

Caractéristiques des facteurs génétique

INTRODUCTION

1. La modélisation du risque génétique de la maladie,
2. la modélisation de la relation entre les fréquences des allèles et des génotypes,
3. et la modélisation de la relation entre les fréquences des allèles entre deux locus.

Modèle génétique de risque

Chaque individu peut porter l'un des trois génotypes S||S S||s et s||s. La pénétrance d'un génotype est la probabilité qu'un individu porteur de ce génotype soit malade. Ainsi, dans le modèle général de risque, trois pénétrances peuvent être calculées, P2, P1 et P0 telles que :

$$P2 = P(\text{Malade sachant S||S})$$

$$P1 = P(\text{Malade sachant S||s})$$

$$P0 = P(\text{Malade sachant s||s})$$

En prenant pour référence le génotype s||s, on définit deux risques relatifs génotypiques :

$$RRG2 = P2 / P0$$

$$RRG1 = P1 / P0$$

Dans le cas d'un modèle dominant, la présence d'au moins une copie de l'allèle S augmente le risque de la maladie avec :

$$P2 = P1 = P0 \times RRG$$

Dans un modèle récessif, deux copies de l'allèle S sont nécessaires pour que le risque de la maladie soit augmenté avec :

$$P2 = P1 \times RRG = P0 \times RRG$$

Dans un modèle additif les relations entre les pénétrances sont :

$$P2 = P1 \times RRG = P0 \times RRG^2$$

Modèle de Hardy-Weinberg

Ce principe a été indépendamment décrit par le physicien Wilhelm Weinberg (1862-1937) et le mathématicien Godfrey Hardy (1877-1947). Une population est dite en équilibre d'Hardy Weinberg:

1. si elle est de taille infinie, c'est-à-dire suffisamment grande pour être gouvernée par les lois de Mendel,
2. si l'espèce est diploïde et dispose d'une reproduction sexuée, si les mariages se font au hasard (panmixie),
3. s'il n'y a ni migration, ni mutation, ni sélection naturelle et que les générations ne se chevauchent pas.

Dans ce cas, la fréquence des allèles et des génotypes reste stable au cours des générations et on peut calculer les fréquences génotypiques directement à partir des fréquences alléliques et inversement.

Déséquilibre de liaison

Le déséquilibre de liaison (DL) correspond à une répartition non aléatoire d'allèles pour un certain nombre de locus. Il traduit donc l'indépendance ou la non indépendance physique de deux loci lors de leur transmission aux gamètes, au cours de la méiose.

On définit alors le déséquilibre de liaison D comme la différence entre les fréquences chromosomiques observées et les fréquences théoriques à l'équilibre et on a alors :

$D > 0$, lorsque les allèles rares (ou fréquents) des deux loci sont préférentiellement associés ensemble et,

$D < 0$, lorsque l'allèle rare de l'un des loci est préférentiellement associé à l'allèle fréquent de l'autre locus.

La valeur maximale ou minimale de D peut être calculée à partir des fréquences alléliques des loci A et B. Ainsi, concernant la valeur maximale de D (D_{\max}), qui est la plus fréquemment utilisée, désignons par A celui des deux loci pour lequel la fréquence de l'allèle rare est la plus faible ($p_{A2} < p_{B2}$) :

$$\text{Si } D > 0, D_{\max} = p_{A2} (1 - p_{B2}) = p_{A2} p_{B1}$$

$$\text{Si } D < 0, D_{\max} = p_{A2} (1 - p_{B1}) = p_{A2} p_{B2}$$

1. Le coefficient D' est égal au rapport D/D_{\max} et est compris entre -1 et +1. Plus la valeur absolue de ce rapport est proche de 1, plus le DL est fort ; la valeur absolue 1 correspondant à un déséquilibre complet.

2. Le coefficient r^2 est égal au rapport $(D)^2/(p_{A1} \times p_{A2} \times p_{B1} \times p_{B2})$. La valeur 1 correspond dans ce cas à un déséquilibre parfait, signifiant non seulement que $D'=1$, mais aussi que les fréquences alléliques des deux loci sont strictement égales. Dans le cas où le coefficient entre 2 SNPs est égal à 1, la présence d'un allèle implique automatiquement la présence de l'autre, l'information recueillie à partir de l'étude d'un des deux polymorphismes est donc strictement identique à celle qui serait fournie par l'étude du polymorphisme avec lequel il est en DL parfait.

Dans la plupart des études, le DL est considéré comme fort si le r^2 est supérieur ou égal à 0,80.

Haplotypes

Les allèles de plusieurs SNPs qui sont proches tendent à rester regroupés lors de la transmission héréditaire : un tel groupe d'allèles de SNPs associés dans une région chromosomique est appelé haplotype.

