

ÉRIC DEMERS

**L'ANALYSE STATISTIQUE DE LA LIAISON  
GÉNÉTIQUE DES SYMPTÔMES POSITIFS ET  
NÉGATIFS DE LA SCHIZOPHRÉNIE**

Mémoire  
présenté  
à la Faculté des études supérieures  
de l'Université Laval  
pour l'obtention  
du grade de maître ès sciences (M.Sc.)

Département de mathématiques et de statistique  
FACULTÉ DES SCIENCES ET DE GÉNIE  
UNIVERSITÉ LAVAL  
QUÉBEC

JUILLET 1997

© Éric Demers, 1997

**Acquisitions and  
Bibliographic Services**

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

**Acquisitions et  
services bibliographiques**

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-25550-6

# Résumé

Dans ce mémoire, on va d'abord décrire les notions de base de la génétique humaine. Ensuite, on expliquera le phénomène de recombinaison et l'analyse de linkage entre le locus d'une maladie et celui d'un marqueur génétique. Un chapitre sera consacré aux études familiales, de jumeaux et d'adoption réalisées antérieurement sur la schizophrénie afin de déterminer s'il s'agit d'une maladie génétique. Par la suite, on décrira la complexité de l'analyse de linkage pour la schizophrénie. Dans un chapitre suivant, on va discuter de l'utilisation des symptômes positifs et négatifs comme nouvelle définition du phénotype. Finalement, des études pratiques d'analyse statistique de la liaison génétique ont été réalisées avec un échantillon de familles provenant du Centre de recherche Université Laval Robert-Giffard.

Éric Demers  
Étudiant

Louis-Paul Rivest  
Directeur de recherche

Chantal Mérette  
Codirectrice de recherche

# Avant-propos

Tout d'abord, j'aimerais remercier mon directeur de recherche, M. Louis-Paul Rivest du Département de mathématiques et de statistique et ma codirectrice, Mme Chantal Mérette du Centre de recherche Université Laval Robert-Giffard qui m'ont bien encadré et aidé par leur grande disponibilité pour la réalisation de ce mémoire. De plus, je tiens à leur dire merci pour le soutien financier qui m'a été accordé.

J'apprécie également les généreux conseils de mes amis, Caty Blanchette et Éric Shink, en ce qui concerne l'utilisation du logiciel L<sup>A</sup>T<sub>E</sub>X. Je remercie tous les membres de ma famille pour leur soutien moral qui m'a été bénéfique.

Finalement, je ne pourrai passer sous silence l'influence de l'un de mes directeurs de conscience, Michel de Montaigne, dont je cite : "Il n'est désir plus naturel que le désir de connaissance".

# Table des matières

<b>Résumé</b>	<b>i</b>
<b>Avant-propos</b>	<b>ii</b>
<b>Table des matières</b>	<b>iii</b>
<b>Liste des figures</b>	<b>vi</b>
<b>Liste des tableaux</b>	<b>viii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Notions de base de la génétique humaine</b>	<b>3</b>
1.1 Les concepts fondamentaux de la génétique humaine . . . . .	3
1.2 Les deux lois génétiques de Mendel . . . . .	7
<b>2 La recombinaison et l'analyse de linkage</b>	<b>10</b>
2.1 La recombinaison . . . . .	10
2.2 Distinction entre distance génétique et distance physique . . . . .	15
2.3 La cartographie du génome humain . . . . .	16
2.4 Analyse de linkage entre le locus d'une maladie et celui d'un marqueur génétique . . . . .	16

2.4.1	Calcul de la fonction de vraisemblance pour une famille en fonction de la fraction de recombinaison $\theta$ . . . . .	17
2.4.2	L'algorithme d'Elston-Stewart . . . . .	23
2.4.2.1	Les trois composantes de la vraisemblance d'un pedigree	23
2.4.2.2	La vraisemblance d'une famille nucléaire . . . . .	30
2.4.2.3	La vraisemblance d'une famille étendue . . . . .	31
2.4.3	Méthode du Lod-score . . . . .	34
2.4.4	Hétérogénéité due à un mélange de familles indissociables . . . .	38
2.4.5	La notion de boucle dans un pedigree . . . . .	41
2.4.5.1	Boucle de consanguinité . . . . .	42
2.4.5.2	Boucle de mariage . . . . .	43
<b>3</b>	<b>La schizophrénie est-elle une maladie génétique?</b>	<b>45</b>
3.1	Les études familiales . . . . .	46
3.2	Les études de jumeaux . . . . .	47
3.3	Les études d'adoption . . . . .	48
<b>4</b>	<b>Complexité de l'analyse de linkage pour la schizophrénie</b>	<b>49</b>
4.1	Indices de complexité . . . . .	49
4.2	Analyse de linkage avec un trait complexe . . . . .	51
4.3	Modes de transmission complexes suggérés . . . . .	52
4.4	Nouvelle définition du phénotype . . . . .	53
<b>5</b>	<b>Utilisation d'une dimension symptomatique comme nouvelle définition du phénotype</b>	<b>55</b>
5.1	Description des symptômes de la schizophrénie . . . . .	55
5.2	Concept des symptômes positifs et négatifs . . . . .	57

<b>6</b>	<b>Études pratiques d'analyse statistique de la liaison génétique pour les deux psychoses majeures</b>	<b>60</b>
6.1	Description de l'échantillon de familles . . . . .	60
6.1.1	Sélection de l'échantillon . . . . .	60
6.1.2	Évaluation du diagnostic . . . . .	63
6.1.3	Scores des symptômes positifs et négatifs . . . . .	64
6.2	Utilisation d'un seuil de coupure sur les scores des symptômes et nouvelle définition du phénotype . . . . .	66
6.2.1	Détermination du seuil de coupure . . . . .	66
6.2.2	Concordance entre les définitions usuelles du phénotype pour chacune des deux psychoses majeures et la nouvelle définition utilisant les scores aux symptômes positifs et négatifs . . . . .	68
6.3	Spécification des composantes nécessaires pour le calcul de la vraisemblance d'un pedigree avec l'algorithme d'Elston-Stewart . . . . .	69
6.3.1	La fréquence des allèles pour les différents marqueurs génétiques utilisés . . . . .	69
6.3.2	Le choix du modèle de transmission génétique . . . . .	73
6.3.3	La pénétrance dépendante de l'âge . . . . .	74
6.4	Utilisation de deux classifications du phénotype à partir des seuils de coupure fixés . . . . .	77
6.5	Résultats des différentes analyses de linkage . . . . .	79
6.6	Résultats du test d'hétérogénéité pour un mélange de familles indissociables . . . . .	86
	<b>Conclusion</b>	<b>91</b>
	<b>Bibliographie</b>	<b>92</b>
	<b>Annexe A : Exemples des trois types de familles des études pratiques avec la légende des diagnostics</b>	<b>99</b>
	<b>Annexe B : Cartes génétiques des chromosomes 6, 11 et 18</b>	<b>105</b>

# Liste des figures

1.1	Paire de chromosomes . . . . .	4
1.2	Représentation des allèles d'un individu pour deux loci . . . . .	5
1.3	Illustration de la première loi de Mendel . . . . .	8
1.4	Illustration de la deuxième loi de Mendel . . . . .	9
2.1	Phénomène de recombinaison . . . . .	11
2.2	Représentation schématique d'un crossing over entre deux paires de chromosomes homologues portant deux gènes . . . . .	12
2.3	Illustration de la recombinaison avec un parent de phase connue et l'autre non informatif . . . . .	14
2.4	Illustration de la recombinaison avec un parent de phase inconnue et l'autre non informatif . . . . .	15
2.5	Représentation schématique d'une analyse de linkage entre le locus d'une maladie et celui d'un marqueur génétique . . . . .	17
2.6	Phénotypes d'une maladie récessive rare ainsi que d'un marqueur représenté par les allèles 1, 2, 3 et 4 . . . . .	18
2.7	Phénotypes et génotypes d'une maladie récessive rare ainsi que d'un marqueur représenté par les allèles 1, 2, 3 et 4 avec des parents de phases inconnues . . . . .	19
2.8	Phénotypes et génotypes d'une maladie récessive rare ainsi que d'un marqueur représenté par les allèles 1, 2, 3 et 4 avec des parents de phases connues . . . . .	22
2.9	Courbes $F(\text{âge})$ et $f(\text{âge})$ pour des individus de génotype susceptible dans le cas de la schizophrénie . . . . .	27
2.10	Représentation d'une famille nucléaire . . . . .	30



2.11	Représentation d'une famille étendue . . . . .	31
2.12	Courbe du Lod-score $Z(\theta)$ pour le premier exemple de la section 2.4.1 .	36
2.13	Courbe du Lod-score $Z(\theta)$ pour le deuxième exemple de la section 2.4.1	37
2.14	Pedigree contenant une boucle de consanguinité . . . . .	42
2.15	Pedigree contenant une boucle de mariage . . . . .	44
6.1	Distribution de fréquences des scores des symptômes positifs et négatifs	66
6.2	Fonction de répartition (F) et fonction de densité (f) de l'âge d'apparition de la maladie pour les individus de génotype susceptible des deux psychoses majeures. Détermination de la pénétrance de 3 individus (dénotés 1, 2 et 3) . . . . .	78
6.3	Première classification du phénotype par rapport à un seuil de coupure fixé . . . . .	78
6.4	Deuxième classification du phénotype par rapport à un seuil de coupure fixé . . . . .	79

# Liste des tableaux

2.1	Matrice de transition génétique pour le cas d'un locus autosomal avec les allèles $A$ et $a$ . . . . .	28
3.1	Risques observés de schizophrénie parmi les membres d'une famille comportant des schizophrènes . . . . .	46
5.1	Identification des symptômes positifs et négatifs . . . . .	58
6.1	Nombre total d'individus et le nombre d'individus affectés, par famille .	62
6.2	Nombre d'individus affectés selon une définition usuelle du phénotype pour chacune des deux psychoses majeures avec les diagnostics du DSM-III-R correspondants . . . . .	65
6.3	Statistiques descriptives des scores des symptômes positifs et négatifs .	67
6.4	Valeur du centile aux seuils de coupure de 0.5 et 1.0 pour les symptômes positifs et négatifs . . . . .	67
6.5	Tableau de fréquences entre une définition usuelle du phénotype correspondant à chacune des deux psychoses majeures et la nouvelle définition basée sur les scores des symptômes positifs et négatifs avec un seuil de coupure donné . . . . .	69
6.6	Proportions calculées entre diverses définitions usuelles du phénotype pour les deux psychoses majeures et la nouvelle définition basée sur le score des symptômes positifs avec des seuils de coupure de 0.5 et 1.0 . .	70
6.7	Proportions calculées entre diverses définitions usuelles du phénotype pour les deux psychoses majeures et la nouvelle définition basée sur le score des symptômes négatifs avec des seuils de coupure de 0.5 et 1.0 .	71
6.8	Fréquence observée des allèles pour les différents marqueurs génétiques des chromosomes 6, 11 et 18, basée sur un échantillon de 120 individus	72
6.9	La pénétrance dépendante de l'âge du modèle dominant . . . . .	75

6.10	La pénétrance dépendante de l'âge du modèle récessif . . . . .	76
6.11	Résultats des analyses de linkage qui ont donné des Lod-scores maximums supérieurs ou égaux à 1 pour le seuil de coupure de 0.5. Les deux classifications sont présentées à la section 6.4 . . . . .	80
6.12	Résultats des analyses de linkage qui ont donné des Lod-scores maximums supérieurs ou égaux à 1 pour le seuil de coupure de 1.0. Les deux classifications sont présentées à la section 6.4 . . . . .	81
6.13	Différences supérieures à 1 des Lod-scores maximums entre les seuils de coupure de 0.5 et 1.0, en valeur absolue, pour les symptômes positifs . .	82
6.14	Différences supérieures à 1 des Lod-scores maximums entre les seuils de coupure de 0.5 et 1.0, en valeur absolue, pour les symptômes négatifs .	83
6.15	Résultats du test d'hétérogénéité pour un mélange de familles indissociables au seuil de coupure de 0.5 . . . . .	87
6.16	Résultats du test d'hétérogénéité pour un mélange de familles indissociables au seuil de coupure de 1.0 . . . . .	88
6.17	Log-vraisemblance bivariée globale maximum avec les estimations des paramètres $\alpha$ et $\theta$ sous différentes hypothèses de l'exemple de la section 6.6 . . . . .	89

# Introduction

La schizophrénie est un trouble psychotique, c'est-à-dire qu'il s'agit d'un trouble mental caractérisé par une altération profonde de la personnalité et des fonctions intellectuelles, et le fait que le sujet n'en reconnaît pas le caractère morbide (Lalonde et al., 1988). Elle comporte toujours, à un moment ou à un autre de la maladie, des idées délirantes, des hallucinations ou certaines perturbations caractéristiques de l'affectivité ou du cours de la pensée (DSM-III-R, 1987). La prévalence de cette maladie psychiatrique dans la population est d'environ 1 %.

Les psychiatres différencient deux catégories de symptômes de la schizophrénie : les symptômes productifs (positifs) et les symptômes déficitaires (négatifs). Les symptômes positifs correspondent aux hallucinations, aux idées délirantes, au comportement bizarre et au trouble de la pensée formelle tandis que les symptômes négatifs se rapportent à la pauvreté affective, l'alogie, l'apathie et l'anhédonie.

Les études familiales, les études de jumeaux et les études d'adoption, qui se succèdent depuis le début du siècle, ont fourni des hypothèses jusqu'ici les plus probantes quant à la nature de la schizophrénie (Lalonde et al., 1988). Ces études sont habituellement axées sur la question fondamentale du départage entre l'inné et l'acquis, c'est-à-dire entre le biologique et l'environnement. Les études familiales suggèrent qu'il y a une forte agrégation familiale de la schizophrénie (Kendler et Scott, 1993). D'autre part, les études de jumeaux et d'adoption suggèrent que les facteurs génétiques ont un rôle majeur dans cette agrégation familiale.

Ainsi, le présent ouvrage s'intéresse à la recherche, dans le génome humain, de la localisation du ou des gène(s) responsable(s) de la schizophrénie. L'analyse de linkage est la méthode pour y répondre. Les symptômes positifs et négatifs seront utilisés comme nouvelle définition du phénotype (expression des gènes) plutôt que de se servir du diagnostic usuel de schizophrénie. Une analyse de linkage sera effectuée entre chacun de ces symptômes et un ensemble de marqueurs génétiques.

Le premier chapitre va décrire les notions de base de la génétique humaine, nécessaires à la compréhension de l'analyse de linkage. C'est au deuxième chapitre qu'on expliquera le phénomène de recombinaison et l'analyse de linkage entre le locus d'une maladie et celui d'un marqueur génétique. Le troisième chapitre sera consacré aux études familiales, de jumeaux et d'adoption réalisées antérieurement sur la schizophrénie. Ensuite, au chapitre quatre, on décrira la complexité de l'analyse de linkage pour cette maladie. Au chapitre cinq, on va discuter de l'utilisation des symptômes positifs et négatifs comme nouvelle définition du phénotype. Finalement, le dernier chapitre consistera en des études pratiques d'analyse statistique de la liaison génétique avec un échantillon de familles provenant du Centre de recherche Université Laval Robert-Giffard.

# Chapitre 1

## Notions de base de la génétique humaine

### 1.1 Les concepts fondamentaux de la génétique humaine

La génétique est la science de l'hérédité (Darnell et al., 1988). Elle comporte au moins deux aspects : la transmission des gènes d'une génération à la suivante et le processus par lequel les gènes sont exprimés pour donner le *phénotype*. Le *phénotype* correspond à l'expression physique observable du *génotype* d'un individu, par exemple la couleur des cheveux (Mange et Mange, 1994). Quant au *génotype*, il est la constitution génétique responsable du *phénotype* d'un individu. Le *phénotype* peut être non seulement qualitatif, telle la couleur des yeux, mais aussi quantitatif, le taux de cholestérol dans le sang en est un exemple (Ott, 1991).

L'information héréditaire est contenue dans les chromosomes (Khoury et al., 1993). Chaque être humain possède 23 paires de chromosomes. Les 22 premières paires de chromosomes sont appelées *autosomes* et la 23<sup>e</sup> paire de chromosomes correspond aux *chromosomes sexuels*. Chez la femme, cette 23<sup>e</sup> paire est XX, tandis que chez l'homme, c'est XY. La plupart des cellules du corps de l'être humain sont *diploïdes*, c'est-à-dire qu'elles ont deux jeux homologues de chromosomes dont l'un provient du père et l'autre de la mère (Mange et Mange, 1994). Par exemple, les cellules des tissus musculaires

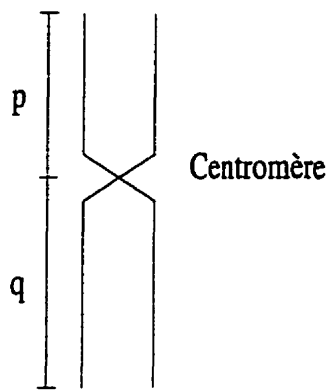


Figure 1.1: Paire de chromosomes

sont *diploïdes*. En ayant deux exemplaires de chaque chromosome, on a par conséquent deux exemplaires de chaque gène. Les gamètes sont des cellules *haploïdes*. Ils consistent en des cellules reproductrices, mâles ou femelles, dont le noyau ne contient qu'un seul chromosome de chaque paire, et qui, en s'unissant à une autre cellule reproductrice de sexe opposé, forme l'oeuf d'où sortira un nouvel être vivant. Chaque chromosome, composé de deux bras, est séparé par le *centromère*; par convention, le *petit bras* est dénoté par p et le *long bras* est dénoté par q, voir la figure 1.1 (Ott, 1991).

Un gène est un segment d'A.D.N. (acide désoxyribonucléique, constituant essentiel des chromosomes) qui code pour une protéine particulière et conditionne la transmission et l'expression d'un caractère héréditaire déterminé (Mange et Mange, 1994). Un gène est donc responsable d'un caractère phénotypique donné. Cependant, plusieurs gènes peuvent porter sur un même phénotype (Ott, 1991). Par exemple, chez l'humain, la pigmentation de la peau est déterminée par quatre gènes. L'endroit sur un chromosome où se situe un gène est appelé le *locus* et le pluriel de *locus* est *loci*. Les formes sous lesquelles un gène peut se présenter sont appelées *allèles*; chacun a une expression physique différente.

On remarque à la figure 1.2 qu'un individu peut avoir, par exemple, les allèles  $A_1$  et  $A_2$  au locus A. Cependant, on pourrait observer d'autres allèles à ce locus : les allèles  $A_3, A_4, \dots, A_n$ . Un génotype est une combinaison d'allèles  $A_i A_j$  parmi les  $n$  allèles constituant le locus A. Ainsi, le génotype d'un individu à un locus donné est constitué

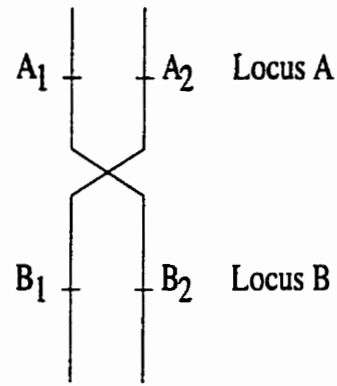


Figure 1.2: Représentation des allèles d'un individu pour deux loci

d'une paire d'allèles, un allèle provenant du père et l'autre de la mère (Mange et Mange, 1994). En ne considérant qu'un seul locus, on parle alors d'un *génotype univarié*, par exemple  $A_1A_2$ . Cependant, lorsque l'on considère un génotype pour deux loci, on parle alors d'un *génotype bivarié*, par exemple  $A_1B_1/A_2B_2$ . À un locus donné, lorsqu'un génotype est constitué de deux allèles identiques, ce génotype est appelé *homozygote*. Par contre, si les deux allèles sont différents, le génotype est appelé *hétérozygote*. Un allèle *dominant* est dit d'un allèle dont l'effet phénotypique se manifeste lorsqu'il est présent à l'état *hétérozygote* aussi bien qu'*homozygote*. Un allèle *récessif* est dit d'un allèle dont l'effet phénotypique se manifeste uniquement lorsqu'il est présent à l'état *homozygote*. Deux allèles sont dits *codominants* lorsque les phénotypes conférés par ces deux allèles sont présents en même temps. Pour le cas d'une maladie due à un allèle *récessif*, on note l'allèle sain par une lettre majuscule et l'allèle malade par une lettre minuscule. Tandis que pour le cas d'une maladie due à un allèle *dominant*, l'allèle sain est noté par une lettre minuscule et l'allèle malade est noté par une lettre majuscule.

Dans une population donnée, les fréquences relatives des différents allèles d'un gène correspondent aux fréquences d'un gène (Ott, 1991). Un gène est dit *polymorphe* lorsqu'il y a présence dans la population d'au moins deux allèles relativement communs de ce gène (Mange et Mange, 1994). En génétique humaine, une mesure du degré de *polymorphisme* d'un gène qui est souvent utilisée est la valeur du PIC ("Polymor-



phism Information Content") (Ott, 1991). Le PIC est calculé de la façon suivante :

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2 = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n p_i p_j (1 - p_i p_j) \quad (1.1)$$

où  $p_i$  est la fréquence de l'allèle  $i$  dans la population, pour un gène avec  $n$  allèles. Si, dans la population, un gène est *polymorphe*, alors deux personnes choisies au hasard dans celle-ci auront vraisemblablement des génotypes différents et une proportion relativement grande de ces génotypes seront hétérozygotes (Mange et Mange, 1994).

Prenons l'exemple du groupe sanguin afin d'illustrer les notions de base de la génétique humaine (Ott, 1991). Au locus du groupe sanguin, on retrouve les allèles  $A$ ,  $B$  et  $O$ . Les génotypes possibles pour le groupe sanguin sont les suivants :  $AA$ ,  $AO$ ,  $BB$ ,  $BO$ ,  $AB$  et  $OO$ . Les génotypes  $AA$ ,  $BB$  et  $OO$  sont homozygotes. Par contre, les génotypes  $AO$ ,  $BO$  et  $AB$  sont hétérozygotes. Le phénotype associé aux génotypes  $AA$  et  $AO$  est  $A$ . Celui pour les génotypes  $BB$  et  $BO$  est  $B$ . Le phénotype est  $AB$  pour le génotype  $AB$ . Et puis, finalement, le phénotype  $O$  correspond au génotype  $OO$ . On observe alors que les allèles  $A$  et  $B$  sont dominants par rapport à l'allèle  $O$ . L'allèle  $O$  est récessif par rapport aux allèles  $A$  et  $B$ . De plus, les allèles  $A$  et  $B$  sont codominants, considérant le phénotype  $AB$ . Le gène responsable du groupe sanguin est très polymorphe (Mange et Mange, 1994).

Sous des conditions d'accouplement aléatoire et en l'absence de perturbations telles la migration, la mutation et la sélection naturelle, une population est appelée à être en équilibre d'Hardy-Weinberg, c'est-à-dire que les fréquences des génotypes sont distribuées selon la proportion des gènes ou allèles dans la population, et cette distribution demeure constante d'une génération à l'autre (Ott, 1991). En d'autres mots, la fréquence des génotypes d'une génération d'enfants dépend seulement de la fréquence des allèles et non de la fréquence des génotypes dans la génération des parents. Par exemple, pour un locus donné, on retrouve les allèles  $A$  et  $a$  en proportion  $p$  et  $q$  (où

$q = 1 - p$ ), respectivement dans la population, alors les fréquences des génotypes auront la distribution suivante :

- $P(AA) = p^2$ ;
- $P(Aa) = 2pq$ ;
- $P(aa) = q^2$ .

On dit alors que cette population est en équilibre d'Hardy-Weinberg.

Les combinaisons d'allèles de différents gènes, reçues par un individu d'un parent, sont appelées *haplotypes* (Mange et Mange, 1994). En se référant de nouveau à la figure 1.2, on observe que l'individu a reçu l'*haplotype*  $A_1-B_1$  d'un de ses parents et l'*haplotype*  $A_2-B_2$  de l'autre parent.

## 1.2 Les deux lois génétiques de Mendel

La découverte des principes de base de l'hérédité a été réalisée par le moine autrichien Gregor Mendel (1822-1884) (Darnell et al., 1988). Le caractère génial de ses travaux tient d'abord au choix d'un organisme vivant simple, le petit pois, dont la croissance et la fécondation sont faciles à assurer. Ensuite, Mendel a choisi d'étudier des caractères héréditaires (couleur et texture des graines et de leurs enveloppes, position des fleurs sur les tiges et longueur des tiges) dont la détermination est facile, si bien que le dénombrement des descendants de chaque type, après les croisements, pouvait se faire sans ambiguïté. Enfin et surtout, il formula ses conclusions en s'appuyant sur des résultats mathématiques rigoureux à partir d'un grand nombre de descendants obtenus par ses croisements expérimentaux.

En se basant sur ses résultats obtenus, Mendel a pu ainsi établir deux lois génétiques concernant la transmission des traits simples (Darnell et al., 1988). Dans le cadre de ses expériences, la couleur des graines est un exemple d'un trait simple.

La première loi de Mendel est la loi de la séparation ou de la ségrégation des allèles d'un même gène. Lors de la formation des gamètes, les deux allèles se séparent et chaque gamète n'en contient qu'un. La figure 1.3 illustre cette première loi de Mendel, en considérant, par exemple, les allèles  $A$  (dominant) et  $a$  (récessif) d'un gène donné.

La deuxième loi de Mendel postule que la loi de la séparation ou de la ségrégation est dite indépendante, c'est-à-dire que la séparation des allèles d'un gène est indépendante de celle des allèles d'autres gènes. La figure 1.4 illustre cette deuxième loi de Mendel. Les allèles  $R$  et  $Y$  sont dominants, tandis que les allèles  $r$  et  $y$  sont récessifs. Quand Mendel croisait deux variétés pures de petits pois pour chacun des deux caractères, par exemple des poids ronds ( $R$ ) et jaunes ( $J$ ) avec des poids ridés ( $r$ ) et verts ( $y$ ), les plants de la deuxième génération montraient quatre combinaisons différentes des deux caractères, c'est-à-dire 9 ronds et jaunes, 3 ronds et verts, 3 ridés et jaunes et 1 ridé et vert. Ces quatre combinaisons correspondaient au rapport phénotypique (9 : 3 : 3 : 1), ce qui correspond à la répartition espérée sous l'hypothèse d'une transmission indépendante des deux caractères, c'est-à-dire sous l'hypothèse que chacun des génotypes du tableau de la figure 1.4 ait une probabilité de 1/16 d'être observé. Ainsi, Mendel conclua que les deux caractères, forme et couleur, étaient distribués l'un et

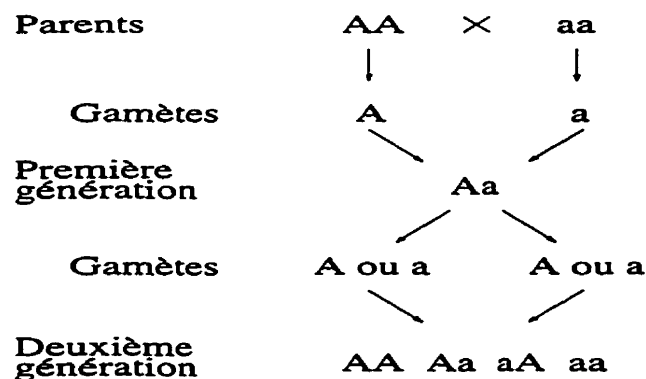


Figure 1.3: Illustration de la première loi de Mendel

l'autre de façon indépendante chez les descendants.

