



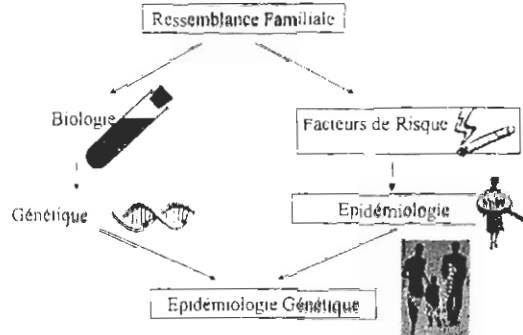
# Epidémiologie génétique

Florence RICHARD  
MCU-PH, Lille2



F.Richard

1



F.Richard

2

## Objectifs

- ① Mettre en évidence la composante familiale d'une maladie
- ② Mettre en évidence des facteurs génétiques impliqués dans le déterminisme du trait étudié
- ③ Estimer les effets de ces facteurs génétiques

F.Richard

3

## Composante familiale des maladies

Florence RICHARD  
MCU-PH, Lille2



F.Richard

4

## Composante familiale des maladies

### Trois étapes

- ① Mise en évidence du caractère familial d'une maladie (excès de malades dans la famille)
- ② Etudier l'origine de cet excès familiale  
Est-il d'origine héréditaire ou environnementale ?
- ③ Etudier la nature de la composante héréditaire.

F.Richard

5

## Composante familiale des maladies

### Trois étapes

- ① Mise en évidence du caractère familial d'une maladie
- ② Etudier l'origine de cet excès familiale
- ③ Etudier la nature de la composante héréditaire.

F.Richard

6

## Caractérisation du caractère familial d'une maladie

### • Etude cas-témoins : deux types d'enquête

- ① Enquête cas-témoins : présence ou absence d'un antécédent familial
- ② Proportion de sujets qui ont au moins un apparenté du premier degré atteint de la maladie

F. Richard

7

## Etude de l'existence d'un antécédent familial dans une étude cas-témoins

	Cas	Témoins
Histoire familiale +	a	b
Histoire familiale -	c	d

F. Richard

8

## Etude de la distribution familiale chez les apparentés de malades et de témoins

- Le facteur étiologique est représenté par l'état de maladie du sujet propositus ou probant (sujet à partir duquel on étudie la famille)
- Le groupe témoin est constitué par la famille d'un sujet sain.

F. Richard

9

## Etude de la distribution familiale chez les apparentés de malades et de témoins

		Cas	Témoins
Apparentés	Malades	a	b
	Sains	c	d
		$\frac{a}{a+c}$	$\frac{b}{b+d}$

prévalence

F. Richard

10

## Etude de la distribution familiale chez les apparentés de malades et de témoins

- On peut remplacer le groupe témoin par la prévalence de la maladie en population générale *qd connue*.
- Prise en compte possible de facteurs de confusion (âge, sexe, ...) *trouble de la probabilité*

F. Richard

11

## Exemple 1 : Mortalité par cancer du cancer du poumon.



Apparentés de :	% de la maladie	RR
Sujets sains non-fumeurs	0.28	1
Sujets malades non-fumeurs	1.11	3.96
Sujets sains fumeurs	1.47	5.25
Sujets malades fumeurs	3.82	13.64

D'après G. Tokuhata J. Natl. Cancer Inst. 1963, 30:298-312

F. Richard

12

*indépendant du fait de fumer ou non d'être malade*

## Fente Palatine



F. Richard

13



## Exemple 2 : Incidence de la fente palatine

Parents	% de parents atteints	/ PG
<b>Premier degré</b>		
Fratric	4.1	× 40
Enfants	3.5	× 35
<b>Second degré</b>		
Oncles et tantes	0.7	× 7
Neveux et nièces	0.8	× 8
<b>Troisième degré</b>		
Cousins germains	0.3	× 3

D'après Carter Br Med Bull, 1969, 25 : 52-57

F. Richard

14

## Composante familiale des maladies

### Trois étapes

- ① Mise en évidence du caractère familial d'une maladie (excès de malades dans la famille)
- ② Etudier l'origine de cet excès familial
- ③ Etudier la nature de la composante héréditaire.

F. Richard

15

## Rappel

- **Génotype** : il s'agit de la constitution génétique d'un individu.
- **Phénotype** : ensemble des caractères observés chez un individu : sa taille, sa couleur de cheveux, sa pression artérielle, ses enzymes, ses protéines, présence ou absence d'une pathologie, ...
- Le phénotype va résulter du génotype d'un individu mais également de l'interaction de son génotype avec des facteurs environnementaux.

F. Richard

16

## Origine de l'excès familial

Pour un phénotype (P) donné, on va définir un effet génétique (G) et un effet environnemental (E)

$$P = G + E$$

Variance de P

$$V_P = V_G + V_E + 2\text{cov}_{(G/E)}$$

F. Richard

17

## Origine de l'excès familial

$$\text{Cov}_{(G/E)} = ???$$

On suppose que cette covariance = 0.

Cela signifie :

- Absence de corrélation entre époux
- Corrélations enfants-enfants = corrélations parents-enfants

F. Richard

18

## Origine de l'excès familial

$COV_{(G/E)} \text{ étant } = 0,$

$$V_P = V_G + V_E$$

On va redécomposer la variance  $V_E$  en

$$V_{EFam} + V_{EGen}$$

*en v. nominale*

$$V_P = V_G + V_{EFamilial} + V_{EGénéral}$$

Hérédité  
biologique      Hérédité  
culturelle

Séparation difficile

F. Richard

19

## Origine de l'excès familial

### Deux types d'enquêtes

- ① La méthode des jumeaux
- ② L'étude des enfants adoptés

F. Richard

20

## La méthode des jumeaux

• Des jumeaux (1/80 grossesse) sont :

- Soit Monozygotes (MZ) (1/3 "Vrais jumeaux") :
  - » Issus du même œuf
  - » Partageant le même patrimoine génétique
- Soit Dizygotes (DZ) (2/3 "Faux jumeaux") :
  - » Issus de deux œufs différents
  - » Partageant en moyenne, 50% de leur patrimoine génétique

• On part de l'hypothèse que les différences d'environnement entre deux jumeaux sont les mêmes pour les MZ et les DZ.

F. Richard

21



## La méthode des jumeaux



Au sein d'une paire on peut avoir

	Premier Jumeau	Second Jumeau		
Concordance	Sain	Sain	}	😊 😊
Discordance	Malade	Sain		😊 😞
Discordance	Sain	Malade		😊 😞
Concordance	Malade	Malade		😞 😞

F. Richard

22



## La méthode des jumeaux



- On essaiera de se restreindre aux paires de jumeaux de même sexe.
- Evident pour les MZ (♂♂ ou ♀♀) !
- Il ne faut pas oublier que pour les DZ, cela ne sera pas toujours vrai (♂♂ ou ♀♀ ou ♂♀).
- Il serait difficile chez des DZ de sexe différent de faire l'hypothèse qu'ils partagent les mêmes différences environnementales à l'intérieur d'une paire que les MZ.

F. Richard

23



## La méthode des jumeaux



- Si une maladie, une malformation, un trait est totalement génétiquement déterminé
  - » 100% des jumeaux MZ devraient être concordants pour la maladie
  - » 50% des jumeaux DZ devraient être concordants pour la maladie si maladie autosomique dominante
- Si une maladie n'a rien de génétique
  - » La concordance au sein des paires MZ devrait être grossièrement égale à la concordance des paires de DZ
- Si une maladie est multifactorielle
  - » Les jumeaux MZ ne devraient pas toujours être concordants mais ils le seront plus souvent que les DZ.

F. Richard

24

## La méthode des jumeaux

- MZ : discordance (1 jumeau malade, 1 jumeau sain) ne peut être que d'origine environnementale.
- DZ : la discordance peut être d'origine génétique et/ou environnementale.
- On va comparer les taux de concordance (TC) d'une maladie des MZ et des DZ.

F Richard

25

## La méthode des jumeaux

		Jumeau	
		Sain	Malade
Jumeau	Sain	X	D (discordants)
	Malade	D	C (concordants)

Remarque: les paires de jumeaux sains ne sont pas prises en compte

F Richard

26

## La méthode des jumeaux

- Deux sortes de concordance
  - Concordance intra paires (Pairwise concordance)

$$TC = \frac{C}{C+D}$$

TC =  $\frac{\text{Nombre de paires de jumeaux où les deux sont atteints de la maladie}}{\text{Nombre de paires de jumeaux où au moins un est atteint de la maladie}}$

F Richard

27

## La méthode des jumeaux

- Deux sortes de concordance
  - "Probandwise Concordance"

$$PC = \frac{2C}{2C+D}$$

Probabilité d'un jumeau d'être atteint sachant que son jumeau est atteint de la maladie (probabilité conditionnelle).

F Richard

28

## La méthode des jumeaux

MZ	a
Concordance	Discordance
DZ	b

a : différence due à l'environnement

b : différence due aux effets des genes.

TCMZ > TCDZ s'il existe une composante héréditaire

F Richard

29

## La méthode des jumeaux

Exemple: décès par maladie coronarié

	Hommes		Femmes	
	MZ	DZ	MZ	DZ
Nb paires	1649	2982	2006	3865
Nb concord	114	164	66	99
Nb discord	335	705	273	611
TC	0.25	0.19	0.20	0.14
p	< 0.001		< 0.05	

Registre des jumeaux suédois, d'ap Marenberg et al., N Engl J Med, 1994; 330: 1041



## La méthode des jumeaux



### Limites de ces études :

- Recueil des informations souvent lent et difficile
- L'effet du milieu peut être différent chez les MZ et les DZ. Traits comportementaux.
- Problème avec les petites séries qui surestiment le taux de concordance, surtout chez les MZ

F. Richard

31

## Les études d'adoption

- L'adoption sépare les enfants de leurs parents biologiques
- Un enfant adopté
  - Par rapport à ses parents adoptifs :
    - » Partage une partie de son environnement
    - » Peu de similarités génétiques
  - Par rapport à ses parents biologiques
    - » Partage en moyenne, 50% de leur patrimoine génétique
    - » A peu **partagé** son environnement

F. Richard

32

## Les études d'adoption

Enquêtes très difficiles et quasiment impossibles en France car elles nécessitent de :

- De recenser des enfants adoptés
- De connaître les parents adoptifs
- De connaître les parents biologiques

F. Richard

33

## Les études d'adoption

Plusieurs approches sont possibles :

- Trouver des personnes adoptées souffrant d'une maladie particulière et rechercher si cette maladie se transmet dans la famille biologique ou dans la famille adoptive
- Démarrer avec un sujet atteint dont les enfants ont été adoptés et rechercher si l'adoption permet de protéger un enfant du risque de la maladie.
- Regarder la **ressemblance** entre des enfants adoptés et leurs parents adoptifs d'une part et entre des enfants et leurs parents biologiques d'autre part

F. Richard

34

## Les études d'adoption

- Si trait est génétiquement déterminé
  - ➔ On s'attend à retrouver plus fréquemment la maladie dans la famille biologique de l'enfant adopté
  - ➔ Si transmission autosomique dominante : 50% des cas
- Si une maladie n'a rien de génétique, si elle est totalement sous l'influence de facteurs environnementaux
  - ➔ Dans ce cas, l'enfant ressemblera plus à ses parents adoptifs.

F. Richard

35

## Les études d'adoption

### Exemple : Enfants Hyperactifs.

% de pères ayant été ou étant hyperactifs dans un échantillon d'enfants hyperactifs

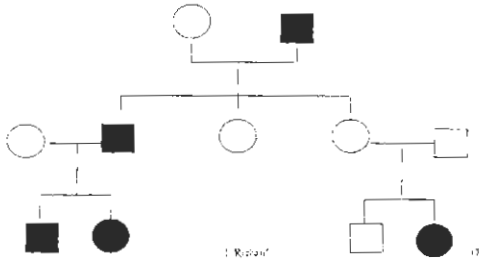
% de Pères hyperactifs	% de Pères hyperactifs
Enfants biologiques	Enfants adoptés
16%	3%

F. Richard

36

Attention : un excès de cas familiaux n'est pas toujours d'origine héréditaire génétique.

Ne pas négliger l'hérédité culturelle



## Composante familiale des maladies

- Trois étapes
  - ① Mise en évidence du caractère familial d'une maladie (excès de malades dans la famille)
  - ② Etudier l'origine de cet excès familiale
  - ③ Etudier la nature de la composante héréditaire

## Etude de la nature de la composante héréditaire d'une maladie

- Analyse de ségrégation
- Objectif : détecter l'existence de facteurs génétiques et estimer les effets respectifs d'un gène transmis de façon mendélienne, de gènes multiples à effet faible et de l'environnement afin d'expliquer les distributions familiales d'un caractère donné
- Permet de préciser le mode héréditaire d'un caractère.

