

# Etudes Génétiques chez l'homme

- Mettre en évidence des facteurs de susceptibilité génétique impliqués dans le déterminisme du trait étudié
  - MALADIE (caractère dichotomique: atteint/sain)
  - TRAIT QUANTITATIF (tension artérielle, taux plasmatique d'une protéine, BMI..)
- Estimer les effets
  - des facteurs génétiques
  - des autres facteurs de risque et leurs interactions avec les facteurs génétiques

**Pathologies concernées:** maladies multifactorielles communes dans la population (cancers, maladies cardio-vasculaires, métaboliques, infectieuses, maladies psychiatriques)

# Composante Génétique D'un Trait: **Étapes**

## 1. **Le trait est-il familial (corrélations familiales)?**

Études du risque pour la maladie chez les apparentés de malades

## 2. **Causes des ressemblances familiales?**

- Exposition environnementale commune aux différents membres de la famille ?
- Une transmission de type culturelle des facteurs de risque?
- Une susceptibilité génétique?

## 3) **Mode de transmission** de cette susceptibilité génétique?

Rôle du(des) gène(s) dans la variabilité du trait (développement de la maladie)

# Le trait est-il familial ?

## 1- Risque familial

**risque associé à une histoire familiale positive de la maladie**

### Etudes épidémiologiques en population

	histoire familiale	
	+	-
cas		
contrôles		

	nbr d'apparentés	
	atteints	non-atteints
cas		
contrôles		

# 1- Risque familial

## risque associé à une histoire familiale positive de la maladie

1. Fréquence de cas vs. témoins ayant une Histoire Familiale “positive” ( $\geq 1$  apparenté atteint)

	HF(+)	HF(-)
Cas	$n_{11}$	$n_{12}$
Témoins	$n_{21}$	$n_{22}$

**Odds Ratio (OR)** =  $n_{11}n_{22}/n_{12}n_{21}$

2. Fréquence d'apparentés atteints chez les cas vs. témoins :

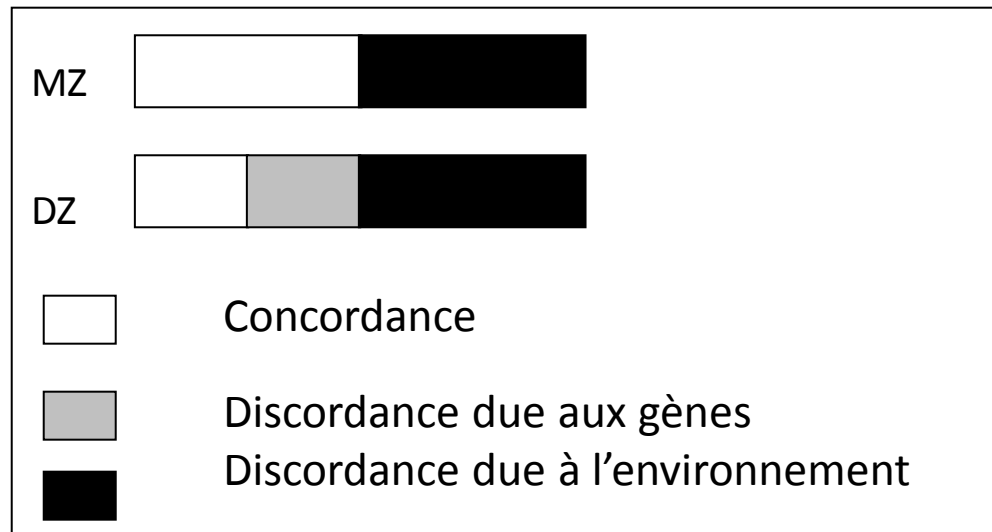
	Nbre d'apparentés		
	A	NA	Total
Cas	$n_{11}$	$n_{12}$	$n_{1.}$
Témoins	$n_{21}$	$n_{22}$	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n_{..}$

**Risque Relatif (RR)** =  $(n_{11}/n_{1.})/(n_{12}/n_{2.})$

# Le trait est-il familial ?

## 2. Méthode des jumeaux

Discerner l'origine génétique ou environnementale de la ressemblance entre apparentés.



Taux de concordance = Prob. (2ème jumeau = atteint | 1er = atteint)

Taux de concordance chez les jumeaux monozygotes (MZ) vs. dizygotes (DZ)

# Le trait est-il familial ?

## 3. CONCENTRATIONS FAMILIALES

Mesure de la ressemblance (variabilité) pour le trait chez des sujets apparentés vs. non-apparentés

Risque pour un apparenté d'un sujet atteint > Risque pour un individu de la population générale

$$\lambda_s = \frac{\text{Risque chez un germain d'un patient}}{\text{Risque chez un individu de la population générale}}$$

Ex: Risques attendus chez les apparentés d'un sujet atteint sous ***l'hypothèse d'une susceptibilité génétique transmise selon un mode dominant:***

- Apparentés du 1er degré: 50%
- Apparentés du 2nd degré: 25%

# RESSEMBLANCES FAMILIALES: Exemples

## Maladies cardio-vasculaires

Taux de concordance chez des jumeaux

	<u>MZ</u>	<u>DZ1</u>	<u>DZ2</u>
Infarctus du myocarde	19.6%	15.5%	7.5%
Hypertension	25%	9.4%	3.8%

