

## Quelques éléments sur l'analyse d'association

---

---

---

---

---

---

---

---

### Déséquilibre de liaison ou gamétique

Non indépendance entre allèles en deux locus sur un même gamète

	B	b	
A	$p_A p_B + D$	$p_A(1-p_B) - D$	$p_A$
a	$(1-p_A)p_B - D$	$(1-p_A)(1-p_B) + D$	$1-p_A$
	$p_B$	$(1-p_B)$	1

---

---

---

---

---

---

---

---

- D : écart à l'attendu aléatoire dans les fréquences gamétiques, si indépendance
- $D = p_{AB} p_{ab} - p_{Ab} p_{AB}$
- D dépend des fréquences alléliques.

### Mesure du déséquilibre de liaison

- Plusieurs mesures possibles, dont
- $r^2 = D^2 / \{p_A p_B (1-p_A) (1-p_B)\}$
- $r^2$  est la corrélation entre allèles aux locus A et B
- très proche de la corrélation des génotypes  $\Rightarrow$  inutile d'observer les gamètes.

---

---

---

---

---

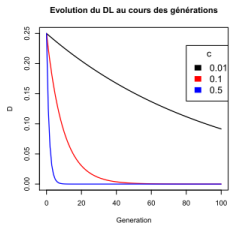
---

---

---

## Multiples causes de DL Même évolution

$$D_{t+1} = D_t(1 - c) = D_0(1 - c)^t$$



Le DL, créé à un moment, disparaît progressivement au fil des générations du fait des recombinaisons

Disparition d'autant plus rapide que les locus sont distants

---

---

---

---

---

---

---

---

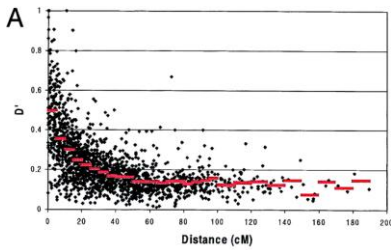
---

---

---

---

## Le LD est une réalité



La dépendance entre les génotypes en 2 locus est d'autant plus forte que ces locus sont proches

---

---

---

---

---

---

---

---

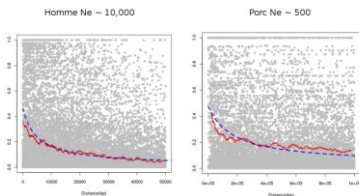
---

---

---

---

## Variabilité entre espèces



(Diapo B Servin)

---

---

---

---

---

---

---

---

---

---

---

---

## A venir dans la suite du cours

- Causes du DL
- (autres) mesures du DL
- Simulations de populations en DL
- Exploitation du DL par diverses approches (GWAS, Genomique, LDA)

---

---

---

---

---

---

---

---

## Le point clé

- Le DL existe et est prouvé entre marqueurs
- Il doit donc exister entre locus marqueurs et QTL
- Donc ce qui se voit sur les marqueurs informe sur les QTL putatifs à proximité
- Quand des marqueurs donnent un signal, un QTL n'est pas loin.....

---

---

---

---

---

---

---

---

## Utilisation du DL pour analyser des données non structurées. Principes

- Les individus sont supposés non ou faiblement apparentés (QTLMap à l'inverse travaille sur des populations expérimentales)
- Les caractères sont tout ou rien (malades/ sains) ou continus
- Beaucoup de marqueurs sont mesurés
- Les QTL sont détectés par leur effets visibles sur les marqueurs avec lesquels ils sont en DL

---

---

---

---

---

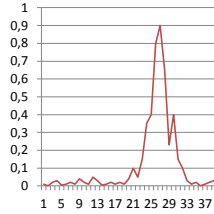
---

---

---

### Idée générale

- Mettre en place une « statistique »  $S(p,g)$  dépendant des valeurs phénotypique et des génotypes aux marqueurs
- Dont les valeurs élevées pour certains emplacements du génome seront des signaux de présence éventuelle de QTLs