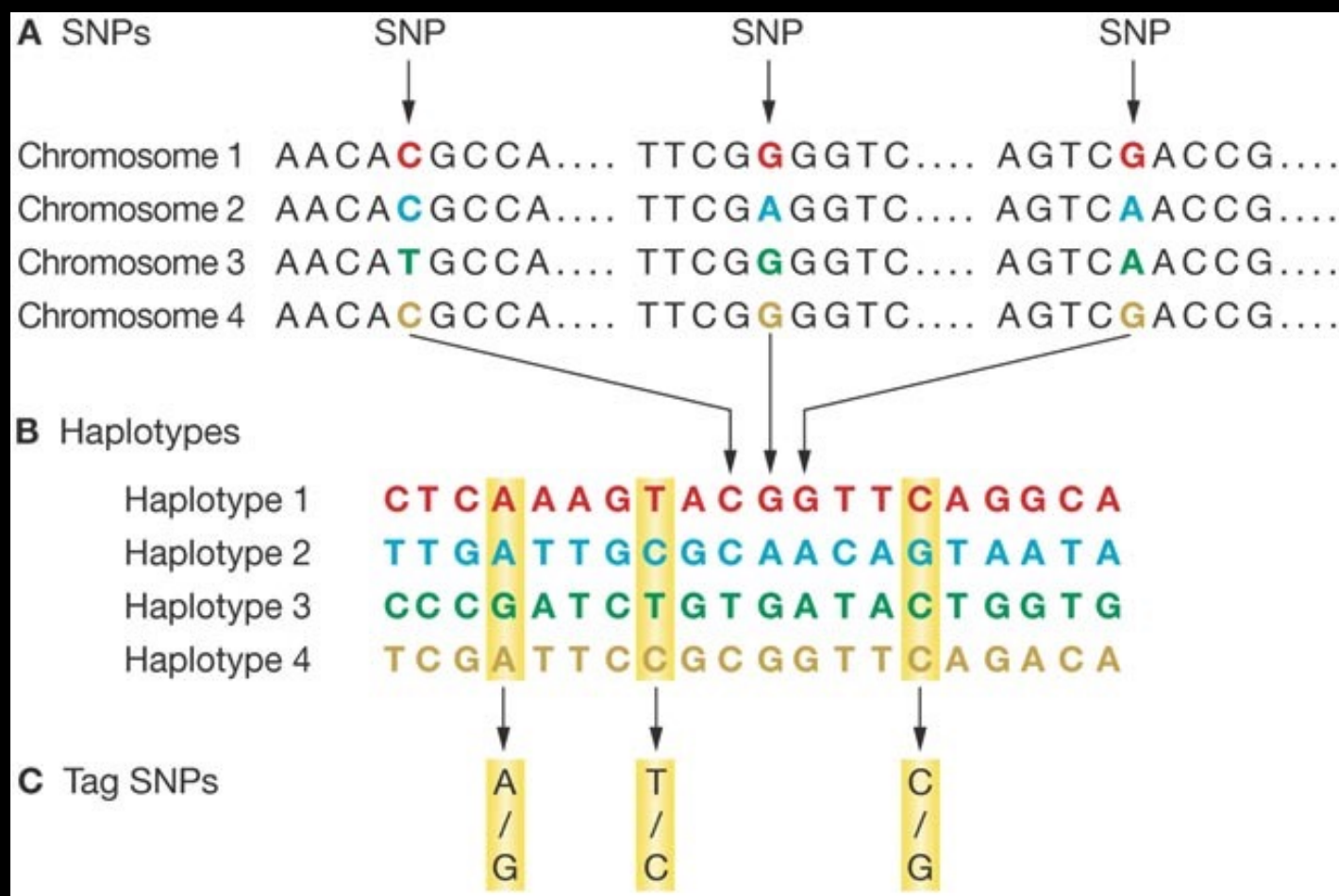
Les fréquences haplotypiques ne peuvent être, en fait, mesurées que dans le cadre d'études familiales. Cependant, des modèles statistiques de maximum de vraisemblance permettent d'estimer les fréquences haplotypiques dans des populations d'individus non apparentés à partir des fréquences des génotypes, permettant ainsi d'évaluer l'impact d'une combinaison particulière de polymorphismes sur un phénotype donné et non pas d'un seul polymorphisme.

Quand des polymorphismes présentent un DL, la transmission des allèles d'un chromosome se fait en réalité par bloc haplotypique. La probabilité de recombinaison est donc très faible au sein d'un même bloc. Ces haplotypes sont généralement séparés par des régions de faible DL représentant des points chauds (« hot spots ») de recombinaison méiotique. Chaque bloc haplotypique est donc défini par la combinaison d'un nombre variable de polymorphismes et chaque gène peut contenir un nombre variable de blocs haplotypiques.

Cette notion est essentielle en étude d'association car si un bloc haplotypique contient 5 SNPs et que ce bloc est associé à un phénotype donné alors le SNP fonctionnel peut être n'importe lequel de ces 5 SNPs.

Tag-SNPs

L'utilisation du DL, et plus particulièrement des blocs haplotypiques, est couramment utilisée en étude d'association afin d'alléger la quantité de SNPs à génotyper. En effet, du fait de la corrélation existant entre les polymorphismes d'un même bloc, il est possible de sélectionner le ou les SNPs les plus informatifs, c'est-à-dire représentant au mieux l'information portée par le bloc haplotypique ; ces SNPs sont appelés tag-SNPs. Les Tag-SNPs permettent ainsi de réduire le nombre de polymorphismes à étudier sans perdre d'information génétique et donc le coût du génotypage.



Reproduced with permission from The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426: 789–796

Principe de la construction de HapMap (d'après le projet HapMap)