## Composante Héréditaire d'une maladie

### Modes de transmissions

- Les maladies à transmission mendélienne (monogénique)
  - transmission autosomique dominante
  - transmission autosomique récessive
  - transmission récessive liée à l'X
- Les maladies mitochondriales
- Les maladies à hérédité complexe
  - polygénique avec gene majeur
  - polygénique sans gene majeur



## Composante Héréditaire d'une maladie

### Hétérogénéité Génétique et Phénotypique des maladies à transmission mendélienne

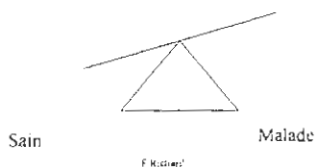
- Un locus et plusieurs allèles
  - Plus de 400 mutations pour la mucoviscidose
- Une maladie et plusieurs locus et plusieurs allèles
  - Au moins 3 locus pour la maladie d'Alzheimer
- Hétérogénéité phénotypique
  - Maladie de Recklinghausen : (tâche café au lait à des tumeurs cérébrales)
- Un même gène et plusieurs maladies
  - Gène de la  $\beta$ -globine : drépanocytose,  $\beta$ -thalassémie

## Composante Héréditaire d'une maladie

### Les maladies à Hérédité Complexe

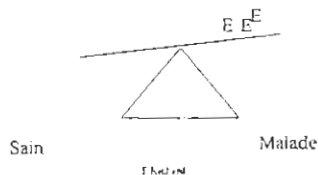
- Transmission non mendélienne
  - Transmission polygénique (oligogénique) avec gene(s) majeur(s)
  - Transmission polygénique (oligogénique) sans gene majeur
- Facteurs de susceptibilité génétique
- Maladies polygéniques
- Rôle de l'environnement

Composante Hérititaire d'une maladie  
Les maladies à Héritéité Complexe



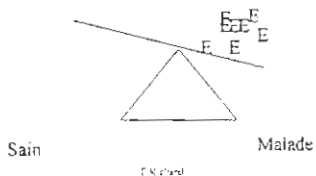
41

Composante Hérititaire d'une maladie  
Les maladies à Héritéité Complexe



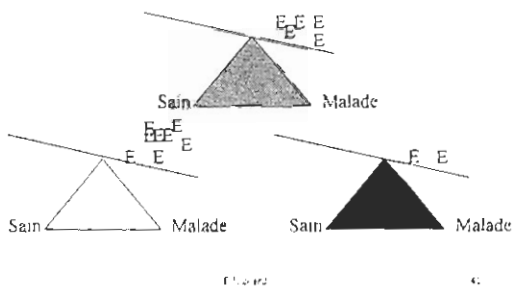
44

Composante Hérititaire d'une maladie  
Les maladies à Héritéité Complexe



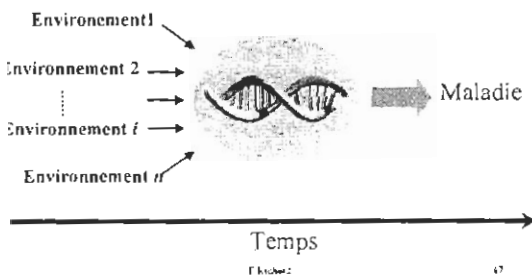
45

Composante Hérititaire d'une maladie  
Les maladies à Héritéité Complexe



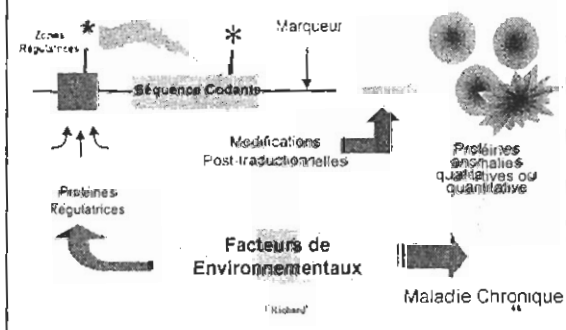
46

Composante Hérititaire d'une maladie  
Les maladies à Héritéité Complexe  
Susceptibilité Individuelle



47

Composante Hérititaire d'une maladie  
Héritéité complexe : Susceptibilité Individuelle





## Composante Héritaire d'une maladie

### De la transmission mendélienne à l'Hérédité Complexe : toutes les situations sont possibles



Génétique (Très rares) Environnementale

## Composante Héritaire d'une maladie : l'exemple de la maladie d'Alzheimer

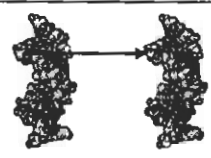
- Formes à transmission autosomique dominante
  - moins de 2% des formes
  - essentiellement à début précoce *devenue : 27 ans!*
  - mutations sur les gènes
    - Précurseur du peptide Amyloïde (17 mutations)
    - Préséniline 1 (80 mutations)
    - Préséniline 2 (8 mutations)
- Formes à transmission non mendéliennes (sporadique) : *Com. plé. de*
  - plus de 90% des cas
  - essentiellement à début tardif
  - Polygénique et multifactorielle
  - Un déterminant génétique majeur, le gène de l'apolipoprotéine E

## Composante Héritaire d'une maladie

### Composantes Génétiques des Maladies

<p><b>Formes Autosomiques Dominantes</b></p>	<p><b>Formes « Sporadiques »</b> Hérédité complexe</p>
--	--

- |  |   |
|--|---|
| <ul style="list-style-type: none"> <li>• Formes familiales à début précoce</li> <li>• Mutations associées à la maladie</li> <li>• Rare</li> <li>• Transmission monogénique mendélienne dominante</li> <li>• Effet individuel important</li> <li>• Faible impact en population</li> <li>• Peu d'interaction avec les facteurs environnementaux</li> </ul> | <ul style="list-style-type: none"> <li>• Formes sporadiques à début tardif</li> <li>• Facteurs de susceptibilité génétique</li> <li>• Commune</li> <li>• Transmission polygénique complexe</li> <li>• Effet individuel faible <i>→ Risque ?</i></li> <li>• Fort impact en population</li> <li>• Modification de l'impact des facteurs environnementaux</li> </ul> |
|--|---|

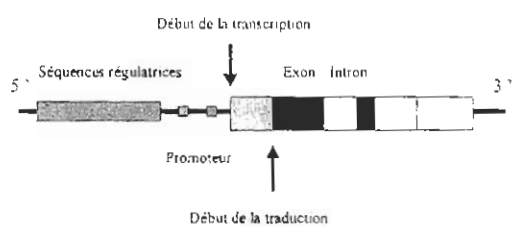


## Polymorphismes et mutations

Florence RICHARD  
MCU-PH, Faculté de Médecine,  
Lille 2



## Gène



## Polymorphisme de l'ADN

- La molécule d'ADN comporte une variation tous les 2 à 300 nucléotides : polymorphisme génétique
- Définition d'un polymorphisme : existence dans une même population de deux formes ou plus de la même séquence d'ADN

Exemple :  
CCG-ATG-GGT  
CCT-ATG-GGT

## Types de polymorphisme de l'ADN

- Insertion/délétion
- Mutations ponctuelles
- Séquences répétées en tandem (VNTR)
  - Satellites
  - Minisatellites
  - Microsatellites

F. Richard

55

## Types de polymorphisme de l'ADN : Insertion-délétion

AGTCTTGC → AGTCTTTTGC : Exemple d'insertion

AGTCTTGC → AGTC--GC : Exemple de délétion

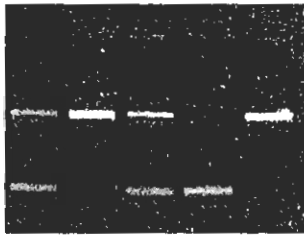
F. Richard

56

## Polymorphisme d'Insertion/délétion du Gène de l'ECA

490 bp

190 bp



ID II ID DD II

57

## Types de polymorphisme de l'ADN : Mutation ponctuelle dans la partie codante

Exemple de mutation synonyme (mutation silencieuse):

CGG-ATG-GGT      CGG : Arginine

CGA-ATG-GGT      CGA : Arginine

F. Richard

58

## Types de polymorphisme de l'ADN : Mutation ponctuelle dans la partie codante

Exemple de mutation non synonyme :

CGG-ATG-GGT      CGG : Arginine

CCA-ATG-GGT      CCA : Proline

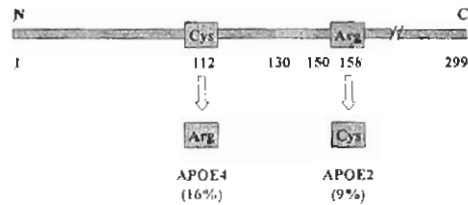
F. Richard

59

## Exemple de Polymorphisme avec double mutation ponctuelle dans la partie codante : Gène de l'APOE

APOE3 (75%)

Site de liaison aux récepteurs



F. Richard

Rall et al., 1982

## Types de polymorphisme de l'ADN : Séquences répétées en tandem

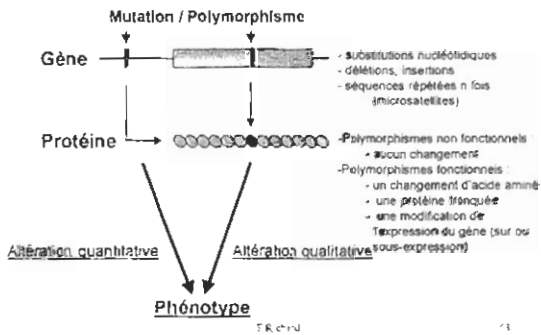
Exemple :

CAG CGG CGG  
 CAG CGG CGG CGG  
 CAG CGG CGG CGG CGG  
 CAG CGG CGG CGG CGG CGG

## Mécanismes de génération de polymorphismes

- Anomalie lors de crossing-over
- Altération d'origine physique ou chimique de l'ADN + Erreurs du système de réparation
- Insertions de séquences mobiles ou virales
- ...

## En résumé : les polymorphismes génétiques



## Génétique des Populations

Florence RICHARD



## Génétique des Populations

### Génétique des populations : Définition

Etude de la fréquence des gènes et des génotypes dans les populations, ainsi que des facteurs qui peuvent les modifier au cours des générations

## Génétique des Populations

### Génétique des populations : Lien avec l'évolution

**Evolution** : Changement dans les fréquences génétiques au sein des populations de génération en génération

La génétique des populations est "intimement" liée à l'histoire de l'humanité.

### Facteurs pouvant modifier les fréquences génétique dans une population

- Mutation
- Sélection naturelle
- Dérive génétique
- Brassage génétique

### Sélection naturelle

- La sélection naturelle est le processus naturel d'élimination des individus d'une population les moins adaptés à leur environnement. Ceux-ci disparaissent peu à peu tandis que les autres survivent et se reproduisent, transmettant un patrimoine génétique avantageux à leurs descendants.
- Exclusion différentielle de certains génotypes de la reproduction

### Génétique des Populations Exemple de sélection naturelle avec influence du milieu

#### Réflectance des populations et Populations



Réflectance de la peau = % de lumière incidente renvoyée par la peau

### Sélection naturelle

- La carte précédente est représentation schématique de la variabilité de la réflectance dans diverses populations : elle est en réalité variable au sein même des populations
- Le réflectance semble grossièrement obéir à la latitude ou à l'ensoleillement (cas de l'Europe)

### Facteurs pouvant modifier les fréquences alléliques dans une population

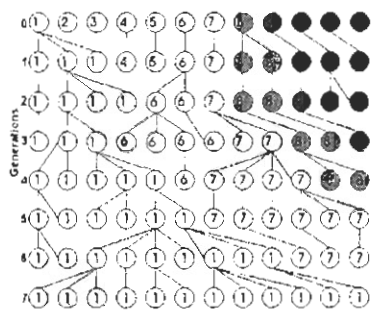
- Mutation
- Sélection naturelle
- Dérive génétique
- Brassage génétique

### Dérive Génétique

- Phénomène purement aléatoire, lié aux fluctuations d'échantillonnage qui va modifier dans une population les fréquences alléliques
- N'est pas le résultat d'une sélection naturelle, d'une mutation, ni d'une migration
- La génération qui remplace celle qui l'a générée présente du fait des hasards de la méiose et de la fécondation des fréquences alléliques différentes.
- Différent de la sélection naturelle (dans lequel va intervenir l'environnement)

## Exemple de Dérive Génétique

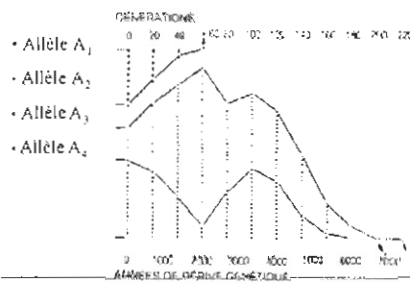
Au départ : 6 Individus - 12 allèles



Les individus porteurs de l'allèle 1 se reproduisent plus

76

## Exemple de Dérive Génétique



Simulation génétique de la descendance sur 220 générations d'une population stable et isolée de 21 individus: les allèles disparaissent les uns après les autres. Il n'en restera qu'un seul (A<sub>1</sub>); on dit qu'il est fixé

## Dérive Génétique

- Si l'on a affaire à une population de grande taille les fréquences alléliques varieront peu d'une génération à l'autre par l'effet du grand nombre de géniteurs. On peut même considérer que sur un temps assez court elles sont stables : absence de dérive génétique.
- Il en est tout autrement d'une population de petite taille dans lequel vont un effet de dérive génétique a plus de probabilité d'arriver.
- Parmi les principales causes de dérives génétiques : l'effet fondateur et l'embouteillage.

## Dérive Génétique

- Les principales causes de dérives génétiques :
  - L'effet fondateur
  - L'effet embouteillage.

F. Richard

75

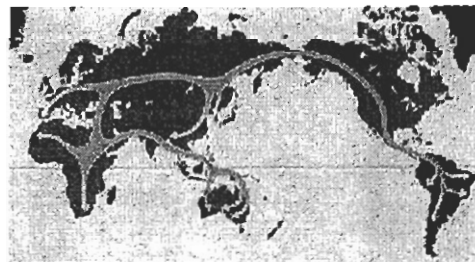
## Dérive génétique : Effet Fondateur

- Lors de son expansion progressive sur tous les continents l'humanité était peu nombreuse (tout au plus quelques centaines de milliers d'Hommes).
- Cette expansion s'est faite par l'émigration de petits groupes issus d'une population plus importante.
- Le groupe ou les groupes d'émigrants constitués de quelques dizaines d'individus représentent un échantillonnage des individus du groupe souche, généralement plus nombreux.

F. Richard

77

## Migration des Homosapiens : Une origine africaine



F. Richard

78

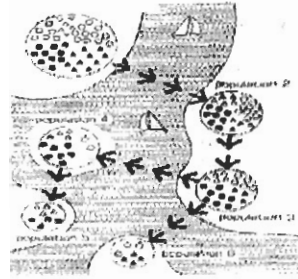
## Dérive génétique : Effet Fondateur

- Cet échantillonnage, effectué au hasard (il n'existe pas de gène de la migration !), n'est pas fidèle aux fréquences alléliques de la population souche.
- Il est même dépourvu de certains allèles, généralement ceux qui sont faiblement représentés dans la population d'origine. Cependant un allèle peu fréquent dans le groupe souche peut être sur-représenté dans le groupe de migrants.

F. Richard

79

## Effet fondateur = effet d'échantillonnage



F. Richard

80

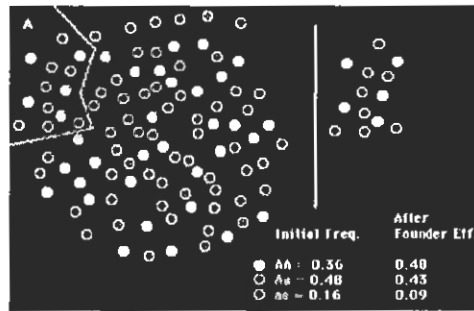
## Dérive génétique : Effet Fondateur

- Si l'on a affaire à une population de grande taille les fréquences alléliques varieront peu d'une génération à l'autre par l'effet du grand nombre de géniteurs. On peut même considérer que sur un temps assez court elles sont stables.
- Il en est tout autrement d'une population de petite taille : groupe de migrants sans brassage génétique
- Donc l'effet fondateur va atteindre des populations isolées de taille limitée.