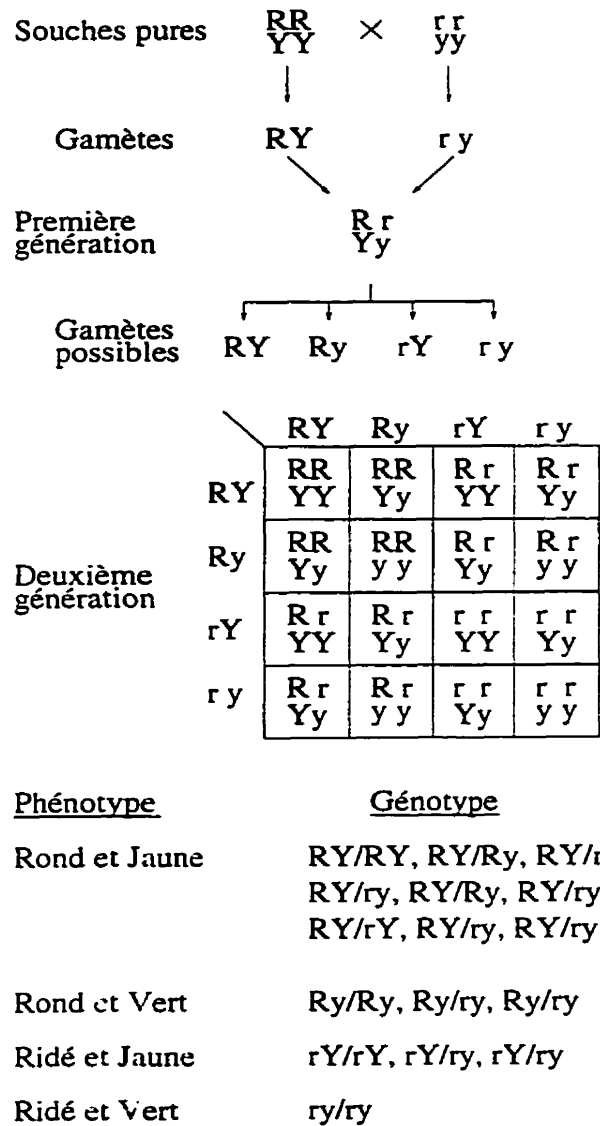


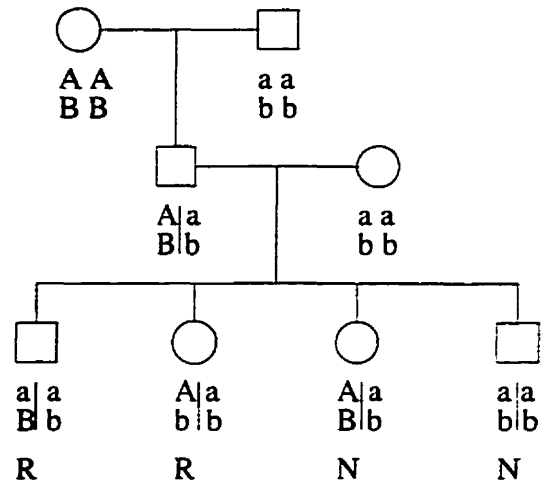
Figure 1.4: Illustration de la deuxième loi de Mendel

# Chapitre 2

## La recombinaison et l'analyse de linkage

### 2.1 La recombinaison

Considérons le cas de deux loci situés sur deux chromosomes différents, retrouvant les allèles  $A$  et  $a$  au premier locus et les allèles  $B$  et  $b$  au deuxième locus. À la figure 2.1, ce premier cas y est illustré. On représente par un cercle, un individu de sexe féminin et par un carré, un individu de sexe masculin. On considère que le père de la deuxième génération est de phase connue, c'est-à-dire que l'on sait l'haplotype reçu de chacun de ses parents. Il a reçu l'haplotype  $a-b$  de son père et l'haplotype  $A-B$  de sa mère. Lorsque la phase est connue, on utilise une ligne verticale pour distinguer les allèles maternels des allèles paternels. Les deux grands-parents de la première génération ainsi que la mère de la deuxième génération possèdent des génotypes non informatifs pour une analyse de linkage, car ils ont au moins un locus homozygote parmi les deux loci considérés. On remarque que les gamètes produits par la mère sont tous  $a-b$ . Selon la deuxième loi de Mendel, les gamètes possibles qui proviendront du père sont les suivants :  $A-B$  ( $1/4$ ),  $A-b$  ( $1/4$ ),  $a-B$  ( $1/4$ ) ou  $a-b$  ( $1/4$ ). Les quatre enfants de la troisième génération ont reçu du père soit l'haplotype  $A-B$ , soit l'haplotype  $A-b$ , soit l'haplotype  $a-B$ , soit l'haplotype  $a-b$ . Les haplotypes  $A-B$  et  $a-b$  chez les enfants sont identiques à ceux que leurs parents ont produits. Il s'agit d'*haplotypes non recombinants*, indiqués par la lettre N (Mange



1. Deux loci sur deux chromosomes différents :	1/4	1/4	1/4	1/4
2. Deux loci loin l'un de l'autre sur un même chromosome :	1/4	1/4	1/4	1/4
3. Deux loci près l'un de l'autre sur un même chromosome (où $\theta < 1/2$ ) :	$\theta$	$\theta$	$1 - \theta$	$1 - \theta$

Figure 2.1: Phénomène de recombinaison

et Mange, 1994). Par contre, les haplotypes  $A-b$  et  $a-B$  correspondent à de nouvelles combinaisons d'allèles chez les enfants, différentes des combinaisons parentales; il s'agit d'*haplotypes recombinants*, indiqués par la lettre R.

Ainsi, dans ce premier cas, la fréquence des haplotypes recombinants est de  $1/2$ . Le pourcentage de recombinaison est de 50 % pour cet exemple de recombinaison interchromosomique.

Considérons maintenant le cas de deux loci situés loin l'un de l'autre sur un même chromosome, à la figure 2.1. C'est le biologiste américain Thomas Hunt Morgan (1866-1945) qui a postulé qu'au cours de la formation des gamètes, il se produit occasionnellement des échanges physiques de parties de chromosomes par un processus qui est appelé *crossing over* (Darnell et al., 1988). Ce processus d'échange entre les chromosomes

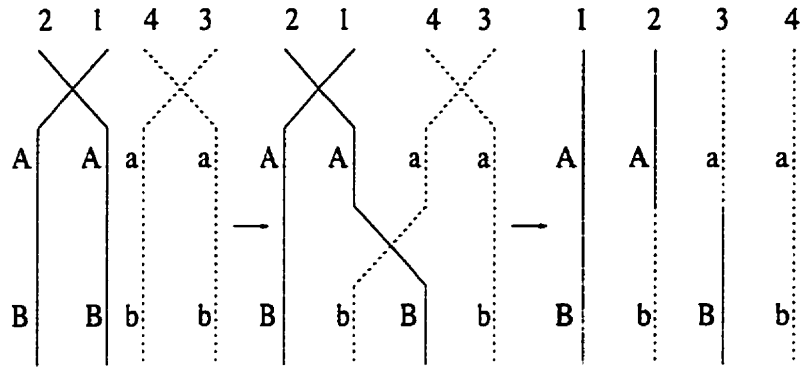


Figure 2.2: Représentation schématique d'un crossing over entre deux paires de chromosomes homologues portant deux gènes

homologues dupliqués est illustré à la figure 2.2. Il faut un nombre impair de *crossing over* dans l'intervalle entre deux gènes pour obtenir la recombinaison. Si les deux loci sont séparés par une grande distance, il y aura alors un nombre élevé d'échanges de parties de chromosomes entre ces deux loci. La probabilité d'en avoir un nombre impair, produisant des gamètes recombinants, sera approximativement égale à celle d'en avoir un nombre pair, produisant des gamètes non recombinants.

Dans ce deuxième cas, la fréquence des haplotypes recombinants est de  $1/2$ . Pour cet exemple de recombinaison intrachromosomique, le pourcentage de recombinaison est de 50 %.

Ce qu'on peut dire au sujet de la recombinaison, c'est que le pourcentage de recombinaison est identique, correspondant à 50 %, dans les cas où les deux gènes sont, soit séparés par une grande distance physique sur un même chromosome, soit situés sur deux chromosomes différents. C'est pour ces raisons que le taux maximal de recombinaison est de 50 %.

Considérons finalement le cas de deux loci situés près l'un de l'autre sur un même chromosome, à la figure 2.1. Il est alors peu probable qu'un crossing over ait lieu entre ces deux loci. Ainsi, les allèles d'un haplotype transmis d'un grand-parent à un parent ont tendance à être transmis de nouveau, constituant le même haplotype,

du parent à la progéniture. En d'autres mots, les allèles de différents gènes semblent être génétiquement liés, et ce phénomène est appelé *linkage génétique*. Ces gènes liés représentent une exception à la deuxième loi de Mendel, à savoir que la séparation des allèles d'un gène est indépendante de celle des allèles d'autres gènes.

Le degré de linkage génétique est mesuré par la *fraction de recombinaison*  $\theta$  qui est la probabilité qu'un gamète produit par un parent soit un recombinant (Ott, 1991). Une *fraction de recombinaison*  $\theta = 1/2$ , entre deux gènes, caractérise des gènes non liés, alors qu'une *fraction de recombinaison*  $\theta < 1/2$ , entre deux gènes, caractérise des gènes liés.

La fraction de recombinaison  $\theta$  est donc un paramètre important qui reflète la proximité entre deux loci. Elle s'interprète comme une *distance génétique*, c'est-à-dire une fréquence d'événements de recombinaison, qui est liée à la *distance physique* entre deux loci. Si on pouvait estimer la fraction de recombinaison  $\theta$ , on connaîtrait ainsi la proximité entre deux loci donnés.

Lorsqu'on connaît le génotype à chacun de deux loci considérés de chacun des membres d'une famille répartis sur au moins trois générations, le nombre d'événements de recombinaison peut être directement compté à l'intérieur de cette famille. Considérons les deux exemples suivants.

Le premier exemple est représenté par le pedigree à la figure 2.3. Au premier locus, on retrouve les allèles  $A$ ,  $B$ ,  $C$  et  $D$  et les allèles 1, 2, 3 et 4 caractérisent le deuxième locus. On considère que le père de la deuxième génération est de phase connue, c'est-à-dire que l'on sait l'haplotype reçu de chacun de ses parents. Dans cet exemple, il a reçu l'haplotype  $A-1$  de son père et l'haplotype  $B-2$  de sa mère. Les deux parents de la première génération ainsi que la mère de la deuxième génération possèdent des génotypes non informatifs pour une analyse de linkage, car ils ont au moins un locus homozygote parmi les deux loci considérés. Ainsi, par exemple, pour la mère de la deuxième génération, chaque gamète  $C-3$  qu'elle produira peut être un recombinant ou



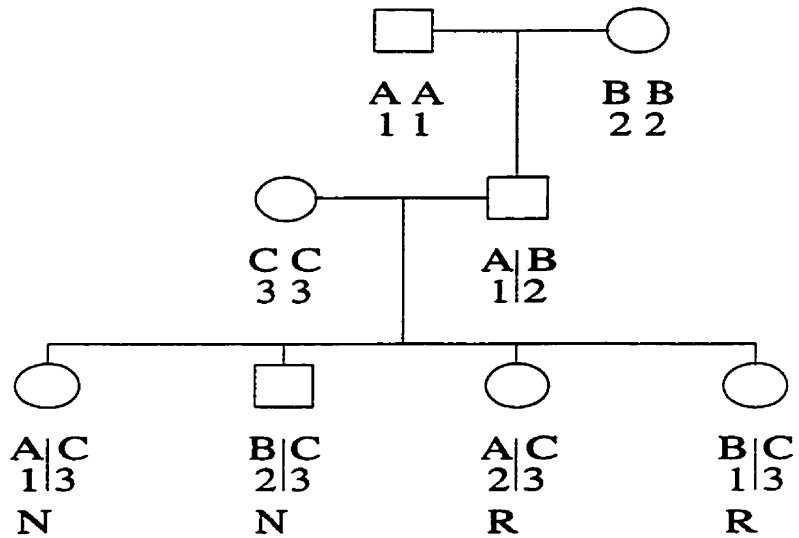


Figure 2.3: Illustration de la recombinaison avec un parent de phase connue et l'autre non informatif

un non-recombinant; on ne peut pas les distinguer.

On remarque que la première fille de la troisième génération a reçu l'haplotype non recombinant  $A-1$  de son père. La deuxième fille de la même génération a reçu l'haplotype recombinant  $A-2$  de son père. Les autres résultats s'obtiennent en faisant le même raisonnement. Il y a deux événements de recombinaison et deux événements de non-recombinaison. Pour cet exemple, la fraction de recombinaison vaut  $1/2$ .

Le deuxième exemple est illustré à la figure 2.4. Considérons le cas où les génotypes des grands-parents de la première génération sont inconnus, ainsi la phase du père de la deuxième génération est inconnue.

Sous la phase I, les résultats sont les mêmes que ceux de l'exemple précédent. Par contre, sous la phase II, on constate que les résultats obtenus représentent l'inverse de ceux de la phase I. Par exemple, sous la phase II, la première fille a reçu l'haplotype recombinant  $A-1$  de son père, alors que sous la phase I, l'haplotype  $A-1$  est considéré non recombinant. Les autres résultats s'obtiennent en procédant de la même façon. La fraction de recombinaison est de  $1/2$  pour chacune des phases.

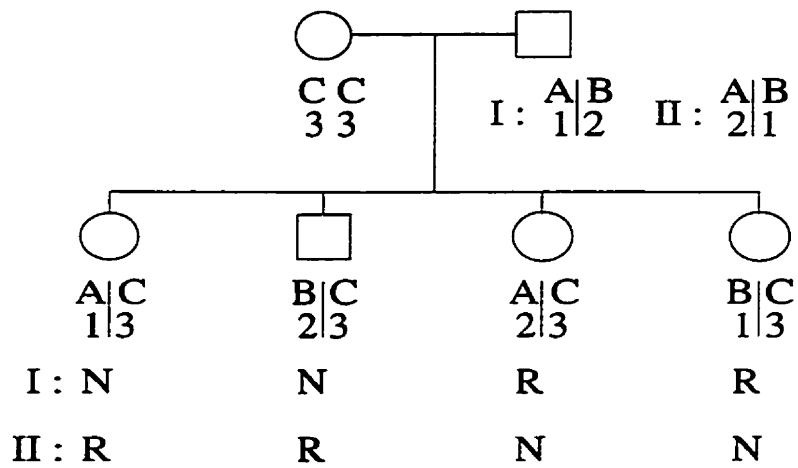


Figure 2.4: Illustration de la recombinaison avec un parent de phase inconnue et l'autre non informatif

## 2.2 Distinction entre distance génétique et distance physique

Il est important de faire la distinction entre distance génétique et distance physique (Ott, 1991). La distance physique entre deux loci est exprimée en nombre de paires de base de l'A.D.N. qui les séparent. L'A.D.N. est constitué de quatre bases : l'adénine (A), la thymine (T), la cytosine (C) et la guanine (G). Les paires de bases de l'A.D.N. sont A-T et C-G. La distance génétique entre deux loci représente plutôt une mesure d'une fréquence d'un événement génétique, la recombinaison. Pour une fraction de recombinaison égale à  $1/2$ , on considère que les deux loci sont situés à une distance infinie l'un de l'autre. Par exemple, ce serait le cas de deux gènes situés sur des chromosomes différents, ou bien de deux gènes étant éloignés l'un de l'autre sur un même chromosome. Pour une fraction de recombinaison inférieure à  $1/2$ , on considère que les deux loci sont proches physiquement l'un de l'autre sur un même chromosome.

Ainsi, les objectifs d'une analyse de linkage sont d'abord, d'estimer la fraction de recombinaison  $\theta$  et ensuite, de tester l'hypothèse nulle postulant que  $\theta = 1/2$  (absence de linkage) par rapport à la contre-hypothèse postulant que  $\theta < 1/2$  (présence de linkage).

## 2.3 La cartographie du génome humain

Le projet du génome humain, qui se poursuit toujours actuellement, consiste en une collaboration internationale d'équipes de chercheurs dont le but ultime est de déterminer la séquence de tous les gènes du génome humain (Darnell et al., 1988). Ce projet nous fournit des marqueurs génétiques dont on connaît les positions sur un chromosome, grâce aux cartographies génétiques et physiques des marqueurs qui ont été réalisées avant le début du séquençage des gènes. Un marqueur génétique est utilisé pour "identifier" un endroit sur une carte génétique.

L'intérêt d'un marqueur n'est donc pas sa fonction mais plutôt son emplacement sur le chromosome. Les marqueurs génétiques sont utilisés pour étudier le phénomène de linkage génétique (Ott, 1991). En particulier, les marqueurs génétiques sont utiles pour identifier l'emplacement d'un gène d'intérêt, tel que le gène d'une maladie, relativement au locus d'un marqueur génétique. L'idée est d'observer le phénomène de linkage génétique entre le marqueur, dont la position est connue, et le gène de la maladie, dont on ne connaît pas encore la position. Si une présence de linkage est détectée (ce qui s'exprime par le rejet de l'hypothèse nulle  $H_0 : \theta = 1/2$ ), alors on pourra conclure que le gène de la maladie se situe à proximité du marqueur génétique.

## 2.4 Analyse de linkage entre le locus d'une maladie et celui d'un marqueur génétique

Lorsqu'on fait une analyse de linkage entre le locus d'une maladie et celui d'un marqueur génétique, on connaît d'abord le phénotype de la maladie ainsi que le génotype au locus du marqueur. Par contre, ce que l'on ne connaît pas et ce dont on cherche à déterminer finalement, c'est le génotype au locus de la maladie. La figure 2.5 illustre sous forme schématique ce type d'analyse de linkage. Ce qui est connu est encadré, cependant ce qui ne l'est pas, est encerclé. En établissant d'abord un modèle héréditaire

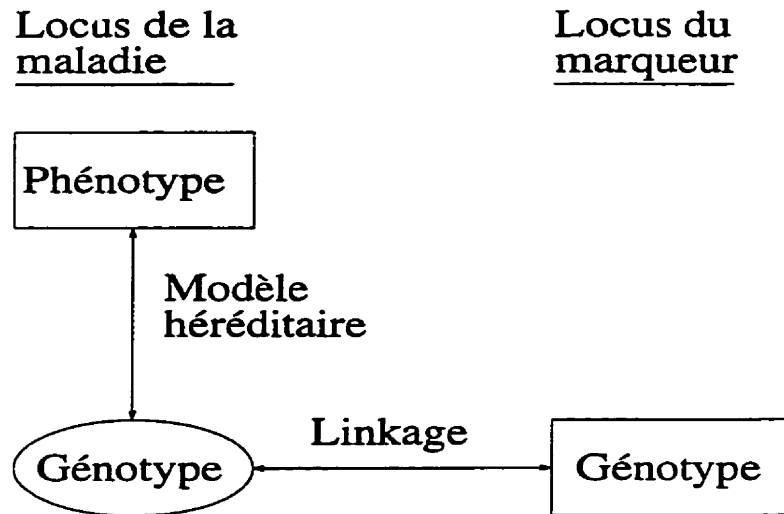


Figure 2.5: Représentation schématique d'une analyse de linkage entre le locus d'une maladie et celui d'un marqueur génétique

concernant le mode de transmission de la maladie, qui fera le lien entre le phénotype et le génotype de la maladie, et en faisant par la suite une analyse de linkage entre le locus de la maladie et celui d'un marqueur, on pourra ainsi déterminer l'emplacement de la maladie et par conséquent, le génotype au locus de la maladie.

#### 2.4.1 Calcul de la fonction de vraisemblance pour une famille en fonction de la fraction de recombinaison $\theta$

La famille qui sera utilisée pour faire deux exemples de calcul d'une fonction de vraisemblance se retrouve à la figure 2.6.

Il s'agit d'une maladie récessive rare, étant donné que la fréquence de l'allèle malade dans la population est très petite. De plus, on suppose que la pénétrance est complète et qu'il n'y a pas de phénocopie. La pénétrance complète signifie que si un individu a un génotype malade, alors il exprimera la maladie. Une phénocopie représente un individu qui a un génotype sain et qui est malgré tout atteint de la maladie. Les individus en noir sur la figure 2.6 sont affectés par cette maladie. Le marqueur génétique utilisé, situé sur le chromosome 3, est composé de quatre allèles également fréquents.

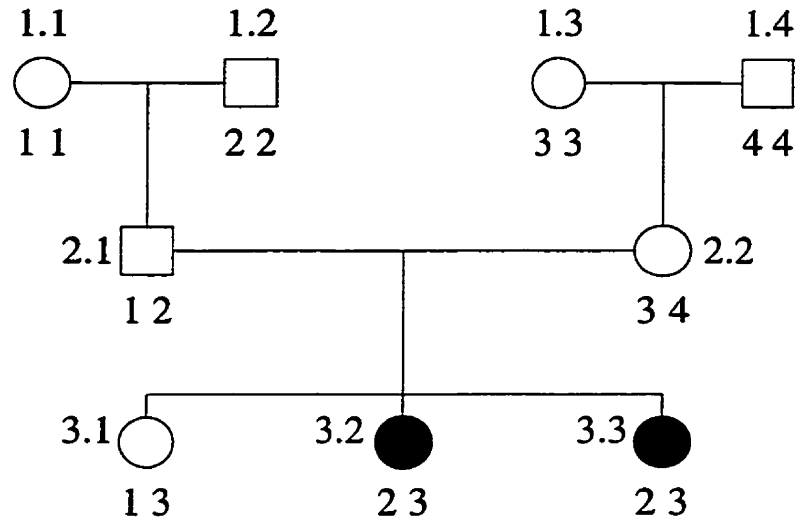


Figure 2.6: Phénotypes d'une maladie récessive rare ainsi que d'un marqueur représenté par les allèles 1, 2, 3 et 4

Le premier exemple de calcul d'une fonction de vraisemblance concernera la famille de la figure 2.6.

La première chose que l'on doit faire est de déduire du mieux possible l'ensemble des génotypes sous-jacents aux phénotypes observés. La figure 2.7 représente les phénotypes et les génotypes de chacun des individus pour ce premier exemple. Les individus 3.2 et 3.3 sont tous les deux homozygotes au locus de la maladie, soit  $dd$ , puisque la maladie est récessive. Ces individus ont deux exemplaires de l'allèle malade  $d$ , c'est-à-dire que l'un provient du père et l'autre de la mère. Ils ont ainsi reçu l'haplotype  $d-2$  de l'individu 2.1 et l'haplotype  $d-3$  de l'individu 2.2. Leur génotype est donc  $d2/d3$ . Les parents des individus 3.2 et 3.3, correspondant aux individus 2.1 et 2.2, sont porteurs de l'allèle malade  $d$  et ils ne sont pas atteints par la maladie, donc ils sont tous les deux hétérozygotes au locus de la maladie, c'est-à-dire  $Dd$ . Cela implique que l'individu 2.1 a reçu, soit l'haplotype  $d-2$  de l'individu 1.2, soit l'haplotype  $D-2$  du même individu, étant donné que les individus 1.1 et 1.2 sont sains. Ainsi, les individus 1.1 et 1.2 ont le génotype  $Dd$  au locus de la maladie. L'individu 2.1 est alors de phase inconnue. Son génotype est, soit  $D1/d2$  (phase I), soit  $d1/D2$  (phase II). Il s'agit d'une situation semblable pour l'individu 2.2. Étant donné que les individus 2.1 et 2.2 sont de phases

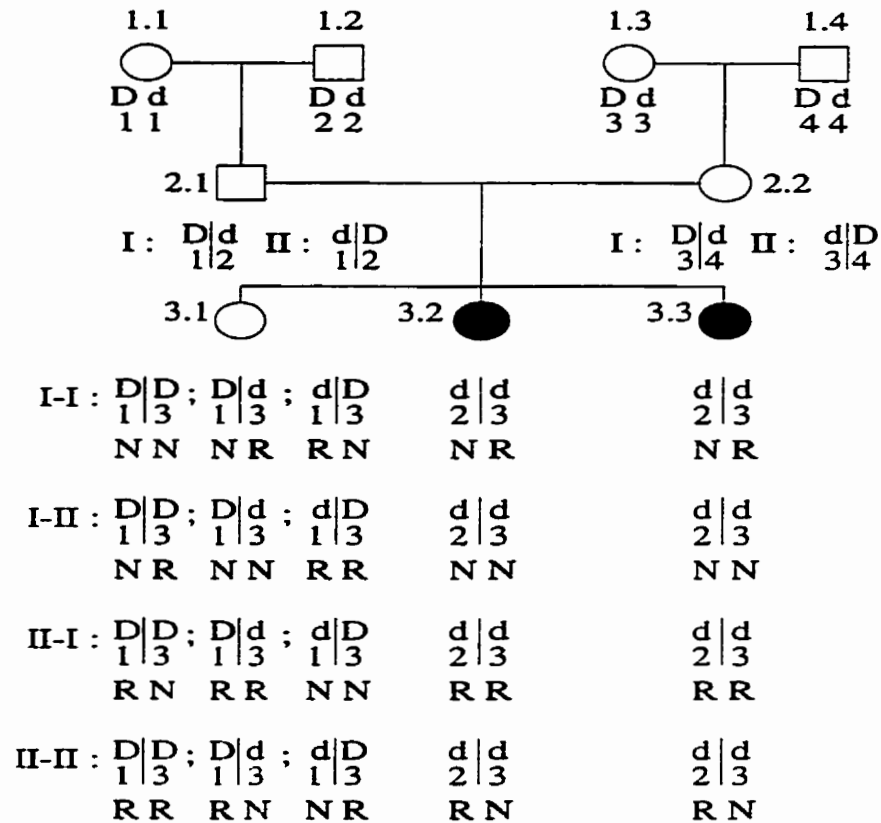


Figure 2.7: Phénotypes et génotypes d'une maladie récessive rare ainsi que d'un marqueur représenté par les allèles 1, 2, 3 et 4 avec des parents de phases inconnues

inconnues, par conséquent il y aura quatre combinaisons possibles des phases. En considérant la phase I de l'individu 2.1 (génotype  $D1/d2$ ) ainsi que la phase I de l'individu 2.2 (génotype  $D3/d4$ ) et que l'individu 3.1 est sain, il y a deux génotypes possibles au locus de la maladie pour l'individu 3.1. Il peut avoir, soit le génotype  $DD$ , soit le génotype  $Dd$ . Au locus du marqueur, il y a un seul génotype possible, c'est-à-dire 13. En tenant compte de la phase I de chacun des parents de cet individu 3.1, il y a trois génotypes bivariés possibles pour cet individu. Ces génotypes bivariés sont :  $D1/D3$  ou  $D1/d3$  ou  $d1/D3$ . Ceux-ci demeurent les mêmes pour chacune des trois combinaisons restantes des phases.

On va maintenant déterminer pour chacun des haplotypes des individus 3.1, 3.2 et 3.3 s'il s'agit d'un recombinant ou d'un non-recombinant. Commençons tout d'abord avec la phase I de chacun de leurs parents et ensuite, le raisonnement sera le même pour

les trois combinaisons restantes des phases. En ce qui concerne le premier génotype possible  $D1/D3$  de l'individu 3.1, l'haplotype  $D-1$  est non recombinant selon son père et l'haplotype  $D-3$  est non recombinant selon sa mère. Pour le deuxième génotype possible  $D1/d3$  du même individu, l'haplotype  $D-1$  est non recombinant selon son père et l'haplotype  $d-3$  est recombinant selon sa mère. Finalement, pour le troisième génotype possible  $d1/D3$  de l'individu 3.1, l'haplotype  $d-1$  est recombinant selon son père et l'haplotype  $D-3$  est non recombinant selon sa mère. Pour les individus 3.2 et 3.3, ayant le même génotype, l'haplotype  $d-2$  est non recombinant selon le père et l'haplotype  $d-3$  est recombinant selon la mère.

À ce stade, on est en mesure de calculer la fonction de vraisemblance de la famille à la figure 2.7. La probabilité qu'un haplotype soit recombinant est représentée par la fraction de recombinaison  $\theta$  et donc, la probabilité qu'un haplotype soit non recombinant est  $(1 - \theta)$ . La fonction de vraisemblance est la suivante :

$$\begin{aligned}
 L(\theta) = & P(\text{génotypes des 3 enfants} \mid \text{phases I-I}) P(\text{phases I-I}) + \\
 & P(\text{génotypes des 3 enfants} \mid \text{phases I-II}) P(\text{phases I-II}) + \\
 & P(\text{génotypes des 3 enfants} \mid \text{phases II-I}) P(\text{phases II-I}) + \\
 & P(\text{génotypes des 3 enfants} \mid \text{phases II-II}) P(\text{phases II-II}).
 \end{aligned}$$

La probabilité de réalisation de chacune des quatre combinaisons possibles des phases est la même, c'est-à-dire  $1/4$ .

Pour la combinaison des phases I-I :

$$\begin{aligned}
 P(\text{génotype de la 1}^{\text{ère}} \text{ fille}) &= (1/3) (1 - \theta)^2 + (1/3) (1 - \theta) \theta + (1/3) \theta (1 - \theta) \\
 \Rightarrow P(\text{génotype de la 1}^{\text{ère}} \text{ fille}) &= 1/3 (1 - \theta^2); \\
 P(\text{génotype de la 2}^{\text{e}} \text{ fille}) &= (1 - \theta) \theta;
 \end{aligned}$$

$$P(\text{génotype de la 3e fille}) = (1 - \theta) \theta.$$

Les probabilités des génotypes des enfants pour les autres combinaisons des phases se calculent de la même façon. On obtient ainsi la fonction de vraisemblance de la famille, en fonction de la fraction de recombinaison  $\theta$ , qui s'exprime comme suit :

$$L(\theta) = K [ 1/3 (1 - \theta^2) (1 - \theta)\theta (1 - \theta)\theta ] [ 1/4 ] +$$

$$K [ 1/3 (1 - \theta + \theta^2) (1 - \theta)^2 (1 - \theta)^2 ] [ 1/4 ] +$$

$$K [ 1/3 (1 - \theta + \theta^2) \theta^2 \theta^2 ] [ 1/4 ] +$$

$$K [ 1/3 (2\theta - \theta^2) \theta(1 - \theta) \theta(1 - \theta) ] [ 1/4 ]$$

$$\Rightarrow L(\theta) = K (1/12) (1 - 5\theta + 12\theta^2 - 14\theta^3 + 7\theta^4)$$

où  $K$  est une constante qui sera éliminée par simplification lors du calcul de  $L'(\theta)$ , on en discutera à la section 2.4.3.

Le deuxième exemple de calcul d'une fonction de vraisemblance concernera la famille à la figure 2.6, cependant en considérant aussi les individus 1.2 et 1.3 atteints par la maladie.

La figure 2.8 représente les phénotypes et les génotypes de chacun des individus pour ce deuxième exemple. Les individus 1.2 et 1.3 ont le génotype  $dd$  au locus de la maladie, car la maladie est récessive. L'individu 2.1 a nécessairement le génotype  $D1/d2$ , car il ne pouvait recevoir de son père que l'haplotype  $d-2$  et recevoir de sa mère, en considérant qu'il n'est pas atteint par la maladie, que l'haplotype  $D-1$ . Il s'agit d'une situation semblable pour l'individu 2.2 ayant le génotype  $d3/D4$ . Les phases des deux parents sont alors connues. Elles correspondent justement à la combinaison des phases I-II de l'exemple précédent. On détermine ensuite pour chacun des haplotypes des enfants s'il est recombinant ou non.