---

---

---

---

---

---

---

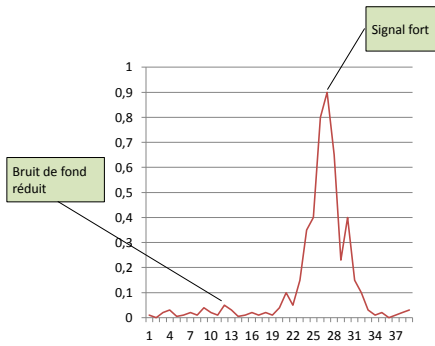
---

---

---

---

---




---

---

---

---

---

---

---

---

---

---

---

---

### Données tout ou rien : relative risk

	Fumeurs	Non fumeurs	
Sains	FS	NFS	S=FS+NFS
Atteints	FA	NFA	A=FA+NFA
	F=FS+FA	NF=NFS+NFA	

Comparer  $FS/S$  à  $FA/A$  ( $d = FA/A - FS/S$ ) en prenant comme statistique  $\frac{d}{\sqrt{\text{var}(d)}}$

Mauvaise idée : Car on cherche à comparer les risques ( $FA/F$  et  $NFA/NF$ ) chez les fumeurs et les non fumeurs et non la proportion de fumeurs dans ces deux groupes

---

---

---

---

---

---

---

---

---

---

---

---

### Données tout ou rien : relative risk

	Fumeurs	Non fumeurs	
Sains	FS	NFS	S=FS+NFS
Atteints	FA	NFA	A=FA+NFA
	F=FS+FA	NF=NFS+NFA	

Mauvaise idée : comparer FS/S à FA/A (d= FA/A -FS/S )

Exemple. Risque selon le comportement constant : FA/F = r<sub>f</sub> = 0.10 et NFA/NF = r<sub>nf</sub> = 0.01  
Le résultat en utilisant d dépend du recensement

	FS	NFS	S
Pop1	5000	5000	10000
Pop2	1000	9000	10000

---

---

---

---

---

---

---

---

---

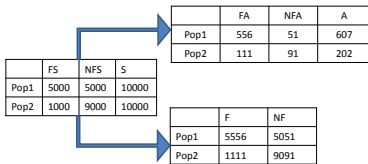
---

### Données tout ou rien : relative risk

	Fumeurs	Non fumeurs	
Sains	FS	NFS	S=FS+NFS
Atteints	FA	NFA	A=FA+NFA
	F=FS+FA	NF=NFS+NFA	

Mauvaise idée : comparer FS/S à FA/A (d= FA/A -FS/S )

Exemple. Risque selon le comportement constant : FA/F = r<sub>f</sub> = 0.10 et NFA/NF = r<sub>nf</sub> = 0.01  
Le résultat dépend du recensement




---

---

---

---

---

---

---

---

---

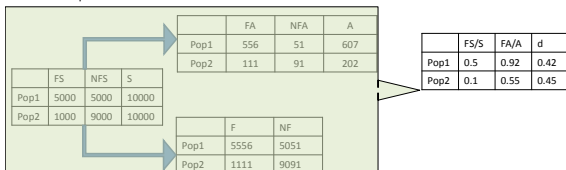
---

### Données tout ou rien : relative risk

	Fumeurs	Non fumeurs	
Sains	FS	NFS	S=FS+NFS
Atteints	FA	NFA	A=FA+NFA
	F=FS+FA	NF=NFS+NFA	

Mauvaise idée : comparer FS/S à FA/A (d= FA/A -FS/S )

Exemple. Risque selon le comportement constant : FA/F = r<sub>f</sub> = 0.10 et NFA/NF = r<sub>nf</sub> = 0.01  
Le résultat dépend du recensement




---

---

---

---

---

---

---

---

---

---

## Il faut utiliser le risque relatif

$$RR = \frac{FA/F}{NFA/NF} = \frac{RF}{RNF}$$

Ou l'Odd Ratio, rapport de la « cote » d'un évènement arrivant au groupe F d'individus, avec celle du même évènement arrivant à un groupe NF.

