

Ainsi pour résumer:

	<b>Origine génétique</b>	<b>Impact sur l'individu</b>	<b>Impact sur la population</b>
<b>Phénotype rare, sévère de l'enfant</b>	Monogénique	Catastrophique	Faible
<b>Phénotype moins sévère, moins rare de l'adulte</b>	Multigénique	Très faible	Important

C'est très scolaire mais c'est ce qu'on pensait il n'y a pas si longtemps. Le fait est que c'est vrai mais ce sont des exceptions finalement.

Deux commentaires sur « monogénique »:

- Pour **démontrer** que quelque chose est **monogénique**, il faudrait qu'on ait un organisme qui ne possède qu'un gène. Et quand ce gène subit une mutation, survient un phénotype. À ce moment là on pourrait parler de monogénisme. Mais il n'existe pas à l'heure actuelle, sur la planète terre, d'organisme avec un seul gène.

- Par exemple, pour la mucoviscidose on a isolé une mutation qui représente 80% des mutations responsables de la maladie. On retrouve une **application complète**, c'est à dire que quand un individu a la mutation, il va être malade. Mais, sur ce fond génétique commun, des gens vont mourir à 5 ans et d'autres à 35/40 ans. Il y a donc bien quelque chose qui se passe, qui intervient. Ça peut être l'environnement ou une seconde couche génétique qui va modifier l'impact de la première mutation. L'idée c'est que les **mutations de novo** qui se produisent, **tombent sur un génome**. Ce dernier a ensuite la capacité ou non d'absorber cette mutation. C'est ça qui va déterminer la **pénétrance d'une mutation**. C'est la capacité du génome, à toutes les positions (*sauf celle de la mutation*), d'absorber, de tamponner les effets de la mutation (*comment on en sait rien*). Cela est rassurant car ça veut dire qu'on n'est pas dépendant d'un seul gène. Les gènes ne sont pas indépendants les uns des autres, ils interagissent. Donc monogénique ça n'existe pas.

NB: Le prof a fait remarqué que si sur un ensemble de phénotypes sur lesquels on travaille on remarque un groupe quasiment monogénique il faut l'isoler car il crée de l'hétérogénéité. En effet, si on a réussi à expliquer un cas par un gène, il faut exclure ce cas de nos études car on a résolu le problème.

## 2. Quel type de génétique?

Quelle question se poser? Ça a de l'importance car ça permet de savoir quelle type de génétique on va faire.

Si le **phénotype est rare et sévère**, on s'oriente vers la **génétique moléculaire**. En revanche quand il est **moins rare et moins sévère** on s'oriente plutôt vers la **génétique épidémiologique**.

### i. Génétique moléculaire.

On a un phénotype cellulaire et on travaille de façon moléculaire.

*Par exemple, le phénotype cellulaire est le fait que les lymphocytes ne réagissent pas en présence de la tuberculose, c'est à dire qu'ils ne produisent pas la molécule IL12 (dans le cas de la BCGite disséminée). On va alors rajouter un morceau de chromosome 2 ou 3 (peu importe c'est un exemple). Au final, quand on va introduire le gène codant pour le récepteur permettant la production de IL12, le lymphocyte va produire IL12. On a alors **montré au niveau cellulaire** (lymphocyte) qu'il y avait sans doute **une perte de fonction** du récepteur de l'IL12 et c'est ça qui cause la BCGite disséminée. On a donc **au niveau du lymphocyte une démonstration du déterminisme**: si on n'a pas le récepteur, on ne peut pas produire IL12.*

En fait, en génétique moléculaire on s'attend à expliquer le phénotype par une protéine ou une fonction disparue. Le **niveau de preuve est très élevé** au **niveau cellulaire**, spontanément les gens vont donner plus de force à l'évidence dans ce cas. Le problème c'est que là on travaille au niveau du lymphocyte et non au niveau de l'individu. Le saut qui est fait est gigantesque. Qu'est ce qui prouve que c'est bien ça qui est responsable de la pathologie de l'enfant? On ne sait pas comment extrapoler du lymphocyte au patient, mais dans l'inconscient ça va se faire de façon spontanée.