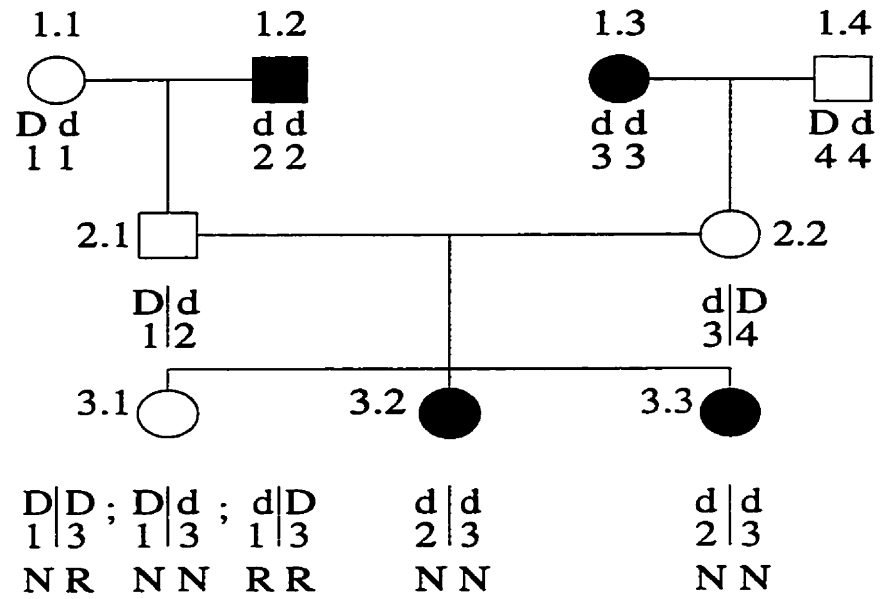


Figure 2.8: Phénotypes et génotypes d'une maladie récessive rare ainsi que d'un marqueur représenté par les allèles 1, 2, 3 et 4 avec des parents de phases connues

La fonction de vraisemblance de ce deuxième exemple est différente de celle obtenue du premier exemple à la figure 2.7, étant donné que les parents sont de phases connues. Cette fonction de vraisemblance est la suivante :

$$L(\theta) = K [ (1/3 (1 - \theta)\theta + 1/3 (1 - \theta)^2 + 1/3 \theta^2) (1 - \theta)^2 (1 - \theta)^2 ]$$

$$\Rightarrow L(\theta) = K (1/3) [ (1 - \theta + \theta^2) (1 - \theta)^4 ].$$

On avait posé comme hypothèses pour les deux exemples précédents que la pénétrance était complète et qu'il n'y avait pas de phénocopie. Ces hypothèses ont permis de simplifier le calcul des fonctions de vraisemblance. Le calcul de la fonction de vraisemblance est encore plus compliqué lorsque la pénétrance est incomplète et/ou il y a des phénocopies. On a alors besoin d'une forme générale pour exprimer la fonction de vraisemblance. C'est en utilisant l'algorithme d'Elston-Stewart qu'on pourra y arriver; ce sera le sujet abordé à la section suivante.

## 2.4.2 L'algorithme d'Elston-Stewart

Cet algorithme consiste en une méthode récursive qui permet de faire des calculs rapides et exacts de la vraisemblance de pedigree entre deux loci (Elston et Stewart, 1971). La vraisemblance d'un modèle statistique pour un ensemble de données est proportionnelle à la probabilité d'observer l'échantillon que l'on possède sous l'hypothèse d'un modèle en particulier.

### 2.4.2.1 Les trois composantes de la vraisemblance d'un pedigree

La vraisemblance d'un pedigree est développée en fonction de trois composantes :

- la distribution des différents génotypes dans la population ( $\Psi_s$ );
- la relation entre le phénotype et le génotype définissant la pénétrance ( $g_u(z_j)$ );
- le mode de transmission génétique ( $p_{stu}$ ).

La distribution des génotypes  $s$  et  $t$  des parents fondateurs ainsi que ceux des époux(es) des membres de la famille correspond à la distribution des différents génotypes dans la population. En effet, on utilise cette distribution lorsque les parents d'un individu n'ont pas été échantillonnés, ou bien qu'on ne connaît pas les génotypes de ses parents. La distribution des génotypes dans la population est déterminée sous équilibre d'Hardy-Weinberg (Ott, 1991). La probabilité qu'un individu ait le génotype  $s$ , étant donné la distribution des différents génotypes dans la population, est notée par  $\Psi_s$ . Par exemple, pour un locus à deux allèles,  $A$  et  $a$ , avec fréquence  $p$  et  $(1 - p)$  dans la population, la distribution génotypique sous équilibre d'Hardy-Weinberg est alors :  $\Psi_{AA} = p^2$ ,  $\Psi_{Aa} = 2p(1 - p)$  et  $\Psi_{aa} = (1 - p)^2$ . Pour un génotype bivarié, la distribution des génotypes en équilibre d'Hardy-Weinberg est simplement obtenue en multipliant la probabilité des génotypes à chacun des deux loci.

Le concept de la pénétrance établit le lien entre le génotype et le phénotype; c'est pour cette raison que la pénétrance a une importance centrale en analyse de linkage (Ott, 1991). La pénétrance est la probabilité qu'un individu exprime le phénotype  $x$  étant donné qu'il a le génotype  $g$ ; il s'agit de la probabilité conditionnelle  $P(x|g)$ . Pour une maladie, elle est la probabilité conditionnelle  $P(\text{statut du phénotype}|g)$ , c'est-à-dire la probabilité qu'un individu soit affecté ou non par la maladie étant donné qu'il a le génotype  $g$ .

Les pénétrances dans le cas d'une maladie récessive, pour les allèles  $A$  et  $a$ , sont les suivantes :

- $P(\text{affecté}|aa) = 1$ ;
- $P(\text{affecté}|Aa) = 0$ ;
- $P(\text{affecté}|AA) = 0$ .

Les pénétrances dans le cas d'une maladie dominante sont :

- $P(\text{affecté}|AA) = 1$ ;
- $P(\text{affecté}|Aa) = 1$ ;
- $P(\text{affecté}|aa) = 0$ .

Toutes ces pénétrances correspondent à des pénétrances complètes. Cependant, il peut arriver que la pénétrance soit incomplète. Pour une maladie récessive, on aurait alors que  $P(\text{affecté}|aa) < 1$  et pour une maladie dominante,  $P(\text{affecté}|AA) = P(\text{affecté}|Aa) < 1$ .

Pour certaines maladies, on observe la présence de phéncopies (Ott, 1991). Une phéncopie est représentée par une pénétrance non nulle qu'un individu exprime la maladie étant donné qu'il a un génotype sain. Par exemple, pour le cas d'une maladie

dominante, on aurait alors que  $P(\text{affecté}|aa) > 0$ . La phénocopie n'est pas le résultat d'une prédisposition génétique mais plutôt due à certains facteurs environnementaux.

Plusieurs maladies génétiques ne sont pas exprimées à la naissance (Ott, 1991). Elles se développent seulement plus tard dans la vie, c'est-à-dire que la pénétrance est dépendante de l'âge. Si un individu est affecté par une maladie, alors il y a deux types d'âge d'observation possibles. Si on sait à quel âge cet individu est devenu affecté par la maladie, on peut alors considérer l'âge du début comme faisant partie du phénotype. Cependant, parfois on ne connaît pas l'âge du début de la maladie de l'individu; donc, dans ce cas, l'âge actuel de cet individu (ou l'âge du dernier examen médical) devrait faire partie du phénotype. Ces deux situations requièrent deux définitions différentes de la pénétrance dépendante de l'âge.

Notons l'âge du début de la maladie par la variable  $X$  avec sa fonction de densité  $f(x)$  qui est connue. Sa fonction de répartition est notée par  $F(x)$ . Considérant un phénotype dichotomique, un individu affecté par une maladie quelconque est noté par  $Z=1$ , tandis que s'il est non affecté,  $Z=0$ .

La probabilité qu'un individu avec génotype  $g$  soit affecté à l'âge  $a'$  où il a été examiné, c'est-à-dire la probabilité qu'il a développé la maladie entre l'âge 0 et  $a'$ , s'exprime comme suit :

$$g_g(Z = 1|x = a') = \gamma_g F(a') \quad (2.1)$$

où

$\gamma_g$  est un paramètre de prédisposition du génotype  $g$  à la maladie, c'est-à-dire la probabilité qu'un individu avec génotype  $g$  développera éventuellement la maladie s'il vit suffisamment longtemps. Pour le cas d'une maladie dominante à pénétrance complète, on aurait alors que  $\gamma_{AA} = \gamma_{Aa} = 1$ . Par contre, si la pénétrance était incomplète, cela correspondrait à  $\gamma_{AA} = \gamma_{Aa} < 1$ .

Pour le cas d'un individu avec génotype  $g$  dont on connaît l'âge au moment où il est devenu affecté, la probabilité conjointe du phénotype et de l'âge du début de la maladie, noté  $a$ , étant donné que cet individu a été examiné à l'âge  $a'$ , est la suivante :

$$g_g(Z = 1, x = a|a') = \gamma_g f(a). \quad (2.2)$$

Concernant les individus non affectés, la pénétrance dépendante de l'âge se définit comme étant la probabilité de ne pas avoir encore développé la maladie à l'âge  $a'$  :

$$g_g(Z = 0|x = a') = 1 - \gamma_g F(a'). \quad (2.3)$$

Afin d'illustrer la notion de la pénétrance dépendante de l'âge, on va utiliser l'exemple de la schizophrénie. Considérons d'abord le cas d'individus qui ont un génotype susceptible, c'est-à-dire qu'ils sont porteurs du gène ou des gènes de la maladie. En se basant sur les données que l'on possède sur cette maladie, celle-ci se manifeste avec une fréquence de 31 % entre 0 et 19 ans, 31 % entre 20 et 24 ans, 17 % entre 25 et 29 ans, 11 % entre 30 et 34 ans et enfin 10 % après 34 ans. Ces pourcentages correspondent à la fonction de densité de la figure 2.9 pour la schizophrénie. Supposons que le paramètre de prédisposition  $\gamma_g$  vaut 0.90.

Pour la schizophrénie, la figure 2.9 illustre la pénétrance dépendante de l'âge pour des individus de génotype susceptible. Par exemple, si un individu est devenu affecté à l'âge de 20 ans et qu'il est maintenant âgé de 30 ans, la pénétrance qui lui sera attribuée correspond à l'équation 2.2 :

$$g_g(Z = 1, x = 20|30) = \gamma_g f(20) = 0.90 \times 0.31 = 0.279.$$

Par contre, pour un individu âgé de 30 ans dont on ne connaît pas l'âge au début de la maladie, sa pénétrance correspond à celle de l'équation 2.1 :

$$g_g(Z = 1|x = 30) = \gamma_g F(30) = 0.90 \times 0.90 = 0.81.$$

Pour un individu non affecté âgé de 32 ans, sa pénétrance correspond à l'équation 2.3 :

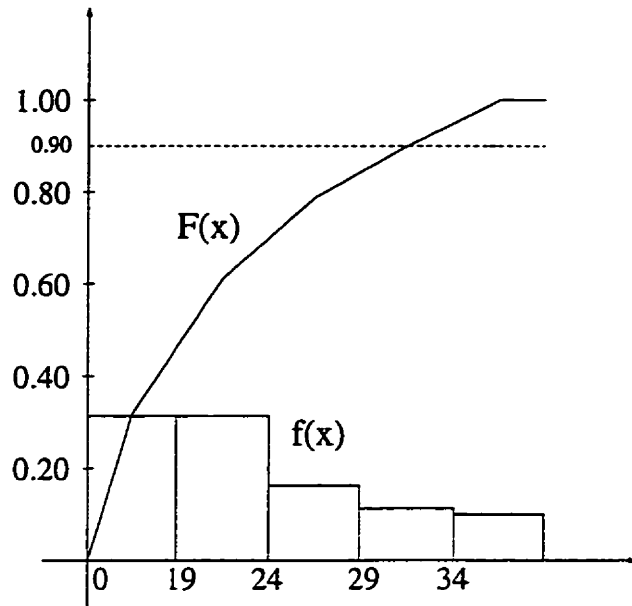


Figure 2.9: Courbes  $F(\text{âge})$  et  $f(\text{âge})$  pour des individus de génotype susceptible dans le cas de la schizophrénie

$$g_g(Z = 0|x = 32) = 1 - \gamma_g F(32) = 1 - 0.90 \times 0.90 = 0.19.$$

Pour la plupart des maladies, lorsque la pénétrance pour des génotypes susceptibles est dépendante de l'âge, la pénétrance pour des génotypes non susceptibles devrait aussi être dépendante de l'âge (Ott, 1991). Un individu affecté de génotype non susceptible correspond à une phénotypie, c'est-à-dire que cet individu exprime la maladie tout en étant non porteur du gène ou des gènes de la maladie.

Le calcul des pénétrances dépendantes de l'âge pour des individus de génotype non susceptible concernant la schizophrénie s'effectue en utilisant les équations 2.1, 2.2 et 2.3. Dans ce cas, la valeur de  $\gamma_g$  est très petite, car la probabilité d'exprimer un jour la schizophrénie est beaucoup plus petite que pour des individus de génotype susceptible.

Le mode de transmission génétique est défini par le paramètre  $p_{stu}$ . Il s'agit de la probabilité qu'un enfant ait le génotype  $u$  étant donné les génotypes  $s$  et  $t$  de ses parents. Pour un locus autosomal, étant composé des allèles  $A$  et  $a$ , le vecteur  $(p_{st1}, p_{st2}, p_{st3})$  des probabilités de chacun des trois génotypes ( $1=AA$ ,  $2=Aa$ ,  $3=aa$ ) conditionnelles aux

		2 <sup>e</sup> parent, t		
		1=AA	2=Aa	3=aa
1 <sup>er</sup> parent, s	1=AA	(1,0,0)	(1/2,1/2,0)	(0,1,0)
	2=Aa	(1/2,1/2,0)	(1/4,1/2,1/4)	(0,1/2,1/2)
	3=aa	(0,1,0)	(0,1/2,1/2)	(0,0,1)

Tableau 2.1: Matrice de transition génétique pour le cas d'un locus autosomal avec les allèles  $A$  et  $a$

génotypes  $s$  et  $t$  des parents, est un élément d'une matrice de transition génétique de dimension trois par trois. Dans ce cas, la matrice de transition génétique est représentée au tableau 2.1.

Chaque élément de cette matrice est obtenu en définissant la relation suivante :

$$(p_{st1}, p_{st2}, p_{st3}) = (\tau_s \tau_t, \tau_s(1 - \tau_t) + \tau_t(1 - \tau_s), (1 - \tau_s)(1 - \tau_t))$$

où

- $\tau_s$  = probabilité qu'un parent avec le génotype  $s$  transmette l'allèle  $A$  à son enfant;
- $(1 - \tau_s)$  = probabilité qu'un parent avec le génotype  $s$  transmette l'allèle  $a$  à son enfant.

Par exemple, pour le cas de deux parents ayant le même génotype  $Aa$ , le vecteur des probabilités conditionnelles est le suivant :

$$(p_{221}, p_{222}, p_{223}) = (1/2 \times 1/2, 1/2 \times 1/2 + 1/2 \times 1/2, 1/2 \times 1/2) = (1/4, 1/2, 1/4).$$

Considérons maintenant le mode de transmission pour le cas de deux loci autosomaux liés. Si le locus 1 possède  $a_1$  allèles et que le locus 2 possède  $a_2$  allèles, on pourra distinguer  $a_1 a_2$  gamètes et  $a_1^2 a_2^2$  génotypes différents, puisque les génotypes hétérozygotes seront distingués pour un locus donné; par exemple, le génotype  $Aa$  sera distingué de

$aA$ , en considérant la provenance des allèles. Soit  $\tau_{i_1 j_1 i_2 j_2 k_1 k_2}$  être la probabilité qu'un parent avec génotype  $i_1 j_1$  au locus 1 et  $i_2 j_2$  au locus 2 transmette l'haplotype  $k_1 k_2$  à son enfant (où  $i_1, j_1, k_1 = 1, 2, \dots, a_1$ ;  $i_2, j_2, k_2 = 1, 2, \dots, a_2$ ). Cette probabilité est appelée *probabilité de transmission* et dépend de la fraction de recombinaison  $\theta$ . Elle s'énonce comme suit :

$$\tau_{i_1 j_1 i_2 j_2 k_1 k_2} = (1/2)(1 - \theta)(\delta_{i_1 k_1} \delta_{i_2 k_2} + \delta_{j_1 k_1} \delta_{j_2 k_2}) + (1/2)\theta(\delta_{i_1 k_1} \delta_{j_2 k_2} + \delta_{j_1 k_1} \delta_{i_2 k_2}) \quad (2.4)$$

où  $\delta_{uv}$  est le delta de Kronecker :

$$\delta_{uv} = 1 \quad \text{si} \quad u = v,$$

$$\delta_{uv} = 0 \quad \text{si} \quad u \neq v.$$

C'est ainsi que les éléments d'une matrice de transition génétique pour deux loci autosomiaux s'expriment sous la forme suivante :

$$P_{s_1 s_2 s'_1 s'_2 t_1 t'_1 t'_2 u_1 u_2 u'_1 u'_2} = T_{s_1 s'_1 s_2 s'_2 u_1 u_2} \times T_{t_1 t'_1 t_2 t'_2 u'_1 u'_2} \quad (2.5)$$

où

- $s_1 s'_1$  = allèles au locus 1 du père;
- $s_2 s'_2$  = allèles au locus 2 du père;
- $u_1 u_2$  = allèles transmis par le père au locus 1 et 2, respectivement;
- $t_1 t'_1$  = allèles au locus 1 de la mère;
- $t_2 t'_2$  = allèles au locus 2 de la mère;
- $u'_1 u'_2$  = allèles transmis par la mère au locus 1 et 2, respectivement.



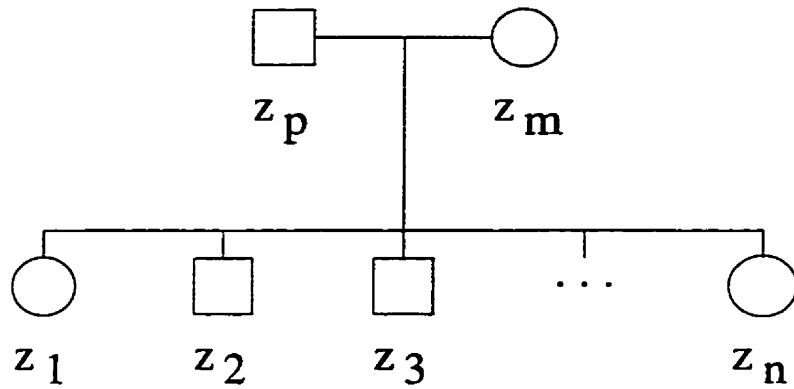


Figure 2.10: Représentation d'une famille nucléaire

#### 2.4.2.2 La vraisemblance d'une famille nucléaire

À la figure 2.10, on a la représentation d'une famille nucléaire avec les phénotypes des individus qui la composent. On suppose que, conditionnellement à leur propre génotype, les phénotypes des enfants sont indépendants les uns des autres. Ainsi, on suppose qu'il n'existe pas de facteurs environnementaux qui auraient comme conséquence de causer une corrélation entre les phénotypes.

La vraisemblance d'une famille nucléaire s'exprime comme suit :

$$\sum_{s=1}^k \Psi_s g_s(z_p) \sum_{t=1}^k \Psi_t g_t(z_m) \left[ \prod_{j=1}^n \sum_{u=1}^k p_{stu} g_u(z_j) \right]. \quad (2.6)$$

On va maintenant montrer comment la vraisemblance d'une famille nucléaire est établie en fonction des trois composantes que l'on vient d'expliquer à la section précédente.

La vraisemblance du  $j^{\text{e}}$  enfant conditionnelle à son génotype  $u$  est :

$$g_u(z_j). \quad (2.7)$$

Cette même vraisemblance conditionnelle aux génotypes  $s$  et  $t$  des parents, parmi  $k$  génotypes possibles, s'exprime comme suit :

$$\sum_{u=1}^k p_{stu} g_u(z_j). \quad (2.8)$$

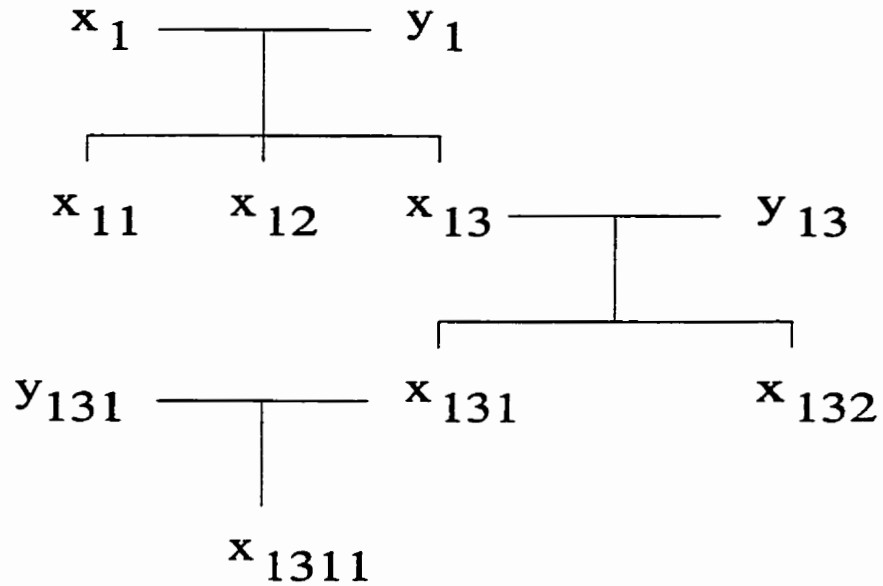


Figure 2.11: Représentation d'une famille étendue

La vraisemblance des enfants de la famille nucléaire étant donné les génotypes  $s$  et  $t$  des parents est :

$$\prod_{j=1}^n \sum_{u=1}^k p_{stu} g_u(z_j). \quad (2.9)$$

Sous l'hypothèse d'accouplement aléatoire, la vraisemblance des parents est alors :

$$\sum_{s=1}^k \Psi_s g_s(z_p) \sum_{t=1}^k \Psi_t g_t(z_m). \quad (2.10)$$

La vraisemblance d'une famille nucléaire est obtenue en multipliant la vraisemblance des parents par la vraisemblance des enfants, ce qui donne bien l'équation 2.6 au début de cette section.

### 2.4.2.3 La vraisemblance d'une famille étendue

À la figure 2.11, on a la représentation d'une famille étendue avec les phénotypes de chacun des individus de cette famille.

On note par  $x$  le phénotype d'un individu relié à quelqu'un dans une génération

précédente, tandis que par  $y$  le phénotype d'un individu non relié, c'est le cas d'un époux ou d'une épouse d'un membre de la famille. On utilise la numérotation  $1, 2, \dots, i_0$  pour différencier les familles. Pour le cas des parents fondateurs d'une famille donnée, on assigne un  $x$  et un  $y$  de façon arbitraire. Notons les mesures prises sur les parents fondateurs de la  $i_0^e$  famille par  $x_{i_0}$  et  $y_{i_0}$ . De plus, notons les mesures prises sur le  $i_1^e$  enfant de ses parents fondateurs par  $x_{i_0 i_1}$ , et sur son époux ou épouse par  $y_{i_0 i_1}$ ; de façon analogue, notons la mesure prise sur le  $i_2^e$  enfant du  $i_1^e$  enfant des  $i_0^e$  parents fondateurs par  $x_{i_0 i_1 i_2}$ , et sur son époux ou épouse par  $y_{i_0 i_1 i_2}$ ; et ainsi de suite.

En utilisant cette notation, on peut ainsi réécrire l'expression 2.6 de la vraisemblance d'une famille nucléaire de la façon suivante :

$$\sum_{s_0=1}^k \Psi_{s_0} g_{s_0}(x_{i_0}) \sum_{t_0=1}^k \Psi_{t_0} g_{t_0}(y_{i_0}) \left[ \prod_{i_1=1}^{n_1} \sum_{s_1=1}^k p_{s_0 t_0 s_1} g_{s_1}(x_{i_0 i_1}) \right]. \quad (2.11)$$

En ce qui concerne la vraisemblance de l'époux(se) du  $i_1^e$  enfant, c'est-à-dire  $y_{i_0 i_1}$ , elle est indépendante de l'expression 2.11 et cette vraisemblance s'écrit comme suit :

$$\sum_{t_1=1}^k \Psi_{t_1} g_{t_1}(y_{i_0 i_1}). \quad (2.12)$$

La vraisemblance d'une famille qui consiste en un couple fondateur, leurs enfants ainsi que les époux(ses) de leurs enfants est donc :

$$\sum_{s_0=1}^k \Psi_{s_0} g_{s_0}(x_{i_0}) \sum_{t_0=1}^k \Psi_{t_0} g_{t_0}(y_{i_0}) \left[ \prod_{i_1=1}^{n_1} \sum_{s_1=1}^k p_{s_0 t_0 s_1} g_{s_1}(x_{i_0 i_1}) \sum_{t_1=1}^k \Psi_{t_1} g_{t_1}(y_{i_0 i_1}) \right]. \quad (2.13)$$

Les phénotypes de la génération suivante, c'est-à-dire la deuxième, dépendront de la génération précédente seulement à travers la transmission des génotypes.

C'est ainsi que, conditionnellement aux génotypes  $s_1$  et  $t_1$ , la vraisemblance des enfants avec phénotype  $x_{i_0 i_1 i_2}$  est :

$$\prod_{i_2=1}^{n_2} \left[ \sum_{s_2=1}^k p_{s_1 t_1 s_2} g_{s_2}(x_{i_0 i_1 i_2}) \right]. \quad (2.14)$$

La vraisemblance d'une famille étendue sur plusieurs générations s'exprime donc par l'expression 2.13 suivie des expressions 2.14 et 2.12 de façon alternée.

Définissons l'opérateur suivant :

$$\Gamma_j = \sum_{s_j=1}^k p_{s_{j-1}t_{j-1}s_j} g_{s_j}(x_{i_0 i_1 \dots i_j}) \sum_{t_j=1}^k \Psi_{t_j} g_{t_j}(y_{i_0 i_1 \dots i_j}). \quad (2.15)$$

Cet opérateur  $\Gamma_j$  est fonction de  $s_{j-1}$  et  $t_{j-1}$  puisqu'il représente la probabilité d'observer les enfants et les conjoints de la génération  $j$  conditionnellement au fait que les parents soient de génotypes  $s_{j-1}$  et  $t_{j-1}$ , respectivement.

On obtient ainsi la vraisemblance d'une famille étendue :

$$\Gamma_0(\Gamma_1(\Gamma_2(\Gamma_3 \dots))) \quad (2.16)$$

en précisant que  $p_{s_{j-1}t_{j-1}s_j} = \Psi_{s_j}$ , quand  $j = 0$ . Cela correspond à la vraisemblance des parents fondateurs d'une famille étendue, c'est-à-dire :

$$\Gamma_0 = \sum_{s_0=1}^k \Psi_{s_0} g_{s_0}(x_{i_0}) \sum_{t_0=1}^k \Psi_{t_0} g_{t_0}(y_{i_0}). \quad (2.17)$$

Notons que l'expression 2.16 ne signifie pas que  $\Gamma_i$  multiplie  $\Gamma_j$ , il s'agit plutôt d'une séquence d'opérations.

La vraisemblance d'un ensemble de familles étendues est obtenue en multipliant la vraisemblance de chacune de ces familles étendues, en s'assurant que ces familles soient indépendantes les unes des autres.

À la section suivante, on montrera comment tenir compte des vraisemblances décrites dans cette section pour calculer le Lod-score d'un pedigree.

### 2.4.3 Méthode du Lod-score

Lorsqu'on fait une analyse de linkage entre deux loci, on cherche entre autres à obtenir une estimation de la fraction de recombinaison  $\theta$  (Ott, 1991). On utilise généralement la méthode du Lod-score pour tester l'hypothèse de linkage entre deux loci. Cette méthode fait intervenir le principe de l'estimation du maximum de la vraisemblance. L'estimateur du maximum de vraisemblance  $\hat{\theta}$  de  $\theta$  est défini comme étant la valeur de  $\theta$  qui maximise la vraisemblance.

En ce qui concerne une analyse de linkage entre deux loci, le test d'hypothèses qui s'applique est le suivant :

$$H_0 : \theta = 1/2 \text{ (absence de linkage)}$$

$$H_1 : \theta < 1/2 \text{ (présence de linkage)}.$$

Comme on l'a expliqué à la section 2.1, on utilise la borne supérieure  $1/2$  pour l'hypothèse alternative  $H_1$ , car cette valeur correspond à la fraction maximale de recombinaison que l'on peut obtenir.