DZ1, DZ2 : jumeaux dizygotes de même sexe, de sexe différent

## Schizophrénie

Taux de concordance:      • MZ = 45%(261) • DZ= 13%(329)

Freq. SZ chez les enfants adoptés de :

- parents biologiques SZ =12.5%
- parents biologiques indemnes = 2.9%

# Ajustement pour des facteurs de l'environnement

**Ex: Proportion de décès par cancer du poumon chez les apparentés de cas et témoins selon les habitudes tabagiques**

	Apparentés	
	Fumeurs	Non fumeurs
contrôles	0,28%	1,5 % (RR = 5,3)
cas	1,1 % (RR = 4,0)	3,8 % (RR = 13,6)



## RESSEMBLANCES FAMILIALES: **Exemples**

$\lambda_s = 100$  à 500 trait mendélien

$\lambda_s = 10$  à 20 (IDDM)

$\lambda_s = 8.5$  (Schizophrénie)

$\lambda_s = 3.5$  (NIDDM)

$\lambda_s = 3$  (Hypertension)

$\lambda_s \cong 2$  (Maladie coronarienne)

# Risque familial – Analyse de survie

## Ex: Maladie de Parkinson

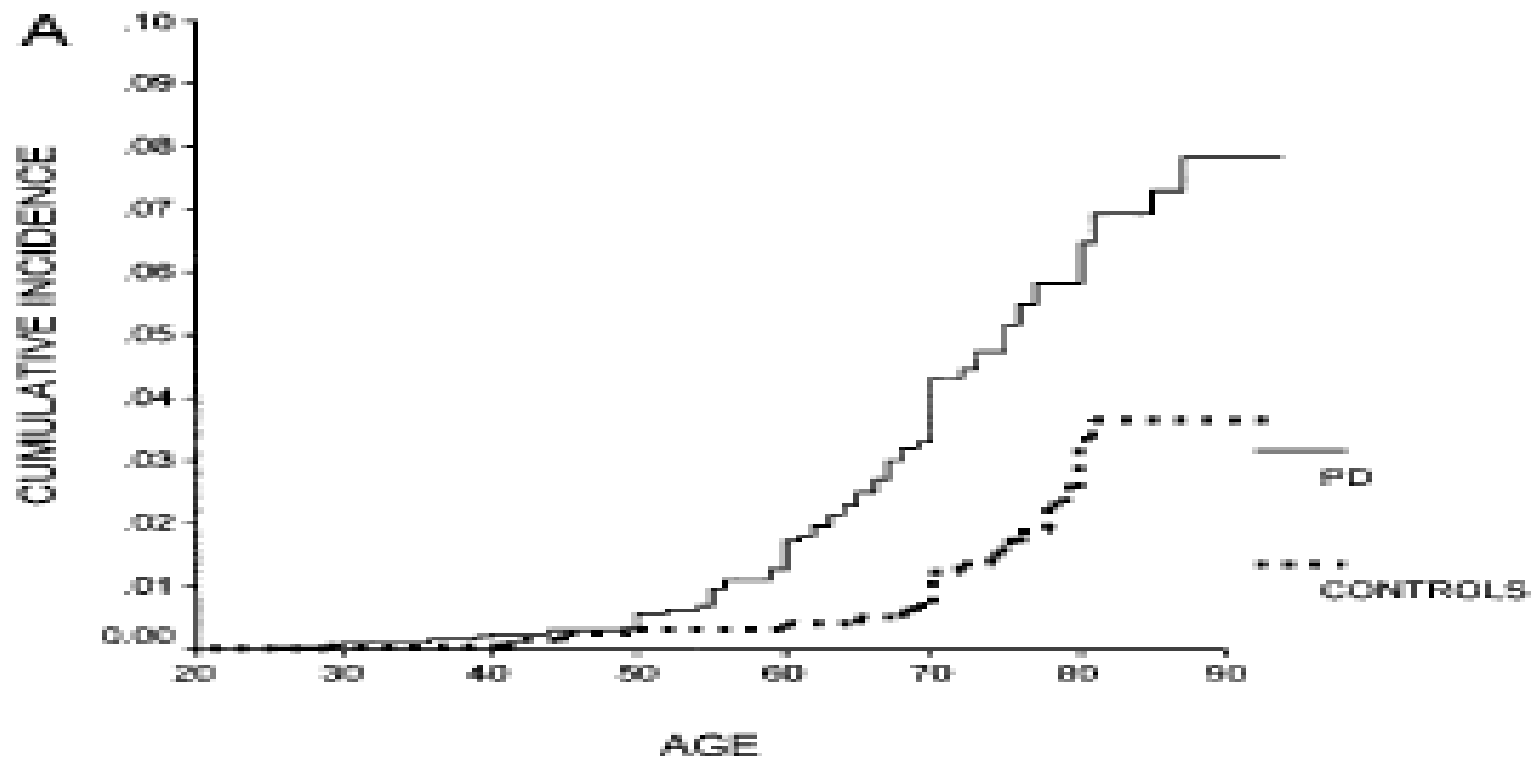
### Incidence cumulée de MP

	<u>App de cas</u>	<u>App de contrôles</u>
à 70 ans	4.3%	1.2%
à 80 ans	6.5%	3.4%

**RR=2.7** [95% CI=1.7- 4.4],  $p < 10^{-4}$

- MP cases (n= 487): 2865 App. 1er degré
- Contrôles (n=409): 2446 App. 1er degré