F. Richard

81

## Effet fondateur = effet d'échantillonnage



F. Richard

82

## Dérive génétique : Effet Fondateur



F. Richard

83

## Exemple d'Effet Fondateur : les Amish de Pennsylvanie

- Communauté religieuse très isolée, les Amish de Pennsylvanie
- Groupe qui compte une dizaine de milliers de membres, issu de trois couples ayant émigré aux États-Unis en 1770.
- Forte fréquence (0.07) du syndrome d'Ellis-van-Creveld, une maladie récessive du développement particulièrement rare ailleurs dans le monde.
- L'interprétation la plus plausible est de penser que l'un des ancêtres fondateurs était porteur de ce trait (probablement à l'état hétérozygote) et qu'il l'a transmis à sa descendance.
- On remarquera que l'effet de la dérive peut aussi contrecarrer l'effet de la sélection dans des populations d'effectif réduit. Les gènes défavorables peints par hasard à des fréquences relativement élevées durant les premières générations seront ensuite contre sélectionnés, mais leur élimination pourra être assez lente.

F. Richard

84

Génétique des Populations **Dérive génétique**  
Exemple d'Effet Fondateur : Maladie de Huntington

- Région du lac Maracaibo (Venezuela)
- Haute prévalence (environ 190 personnes dans la zone concernée) de la maladie de Huntington (maladie autosomique dominante)
- Toutes les victimes sont des descendants d'une femme nommée Maria Concepcion Soto qui arriva dans la région au cours du 19ème siècle.

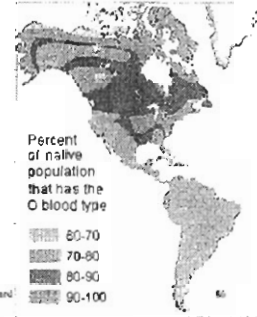


F. Richard

85

Génétique des Populations **Dérive génétique**  
Exemple d'Effet Fondateur : Groupe Sanguin dans la population des indiens d'Amérique

- Dans la population native, des indiens
- It is also possible to find the results of the founder effect even though the original ancestors are unknown. For example, South and Central American Indians were nearly 100% type O for the ABO blood system. Since nothing in nature seems to strongly select for or against this trait, it is likely that most of these people are descended from a small band of closely related "founders" who also shared this blood type. They migrated into the region from the north, probably by the end of the

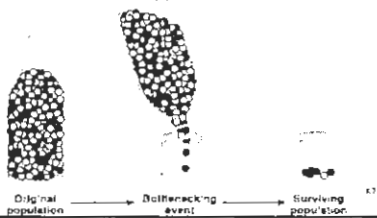


F. Richard

86

Génétique des Populations **Dérive génétique : l'effet embouteillage**

- L'effet embouteillage
  - Inéresse des populations de grande taille
  - soumise à un désastre (catastrophe naturelle, famine, ...)
  - Par hasard, parmi les survivants, un (ou plusieurs) allèles vont être sur ou sous représentés comparés à la population de départ.



Génétique des Populations **Dérive génétique : l'effet embouteillage**

- L'effet embouteillage
  - La dérive génétique continue jusqu'à ce que la population redevienne assez importante pour minimiser le risque de fluctuation d'échantillonnage.



Génétique des Populations **Facteurs participant à l'évolution**

- Mutation
- Sélection naturelle
- Dérive génétique
- Brassage génétique

F. Richard

89

Génétique des Populations **Migration**

- Au sein d'une espèce, flux de gènes d'une population à une autre que ce soit sous forme d'individus ou de gamètes
- Limite les effets fondateurs
- Favorise le maintien de la variabilité par le brassage génétique

F. Richard

90

### Brassages Génétiques

- Ils empêchent l'isolement génétique.
- Ralentissent la dérive génétique en réintroduisant des allèles dans la population
- Résultat de migration

### Brassage Génétique



### Pressions évolutives modifiant les fréquences alléliques

- Mutation
  - Sélection naturelle
  - Dérive génétique
  - Brassage génétique
- Ces 4 éléments vont interagir ensemble.

### Des histoires de migration



### Des flux de gènes et du brassage génétique

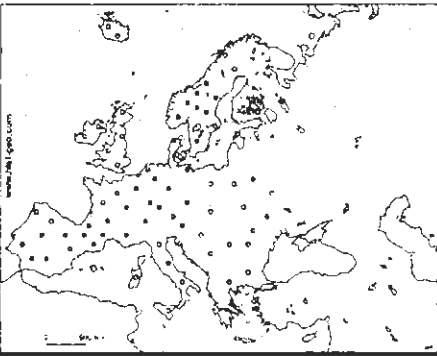


### Des flux de gènes et du brassage génétique

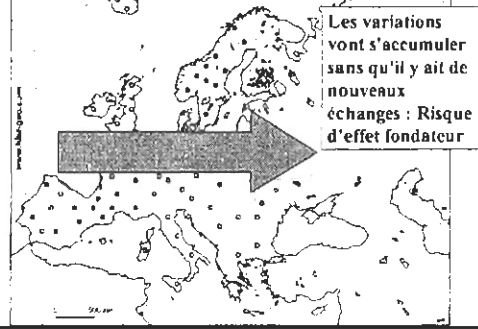




## Des flux de gènes et du brassage génétique



## Mais dans des populations discontinues : le brassage génétique va être bloqué



## Précision

Le reste de l'exposé ne concerne que le cas des polymorphismes situés sur un gène autosomal, avec une transmission mendélienne.

## La Loi de Hardy-Weinberg : "Population génétiquement stable"

## En absence d'évolution :

- Les fréquences alléliques et génotypiques restent inchangées au fil des générations
- On peut prévoir les fréquences génotypiques à partir de la connaissance des fréquences alléliques et vice versa.

## Hypothèse de la Loi de Hardy-Weinberg

- Population d'une grande dimension
- Panmixie (les mariages se font au hasard. Ex organisme asexué avec reproduction externe)
- Les hommes et les femmes ont des fréquences alléliques similaires
- Absence de migration
- Absence de mutation (Habituellement, taux de mutation  $< 10^{-5}$  /génération)
- Absence de consanguinité
- Absence de dérive génétique
- Absence d'effet fondateur

## Fréquences Génotypiques

Soit une Locus X, biallélique : A et a

Génotypes	N	p
AA	$n_{AA}$	$n_{AA}/N$
Aa	$n_{Aa}$	$n_{Aa}/N$
aa	$n_{aa}$	$n_{aa}/N$

## Comptage des allèles

- Chaque individu a deux allèles

⇒ Dans une population de N individus :

$2 \times N$  allèles.

- Nombres d'allèles A et a

$$A = 2n_{AA} + n_{Aa}$$

$$a = 2n_{aa} + n_{Aa}$$

F. Richard

103

## Calcul des fréquences alléliques

- Dans une population de N individus

- Nombres d'allèles A et a

$$A = 2n_{AA} + n_{Aa}$$

$$a = 2n_{aa} + n_{Aa}$$

- Calcul des fréquences

$$p_A = p = (2n_{AA} + n_{Aa})/2N$$

$$p_a = q = (2n_{aa} + n_{Aa})/2N$$

F. Richard

104

## Génétique des Populations

### La Loi de Hardy-Weinberg: Exemple

Au locus, on observe 2 allèles A et a de fréquences alléliques respectives en population générale p et q

» où  $p + q = 1$

- Si les Fréquences génotypiques :

$$P_{AA} \quad P_{Aa} \quad P_{aa}$$

$$p^2 \quad 2pq \quad q^2$$

» où  $p^2 + 2pq + q^2 = 1$

Alors la population est dite en équilibre de Hardy-Weinberg

F. Richard

105

## Génétique des Populations

### Tester l'équilibre de Hardy-Weinberg dans une population

$H_0$  : la population est en équilibre de Hardy-Weinberg

$H_1$  : la population n'est pas en équilibre de Hardy-Weinberg

Cela va revenir à comparer une distribution observée

$P_{AA}$        $P_{Aa}$        $P_{aa}$   
à  
une distribution théorique  $\Rightarrow \chi^2$

$$p^2 \quad 2pq \quad q^2$$

F. Richard

106

## Génétique des Populations

### Tester l'équilibre de Hardy-Weinberg dans une population de N sujets

Exemple : pour un marqueur biallélique, 2 allèles A et a.

On observe

$n_{AA}$  sujets porteurs du génotype AA

$n_{Aa}$  sujets porteurs du génotype Aa

$n_{aa}$  sujets porteurs du génotype aa

- 1ère étape : Comptage des allèles
- 2ème étape : Calcul des fréquences alléliques p et q

F. Richard

107

## Génétique des Populations

### Tester l'équilibre de Hardy-Weinberg dans une population de N sujets

- 3ème étape : En déduire les fréquences génotypiques attendues sous l'hypothèse d'Hardy-Weinberg :

$$AA \quad Aa \quad aa$$

$$p^2 \quad 2pq \quad q^2$$

- 4ème étape : En déduire un nombre attendu de sujets porteurs respectivement des génotypes

$$AA \quad Aa \quad aa$$

$$p^2 \times N \quad 2pq \times N \quad q^2 \times N$$

F. Richard

108

### Tester l'équilibre de Hardy-Weinberg dans une population de N sujets

- 5ème étape: Comparer ces nombres attendus aux nombres observés.

$$\begin{array}{ccc} n_{AA} & n_{Aa} & n_{aa} \\ p^2 \times N & 2pq \times N & q^2 \times N \end{array}$$

$$\chi^2 = \sum \frac{(\text{Attendu} - \text{observé})^2}{\text{Attendu}} \text{ssi le plus petit } C \geq 5$$

- Attention : il n'y a pas de réel consensus sur le nombre de ddl. Pour certains : nombre d'allèles - 1, pour d'autres : nombre de génotypes - 1
- Le plus souvent, on espère ne pas rejeter H0, dans ce cas, on peut faire l'hypothèse que la population est en équilibre de Hardy-Weinberg

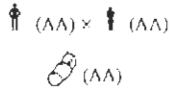
### Déviations de la Loi de Hardy-Weinberg

- Petite population
- Taux élevé de mutation dans une population
- Sélection d'une ou d'une combinaison de génotypes (la sélection peut être positive ou négative)
  - Car génotype ou combinaison non viable (dépens ultra-ultra)

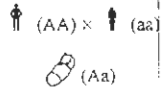
### Déviations de la Loi de Hardy-Weinberg

#### Des mariages qui ne se font pas au hasard

- Autogamie : Les époux se ressemblent (↗ homozygotes)



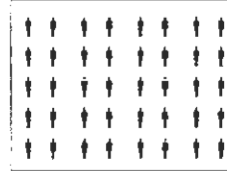
- Hétérogamie : les époux ne se ressemblent pas (↗ hétérozygotes)



- Consanguinité (↗ homozygotes)

### Déviations de la Loi de Hardy-Weinberg

#### Migration Massive dans une population génotypiquement différente



Population d'accueil 1 : Tous les individus sont AA



Population immigrante 2 : Tous les individus sont aa

A la 1ère génération, pas de HW, en effet absence d'hétérozygote Aa  
Le plus souvent, équilibre de HW atteint après quelques générations.

### Equilibre de Hardy-Weinberg : Un exemple

Association at LRP gene locus with sporadic late-onset Alzheimer's disease.

Lambert et al. Lancet, 1998;351:1787

### LRP gene

- Polymorphisme silencieux
- Deux allèles : C et T, situé dans l'exon 3.
- Etude cas-témoins
- On s'intéresse au groupe témoin.

Fréquences Génotypiques

LRP dans le groupe témoin

Génotypes	N	p	p
CC	407	407/596	0.68
CT	168	168/596	0.28
TT	21	21/596	0.04
<b>Total</b>	<b>596</b>	<b>1.0</b>	<b>1.0</b>

0,77  
0,21  
0,02

558

Comptage des allèles

- Dans une population de 596 individus :  
 $2 \times 596$  soit 1192 allèles.
- Nombres d'allèles C et T  
 $C = 2n_{CC} + n_{CT} = [(2 \times 407) + 168] = 982$   
 $T = 2n_{TT} + n_{CT} = [(2 \times 21) + 168] = 210$
- Remarque  
 $982 + 210 = 1192$  allèles.

975  
161  
1116

Calcul des fréquences alléliques

- $P_C = p = (2n_{CC} + n_{CT}) / 2N = 982 / 1192 = 0.82$
- $P_T = q = (2n_{TT} + n_{CT}) / 2N = 210 / 1192 = 0.18$
- Remarque :  $p + q = 0.82 + 0.18 = 1$

0,87  
0,13

Tester l'équilibre de Hardy-Weinberg dans une population

- $H_0$  : la population est en équilibre de Hardy-Weinberg
- $H_1$  : la population n'est pas en équilibre de Hardy-Weinberg

Cela va revenir à comparer une distribution observée

$P_{CC}$	$P_{CT}$	$P_{TT}$
à		
une distribution théorique		
$p^2$	$2pq$	$q^2$

Tester l'équilibre de Hardy-Weinberg dans une population de N sujets

- Déduire les fréquences génotypiques attendues sous l'hypothèse d'Hardy-Weinberg :

CC	CT	TT
$p^2$	$2pq$	$q^2$
$0.82^2$	$2 \times 0.82 \times 0.18$	$0.18^2$
0.6724	0.2952	0.0324

- Remarque :  $p^2 + 2pq + q^2 = 0.6724 + 0.2952 + 0.0324 = 1$

0,7569    0,2262    0,0169

Tester l'équilibre de Hardy-Weinberg dans une population de N sujets

- Déduire un nombre attendu de sujets porteurs respectivement des génotypes

CC	CT	TT
$p^2 \times N$	$2pq \times N$	$q^2 \times N$
$0.6724 \times 596$	$0.2952 \times 596$	$0.0324 \times 596$
400.75	175.94	19.31

- Remarque :  $400.75 + 175.94 + 19.31 = 596$

844,70    292,64    19,86

422,35    126,82    9,63 > S  
20,1  
 $0,075 + 0,057 + 0,261 = 0,39$

## Tester l'équilibre de Hardy-Weinberg dans une population de N sujets

- Comparer les nombres attendus aux nombres observés.

Génotypes	CC	CT	TT
Observés :	407	168	21
Attendus sous HW :	400,75	175,94	19,31

$$\chi^2 = \sum \frac{(\text{Attendu} - \text{observé})^2}{\text{Attendu}}$$

Ici le plus petit attendu = 19,31 donc  $\geq 5$ .

F. Richard

121

Lancet, 1998, 351:1787.