Les chances en faveur de l'hypothèse alternative  $H_1$  par rapport à l'hypothèse nulle  $H_0$  sont exprimées par le rapport des vraisemblances suivant :

$$L^*(\theta) = \frac{L_{H_1}(\theta)}{L_{H_0}(\theta = 1/2)}. \quad (2.18)$$

Plutôt que de chercher à maximiser  $L^*(\theta)$ , il est davantage pratique d'utiliser le Lod-score  $Z(\theta)$ . Il s'exprime de la façon suivante :

$$Z(\theta) = \log_{10} \left[ \frac{L_{H_1}(\theta)}{L_{H_0}(\theta = 1/2)} \right]. \quad (2.19)$$

En ayant comme objectif d'obtenir  $\hat{\theta}$ , c'est donc  $Z(\theta)$  qui sera maximisé. La théorie des tests du rapport des vraisemblances (Casella et Berger, 1990) établit qu'asymptotiquement, et ce sous  $H_0$ ,

$$2 \times \log_e 10 \times Z(\hat{\theta}) \rightsquigarrow \chi_1^2. \quad (2.20)$$

Généralement, la valeur critique utilisée pour effectuer le test d'hypothèses, notée  $Z_0$ , est égale à 3 pour les loci autosomiaux, et à 2 pour les loci des chromosomes sexuels (Ott, 1991). Ainsi, en considérant la valeur critique  $Z_0 = 3$  pour les loci autosomiaux, on déclare qu'il y a une évidence significative de linkage entre deux loci lorsque  $Z(\hat{\theta}) \geq 3$ , c'est-à-dire lorsque la  $\chi_1^2 \geq 2 \times \log_e 10 \times 3 = 13.8$ . Le seuil de signification asymptotique  $\alpha$  associé à ce test du rapport de vraisemblance pour linkage est le suivant :

$$P_{H_0}(\chi_1^2 \geq 13.8) = 0.0001. \quad (2.21)$$

Le seuil de signification pour un test de linkage est beaucoup plus bas que les seuils de signification de 0.05 ou 0.01 habituellement utilisés pour des tests statistiques. Smith (1953) en explique la raison : lorsqu'on effectue un test d'hypothèses à un niveau de confiance de 95 %, cela signifie que l'on déclare faussement une présence de linkage dans 5 % des cas. Or, la probabilité a priori qu'une paire de loci soit située sur un même chromosome (présence de linkage) est également d'environ 5 % (Smith, 1959). Ainsi, un test d'hypothèses d'une puissance utopique de 100 % détectera autant de faux linkages que de réels, ce qui peut s'expliquer de façon statistique (Pagé, 1993). Soit,

- $P(\text{rejeter } H_0 \mid H_0) = 1 - (\text{niveau de confiance du test}) = 5 \%$ ;
- $P(\text{rejeter } H_0 \mid H_1) = \text{puissance du test} = 100 \%$ ;
- $P(H_1) \approx 5 \% \Rightarrow P(H_0) \approx 95 \%$ .

Par conséquent,

$$\begin{aligned} P(\text{rejeter } H_0) &= P(\text{rejeter } H_0 \mid H_0) P(H_0) + P(\text{rejeter } H_0 \mid H_1) P(H_1) \\ &= (0.05 \times 0.95) + (1 \times 0.05) \\ &\quad \text{“mauvais” rejet} \quad \text{“bon” rejet} \end{aligned}$$

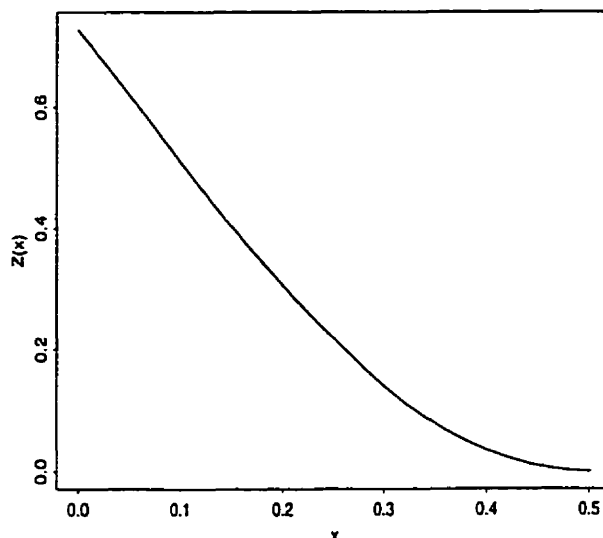


Figure 2.12: Courbe du Lod-score  $Z(\theta)$  pour le premier exemple de la section 2.4.1

La proportion de “mauvais” rejet est alors :

$$\frac{(0.05 \times 0.95)}{(0.05 \times 0.95) + (1 \times 0.05)} = \frac{0.0475}{0.0975} \approx 50\%.$$

À mesure que la puissance du test diminuera, la proportion de “mauvais” rejet augmentera davantage. Smith (1953) proposa donc de choisir un seuil de signification plus bas qu’à l’habitude afin d’élever la probabilité a posteriori d’avoir détecté un vrai linkage à 95 %.

On va maintenant appliquer cette méthode du Lod-score aux deux exemples de la section 2.4.1. Rappelons que la fonction de vraisemblance du premier exemple était la suivante :

$$L(\theta) = K(1/12)(1 - 5\theta + 12\theta^2 - 14\theta^3 + 7\theta^4)$$

$$\Rightarrow L^*(\theta) = \frac{K(1/12)(1 - 5\theta + 12\theta^2 - 14\theta^3 + 7\theta^4)}{K(1/12)(1 - 5(1/2) + 12(1/2)^2 - 14(1/2)^3 + 7(1/2)^4)}$$

$$\Rightarrow Z(\theta) = \log_{10} [(16/3)(1 - 5\theta + 12\theta^2 - 14\theta^3 + 7\theta^4)].$$

Le graphique de ce Lod-score  $Z(\theta)$  en fonction de la fraction de recombinaison  $\theta$  (à des

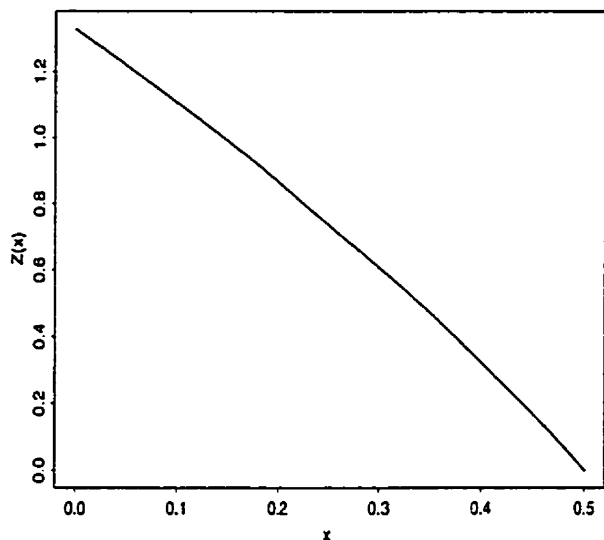


Figure 2.13: Courbe du Lod-score  $Z(\theta)$  pour le deuxième exemple de la section 2.4.1

intervalles de 0.02) est représenté à la figure 2.12. On remarque que  $Z(\theta)$  est maximisé à  $\hat{\theta} = 0$ . Ainsi, on obtient que  $Z(\hat{\theta}) = 0.73$ . Étant donné que  $Z(\hat{\theta}) < 3$ , on conclut alors qu'il n'y a pas d'évidence significative de linkage entre le locus de la maladie et celui du marqueur génétique pour ce premier exemple.

En ce qui concerne la fonction de vraisemblance du deuxième exemple de la section 2.4.1, rappelons qu'elle était la suivante :

$$L(\theta) = K(1/3)(1 - \theta + \theta^2)(1 - \theta)^4$$

$$\Rightarrow L^*(\theta) = \frac{K(1/3)(1 - \theta + \theta^2)(1 - \theta)^4}{K(1/3)(1 - (1/2) + (1/2)^2)(1 - (1/2))^4}$$

$$\Rightarrow Z(\theta) = \log_{10} [(64/3)(1 - \theta + \theta^2)(1 - \theta)^4].$$

Le graphique de ce Lod-score  $Z(\theta)$  en fonction de la fraction de recombinaison  $\theta$  (à des intervalles de 0.02) est représenté à la figure 2.13. On constate que  $\hat{\theta} = 0$  maximise  $Z(\theta)$ . Ainsi, on obtient que  $Z(\hat{\theta}) = 1.33$ . Étant donné que  $Z(\hat{\theta}) < 3$ , on conclut alors qu'il n'y a pas d'évidence significative de linkage entre le locus de la maladie et celui du marqueur génétique pour ce deuxième exemple.



#### 2.4.4 Hétérogénéité due à un mélange de familles indissociables

Un phénotype, en particulier une maladie, est génétiquement hétérogène lorsqu'il y a une étiologie génétiquement différente chez des individus différents. Par exemple, chez certains individus une maladie est due à un gène unique alors que chez d'autres la maladie est causée par un facteur environnemental (Ott, 1991). Dans le contexte d'une analyse de linkage, on distingue deux types d'hétérogénéité : l'hétérogénéité allélique et l'hétérogénéité non allélique ou génétique. Avec l'hétérogénéité allélique, les individus diffèrent les uns des autres en ayant des allèles différents au même locus responsable de la maladie. Cependant, avec l'hétérogénéité génétique, la maladie est causée par des loci différents.

Un exemple d'hétérogénéité allélique est la fibrose kystique : un grand nombre d'allèles mutés différents ont été identifiés chez des individus affectés, tous au même locus sur le chromosome 7. Un exemple d'hétérogénéité génétique est la maladie appelée "Charcot-Marie-Tooth disease" qui peut être causée par au moins un locus sur le chromosome 1 ou un locus sur le chromosome 17. L'hétérogénéité allélique n'affecte en rien les résultats d'une analyse de linkage. L'hétérogénéité génétique (existence de plus d'un gène) aura cependant un impact important. En effet, lorsque le marqueur génétique relié au premier gène sera étudié, il y aura alors deux types de familles : les familles à l'intérieur desquelles il y a linkage entre le locus de la maladie et celui du marqueur ( $\theta_1 < 1/2$ ) et les familles pour lesquelles la maladie n'est pas liée à ce gène et donc à l'intérieur desquelles  $\theta_2 = 1/2$ . Ces deux types de familles sont souvent mélangés de telle façon qu'il est impossible de déterminer a priori quelles sont les familles appartenant au premier ou au second type. On dit alors que les familles forment un *mélange*. La contre-hypothèse qui est alors considérée s'exprime comme suit :

$H_1$  : Une proportion de familles ont une fraction de recombinaison  $\theta$  inférieure à  $1/2$ , tandis que pour d'autres, elle est égale à  $1/2$ .

De façon plus précise, cette hypothèse représente une situation d'hétérogénéité telle que, dans certaines familles, une maladie est causée par un gène localisé dans le voisinage du locus d'un marqueur étudié tandis que, chez d'autres familles, la même maladie est due à un gène malade localisé ailleurs, ou bien la maladie n'est pas due à un gène. Dans cette situation, l'évidence de linkage entre le locus de la maladie et celui du marqueur existera uniquement pour le premier type de famille.

Considérant le modèle d'hétérogénéité due à un mélange de familles indissociables, on pose alors que :

- $\alpha$  : la proportion des familles de type 1;
- $(1 - \alpha)$  : la proportion des familles de type 2;
- $\theta_1$  : la fraction de recombinaison des familles de type 1;
- $\theta_2 = 1/2$  : la fraction de recombinaison des familles de type 2.

Si on ne connaît pas a priori une classification plausible de l'ensemble des familles, alors la vraisemblance bivariée pour la  $i^e$  famille s'exprime de la façon suivante :

$$L_i(\alpha, \theta_1) = \alpha L_i(\theta_1) + (1 - \alpha) L_i(1/2) \quad (2.22)$$

où  $L_i(\theta_1)$  est la vraisemblance évaluée à  $\theta = \theta_1$ . Afin de construire la statistique du rapport de vraisemblance, on peut diviser, sans perte de généralité, l'équation 2.22 par  $L_i(1/2)$ . Alors, on obtient l'expression suivante :

$$L_i^*(\alpha, \theta_1) = \alpha L_i^*(\theta_1) + (1 - \alpha). \quad (2.23)$$

En considérant l'ensemble des familles, la log-vraisemblance ajustée de façon globale s'exprime comme suit :

$$\log_{10} L(\alpha, \theta_1) = \sum_i \log_{10} [L_i^*(\alpha, \theta_1)] \quad (2.24)$$

qui est une fonction de deux paramètres, c'est-à-dire  $\alpha$  et  $\theta_1$ . Le test d'hypothèses pour l'hétérogénéité génétique est basé sur ces deux derniers paramètres. Ce test est le suivant :

$$H_0 : \alpha = 1 \text{ (les familles partagent un } \theta \text{ commun)}$$

$$H_1 : 0 < \alpha < 1, \theta_1 < 1/2 \text{ (hétérogénéité génétique entre les familles).}$$

Sous  $H_0$ , on doit donc calculer l'estimation dite restreinte de  $\theta$ , notée  $\hat{\theta}_r$ , en supposant l'homogénéité. Tandis que sous  $H_1$ ,  $\hat{\alpha}$  et  $\hat{\theta}_1$  sont les deux valeurs qui, simultanément, maximisent la log-vraisemblance bivariée globale présentée à l'équation 2.24. L'équation 2.25 représente la statistique utilisée pour tester l'hypothèse d'homogénéité entre les familles.

$$\chi^2 = 2 \left[ \log_e L(\hat{\alpha}, \hat{\theta}_1) - \log_e L(1, \hat{\theta}_r) \right]. \quad (2.25)$$

Sous  $H_0$ , cette statistique suit asymptotiquement une  $\chi_1^2$  avec une probabilité de 1/2 et vaut 0 avec une probabilité de 1/2 (test unilatéral).

Lorsqu'on a de bonnes estimations des paramètres  $\alpha$  et  $\theta_1$ , il est alors possible de calculer, conditionnellement aux observations que l'on possède, la probabilité que la  $i^*$  famille appartienne au type 1, c'est-à-dire aux familles pour lesquelles le locus de la maladie est lié à celui d'un marqueur. Cette probabilité conditionnelle estimée est la suivante :

$$w_i(\hat{\alpha}, \hat{\theta}_1) = \frac{\hat{\alpha} L_i^*(\theta_1)}{\hat{\alpha} L_i^*(\theta_1) + 1 - \hat{\alpha}}. \quad (2.26)$$

Lorsqu'on connaît les valeurs des paramètres  $\alpha$  et  $\theta_1$ , il semblerait que  $w_i > 1/2$  représenterait la règle de classification optimale pour identifier les familles de type lié (Goldstein et Dillon, 1978). Cependant, en utilisant les estimations des paramètres  $\alpha$  et  $\theta_1$ , pour des petites tailles d'échantillon, la règle de classification 2.27 est généralement plus exacte :

$$w_i(\hat{\alpha}, \hat{\theta}_1) > \hat{\alpha} \Rightarrow \text{la } i^* \text{ famille appartient au type lié.} \quad (2.27)$$

L'inégalité 2.27 est équivalente à  $Z_i(\hat{\theta}_1) > 0$ , où  $Z_i$  est le Lod-score de la  $i^e$  famille évalué en l'estimation de  $\theta_1$  et obtenu à partir de l'ensemble des familles. Cette seconde règle de classification est intuitivement compréhensible; lorsqu'une famille est non informative pour une analyse de linkage, elle obtient un  $w_i$  égal à  $\alpha$ , soit la probabilité théorique (non conditionnelle). Une évidence pour le linkage, c'est-à-dire qu'une famille appartienne au type lié, devrait donc hausser la valeur de  $w_i$  au-dessus de celle de  $\alpha$ , d'où la règle  $w_i > \hat{\alpha}$ .

### 2.4.5 La notion de boucle dans un pedigree

Une boucle est présente dans un pedigree lorsqu'il est possible de partir d'un individu quelconque dans un dessin d'un pedigree, de quitter cet individu en suivant une séquence de lignes reliées et de retourner à l'individu original sans revenir sur ses pas (Terwilliger et Ott, 1994). Si une boucle existe dans un pedigree, celle-ci va piéger l'algorithme d'Elston-Stewart dans une boucle infinie.

Pour résoudre ce problème, on doit choisir un individu dans la boucle qui est à la fois un enfant et un parent, dont cet individu sera "doublé". Ainsi, cela "brise" la boucle, car l'algorithme d'Elston-Stewart considère ces deux individus doublés avec le même génotype (incluant la phase), mais tout en demeurant des individus séparés. On doit toujours briser une boucle à l'individu ayant le génotype le moins ambigu, incluant la phase. Cela permet de réduire de beaucoup le temps de calcul de la vraisemblance du pedigree à l'aide de l'algorithme.

On distingue deux types de boucles : les boucles de consanguinité et les boucles de mariage.

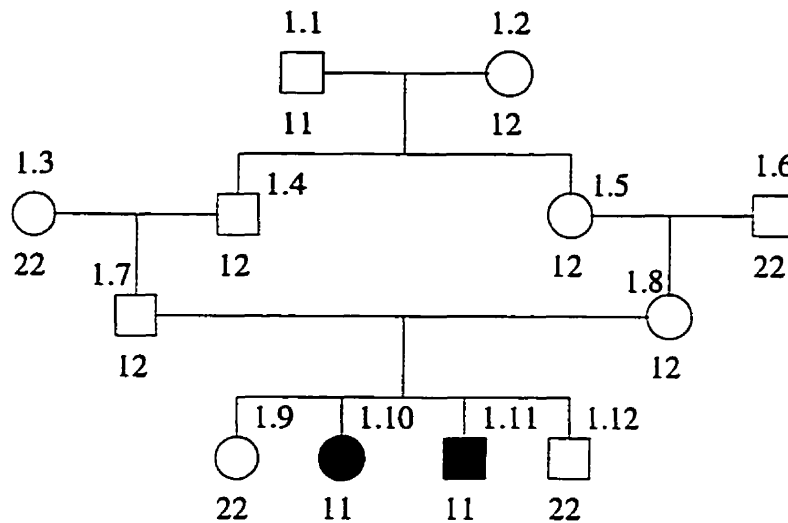


Figure 2.14: Pedigree contenant une boucle de consanguinité

#### 2.4.5.1 Boucle de consanguinité

Dans une boucle de consanguinité, on retrouve au moins deux partenaires ayant un sang commun.

Considérons le pedigree de la figure 2.14, dont la maladie est récessive rare et le marqueur génétique possède trois allèles. Étant donné que la fréquence de l'allèle malade est très petite, il est plus vraisemblable que cet allèle soit présent chez un seul des deux parents fondateurs. Ainsi, cela permet de dire que les individus 1.4 et 1.5 seraient les plus vraisemblables porteurs de l'allèle malade. Dans ce cas, on peut briser la boucle à l'individu 1.4, 1.5, 1.7 ou 1.8. On se rappelle qu'il faut toujours briser la boucle à l'individu possédant le génotype le moins ambigu. Les individus 1.7 et 1.8 doivent être hétérozygotes au locus de la maladie; alors on sait que leurs génotypes sont plus exacts que ceux des individus 1.4 et 1.5, car l'allèle malade ne provient pas spécifiquement d'un de ces deux individus. Par conséquent, on peut briser la boucle à l'individu 1.7 ou 1.8.

### 2.4.5.2 Boucle de mariage

Dans une boucle de mariage, il n'y a pas d'accouplement entre des individus ayant un sang commun. Une boucle de mariage souvent rencontrée correspond à celle où deux frères sont mariés à deux soeurs.

Considérons justement cette situation par le pedigree de la figure 2.15. Il s'agit de la même maladie récessive rare et du même marqueur génétique ayant trois allèles qu'à l'exemple utilisé à la section précédente. Ces mariages créent une boucle dans le pedigree, car on peut partir de l'individu 1.5 et revenir à lui par le chemin des individus 1.5 - 1.8 - 1.10 - 1.7 - 1.6 - 1.9 - 1.5. Cette boucle doit être brisée avant de faire une analyse de linkage de ce pedigree. On peut briser la boucle à l'un des individus 1.5, 1.6, 1.7 ou 1.8. Dans cet exemple, on sait que chacun des individus 1.5, 1.6, 1.7 et 1.8 est porteur de l'allèle malade mais, parmi eux, ceux qui sont hétérozygotes au locus du marqueur portent en quelque sorte moins d'information concernant la phase. Évidemment, les individus homozygotes (1.6 et 1.8) sont non informatifs pour une analyse de linkage; on sait que chez l'individu 1.8 les allèles malade et normal sont chacun dans un haplotype avec l'allèle 3 du marqueur. Ainsi, il peut être légèrement préférable de briser la boucle, soit à l'individu 1.6, soit à l'individu 1.8 parce qu'ils sont homozygotes au locus du marqueur, alors que les individus 1.5 et 1.7 sont hétérozygotes et de phases inconnues.

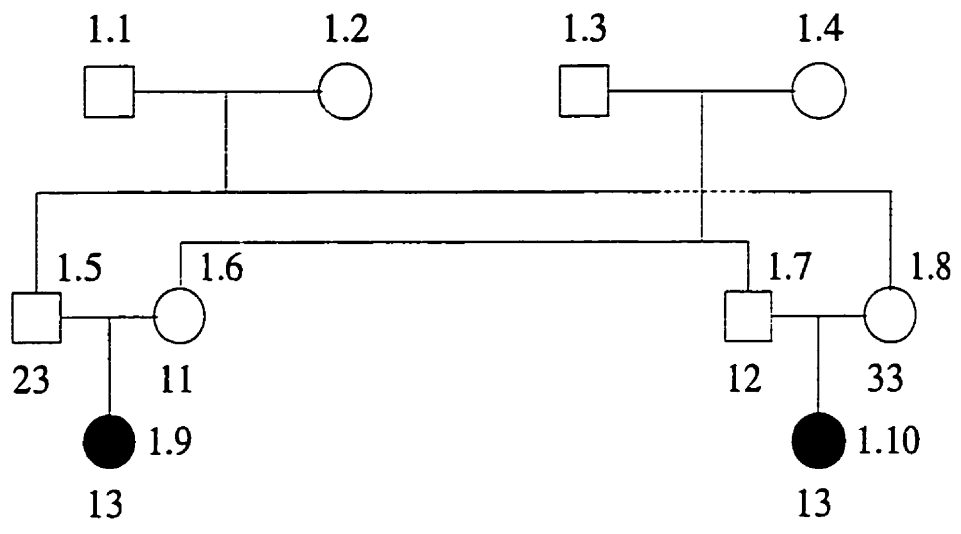


Figure 2.15: Pedigree contenant une boucle de mariage

# Chapitre 3

## La schizophrénie est-elle une maladie génétique?

La schizophrénie est un trouble psychotique, c'est-à-dire qu'il s'agit d'un trouble mental caractérisé par une altération profonde de la personnalité et des fonctions intellectuelles, et le fait que le sujet n'en reconnaît pas le caractère morbide (Lalonde et al., 1988). Elle comporte toujours, à un moment ou à un autre de la maladie, des idées délirantes, des hallucinations ou certaines perturbations caractéristiques de l'affectivité ou du cours de la pensée (DSM-III-R, 1987). La fréquence de la schizophrénie est apparemment identique chez les hommes et les femmes. Normalement, la période de risque pour le développement de la schizophrénie est à partir de l'âge de 15 ans jusqu'à 45 ans (Gottesman et Bertelsen, 1989). En général, le risque de la schizophrénie pour la population est de 1 %.

Les études familiales, les études de jumeaux et les études d'adoption, qui se succèdent depuis le début du siècle, ont fourni des hypothèses jusqu'ici les plus probantes quant à la nature de la schizophrénie (Lalonde et al., 1988). Ces études sont habituellement axées sur la question fondamentale du départage entre l'inné et l'acquis, c'est-à-dire entre le biologique et l'environnement.



### 3.1 Les études familiales

En ce qui concerne les études familiales, si des facteurs génétiques transmissibles contribuent au développement de la schizophrénie, ce trouble devrait s'agréger dans les familles comportant des individus affectés à un plus haut risque que celui de 1 % pour la population en général (Gottesman, 1991). De plus, si ces facteurs génétiques sont présents, le risque de morbidité devrait s'accroître avec le degré de parenté. Ainsi, l'agrégation familiale de la schizophrénie est une condition essentielle pour une explication génétique de cette maladie (Gottesman et Shields, 1972).

En regroupant plus de vingt études familiales de l'Europe de l'Ouest réalisées entre 1920 et 1987, on a ainsi obtenu les risques observés de schizophrénie, présentés au tableau 3.1, parmi les membres d'une famille comportant des schizophrènes (Gottesman et McGue, 1989). On y observe que le risque est de 7.8 % pour un lien de parenté de premier degré, c'est-à-dire chez les pères et mères de schizophrènes ainsi que chez leurs frères et soeurs, que le risque est de 2.7 % pour un lien de parenté de deuxième degré, c'est-à-dire chez les grands-parents, les petits-enfants, les oncles, les tantes, les neveux, les nièces de schizophrènes et que le risque est de 1.6 % pour un lien de parenté de troisième degré, c'est-à-dire chez les arrière-grands-parents, les cousins(es) de schizophrènes. Ainsi, le risque de schizophrénie semble bien s'accroître avec le degré de parenté à des sujets schizophrènes. Ces études familiales suggèrent donc qu'il y a une agrégation familiale

Lien de parenté	Risque observé (%)
Enfants de deux schizophrènes	36.6
Premier degré	7.8
Deuxième degré	2.7
Troisième degré	1.6

Tableau 3.1: Risques observés de schizophrénie parmi les membres d'une famille comportant des schizophrènes

de la schizophrénie. D'autre part, un autre résultat intéressant de ces études est celui du risque observé de 36.6 % pour les enfants de deux parents schizophrènes.

### 3.2 Les études de jumeaux

Pour les études de jumeaux, on utilise la notion de concordance (Khoury et al., 1993). Celle-ci se définit comme étant la probabilité conditionnelle que le deuxième jumeau soit affecté étant donné que le premier jumeau est affecté. L'évaluation de la concordance schizophrénique entre des jumeaux monozygotes (identiques) et des jumeaux dizygotes (non identiques) vise à distinguer l'apport génétique proprement dit de l'influence de l'environnement sociofamilial (Lalonde et al., 1988). La plus forte évidence de l'implication de facteurs génétiques dans l'étiologie de la schizophrénie vient des études de jumeaux. Si un facteur génétique est présent, la concordance devrait en effet être plus élevée chez les jumeaux monozygotes, qui partagent 100 % de leurs gènes, que chez les jumeaux dizygotes, qui partagent en moyenne 50 % de leurs gènes tout comme pour des frères ou soeurs.

Les meilleures études de jumeaux indiquent un taux de concordance moyen de 47 % chez les jumeaux monozygotes et de 15 % chez les jumeaux dizygotes (Lalonde et al., 1988). Il s'agit là d'une différence hautement significative qui appuie fortement l'hypothèse de facteurs génétiques. Le taux de concordance pour les jumeaux monozygotes nous fournit une estimation directe de la pénétrance de la schizophrénie pour un environnement donné (Lander et Schork, 1994).

Une étude de paires de jumeaux discordantes pour la maladie, réalisée par Gottesman et Bertelsen (1989), a démontré, entre autres choses, que la probabilité de développer la schizophrénie, en corrigeant pour l'âge, chez les enfants de jumeaux atteints, était de 16.8 %, alors que la probabilité correspondante chez les enfants de jumeaux non schizophrènes était de 17.4 %. En d'autres termes, la probabilité de développer la schizophrénie chez les enfants du jumeau atteint et celui non atteint est pratiquement

identique. Il s'agit d'un résultat qui vient appuyer nettement les hypothèses génétiques. Par contre, le taux élevé de discordance semble indiquer que les facteurs génétiques ne sont pas suffisants pour expliquer toute la schizophrénie.

### **3.3 Les études d'adoption**

Les études d'adoption sont une méthodologie apte à départager les apports de l'hérédité et de l'environnement dans l'étiologie de la schizophrénie (Lalonde et al., 1988). La plus importante des études d'adoption, concernant des enfants adoptés provenant de mères schizophrènes ou saines, a été réalisée en Finlande sous la direction de M. Tienari (1991) (Kendler et Scott, 1993). Les 361 familles retenues, qui contenaient un enfant adopté provenant d'une mère schizophrène ou saine, ont été contactées pour réaliser cette étude d'adoption. Les résultats étaient disponibles pour 144 enfants de mères schizophrènes et 178 enfants de mères saines. La fréquence de la schizophrénie était significativement plus grande chez les enfants de mères schizophrènes ( $7/144 = 4.9\%$ ) que chez les enfants de mères saines ( $2/178 = 1.1\%$ ), le seuil observé correspondant était de 0.02 pour un test unilatéral. Cela indique qu'un facteur génétique peut être inféré, puisque le rôle des facteurs environnementaux est partiellement neutralisé par le processus d'adoption (Khoury et al., 1993).

Ce que l'on retient de ces trois types d'études, c'est que les études familiales suggèrent qu'il y a une forte agrégation familiale de la schizophrénie (Kendler et Scott, 1993). De plus, les études de jumeaux et d'adoption suggèrent que les facteurs génétiques ont un rôle majeur dans cette agrégation familiale.

# Chapitre 4

## Complexité de l'analyse de linkage pour la schizophrénie

Si la schizophrénie est fortement influencée par des facteurs génétiques, les prochaines questions qui nous viennent à l'esprit sont : Combien de gènes sont-ils impliqués? Où se situe(nt) ce ou ces gène(s) dans le génome humain?

Certaines maladies humaines sont causées par des gènes uniques, c'est-à-dire que leur mode de transmission correspond à une transmission mendélienne classique avec un seul gène (Kendler et Scott, 1993). Par exemple, c'est le cas de la maladie d'Huntington, de la dystrophie musculaire et de la fibrose kystique.

### 4.1 Indices de complexité

Lorsqu'on fait une analyse de linkage pour la schizophrénie, il faut tenir compte de plusieurs problèmes de base existants. Ces problèmes rendent ainsi cette analyse de linkage plus complexe. Cette complexité peut s'expliquer par au moins cinq problèmes de base.