## Incidence cumulée de PD chez les app. 1er degré de patients PD et de contrôles



# RESSEMBLANCES FAMILIALES: Exemples

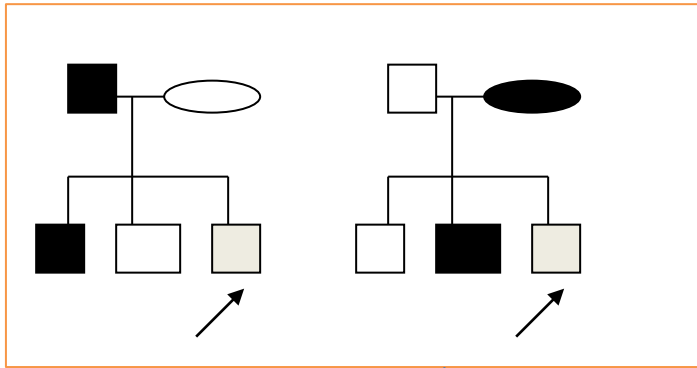
## Traits Cardio-vasculaires chez l'homme

	<u>Heritabilité</u>
total cholesterol	40-60%
HDL cholesterol	40-60%
LDL cholesterol	50-60%
Triglycerides	35-50%
Lp(a)	70-90%
Blood pressure	15-40%
Corpulence	25-60%
fasting glycemia	25-30%

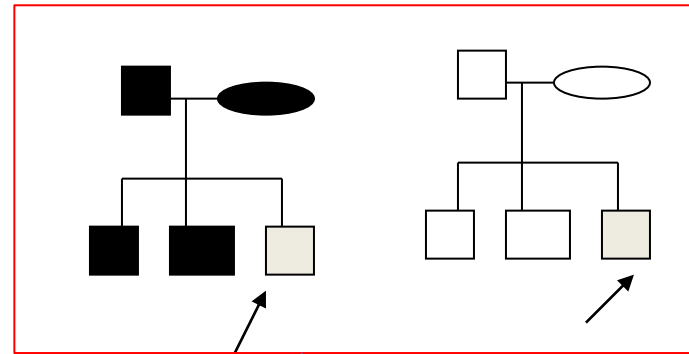
# CONCENTRATIONS FAMILIALES: Limites

- Un même risque  $\leftrightarrow$  différentes distributions du trait dans les familles des cas (proposants)
- Ne démontre pas l'origine génétique des concentrations familiales (autres causes: facteurs environnement communs)

## • Echantillon 1



## • Echantillon 2

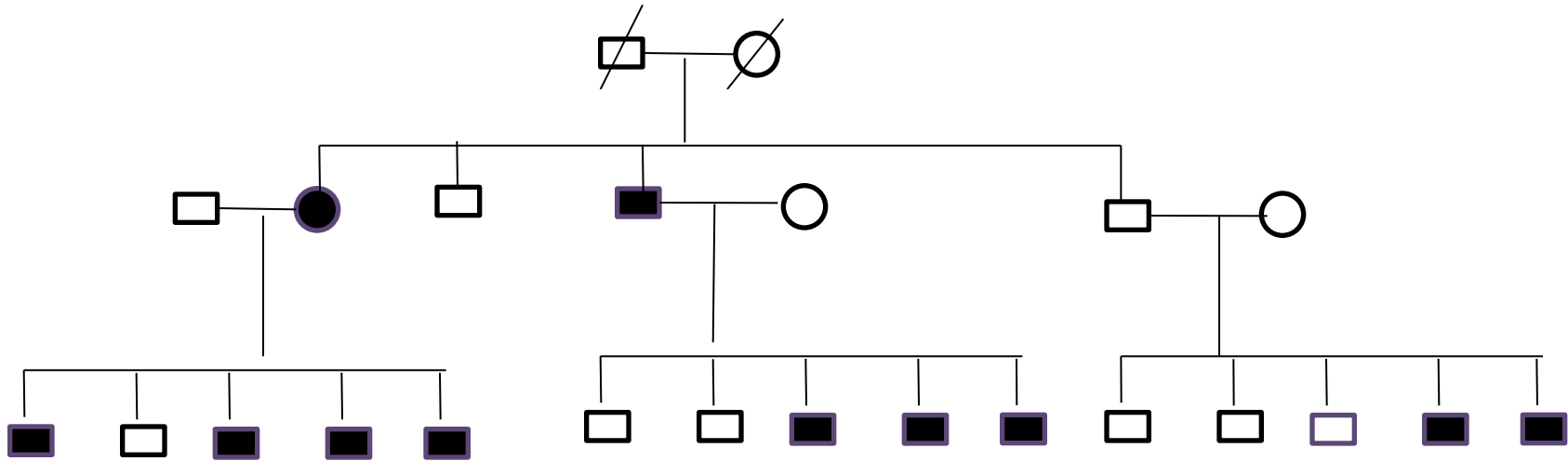


$\Rightarrow$  Même risque  $\sim 50\%$



**Infinité de modèles génétiques**

# Etude de la composante génétique d'un phénotype



## ➤ Analyse de ségrégation

Distribution du trait dans les familles avec prise en compte des facteurs de risque mesurés (âge, sexe, autres facteurs épidémiologique..)