$$OR = \frac{RF}{1-RF} \bigg/ \frac{RNF}{1-RNF}$$

---

---

---

---

---

---

---

---

---

---

## Données tout ou rien : relative risk

	Fumeurs	Non fumeurs	
Sains	FS	NFS	S=FS+NFS
Atteints	FA	NFA	A=FA+NFA
	F=FS+FA	NF=NFS+NFA	

Mauvaise idée : comparer FS/S à FA/A (d= FA/A -FS/S)

Exemple. Risque selon le comportement constant : FA/F = r<sub>f</sub> = 0.10 et NFA/NF = r<sub>nf</sub> = 0.01  
Le résultat dépend du recensement

The diagram illustrates the calculation of relative risk from raw data for two populations (Pop1 and Pop2). It shows three tables: a raw data table, a table of relative frequencies (FS/S, FA/A, d), and a table of relative risks (FA/F, NFA/NF, RR).

	FS	NFS	S
Pop1	5000	5000	10000
Pop2	1000	9000	10000

	FA	NFA	A
Pop1	556	51	607
Pop2	111	91	202

	F	NF
Pop1	5556	5051
Pop2	1111	9091

	FS/S	FA/A	d
Pop1	0.5	0.92	0.42
Pop2	0.1	0.55	0.45

	FA/F	NFA/NF	RR
Pop1	0.1	0.01	10
Pop2	0.1	0.01	10

---

---

---

---

---

---

---

---

---

---

## Application à la recherche de QTLs

On construit un tableau Etat x Génotype pour chaque marqueur à tester

	AA	AB	BB	
Sains	s <sub>0</sub>	s <sub>1</sub>	s <sub>2</sub>	s
Atteints	r <sub>0</sub>	r <sub>1</sub>	r <sub>2</sub>	r
	n <sub>0</sub>	n <sub>1</sub>	n <sub>2</sub>	n

Les risques sont R<sub>i</sub> = s<sub>i</sub> / n<sub>i</sub>

Le risque relatif de AA par rapport à AB est RR =  $\frac{R_{AA}}{R_{AB}}$

---

---

---

---

---

---

---

---

---

---

### Comment tester l'effet d'un marqueur ?

	AA	AB	BB	
Sains	$s_0$	$s_1$	$s_2$	$s$
Atteints	$r_0$	$r_1$	$r_2$	$r$
	$n_0$	$n_1$	$n_2$	$n$

- Test de chi2 de la table des observations  $T = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$
- Test de Fisher (adapté aux effectifs faibles)
- Test Armitage (prend compte de tendances. Puissant si les hypothèses de la tendance sont avérées)

$$T_x = \frac{(\frac{1}{n} \sum_{i=0}^2 x_i (sr_i - rs_i))^2}{(rs/n^2) \sum_{i=0}^2 n_i x_i (1 - x_i)}, \quad T_x \sim \chi_1^2$$

---

---

---

---

---

---

---

---

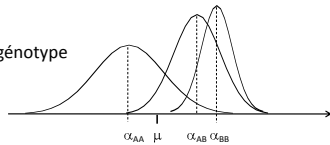
---

---

### Données continues : ANOVA

Modélisation de l'effet du génotype ( $g_i = AA, AB \text{ ou } BB$ )

$$y_i = \mu + \alpha_{g_i} + e_i \quad (1)$$



Modélisation de l'effet du nombre d'allèles B ( $x_i = 0, 1 \text{ ou } 2$ )

$$y_i = \mu + \beta x_{g_i} + e_i \quad (2)$$

On utilise les tests usuels d'effets de facteurs sur une mesure  
 Test de Fisher à (nb alleles - 1, N-nb alleles) ddl pour (1)  
 Test de Fisher à (1, N-1) ddl pour (2)

---

---

---

---

---

---

---

---

---

---

### Utilisation simultanée de plusieurs marqueurs

- Les analyses précédentes peuvent être étendues à l'utilisation d'haplotypes (ou de génotypes)
- Changement dans le nombre d'éléments dans la matrice ou dans le nombre de niveaux dans l'ANOVA
- Utilité espérée : meilleure prédiction de l'état IBD au QTL putatif  $\Rightarrow$ 
  - Plus de puissance
  - plus de précision