### ii) Génétique épidémiologique

Dans ce cas on va faire des **biostatistiques** et de **l'observation**. En effet, aucun des individus n'a suffisamment d'information génétique pour qu'on puisse répondre à la question juste en l'étudiant lui. Pour une cause génétique complexe, on ne s'attend pas à ça. Du coup, comme il y a

beaucoup de cas, on va les mettre ensemble et regarder si la fréquence d'un variant diffère entre les cas et le témoin.

On a un échantillon 1 avec 20% du variant chez les cas et 8% chez les témoins, et on répète avec d'autres échantillons. *Par exemple quand on travaille avec la tuberculose de l'adulte, on va remarquer que le risque est multiplié par 3 chez les porteurs du variant.*

On répète l'expérience 10 fois et 10 fois on retrouve la même tendance. L'évidence commence à s'imposer mais la **nature de la preuve** est quelque part **expérimentale**. On va alors parler de **p-value**. En fait c'est la probabilité que le hasard, qu'il y ait une telle différence, est faible.

Quand on travaille sur des **p-values**, sur des groupes d'individus, on est certain d'avoir une **approche holistique**. C'est à dire qu'on ne sait pas pourquoi ce variant est sur représenté chez les cas qui ont la tuberculose, mais ce qui est sûr c'est qu'en étudiant les cas de tuberculoses dans un environnement naturel, la variant sera représenté. Le **niveau de preuve** d'un point de vue causalité est donc **peut être faible** mais l'**association** est absolument **convaincante**. Elle existe en vrai, dans la nature chez l'homme en entier. C'est parfois compliqué à faire intégrer aux gens. Par exemple, il a fallu 40 ans pour établir la causalité entre tabac et cancer du poumon.

NB: Le prof insiste sur le fait que c'est pointu mais très important.

Si on veut avoir un pouvoir prédictif sur la génétique, il faut alors demander dans une famille s'il y a des antécédents de pathologies. Si la réponse est oui, c'est beaucoup plus prédictif que 20 marqueurs. Pourquoi? Parce que si la réponse est oui, c'est tout le génome qu'on prend en compte. C'est pour ça que le prof préfère travailler dans les familles où il y a deux, trois cas, car si c'est génétique (cela peut être environnemental), l'homogénéité va être bien.

Avant ces deux approches étaient totalement séparées. D'un côté il fallait trouvé un phénotype cellulaire et de l'autre un nombre d'individu qui soit convenable pour répondre à la question. Maintenant, le séquençage du génome a radicalement changé les choses, on parle maintenant de quelque chose de commun: l'ADN. En fait, la base de donnée qu'on va interroger est la même dans les deux cas, c'est **la question qu'on va posée qui diffère**. On va interroger la base de données de manière différente si la maladie est monogénique ou complexe. Pour les maladies complexes, on peut éliminer entre guillemets ce qui se rapporte aux maladies monogéniques. C'est une révolution, on parle maintenant le même langage.

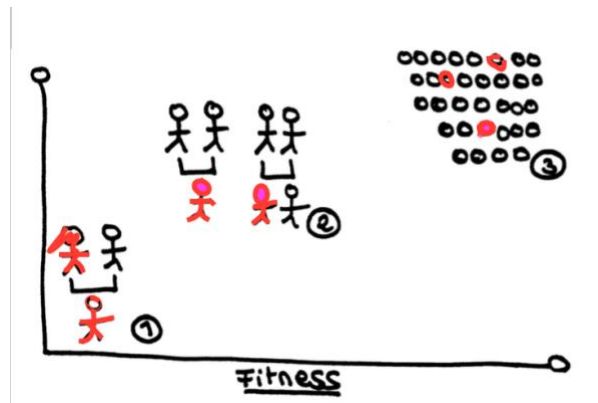
<p><u>Mot du RT</u>: le prof précise qu'à l'examen on va avoir 10 c'est sûr. Tout le monde ne va pas se lancer dans la recherche, la médecine on la fait pour les autres et la recherche on la fait pour soi.</p>
---