Premièrement, la plupart des troubles mendéliens sont à pénétrance complète (Kendler et Scott, 1993). Pour la schizophrénie, la rareté de familles avec transmission mendélienne apparente, un taux de concordance pour les jumeaux monozygotes bien inférieur

à 100 % et l'évidence que les enfants d'un jumeau monozygote non affecté, son jumeau étant affecté, soient substantiellement à risques accrus pour la schizophrénie comparativement à la population en général, suggèrent que ce trouble est à pénétrance incomplète. C'est donc dire qu'un individu peut avoir une prédisposition génétique élevée à la schizophrénie sans jamais la manifester.

Deuxièmement, dans la plupart des conditions mendéliennes, pour tous les individus qui expriment des symptômes typiques de la maladie, ces symptômes sont dus aux effets directs du gène de la maladie (Kendler et Scott, 1993). Cependant, pour la schizophrénie, cela n'est pas nécessairement vrai. Certains cas existent dont les symptômes s'apparentent à ceux de la schizophrénie qui ont été causés par des conditions métaboliques ou neurologiques ou par l'abus de drogues. Ces cas sont appelés des phénocopies.

Troisièmement, la plupart des conditions mendéliennes sont rares et étiologiquement homogènes, c'est-à-dire que tous les cas émanent d'anormalités dans le même gène (Kendler et Scott, 1993). Différentes mutations dans le même gène peuvent causer des variantes d'un trouble donné, mais cela n'affecte pas l'analyse de linkage parce que de pareilles variantes alléliques se produisent au même locus. Ce phénomène décrit l'hétérogénéité allélique. En ce qui concerne la schizophrénie, on envisage plutôt la présence d'hétérogénéité génétique. Cela signifie que différentes mutations dans différents gènes spécifiques influencent probablement le risque de différentes formes de la schizophrénie qui sont présentement cliniquement indistinctes bien qu'étiologiquement distinctes.

Quatrièmement, certains troubles peuvent requérir la présence simultanée de différentes mutations dans différents gènes non spécifiques, constituant ainsi des interactions entre différents gènes (Lander et Schork, 1994). On parle alors de transmission polygénique. Cette transmission polygénique est facilement démontrable en faisant des croisements avec des animaux, mais il est plus difficile de la démontrer directement chez les humains. Récemment, il a été démontré, chez l'humain, qu'une forme de la maladie appelée *rétinite pigmentaire* est due à une transmission digénique, requérant la présence

de mutations hétérozygotes dans les gènes RDS et ROM1. Le mode de transmission polygénique est un mode possible pouvant expliquer la schizophrénie.

Cinquièmement, pour la plupart des troubles mendéliens, il y a une discontinuité évidente entre les individus affectés et les individus non affectés (Kendler et Scott, 1993). Ce n'est pas aussi simple pour la schizophrénie. L'établissement de son diagnostic constitue l'une des difficultés majeures de la recherche psychiatrique. Cette difficulté est reliée à l'absence de critères diagnostiques objectifs tels que lésion observable, marqueur biologique ou même dysfonction psychophysiologique spécifique. Le présent mémoire traitera directement ce problème de diagnostic et par le fait même celui de la définition du phénotype en proposant d'utiliser une dimension symptomatique plutôt qu'un diagnostic pour définir le phénotype (section 4.4).

En se basant sur les observations précédentes, tous les chercheurs s'entendent pour dire que la schizophrénie diffère du trouble mendélien classique (Lander et Schork, 1994). La schizophrénie est un trouble complexe, à savoir un trouble se référant à un phénotype qui ne montre pas une transmission mendélienne classique récessive ou dominante due à un seul gène.

## **4.2 Analyse de linkage avec un trait complexe**

À l'origine, l'analyse de linkage appliquée à l'être humain fut développée pour les troubles mendéliens simples (Kendler et Scott, 1993). Comme on l'a expliqué précédemment, la schizophrénie ne fait pas partie de ces troubles simples. L'analyse de linkage est à la fois qualitativement et quantitativement plus difficile pour la schizophrénie que pour des troubles mendéliens simples, qualitativement parce que les méthodes statistiques développées pour des troubles mendéliens classiques doivent être modifiées avant d'être appliquées à la schizophrénie et quantitativement parce que les tailles d'échantillons nécessaires sont vraisemblablement plus grandes que celles pour des troubles mendéliens classiques.

L'incapacité de détecter une évidence définitive de linkage dans les études réalisées jusqu'à maintenant a forcé les chercheurs à réévaluer leurs hypothèses et à planifier des études plus puissantes en considérant des hypothèses plus complexes et réalistes au sujet du mode de transmission de la schizophrénie (Cloninger, 1994). L'hétérogénéité génétique fut proposée comme étant une explication plausible concernant la transmission non mendélienne apparente des troubles psychiatriques.

Récemment, en catégorisant l'âge du début de la maladie d'Alzheimer, étant une maladie complexe, ainsi que son mode de transmission, cette procédure a permis de mettre en évidence l'hétérogénéité génétique de cette maladie (Cloninger, 1994). Plus de 50 % des maladies d'Alzheimer hâtives, dont le mode de transmission est autosomal-dominant, sont liées à un gène défectueux sur le chromosome 14. Tandis que la forme familiale et sporadique la plus commune des maladies d'Alzheimer tardives a été associée à un gène défectueux situé sur le chromosome 19. Malheureusement, l'âge du début de la maladie n'est pas fortement familial pour la schizophrénie, à savoir que les familles de schizophrènes ne se différencient pas quant à l'âge du début de la maladie.

### 4.3 Modes de transmission complexes suggérés

Jusqu'à maintenant, deux grandes classes de modèles génétiques ont été retenues concernant le mode de transmission de la schizophrénie grâce aux données que l'on possède sur cette maladie (Cloninger, 1994). La première classe consiste en un modèle monogénique et postule une transmission mendélienne par un gène unique et suffisant de type récessif ou dominant. La deuxième classe comprend les modèles d'hétérogénéité génétique qui décrivent deux ou plusieurs gènes agissant de façon indépendante, ou encore en interaction pour causer la maladie. De plus, les facteurs environnementaux seraient un élément à ajouter à chacun de ces modèles.

Certains suggèrent que les analyses de linkage à venir devraient donc tenir compte des effets possibles des interactions multiples entre deux ou plusieurs gènes et des facteurs

environnementaux, c'est-à-dire considérer simultanément les effets de l'épistasie et de l'interaction gène-environnement (Cloninger, 1994).

Les méthodes d'analyses non paramétriques pourraient être utiles pour la détection de linkage. Pour faire une analyse de linkage non paramétrique, on n'a pas besoin de spécifier un mode de transmission génétique. Il s'agit alors d'une méthode plus robuste qu'une analyse de linkage paramétrique quand le mode de transmission est incertain. Cependant, les méthodes non paramétriques sont moins puissantes lorsque le mode de transmission est connu. Ce mémoire s'attardera toutefois uniquement aux analyses statistiques paramétriques.

## 4.4 Nouvelle définition du phénotype

La possibilité d'interactions multiples dans l'étiologie de la schizophrénie suggère que ce serait aussi utile de considérer l'analyse de la schizophrénie comme étant un phénotype multidimensionnel (Cloninger, 1994). Celui-ci serait constitué de dimensions symptomatiques multiples qui interagiraient pour créer un spectre variable de maladies. Il semblerait que ces dimensions ou composantes symptomatiques auraient une base génétique plus simple que la schizophrénie considérée comme un tout hétérogène.

De tels troubles complexes multidimensionnels sont aussi suggérés par des observations faites au sujet de la comorbidité à l'intérieur des individus et à l'intérieur des familles (Cloninger, 1994). Un exemple de comorbidité serait que, chez un schizophrène donné, il y aurait la présence de symptômes psychotiques propres à la schizophrénie et de symptômes dépressifs n'appartenant pas à la schizophrénie mais plutôt aux maladies affectives.

On s'attend à ce que l'hétérogénéité génétique soit à la fois importante à l'intérieur des familles et entre les familles lorsque les multiples facteurs environnementaux et génétiques interagissent pour produire un spectre variable de maladies (Cloninger,



1994). Lorsqu'il y a présence de ces interactions multifactorielles durant le développement d'un individu, différentes combinaisons de facteurs étiologiques produisent une plus grande hétérogénéité symptomatique et de la comorbidité à l'intérieur des individus et à l'intérieur des familles. Par exemple, des études de jumeaux et d'adoption ont permis de montrer récemment que les troubles de la personnalité peuvent être décomposés en sept composantes qui sont génétiquement indépendantes les unes des autres. Conséquemment, ça devrait être utile d'associer la variabilité dans les gènes candidats, c'est-à-dire les gènes susceptibles de causer la schizophrénie, à de telles composantes individuelles plutôt qu'à la schizophrénie considérée comme un ensemble de symptômes hétérogènes.

Il deviendra alors intéressant de s'attarder davantage aux composantes symptomatiques de la schizophrénie, faisant partie des phénotypes multidimensionnels de cette maladie, qui sont par définition plus simplement héritées et analytiquement traitables (Cloninger, 1994). Par conséquent, ces composantes permettront de produire vraisemblablement davantage de résultats qui pourront être répétés.

Le chapitre suivant décrit l'approche utilisée dans ce mémoire pour définir le phénotype à partir des composantes symptomatiques de la schizophrénie.

# Chapitre 5

## Utilisation d'une dimension symptomatique comme nouvelle définition du phénotype

### 5.1 Description des symptômes de la schizophrénie

Les composantes symptomatiques dont on a discuté au chapitre précédent correspondent aux dimensions psychopathologiques qui caractérisent la schizophrénie. Ces dimensions sont au nombre de huit (Andreasen, 1987). Il s'agit de l'alogie, de l'anhédonie, de l'apathie, du comportement bizarre, des hallucinations, des idées délirantes, de la pauvreté affective et du trouble de la pensée formelle.

L'*alogie* reflète une pensée et une capacité cognitive appauvries. Les processus de pensée semblent vides, emphatiques ou lents. Les deux manifestations majeures sont la pauvreté du discours, restriction quantitative et la pauvreté du contenu du discours, restriction qualitative.

L'*anhédonie* est un symptôme qui regroupe les difficultés du schizophrène à éprouver de l'intérêt ou du plaisir. Cela peut se traduire par une perte d'intérêt pour les activités agréables ou une incapacité à éprouver du plaisir au cours d'activités habituellement considérées comme agréables.

L'*apathie* se caractérise par un manque d'énergie, d'entrain, d'intérêt. Les malades sont incapables de se mobiliser pour entreprendre ou terminer plusieurs types d'activités. Ce symptôme amène souvent une détérioration sociale et économique chez le malade.

Un *comportement bizarre* est un comportement qui ne correspond pas aux normes sociales et culturelles. Il peut s'agir, entre autres, de l'habillement de l'individu, de sa conduite sociale et sexuelle.

Les *hallucinations* correspondent à des modifications anormales de la perception. Ce sont des perceptions erronées qui surviennent en l'absence de tout stimulus extérieur. Elles peuvent se manifester dans toutes les modalités sensorielles, aussi bien auditives que gustatives, olfactives, tactiles ou visuelles.

Les *idées délirantes* consistent en une anomalie du contenu de la pensée. Ce sont des croyances personnelles erronées qui ne trouvent pas d'explication dans le contexte culturel ou religieux du malade. Il peut s'agir, entre autres, d'idées délirantes de jalousie, de culpabilité.

La *pauvreté affective* se manifeste par un appauvrissement caractéristique de l'expression de la réactivité et de la sensibilité émotionnelle. Prenons l'exemple de la pauvreté du contact visuel. Dans ce cas, le malade évite de regarder l'autre ou d'utiliser ses yeux pour s'exprimer.

Le *trouble de la pensée formelle* se caractérise par un discours facile mais, pour des raisons diverses, il est porteur de peu d'information. Le malade a tendance à passer de façon improvisée d'un sujet à un autre, à être facilement distrait par l'environnement immédiat, à juxtaposer des mots sur leur seule ressemblance phonétique ou sémantique même si aucune signification n'en émerge, ou à ignorer une question posée pour répondre à une autre.

## 5.2 Concept des symptômes positifs et négatifs

Dans une publication autorisée par le ministre de la Santé nationale et du Bien-être social intitulée *La schizophrénie : guide à l'intention des familles*, les psychiatres différencient deux catégories de symptômes de la schizophrénie : les symptômes productifs (positifs) et les symptômes déficitaires (négatifs). On caractérise de *positifs* les symptômes qu'on retrouve chez les individus ayant une plus forte expressivité dans leurs comportements que chez les individus non atteints. Par contre, on caractérise de *négatifs* les symptômes chez les individus manifestant une plus faible expressivité. Les huit symptômes décrits à la section précédente sont regroupés en symptômes positifs et négatifs au tableau 5.1 (Andreasen, 1987).

Le concept des symptômes positifs et négatifs fut introduit par deux neurologues anglais du XIX<sup>e</sup> siècle, John Reynolds et Hughlings Jackson (Roy et DeVriendt, 1994). Décrivant les manifestations de l'épilepsie, ils postulèrent que les symptômes positifs refléteraient l'hyperfonctionnement de certaines régions du cerveau normalement inhibées, alors que les symptômes négatifs seraient sous-tendus par la perte de fonctions normales associées à d'autres régions du cerveau. Ainsi, la distinction entre ces deux catégories de symptômes était initialement basée sur la phénoménologie clinique (Minas et al., 1992).

En 1980, Tim Crow proposa deux types de schizophrénie qui sont basés sur la distinction entre les symptômes positifs et négatifs (Roy et DeVriendt, 1994). Le type I est caractérisé par deux des quatre symptômes positifs (tableau 5.1), soit les hallucinations et les idées délirantes, qui seraient sous-tendus par un dérèglement biochimique. Par contre, le type II est caractérisé par deux items rattachés à deux symptômes négatifs, soit l'émoussement affectif (item de la pauvreté affective) et la pauvreté du discours (item de l'alogie), qui seraient sous-tendus par des lésions anatomiques. Crow ne considérait pas ces deux types comme étant mutuellement exclusifs, mais comme des di-

Symptômes positifs	Symptômes négatifs
Hallucinations	Pauvreté affective
Idées délirantes	Alogie
Comportement bizarre	Apathie
Trouble de la pensée	Anhédonie

Tableau 5.1: Identification des symptômes positifs et négatifs

mensions indépendantes sous-jacentes à différents processus pathologiques mentionnés précédemment (Liddle, 1987).

Andreasen adopta un point de vue différent concernant la relation entre les symptômes positifs et négatifs (Liddle, 1987). Elle considère ces symptômes comme étant une caractéristique de deux maladies différentes. Andreasen développa deux échelles d'appréciation de symptômes observés chez les patients présentant une affection psychiatrique importante : l'Échelle d'appréciation des symptômes positifs ("The Scale for the Assessment of Positive Symptoms (SAPS)") et l'Échelle d'appréciation des symptômes négatifs ("The Scale for the Assessment of Negative Symptoms (SANS)") (Peralta et al., 1992). La SANS est composée de 25 items répartis en 5 symptômes négatifs (Andreasen, 1987). La SAPS est composée de 34 items répartis en 4 symptômes positifs. La cotation est constante pour chaque item : 0 (absent), 1 (discutable), 2 (léger), 3 (moyen), 4 (important) et 5 (sévère). Comparativement au modèle de Crow, celui d'Andreasen est nettement moins restrictif, c'est-à-dire qu'il comporte un plus grand nombre d'items pour caractériser les symptômes positifs et négatifs que celui de Crow.

La SANS et la SAPS ont donné lieu à des études de fidélité interjuge satisfaisantes (Andreasen, 1987). Une des raisons en est peut-être que ces échelles sont fondées sur une observation clinique soigneuse et détaillée. À la place d'essayer de deviner ce que le patient pense ou ressent, le clinicien observe et pose des questions.

La perspective dimensionnelle d'Andreasen est devenue de plus en plus employée dans l'étude des symptômes positifs et négatifs de la schizophrénie (Roy et DeVriendt,

1994). Les approches dimensionnelles postulent que les symptômes positifs et négatifs sont des phénomènes indépendants, sous-tendus par des pathophysiologies distinctes, mais qui peuvent se manifester simultanément chez un sujet. Dans plusieurs études antérieures, les corrélations entre les symptômes positifs et négatifs ont été faibles ou nulles, suggérant que les symptômes positifs et négatifs sont grandement indépendants (Tsuang, 1993). Ainsi, selon le modèle dimensionnel, les schizophrènes diffèrent les uns des autres quant à la sévérité des symptômes positifs et négatifs sans pour autant appartenir à des catégories distinctes (Roy et DeVriendt, 1994). Les différences entre les schizophrènes sont conçues comme étant quantitatives plutôt que qualitatives.

C'est en se basant sur cette littérature que ce mémoire s'appuie pour effectuer une analyse de liaison génétique entre chacune des dimensions symptomatiques, qui deviendront des phénotypes, et un ensemble de marqueurs génétiques. Les résultats des études pratiques de ce mémoire sont présentés au chapitre suivant.

# Chapitre 6

## Études pratiques d'analyse statistique de la liaison génétique pour les deux psychoses majeures

### 6.1 Description de l'échantillon de familles

La banque de données utilisée pour la réalisation des différentes analyses de ce chapitre provient du Centre de recherche Université Laval Robert-Giffard. L'échantillon comporte 20 familles dont le nombre de générations varie de deux à quatre. Ces 20 familles sont originaires de quatre régions de l'Est du Québec : la ville de Québec et ses environs, la région du Saguenay-Lac-St-Jean, les Iles-de-la-Madeleine et la région de l'Acadie au Nouveau-Brunswick. Chacune de ces familles est composée d'un certain nombre d'individus affectés appartenant à l'une ou l'autre des deux psychoses majeures à l'étude, soit la schizophrénie, soit le trouble bipolaire.

#### 6.1.1 Sélection de l'échantillon

La collecte des données a débuté en 1990 et se poursuit toujours. La sélection de chacune des familles s'est effectuée à partir d'un ou de plusieurs probands. Le proband correspond à l'individu qui est à l'origine de l'identification de la famille. Les critères d'entrée pour qu'une famille fasse partie de l'échantillon sont les suivants :

1. Au moins 4 individus affectés et apparentés au premier, deuxième ou troisième degré de parenté.
2. Fratrie nombreuse dans les branches affectées (plus que 6 enfants).
3. Bonne coopération attendue d'au moins 85 % des membres de la famille.
4. Transmission unilatérale de la maladie dans les générations antérieures, c'est-à-dire qu'un seul parent soit affecté; apparence de transmission verticale sur plusieurs générations.

Ensuite, pour faire partie de l'étude, les membres de chaque famille devaient alors donner leur consentement.

Parmi les 20 familles, il y en a 7 dont les individus affectés sont majoritairement du trouble bipolaire, 7 composées en majorité de schizophrènes et les 6 autres ont environ autant d'individus atteints pour chacune des deux psychoses majeures. Le tableau 6.1 identifie chacune des familles selon leur type (SZ = schizophrène, BP = trouble bipolaire, MIXTE = à la fois SZ et BP), le nombre total d'individus qui la composent et le nombre d'individus affectés. À l'annexe A, on retrouve un exemple de chacun de ces types de familles ainsi que la légende des diagnostics selon le DSM-III-R. La famille 255 est composée de 12 individus affectés de schizophrénie et de 6 individus de trouble bipolaire; par conséquent, cette famille est de type SZ. Quant à la famille 125, on compte 2 schizophrènes et 13 individus affectés de trouble bipolaire; donc, cette famille correspond au type BP. Finalement, la famille 151 est de type MIXTE, car elle contient 3 schizophrènes et 4 individus affectés de trouble bipolaire.

L'ensemble de ces 20 familles totalise 628 individus. Chacun de ces individus a été soumis à une évaluation de son état de santé mentale; 185 de ceux-ci ont obtenu un diagnostic appartenant à un type de désordre de l'une ou l'autre des deux psychoses majeures étudiées.



Identification de la famille	Type de famille	Nombre total d'individus	Nombre d'individus affectés
101	MIXTE	38	13
103	MIXTE	29	14
105	BP	21	6
105A	MIXTE	29	7
110	BP	17	6
111	BP	25	9
125	BP	34	15
151	MIXTE	30	10
182	BP	25	10
208	SZ	13	3
210	SZ	78	6
211	BP	36	11
211A	SZ	16	3
212	BP	49	13
218	MIXTE	12	3
220	SZ	35	10
228	SZ	25	10
230	MIXTE	23	8
250	SZ	36	10
255	SZ	57	18
TOTAL		628	185

Tableau 6.1: Nombre total d'individus et le nombre d'individus affectés, par famille

## 6.1.2 Évaluation du diagnostic

Les 628 individus de l'échantillon de familles ont été rencontrés et évalués séparément. Lorsqu'il y avait lieu d'établir un diagnostic, on s'est basé sur trois sources d'information :

1. Le dossier médical du patient.
2. Les questionnaires "Échelle d'appréciation des symptômes positifs (SAPS)" et "Échelle d'appréciation des symptômes négatifs (SANS)" d'Andreasen ont été remplis pour chaque individu permettant d'évaluer la présence et le degré de sévérité des symptômes positifs et négatifs. Pour chacun des items de ces symptômes, le choix de réponses était coté de 0 à 5, dont la cote 5 correspondait à une situation avec sévérité élevée du symptôme.
3. Une entrevue "Structured Clinical Interview for DSM-III-R (SCID)" fut réalisée en la présence d'un membre de la même famille que le patient. Ce membre devait être capable de très bien décrire le comportement du patient lors des épisodes aigus de la maladie car, au moment des entrevues, il a été observé que les patients avaient tendance à sous-estimer leurs symptômes.

Ces trois sources d'information furent collectées par des assistants de recherche qualifiés et expérimentés, sans qu'ils ne connaissent les génotypes des marqueurs génétiques du patient. Sous la supervision d'un médecin, les assistants de recherche ont d'abord examiné toute cette information et un diagnostic a été posé selon les critères du DSM-III-R. Par la suite, une deuxième évaluation du diagnostic des patients a été effectuée par quatre psychiatres du CDR (Conseil Diagnostique de Recherche). Ils avaient obtenu toute l'information sur chacun des patients afin de poser un second diagnostic à l'aveugle, c'est-à-dire sans connaître les génotypes des marqueurs du patient, le diagnostic du proband, le type de famille auquel le patient appartenait et les diagnostics obtenus des assistants de recherche des autres membres de la famille. Les quatre psychiatres tentaient d'obtenir un consensus sur le diagnostic de chacun des patients. Si les deux dia-

gnostics différaient, le diagnostic résultant du consensus des quatre psychiatres (CDR) était préféré à celui posé par les assistants de recherche.

Le tableau 6.2 montre la répartition des 185 patients de l'échantillon selon différentes définitions usuelles des deux psychoses majeures et le regroupement des diagnostics du DSM-III-R correspondants. Les définitions de schizophrénie stricte et large s'appliquent aux familles de types SZ et MIXTE (voir tableau 6.1). Les définitions de trouble bipolaire strict et large s'appliquent aux familles de types BP et MIXTE. Tandis que les définitions de gène commun strict et large s'appliquent aux trois types de famille, c'est-à-dire à l'échantillon total.

### **6.1.3 Scores des symptômes positifs et négatifs**

Les quatre symptômes positifs des deux psychoses majeures étudiées correspondent aux hallucinations, aux idées délirantes, au comportement bizarre et au trouble de la pensée formelle alors que les symptômes négatifs se rapportent à la pauvreté affective, l'alogie, l'apathie et l'anhédonie. Pour chacun des 185 individus affectés de l'échantillon, des scores moyens pour les syndromes positif et négatif des deux maladies à l'étude ont été définis à partir des moyennes des cotes obtenues aux items du questionnaire d'Andreasen pour chacun des symptômes qui les caractérisaient. On a considéré les scores obtenus des symptômes positifs et négatifs se rapportant aux épisodes aigus de la maladie.

La figure 6.1 représente la distribution de fréquences des scores des symptômes positifs et négatifs. D'abord, on remarque que l'étendue des scores des symptômes négatifs est plus grande que celle des scores des symptômes positifs. De plus, il y a un plus grand nombre d'individus ayant un score 0 pour les symptômes négatifs ( $N=40$ ) comparativement à celui pour les symptômes positifs ( $N=28$ ). Le tableau 6.3 comporte quelques statistiques descriptives de ces symptômes. La moyenne, l'écart-type, la médiane et la valeur maximale des scores pour les symptômes négatifs sont plus grands que ceux

Définition usuelle des phénotypes correspondant aux deux psychoses majeures	Diagnostics selon le DSM-III-R	Nombre d'individus
Schizophrénie stricte	Schizophrénie	54
Schizophrénie large	Schizophrénie stricte + Trouble schizophréniforme + Trouble schizotypique	58
Trouble bipolaire strict	Trouble bipolaire 1 sans caractéristique psychotique + Trouble bipolaire 1 avec caractéristique psychotique congruente + Trouble bipolaire 1 avec caractéristique psychotique non congruente	41
Trouble bipolaire large	Trouble bipolaire strict + Trouble bipolaire 2 sans caractéristique psychotique + Trouble bipolaire 2 avec caractéristique psychotique congruente + Trouble bipolaire 2 avec caractéristique psychotique non congruente + Dépression majeure récurrente sans caractéristique psychotique + Dépression majeure récurrente avec caractéristique psychotique + Dépression majeure récurrente avec caractéristique psychotique congruente	56
Gène commun strict	Schizophrénie stricte + Trouble bipolaire strict + Trouble schizo-affectif	117
Gène commun large	Schizophrénie large + Trouble bipolaire large + Trouble schizo-affectif + Psychose réactionnelle brève	142

Tableau 6.2: Nombre d'individus affectés selon une définition usuelle du phénotype pour chacune des deux psychoses majeures avec les diagnostics du DSM-III-R correspondants

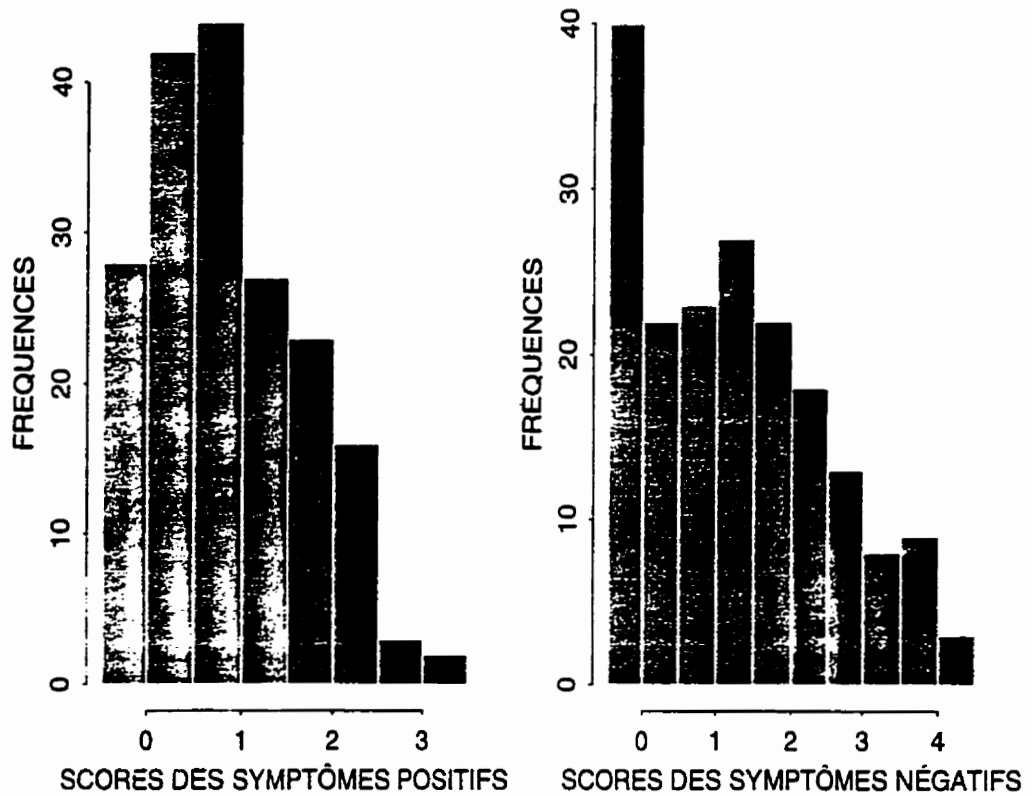


Figure 6.1: Distribution de fréquences des scores des symptômes positifs et négatifs pour les symptômes positifs. Ainsi, il semblerait que le degré de sévérité des symptômes négatifs serait plus élevé que celui des symptômes positifs.

## 6.2 Utilisation d'un seuil de coupure sur les scores des symptômes et nouvelle définition du phénotype

### 6.2.1 Détermination du seuil de coupure

Pour chacun des 185 patients de l'échantillon, on déterminera si le patient a le phénotype *affecté* ou non selon un seuil de coupure déterminé pour chacun des symptômes positifs

	Scores	
	Symptômes positifs	Symptômes négatifs
Nombre d'observations	185	185
Moyenne	0.90	1.34
Écart-type	0.79	1.17
Médiane	0.79	1.17
Valeur minimale	0	0
Valeur maximale	3.47	4.38

Tableau 6.3: Statistiques descriptives des scores des symptômes positifs et négatifs

Seuil de coupure	Centile	
	Symptômes positifs	Symptômes négatifs
0.5	38 (70/185)	34 (62/185)
1.0	62 (114/185)	46 (85/185)

Tableau 6.4: Valeur du centile aux seuils de coupure de 0.5 et 1.0 pour les symptômes positifs et négatifs

et négatifs. Deux seuils de coupure ont été choisis en fonction du système de cotation des symptômes positifs et négatifs et de la valeur du centile de ces seuils. Je rappelle que le système de cotation décrit à la section 5.2 des items des symptômes positifs et négatifs demeure le suivant : 0 (absent), 1 (discutable), 2 (léger), 3 (moyen), 4 (important) et 5 (sévère). Les seuils de coupure de 0.5 et 1.0 furent choisis comme nouvelle définition du phénotype. Le seuil de coupure de 1.0 correspond à une sévérité *discutable* tandis que le seuil de coupure de 0.5 se rapporte à une sévérité au milieu de celle *absente et discutable*. Il est certain que d'autres seuils de coupure auraient pu tout aussi bien être choisis.