---

---

---

---

---

---

---

---

---

---

## Puissances de ces méthodes

Le cas des données tout ou rien

Nombre de control/cas

Mj	Relative risk								
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	10.0
Case-control sets (n=10):									
1	13	8	6	5	4	4	4	4	7
2	31	23	21	20	19	20	21	21	38
4	53	46	44	43	44	46	47	49	71
8	72	67	66	66	67	69	70	72	88
16	85	82	82	82	83	84	85	86	94
32	93	91	91	91	92	92	93	93	98
Case-control sets (n=20):									
1	19	13	11	11	11	12	13	15	43
2	39	32	30	31	33	36	40	43	82
4	60	54	54	56	59	62	66	70	95
8	76	73	74	75	78	80	83	85	99
16	87	86	86	87	89	90	92	93	99
32	94	93	93	94	95	95	96	97	100

(Pang, Occup Environ Med 1999; 56)

---

---

---

---

---

---

---

---

---

---

---

---

## Puissances de ces méthodes

Le cas des données continues

No. SNPs included in model	Marker density (no. SNPs in 11 cM region)									Population de 500 individus
	6			10			20			
	Geno	Haplo	IBD	Geno	Haplo	IBD	Geno	Haplo	IBD	
	Power to detect QTL (%)									
1	26	-	18	31	-	21	34	-	22	
2	25	23	25	28	27	30	31	28	34	
4	24	15	28	28	18	32	30	19	31	
6	-	-	27	-	-	34	-	-	32	
8	-	-	-	-	-	32	-	-	-	

Régression sur haplotypes      Régression sur génotypes      Modèle Mixte de M&G

Zhao et al, Genetics, 2007

---

---

---

---

---

---

---

---

---

---

---

---

## Limites de ces méthodes

- Contrôle difficile de l'erreur de première espèce du fait du très grand nombre de tests
- Non robustesse aux structures cachées, notamment génétiques

---

---

---

---

---

---

---

---

---

---

---

---



## Contrôle de l'erreur de première espèce

Un bon test statistique ?

permettrait de décider qu'une hypothèse particulière (H0) est vraisemblable parmi un ensemble d'hypothèses possibles (H1) en étant sévère sur un rejet injustifié de H0 tout en voulant se donner de bonnes chances de jeter H0 si elle est fausse

Nature Conclusion	H0	H1
H0	Contrôle de H0 (1- $\alpha$ )	Erreur de 2 <sup>ème</sup> espèce ( $\beta$ )
H1	Erreur de 1 <sup>ère</sup> espèce ( $\alpha$ )	Puissance (1- $\beta$ )

Dans le contexte : H0 = le marqueur na pas d'effet sur le caractère (donc pas de QTL à proximité)

---

---

---

---

---

---

---

---

---

---

---

---

## Tests multiples

- On teste H0<sub>1</sub> : le SNP1 n'a pas d'effet et H0<sub>2</sub> : le SNP2 n'a pas d'effet

---

---

---

---

---

---

---

---

---

---

---

---

## Tests multiples

- On teste H0<sub>1</sub> : le SNP1 n'a pas d'effet et H0<sub>2</sub> : le SNP2 n'a pas d'effet
- Test statistique :  $S(p, g_1)$  et  $S(p, g_2)$   
Par exemple  $S(p, g_1)$  = moyenne des phénotypes des AA - moyenne des phénotypes des AB

---

---

---

---

---

---

---

---

---

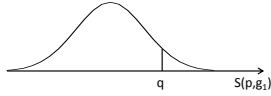
---

---

---

### Tests multiples

- On teste  $H_{01}$  : le SNP1 n'a pas d'effet  
et  $H_{02}$  : le SNP2 n'a pas d'effet
- Test statistique :  $S(p, g_1)$  et  $S(p, g_2)$   
Par exemple  $S(p, g_i)$  = moyenne des phénotypes des AA - moyenne des phénotypes des AB
- $H_{01}$  est rejetée si  $S(p, g_1) > q$  et  $H_{02}$  si  $S(p, g_2) > q$