## Calcul du Khi<sup>2</sup>

$$\chi^2 = \sum \frac{(\text{Attendu} - \text{observé})^2}{\text{Attendu}} \text{ ssi le plus petit } C \geq 5$$

$$\chi^2 = \sum \frac{(407-400,75)^2}{400,75} + \frac{(168-175,94)^2}{175,94} + \frac{(21-19,31)^2}{19,31}$$

$$\chi^2 = 0,60 \Rightarrow 0,39$$

0,60 à comparer à la valeur seuil d'un  $\chi^2$  à 1ddl : 3,841

0,60 < 3,841 donc non rejet de  $H_0$

F. Richard

122

Lancet, 1998, 351:1787

## Identifier les facteurs génétiques

Florence RICHARD  
MCU-PH, Lille2



F. Richard

## Identification des facteurs génétiques

### Comment ?

- ① Etudes d'association
- ② Etudes de liaison
- ③ Etude d'association-liaison

F. Richard

124

## Etudes d'association

Florence RICHARD  
MCU-PH, Lille2

F. Richard

125

## Etudes d'association en population

- Intéressent des populations de sujets indépendants : sans liens familiaux
- Font appel aux techniques classiques de l'épidémiologie
- Les plus fréquentes : Etudes cas-témoins - Comparaison de la distribution d'un marqueur génétique dans :
  - Groupe de sujets atteints de la maladie : cas
  - Groupe de sujets indemnes de la maladie : témoins

F. Richard

126

## Etudes d'association en population

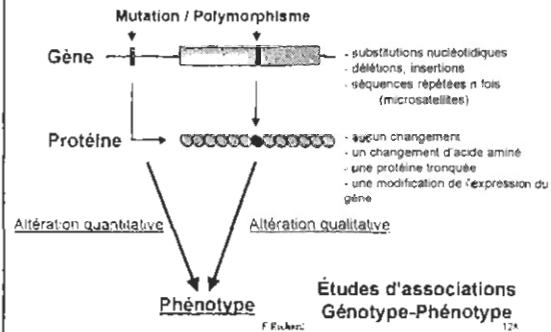
Repose sur :

- Des études cas-témoins (le plus souvent)
- Des polymorphismes
- Des déséquilibres de liaison
- Une stratégie Gène-Candidat

F. Richard

127

## Les polymorphismes génétiques



F. Richard

128

## Le Déséquilibre de Liaison Gamétique

Exemple de deux polymorphismes bialléliques situés sur deux locus  $\alpha$  et  $\beta$



F. Richard

129

## Le Déséquilibre de Liaison Gamétique

	Fréquences gamétiques ou chromosomiques		Fréquences alléliques
	A	a	
B	$P(AB)$	$P(aB)$	$P(B)$
b	$P(Ab)$	$P(ab)$	$P(b)$
Fréquences alléliques	$P(A)$	$P(a)$	

F. Richard

130

## L'Equilibre de Liaison Gamétique

Si la distribution des allèles des locus a et b se fait au hasard, on s'attend à avoir

	$P(AB)$	=	$P(A) \times P(B)$
Et	$P(Ab)$	=	$P(A) \times P(b)$
Et	$P(aB)$	=	$P(a) \times P(B)$
Et	$P(ab)$	=	$P(a) \times P(b)$

F. Richard

131

## Le Déséquilibre de Liaison Gamétique

Si la distribution des allèles des locus a et b ne se fait au hasard

	$P(AB)$	$\neq$	$P(A) \times P(B)$
Ou	$P(Ab)$	$\neq$	$P(A) \times P(b)$
Ou	$P(aB)$	$\neq$	$P(a) \times P(B)$
Ou	$P(ab)$	$\neq$	$P(a) \times P(b)$

F. Richard

132

## Le Déséquilibre de Liaison Gamétique

Si la distribution des allèles des locus a et b ne se fait au hasard, le déséquilibre de liaison va pouvoir être "mesuré" par D

D = différence entre les fréquences chromosomiques observées et les fréquences chromosomiques théoriques

$$D(AB) = P(AB) - P(A) \times P(B)$$

$$D(Ab) = P(Ab) - P(A) \times P(b)$$

$$D(aB) = P(aB) - P(a) \times P(B)$$

$$D(ab) = P(ab) - P(a) \times P(b)$$

131

## Le Déséquilibre de Liaison Gamétique

Interprétation

Si  $D(AB) > 0$  A et B sont associés plus souvent que ne le voudrait le hasard

Si  $D(AB) < 0$  A et B sont associés moins souvent que ne le voudrait le hasard

Si  $D(AB) = 0$  Pas de déséquilibre de liaison

F. Richard

134

## Le Déséquilibre de Liaison Gamétique

Interprétation

$D(AB) \neq 0$  quand :

- Sélection de la population (ex : l'association allèle A et B est létale ou non viable)
- Mélange de deux populations (une population AB et ab) : ce déséquilibre va diminuer au fur et à mesure des générations sauf si les deux locus sont tellement proche l'un de l'autre que ce déséquilibre va se maintenir sur plusieurs générations
- Si a et b sont proches physiquement

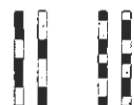
F. Richard

135

## Le Déséquilibre de Liaison Gamétique



Un individu : 2 exemplaires du même chromosome



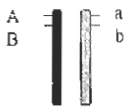
Exemple de 6 gamètes possiblement produits par l'individu

F. Richard

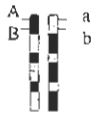
136

## Le Déséquilibre de Liaison Gamétique

Si les locus A et B sont proches l'un de l'autre : ils se transmettront plus souvent ensemble : effet ancestral



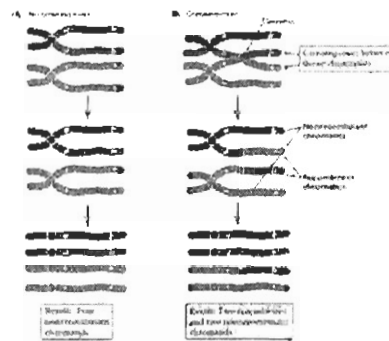
Un Individu : 2 exemplaires du même chromosome



Exemple de 6 gamètes possiblement produits par l'individu

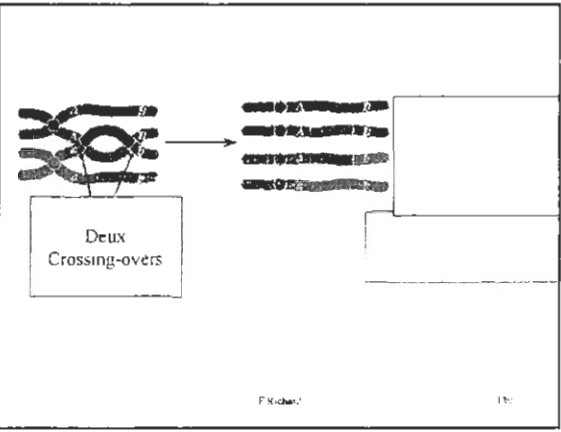
F. Richard

137



F. Richard

138



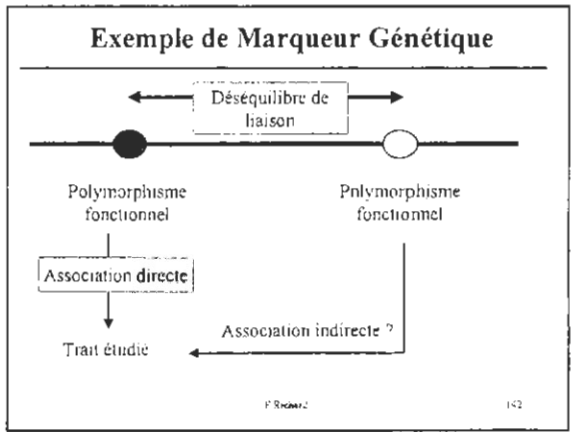
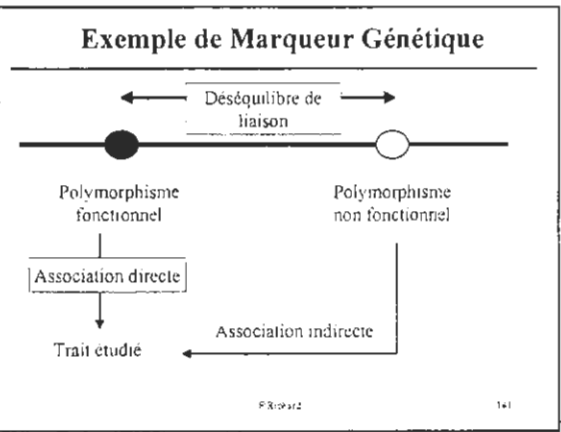
### Déséquilibre de liaison : exemple

Recherche de polymorphismes

numéro	position	localisation	changement	ref.	fréquence
1	-1358	promoteur	C/T	rs708286	47.1%
2	-1132	promoteur	T/C	rs202892	41%
3	-849	promoteur	G/C	rs203985	43.0%
4	-787	promoteur	G/A	rs203984	13.1%
5	-761	promoteur	G/T	rs202894	nc*
6	-738	promoteur	A/	rs202895	12.3%
7	-500	promoteur	T/C	rs202896	23.5%
8	+2871	intron 1	G/T	rs2186358	8.8%
9	+2786	exon 2	G/C	rs1551169	28.8%
10	+2821	intron 2	T/C	rs202897	3.4%
11	+2837	intron 3	A/G	rs2186356	9.2%
12	+2843	intron 7	G/A	rs2039355	20.3%
13	+3079	intron 7	T/A	rs202898	17.4%
14	+3255	intron 8	G/A	rs202899	41%
15	+3492	intron 9	G/A	rs2039353	13.2%
16	+3562	intron 9	G/A	rs20289405	33.8%
17	+3916	intron 9	C/T	rs20289404	35.8%
18	+3942	intron 9	A/G	rs20289403	27.8%
19	+4416	intron 9	G/A	rs20289402	34.8%
20	+4831	intron 9	C/T	rs20289401	41%
21	+5039	intron 9	A/G	rs2186357	9.8%
22	+5129	intron 9	C/T	rs2039354	32.1%
23	+5185	intron 9	G/A	rs2039352	34.9%
24	+5767	3'-UTR	C/T	rs2186359	13.1%
25	+5769	3'-UTR	C/A	rs2028990	41%

Analyse des déséquilibres de liaison (Logiciel Gold)

F. Reichert 141

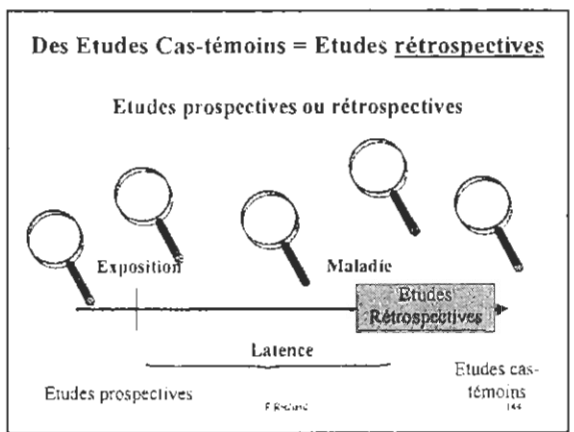


### Des Etudes Cas-témoins = Etudes rétrospectives

Cas : sujets atteints de la maladie

Témoins : sujets indemnes de la maladie, représentatif de la population dont sont issus les cas

F. Reichert 144





## Etudes Cas-témoins

- Description détaillée :
  - La définition de la maladie utilisée
  - Recrutement des cas :
    - Cas incident ou prévalent
    - Cas hospitaliers, en ville, registre
    - Cas consécutifs, tirage au sort,
    - Origine ethnique de la population
- Description détaillée des caractéristiques des cas :
  - Age, sexe,
  - Autres (dépend de la maladie étudiée)

F. Richard

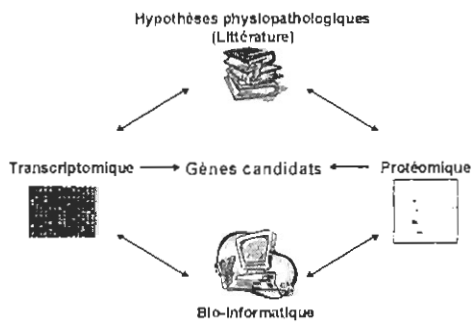
145

## Quel(s) Gène(s) ?

- Criblage systématique
  - Mais environ 25 000 gènes et 14 million de polymorphismes SNP.
- Stratégie Gène Candidat
  - Définition d'un Gène candidat :
    - Gène qui par sa fonction peut être impliqué dans le processus étiopathogénique de la maladie
    - Gène codant pour des protéines susceptibles d'être impliquées dans la physiopathologie de la maladie
  - Ex: Gènes impliqués dans le métabolisme des lipides pour les maladies cardiovasculaires

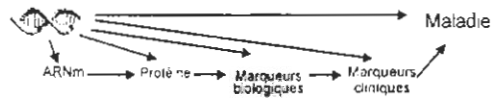
146

## Sources de gènes candidats



## Etude des phénotypes intermédiaires

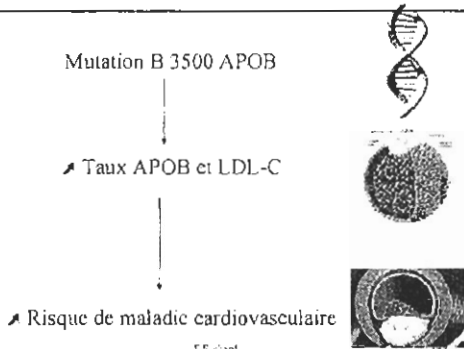
Il est important afin de valider l'hypothèse d'un gène candidat de pouvoir associer un polymorphisme non seulement à la maladie mais également à des phénotypes intermédiaires



F. Richard

148

## Phénotype intermédiaire : un exemple



F. Richard

## Etudes Cas-témoins : Mesure du risque

Polymorphisme biallélique: Allèle A et a (2n allèles) ·  
Etude des fréquences alléliques

	Cas	Témoins
Fréquence allélique A	$p_1$	$p_2$
Fréquence allélique a	$q_1$	$q_2$

$p_1 \neq p_2 ? \rightarrow$  Test du  $\chi^2$  à un ddl

F. Richard

150

## Etudes Cas-témoins : Mesure du risque

Polymorphisme bi-allélique: Allèle A et a : Etude des fréquences génotypiques

	Cas	Témoins
AA	a	b
Aa	c	d
aa	e	f
Total	$n_1$	$n_2$

F. Richard

131

## Etudes Cas-témoins : Mesure du risque

Polymorphisme bi-allélique: Allèle A et a : Etude des fréquences génotypiques ( $f_i$ )

	Cas		Témoins	
	n	$f_i$	n	$f_i$
AA	a	$a/n_1$	b	$b/n_2$
Aa	c	$c/n_1$	d	$d/n_2$
aa	e	$e/n_1$	f	$f/n_2$
Total	$n_1$		$n_2$	

La distribution des fréquences génotypiques est-elle différentes entre les cas et les témoins? → Test du  $\chi^2$  à 2ddl

132

## Etudes Cas-témoins : Mesure du risque

**Odds Ratio (OR)** : Mesure du risque chez les sujets porteurs du génotype AA

Groupe de référence : aa

$$OR = \frac{af}{be}$$

F. Richard

133

## Etudes Cas-témoins : Mesure du risque

Intervalle de confiance à 95% de l'OR

**Méthode exacte**

les sujets porteurs du génotype AA vs aa

$$IDC\ 95\% = OR e^{\left( \pm 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{e} + \frac{1}{f}} \right)}$$

F. Richard

134

## Etudes Cas-témoins : Mesure du risque

**Odds Ratio (OR)** : Mesure du risque chez les sujets porteurs du génotype Aa

Groupe de référence : aa

$$OR = \frac{cf}{de}$$

F. Richard

135

## Etudes Cas-témoins : Mesure du risque

Intervalle de confiance à 95% de l'OR

**Méthode exacte**

les sujets porteurs du génotype Aa vs aa

$$IDC\ 95\% = OR e^{\left( \pm 1.96 \sqrt{\frac{1}{c} + \frac{1}{d} + \frac{1}{e} + \frac{1}{f}} \right)}$$

F. Richard

136

## Etudes Cas-témoins = OR Interprétation

- OR = 1 ( ]1- + ∞[ ) Association entre le génotype et la maladie
- OR > 1 ( ]1- + ∞[ ) Association entre le génotype et la maladie. Le génotype est plus fréquent chez les cas que chez les témoins (Facteur de risque?)
- OR < 1 ( ]0-1[ ) Association entre le génotype et la maladie. Le génotype est moins fréquent chez les cas que chez les témoins (Facteur protecteur?)