➔ Tester des modes de transmission du trait

# Analyse de ségrégation

- **BUT**

Mettre en évidence la **ségrégation mendélienne d'un gène** « majeur » parmi l'ensemble des facteurs génétiques et non-génétiques causant les transmissions familiales du trait étudié

- **PRINCIPE**

Principe du **maximum de vraisemblance**

Calcul de la vraisemblance du modèle étant donné les observations.

Recherche du modèle le plus simple (parcimonieux) pouvant expliquer les distributions familiales du trait (en particulier, transmission d'un gène majeur).

# Méthode du Maximum de Vraisemblance

- **Vraisemblance du modèle (M) sachant les observations Y**

- Calcul de la probabilité des observations sous différents modèles (vraisemblance)

$$L( M (\Phi) | Y) \approx P( Y | M (\Phi))$$

$\Phi$  : ensemble des paramètres de M

- Pour chaque modèle considéré, chercher les valeurs des paramètres ( $\Phi'$ ) qui maximisent la vraisemblance

- **Tests d'hypothèses: Rapport du max de vraisemblance**

- Construction de modèles emboîtés: Hypothèses testées [ $M(\Phi'')$ ] sont des cas particuliers du modèle général  $M(\Phi')$

$$\lambda = L \max_{\Phi''} ( M (\Phi'') | Y) / L \max_{\Phi'} ( M (\Phi') | Y)$$

$-2\ln \lambda \sim \chi^2$  , ddl = nbr paramètres estimés sous  $\Phi'$  – nbr paramètres estimés sous  $\Phi''$



# MODELES MATHEMATIQUES DEVELOPPES POUR L'ANALYSE DE SEGREGATION

## – Modèles MIXTES-UNIFIES

Gène majeur + Composante Polygénique + Milieu

Modèles basés sur la décomposition de la variance ; corrélations non dues au gène majeur sont spécifiées par un mécanisme causal

## – Modèles REGRESSIFS

Gène majeur + Corrélations Familiales + Milieu

Modèles basées sur régression ; corrélations familiales de mécanisme général

# Modèle Mixte-Unifié

Y résulte des effets indépendants & additifs de :

- Gène majeur (GM)
- Composante Polygénique ( G)
- Milieu aléatoire (E)

Hypothèses du modèle:

- Pas d'interactions Gène-Milieu
- Schéma particulier des corrélations familiales (polygénique)
  - Pas de corrélations entre conjoints
  - Corr. Père-enfant = corr. Mère-enfant
  - Corr. Frère-Sœur  $\geq$  corr. Parents-Enfants
- Effets facteurs de risque estimés avant l'analyse de ségrégation (fixés)

Calcul de la vraisemblance complexe sur le plan numérique (intégrales multiples)

# Modèle Gène Majeur

- Locus autosomal di-allélique: A (q); a (1-q)

3 Génotypes:	<u>aa</u>	<u>Aa</u>	<u>AA</u>
Freq.	$(1-q)^2$	$2q(1-q)$	$q^2$
$E(Y \mid G_{Mi})$	$\mu_{aa}$	$\mu_{Aa}$	$\mu_{AA}$
$V(Y \mid G_{Mi})$	$\sigma^2$	$\sigma^2$	$\sigma^2$