---

---

---

---

---

---

---

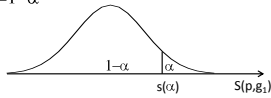
---

---

---

### Tests multiples

- On teste  $H_{01}$  : le SNP1 n'a pas d'effet  
et  $H_{02}$  : le SNP2 n'a pas d'effet
- Test statistique :  $S(p, g_1)$  et  $S(p, g_2)$   
Par exemple  $S(p, g_i)$  = moyenne des phénotypes des AA - moyenne des phénotypes des AB
- $H_{01}$  est rejetée si  $S(p, g_1) > q$  et  $H_{02}$  si  $S(p, g_2) > q$
- Un faux rejet de  $H_{01}$  au niveau  $\alpha$  est contrôlé si  $q = s(\alpha)$   
tel que  $\text{prob}(S(p, g_1) < s(\alpha)) = 1 - \alpha$




---

---

---

---

---

---

---

---

---

---

### Tests multiples

- On teste  $H_{01}$  : le SNP1 n'a pas d'effet  
et  $H_{02}$  : le SNP2 n'a pas d'effet
- Test statistique :  $S(p, g_1)$  et  $S(p, g_2)$   
Par exemple  $S(p, g_i)$  = moyenne des phénotypes des AA - moyenne des phénotypes des AB
- $H_{01}$  est rejetée si  $S(p, g_1) > q$  et  $H_{02}$  si  $S(p, g_2) > q$
- Un faux rejet de  $H_{01}$  au niveau  $\alpha$  est contrôlé si  $q = s(\alpha)$   
tel que  $\text{prob}(S(p, g_1) < s(\alpha)) = 1 - \alpha$
- On veut contrôler  $H_{01}$  et  $H_{02}$ ,  
donc on cherche  $Q$  tel que  $\text{prob}(S(p, g_1) < Q \text{ et } S(p, g_2) < Q) = 1 - \alpha$

---

---

---

---

---

---

---

---

---

---

## Tests multiples

- On teste  $H_{01}$  : le SNP1 n'a pas d'effet  
et  $H_{02}$  : le SNP2 n'a pas d'effet
- Test statistique :  $S(p, g_1)$  et  $S(p, g_2)$   
Par exemple  $S(p, g_i)$  = moyenne des phénotypes des AA - moyenne des phénotypes des AB
- $H_{01}$  est rejetée si  $S(p, g_1) > q$  et  $H_{02}$  si  $S(p, g_2) > q$
- Un faux rejet de  $H_{01}$  au niveau  $\alpha$  est contrôlé si  $q = s(\alpha)$   
tel que  $\text{prob}(S(p, g_1) < s(\alpha)) = 1 - \alpha$
- On veut contrôler  $H_{01}$  et  $H_{02}$ ,  
donc  $\text{prob}(S(p, g_1) < Q \text{ et } S(p, g_2) > Q) = 1 - \alpha$
- Soit  $\text{prob}(S(p, g_1) < Q) \cdot \text{prob}(S(p, g_2) < Q) = 1 - \alpha$

---

---

---

---

---

---

---

---

---

---

---

---

## Tests multiples

- On teste  $H_{01}$  : le SNP1 n'a pas d'effet  
et  $H_{02}$  : le SNP2 n'a pas d'effet
- Test statistique :  $S(p, g_1)$  et  $S(p, g_2)$   
Par exemple  $S(p, g_i)$  = moyenne des phénotypes des AA - moyenne des phénotypes des AB
- $H_{01}$  est rejetée si  $S(p, g_1) > q$  et  $H_{02}$  si  $S(p, g_2) > q$
- Un faux rejet de  $H_{01}$  au niveau  $\alpha$  est contrôlé si  $q = s(\alpha)$   
tel que  $\text{prob}(S(p, g_1) < s(\alpha)) = 1 - \alpha$
- On veut contrôler  $H_{01}$  et  $H_{02}$ ,  
donc  $\text{prob}(S(p, g_1) < Q \text{ et } S(p, g_2) > Q) = 1 - \alpha$
- Soit  $\text{prob}(S(p, g_1) < Q) \cdot \text{prob}(S(p, g_2) < Q) = 1 - \alpha$
- Donc  $\text{prob}(S(p, g_1) < Q) = \sqrt{1 - \alpha} \approx 1 - 1/2\alpha \Rightarrow Q = s(\alpha/2)$

---

---

---

---

---

---

---

---

---

---

---

---

## Tests multiples

Donc le contrôle global de l'erreur de première espèce quand on teste deux SNP est obtenu en prenant le niveau  $\alpha/2$  plutôt que  $\alpha$ .