F. Richard

157

## Etudes Cas-témoins : Mesure du risque


Polymorphisme biallélique: Allèle A et a

	« Dominant »	« Récessif »	« Codominant »
AA	OR > 1	OR > 1	OR > 1
Aa	OR > 1	OR = 1	OR > 1
aa	1	1	1
	$OR_{AA} \approx OR_{Aa}$		$OR_{Aa} < OR_{AA}$
	$\chi^2 \text{ à } 1 \text{ddl}$	$\chi^2 \text{ à } 1 \text{ddl}$	$\chi^2 \text{ à } 1 \text{ddl}$

F. Richard

158

## Quand un OR Est-il différent de 1 ?

- Le plus souvent, on conclut avec un risque alpha à 0.05 ( $p \leq 0.05$ )
-  Qu'est ce qu'un risque alpha à 0.05?
- Lorsque l'on conclut qu'un OR est significativement différent de 1, on se trompe 5 fois sur 100

F. Richard

159

## Comment savoir si un OR est différent de 1 ?

### Deux méthodes :

- Test du  $\chi^2$  ou régression logistique
- Intervalle de confiance à 95%

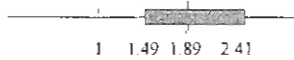
F. Richard

160

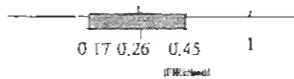
## Etudes Cas-témoins : IDC95%

Exemple d'OR significativement  $\neq 1$  avec l'IDC à 95%

- OR = 1.89 IDC95% [1.49-2.41]



- OR = 0.26 IDC95% [0.17-0.45]



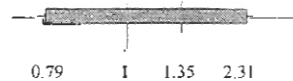
F. Richard

161

## Etudes Cas-témoins : IDC 95%

Exemple d'OR non significativement  $\neq 1$  avec l'IDC à 95%

- OR = 1.35 IDC95% [0.79-2.31]



F. Richard

162

## Études Cas-témoins : Les biais

**Définition** Ecart ou erreur présentant une direction privilégiée avec comme conséquence :

- Diminution de la puissance de l'étude
- Sous-estimation ou surestimation d'une association
- Comme dans n'importe quel type d'étude épidémiologique, on essaiera d'éviter les biais.

F. Richard

103

## Études Cas-témoins : Les Principaux Types de Biais

### ① Biais de sélection

Résulte de la façon dont l'échantillon a été choisi au sein de la population

### ② Biais de classement

Erreurs dans la mesure de l'exposition et/ou de la maladie

### ③ Facteurs de confusion

Liés à l'influence de tiers facteurs sur l'association entre l'exposition et la maladie

F. Richard

104

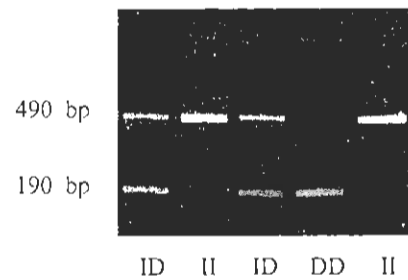
## Biais et Particularités de l'utilisation de marqueurs génétiques dans les études cas-témoins : les témoins.

- Les témoins sont-ils de "bons" témoins?
  - Si oui : population en équilibre de Hardy-Weinberg
- Leurs fréquences alléliques et génotypiques doivent être celles attendues dans la population dont ils sont issus (Ex de l'ACT).

F. Richard

105

## Polymorphisme du Gène de l'ECA Détermination du génotype

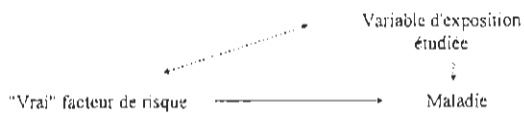


F. Richard

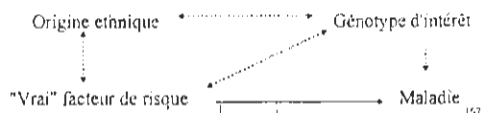
106

## Biais et Particularités de l'utilisation de marqueurs génétiques dans les études cas-témoins : Biais de Stratification de population

### Facteur de Confusion



### Biais de stratification



D'après INCI, 2000, 92-1151

F. Richard

107

## Biais et Particularités de l'utilisation de marqueurs génétiques dans les études cas-témoins : Biais de Stratification de population

### Haplotype HLA

Population amérindienne PIMA		Population caucasienne	
Prévalence de DNID = 40%		Prévalence de DNID = 15%	
Gm3,5,13,14	+	Gm3,5,13,14	+
≈ 1%	≈ 69%	≈ 66%	≈ 34%

### Etude sans prendre en compte l'origine ethnique des sujets

Haplotype	Cas	Témoins	OR = 0.27 [0.18-0.40]
Gm3,5,13,14			
+	7.8%	29.0%	
-	92.2%	71.0%	

D'après Lancet, 2003, 361:598

F. Richard

108

**Biais et Particularités de l'utilisation de marqueurs génétiques dans les études cas-témoins : Biais de Stratification de population**

**Haplotype HLA**

Population amérindienne		Population caucasienne	
Prévalence de DNID ≈ 40%		Prévalence de DNID ≈ 15%	
Gm3,5,13,14	-	Gm3,5,13,14	-
≈1%	≈99%	≈66%	≈34%

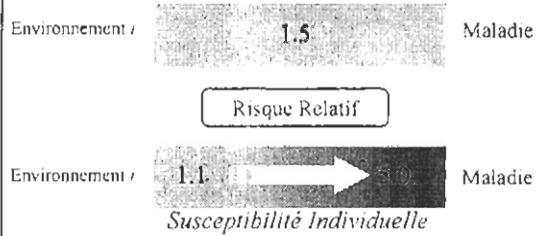
Disparition de l'association lorsque l'on restreint l'analyse à la population indienne

D'après Lancet, 2003; 361:598



157

**Recherche d'Interaction Gène-Environnement**



F. Richard

158

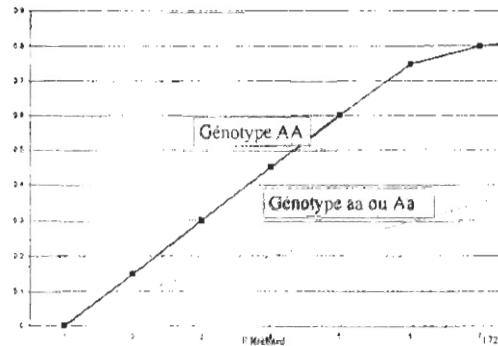
**Interaction : définition**

- Il existe une interaction entre un facteur de risque A et un tiers facteur B, si la relation entre A et la maladie n'est pas la même selon les niveaux du facteur B.
- Ce tiers facteur B peut être :
  - un facteur environnemental : interaction gène-environnement
  - Un autre facteur de susceptibilité génétique : interaction Gène-Gène



159

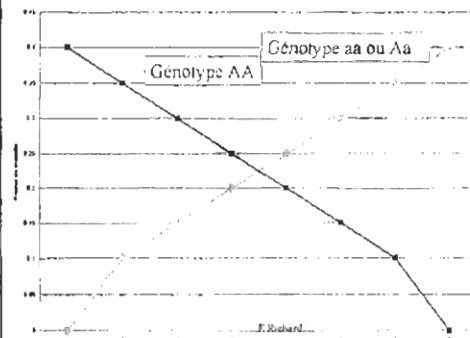
**Exemple: Interaction quantitative**



F. Richard

162

**Exemple: Interaction qualitative**



171

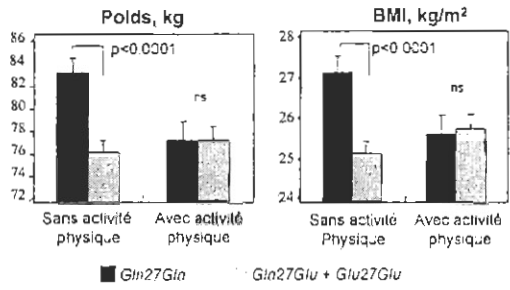
**Interaction exemple : Le Récepteur  $\beta_2$ -Adrénérique**

- L'activation et l'inhibition de la lipolyse sont deux voies majeures de régulation du stockage de la graisse corporelle
- Ces voies sont modulées par des facteurs environnementaux (activité physique) et des déterminants génétiques
- Les récepteurs  $\beta_1$ ,  $\beta_2$ - et  $\beta_3$ -adrénériques stimulent la lipolyse alors que le récepteur  $\alpha_2$ -l'inhibe
- Il existe plusieurs polymorphismes dans le gène humain codant le récepteur  $\beta_2$ -adrénérique (*BAR-2*) dont les polymorphismes Gly16Arg et Gln27Glu

F. Richard

174

## Effets du Polymorphisme *BAR-2* Gln27Glu chez les Hommes selon l'Activité Physique



Meirhaeghe et al., Lancet, 1999

175

## Interaction exemple : Conclusion

- L'activité physique contrebalance l'effet délétère d'une prédisposition génétique à augmenter le poids corporel, la quantité de masse grasse et le risque d'obésité
- Ainsi, l'activité physique serait-elle fortement bénéfique chez les hommes obèses Gln27Gln désirant réduire leur poids
- Une interaction gène-environnement de ce type illustre le fait que l'obésité est le fruit d'interactions gène-environnement

F. Richard

175

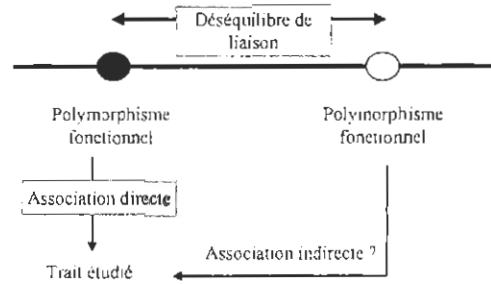
## Interprétation d'une étude d'association.

- Liens de causalité
- Fonctionnalité du polymorphisme
  - Non : autre polymorphisme fonctionnel en déséquilibre de liaison situé à proximité
  - Oui
    - C'est le bon polymorphisme (il faudra le prouver)
    - Un autre polymorphisme fonctionnel peut se trouver à proximité et être en déséquilibre de liaison

F. Richard

177

## Exemple de Marqueur Génétique



F. Richard

177

## Que signifie une association significative?

- Relation réelle et causale
- Relation réelle et non causale
  - Ex : L'association est la conséquence de la maladie
- Relation due au Hasard (erreur de type 1)
- Relation due à un biais.

F. Richard

178

## Décision et risque d'erreur

		La réalité	
		Association	Pas d'association
Conclusion de l'étude	Rejet de H <sub>0</sub>	Correcte Puissance	Erreur de type 1 (alpha = 0.05)
	Non Rejet de H <sub>0</sub>	Erreur de type 2 (Beta ≥ 0.20)	Correcte

F. Richard

180

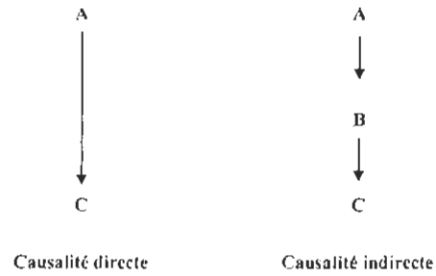
## Les Erreurs en Epidémiologie

- Le hasard : variation aléatoire
- Le biais

F. K. J. et al.

181

## Causalité directe ou indirecte?



F. K. J. et al.

182

## Types de relation de causalité

Dans une relation de causalité entre un facteur A et une maladie M, le facteur A peut être :

- Nécessaire et suffisant
- Nécessaire mais non suffisant
- Suffisant mais non nécessaire
- Ni suffisant, ni nécessaire

F. K. J. et al.

183

## Critères de Présomption de causalité

- ① Relation temporelle
- ② Force de l'association
- ③ Relation dose-effet
- ④ Spécificité de la relation
- ⑤ Constance de l'association et reproductibilité
- ⑥ Effet de l'arrêt de l'exposition
- ⑦ Absence de biais potentiel
- ⑧ Plausibilité biologique
- ⑨ Cohérence avec l'ensemble de connaissances

F. K. J. et al.

184

## Les études d'association

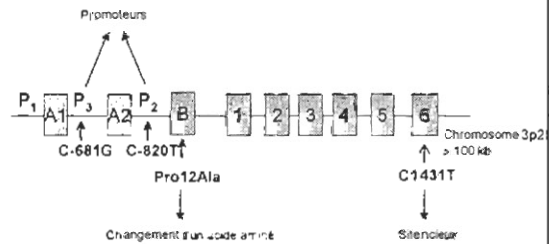
- Recherche de SNPs dans le gène
- Etudes d'haplotypes
- Populations de grande taille
- Réplication dans des populations indépendantes
- Si possible, réplication dans une étude familiale
- Analyse des interactions gène-environnement
- Etude de la fonctionnalité des polymorphismes

Freely associating - Editors - Nat Genet 1999;22: 1-2

185

## Locus du gène *PPAR* $\gamma$

(Peroxisome proliferator-activated receptor)



F. K. J. et al.

186

## Analyse haplotypique du gène PPARγ et syndrome métabolique

Haplotypes	Témoins 826	SynX 268	$\rho^*$ effet global	$\rho^*$ GTGC vs CCCC†
ref CCCC	0.713	0.700		
GCCC	0.118	0.121		
GTGT	0.084	0.072	0.045	0.002
CCCT	0.034	0.028		
GTGC	<b>0.022</b>	<b>0.050</b>		
GCCT	0.017	0.017		

Seuls les haplotypes de fréquence > 1% sont représentés

\* Ajusté sur l'âge, le sexe, le tabac, l'alcool, l'activité physique  
† OR GTGC vs CCCC ajusté=2.37 [1.42-3.95]

Enquête de population, 1155 hommes et femmes âgés de 35 à 65 ans



## Enquête cas-témoins : Un exemple

Association at LRP gene locus with sporadic late-onset Alzheimer's disease.