Distribution de  $Y \mid G_{Mi} \sim N(\mu_{G_{Mi}}, \sigma^2)$

**Moyenne du gène majeur:  $E(GM) = (1-q)^2 \mu_{aa} + 2q(1-q) \mu_{Aa} + q^2 \mu_{AA}$**

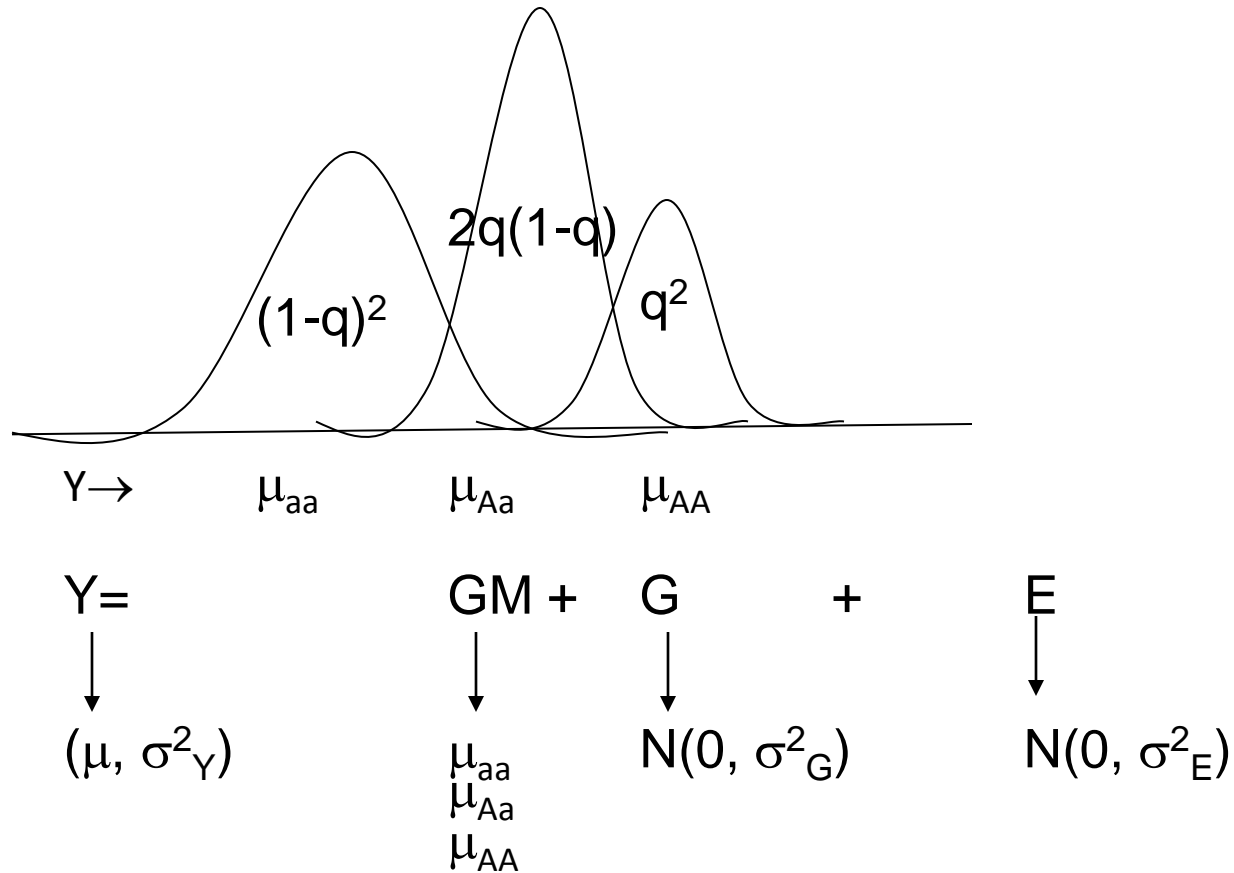
**Variance du gène majeur  $= (1-q)^2 \mu_{aa}^2 + 2q(1-q) \mu_{Aa}^2 + q^2 \mu_{AA}^2 - \mu^2$**

**Moyenne générale:  $\mu = E(Y) = E(MG) + E(G) + E(E) = E(MG)$**

Variance totale:  $\sigma_Y^2 = \sigma_{GM}^2 + \sigma^2$

Variance résiduelle:  $\sigma^2 = \sigma_G^2 + \sigma_E^2$

# Modèle Mixte-Unifié



# MODELE MIXTE UNIFIE -- Cas d'une maladie

S : variable sous-jacente non observable = *susceptibilité à la maladie*

S dans la population  $\sim N(0,1)$

L'affection =  $P(Y=1) = P(S \geq Z)$

Sur l'échelle de susceptibilité, valeur du seuil Z est déterminée par la fréquence de la maladie dans la population,  $k_p$ .

$$k_p = 1 - \Phi(Z / \sigma) = \Phi(-Z / \sigma)$$

$$\text{où } \Phi(x) = \text{cumulative normale} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) du$$