Si on a 50 000 SNP, il faut prendre le niveau  $\alpha/50000$ ....

Le seuil de rejet devient très élevé et la puissance faible

---

---

---

---

---

---

---

---

---

---

---

---

## Tests multiples

Donc le contrôle global de l'erreur de première espèce quand on teste deux SNP est obtenu en prenant le niveau  $\alpha/2$  plutôt que  $\alpha$

Si on a 50 000 SNP, il faut prendre le niveau  $\alpha/50000$ ....

$$\alpha = \text{GWER (Genome Wise Error Rate)} = \text{prob}(\text{nb faux positif} > 0)$$

Le seuil de rejet devient très élevé et la puissance faible

Faut il être aussi intransigeant dans le contrôle des erreurs ?

FDR (Benjamini et Hochberg, 1995) :  $E[\text{nb faux positif} / \text{nb positifs}]$

$\text{GWER}_k$  (Chen et Storey, Genetics, 2006) :  $\text{prob}(\text{nb faux positifs} > k)$

## Tests multiples

Donc le contrôle global de l'erreur de première espèce quand on teste deux SNP est obtenu en prenant le niveau  $\alpha/2$  plutôt que  $\alpha$

Si on a 50 000 SNP, il faut prendre le niveau  $\alpha/50000$ ....

$$\alpha = \text{GWER (Genome Wise Error Rate)} = \text{prob}(\text{nb faux positif} > 0)$$

Le seuil de rejet devient très élevé et la puissance faible

Faut il être aussi intransigeant dans le contrôle des erreurs ?

FDR (Benjamini et Hochberg, 1995) :  $E[\text{nb faux positif} / \text{nb positifs}]$

$\text{GWER}_k$  (Chen et Storey, Genetics, 2006) :  $\text{prob}(\text{nb faux positifs} > k)$

En outre les tests ne sont pas indépendants

## Non indépendance entre les tests

- Du fait du DL entre marqueurs proches (SNP1 et SNP2), les statistiques de test  $S(p, g_1)$  et  $S(p, g_2)$  ne sont pas indépendantes
- Donc

$$\text{prob}(S(p, g_1) < Q \text{ et } S(p, g_2) < Q) > \text{prob}(S(p, g_1) < Q) \cdot \text{prob}(S(p, g_2) < Q)$$

- Si SNP1 et SNP2 sont totalement liés,  $S(p, g_1) = S(p, g_2)$ , donc  $\text{prob}(S(p, g_1) < Q \text{ et } S(p, g_2) < Q) = \text{prob}(S(p, g_1) < Q)$

⇒ Un GWER de  $\alpha$  sera obtenu avec des seuils plus grands que  $s(\alpha/50000)$

### Limites de ces méthodes

- Contrôle difficile de l'erreur de première espèce du fait du très grand nombre de tests
- Non robustesse aux structures cachées, notamment génétiques

---

---

---

---

---

---

---

---

### Confusion entre effets d'un QTL et de la structure génétique de la population

- Structure génétique =
  - Les animaux appartiennent à des races (rameaux, lignées...) variés
  - Les individus sont apparentés et la population agglomère des familles