Lambert et al. Lancet, 1998;351:1787

F. Richard

188

## Etudes Cas-témoins : LRP



Polymorphisme biallélique: Allèle C et T : Etude des fréquences alléliques

	Cas	Témoins
Allèle C	975 (0.87)	982 (0.82)
Allèle T	141 (0.13)	210 (0.18)

$p_1 \neq p_2$  ? → Test du  $\chi^2$  à un ddl

F. Richard

189

Lancet, 1998; 351:1787

## Etudes Cas-témoins : LRP



Test du  $\chi^2$  à un ddl  $H_0: p_1 = p_2$   $H_1: p_1 \neq p_2$

	Cas (O A)	Témoins (O A)	Total
Allèle C	975 [(1957 - 1116)/2308]	982 [(1957 - 1192)/2308]	1957
Allèle T	141 [(351 - 1116)/2308]	210 [(351 - 1192)/2308]	351
Total	1116	1192	2308

$$\chi^2 = \sum \frac{(\text{Attendu} - \text{observé})^2}{\text{Attendu}} \text{ ssi le plus petit } C \geq 5$$

(Observé | Attendu)

F. Richard

190

Lancet, 1998; 351:1787

## Etudes Cas-témoins : LRP



Test du  $\chi^2$  à un ddl  $H_0: p_1 = p_2$   $H_1: p_1 \neq p_2$

	Cas (O A)	Témoins (O A)	Total
Allèle C	975 [946.3]	982 [1010.7]	1957
Allèle T	141 [169.7]	210 [181.3]	351
Total	1116	1192	2308

Rq : 1010.7 + 946.3 = 1957, ...

le plus petit C (=169.7)  $\geq 5$

F. Richard

191

Lancet, 1998; 351:1787

## Etudes Cas-témoins : LRP : fréquences alléliques



	Cas (O A)	Témoins (O A)	Total
Allèle C	975 [946.3]	982 [1010.7]	1957
Allèle T	141 [169.7]	210 [181.3]	351
Total	1116	1192	2308

$$\chi^2 = \sum \frac{(A-O)^2}{A} = \frac{(975-946.3)^2}{946.3} + \frac{(982-1010.7)^2}{1010.7} + \frac{(141-169.7)^2}{169.7} + \frac{(210-181.3)^2}{181.3} = 11.08$$

$\chi^2 = 11.08$  à comparer avec la valeur seuil d'un  $\chi^2$  à 1ddl pour un risque alpha à 0.05 soit à 3.84  $11.08 > 3.84$  donc rejet de  $H_0$

F. Richard

192

Lancet, 1998; 351:1787



### Etudes Cas-témoins et LRP : Mesure du risque

Etude des fréquences génotypiques Test du  $\chi^2$  à 2 ddl

$H_0$  : Pas de différence de distribution entre les cas et les témoins

$H_1$  : Différence de distribution entre les cas et les témoins

	Cas n (p)	Témoins n (p)	Total
CC	428 (0.77)	407 (0.68)	835
CT	119 (0.21)	168 (0.28)	287
TT	11 (0.02)	21 (0.04)	32
Total	558 (1.0)	596 (1.0)	1154

F. Richard

195

Lancet. 1998; 351:1787.

### Etudes Cas-témoins et LRP : Mesure du risque

Etude des fréquences génotypiques Test du  $\chi^2$  à 2 ddl

$H_0$  : Pas de différence de distribution entre les cas et les témoins

$H_1$  : Différence de distribution entre les cas et les témoins

	Cas O [A]	Témoins O [A]	Total
CC	428 [403.75]	407 [431.25]	835
CT	119 [138.78]	168 [148.22]	287
TT	11 [15.47]	21 [16.53]	32
Total	558	596	1154

F. Richard

195

Lancet. 1998; 351:1787.

### Etudes Cas-témoins et LRP : Mesure du risque

Etude des fréquences génotypiques Test du  $\chi^2$  à 2 ddl

	Cas O [A]	Témoins O [A]	Total
CC	428 [403.75]	407 [431.25]	835
CT	119 [138.78]	168 [148.22]	287
TT	11 [15.47]	21 [16.53]	32
Total	558	596	1154

$$\chi^2 = \sum \frac{(A-O)^2}{A} \text{ le plus petit } C = (-15.47) \geq 5$$

Suit un  $\chi^2$  à  $(1-1) \times (2-1)$  ddl, ici un  $\chi^2$  à 2ddl.

F. Richard

195

Lancet. 1998; 351:1787.

### Etudes Cas-témoins et LRP : Mesure du risque

Etude des fréquences génotypiques Test du  $\chi^2$  à 2 ddl

	Cas O [A]	Témoins O [A]	Total
CC	428 [403.75]	407 [431.25]	835
CT	119 [138.78]	168 [148.22]	287
TT	11 [15.47]	21 [16.53]	32
Total	558	596	1154

$\chi^2 = \sum \frac{(A-O)^2}{A} = 10.78$  à comparer avec la valeur seuil d'un  $\chi^2$  à 2 ddl pour un risque  $\alpha = 0.05$  soit 5.991 - 11.08 - 5.991 donc rejet de  $H_0$ .

F. Richard

195

Lancet. 1998; 351:1787.

### Etudes Cas-témoins et LRP : Mesure du risque

**Odds Ratio (OR) :** Mesure du risque chez les sujets porteurs du génotype CC

Groupe de référence : TT

$$OR = \frac{af}{be} = \frac{428 \times 21}{407 \times 11} = 2.01$$

F. Richard

195

Lancet. 1998; 351:1787.

### Etudes Cas-témoins et LRP : Mesure du risque

**Odds Ratio (OR) :** Mesure du risque chez les sujets porteurs du génotype CT

Groupe de référence : TT

$$OR = \frac{cf}{de} = \frac{119 \times 21}{168 \times 11} = 1.35$$

F. Richard

195

Lancet. 1998; 351:1787.

## Etudes Cas-témoins et LRP : Mesure du risque



	« Dominant »	« Récessif »	« Codominant »
CC	OR > 1	OR > 1	OR > 1
CT	OR > 1	OR = 1	OR > 1
TT	1	1	1
	$OR_{CC} \approx OR_{CT}$		$OR_{CT} < OR_{CC}$
	$\chi^2$ à 1ddl	$\chi^2$ à 1ddl	$\chi^2$ à 1ddl = 10,72

F Richard

199

Lancet, 1998; 351:1787.

## Etudes Cas-témoins et LRP : Mesure du risque



	« Dominant »	« Récessif »	« Codominant »
CC	OR > 1	OR > 1	OR > 1
CT	OR > 1	OR = 1	OR > 1
TT	1	1	1
	$OR_{CC} \approx OR_{CT}$		$OR_{CT} < OR_{CC}$
	$\chi^2$ à 1ddl	$\chi^2$ à 1ddl	$\chi^2$ à 1ddl = 10,72

F Richard

200

Lancet, 1998; 351:1787.

## Etudes Cas-témoins et LRP : Mesure du risque



Etude des fréquences génotypiques sous l'hypothèse que C a un effet dominant

	Cas	Témoins
	n (p)	n (p)
CC + CT	547 (0,98)	575 (0,96)
TT	11 (0,02)	21 (0,04)

$$\chi_1^2 = \sum \frac{(A-O)^2}{A} = 2,58 < 3,84 \text{ (Rq : le plus petit C = 15)}$$

Test du  $\chi^2$  à 1 ddl

$$OR = \frac{547 \times 21}{575 \times 11} = 1,82 \quad OR = 1,82 [0,83-4,05]$$

F Richard

Lancet, 1998; 351:1787.

## Etudes Cas-témoins et LRP : Mesure du risque



	« Dominant »	« Récessif »	« Codominant »
CC	OR > 1	OR > 1	OR > 1
CT	OR > 1	OR = 1	OR > 1
TT	1	1	1
	$OR_{CC} \approx OR_{CT}$		$OR_{CT} < OR_{CC}$
	$\chi^2$ à 1ddl	$\chi^2$ à 1ddl	$\chi^2$ à 1ddl = 10,72

Lancet, 1998; 351:1787.

F Richard

201

## Etudes Cas-témoins et LRP : Mesure du risque



Etude des fréquences génotypiques sous l'hypothèse que C a un effet récessif

	Cas	Témoins
	n (p)	n (p)
CC	428 (0,77)	407 (0,68)
CT + TT	130 (0,23)	189 (0,32)

$$\chi_1^2 = \sum \frac{(A-O)^2}{A} = 10,20 \gg 3,84 \text{ (Rq : le plus petit C = 154,2)}$$

Test du  $\chi^2$  à 1 ddl

$$OR = \frac{428 \times 189}{407 \times 130} = 1,53 \quad OR = 1,53 [1,17-2,00] \text{ (Cf article)}$$

F Richard

Lancet, 1998; 351:1787.

## Analyse de Liaison Génétique

Florence RICHARD

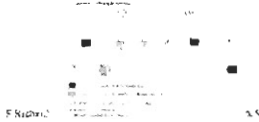


F Richard

204

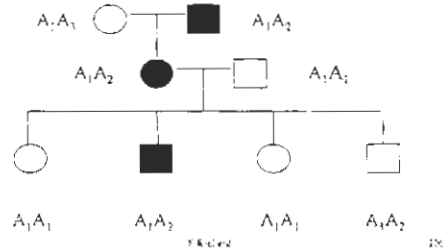
Analyse de Liaison Génétique  
Principe

- L'analyse de liaison génétique consiste à suivre la ségrégation conjointe de deux locus d'une génération à une autre.
- Nécessite d'avoir des informations à un niveau familial sur au moins deux générations.



F. Richard 2004 205

Analyse de Liaison Génétique  
Coségrégation



F. Richard 2004 206

Analyse de Liaison Génétique  
But

- Tester l'indépendance de transmission des allèles aux deux locus d'intérêt et si tel n'est pas le cas
- Localiser un locus par rapport à un autre
- Rq: un de ces locus peut être une maladie un des allèles du marqueur génétique est-il transmis de façon préférentielle avec la maladie. Si oui, la mutation responsable de la maladie se situe probablement à proximité de ce marqueur

F. Richard 2004 207

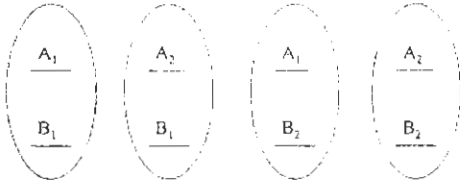
Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

- Prenons un sujet, double hétérozygote pour deux locus A et B.
- Le sujet est :  $A_1A_2$  pour le locus A  
 $B_1B_2$  pour le locus B
- Si ces deux locus A et B sont proches l'un de l'autre, la transmission des allèles ne se fera pas de façon indépendante: les deux locus sont dits liés génétiquement
- Si les deux locus sont sur deux chromosomes différents ou sur le même chromosome mais à distance, la transmission des allèles se fera de manière indépendante

F. Richard 2004 208

Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

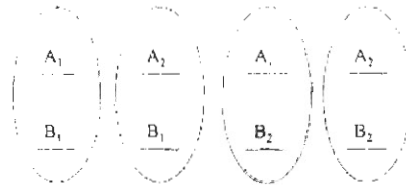
Notre sujet peut produire 4 types de gamètes.



F. Richard 2004 209

Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

Si les deux locus A et B sont indépendants



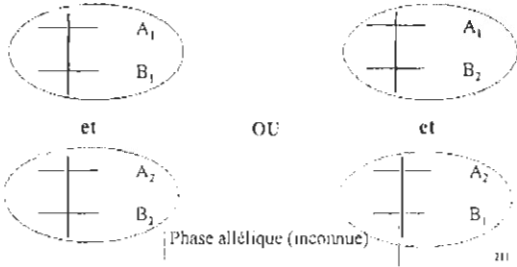
Probabilités 1/4 1/4 1/4 1/4

F. Richard 2004 210

Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

Si les deux locus A et B sont génétiquement liés

Il produira de façon préférentielle les gamètes :



Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

Si les deux locus A et B sont liés : supposons que l'on connaisse la phase allélique du sujet :  $A_1B_1/A_2B_2$



•  $A_1$  et  $B_1$  sont dits en couplage (idem pour  $A_2B_2$ )

et

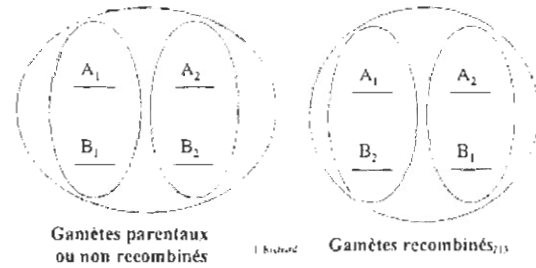


•  $A_1$  et  $B_2$  sont dits en révulsion (idem pour  $A_2$  et  $B_1$ )

Le sujet produira de façon "préférentielle" ces deux types de gamètes

Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

Les deux locus A et B sont liés et la phase allélique du sujet :  $A_1B_1/A_2B_2$



Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

Les deux locus A et B sont liés et la phase allélique du sujet :  $A_1B_1/A_2B_2$

Plus A et B sont proches,

- Plus on observera de gamètes de type  $A_1B_1$  et  $A_2B_2$
- Plus les gamètes de type  $A_1B_2$  et  $A_2B_1$  seront rares

Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

$$\text{Taux de Recombinaison } \theta = \frac{\text{Nombre de gamètes recombines}}{\text{Nombre de gamètes}}$$

Analyse de Liaison Génétique  
Taux de Recombinaison  $\theta$

Si les deux locus A et B sont indépendants:

Le sujet produira autant de gamètes recombines que de gamètes non recombines

$$\text{Taux de Recombinaison } \theta = \frac{\text{Nombre de gamètes recombines}}{\text{Nombre de gamètes}}$$

$$\text{Taux de Recombinaison } \theta = \frac{1}{2}$$

## Analyse de Liaison Génétique

### Taux de Recombinaison $\theta$

Si les deux locus A et B sont génétiquement liés :