S résulte des effets génétiques et de milieu :  $E(S) = 0 = E(GM) + E(G) + E(E)$

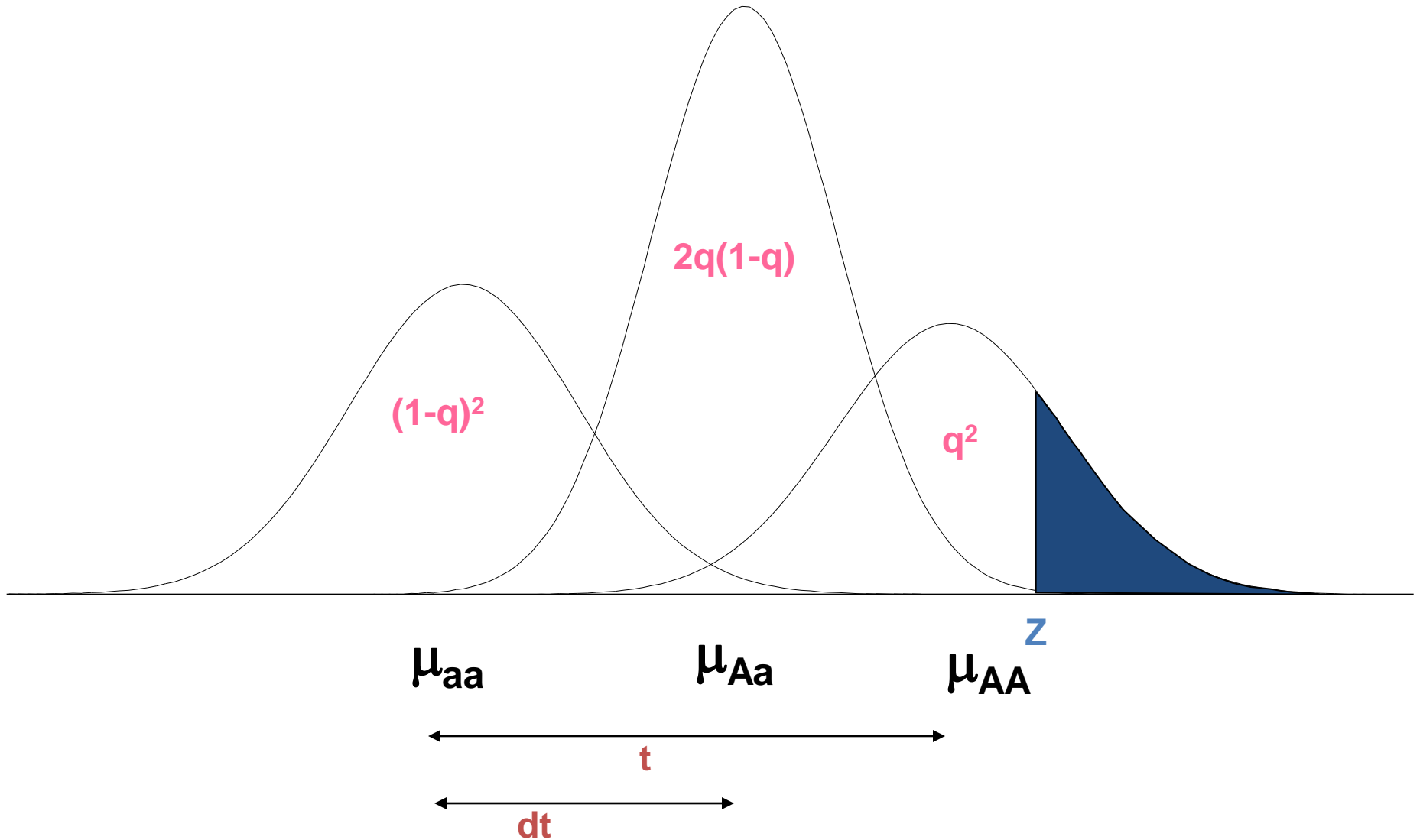
$$V(S) = \sigma^2 = 1 = \sigma_{GM}^2 + \sigma_G^2 + \sigma_E^2$$

G et E :  $N(0, \sigma_G^2)$   $N(0, \sigma_E^2)$

GM :  $N(0, \sigma_{GM}^2)$

- $H = \text{héritabilité} = \sigma_G^2 / \sigma^2$
- Loi de  $S | GM_i = N(\mu_{GMi}, \sigma_C^2 + \sigma_E^2) = N(\mu_{GMi}, 1 - \sigma_{GM}^2)$

# Modèle « mixte » à seuil – Susceptibilité génétique



# VRAISEMBLANCE D'UNE HYPOTHÈSE

Vraisemblance d'un modèle  $\propto$  Probabilité des observations (phénotypes  $Y$ ) sous le **modèle** ( $M$ ) & sous un **mode de sélection** ( $S$ ) des données :

$$P(Y|M, S) = P(Y|M) P(S|Y, M) / P(S|M)$$

Fonction de quatre composantes principales:

1. Distribution conjointe des génotypes dans la population
2. Relation phénotype-génotype
3. Loi de transmission du génotype d'une génération à l'autre
4. Mode d'échantillonnage des familles

# DISTRIBUTIONS DES 3 COMPOSANTES PRINCIPALES

## 1. MODELE MONOGENIQUE GENERALE (S=GM +E)

Locus au gène majeur = diallélique A et a → 3 génotypes ( $GM_i$ ,  $i=AA, Aa$  ou  $aa$ )

### 1. Distribution conjointe des génotypes dans la population : $P(GM_i)$

Distribution = Hardy Weinberg (les mariages sont aléatoires):

$$P_{GM} = P(AA)=q^2, P(Aa)=2q(1-q) \text{ et } P(aa)=(1-q)^2$$

### 2. Distribution du phénotype conditionnellement au génotype : $P(Y/GM_i)$

$Y=1$  (Malade) /  $0$  (Non malade)

$$P(Y=1/ GM_i) = \textit{pénétrance } f_i$$

$k_p$ = prévalence de la maladie dans la population =  $\sum_i P(GM_i) \times f_i$



# DISTRIBUTIONS DES 3 COMPOSANTES PRINCIPALES

## 1. MODELE MONOGENIQUE GENERALE

### 3. Distribution des génotypes des enfants conditionnellement aux génotypes parentaux

$GM_e, GM_f, GM_m$  = génotypes de l'enfant e, du père et de la mère.

$P(GM_e / GM_f, GM_m) = \text{fct } \textbf{des taux de transmission} (t_{k,GM_i}).$

$$t_{k,GM_i} = P(\text{transmettre allèle } k / GM_i)$$

Gène mendélien : valeurs des taux de transmission = celles de la loi de l'hérédité mendélienne :  $t_{A/AA}=1$ ;  $t_{A/Aa}=0.5$  et  $t_{A/aa}=0$ .