---

---

---

---

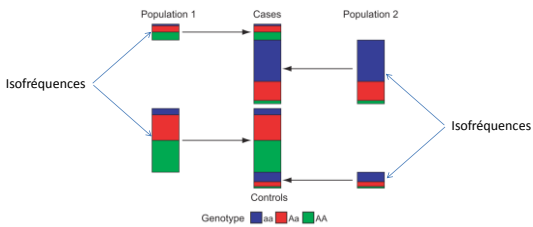
---

---

---

---

### Mélange racial non repéré



Marchini et al, Nat Genet, 2004

---

---

---

---

---

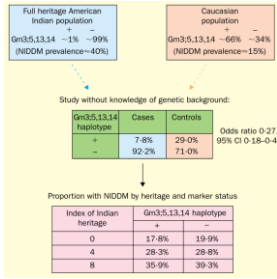
---

---

---

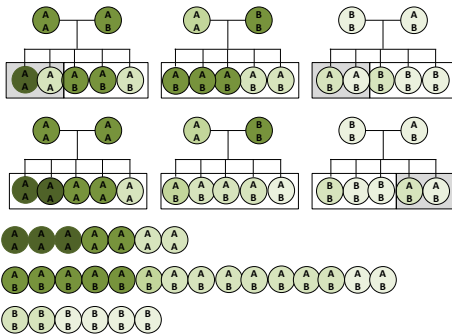
### Mélange racial non repéré

Exemple de la sensibilité au diabète chez les indiens US



Cardon et Palmer, The Lancet, 2003

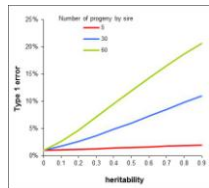
### Structures familiales



### Conséquence : perte de robustesse

Rejet par erreur de l'hypothèse d'absence d'effet d'un SNP du fait de la structuration en famille

Exemple : modèle de régression (Teysseire et al, 2012)



## Prendre en compte la structuration

Structurer en familles

On construit la population pour pouvoir raisonner intra famille, donc « toutes choses égales par ailleurs »

Corriger pour la structure

Modifier ou étalonner la statistique de test

---

---

---

---

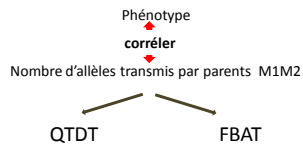
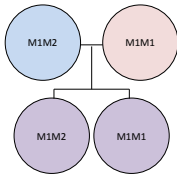
---

---

---

---

## Structurer en familles




---

---

---

---

---

---

---

---

## QTD

Dérivé du TDT de Spielman et al (1993)  
(mesure un déséquilibre de transmission par des parents hétérozygotes A1A2)

$T = \frac{(b-c)^2}{b+c}$	Allèles transmis		Allèles non transmis		Total
	A1	a	b	a+b	
	A2	c	d	c+d	
	a+c	b+d		2n	

---

---

---

---

---

---

---

---

## QTD

Dérivé du TDT de Spielman et al (1993)  
(mesure un déséquilibre de transmission  
par des parents hétérozygotes A1A2)

$T = \frac{(b-c)^2}{b+c}$	Allèles transmis		Allèles non transmis		Total
	A1	A2	a	b	
			c	d	c+d
			a+c	b+d	2n

$$o_i = b \text{ ou } c = \text{nombre observé d'allèles transmis aux cas}$$

$$e_i = \frac{b+c}{2} = \text{nombre espéré si indépendance}$$

$$T = \sum_i \frac{(o_i - e_i)^2}{e_i}$$

---

---

---

---

---

---

---

---

---

---

## QTD

Dérivé du TDT de Spielman et al (1993)  
(mesure un déséquilibre de transmission  
par des parents hétérozygotes A1A2)

$T = \frac{(b-c)^2}{b+c}$	Allèles transmis		Allèles non transmis		Total
	A1	A2	a	b	
			c	d	c+d
			a+c	b+d	2n

Extension de Allison (1997), Fulker et al (1999), Abecassis et al (2000)

$$Y_{ij} = \mu + (P_{ij} - \bar{P})\beta_p + (M_{ij} - \bar{P}_i)\beta_w + e_{ij}$$

Génotype de ij
E(M<sub>ij</sub>/parents)
On teste β<sub>w</sub>=0

---

---

---

---

---

---

---

---

---

---

## FBAT

FBAT est une corrélation entre le phénotype  
et l'écart de transmission intra famille  
Si le marqueur est lié à un QTL  
➡ Les phénotypes extrêmes se trouvent chez les déviants