### C. Comment répondre à la question?

Le principale c'est d'avoir des idées. C'est très facile de les tester. En effet, il a plusieurs outils à notre disposition et les calculs sont fait par des logiciels pour nous. Tout existe déjà. Le problème c'est de savoir quel outil utilisé en fonction de la question que l'on se pose. Cette question va générer les données sur lesquelles on va travailler.

L'outil va nous donner un résultat. Comment on va l'interpréter? Où sont les limites de l'interprétation? Savoir des formules par coeur, jamais ça ne va tomber le jour de l'examen.

Les outils qui vont être utilisés au final sont très simples.



1. Génétique clinique (au sein de la famille). On va faire du séquençage, l'interprétation des résultats et très important.
2. Génétique épidémiologique (quelque chose de plus fréquent). On va faire des études plus communes. On va travailler sur des familles. On ne cherche pas de mutations perte de fonction. La seule chose importante est qu'il nous faut un variant ayant une fréquence relativement élevé (plus de 1%), pour avoir de la puissance statistique.
3. Génétique des populations (ensemble des individus de la planète ou une population)

Comment savoir par quel gène commencer parmi les 200 qui sont possibles (pour les cas 1 et 2)? La réponse est donnée par la génétique des populations parce qu'elle va nous dire, par exemple, que le gène 3 est très conservé et que le 17 lui n'est pas conservé. Ça veut dire que s'il y a une mutation sur le 3 cela a probablement un effet majeur. On commencera donc par celui-ci.



# FICHE RECAPITULATIVE

## I. Quand/ pourquoi la génétique ?

**Problématique :** Pourquoi des gens exposés de la même façon n'ont pas les mêmes maladies ? La question de la génétique se pose donc quand il y a une variabilité phénotypique.

**Objectif :** Expliquer ces variations avec la génétique

### Problèmes :

- Existence d'une variabilité résiduelle, il faut éliminer toutes les composantes environnementales pour pouvoir étudier le rôle de la génétique.

- Ce modèle « Phénotype = environnement + génétique + interactions gènes/environnement » n'est pas toujours exact. La nature est conditionnée, par exemple les maladies infectieuses sont purement d'origine environnementale (il n'y a pas de maladie s'il n'y a pas de moustique dans le cas du paludisme) alors que d'autres maladies sont purement d'origine génétique comme la phénylcétonurie.

## II. Quelle question poser à la génétique ?

Deux phénotypes très différents :

	Origine génétique	Impact sur l'individu	Impact sur la population
<b>Phénotype rare, sévère de l'enfant</b> Ex : BCGite	Monogénique	Catastrophique	Faible
<b>Phénotype moins sévère, moins rare de l'adulte</b> Ex : Tuberculose pulmonaire	Multigénique	Très faible	Important

Si le phénotype est rare et sévère, on s'oriente vers la génétique moléculaire. En revanche quand il est moins rare et moins sévère on s'oriente plutôt vers la génétique épidémiologique.

En génétique moléculaire on s'attend à expliquer le phénotype par une protéine ou une fonction disparue. Le niveau de preuve est très élevé au niveau cellulaire.

La génétique épidémiologique est plus complexe, on étudie plusieurs variants. On va faire des biostatistiques et de l'observation. La nature de la preuve est quelque part expérimentale. On va alors parler de p-value.

## III. Comment répondre à la question ? Comment on teste nos idées ?

- Si on étudie la génétique clinique, les outils d'étude sont les familles.

- Si on étudie la génétique épidémiologique, il faut créer des groupes homogènes dans les populations.