Le sujet produira plus de gamètes non recombinés que de gamètes recombinés

$$\text{Taux de Recombinaison } \theta = \frac{\text{Nombre de gamètes recombinés}}{\text{Nombre de gamètes}}$$

$$\text{Taux de Recombinaison } \theta < \frac{1}{2}$$

Plus la distance entre A et B  $\rightarrow 0$ , plus  $\theta \rightarrow 0$

F. R. 1992

217

## Analyse de Liaison Génétique

### Taux de Recombinaison $\theta$

- Plus la distance entre A et B  $\rightarrow 0$ , plus  $\theta \rightarrow 0$
- Si  $\theta \rightarrow 0$ , on dit qu'il y a une liaison génétique entre les deux marqueurs

Donc :

Si  $0 < \theta < 1/2$  Liaison génétique

Si  $\theta = 1/2$  Indépendance génétique

F.R. 1992

218

## Analyse de Liaison Génétique

### Méthode des Lod scores

- Méthode d'analyse permettant de détecter une liaison génétique
- Analyse 2 points : analyse la liaison entre deux locus
  - Objectif
    - Positionner un locus par rapport à une autre
    - L'un de ces locus peut être une maladie
    - Aide à la localisation chromosomique

F. R. 1992

219

## Analyse de Liaison Génétique

### Méthode des Lod scores

$H_0$  : Transmission indépendante de deux locus, définie par  $\theta = 1/2$

$H_1$  : Liaison génétique :  $0 < \theta < 1/2$

- Pour une famille  $i$ , on calcule le taux de recombinaison entre les deux locus :  $\theta_i$
- Sous  $H_0$ ,  $L_i(\theta = 1/2)$       Sous  $H_1$ ,  $L_i(\theta = \theta_i)$

F. R. 1992

220

## Analyse de Liaison Génétique

### Méthode des Lod scores

- Le lod score en  $\theta_i$  de cette famille  $i$ , noté  $Z_i(\theta_i)$  est donné par le logarithme à base 10 du rapport de ces vraisemblances
- Le lod score d'un échantillon de plusieurs familles est égale à la somme des lod scores de chaque famille.

$$Z(\theta_i) = \text{Log}_{10} \left[ \frac{L(\theta_i)}{L(1/2)} \right]$$

F. R. 1992

221

## Analyse de Liaison Génétique

### Méthode des Lod scores : Test de Liaison de Morton (1955)

- Test de type séquentiel
- On accumule de l'information (nombre de familles dans l'échantillon) jusqu'au moment où il est possible de trancher entre :
  - $H_0$  :  $\theta = 1/2$       Indépendance
  - $H_1$  :  $0 < \theta_i < 1/2$       Liaison génétique

Si	$Z(\theta_i) \geq 3$	Rejet de $H_0$ : Liaison génétique
Si	$Z(\theta_i) \leq -2$	Rejet de la liaison à $\theta_i$
Si	$-2 < Z(\theta_i) < 3$	Pas de conclusion, on rajoute des familles

222

**Analyse de Liaison Génétique**  
Méthode des Lod scores : Test de Liaison

- Analyses en deux étapes
  - 1) permettre de conclure ou de rejeter la liaison entre deux locus :  $H_0: \theta = 1/2$
  - 2) Estimer le taux de recombinaison qui maximise la fonction du lod score. On va tester la liaison pour un ensemble de valeurs comprise en 0 et 1/2
    - $H_1: 0 < \theta_1 < 0.4$
    - $H_1: 0 < \theta_1 < 0.3, \dots$
- L'objectif est de préciser la distance entre les deux locus
- Attention : Ces seuils +3 et -2 ont été déterminés en considérant que  $H_0$  est testée contre une seule hypothèse alternative. <sup>223</sup>

**Analyse de Liaison Génétique**  
Equivalences théoriques entre distance génétique et distance physique

Distance physique en kilobases	Distance génétique en centimorgan kilobases	$\theta$	Probabilité de recombinaison
10 000	10	0.10	$10^{-1}$
5 000	5	0.05	$5 \times 10^{-2}$
1 000	1	0.01	$10^{-2}$
100	0.1	0.001	$10^{-3}$
10	Non mesurable	Non mesurable	$10^{-5}$
1	Non mesurable	Non mesurable	$10^{-5}$

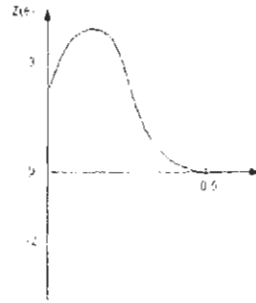
Nb : 1 centimorgan = 1% de recombinaison <sup>224</sup>

**Analyse de Liaison Génétique**  
Analyse de Liaison Multipoint

- Etude de la ségrégation conjointe de plus de 2 marqueurs génétiques
- Objectifs
  - 1) Détecter la liaison et inférer un ordre pour plusieurs locus
  - 2) Tester la liaison et inférer l'ordre d'un locus par rapport à plusieurs autres marqueurs déjà ordonnés

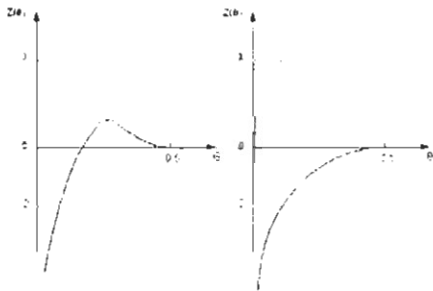
<sup>225</sup>

**Analyse de Liaison Génétique : Liaison**  
Méthode des Lod scores : Test de Liaison de Morton (1955)



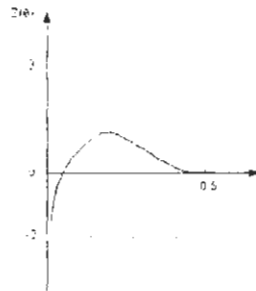
<sup>226</sup>

**Analyse de Liaison Génétique : Exclusion**  
Méthode des Lod scores : Test de Liaison de Morton (1955)



<sup>227</sup>

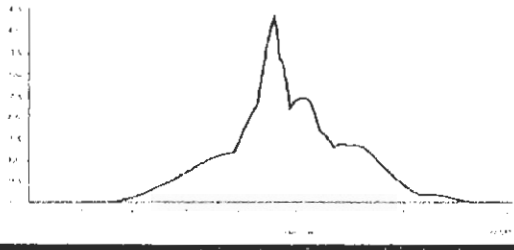
**Analyse de Liaison Génétique : On ne peut pas conclure !**  
Méthode des Lod scores : Test de Liaison de Morton (1955)



<sup>228</sup>

## Exemple de carte

Genetic linkage graph based on centiMorgans



## Analyse de Liaison Génétique Exemple 1: Epilepsie



- Zara F et al. Am J Hum Genet, 2000; 66 : 1552-57
- Epilepsie myoclonique infantile
- Grande famille
- Forme autosomique récessive

F. Richard

231

## Analyse de Liaison Génétique Exemple 1 : Epilepsie



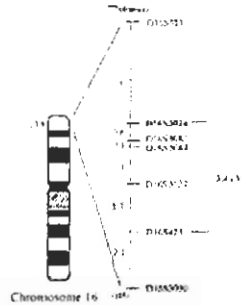
LOD Scores Deux points - marqueurs du chromosome 16

Locus	LOD à $\theta$		
	0	0.05	0.2
D16S3024	$-\infty$	1.50	1.72
D16S3082	1.81	1.61	1.41
D16S3084	2.94	2.61	2.27
D16S3027	4.48	4.05	3.61
D16S4237	$-\infty$	2.89	2.77

F. Richard

231

## Analyse de Liaison Génétique Exemple 1 : Epilepsie



Zara F et al. Am J Hum Genet, 2000; 66 : 1552-57

232

## Analyse de Liaison Génétique Exemple 2: Migraine sans aura



- Soragna et al. Am J Hum Genet, 2003; 72: 161-167
- Migraine sans aura
- Une grande famille
- Liaison avec le chromosome 14.

F. Richard

231

## Analyse de Liaison Génétique Exemple 2: Migraine sans aura

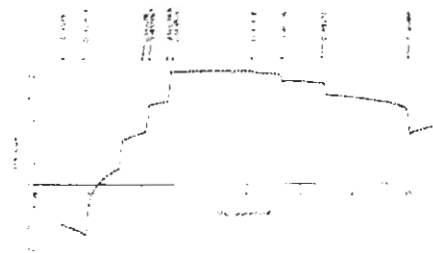


Figure 2. Genetic linkage graph based on centiMorgans for migraine without aura. The x-axis represents centiMorgans and the y-axis represents LOD scores. The peak is at approximately 1.5 cM with a LOD score of 4.5.

## Méthode des Germais

### Principe

Sous  $H_0$  : indépendance entre la maladie et le marqueur

$P(\text{IBD} = 2) = 1/4$  Complète concordance entre les germains

$P(\text{IBD} = 1) = 1/2$  Un allèle en commun

$P(\text{IBD} = 0) = 1/4$  Complète discordance entre les germains

La distribution observée est-elle significativement différente de la distribution attendue sous  $H_0$  ?

Rq : méthode de plus faible puissance que la précédente

Nécessite un grand nombre de paires de germains

F.Richard

241

## Sibpairs IBD des Hétérozygotes : Maladie d'Alzheimer et chromosome 10



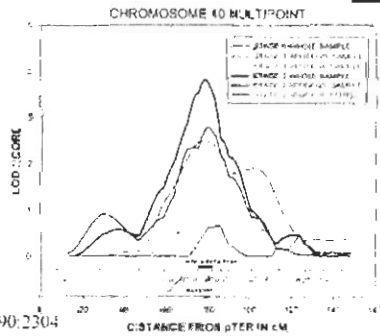
Locus	Position cM	IBD	IBD APOEε4+	IBD APOEε4-
1423	47.75	0.50	0.50	0.50
1227	75.57	0.56	0.53	0.57
1211	82.21	0.64	0.64	0.69
2475	103.43	0.56	0.57	0.54
2470	112.58	0.52	0.52	0.54
1269	131.8	0.50	0.50	0.50

Science, 2000, 290:2304

F.Richard

242

## Sibpairs IBD des Hétérozygotes : Maladie d'Alzheimer et chromosome 10



Science, 2000, 290:2304

F.Richard

244

## Analyses d'association-liaison

### Localisation des gènes par analyse de liaison Génétique

- Recherche orientée vers des régions candidates : confirmer une liaison génétique d'une région candidate avec le phénotype étudié
- Recherche au hasard sur le génome : localisation du gène d'une maladie sans a priori sur sa fonction (clonage positionnel)

F.Richard

245

### Etudes d'Association Liaison : Transmission Disequilibrium Test (TDT)

#### Principe

- Ce test va utiliser à la fois l'association et la liaison
- Il utilise l'information de familles nucléaires



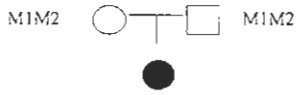
F.Richard

247



## Transmission Disequilibrium Test (TDT) Principe

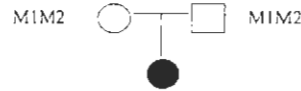
- Dans cette famille, on va génotyper tous les membres soit :
  - l'enfant malade
  - le père
  - la mère



F. Richard

247

## Transmission Disequilibrium Test (TDT) Principe



- M1M1: les deux parents ont transmis M1, aucun n'a transmis M2
- M2M2: les deux parents ont transmis M2, aucun n'a transmis M1
- M1M2: un des parents a transmis M1, l'autre M2

F. Richard

248

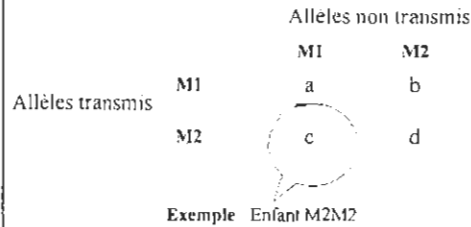
## Transmission Disequilibrium Test (TDT) Principe

- Pour chaque parent, on a donc deux allèles :
  - l'un transmis
  - l'autre transmis
- Le groupe de référence va être constitué d'un groupe témoin intra-familiale.

F. Richard

249

## Transmission Disequilibrium Test (TDT) Principe



F. Richard

250

## Transmission Disequilibrium Test (TDT) Principe

- Test statistique
  - H0 : Pas de liaison et/ou pas d'association
  - Le rejet de H0 donne des arguments en faveur d'une liaison (en présence d'un déséquilibre gamétique) et d'une association (en présence d'une liaison)

### Test de Mac-Nemar

$$\chi^2 = \frac{(b-c)^2}{(b+c)} \rightarrow \chi^2_{1ddl} \text{ H}_0$$

Condition d'application  $b+c > 10$

F. Richard

251

## TDT : Facteur de susceptibilité du diabète type 1 sur le chromosome 10p11-q11

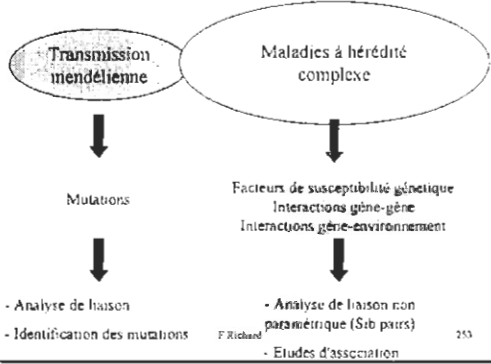
Marqueur	Allèle	%	Transmis	Non Transmis	% Transmis	p
D10S611	1	20	137	145	49%	ns
	2	30	184	204	47%	
	3	45	231	199	54%	
D10S193	5	30	188	192	49%	0.04
	7	32	229	187	55%	

Hum Mol Genet, 1997;6:1011.