---

---

---

---

---

---

---

---

---

---



## Corriger pour la structure

Genomic control

Étalonner la stat de test

Ajouter la transmission

Se rapprocher des FBAT

Modéliser les structures

Améliorer le modèle

---

---

---

---

---

---

---

---

---

---

## Le contrôle génomique

Devlin et Roeder (1999) Bacalu et al (2002)

Table 1  
Genotype distribution

	A1 alleles			Total
	0	1	2	
Case	$r_0$	$r_1$	$r_2$	$R$
Control	$n_0$	$n_1$	$n_2$	$S$
Total	$m_0$	$m_1$	$m_2$	$N$

En l'absence de structure, le test Armitage  

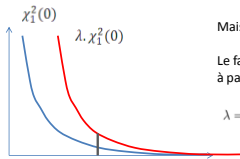
$$T = \frac{N\{N(r_1 + 2r_2) - R(n_1 + 2n_2)\}^2}{R(1-R)\{N(n_1 + 4n_2) - (n_1 + 2n_2)^2\}}$$

suit, sous H0, une loi de  $\chi^2_1(0)$

Mais avec structure,  $T$  suit une  $\lambda \cdot \chi^2_1(0)$

Le facteur d'inflation,  $\lambda$ , est estimable à partir des marqueurs neutres

$$\lambda = \text{Mediane}(T_1, T_2, \dots, T_k) / 0.455$$



Extension à la régression par Bacanu et al (2002)

---

---

---

---

---

---

---

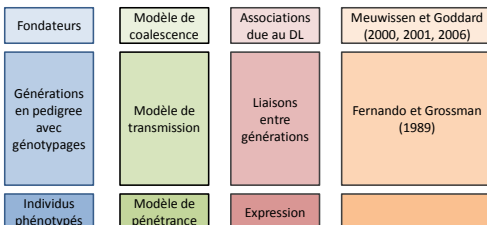
---

---

---

## Ajouter la transmission

Meuwissen et al (2002)




---

---

---

---

---

---

---

---

---

---

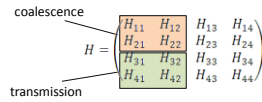
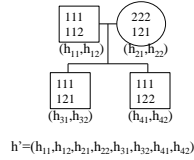
# Ajouter la transmission

Meuwissen et al (2002)

$$Y = X\beta + Z_h h + Zu + e$$

Effets d'haplotypes      Effets individuels

$$V(Y) = Z'_h H Z_h \sigma_h^2 + ZAZ' \sigma_u^2 + I\sigma_e^2$$




---

---

---

---

---

---

---

---

---

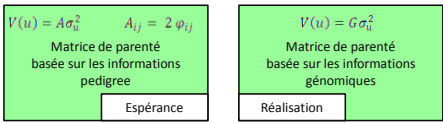
---

# Corriger pour les structures

Modèle mixte (Henderson , 1975 et beaucoup d'autres)

$$Y = X\beta + Ws + u + e$$

Effets de nuisance    Effet SNP    Effet individuel    Résiduelle



Difficulté : Estimer les variances et les effets fixes pour chaque position

---

---

---

---

---

---

---

---

---

---

# Autour du modèle mixte

Accélération

- GRAMMAR (Aulchenko et al, 2007)    step G1 :  $Y = X\beta + u + e \Rightarrow \hat{u}$   
step G2 :  $\hat{u} = Ws + e$
- FASTA (Aulchenko et al, 2011)    step F1 :  $Y = X\beta + u + e \Rightarrow \hat{\sigma}_u^2$   
step F2 :  $Y_{\hat{\sigma}_u^2} = X\beta + Ws + u + e$
- EMMA (Kang et al, 2008)    décomposition spectrale de la matrice  $Z'GZ + \lambda I$
- EMMAX (Kang et al, 2010)    step F1, selon EMMA  
step F2
- TASSEL (Zhang et al, 2010)    step T0 : compression des effets  $u$  par clusterization  
step F1, « P3D »  
step F2

---

---

---

---

---

---

---

---

---

---