Le tableau 6.4 représente le centile des seuils de coupure de 0.5 et 1.0 pour les symptômes positifs et négatifs. On remarque que les centiles des deux seuils de coupure pour les deux catégories de symptômes sont assez différents. Ainsi, on pourra vérifier si le seuil de coupure a un effet sur le Lod-score.

## 6.2.2 Concordance entre les définitions usuelles du phénotype pour chacune des deux psychoses majeures et la nouvelle définition utilisant les scores aux symptômes positifs et négatifs

Il est important de vérifier le taux de concordance entre les définitions usuelles du phénotype et la nouvelle définition utilisant une classification des patients selon leurs scores obtenus des symptômes positifs et négatifs. On voudrait bien que ce taux de concordance soit inférieur à 100 % afin de justifier l'utilisation de la nouvelle définition du phénotype.

Le tableau 6.5 représente un tableau de fréquences entre une définition usuelle du phénotype pour les deux psychoses majeures et la nouvelle définition basée sur le score des symptômes avec un seuil de coupure donné. Différentes proportions peuvent être calculées à partir de ce tableau. Les diverses proportions qui ont été considérées sont les suivantes :

- $p_1 = \frac{a + d}{a + b + c + d}$  : taux de concordance entre une définition usuelle et la nouvelle définition (Fleiss, 1981);
- $p_2 = \frac{a}{a + c}$  : proportion des patients ayant un score supérieur au seuil de coupure qui sont à la fois caractérisés par la définition usuelle du phénotype;
- $p_3 = \frac{a}{a + b}$  : proportion des patients caractérisés par la définition usuelle du phénotype qui ont aussi un score supérieur au seuil de coupure.

Le tableau 6.6 contient les trois proportions décrites précédemment avec des seuils de coupure de 0.5 et 1.0 pour les symptômes positifs tandis que, pour les symptômes négatifs, il faut se référer au tableau 6.7. Que ce soit pour les symptômes positifs ou pour les symptômes négatifs, tous les taux de concordance ( $p_1$ ) sont inférieurs à 100 %. Cela indique que la nouvelle définition du phénotype utilisant un seuil de coupure est différente des définitions usuelles du phénotype selon les diagnostics du DSM-III-R.

Phénotype 1		Phénotype 2		
		Score des symptômes		
		affecté ( $\geq$ seuil)	non affecté ( $<$ seuil)	
Définition usuelle	affecté	a	b	a + b
	non affecté	c	d	c + d
		a + c	b + d	a + b + c + d

Tableau 6.5: Tableau de fréquences entre une définition usuelle du phénotype correspondant à chacune des deux psychoses majeures et la nouvelle définition basée sur les scores des symptômes positifs et négatifs avec un seuil de coupure donné

Ainsi, on utilisera cette nouvelle définition du phénotype pour procéder aux différentes analyses statistiques de la liaison génétique sur l'échantillon de familles. En général, les taux de concordance pour les symptômes positifs au tableau 6.6 sont légèrement supérieurs à ceux des symptômes négatifs au tableau 6.7. Pour les symptômes positifs et négatifs ainsi que pour les deux seuils de coupure, on remarque que les proportions  $p_2$  et  $p_3$  des définitions trouble bipolaire strict et large sont plus petites que celles des définitions schizophrénie stricte et large.

### 6.3 Spécification des composantes nécessaires pour le calcul de la vraisemblance d'un pedigree avec l'algorithme d'Elston-Stewart

#### 6.3.1 La fréquence des allèles pour les différents marqueurs génétiques utilisés

Pour obtenir la fréquence des allèles de différents marqueurs génétiques, on a utilisé un échantillon de 60 individus non liés et sans histoire psychiatrique au premier degré de parenté. Leurs conjoints(es) atteints(es) ou non de la maladie font aussi partie de l'échantillon. Ainsi, on a utilisé 240 chromosomes pour estimer la fréquence des allèles de chacun des marqueurs génétiques.



Type d'échantillon	Seuil de 0.5			Seuil de 1.0		
	$p_1$	$p_2$	$p_3$	$p_1$	$p_2$	$p_3$
<b>Échantillon des schizophrènes</b>						
Schizophrénie stricte	69.6 (80/115)	62.0	90.7	74.8 (86/115)	71.2	77.8
Schizophrénie large	71.3 (82/115)	65.8	89.7	74.8 (86/115)	74.6	75.9
<b>Échantillon des troubles bipolaires</b>						
Trouble bipolaire strict	58.4 (73/125)	42.7	78.1	54.4 (68/125)	30.0	29.3
Trouble bipolaire large	49.6 (62/125)	45.3	60.7	42.4 (53/125)	30.0	21.4
<b>Échantillon total</b>						
Gène commun strict	83.2 (154/185)	87.1	86.3	72.4 (134/185)	95.8	59.0
Gène commun large	76.2 (141/185)	92.2	75.4	61.1 (113/185)	98.6	50.0

Tableau 6.6: Proportions calculées entre diverses définitions usuelles du phénotype pour les deux psychoses majeures et la nouvelle définition basée sur le score des symptômes positifs avec des seuils de coupure de 0.5 et 1.0

Type d'échantillon	Seuil de 0.5			Seuil de 1.0		
	$p_1$	$p_2$	$p_3$	$p_1$	$p_2$	$p_3$
<b>Échantillon des schizophrènes</b>						
Schizophrénie stricte	67.0 (77/115)	59.1	96.3	72.2 (83/115)	64.1	92.6
Schizophrénie large	68.7 (79/115)	62.5	94.8	72.2 (83/115)	66.7	89.7
<b>Échantillon des troubles bipolaires</b>						
Trouble bipolaire strict	51.2 (64/125)	36.5	65.9	52.0 (65/125)	32.1	41.5
Trouble bipolaire large	47.2 (59/125)	43.4	57.1	46.4 (58/125)	39.6	37.5
<b>Échantillon total</b>						
Gène commun strict	73.5 (136/185)	77.0	82.9	70.3 (130/185)	80.4	70.1
Gène commun large	73.0 (135/185)	86.5	76.8	66.5 (123/185)	89.2	64.1

Tableau 6.7: Proportions calculées entre diverses définitions usuelles du phénotype pour les deux psychoses majeures et la nouvelle définition basée sur le score des symptômes négatifs avec des seuils de coupure de 0.5 et 1.0

Marqueur génétique	Nombre d'allèles	Fréquence des allèles
<b>Chromosome 6</b>		
D6S89	14	0.008, 0.093, 0.130, 0.033, 0.069, 0.114, 0.248, 0.020, 0.037, 0.016, 0.191, 0.004, 0.004
D6S260	17	0.004, 0.004, 0.014, 0.005, 0.004, 0.032, 0.032, 0.082, 0.0137, 0.142, 0.092, 0.119, 0.045, 0.105, 0.050, 0.055
D6S274	12	0.009, 0.009, 0.032, 0.116, 0.207, 0.106, 0.042, 0.032, 0.065, 0.130, 0.004
D6S277	12	0.008, 0.004, 0.008, 0.120, 0.215, 0.152, 0.072, 0.223, 0.020, 0.004, 0.003
D6S296	22	0.012, 0.004, 0.048, 0.048, 0.004, 0.004, 0.004, 0.016, 0.0333, 0.068, 0.020, 0.128, 0.008, 0.004, 0.004, 0.012, 0.012, 0.003, 0.056, 0.008
D6S309	11	0.003, 0.008, 0.193, 0.094, 0.031, 0.185, 0.114, 0.083, 0.008, 0.063
D6S334	12	0.043, 0.055, 0.024, 0.094, 0.063, 0.094, 0.142, 0.016, 0.012, 0.008, 0.177
<b>Chromosome 11</b>		
D11S35	7	0.112, 0.293, 0.208, 0.101, 0.147, 0.136, 0.003
D11S876	14	0.003, 0.004, 0.023, 0.047, 0.154, 0.058, 0.093, 0.097, 0.050, 0.085, 0.123, 0.085, 0.003
D11S938	6	0.009, 0.184, 0.197, 0.038, 0.563, 0.009
<b>Chromosome 18</b>		
D18S57	15	0.023, 0.078, 0.163, 0.070, 0.039, 0.063, 0.097, 0.289, 0.031, 0.051, 0.008, 0.003, 0.003, 0.004
D18S65	7	0.070, 0.004, 0.393, 0.136, 0.235, 0.159, 0.003
D18S69	9	0.023, 0.273, 0.265, 0.027, 0.187, 0.004, 0.210, 0.008, 0.004
D18S456	8	0.004, 0.003, 0.018, 0.262, 0.334, 0.348, 0.027, 0.004
D18S457	10	0.008, 0.062, 0.039, 0.047, 0.120, 0.678, 0.003, 0.004, 0.035
D18S472	9	0.012, 0.004, 0.120, 0.008, 0.074, 0.035, 0.019, 0.631, 0.004

Tableau 6.8: Fréquence observée des allèles pour les différents marqueurs génétiques des chromosomes 6, 11 et 18, basée sur un échantillon de 120 individus

Pour les études pratiques de ce mémoire, on a retenu 16 marqueurs génétiques répartis sur les chromosomes 6, 11 et 18. On retrouve au tableau 6.8 le nombre d'allèles ainsi que la fréquence observée des allèles des 16 marqueurs génétiques des chromosomes 6, 11 et 18, basée sur l'échantillon de 120 individus. Ce sont ces fréquences des allèles qui permettront d'estimer la distribution des différents génotypes au locus de chacun de ces 16 marqueurs génétiques pour les parents fondateurs ainsi que pour les époux(es) des membres de chacune des 20 familles à l'étude. Cette distribution correspond à la composante  $\Psi$ , de l'algorithme d'Elston-Stewart décrit à la section 2.4.2.

À l'annexe B, on a les cartes génétiques des chromosomes 6, 11 et 18. À gauche de chaque chromosome, les distances génétiques sont cumulées à partir du centromère et exprimées en morgans (Ott, 1991). Un morgan est l'unité de distance génétique. Chez l'humain, une distance génétique de 1 centimorgan (0.01 morgan) correspond approximativement à 1 million de paires de bases. À droite de chaque chromosome, on retrouve les distances génétiques en centimorgans entre différents marqueurs génétiques. Au tableau 6.8, les marqueurs du chromosome 6 sont situés sur le bras court (p) et les marqueurs des chromosomes 11 et 18 sur le bras long (q).

### 6.3.2 Le choix du modèle de transmission génétique

Afin d'effectuer une analyse de linkage paramétrique, il est nécessaire de spécifier un modèle de transmission génétique. Ce modèle pour la schizophrénie et le trouble bipolaire n'est pas bien connu. Cependant, on sait qu'il est possible de faire une analyse de linkage même avec un modèle génétique inexact. Il est toutefois important que les paramètres génétiques (pénétrance, fréquence des allèles) au locus de la maladie soient approximativement corrects (Greenberg et al., 1996). Aussi, on a suivi en particulier la recommandation de ces auteurs qui ont étudié l'impact d'une erreur dans la spécification des paramètres du modèle de transmission sur l'estimation de la fraction de recombinaison dans une analyse de linkage paramétrique :

“Therefore, if we consider at least a dominant and a recessive model in the analysis, we are much more likely to detect a true linkage than if we limit ourselves to a single genetic model. Also, if we use both a dominant and a recessive model, then our *power* to detect linkage is almost the same as if we had used a single correct model.”

Ainsi, dans notre étude, on a considéré un modèle dominant et récessif pour les analyses de linkage des deux psychoses majeures.

### 6.3.3 La pénétrance dépendante de l'âge

Concernant la schizophrénie et le trouble bipolaire, la pénétrance est dépendante de l'âge. Pour définir la distribution des âges de début de la schizophrénie et du trouble bipolaire, on s'est servi de cinq articles en épidémiologie. En moyenne, l'âge de début du trouble bipolaire est supérieur de cinq années à celui de la schizophrénie pour chacune des catégories d'âge (0–19, 20–24, 25–29, 30–34,  $\geq 35$ ). Les analyses de linkage pour les deux psychoses majeures ont été réalisées en ne considérant que la distribution des âges de début de la schizophrénie.

La pénétrance dépendante de l'âge pour les deux psychoses majeures du modèle dominant se retrouve au tableau 6.9 tandis qu'il faut se référer au tableau 6.10 pour celle du modèle récessif. Par exemple, au locus des deux maladies, on observe les allèles  $A$  (dominant) et  $a$  (récessif). Supposons que, pour ces deux modèles, les paramètres de prédisposition du génotype susceptible  $g$  de ces maladies sont les suivants :

- $\gamma_{AA} = \gamma_{Aa} = 0.70$  (modèle dominant);
- $\gamma_{aa} = 0.70$  (modèle récessif).

Ainsi, parmi les individus de génotype susceptible pour les deux psychoses majeures,

Connaissance de l'âge de début	Catégorie d'âge	Génotype		
		aa	Aa	AA
Non	0-19	0.00115	0.21700	0.21700
	20-24	0.00230	0.43400	0.43400
	25-29	0.00293	0.55300	0.55300
	30-34	0.00333	0.63000	0.63000
	≥ 35	0.00370	0.70000	0.70000
Oui	0-19	0.00115	0.21700	0.21700
	20-24	0.00115	0.21700	0.21700
	25-29	0.00063	0.11900	0.11900
	30-34	0.00040	0.07700	0.07700
	≥ 35	0.00037	0.07000	0.07000

Tableau 6.9: La pénétrance dépendante de l'âge du modèle dominant

70 % d'entre eux exprimeront la maladie s'ils vivent suffisamment longtemps pour chacun des deux modèles.

Pour les deux modèles en question, les fréquences de l'allèle malade qui ont été choisies sont les suivantes :

- $P(A) = 0.01 \Rightarrow P(a) = 0.99$  (modèle dominant);
- $P(a) = 0.10 \Rightarrow P(A) = 0.90$  (modèle récessif).

La prévalence d'une maladie dans une population, où la pénétrance est indépendante de l'âge, se calcule de la façon suivante :

$$\text{Prévalence} = P(\text{affecté}) = P(\text{affecté} | AA) P(AA) + 2 P(\text{affecté} | Aa) P(Aa) + P(\text{affecté} | aa) P(aa).$$

La prévalence des deux psychoses majeures dans la population pour les modèles dominant et récessif se calcule à partir des valeurs aux tableaux 6.9 et 6.10, utilisant la formule précédente mais ajustée afin de tenir compte de l'âge. Les modèles dominant et

Connaissance de l'âge de début	Catégorie d'âge	Génotype		
		aa	Aa	AA
Non	0-19	0.217000	0.000248	0.000248
	20-24	0.434000	0.000496	0.000496
	25-29	0.553000	0.000632	0.000632
	30-34	0.630000	0.000720	0.000720
	≥ 35	0.700000	0.000800	0.000800
Oui	0-19	0.217000	0.000248	0.000248
	20-24	0.217000	0.000248	0.000248
	25-29	0.119000	0.000136	0.000136
	30-34	0.077000	0.000088	0.000088
	≥ 35	0.070000	0.000080	0.000080

Tableau 6.10: La pénétrance dépendante de l'âge du modèle récessif

récessif prédisent une prévalence de 0.0032 et 0.0015, respectivement. Comparativement à la fréquence de l'allèle malade du modèle dominant, celle de l'allèle malade du modèle récessif doit être plus élevée pour respecter la prévalence réelle des deux psychoses majeures dans la population qui est d'environ 1 %. On remarque que les prévalences prédites par des modèles dominant et récessif sont inférieures au 1 % afin de représenter le fait que ces maladies sont probablement dues à plus d'un gène (Risch, 1990).

On peut aussi calculer le taux de phénotypie pour chacun des deux modèles choisis. Par exemple, pour le modèle dominant, le taux de phénotypie (proportion des phénotopies, c'est-à-dire des individus affectés sans avoir le gène, parmi tous les individus affectés), où la pénétrance est indépendante de l'âge, s'exprime comme suit :

$$\frac{P(\text{affecté}|aa)P(aa)}{P(\text{affecté}|AA)P(AA) + 2P(\text{affecté}|Aa)P(Aa) + P(\text{affecté}|aa)P(aa)}$$

On note que le numérateur représente la fréquence des phénotopies tandis que le dénominateur correspond à la prévalence des deux maladies. Ainsi, en tenant compte de l'âge, les taux de phénotypie des modèles dominant et récessif sont de 0.2067 et 0.1027, respectivement.

La figure 6.2 illustre la pénétrance dépendante de l'âge pour les individus de génotype susceptible des deux maladies des modèles dominant et récessif. Pour le modèle dominant, les génotypes susceptibles sont  $AA$  et  $Aa$  tandis que, pour le modèle récessif, le génotype susceptible est  $aa$ . Il y a trois cas indiqués par un chiffre à la figure 6.2 qui peuvent se produire. La détermination de leur pénétrance s'effectue à l'aide de l'équation 2.1, 2.2 ou 2.3 de la section 2.4.2.1 :

1. Par exemple, pour le premier cas, un individu âgé de 32 ans devenu affecté à l'âge de 22 ans, sa pénétrance s'écrit de la façon suivante :

$$g_g(Z = 1, a = 22 | a' = 32) = \gamma_g f(22) = 0.70 \times 0.31 = 0.217.$$

2. Pour le deuxième cas, un individu affecté âgé de 33 ans dont on ne connaît pas l'âge au début de la maladie, sa pénétrance sera la suivante :

$$g_g(Z = 1 | a' = 33) = \gamma_g F(33) = 0.70 \times 0.90 = 0.63.$$

3. Enfin, pour le troisième cas, un individu non affecté âgé de 18 ans, sa pénétrance correspondra à ce qui suit :

$$g_g(Z = 0 | a' = 18) = 1 - \gamma_g F(18) = 1 - 0.70 \times 0.31 = 0.783.$$

Le calcul des pénétrances dépendantes de l'âge pour les individus de génotype non susceptible des deux psychoses majeures s'effectue de la même façon que les trois cas présentés ci-dessus.

## 6.4 Utilisation de deux classifications du phénotype à partir des seuils de coupure fixés

Pour chacun des 185 patients de l'échantillon, on a déterminé le statut du phénotype selon un seuil de coupure fixé pour chaque catégorie de symptômes (positifs et négatifs).



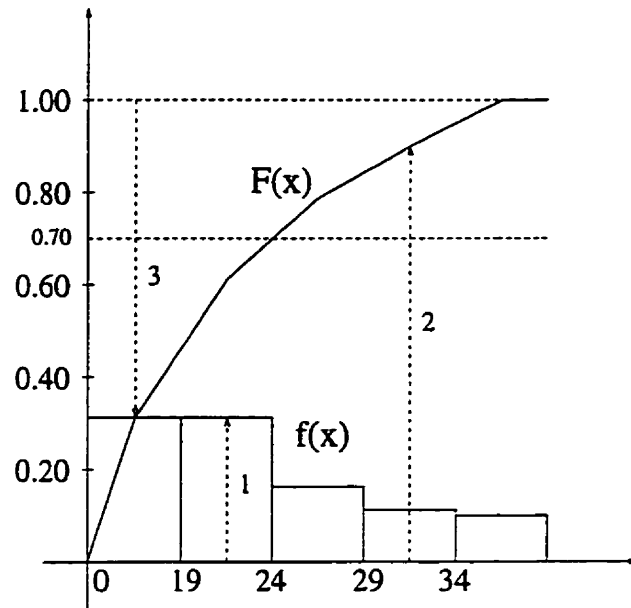


Figure 6.2: Fonction de répartition ( $F$ ) et fonction de densité ( $f$ ) de l'âge d'apparition de la maladie pour les individus de génotype susceptible des deux psychoses majeures. Détermination de la pénétrance de 3 individus (dénotés 1, 2 et 3)

Rappelons que les seuils de coupure sont de 0.5 et 1.0. Pour la première classification en trois catégories, la figure 6.3 illustre comment on détermine le phénotype d'un patient, à partir de son score aux symptômes positifs et négatifs, par rapport à un seuil de coupure fixé. D'après cette première classification, un individu est considéré *non affecté* s'il n'avait pas de score, *inconnu* si son score se situait entre 0 inclusivement et le seuil de coupure, et *affecté* si son score était supérieur ou égal au seuil de coupure. Une deuxième classification en deux catégories, *inconnu* et *affecté*, a aussi été utilisée comme stratégie pour diminuer l'impact des estimateurs de la pénétrance sur les résultats. Cette deuxième classification est illustrée à la figure 6.4. Les deux classifications considérées sont les suivantes : la première classification (non affecté, inconnu, affecté) notée na-i-a, et la deuxième classification (inconnu, affecté) notée i-a.

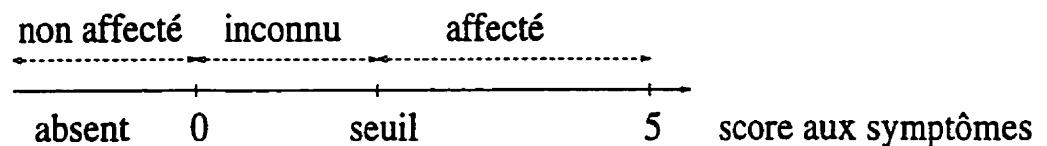


Figure 6.3: Première classification du phénotype par rapport à un seuil de coupure fixé

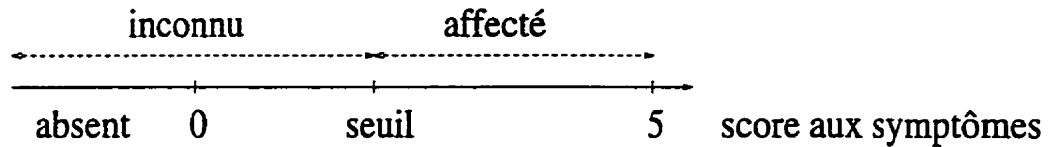


Figure 6.4: Deuxième classification du phénotype par rapport à un seuil de coupure fixé

## 6.5 Résultats des différentes analyses de linkage

Les analyses de linkage ont été effectuées pour les symptômes positifs et négatifs selon :

1. Les seuils de coupure de 0.5 et 1.0;
2. Les modèles dominant et récessif;
3. Les deux classifications du phénotype :
  - (a) non affecté, inconnu, affecté;
  - (b) inconnu, affecté.

Il y a présence d'une boucle dans les familles avec les numéros d'identification 101, 228 et 255. Pour les familles 101 et 228, il s'agit d'une boucle de mariage, tandis que pour la famille 255, c'est une boucle de consanguinité. Veuillez vous référer à la section 2.4.5 pour avoir plus d'information en ce qui a trait à la notion de boucle dans une famille. Pour les analyses de linkage, on a utilisé le programme informatique MLINK.

Le tableau 6.11 contient les Lod-scores maximums, avec la fraction de recombinaison ( $\theta$ ), supérieurs ou égaux à 1 pour le seuil de coupure de 0.5, tandis que le tableau 6.12 concerne le seuil de coupure de 1.0. On remarque qu'aucun Lod-score maximum n'excède la valeur critique  $Z_0 = 3$ .

Suite à ces résultats des analyses de linkage, on a décidé de faire d'autres analyses plus approfondies. On a fait la différence des Lod-scores maximums entre les seuils

Famille	Symptômes	Modèle	Classification	Marqueur	Lod-score maximum ( $\theta$ )
182	positifs	dominant	i-a	D6S274	1.32(.00)
212	positifs	dominant	i-a	D6S260	1.01(.00)
212	positifs	dominant	i-a	D6S309	1.13(.00)
151	positifs	récessif	na-i-a	D6S277	1.26(.00)
151	positifs	récessif	na-i-a	D6S296	1.22(.00)
151	positifs	récessif	na-i-a	D6S309	1.05(.00)
212	positifs	récessif	na-i-a	D6S309	1.07(.00)
250	négatifs	dominant	i-a	D6S89	1.07(.00)
250	négatifs	dominant	i-a	D6S260	1.07(.00)
250	négatifs	dominant	i-a	D6S274	1.07(.00)
250	négatifs	dominant	i-a	D6S334	1.07(.00)
228	positifs	dominant	na-i-a	D11S35	1.49(.00)
228	positifs	dominant	i-a	D11S35	1.01(.00)
228	positifs	dominant	na-i-a	D11S876	1.51(.00)
228	positifs	dominant	i-a	D11S876	1.03(.00)
255	positifs	dominant	na-i-a	D11S35	2.03(.00)
255	positifs	dominant	i-a	D11S35	1.29(.05)
255	positifs	récessif	na-i-a	D11S35	2.39(.00)
255	positifs	récessif	i-a	D11S35	1.47(.00)
182	négatifs	dominant	i-a	D11S876	1.00(.10)
255	négatifs	dominant	na-i-a	D11S35	2.16(.00)
255	négatifs	dominant	i-a	D11S35	1.35(.05)
255	négatifs	récessif	na-i-a	D11S35	1.20(.05)
111	positifs	dominant	na-i-a	D18S69	1.00(.00)
151	positifs	récessif	na-i-a	D18S456	1.03(.00)
151	positifs	récessif	i-a	D18S456	1.18(.00)
250	positifs	récessif	na-i-a	D18S457	1.60(.00)
250	positifs	récessif	na-i-a	D18S472	1.12(.00)
111	négatifs	dominant	na-i-a	D18S69	1.00(.00)
230	négatifs	récessif	na-i-a	D18S456	1.13(.00)
250	négatifs	récessif	na-i-a	D18S457	1.27(.00)
250	négatifs	récessif	na-i-a	D18S472	1.17(.00)

Tableau 6.11: Résultats des analyses de linkage qui ont donné des Lod-scores maximums supérieurs ou égaux à 1 pour le seuil de coupure de 0.5. Les deux classifications sont présentées à la section 6.4

Famille	Symptômes	Modèle	Classification	Marqueur	Lod-score maximum ( $\theta$ )
250	positifs	dominant	i-a	D6S89	1.10(.00)
250	positifs	dominant	i-a	D6S260	1.10(.00)
250	positifs	dominant	i-a	D6S274	1.10(.00)
250	positifs	dominant	i-a	D6S334	1.10(.00)
151	négatifs	dominant	na-i-a	D6S296	2.13(.00)
182	négatifs	dominant	i-a	D6S274	1.07(.00)
250	négatifs	dominant	i-a	D6S89	1.07(.00)
250	négatifs	dominant	i-a	D6S260	1.07(.00)
250	négatifs	dominant	i-a	D6S274	1.07(.00)
250	négatifs	dominant	i-a	D6S334	1.07(.00)
151	négatifs	récessif	na-i-a	D6S296	1.63(.00)
151	négatifs	récessif	na-i-a	D6S309	1.39(.00)
211	négatifs	récessif	i-a	D6S89	1.16(.00)
211	négatifs	récessif	na-i-a	D6S260	1.28(.00)
211	négatifs	récessif	i-a	D6S260	1.16(.00)
211	négatifs	récessif	na-i-a	D6S274	1.28(.00)
211	négatifs	récessif	i-a	D6S274	1.16(.00)
211	négatifs	récessif	na-i-a	D6S334	1.28(.00)
211	négatifs	récessif	i-a	D6S334	1.16(.00)
228	positifs	dominant	na-i-a	D11S35	1.50(.00)
228	positifs	dominant	i-a	D11S35	1.05(.00)
228	positifs	dominant	na-i-a	D11S876	1.51(.00)
228	positifs	dominant	i-a	D11S876	1.07(.00)
255	positifs	récessif	na-i-a	D11S35	1.32(.00)
255	négatifs	dominant	na-i-a	D11S35	1.96(.00)
255	négatifs	dominant	i-a	D11S35	1.47(.00)
255	négatifs	récessif	na-i-a	D11S35	2.50(.00)
255	négatifs	récessif	i-a	D11S35	1.55(.00)
101	positifs	récessif	i-a	D18S457	1.11(.00)
250	positifs	récessif	na-i-a	D18S457	1.38(.00)
250	positifs	récessif	na-i-a	D18S472	1.09(.00)
101	négatifs	dominant	i-a	D18S457	1.41(.00)
111	négatifs	dominant	na-i-a	D18S69	1.00(.00)
230	négatifs	récessif	na-i-a	D18S65	1.14(.00)
230	négatifs	récessif	na-i-a	D18S456	1.21(.00)
230	négatifs	récessif	na-i-a	D18S472	1.06(.00)
250	négatifs	récessif	na-i-a	D18S457	1.27(.00)
250	négatifs	récessif	na-i-a	D18S472	1.17(.00)

Tableau 6.12: Résultats des analyses de linkage qui ont donné des Lod-scores maximums supérieurs ou égaux à 1 pour le seuil de coupure de 1.0. Les deux classifications sont présentées à la section 6.4

Famille	Modèle	Classification	Marqueur	$Z_{0.5}(\hat{\theta})$	$Z_{1.0}(\hat{\theta})$	Différence
182	dominant	i-a	D6S274	1.32	0	1.32
151	récessif	na-i-a	D6S277	1.26	0	1.26
151	récessif	na-i-a	D6S296	1.22	0	1.22
151	récessif	na-i-a	D6S309	1.05	0	1.05
255	dominant	na-i-a	D11S35	2.03	0.09	1.94
255	dominant	i-a	D11S35	1.29	0.09	1.20
255	récessif	na-i-a	D11S35	2.39	1.32	1.07
151	récessif	na-i-a	D18S456	1.08	0	1.08
151	récessif	i-a	D18S456	1.18	0	1.18

Tableau 6.13: Différences supérieures à 1 des Lod-scores maximums entre les seuils de coupure de 0.5 et 1.0, en valeur absolue, pour les symptômes positifs

de coupure de 0.5 et 1.0 en valeur absolue. On a seulement retenu les différences qui étaient supérieures à 1. Le tableau 6.13 correspond à ces différences pour les symptômes positifs, tandis que le tableau 6.14 se rapporte aux symptômes négatifs. Par exemple, au tableau 6.11, on a obtenu un Lod-score maximum de 1.20 à  $\hat{\theta} = 0.05$  pour la famille 255 en ce qui concerne les symptômes négatifs, le modèle récessif, la première classification (na-i-a), le marqueur D11S35 et le seuil de coupure de 0.5. Pour la même famille et dans les mêmes conditions avec le seuil de coupure de 1.0, le Lod-score maximum était de 2.50 à  $\hat{\theta} = 0$ . au tableau 6.12. Il s'agit d'une différence de 1.30 entre ces Lod-scores maximums des seuils de coupure de 0.5 et 1.0. Pour cet exemple, on constate ainsi que le seuil de coupure a un effet sur la valeur du Lod-score maximum.