F. Richard

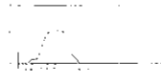
252

## Epidémiologie Génétique et maladies "génétiques"



## Limites méthodologiques

### • études de liaisons génétiques



- très bien adaptées pour les modes de ségrégation monogéniques
- moins efficace pour les maladies multifactorielles
- régions chromosomiques d'intérêt très étendues

### • études d'associations :

- très efficace pour mettre en évidence des impacts même faibles
- très efficace pour mettre en évidence des interactions entre plusieurs facteurs étudiés
- problèmes dus à la qualité et la taille des populations étudiés
- problèmes liés aux polymorphismes testés

F. Richard

254

## Epidémiologie Génétique et maladies "génétiques"

### A quoi cela sert?

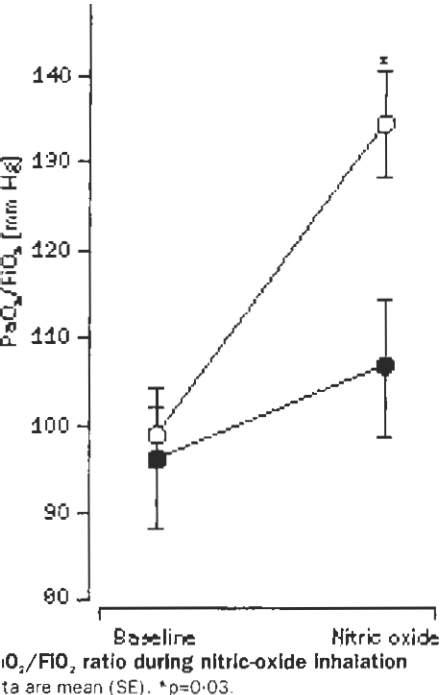
1) Recherche et meilleur compréhension de la physiopathologie de la maladie + nouvelles voies thérapeutiques :

ex mutations sur les gènes de l'APP, PS1, PS2 et maladie d'Alzheimer, facteur de susceptibilité génétique de la Maladie d'Alzheimer.

2) Prédiction

F. Richard

255



nitric oxide was measured as the percentage increase in  $\text{PaO}_2/\text{FiO}_2$  ( $\Delta\text{PaO}_2/\text{FiO}_2$ ). If more than one nitric-oxide dose is inhaled, the largest  $\Delta\text{PaO}_2/\text{FiO}_2$  was recorded. Patients with a  $\Delta\text{PaO}_2/\text{FiO}_2$  of less than 20% were classified as non-responders.<sup>2,3</sup>

The oxygenation response to inhaled nitric oxide was significantly decreased in patients with genotype B who possess blood group B or AB ( $\Delta\text{PaO}_2/\text{FiO}_2$  13 [4]%, n=27) compared with patients without genotype B who express blood group O or A ( $\Delta\text{PaO}_2/\text{FiO}_2$  36 [4]%, n=111, p=0.03, ANOVA and post-hoc Scheffé's test; figure). 67% of patients with blood group B or AB were non-responders to inhaled nitric oxide compared with 27% of patients with blood group O or A (p<0.001, two-tailed Fisher's exact test). The oxygenation response to inhaled nitric oxide did not differ between men (32 [5]%, n=91) and women (32 [6]%, n=47), between rhesus-positive (35 [4]%, n=109) and rhesus-negative patients (21 [5]%, n=29).

These results suggest a relation between the oxygenation response to inhaled nitric oxide and the ABO blood-group system. Possible explanations include a genetic linkage between the ABO gene locus (9q34) and another, as yet unknown, gene locus that is involved in the pulmonary vascular response to inhaled nitric oxide. Alternatively, ABO blood group may modify the severity or course of ARDS and/or responsiveness to nitric oxide. For example, patients with ARDS associated with sepsis more frequently fail to respond to inhaled nitric oxide than patients with ARDS associated with other processes.<sup>2</sup> Antibodies to blood group B are reported to possess antibacterial properties.<sup>5</sup> Since these antibodies are naturally present in individuals with blood groups O or A, who responded better to inhaled nitric oxide, it is possible that ABO blood group may modify the underlying process contributing to ARDS and, thereby, modulate responsiveness to inhaled nitric oxide.

This study was supported by USPHS grants HL 42397 (W M Z and K D) and HL 55377 (K D B), and by the Deutsche Forschungsgemeinschaft SFB 211/4-1 (J W). K D B is an established investigator of the American

- determinants of the response to inhaled nitric oxide in patients with acute respiratory distress syndrome. *Anesthesiology* 1997; 87: 297-307.
- Dellinger RP, Zimmerman JL, Taylor RW, et al. Effects of inhaled nitric oxide in patients with acute respiratory distress syndrome: results of a randomized phase II trial. Inhaled Nitric Oxide in ARDS Study Group. *Crit Care Med* 1998; 26: 15-23.
  - Murray JF, Matthay MA, Luce JM, Flick MR. An expanded definition of the adult respiratory distress syndrome. *Am Rev Respir Dis* 1988; 138: 720-23.
  - Springer GF. Importance of blood-group substances in interactions between man and microbes. *Ann N Y Acad Sci* 1970; 169: 134-52.

Department of Anesthesia and Critical Care (J Weimann) and the Cardiovascular Research Center and Cardiology Division of the Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA; and Klinik für Anaesthesiologie, Ruprecht-Karls Universität Heidelberg, Heidelberg, Germany

## Association at *LRP* gene locus with sporadic late-onset Alzheimer's disease

Jean-Charles Lambert, Fabienne Wavrant-De Vrièze, Philippe Amouyel, Marie-Christine Chartier-Harlin

The  $\epsilon 4$  allele of the apolipoprotein E gene (*APOE*) is the only recognised genetic susceptibility factor for late-onset Alzheimer's disease (AD). However, the mechanism by which the *APOE* alleles affect disease development is not fully understood. The LDL receptor-related protein (*LRP*), located on chromosome 12, is the main apoE receptor in the brain. *LRP* mediates neurite outgrowth in an apoE-isoform-dependent manner. *LRP* is also responsible for the endocytosis of secreted APP, another molecule central to the pathogenesis of AD. These observations suggest a potential for this apoE receptor to have a role in the pathogenesis of AD. Four studies have investigated the association of a TTTC repeat polymorphism in the 5' end of the *LRP* gene with AD, two of them with negative results.<sup>1,3</sup> These discrepancies may be due to sample limitations (n<180 AD cases for all studies), different genetic background (not all whites, a mixture of familial or sporadic AD cases in some studies), or linkage disequilibrium with another gene. A report describing a silent CT polymorphism in exon 3 favours the hypothesis that variability in the *LRP* locus is associated with familial late-onset AD.<sup>4</sup> We tested this polymorphism in a large white sample composed of 558 sporadic probable AD—ie, cases without any formal pattern of family inheritance—(mean age 71.8 [SD 8.1] years; mean age at onset 68.6 [8.2] years; 38.2% men) and 596 controls (mean age 72.7 [8.3] years; 37.2% men). Patients characterised according to NINCDS-ADRDA criteria, were recruited in hospital. Controls were defined as those without DSM-III-R dementia criteria and with intact cognitive functions. Informed consent was obtained. Statistical analyses were done with SAS software (SAS institute, Cary, NC). Categorical data were tested with the Pearson  $\chi^2$  test. Multivariate logistic regression analyses were used to test for interactions between *LRP* genotypes and

	n	Genotype distribution*			Allele distribution†	
		CC	CT	TT	C	T
Control	596	407 (0.68)	168 (0.28)	21 (0.04)	982 (0.82)	210 (0.18)
AD	558	428 (0.77)	119 (0.21)	11 (0.02)	975 (0.87)	141 (0.13)

\*p=0.004, †p<0.001.

### Genotype and allele distributions of the *LRP* exon 3 CT polymorphism

CI 1.2-2.0],  $p=0.002$ ) compared with controls. Although no statistical interaction could be detected between *LRP* polymorphism and age at onset, the association of the *LRP* CC genotype was more pronounced in late-onset AD cases (1.7 [1.2-2.3],  $p=0.001$ , age at onset >65 years) than in cases with earlier ages at onset (1.2 [0.7-1.9],  $p=0.520$ , age at onset  $\leq 65$  years). Similarly, the level of association between the *LRP* CC genotype and AD tended to be stronger in those carrying at least one  $\epsilon 4$  allele (2.2 [1.3-3.7],  $p=0.006$ ), than in people without an  $\epsilon 4$  allele (1.5 [1.0-2.3],  $p=0.080$ ), despite the lack of significance of the statistical interaction.

Kang and colleagues<sup>4</sup> reported that the *LRP* CC genotype was a risk factor for familial late-onset AD and described a similar trend in sporadic late-onset AD cases, lack of significance in this latter group probably being due to the small sample size. Our results extend the possible effect of the *LRP* CC genotype to sporadic late-onset AD cases. However, we cannot reject the hypothesis that linkage disequilibrium exists between this *LRP* polymorphism and another polymorphism within *LRP* or neighbouring genes. In the light of linkage data implicating a new locus on chromosome 12 for late-onset AD,<sup>5</sup> the genetic association of *LRP* variability with both familial and sporadic late-onset AD may confirm involvement of the *LRP* locus in the disease.

We thank Jordi Pérez-Tur for his helpful discussion. J-CL is a recipient of the Ministère de l'Enseignement Supérieur et de la Recherche. This work was supported by the Institut National pour la Santé et la Recherche Médicale, the Institut Pasteur de Lille, the Conseil Régional du Nord-Pas de Calais (axe régional de recherche sur les maladies neurodégénératives et vieillissement cérébral) and the Fondation pour la Recherche Médicale.

London CL, Talbot CJ, Craddock NJ, et al. Genetic association studies between dementia of Alzheimer's type and the three receptors for apolipoprotein E in Caucasian population. *Neurosci Lett* 1997; 222: 187-90.

Wavrant-De Vrièze F, Pérez-Tur J, Lambert J-C, et al. Association between the low-density lipoprotein receptor-related protein (LRP) gene and Alzheimer's disease. *Neurosci Lett* 1997; 227: 68-70.

Clarworthy AE, Gomez-Isla T, Rebeck GW, Wallace RB, Hyman BT. Lack of association of a polymorphism in the low-density lipoprotein receptor-related protein gene with Alzheimer disease. *Arch Neurol* 1997; 54: 1289-92.

Kang DE, Saitoh T, Chen X, et al. Genetic association of the low-density lipoprotein receptor-related protein gene (LRP), an apolipoprotein E receptor with late-onset Alzheimer's disease. *Neurology* 1997; 49: 56-61.

Pericak-Vance MA, Bass MP, Yamaoka LH, et al. Complete genomic screen in late-onset familial Alzheimer disease. *JAMA* 1997; 278: 1237-41.

SERM CJF 95-05, Institut Pasteur de Lille, BP 245, 59019 Lille, France  
4 C Chartier-Harlin

## Could lightning injury be magnetically induced?

Michael Cherington, Howard Wachtel, Philip R Yarnell

Lightning injuries are usually thought to be caused by direct strike, side flash, or ground current,<sup>1,2</sup> none of which explain lightning-related deaths where current does not apparently enter or leave the patient's body.

In June, 1996, a 32-year-old man and three other golfers stood under a tree during a lightning storm. When lightning struck the tree, the patient suffered a cardiac arrest. Cardiopulmonary resuscitation was given by a physician who was nearby. The patient was taken to hospital but remained comatose and died on the 18th day. Three other golfers under the same tree survived. One had surface burns on his head,

evidence of external burns from lightning currents. Lightning-related hypoxic encephalopathy secondary to cardiac arrest is often fatal.<sup>3</sup> Most lightning casualties have external signs of damage from contact with electrical current, such as skin burns and Lichtenberg figures or "ferning".<sup>4,5</sup> However, some patients have a cardiac arrest as the result of a lightning strike without external signs of electrical burns.

Lightning "bolts" have very high peak currents ( $\geq 100\,000$  amps) which rise in  $\mu$ s and decay more slowly. They may produce intense nearby magnetic fields (several millitesla at a distance of about 1 m), which may induce large but short-lived (<1 ms) currents in a human body. The induced current wave-form is proportional to the change per time of the magnetic field. The lightning may induce a loop current within the human torso without evidence of current entering the body. If these currents occur during a vulnerable part of the cardiac cycle, they could cause asystole or ventricular fibrillation.\*

This hypothesis may be tested on mice or rats exposed to rapidly rising magnetic field pulses that approximate the dynamics of lightning currents within cylindrical chambers enclosed by solenoid windings. This method would necessitate the use of fields on the order or ten times as great as the natural lightning fields. Our proposed mechanism may explain some unwitnessed and unexplained "heart attacks" among hikers found dead in the mountains.

Presented in part at the American Association of Physics Teachers Meeting in Denver, Colorado, August 14, 1997.

We thank Vincent W Vanek.

\*Lightning stroke with peak current 100 000 amp gives peak magnetic field (B<sub>r</sub>):

$$B_r = \frac{\mu I_p}{2\pi R}$$

Where  $\mu$  = magnetic permeability of air  $4\pi(10^{-7})$  Webers/A·m  $I_p$  = peak current (amps).  $R$  = distance from the lightning stroke (m).

For a distance of 1.0 m,  $B_r = 2(10^{-2})$  Tesla. Lightning current rises to its peak in about  $10^{-6}$  s, so the rate of rise of dB/dt is  $2(10^6)$  Tesla/s. Electrical field induced in a cylindrical object (human torso)

$$E = \left(\frac{a}{2}\right) \frac{dB}{dt}$$

Where  $E$  = electrical field (v/m)  $a$  = radius of the induced current path (m). Highest fields would be at the perimeter of the torso and the lowest ones produced in the centre. With an estimate of  $a=0.1$  m yields  $E$  of about 1000 V/m for  $dB/dt=2(10^6)$  Tesla, and current density of about 1000 A/m<sup>2</sup> through the ventricles (tissue resistivity 1 ohm/m). Long current pulses (of ms) or for 60 Hz, such a current density would lead to arrhythmias. For current pulses lasting a few  $\mu$ s, electrical impedance of cardiac-cell membranes is due almost entirely to the membrane capacitance ( $C_m$ ) which is on the order of 0.01 F/m<sup>2</sup>, the change in membrane potential ( $\Delta V_m$ ) is

$$\Delta V_m = \frac{J(\Delta t)}{C_m}$$

For  $J=1000$  A/m<sup>2</sup>, and the duration of the current pulse,  $\Delta t=10^{-6}$  s, predicts a  $\Delta V_m$  of 100 mV. If the cardiac cell is in a refractory state this current is not likely to re-trigger a discharge in the active phase corresponding to the ECG T wave. Ventricular fibrillation or asystole may result.

- 1 Kleinschmidt-DeMasters BK. Neuropathology of lightning-strike injuries. *Semin Neurol* 1995; 15: 323-28.
- 2 Cooper MA. Emergent care of lightning and electrical injuries. *Semin Neurol* 1995; 15: 268-78.
- 3 Wetli CV. Keraunopathology: an analysis of 45 fatalities. *Am J Forensic Med Path* 1996; 17: 89-98.
- 4 Cherington M, Yarnell PR, London SF. Neurological complications of lightning injuries. *West J Med* 1995; 162: 413-17.
- 5 ten Duis HJ, Klasen HJ, Nijsten MWN, Pietronero L. Superficial lightning injuries: their "fractal" shape and origin. *Burns* 1987; 13: 141-46.

Lightning Data Center, St Anthony Hospital, Denver, CO 80204, USA (M Cherington); Department of Neurology, University of Colorado School of Medicine, Denver; and Electrical and Computer Engineering Department, Boulder, CO