# Probabilités de transmission Gène Majeur

## Lois de Mendel

$$\tau_{k,GMj} = P(\text{Allèle } k \text{ transmis} \mid GM=Gj)$$

$$\tau_{A,AA} = 1 - \tau_{a,AA}$$

$$\tau_{A,AA} = 1$$

$$\tau_{A,Aa} = 1 - \tau_{a,Aa}$$

$$\tau_{A,Aa} = \frac{1}{2}$$

$$\tau_{A,aa} = 1 - \tau_{a,aa}$$

$$\tau_{A,aa} = 0$$

$$\rightarrow P(GMe \mid GMf, GMm)$$

GMm		GMf		
		1=AA	2=Aa	3=aa
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}

# DISTRIBUTIONS DES 3 COMPOSANTES PRINCIPALES

## 1. MODELE MONOGENIQUE GENERALE (S= GM +E)

Vraisemblance du modèle monogénique pour une famille nucléaire

$$L(Y_f, Y_m, Y_1..Y_n) = \sum_{GM_f} P(Y_f/GM_f) \times P(GM_f) \sum_{GM_m} P(Y_m/GM_m) P(GM_m) \\ \prod_n \sum_{GM_e} P(Y_j/GM_e) P(GM_e/GM_f, GM_m)$$

Paramètres du modèle

- Moyenne générale:  $\mu$
- Variance totale:  $\sigma^2_Y$
- Fréquence génique:  $q$
- 2 moyennes génotypiques ou
  - « déplacement: displacement » :  $t = \mu_{AA} - \mu_{aa}$
  - « dominance »:  $d = (\mu_{Aa} - \mu_{aa}) / (\mu_{AA} - \mu_{aa})$
- 3 taux de transmission

# DISTRIBUTIONS DES 3 COMPOSANTES PRINCIPALES

## 2. MODELE POLYGENIQUE ( $S = C + E$ )

### 1. Distribution conjointe des polygénotypes dans la population :

$$P(g_k) = \phi(g_k) dg$$

$g_k$  = polygénotype d'un individu

$\phi(g_k)$  = densité d'une normale  $(0, \sigma_G^2) = \{1/\sigma_G \sqrt{2\pi}\} \exp[-1/2(g_k^2/\sigma_G^2)] dg$

### 2. Distribution du phénotype conditionnellement au polygénotype : $P(Y/G_k)$

$$S|g_k \sim N(g_k, \sigma_E^2)$$

$$P(Y=1 | g_k) = P(S \geq Z | g_k) = 1 - \Phi([Z - g_k] / \sigma_E)$$

$$P(Y=0 | g_k) = P(S < Z | g_k) = \Phi([Z - g_k] / \sigma_E)$$

- Prévalence,  $k_p = \int_{-\infty}^{+\infty} \Phi([g - Z] / \sigma_E) f(g) dg$

# DISTRIBUTIONS DES 3 COMPOSANTES PRINCIPALES

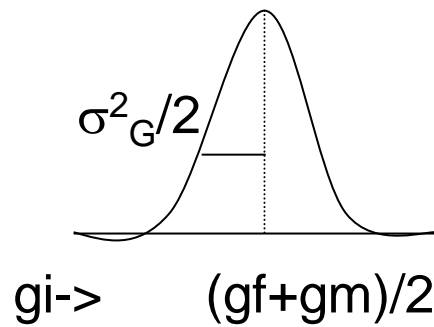
## 2. MODELE POLYGENIQUE ( $S = C + E$ )

### 3. Distribution des polygénotypes des enfants conditionnellement aux polygénotypes parentaux

$$P(g_e / g_f, g_m) = \phi(g_e) dg$$

$g_e, g_f, g_m$  = polygénotypes de l'enfant, du père et de la mère.

$\phi(g_e)$  = densité d'une normale ( $\frac{1}{2} \{g_f + g_m\}, \frac{1}{2}\sigma_{CG}^2$ )



# DISTRIBUTIONS DES 3 COMPOSANTES PRINCIPALES

## 2. MODELE POLYGENIQUE (S= C +E)

Vraisemblance du modèle pour une famille nucléaire

$$L(Y_f, Y_m, Y_1..Y_n) = \int_{-\infty}^{+\infty} P(g_f) P(Y_f/g_f) \int_{-\infty}^{+\infty} P(g_m) P(Y_m/g_m) \\ \prod_n \int_{-\infty}^{+\infty} P(Y_j/g_e) P(g_e/g_f, g_m) dg_f dg_m dg_i$$

Paramètre :  $H = \sigma^2_G$

## Vraisemblance du Modèle Mixte-Unifié ( $S = GM + G + E$ )

$$L(GM, G | F) \approx P(Y_f, Y_m, Y_1, \dots, Y_n | GM, G)$$

$$L = \int f(g_f) \Sigma_{GM_f} P(GM_f) f(Y_f | GM_f, g_f) \int f(g_m) \Sigma_{GM_m} P(GM_m) f(Y_m | GM_m, g_m) \\ \Pi_i [\int f(g_i | g_f, g_m) \Sigma_{G_i} P(GM_i | GM_f, GM_m) f(Y_i | GM_i, g_i)] dg_f, dg_m, dg_i$$

$$f(Y | GM, G) = \phi(\{\mu_G + g\}, \sigma^2) \qquad \sigma^2 = \sigma_E^2 = \sigma_Y^2 - \sigma_{GM}^2 - \sigma_G^2$$

$$f(Y | GM, G) = 1/\sigma \sqrt{2\pi} \exp[-1/2 (\{Y - \mu_G - g\}/\sigma)^2]$$

Ou bien (selon Lalouel et Morton, 1981):

$$L = \int f(g_f) \Sigma_{GM_f} P(GM_f) f(Y_f | GM_f, g_f) \int f(g_m) \Sigma_{GM_m} P(GM_m) f(Y_m | GM_m, g_m) \\ \Pi_i [\Sigma_{GM_i} P(GM_i | GM_f, GM_m) f(Y_i | GM_i, g_f, g_m)] dg_f, dcm$$

# Paramètres du Modèle Mixte-Unifié

Moyenne du trait:  $\mu$

Variance totale :  $\sigma^2_Y$

Fréquence génique:  $q$

Displacement :  $t = \mu_{AA} - \mu_{aa}$

Dominance:  $d = (\mu_{Aa} - \mu_{aa}) / (\mu_{AA} - \mu_{aa})$

Héritabilité:  $H$

variante 2 héritabilités :  $H_A$  chez les parents et  $H_K$  chez les enfants

+ 3 probabilités de transmission:

Modèle mixte unifié (Lalouel et al., 1983)



# Mode d'échantillonnage des familles

- Familles sont généralement recensées via des sujets atteints (valeur du trait) → Distributions biaisées

Exemple: Maladie autosomale récessive

Distribution des familles de 2 enfants with heterozygous parents : 4 types



$$1/4 \times 1/4 = 1/16$$



$$1/4 \times 3/4 = 3/16$$



$$3/4 \times 1/4 = 3/16$$



$$3/4 \times 3/4 = 9/16$$

**La 4<sup>ème</sup> (la plus fréquente dans la population) est absente de l'échantillon**

→ Il est nécessaire de corriger la vraisemblance du modèle pour le mode de recensement des familles

# VRAISEMBLANCE D'UNE HYPOTHÈSE

Vraisemblance d'un modèle  $\propto$  Probabilité des observations (phénotypes  $Y$ ) sous le **modèle** ( $M$ ) & sous un **mode de sélection** ( $S$ ) des données :

$$P(Y | M, \mathbf{S}) = P(Y | M) P(\mathbf{S} | Y, M) / P(\mathbf{S} | M)$$

Fonction de quatre composantes principales:

1. Distribution conjointe des génotypes dans la population
2. Relation phénotype-génotype
3. Loi de transmission du génotype d'une génération à l'autre
4. **Mode d'échantillonnage des familles**

# MODELES CONSIDERES

- 1) Absence de corrélations familiales (M1)  
Pas GM ( $q=t=d=0$ ) + pas Comp Polygénique ( $H=0$ )  
Paramètres estimés:  $\mu, \sigma^2_Y$
- 2) Absence de GM, présence composante polygénique (M2)  
Pas GM ( $q=t=d=0$ )  
Paramètres estimés:  $\mu, \sigma^2_Y, H$
- 3) Absence composante polygénique mais présence GM (M3)  
pas Comp Polygénique ( $H=0$ )  
Paramètres estimés:  $\mu, \sigma^2_Y, q, d, t$
- 4) Présence GM & comp. Polygénique (M4)  
Paramètres estimés:  $\mu, \sigma^2_Y, q, d, t, H$

# TRAIT Quantitatif/Susceptibilité

## – TESTS D’HYPOTHESES MODELE MIXTE-UNIFIE

Modèle	$\mu$	$\sigma^2_Y$	q	t	d	h	$\tau_{A,AA}$	$\tau_{A,Aa}$	$\tau_{A,qq}$	ddl
M1	*	*	(0)	(0)	(0)	(0)	-	-	-	2
M2	*	*	(0)	(0)	(0)	*	-	-	-	3
M3	*	*	*	*	*	(0)	(1)	( $\frac{1}{2}$ )	(0)	5
M4	*	*	*	*	*	*	(1)	( $\frac{1}{2}$ )	(0)	6
M5	*	*	*	*	*	*	*	*	*	9
M6	*	*	*	*	*	*	*	$=(\tau_{A,AA})$	$=(\tau_{A,AA})$	7

\*Paramètres estimés; () valeurs des paramètres fixées; -paramètres n’ayant pas de sens sous le modèle considéré; ddl: nombre de paramètres du modèle

# TESTS D'HYPOTHESES

## Mise en évidence Effets GM

### 1. **Y-a-t-il des corrélations familiales pour le trait ?** (M1 vs M2)

Si significatif (M1 rejeté) → Présence corr. Familiales

### 2. **Un GM ségrège-t-il dans les familles ?** (M2 vs M4)

Si significatif (M2 rejeté) → Présence gène (facteur) majeur

### 3. **Y-a-t-il une Comp. Polyg en plus de GM ?** (M3 vs M4)

Si significatif (M3 rejeté) → Présence Comp Polyg en plus gène majeur

# TESTS D'HYPOTHESES

Si effets GM mis en évidence: **EST-CE BIEN UN GENE?**

Deux Tests effectués:

A) **Se transmet-il selon lois Mendel ?** (M4 vs M5)

*Transmission Mendélienne vs Modèle général de Transmission*

B) **Se transmet-il des parents aux enfants?** (M6 vs M5)

*Absence de Transmission Parents-Enfants vs Modèle général de Transmission*

Si A est non significatif (M4 non rejeté) & B significatif (M6 rejeté)

⇒ Très vraisemblable que le facteur majeur détecté soit un gène