On s'est alors demandé si le Lod-score maximum augmenterait en élevant ou en baissant le seuil de coupure. Au tableau 6.13, pour les symptômes positifs, où les  $Z_{0.5}(\hat{\theta})$  étaient supérieurs aux  $Z_{1.0}(\hat{\theta})$ , on a fait de nouvelles analyses de linkage en ajoutant un à un les individus ayant des scores inférieurs à 0.5, afin de voir si le Lod-score maximum augmenterait. Au tableau 6.14, par rapport aux symptômes négatifs, où les  $Z_{1.0}(\hat{\theta})$  étaient supérieurs aux  $Z_{0.5}(\hat{\theta})$ , on a enlevé un à un les individus ayant des scores supérieurs à 1.0, et refait à chaque fois l'analyse de linkage, afin de voir si le Lod-score maximum augmenterait. Pour les symptômes négatifs, on a obtenu le meilleur résultat

Famille	Modèle	Classification	Marqueur	$Z_{0.5}(\hat{\theta})$	$Z_{1.0}(\hat{\theta})$	Différence
151	dominant	na-i-a	D6S296	0.39	2.13	1.74
211	récessif	na-i-a	D6S260	0	1.28	1.28
211	récessif	na-i-a	D6S274	0	1.28	1.28
211	récessif	na-i-a	D6S334	0	1.28	1.28
211	récessif	i-a	D6S89	0	1.16	1.16
211	récessif	i-a	D6S260	0	1.16	1.16
211	récessif	i-a	D6S274	0	1.16	1.16
211	récessif	i-a	D6S334	0	1.16	1.16
255	récessif	na-i-a	D11S35	1.20	2.50	1.30
230	récessif	na-i-a	D18S65	0.01	1.14	1.13

Tableau 6.14: Différences supérieures à 1 des Lod-scores maximums entre les seuils de coupure de 0.5 et 1.0, en valeur absolue, pour les symptômes négatifs

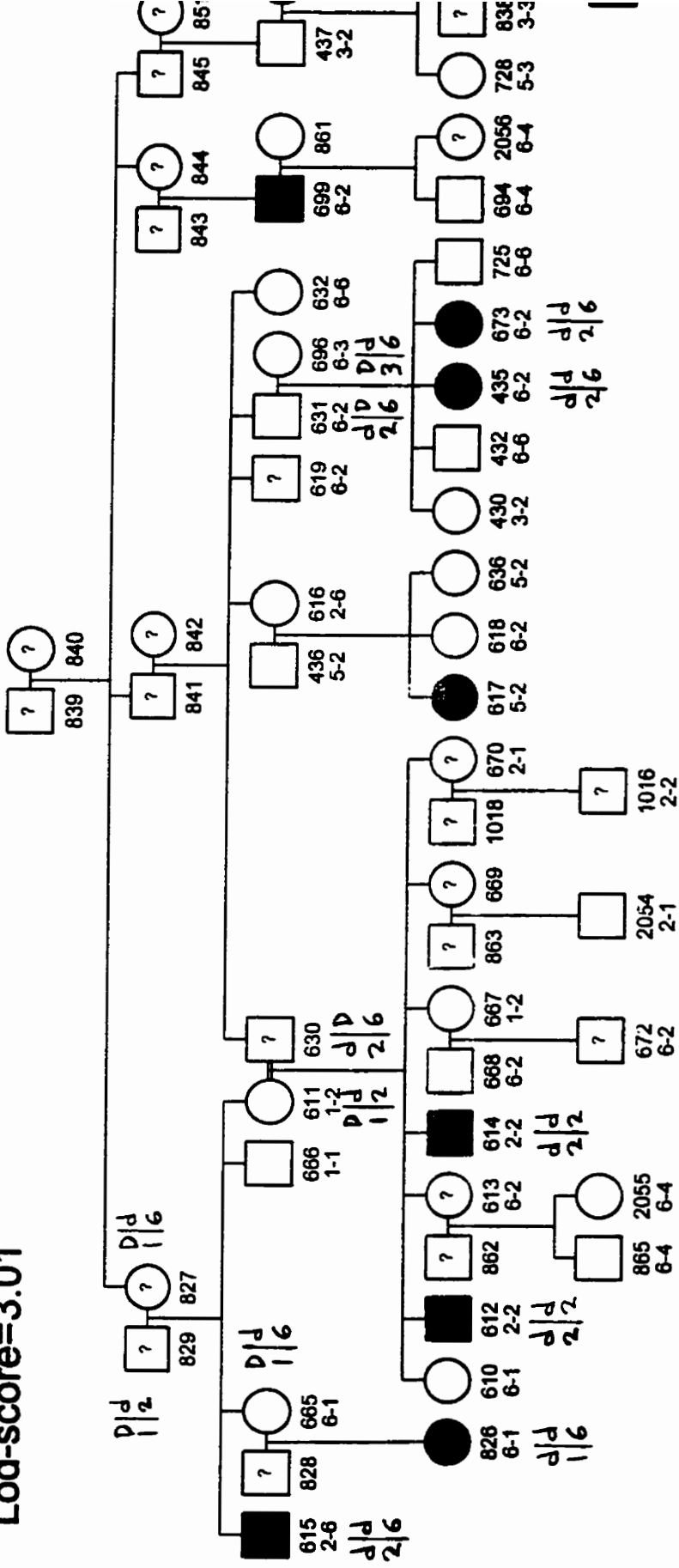
avec un Lod-score maximum de 3.01 à  $\hat{\theta} = 0$  supérieur à la valeur critique  $Z_0 = 3$ . Ce Lod-score maximum de 3.01 a été obtenu avec 9 individus de la famille 255 ayant des scores supérieurs à 1.5 avec le modèle récessif, le marqueur génétique D11S35 et la première classification (non affecté, inconnu, affecté).

À la page suivante, on retrouve le dessin de la famille 255 avec le numéro d'identification des membres de cette famille à la première ligne et le génotype au locus du marqueur D11S35 à la deuxième ligne. Les individus *non affectés* sont en blanc, les *inconnus* ont un point d'interrogation et les *affectés* sont en noir, selon un seuil de coupure de 1.5. Les 9 individus *affectés* ont les numéros d'identification suivants : 615, 699, 826, 612, 614, 617, 435, 673 et 837.

Étant donné que  $\hat{\theta} = 0$ , il n'y aurait pas d'haplotype recombinant avec l'allèle du marqueur génétique. En considérant l'hypothèse que les haplotypes sont non recombinants avec l'allèle du marqueur et que le modèle est récessif, on va déduire quelques génotypes sous-jacents aux phénotypes observés dans quelques branches.

L'individu 826 est affecté, ainsi il a le génotype *dd* au locus de la maladie, sous le modèle récessif. Son génotype est donc *d1/d6*. Par le même raisonnement, on déduit

**Pedigree 255**  
**Marqueur D11S35**  
**Lod-score=3.01**



que l'individu 615 a le génotype  $d2/d6$ . Les individus 612 et 614 ont le génotype  $d2/d2$  et les individus 435 et 673 ont le génotype  $d2/d6$ . À propos de ces 6 individus affectés, on remarque qu'il y a au moins la présence de l'allèle 2 ou 6 dans leur génotype. Cela indique que ces deux allèles seraient susceptibles de causer la maladie. L'individu 665 est non affecté; son génotype est soit  $D1/d6$  (phase I), soit  $d1/D6$  (phase II). Les individus 829 et 827 sont inconnus, cependant on sait qu'ils sont porteurs d'au moins un allèle  $d$ . Au locus de la maladie, leur génotype est soit  $Dd$ , soit  $dd$ . Considérons, par exemple, le génotype  $Dd$ . Ces individus sont porteurs de l'allèle 1 au locus du marqueur, car l'individu 666 (leur fils) a le génotype 11. En tenant compte du génotype 26 de l'individu 615 et en supposant qu'il n'y ait pas d'haplotype recombinant avec l'allèle 2 ou 6, on en déduit que le génotype de l'individu 829 serait possiblement  $D1/d2$  et que celui de l'individu 827,  $D1/d6$ . Ainsi, le génotype de l'individu 665 serait  $D1/d6$ . On peut alors penser que l'haplotype  $d-6$  de l'individu 827 a été transmis aux individus 615 et 665 et que l'individu 826 a reçu cet haplotype de l'individu 665.

Sachant que l'individu 612 a le génotype 22 au locus du marqueur, les individus 611 et 630 sont alors porteurs de l'allèle 2. En considérant le génotype 16 de l'individu 610, on en déduit que l'individu 613 a reçu l'allèle 2 de l'individu 611 (sa mère) et l'allèle 6 de l'individu 630 (son père). Par conséquent, le génotype de l'individu 611 serait  $D1/d2$  et celui de l'individu 630,  $d2/D6$ . En tenant compte toujours de l'hypothèse que l'haplotype avec l'allèle 2 ou 6 est non recombinant, on déduit le génotype  $d2/D6$  pour l'individu 631 et le génotype  $D3/d6$  pour l'individu 696.

En utilisant chacune des quatre définitions usuelles du phénotype applicable à la famille 255, soit la schizophrénie stricte et large, et le gène commun strict et large avec le modèle récessif, le marqueur D11S35 et la première classification (na-i-a), le Lod-score maximum obtenu le plus élevé de ces quatre définitions est seulement de 1.10 à  $\hat{\theta} = 0.05$  pour la schizophrénie stricte. Avec cette définition de schizophrénie stricte, il y a 10 individus affectés dans la famille 255. Les numéros d'identification de ces individus sont les suivants : 615, 630, 619, 826, 612, 613, 614, 670, 435 et 673.



Il y a une grande différence entre le Lod-score de 3.01, obtenu avec la nouvelle définition du phénotype (symptômes négatifs), et celui de 1.10, obtenu avec la définition usuelle basée sur les diagnostics du DSM-III-R (schizophrénie stricte). Cela nous indique que l'utilisation des symptômes positifs et négatifs avec un seuil de coupure pour définir le phénotype des deux psychoses majeures constituerait une alternative intéressante au lieu de considérer ces deux psychoses majeures comme un ensemble de symptômes hétérogènes. Ainsi, pour les futures analyses de linkage, il sera intéressant de considérer les symptômes négatifs avec un seuil de coupure de 1.5 et d'utiliser d'autres marqueurs génétiques.

## 6.6 Résultats du test d'hétérogénéité pour un mélange de familles indissociables

À la section 2.4.4, l'hétérogénéité due à un mélange de familles indissociables a été décrite. Pour les analyses de cette hétérogénéité, on a utilisé le programme informatique HOMOG. On a exécuté ce programme lorsque la somme des Lod-scores maximums positifs sur l'ensemble des familles pour une fraction de recombinaison fixée était supérieure à 3. Du point de vue pratique, le choix de ce critère nous évitait de faire toutes les analyses d'hétérogénéité dont certaines d'entre elles n'étaient pas nécessaires.

Rappelons que le test d'hypothèses pour l'hétérogénéité est le suivant :

$$H_1 : \alpha = 1, \theta_r < 1/2 \text{ (homogénéité entre les familles)}$$

$$H_2 : 0 < \alpha < 1, \theta_1 < 1/2 \text{ (hétérogénéité génétique entre les familles).}$$

Rappelons que la fonction de vraisemblance sous un modèle d'hétérogénéité s'écrit de la façon suivante :

$$L_{hété.} = \prod_{i=1}^n L_i(\alpha, \theta_1) = \prod_{i=1}^n [\alpha L_i(\theta_1) + (1 - \alpha) L_i(1/2)]$$

où

Symp.	Modèle	Clas.	Marqueur	H <sub>1</sub>	H <sub>2</sub>		χ <sup>2</sup>		Lod-score	
				$\hat{\theta}_r$	$\hat{\alpha}$	$\hat{\theta}_1$	hété.	total	homo.	hété.
pos.	dominant	na-i-a	D11S35	0.50	0.20	0.00	6.60	6.60	0.00	1.48
pos.	dominant	i-a	D11S35	0.30	0.25	0.05	2.34	2.99	0.14	0.68
pos.	récessif	na-i-a	D11S35	0.30	0.10	0.00	2.90	5.11	0.48	1.11
pos.	récessif	i-a	D11S35	0.20	0.30	0.00	0.40	4.64	0.92	1.00
nég.	dominant	na-i-a	D11S35	0.25	0.20	0.00	0.50	5.57	1.10	1.21
nég.	récessif	i-a	D11S35	0.20	1.00	0.20	0.00	4.65	1.01	1.00
pos.	récessif	na-i-a	D18S456	0.20	0.65	0.15	0.19	4.66	0.97	1.00
pos.	récessif	i-a	D18S456	0.15	0.65	0.10	0.18	3.55	0.73	0.77
nég.	récessif	na-i-a	D18S456	0.25	0.55	0.15	0.37	4.15	0.82	0.90
pos.	récessif	i-a	D18S472	0.15	1.00	0.15	0.00	7.28	1.58	1.58

Tableau 6.15: Résultats du test d'hétérogénéité pour un mélange de familles indissociables au seuil de coupure de 0.5

- $n$  est le nombre de familles participant à l'étude;
- $\alpha$  est la proportion des familles avec linkage;
- $L_i(\theta_1)$  est la vraisemblance de la famille  $i$  à une fraction de recombinaison de  $\theta_1$ ;
- $L_i(1/2)$  est la vraisemblance des familles avec absence de linkage.

Les résultats du test d'hétérogénéité pour le seuil de coupure de 0.5 se retrouvent au tableau 6.15, tandis que ceux pour le seuil de coupure de 1.0, au tableau 6.16. Pour chaque résultat, on a la catégorie de symptômes, le type de modèle, la classification ainsi que le marqueur génétique.

Lorsqu'on envisage l'hétérogénéité génétique, on doit alors se poser la question suivante : est-il nécessaire de complexifier le modèle de transmission génétique posé en lui ajoutant le paramètre  $\alpha$ ?

Illustrons à l'aide d'un exemple le calcul du Lod-score sous hétérogénéité. On va considérer le cas, au tableau 6.16, avec les symptômes négatifs, le modèle dominant, la première classification (na-i-a), le marqueur D11S35 et le seuil de coupure de 1.0. Le

Symp.	Modèle	Clas.	Marqueur	H <sub>1</sub>	H <sub>2</sub>		χ <sup>2</sup>		Lod-score	
				$\hat{\theta}_r$	$\hat{\alpha}$	$\hat{\theta}_1$	hété.	total	homo.	hé
nég.	dominant	i-a	D6S89	0.25	0.30	0.00	1.28	3.13	0.40	0.0
nég.	récessif	na-i-a	D6S89	0.50	0.25	0.00	2.39	2.39	0.00	0.0
nég.	récessif	i-a	D6S89	0.15	0.55	0.05	0.31	4.87	0.99	1.0
nég.	dominant	i-a	D6S260	0.25	0.40	0.00	1.48	3.88	0.52	0.0
nég.	récessif	na-i-a	D6S260	0.40	0.25	0.00	3.39	3.39	0.00	0.0
nég.	récessif	i-a	D6S260	0.10	0.95	0.10	0.02	7.58	1.64	1.0
nég.	dominant	i-a	D6S274	0.15	0.45	0.00	2.06	7.17	1.11	1.0
nég.	récessif	na-i-a	D6S274	0.50	0.20	0.00	2.50	2.50	0.00	0.0
nég.	dominant	na-i-a	D6S296	0.30	0.20	0.00	4.09	5.61	0.33	1.0
nég.	récessif	na-i-a	D6S296	0.45	0.15	0.00	2.93	2.93	0.00	0.0
nég.	dominant	i-a	D6S334	0.20	0.50	0.00	3.63	6.62	0.65	1.0
pos.	récessif	na-i-a	D11S35	0.20	0.80	0.15	0.42	6.96	1.42	1.0
pos.	récessif	i-a	D11S35	0.10	0.75	0.05	0.79	9.04	1.79	1.0
nég.	dominant	na-i-a	D11S35	0.25	0.30	0.05	0.35	5.28	1.07	1.0
nég.	récessif	na-i-a	D11S35	0.30	0.10	0.00	3.70	4.90	0.26	1.0
pos.	dominant	na-i-a	D11S876	0.50	0.20	0.00	3.19	3.19	0.00	0.0
pos.	récessif	na-i-a	D11S876	0.20	1.00	0.20	0.00	7.74	1.68	1.0
pos.	récessif	i-a	D11S876	0.15	0.80	0.10	0.24	6.74	1.41	1.0
nég.	récessif	na-i-a	D18S472	0.15	1.00	0.15	0.00	8.93	1.94	1.0

Tableau 6.16: Résultats du test d'hétérogénéité pour un mélange de familles indissociables au seuil de coupure de 1.0

Hypothèse	$\log_e L(\hat{\alpha}, \hat{\theta})$	$\hat{\alpha}$	$\hat{\theta}$
$H_2$ : Linkage, hétérogénéité	2.6392	0.30	0.05
$H_1$ : Linkage, homogénéité	2.4638	(1)	0.25
$H_0$ : Absence de linkage	(0)	(0)	(0.50)

Tableau 6.17: Log-vraisemblance bivariée globale maximum avec les estimations des paramètres  $\alpha$  et  $\theta$  sous différentes hypothèses de l'exemple de la section 6.6

le tableau 6.17 représente la log-vraisemblance bivariée globale maximum avec les estimations des paramètres  $\alpha$  et  $\theta$  sous différentes hypothèses de cet exemple, résultats obtenus du programme HOMOG. Les valeurs entre parenthèses de ce tableau sont fixées. Pour tester l'hypothèse  $H_1$  contre l'hypothèse  $H_2$ , la valeur de la statistique utilisée est la suivante :

$$\begin{aligned} \lambda^2 &= 2 [\log_e L(\hat{\alpha}, \hat{\theta}_1) - \log_e L(1, \hat{\theta}_r)] \\ &= 2 [\log_e L(0.30, 0.05) - \log_e L(1, 0.25)] \\ &= 2 [2.6392 - 2.4638] = 0.35. \end{aligned}$$

Le seuil observé est alors d'environ 0.28 (test unilatéral). Il est possible de transformer cette valeur du  $\chi^2$  en un Lod-score. La méthode de calcul est la suivante :

$$\log_e L_{H_2}(0.30, 0.05) - \log_e L_{H_0}(0.50) = \log_e \left[ \frac{L_{H_2}(0.30, 0.05)}{L_{H_0}(0.50)} \right] = 2.6392$$

en utilisant l'identité qui suit :

$$\log_{10}(X) = \frac{\log_e(X)}{\log_e(10)}$$

on obtient donc,

$$Z_{\text{hété.}} = \log_{10} \left[ \frac{L_{H_2}(0.30, 0.05)}{L_{H_0}(0.50)} \right] = \frac{2.6392}{\log_e(10)} = 1.15.$$

Le calcul du Lod-score sous homogénéité s'effectue de façon similaire. On remarque qu'aucun Lod-score sous hétérogénéité des tableaux 6.15 et 6.16 n'excède la valeur critique  $Z_0 = 3$ . Ainsi, aucun résultat significatif n'a été obtenu pour le test d'hétérogénéité pour un mélange de familles indissociables.

# Conclusion

Pour conclure, il faut se rappeler que la schizophrénie et le trouble bipolaire sont des maladies complexes, car la pénétrance est dépendante de l'âge, il y a présence de phénotypes, il y aurait de l'hétérogénéité génétique. On note aussi que le modèle de transmission génétique diffère de la transmission mendélienne classique récessive ou dominante due à un seul gène et ce modèle n'est pas bien connu. De plus, la définition du phénotype de ces deux psychoses majeures constitue l'une des principales difficultés de la recherche psychiatrique. Il faudrait non seulement considérer les facteurs génétiques lors des analyses de linkage mais aussi les facteurs environnementaux.

Tous ces éléments rendent ainsi l'analyse de linkage plus complexe pour ces deux maladies psychiatriques, ce qui explique bien la difficulté de détecter une évidence définitive de linkage dans les études réalisées jusqu'à maintenant. Dans ce mémoire, les analyses de linkage ont été effectuées avec seulement 16 marqueurs génétiques répartis sur les chromosomes 6, 11 et 13 en considérant les symptômes positifs et négatifs comme phénotypes, c'est-à-dire comme expressions des gènes recherchés. Plusieurs marqueurs génétiques n'ont pas été utilisés et un nombre considérable de marqueurs n'ont pas encore été identifiés. Il est alors évident que les travaux de recherche ayant pour but de localiser les gènes susceptibles de causer la schizophrénie et le trouble bipolaire ne font que commencer.

# Bibliographie

- Abe, K., Oda, N., Ikenaga, K. & Yamada, T. (1993). Twin study on night terrors, fears and some physiological and behavioural characteristics in childhood, *Psychiatric Genetics* **3**: 39–43.
- Andreasen, N. C. (1983). *The Scale for the Assessment of Negative Symptoms (SANS)*, University of Iowa.
- Andreasen, N. C. (1984). *The Scale for the Assessment of Positive Symptoms (SAPS)*, University of Iowa.
- Andreasen, N. C. (1987). L'utilisation de la SANS et de la SAPS, *Psychiatrie et Psychobiologie* **2**(6): 412–437. Traduction française : Y. Lecrubier et P. Boyer.
- APA (1987). *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)*. APA : American Psychiatric Association.
- Arndt, S., Alliger, R. J. & Andreasen, N. C. (1991). The distinction of positive and negative symptoms : the failure of a two-dimensional model, *British Journal of Psychiatry* **158**: 317–322.
- Burke, J. G., Murphy, B. M., Bray, J. C., Walsh, D. & Kendler, K. S. (1996). Clinical similarities in siblings with schizophrenia, *American Journal of Medical Genetics* **67**: 239–243.
- Casella, G. & Berger, R. L. (1990). *Statistical Inference*, Duxbury Press.
- Cloninger, C. R. (1994). Turning point in the design of linkage studies of schizophrenia, *American Journal of Medical Genetics* **54**: 83–92.

- Crow, T. J. (1994). Aetiology of schizophrenia, *Current Opinion in Psychiatry* **7**: 39–42.
- Darnell, J. E., Lodish, H. F. & Baltimore, D. (1988). *La cellule : biologie moléculaire*, Éditions Vigot.
- Davison, M. L. (1985). Multidimensional scaling versus components analysis of test intercorrelations, *Psychological Bulletin* **97**: 94–105.
- DeLisi, L. E., Goldin, L. R., Maxwell, E., Kazuba, D. M. & Gershon, E. S. (1987). Clinical features of illness in siblings with schizophrenia or schizoaffective disorder, *Archives of General Psychiatry* **44**: 891–896.
- Dunteman, G. H. (1989). *Principal Components Analysis : Quantitative Applications in the Social Sciences*, 07–069, Sage.
- Elston, R. C. (1981). *Advances in Human Genetics*, Vol. 11, Plenum, chapter 2, pp. 63–120.
- Elston, R. C., George, V. T. & Severtson, F. (1992). The Elston-Stewart algorithm for continuous genotypes and environmental factors, *Human Heredity* **42**: 16–27.
- Elston, R. C. & Stewart, J. (1971). A general model for the genetic analysis of pedigrees, *Human Heredity* **21**: 523–542.
- Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*, John Wiley.
- Goldstein, M. & Dillon, W. R. (1978). *Discrete Discriminant Analysis*, John Wiley.
- Gottesman, I. I. (1991). *Schizophrenia Genesis : The Origins of Madness*, W. H. Freeman and Company.
- Gottesman, I. I. & Bertelsen, A. (1989). Confirming unexpressed genotypes for schizophrenia : risks in the offspring of Fisher's Danish identical and fraternal discordant twins, *Archives of General Psychiatry* **46**: 867–872.
- Gottesman, I. I. & Shields, J. (1972). *Schizophrenia and Genetics : A Twin Study Vantage Point*, Academic Press.



- Greenberg, D. A., Hodge, S. E., Vieland, V. J. & Spence, M. A. (1996). Affecteds-only linkage methods are not a panacea, *American Journal of Human Genetics* **58**: 892-895.
- Haldane, J. B. S. & Smith, C. A. B. (1947). A new estimate of the linkage between the genes for colour-blindness and haemophilia in man, *Annals of Eugenics* **14**: 10-31.
- Johnson, N. L. & Kotz, S. (1970). *Continuous Univariate Distributions*, Vol. 2, Houghton Mifflin.
- Kendler, K. S. & Diehi, S. R. (1993). The genetics of schizophrenia : a current, genetic-epidemiologic perspective, *Schizophrenia Bulletin* **19**: 261-285.
- Khoury, M. J., Beaty, T. H. & Cohen, B. H. (1993). *Fundamentals of Genetic Epidemiology*, Oxford University Press.
- Klimidis, S., Stuart, G. W., Minas, I. H., Copolov, D. L. & Singh, B. S. (1993). Positive and negative symptoms in the psychoses : re-analysis of published SAPS and SANS global ratings, *Schizophrenia Research* **9**: 11-18.
- Lalonde, Grunberg *et al.* (1988). *Psychiatrie clinique : approche bio-psycho-sociale*, Gaëtan Morin éditeur ltée.
- Lander, E. S. (1988). Splitting schizophrenia, *Nature* **336**: 105-106.
- Lander, E. S. & Schork, N. J. (1994). Genetic dissection of complex traits, *Science* **265**: 2037-2048.
- Lange, K. & Elston, R. C. (1975). Extensions to pedigree analysis. I. Likelihood calculations for simple and complex pedigrees, *Human Heredity* **25**: 95-105.
- La schizophrénie : Guide à l'intention des familles* (n.d.). Publication autorisée par le ministre de la Santé nationale et du Bien-être social.
- Lebart, L., Morineau, A. & Warwick, K. M. (1984). *Multivariate Descriptive Statistical Analysis*, John Wiley.

- Levinson, D. F., Mowry, B. J., Sharpe, L. & Endicott, J. (1996). Penetrance of schizophrenia-related disorders in multiplex families after correction for ascertainment, *Genetic Epidemiology* **13**: 11–21.
- Liddle, P. F. (1987). The symptoms of chronic schizophrenia : a re-examination of the positive-negative dichotomy, *British Journal of Psychiatry* **151**: 145–151.
- Liddle, P. F., Barnes, T. R. E., Speller, J. & Kibel, D. (1993). Negative symptoms as a risk factor for tardive dyskinesia in schizophrenia, *British Journal of Psychiatry* **163**: 776–780.
- Long, J. S. (1983). *Confirmatory Factor Analysis : Quantitative Applications in the Social Sciences*, 07–033, Sage.
- Mange, E. J. & Mange, A. P. (1994). *Basic Human Genetics*, Sinauer Associates, Inc., Publishers.
- Maziade, M., Roy, M.-A., Martinez, M., Cliche, D., Fournier, J.-P., Garneau, Y., Nicole, L., Montgrain, N., Dion, C., Ponton, A.-M., Potvin, A., Lavallée, J.-C., Pirès, A., Bouchard, S., Boutin, P., Brisebois, F. & Mérette, C. (1995). Negative, psychoticism and disorganized dimensions in patients with familial schizophrenia or bipolar disorder : continuity and discontinuity between the major psychoses, *American Journal of Psychiatry* **152**: 1458–1463.
- McGue, M. & Gottesman, I. I. (1989). A single dominant gene still cannot account for the transmission of schizophrenia, *Archives of General Psychiatry* **46**: 478–479.
- McGuffin, P., Owen, M. J. & Farmer, A. E. (1995). Genetic basis of schizophrenia, *Lancet* **346**(september 9): 678–682.
- Minas, I. H., Stuart, G. W., Klimidis, S., Jackson, H. J., Singh, B. S. & Copolov, D. L. (1992). Positive and negative symptoms in the psychoses : multidimensional scaling of SAPS and SANS items, *Schizophrenia Research* **8**: 143–156.

- Morton, N. E. (1955). Sequential tests for the detection of linkage, *American Journal of Human Genetics* **7**: 277-318.
- Morton, N. E. (1978). Analysis of crossingover in man, *Cytogenetics and Cell Genetics* **22**: 15-36.
- Ott, J. (1974). Estimation of the recombination fraction in human pedigrees : efficient computation of the likelihood for human linkage studies, *American Journal of Human Genetics* **26**(5): 588-597.
- Ott, J. (1991). *Analysis of Human Genetic Linkage*, The Johns Hopkins University Press.
- Pagé, V. (1993). *L'analyse de linkage : fondements et applications*, Mémoire de maîtrise, Université Laval.
- Peralta, V. & Cuesta, M. J. (1995). Negative symptoms in schizophrenia : a confirmatory factor analysis of competing models, *American Journal of Psychiatry* **152**: 1450-1457.
- Peralta, V., de Leon, J. & Cuesta, M. J. (1992). Are there more than two syndromes in schizophrenia ? A critique of the positive-negative dichotomy, *British Journal of Psychiatry* **161**: 335-343.
- Persico, A. M., Wang, Z. W., Black, D. W., Andreasen, N. C., Uhl, G. R. & Crowe, R. R. (1995). Exclusion of close linkage of the dopamine transporter gene with schizophrenia spectrum disorders, *American Journal of Psychiatry* **152**: 134-136.
- Risch, N. (1990). Linkage strategies for genetically complex traits. I. Multilocus models, *American Journal of Human Genetics* **46**: 222-228.
- Roy, M.-A. & DeVriendt, X. (1994). Symptômes positifs et négatifs de la schizophrénie : une mise à jour, *Revue Canadienne de Psychiatrie* **39**(7): 407-414.




















- Schexnayder, L. W., Hirschowitz, J., Sautter, F. J. & Garver, D. L. (1995). Predictors of response to lithium in patients with psychoses, *American Journal of Psychiatry* **152**: 1511–1513.
- Sham, P. C., Zerbin-Rüdin, E. & Kendler, K. S. (1995). A review and reanalysis of Bruno Schulz's, *Erkrankungsalter schizophrener Eltern und Kinder* [Age at onset of illness in schizophrenic parents and offspring] : *Zeitschrift für die gesamte Neurologie und Psychiatrie* , 168, 709–721, 1940, *Psychiatric Genetics* **5**: 57–62.
- Smith, C. A. B. (1953). The detection of linkage in human genetics, *Journal of the Royal Statistical Society* **14**: 153–192.
- Smith, C. A. B. (1959). Some comments on the statistical methods used in linkage investigations, *American Journal of Human Genetics* **11**: 289–304.
- Smith, C. A. B. (1963). Testing for heterogeneity of recombination fraction values in human genetics, *Annals of Human Genetics* **27**: 175–182.
- Stewart, J. (1992). Genetics and biology : a comment on the significance of the Elston-Stewart algorithm, *Human Heredity* **42**: 9–15.
- Terwilliger, J. D. & Ott, J. (1994). *Handbook of Human Genetic Linkage*, The Johns Hopkins University Press.
- Tsuang, M. T. (1993). Genotypes, phenotypes, and the brain : a search for connections in schizophrenia, *British Journal of Psychiatry* **163**: 299–307.
- Tsuang, M. T., Winokur, G. & Crowe, R. R. (1980). Morbidity risks of schizophrenia and affective disorders among first degree relatives of patients with schizophrenia, mania, depression and surgical conditions, *British Journal of Psychiatry* **137**: 497–504.
- Weeks, D. E., Brzustowicz, L., Squires-Wheeler, E., Cornblatt, B., Lehner, T., Stephanovich, M., Bassett, A., Gilliam, T. C., Ott, J. & Erlenmeyer-Kimling,

L. (1990). Report of a workshop on genetic linkage studies in schizophrenia, *Schizophrenia Bulletin* **16**: 673–685.

## **Annexe A**

**Exemples des trois types de familles des études pratiques avec la légende des diagnostics**

# Légende des diagnostics

	Schizophrenia		Bipolar type I
	Af Schizoaffective		Pc Bipolar type I, mood congruent psychosis
	Fo Schizophreniform		Nc Bipolar type I, mood incongruent psychosis
	Tp Schizotypal		Bipolar type II
	BR Breif reactional psychosis		Pc Bipolar type II, mood congruent psychosis
	DL Délirant		Nc Bipolar type II, mood incongruent psychosis
	Ps Non-specific psychosis		Recurrent major depression
			Pc Recurrent major depression, mood congruent psychosis
			Nc Recurrent major depression, mood incongruent psychosis
			Major depression
			Pc Major depression, mood congruent psychosis
			Nc Major depression, mood incongruent psychosis

- OT Thymic organic disorder
- OM Mental organic disorder
- Pn Panic disorder
- Ax Anxiety
- oc Obsessive-compulsive
- Dy Dysthymia
- xx Other diagnosis
- AA Alcoholism



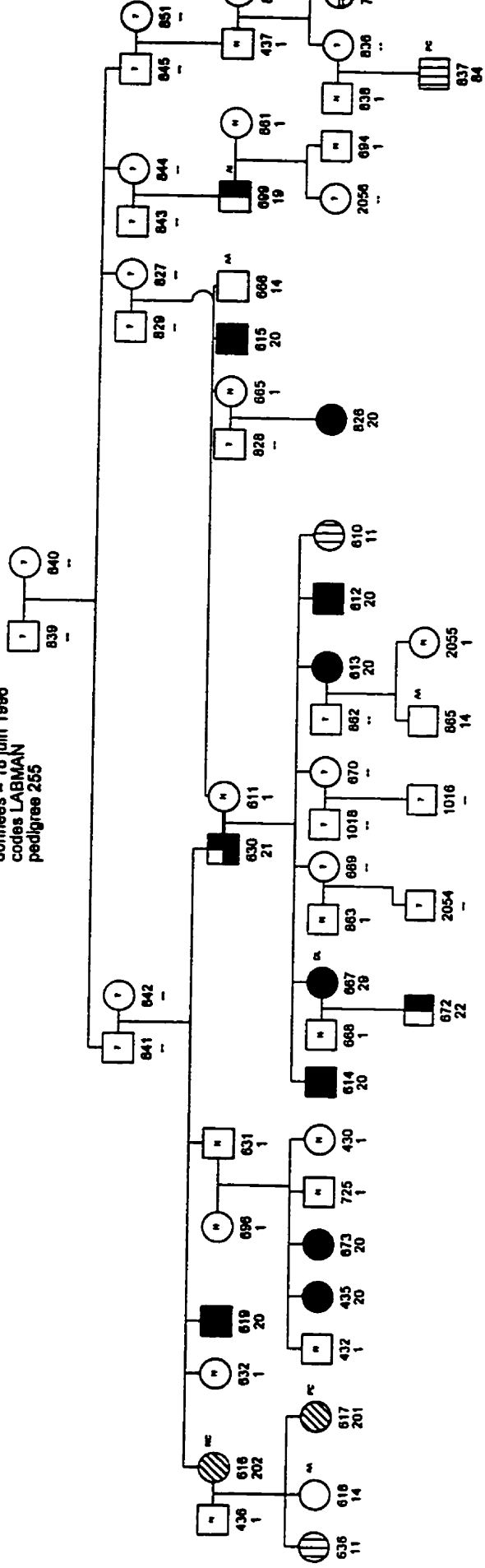
Unaffected



Unknown

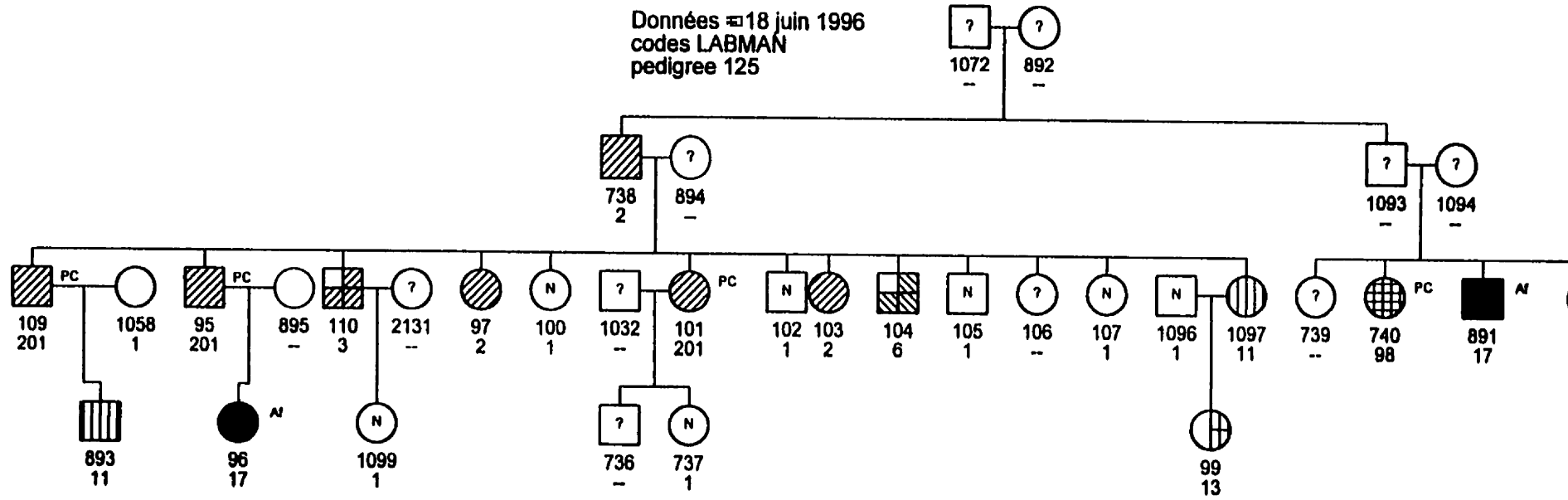


données = 18 juin 1986  
codes LABMAN  
pedigree 255



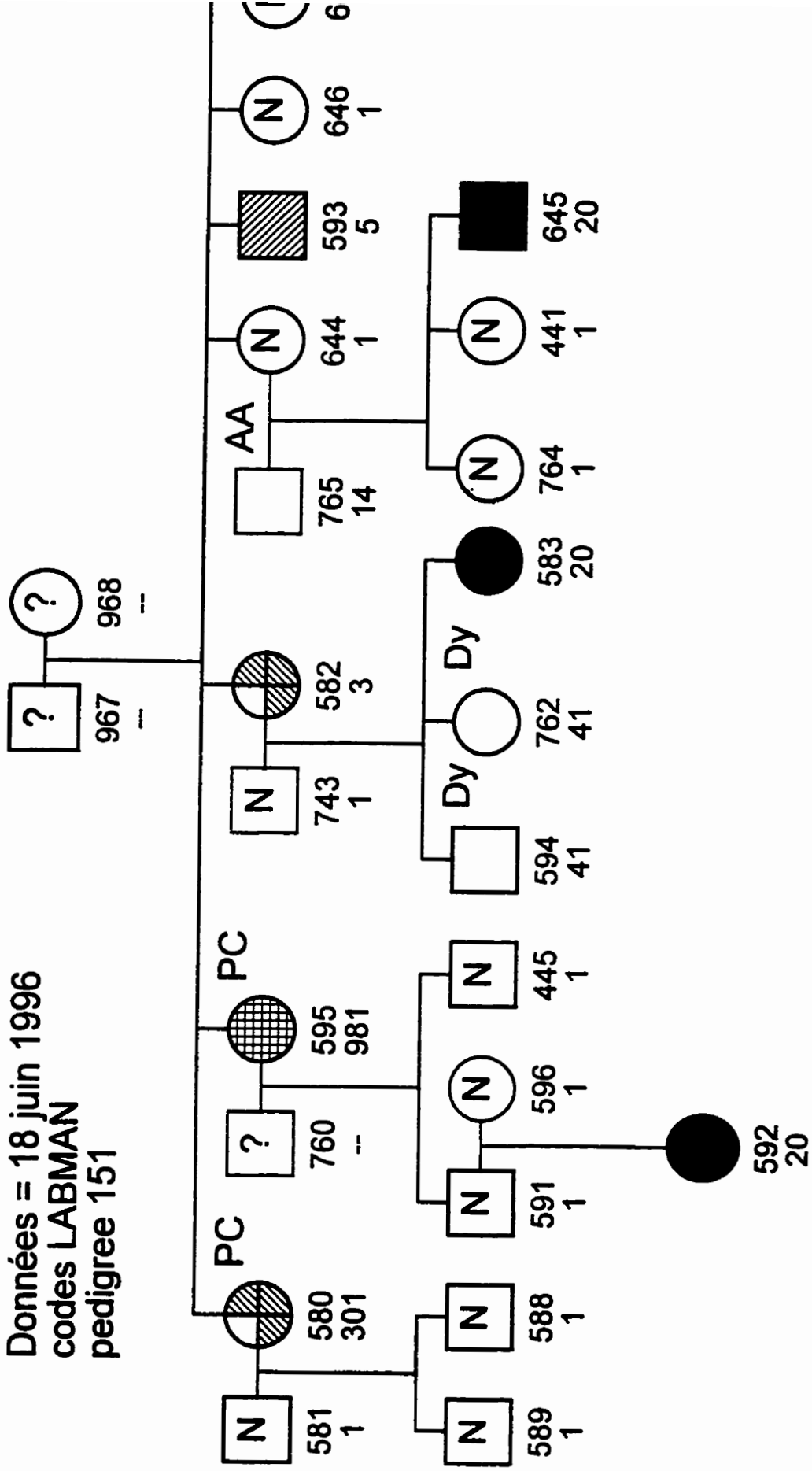
FAMILLE 155 DE TYPE SZ

Données 18 juin 1996  
codes LABMAN  
pedigree 125



FAMILLE 125 DE TYPE BP

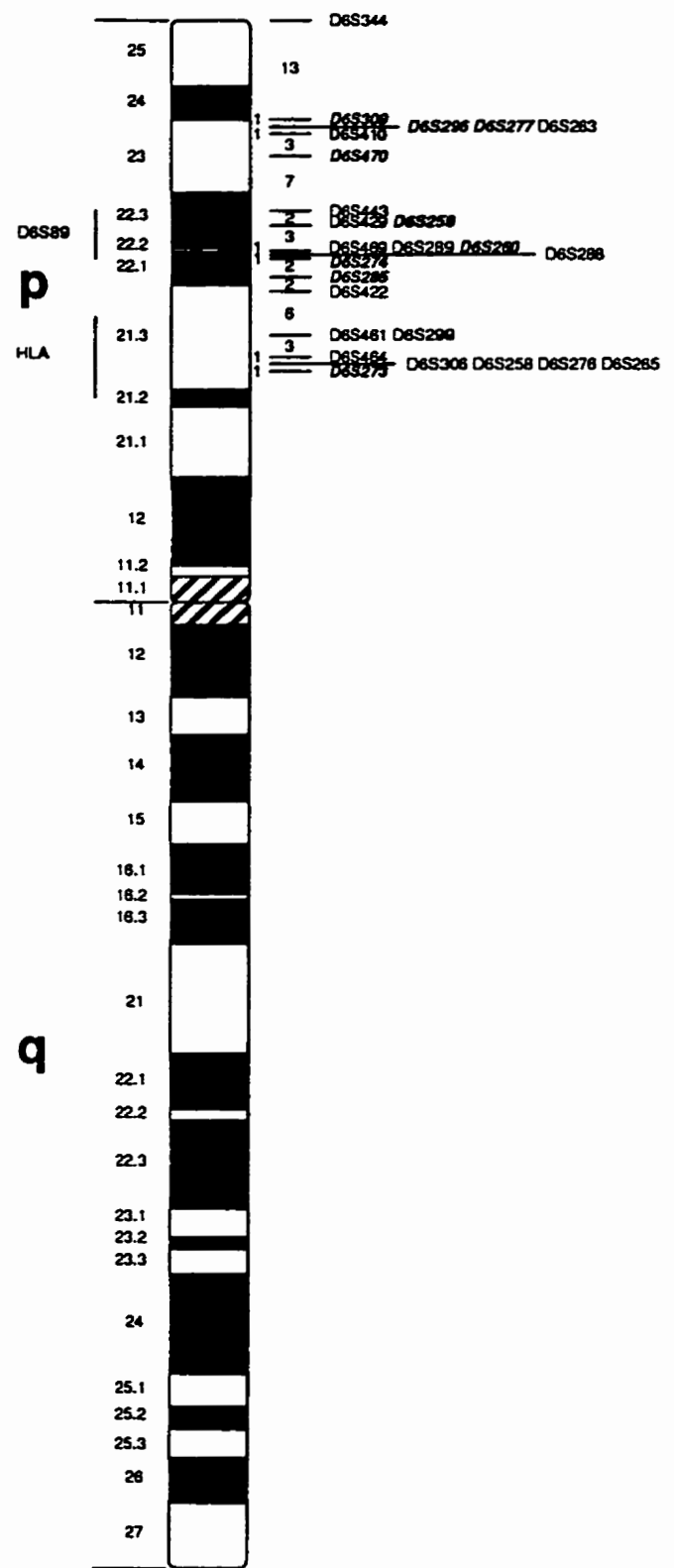
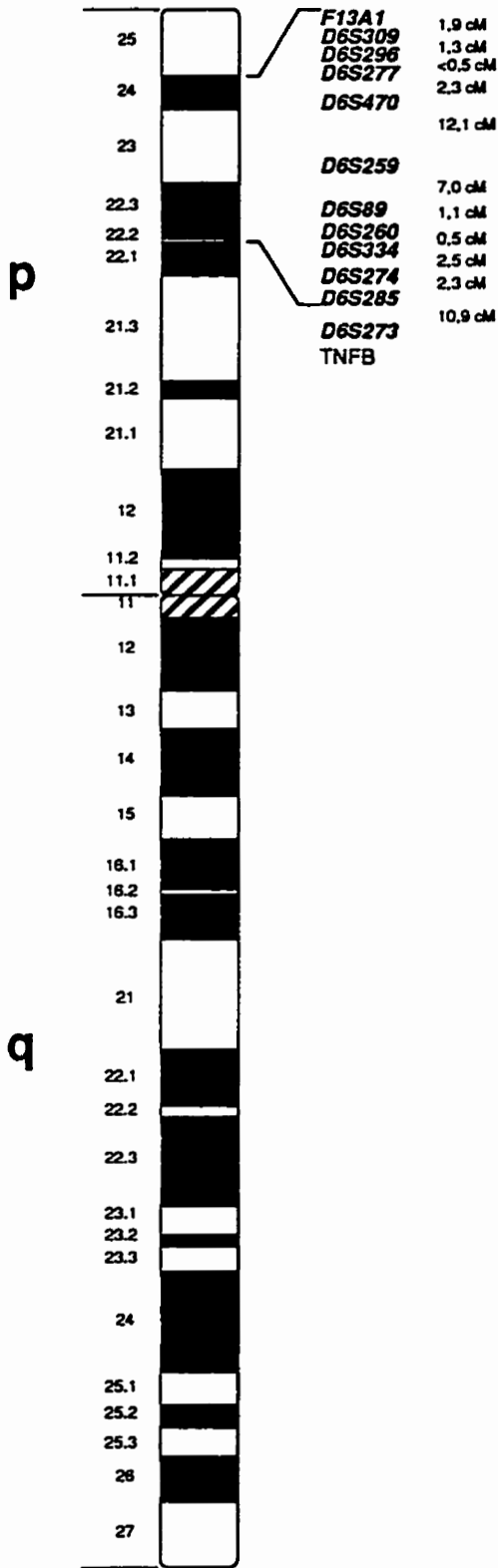
Données = 18 juin 1996  
 codes LABMAN  
 pedigree 151



FAMILLE 151 DE TYPE MIXTE

# **Annexe B**

## **Cartes génétiques des chromosomes 6, 11 et 18**

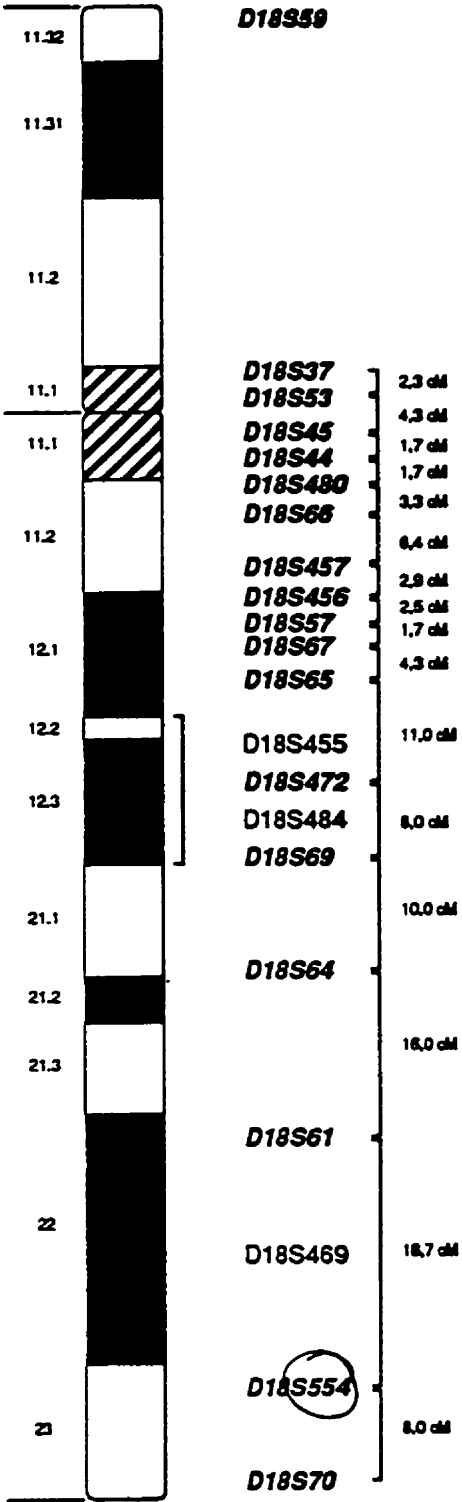


**Légende:**

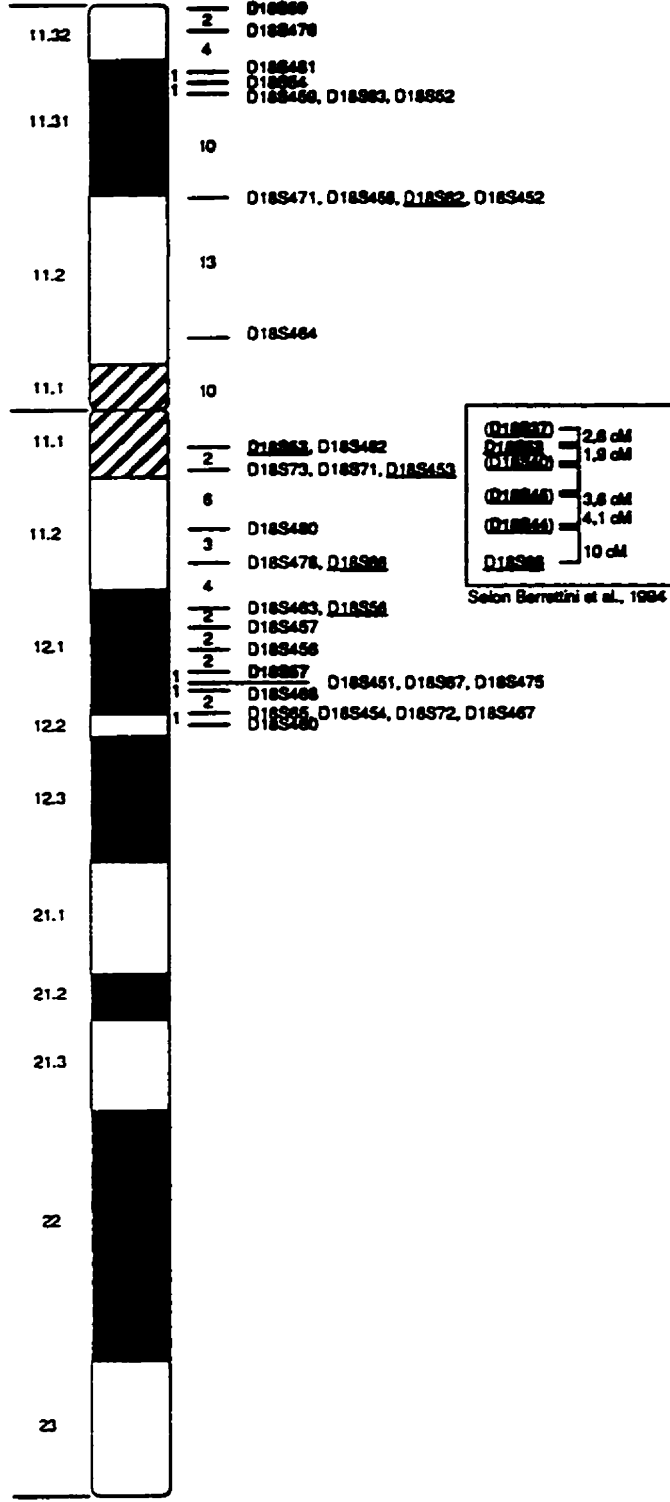
**Marqueur complété**  
**Marqueur en progression**  
Marqueur référence  
 Marqueur disponible



**p**



**q**

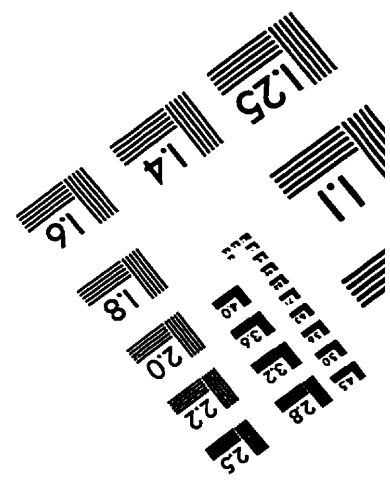
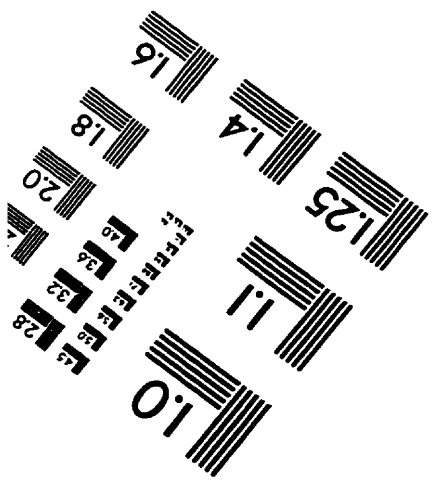
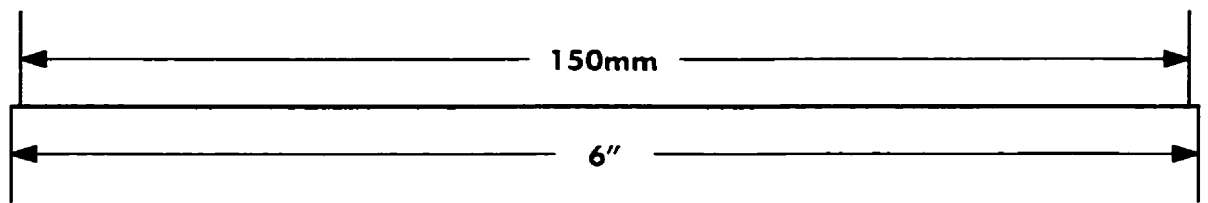
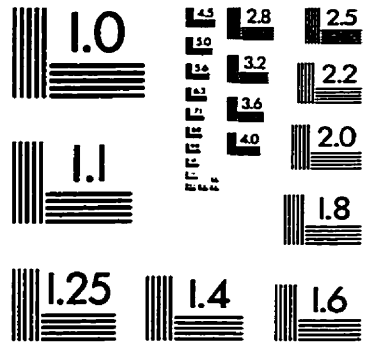
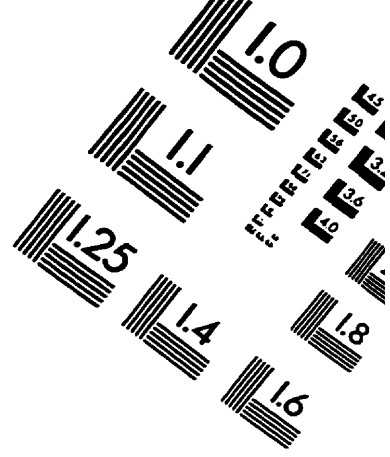
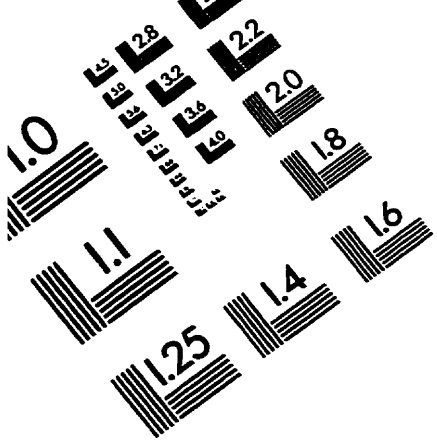


**D18S37** 2,8 cM  
**D18S53** 1,8 cM  
**D18S45** 3,6 cM  
**D18S44** 4,1 cM  
**D18S66** 10 cM  
Selon Berrettini et al., 1994

**Légende:**  
**Marqueur complété**  
**Marqueur en progression**  
**Marqueur référence**  
**Marqueur disponible**

**Légende:**  
**Marqueur de l'étude**  
**Marqueurs Berrettini et al., 1994**  
 (Marqueurs non-positionnés par Gyapay et al., 1994)

# TEST TARGET (QA-5)



**APPLIED IMAGE . Inc**  
1653 East Main Street  
Rochester, NY 14609 USA  
Phone: 716/482-0300  
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved