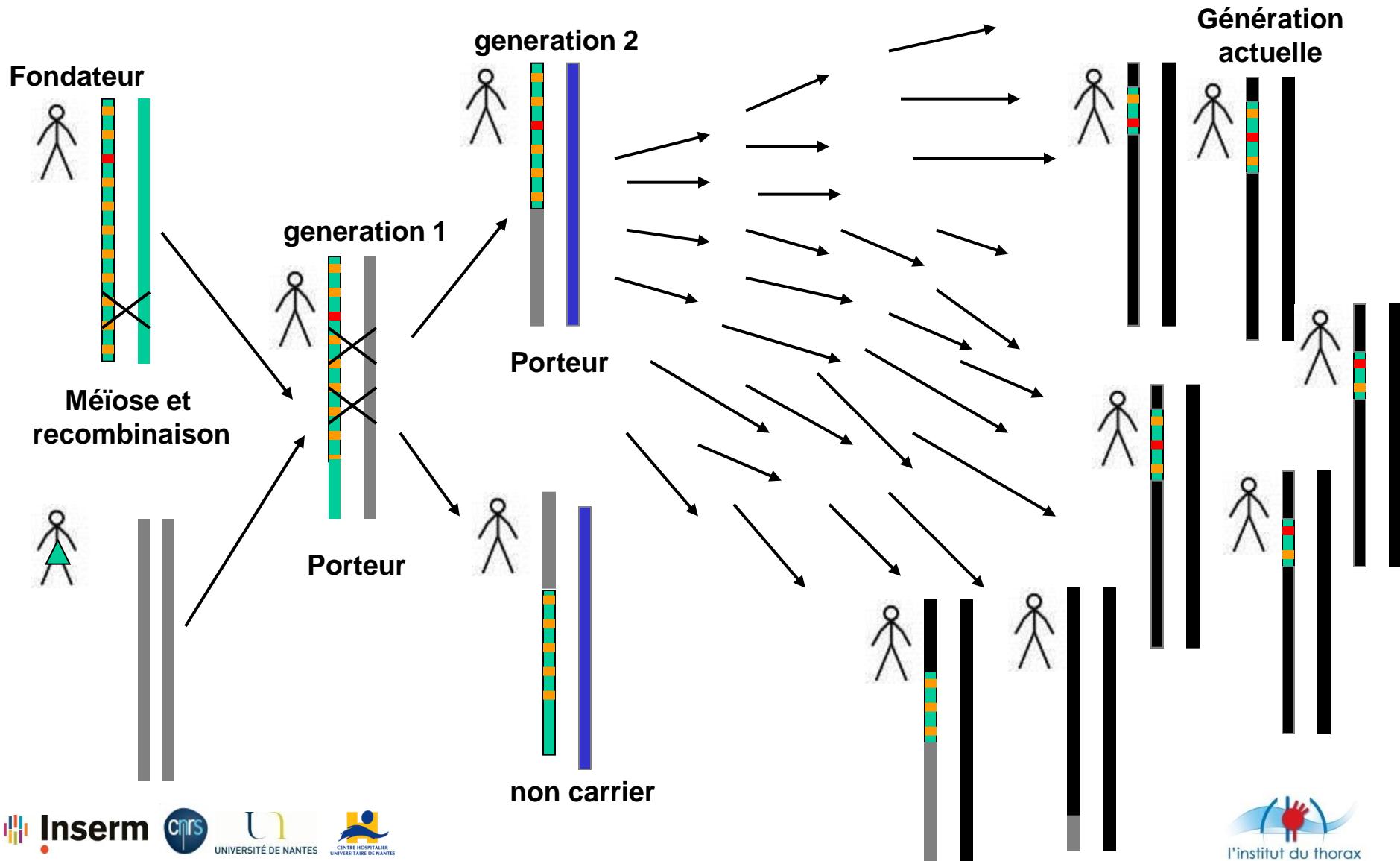


Etudes d'association pangénomiques

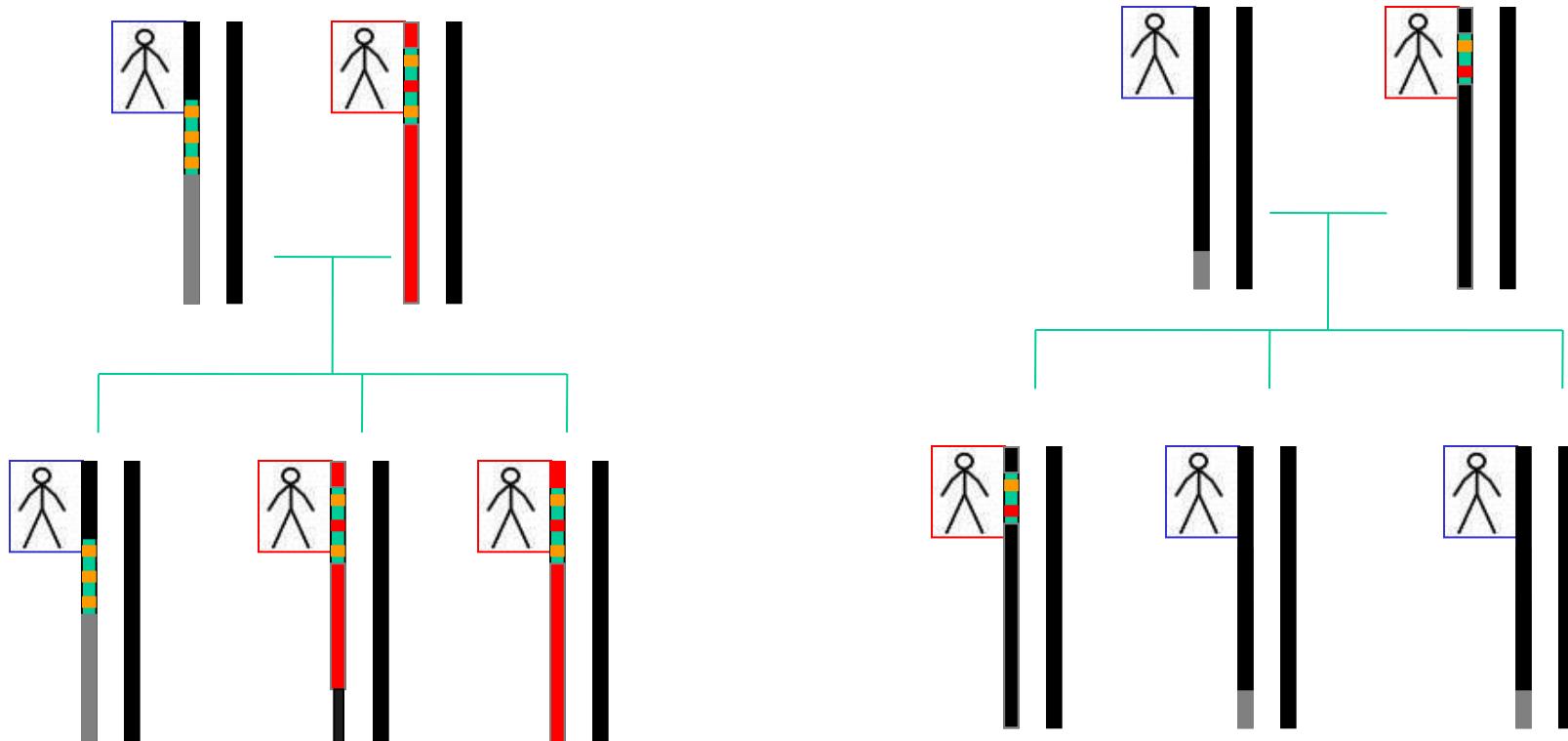


Christian DINA
christian.dina@univ-nantes.fr

Transmission d'un allèle ancien



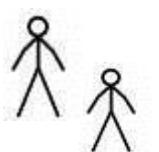
Etude familiale



Peu de recombinations

Cas-Témoins

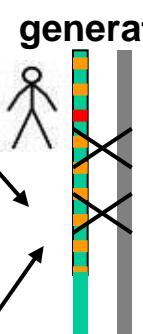
Fondateur



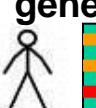
Méiose et
recombinaison



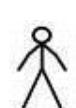
generation 1



generation 2



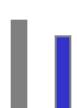
Porteur



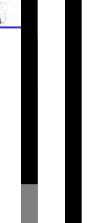
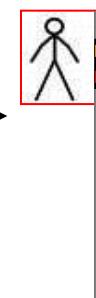
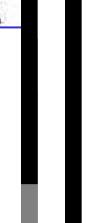
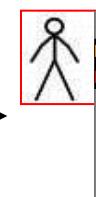
non carrier



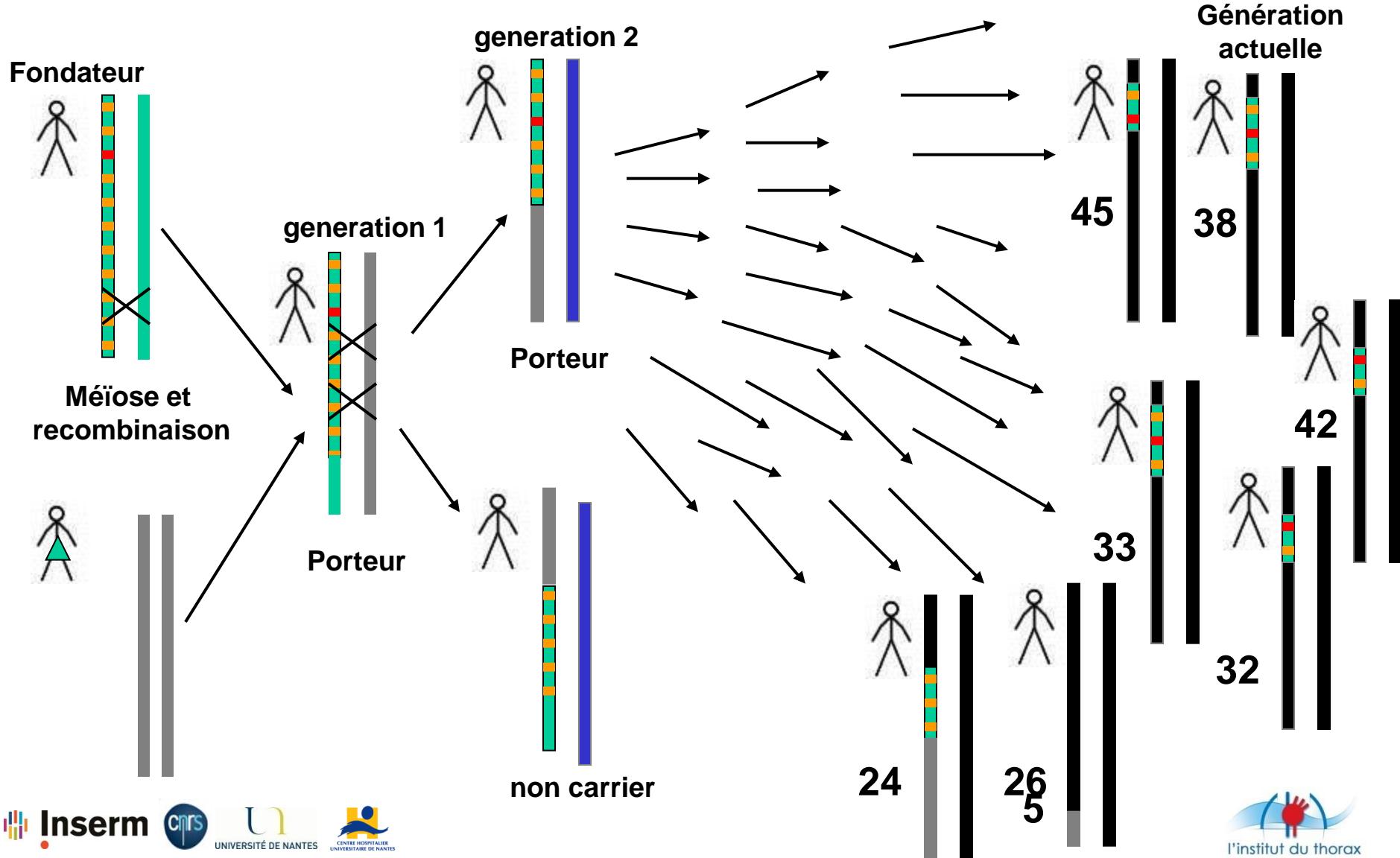
Porteur



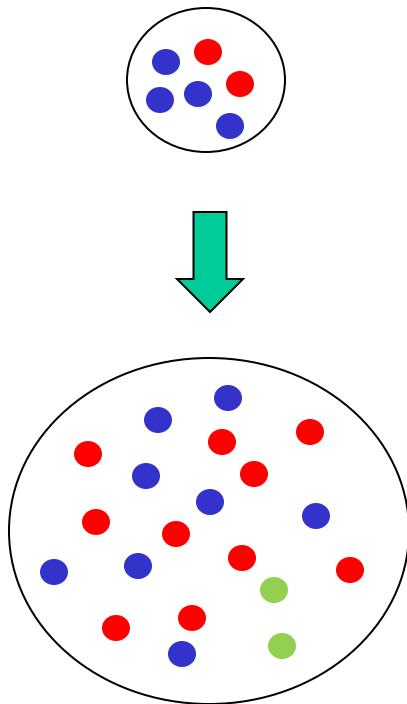
Génération
actuelle



Traits quantitatifs



Common Variant Common Disease



Petite population initiale

Hétérogénéité réduite

Expansion rapide ~ 100.000 y

Effets modestes

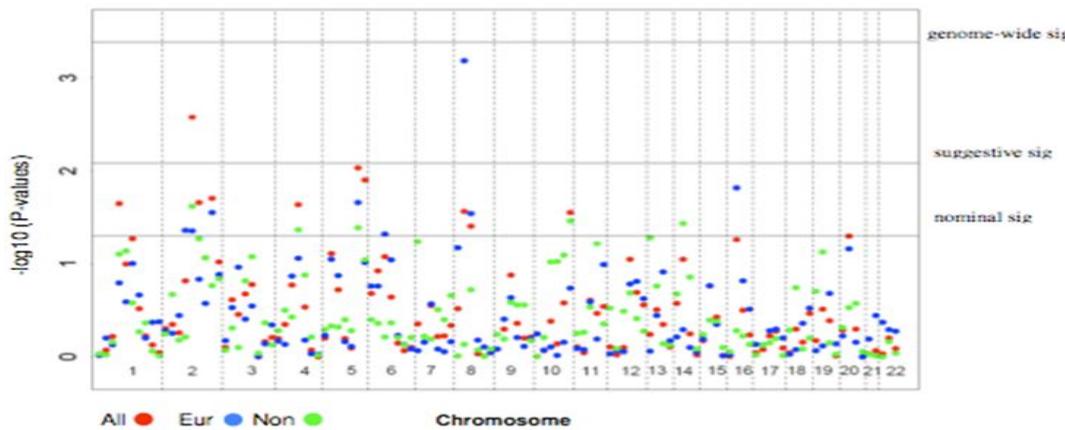
Etudes d'association pangénomiques

Pourquoi ?

Historique

- Analyse génétique = Familles et liaison
- Passage aux traits complexes et fréquents

Linkage studies inconclusive



Meta-analysis of 32 linkage studies in SCZ (3200 pedigrees, 7500 affecteds).

Lewis, Ng, and Levinson, October 2007

8

Hypothèse CD/CV

The New Genomics: Global Views of Biology

Eric S. Lander

On the allelic spectrum of human disease

David E. Reich and Eric S. Lander

Population genetics—making sense
out of sequence

Aravinda Chakravarti

Hypothèse CD/CV

The New Genomics Global Views of Biology

Eric S. Lander

Policy Forum

On the allelic spectrum of human disease

Opinion

David E. Reich and Eric S. Lander

Population genetics—making sense out of sequence

Aravinda Chakravarti

REVIEW

10

Hypothèse CD/CV

The New Genomics: Global Views of Biology

Eric S. Lander

In addition, he provided “tantalizing examples” of common alleles with large effects

APOE : grande proportion de la survenue de la maladie d'Alzheimer

MTHFR : niveaux d'homocystéine

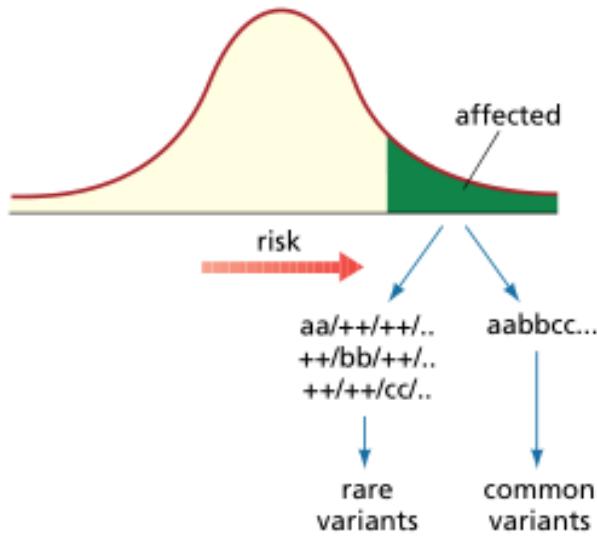
ACE : risque cardiaque

CKR-5 : résistance au VIH

Hypothèse CD/CV

Population genetics—making sense out of sequence

Aravinda Chakravarti



Architecture la plus probable :
- effets « épistatiques »

Si c'est le cas, il faut que chaque allèle soit fréquent.

Et donc il faut que ces variants soient anciens.

Historique

The Future of Genetic Studies of Complex Human Diseases

Neil Risch and Kathleen Merikangas

| Genotypic risk ratio (γ) | Frequency of disease allele A (p) | Probability of allele sharing (Y) | No. of families required (N) | Probability of transmitting disease allele A P(tr-A) | Association | | | |
|-----------------------------|--------------------------------------|--------------------------------------|---------------------------------|---|-------------|--------|-----------|------|
| | | | | | Singletons | | Sib pairs | |
| 4.0 | 0.01 | 0.520 | 4260 | 0.800 | 0.048 | 1098 | 0.112 | 235 |
| | 0.10 | 0.597 | 185 | 0.800 | 0.346 | 150 | 0.537 | 48 |
| | 0.50 | 0.576 | 297 | 0.800 | 0.500 | 103 | 0.424 | 61 |
| | 0.80 | 0.529 | 2013 | 0.800 | 0.235 | 222 | 0.163 | 161 |
| 2.0 | 0.01 | 0.502 | 296,710 | 0.667 | 0.029 | 5823 | 0.043 | 1970 |
| | 0.10 | 0.518 | 5382 | 0.667 | 0.245 | 695 | 0.323 | 264 |
| | 0.50 | 0.526 | 2498 | 0.667 | 0.500 | 340 | 0.474 | 180 |
| | 0.80 | 0.512 | 11,917 | 0.667 | 0.267 | 640 | 0.217 | 394 |
| 1.5 | 0.01 | 0.501 | 4,620,807 | 0.600 | 0.025 | 19,320 | 0.031 | 7776 |
| | 0.10 | 0.505 | 67,816 | 0.600 | 0.197 | 2218 | 0.253 | 941 |
| | 0.50 | 0.510 | 17,997 | 0.600 | 0.500 | 949 | 0.490 | 484 |
| | 0.80 | 0.505 | 67,816 | 0.600 | 0.286 | 1663 | 0.253 | 941 |

Comparison of linkage and association studies. Number of families needed for identification of a disease gene.

Si des variants fréquents à effets modérés sont responsables des pathologies complexes : alors l'analyse d'association est plus puissante que l'analyse de liaison.

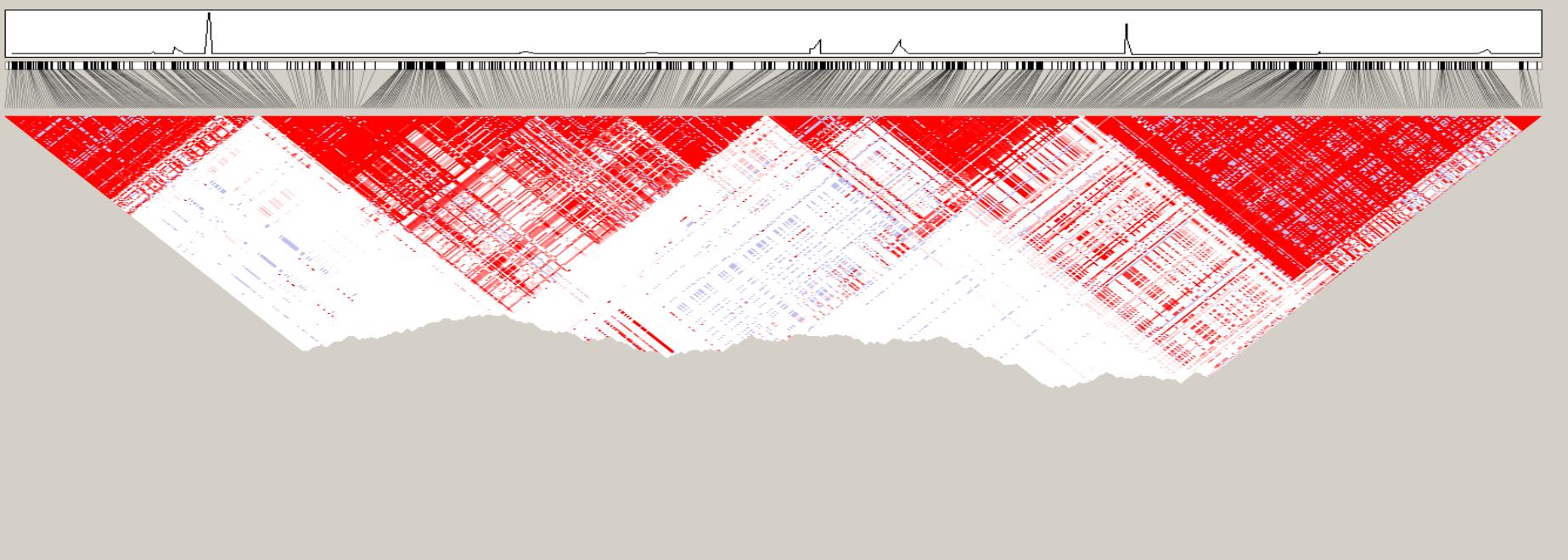
CD/CV - conclusion

- Repose sur l'absence de résultats probands en analyse de liaison
- L'extension rapide de la taille de la population humaine – beaucoup de variants génétiques fréquents
- Si pathologie fréquente, allèles causaux fréquents
- Exemples de variants fréquents à effets forts
- Objectif principal était le fine-mapping

Structure du génome au secours des EAP

Diversité haplotypique réduite

Structure du genome

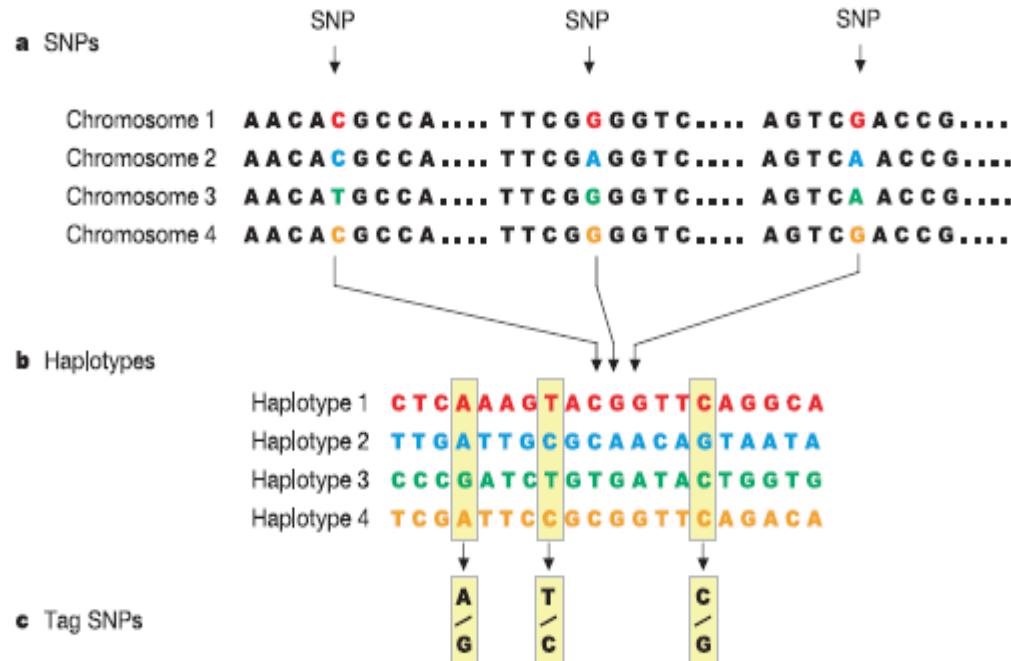


■ Deséquilibre de liaison très fort

Irrégularité de la recombinaison (hot-spots)

Forte corrélation entre SNPs – un SNP « tagge » d'autres SNPs

HapMap principe

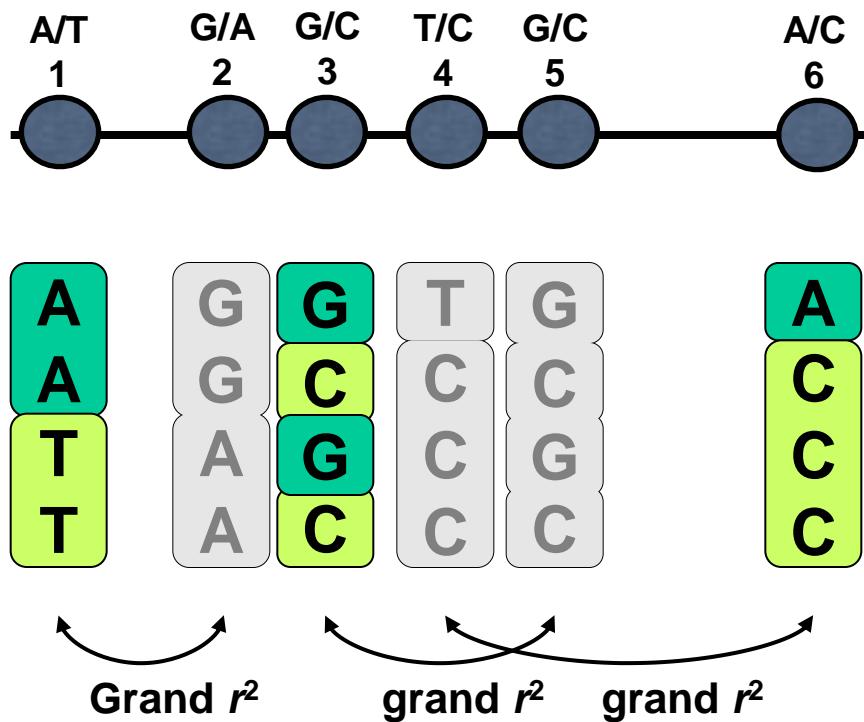


3.1 M SNPs exploitables

On peut représenter 80-90% des SNPs communs avec

~300,000 tag SNPs pour pop. Européennes où Asiatiques
~500,000 tag SNPs pour pop. Africaines

Marquage - Tagging



Tags:

SNP 1
SNP 3
SNP 6

3 SNPs

Tests d'association:

SNP 1
SNP 3
SNP 6

Carlson et al. (2004) AJHG 74:106

HapMap

Permet d'avoir l'information sur un
grand nombre de SNPs

Mais ne génotypant un petit nombre

Catalogue : HapMap 2

- 90 individus Yoruba (30 trios parent-parent-enfant trios) de Ibadan, Nigeria (YRI)
 - 90 individus (30 trios) d'origine Européenne : Utah (CEU)
 - 45 Han Chinois de Beijing (CHB)
 - 45 Japonais de Tokyo (JPT)
 - 3.1 M de SNPs
- 270 au total**

Données 1000 Génomes



| Code | Population | Continent | HM 1 - 2 | HM 3 | 1000 G |
|------|--|-----------|----------|------|--------|
| ASW | Africains Américains, Etats-Unis | Afr | | Oui | Oui |
| CEU | Européens de l'Utah, Etats-Unis | Eur | Oui | Oui | Oui |
| CHB | Chinois à Pékin, Chine | Asi | Oui | Oui | Oui |
| CHD | Chinois à Denver, Etats-Unis | Asi | | Oui | Oui |
| GIH | Indiens Gujarati à Houston, Etats-Unis | Asi | | Oui | Oui |
| JPT | Japonais à Tokyo, Japon | Asi | Oui | Oui | Oui |
| LWK | Luhya à Webuye, Kenya | Afr | | Oui | Oui |
| MXL | Mexican à Los Angeles | Afr | | Oui | Oui |
| MKK | Massai à Kinyawa, Kenya | Afr | | Oui | Oui |
| TSI | Toscans d'Italie | Eur | | Oui | Oui |
| YRI | Yoruba à Ibadan, Nigeria | Afr | Oui | Oui | Oui |
| PEL | Péruviens à Lima | Amr | | | Oui |
| PUR | Porto-Ricains, Porto-Rico | Amr | | | Oui |
| CLM | Colombiens de Medellin, Colombie | Amr | | | Oui |
| GBR | Britanniques, Grande-Bretagne | Eur | | | Oui |
| FIN | Finlandais, Finlande | Eur | | | Oui |
| IBS | Ibériques Espagnols, Espagne | Eur | | | Oui |

Tableau 1 : Populations utilisées dans HapMap et 1000 génomes

Panel de données de référence plus important
Environ 10 M des SNPs communs

Mieux que le tagging ...

... l'imputation

Imputation – Qu'est-ce et que fait-ce ?

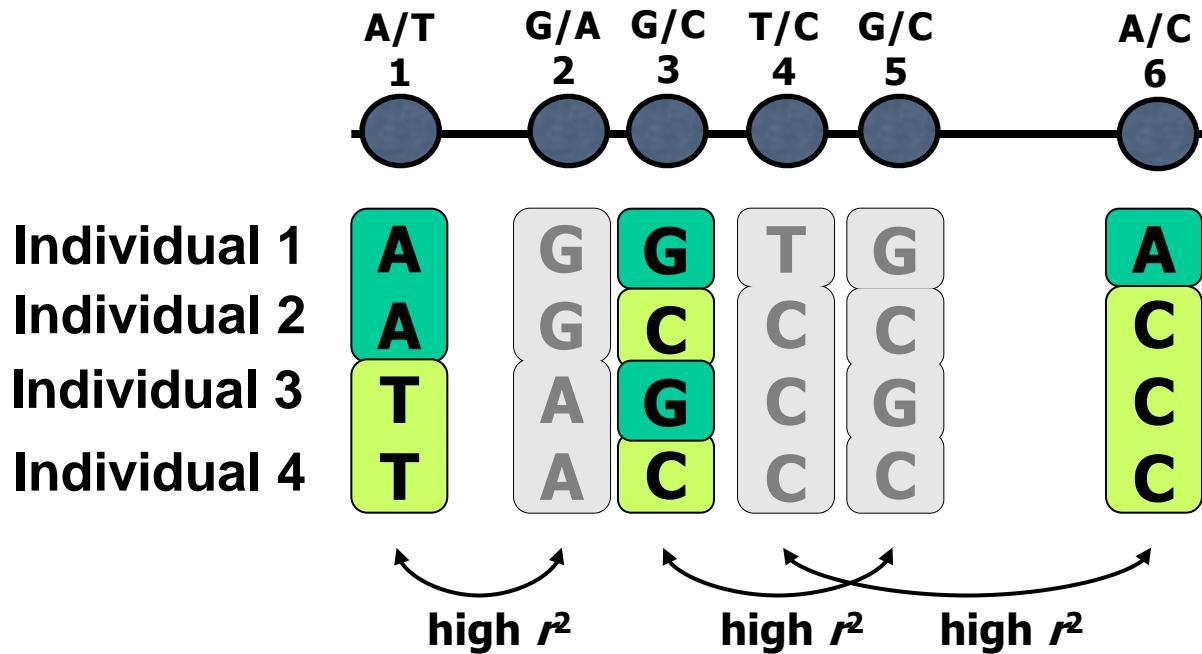
- *Imputation exploits “available information about patterns of correlation among typed and untyped SNPs in a reference panel of densely-genotyped individuals (e.g. HapMap samples) to explicitly predict, or impute, the genotypes at untyped SNPs in a study sample”.*

--Guan et al., *PLoS Genetics*, Dec. 2008

- **Bénéfices de l'Imputation:**

1. Augmente la puissance de détection d'associations
2. Augmente la précision de l'association
3. *Méthode capitale pour combiner les données générées sur des puces différentes.*

Qu'est-ce que l'imputation ?



| Chip | Genotyped (G) or Imputable (I) | | | | | |
|-----------------------|--------------------------------|---|---|---|---|---|
| Omni | G | I | G | I | I | G |
| 660W/610Q / 1M-Duo | | | | | | |
| Affy 6.0 | | | | | | |

**Après
Imputation:**
Les génotypes
pour les 6 SNPs
sur les 3 puces
prêts pour
analyse!

Trois étapes

- Phasage de nos données
- Identification dans le panel de référence
- Imputation proprement dite

Phasage

Observed Genotypes

..... A

..... G

..... C

..... A

Etude

Reference Haplotypes

C G A G A T C T C C T T C T T C T G T G C
C G A G A T C T C C C G A C C T C A T G G
C C A A G G C T C T T T T C T T C T G T G C
C G A A G G C T C T T T T C T T C T G T G C
C G A G A C T C T C C G A C C T T A T G C
T G G G A T C T C C C G A C C T C A T G G
C G A G A T C T C C C G A C C T T G T G C
C G A G A C T C T T T T C T T T T G T A C
C G A G A C T C T C C G A C C T C G T G C
C G A A G G C T C T T T T C T T C T G T G C

HapMap/
1K genomes

- <http://www.shapeit.fr/>, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
- <http://faculty.washington.edu/browning/beagle/beagle.html>
- <http://www.sph.umich.edu/csg/abecasis/MACH/download/>

Gonçalo Abecasis

Identification

Haplotypes de l'étude



Haplotypes de référence

A diagram illustrating multiple reference haplotypes. The first two are highlighted in pink. The last one is highlighted in yellow. The sequence consists of alternating red and green bases.

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|----|---|
| C | G | A | G | A | T | C | T | C | C | T | T | T | C | T | G | T | G | C | | |
| C | G | A | G | A | T | C | T | C | C | C | G | A | C | C | T | C | A | T | GG | |
| C | C | A | A | G | C | T | C | T | T | T | T | C | T | T | C | T | G | T | G | C |
| C | G | A | A | G | C | T | C | T | T | T | T | C | T | T | C | T | G | T | G | C |
| C | G | A | G | A | C | T | C | G | C | G | A | C | C | T | T | A | T | G | C | |
| T | G | G | G | A | T | C | T | C | C | C | G | A | C | C | T | C | A | T | GG | |
| C | G | A | G | A | T | C | T | C | C | C | G | A | C | C | T | T | G | T | G | C |
| C | G | A | G | A | C | T | C | T | T | T | T | C | T | T | T | G | T | A | C | |
| C | G | A | G | A | C | T | C | T | C | C | G | A | C | C | T | C | G | T | G | C |
| C | G | A | A | G | C | T | C | T | T | T | C | T | T | T | C | T | G | T | G | C |

- <http://www.shapeit.fr/>, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
- <http://faculty.washington.edu/browning/beagle/beagle.html>
- <http://www.sph.umich.edu/csg/abecasis/MACH/download/>

Imputation

Observed Genotypes

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| c | g | a | g | A | t | c | t | c | c | c | g | A | c | c | t | c | A | t | g | g |
| c | g | a | a | G | c | t | c | t | t | t | C | t | t | t | t | c | A | t | g | g |

Reference Haplotypes

| | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | G | A | G | A | T | C | T | C | C | T | T | C | T | T | C | T | G | T | G | C | |
| C | G | A | G | A | T | C | T | C | C | C | G | A | C | C | T | C | A | T | G | G | |
| C | C | A | A | G | C | T | C | T | T | T | T | C | T | T | C | T | G | T | G | C | |
| C | G | A | A | G | C | T | C | T | T | T | T | C | T | T | C | T | G | T | G | C | |
| C | G | A | G | A | C | T | C | T | C | C | G | A | C | C | T | T | A | T | G | C | |
| T | G | G | G | A | T | C | T | C | C | C | G | A | C | C | T | C | A | T | G | G | |
| C | G | A | G | A | T | C | T | C | C | C | G | A | C | C | T | T | G | T | G | C | |
| C | G | A | G | A | C | T | C | T | T | T | T | C | T | T | T | T | G | T | A | C | |
| C | G | A | G | A | C | T | C | T | C | C | G | A | C | C | T | C | G | T | G | C | |
| C | G | A | A | G | C | T | C | T | T | T | T | C | T | T | T | C | T | G | T | G | C |

- <http://www.shapeit.fr/>, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
- <http://faculty.washington.edu/browning/beagle/beagle.html>
- <http://www.sph.umich.edu/csg/abecasis/MACH/download/>

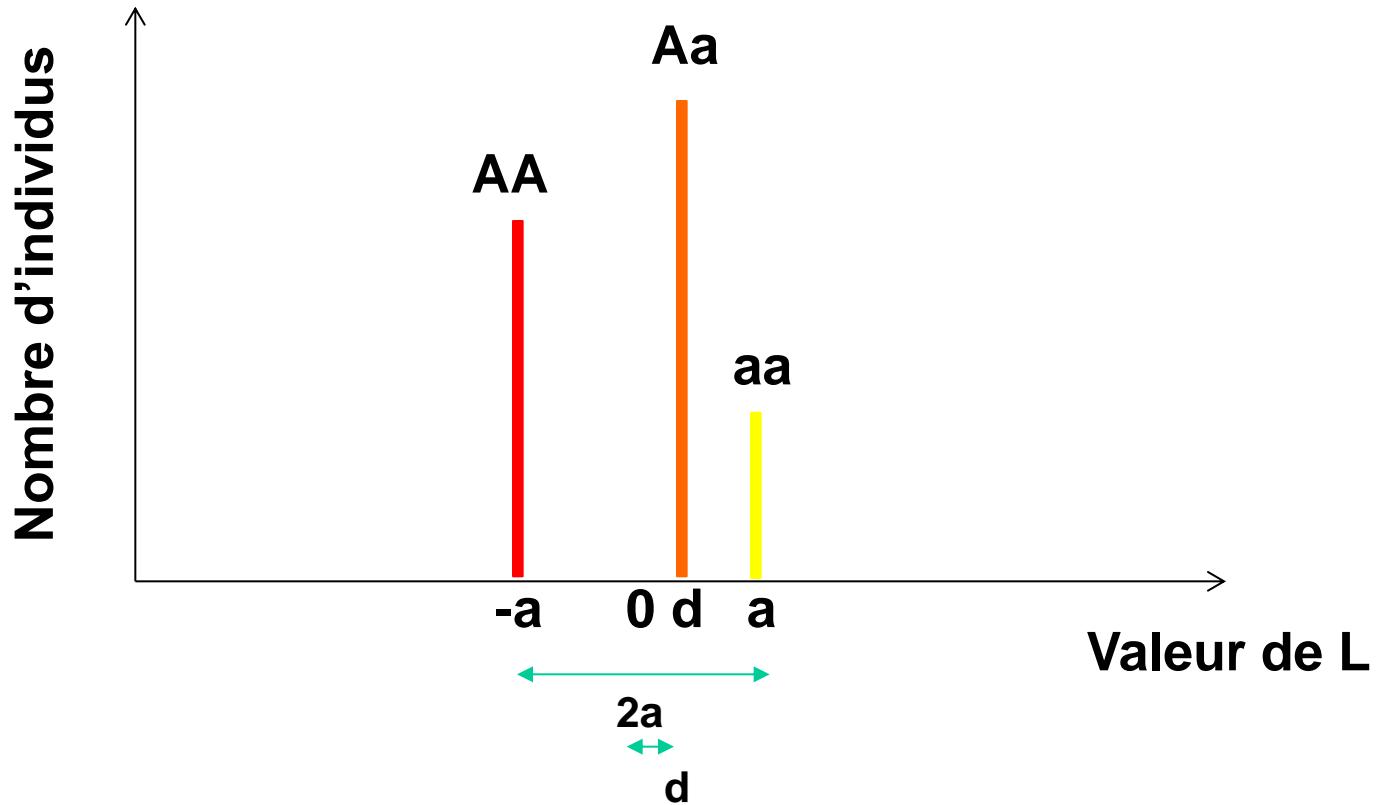
Conséquences

- Les SNPs imputés ne sont pas « certains ». On a des probabilités
- Si le vrai génotype est AG, on peut avoir :
 - $P(AA) = 0.05$
 - $P(AG) = 0.85$
 - $P(GG) = 0.10$
- Incertitude doit être prise en compte dans la statistique d'association

L'héritabilité

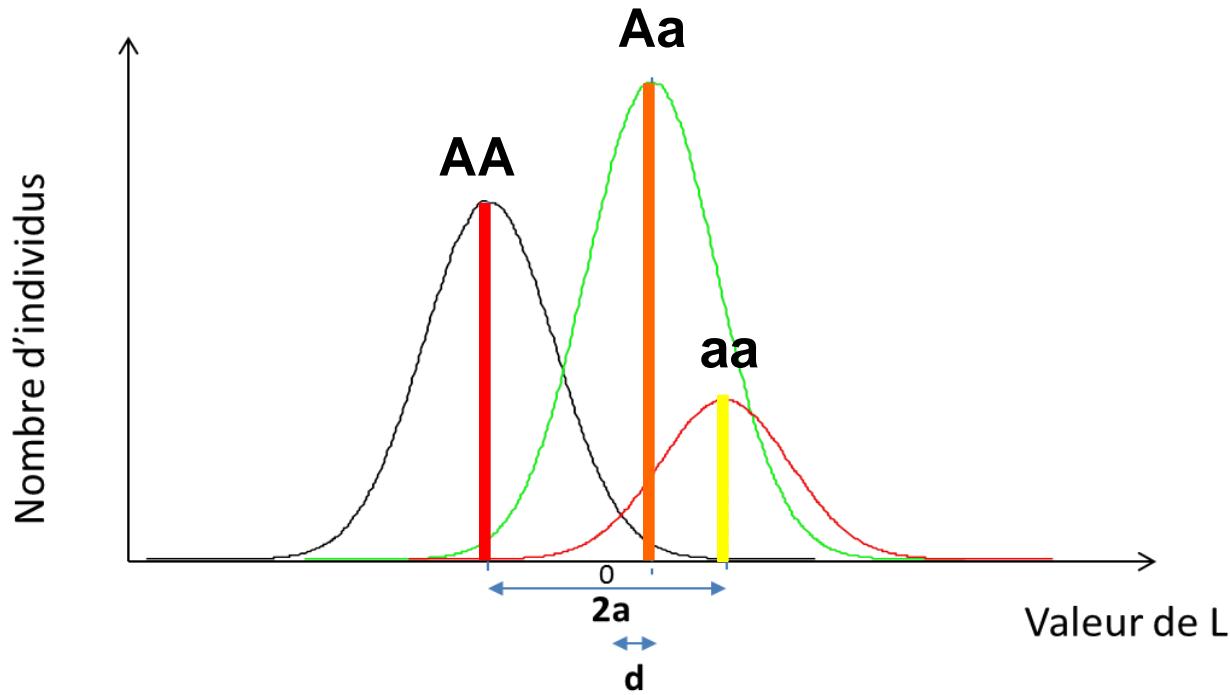
Le but à atteindre

Effet d'un variant génétique



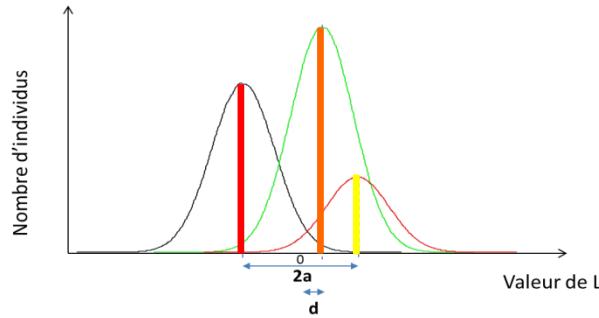
Effet génétique seul

Effet d'un variant génétique

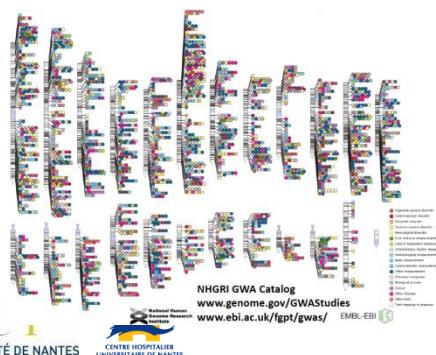


L'héritabilité est la proportion de la variance expliquée par des facteurs génétiques

Objectif de cette section



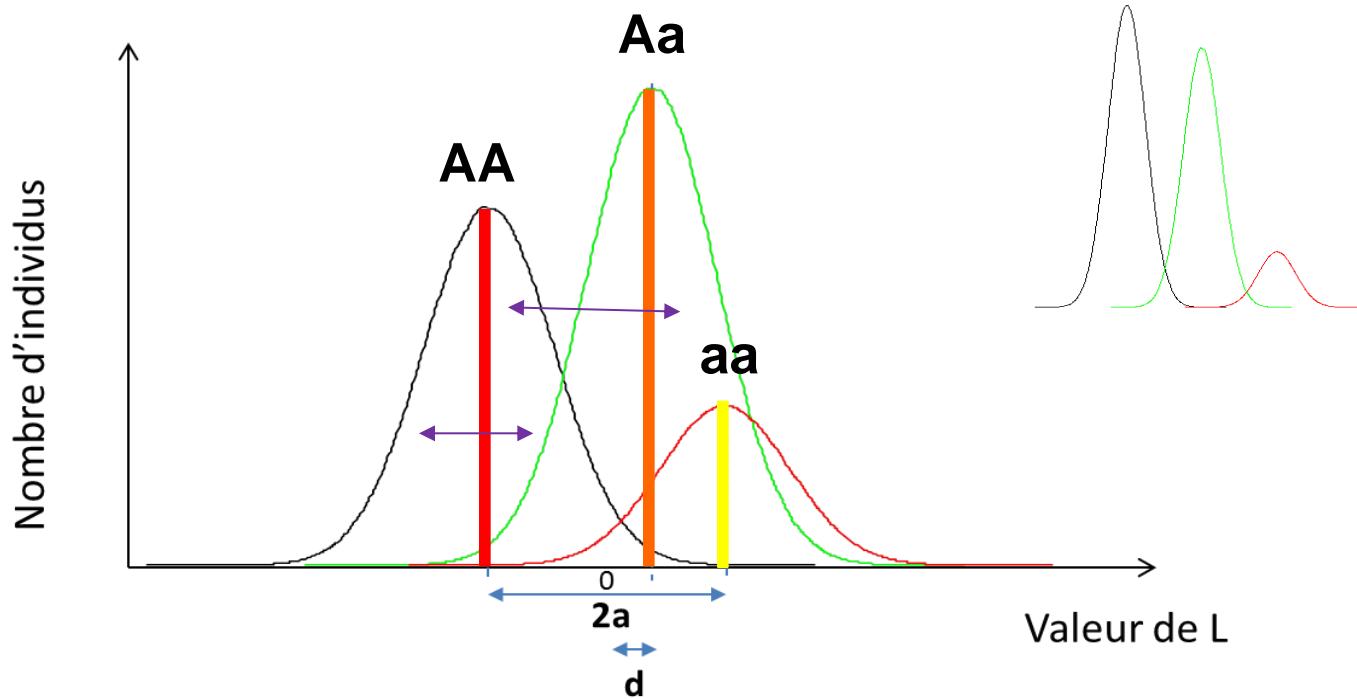
Héritabilité / Importance des gènes dans un phénotype



Que reste-t-il à étudier ?

Effet d'un variant

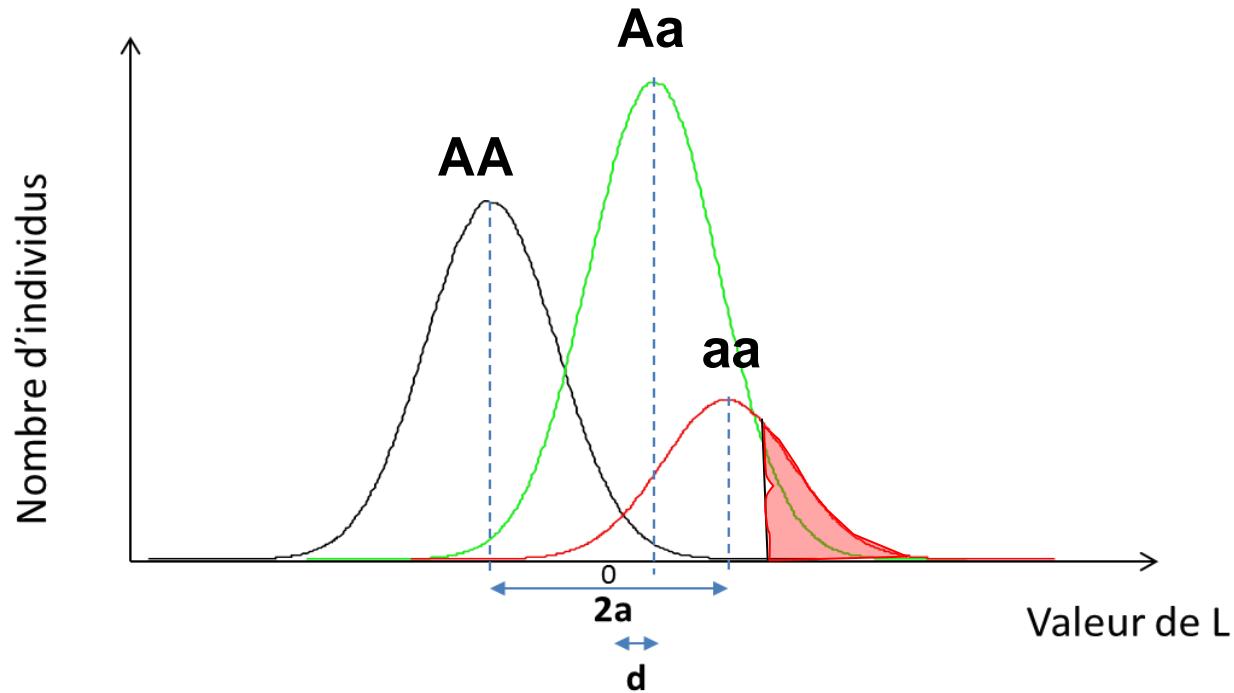
$$h^2 = 80\%$$



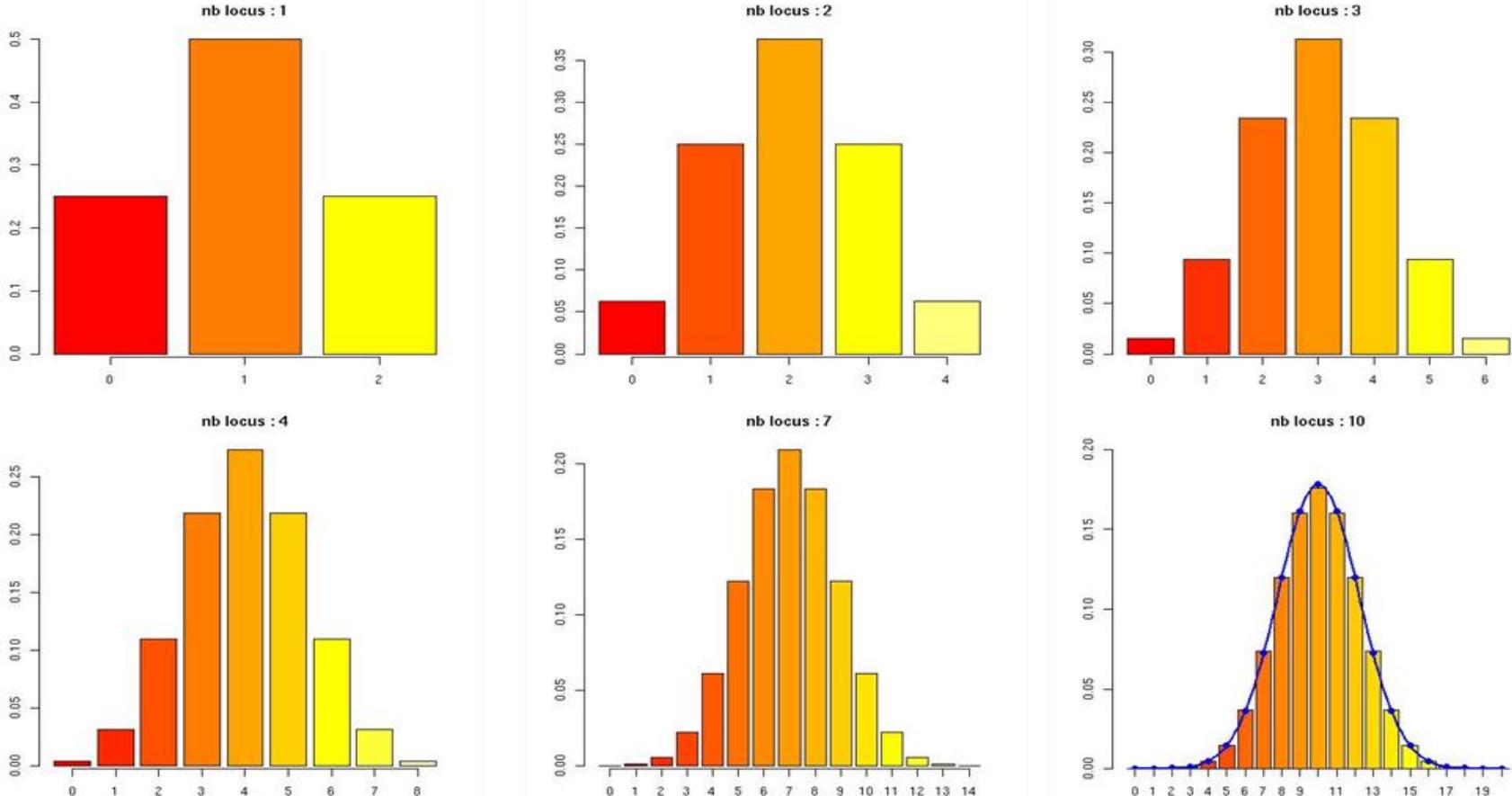
Héritabilité est la proportion de la variance expliquée par des facteurs génétiques

$$h^2 = V_{gen} / V_p$$

Effet d'un variant – Phénotype binaire



Composante polygénique



Hypothèse d'effets additifs

36

Moyenne et Variance

| | AA | Aa | aa | Population |
|----------------------------|-------------|------------|------------|--------------------------------------|
| Fréquence | q^2 | $2pq$ | p^2 | |
| Moyenne / Génotype | -a | d | +a | |
| Contribution à la moyenne | μ_{AA} | μ_{Aa} | μ_{aa} | |
| | $-aq^2$ | $2dpq$ | ap^2 | $M = a(p - q) + 2pqd$ |
| Contribution à la variance | $(-a^2)q^2$ | $2pqd^2$ | $(a^2)p^2$ | $V = 2pq[a + d(p - q)]^2 + [2pqd]^2$ |

Tableau 1 : modélisation d'un phénotype

$$\text{Var}(L) = (a-m)^2 p^2 + (d-m)^2 2pq + (-a-m)^2 q^2$$

$$= 2pq[a+(q-p)d]^2 + (2pqd)^2$$

$$= V_A \quad + \quad V_D$$

But

Nous voulons, à partir des variations entre apparentés d'un trait quantitatif, déduire les différentes variances définies au cours précédent.

Estimation de la variance génétique et de l'héritabilité.

Compréhension et utilisation du modèle biométrique et de la décomposition de la variance.

Outils

Nous pouvons utiliser plusieurs mesures de ressemblance entre apparentés :

$$\begin{aligned} Cov(X, Y) &= E(X - \mu_X)(Y - \mu_Y) = \frac{\sum_i (x_i - \mu_X)(y_i - \mu_Y)}{n-1} \\ &= \sum_i (x_i - \mu_X)(y_i - \mu_Y)f(x_i, y_i) \end{aligned}$$

- corrélation :

- coefficient de corrélation : $\text{cov}(X, Y)/(\text{var}X \cdot \text{var}Y)$
- coefficient de la régression : $\text{cov}(X, Y)/\text{var}X$
ou
 $\text{cov}(X, Y)/\text{var}Y$

Covariance gène majeur bi-allélique

Contribution du locus à Cov (X, Y)

| $Cov(X, Y) = \sum_i (x_i - \mu_X)(y_i - \mu_Y)f(x_i, y_i)$ | | | |
|--|---------------|-------------|--|
| $AA (a-m)$ | $Aa (d-m)$ | $aa (-a-m)$ | |
| $(a-m)^2$ | | | |
| $(a-m)(d-m)$ | $(d-m)^2$ | | |
| $(a-m)(-a-m)$ | $(d-m)(-a-m)$ | $(-a-m)^2$ | |

Jumeaux Monozygotes

Contribution du gène majeur Cov (X,Y) – jumeaux MZ

$$Cov(X, Y) = \sum_i (x_i - \mu_X)(y_i - \mu_Y) f(x_i, y_i)$$

| | AA (a-m) | Aa (d-m) | aa (-a-m) |
|------------------|-----------------|-----------------|------------------|
| AA (a-m) | $p^2(a-m)^2$ | | |
| Aa (d-m) | $0 (a-m)(d-m)$ | $2pq(d-m)^2$ | |
| aa (-a-m) | $0 (a-m)(-a-m)$ | $0 (d-m)(-a-m)$ | $q^2(-a-m)^2$ |

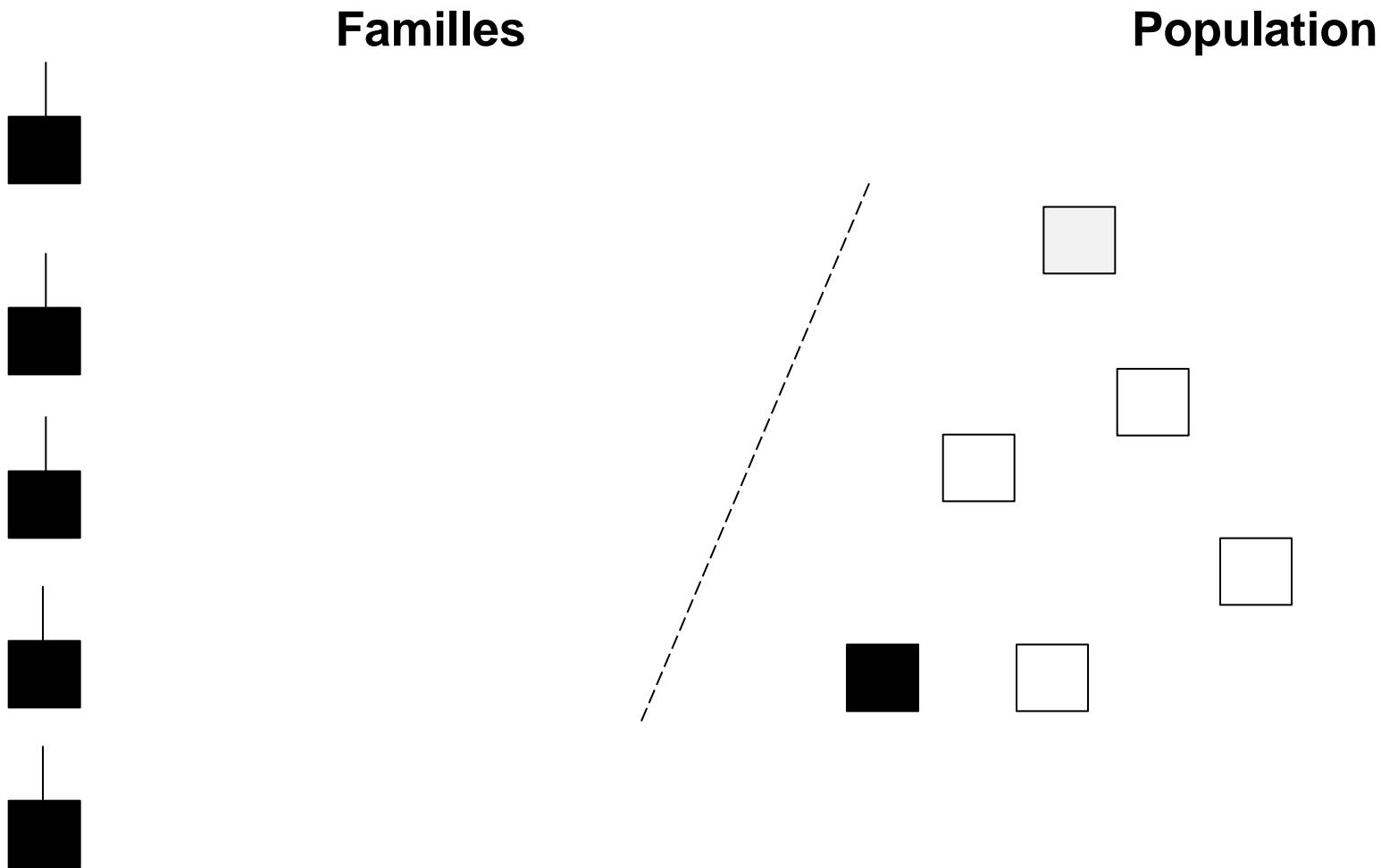
$$\text{Covar } (X_i, X_j) = (a-m)^2 p^2 + (d-m)^2 2pq + (-a-m)^2 q^2$$

$$= 2pq[a+(q-p)d]^2 + (2pqd)^2 \quad = V_A + V_D$$

Héritabilité – Conclusion

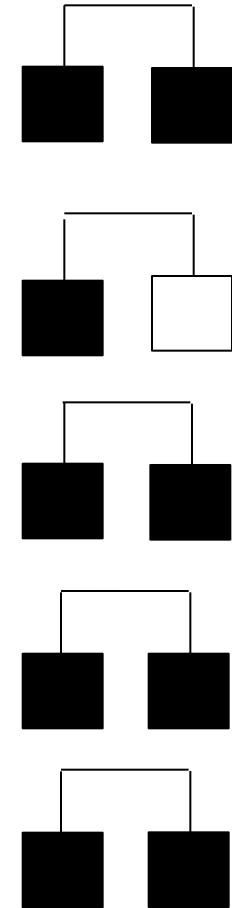
- On peut estimer l'héritabilité à partir de ;
 - Variation phénotypique entre individus
 - Variation génétique entre individus
 - Ici par exemple entre deux jumeaux
- On peut estimer l'héritabilité :
 - À partir des polymorphismes génétiques, si on les connaît – Régression linéaire sur tous les variants génétiques

Risque Relatif d'un apparenté



Risque Relatif d'un apparenté

Familles

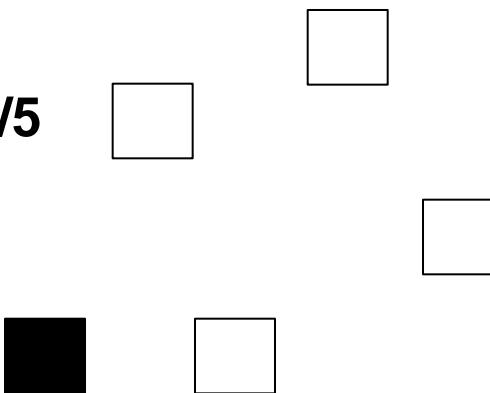


Population

4/5

1/5

$$\lambda_s = 4/1$$



Héritabilité, c'est relatif

- Alcaptonurie
 - accumulation d'acide homogentisique (AHG) et de son produit d'oxydation (acide benzoquinone acétique, ABQA)
 - coloration foncée des urines après émission ... une maladie articulaire sévère avec atteinte axiale et périphérique (arthropathie ochronotique).
 - La prévalence est estimée à 1-9/1 000 000.
- 1898, Archibald Garrod (erreurs innées du métabolisme)
- Gène et mutations trouvées en 1992 (liaison) et 1995.
- régimes réduits en protéines - spécialement en acides aminés, phénylalanine (trouvé dans l'aspartame) et tyrosine.
Héritabilité se trouve réduite.

Tests Multiples

Test d' une hypothèse

- On définit H_0
- On choisit une statistique
- On détermine une région de rejet

| | H_0 acceptée | H_0 rejetée |
|--------------|----------------|---------------|
| H_0 vraie | $1-\alpha$ | α |
| H_0 fausse | β | $1-\beta$ |

En général, α est fixé à 0.05

Pourquoi est-ce un problème ?

- Si on a m tests, $m \times \alpha$ vont être rejetés par hasard:
 - 100 tests à 0.05 : 5 faux positifs
- Pour les GWAs, m est en général très haut 1M
 - $1M \times 0.05 = 50\,000$
- On ne peut se permettre 50 000 résultats positifs même si, par exemple, 1000 sont des vrais positifs

Méthodes standard

- Vraie p-value

$$p_j^* = 1 - (1 - p_j)^m$$

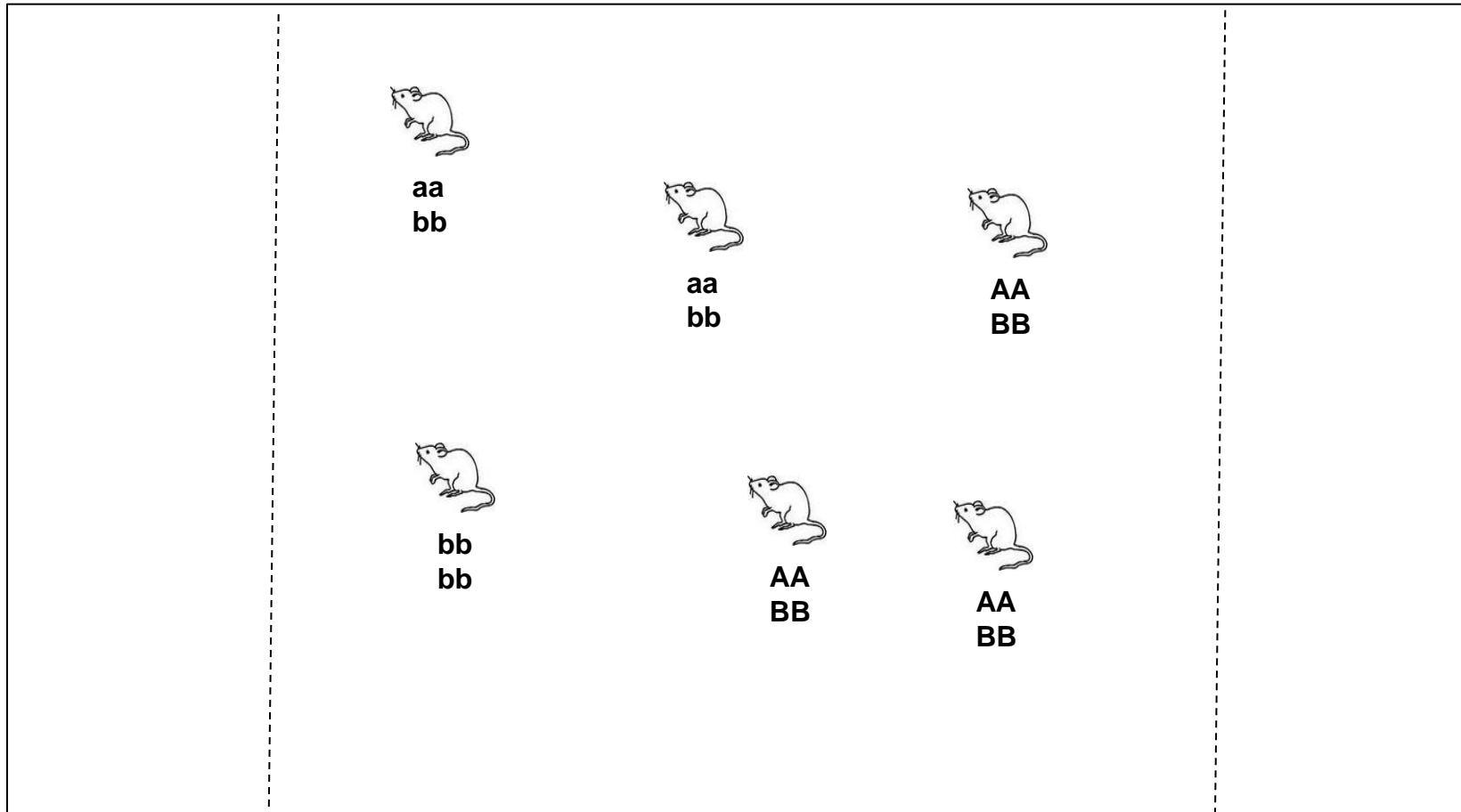
m étant le nombre de tests

On doit prendre un $p_j = 0.05/m$

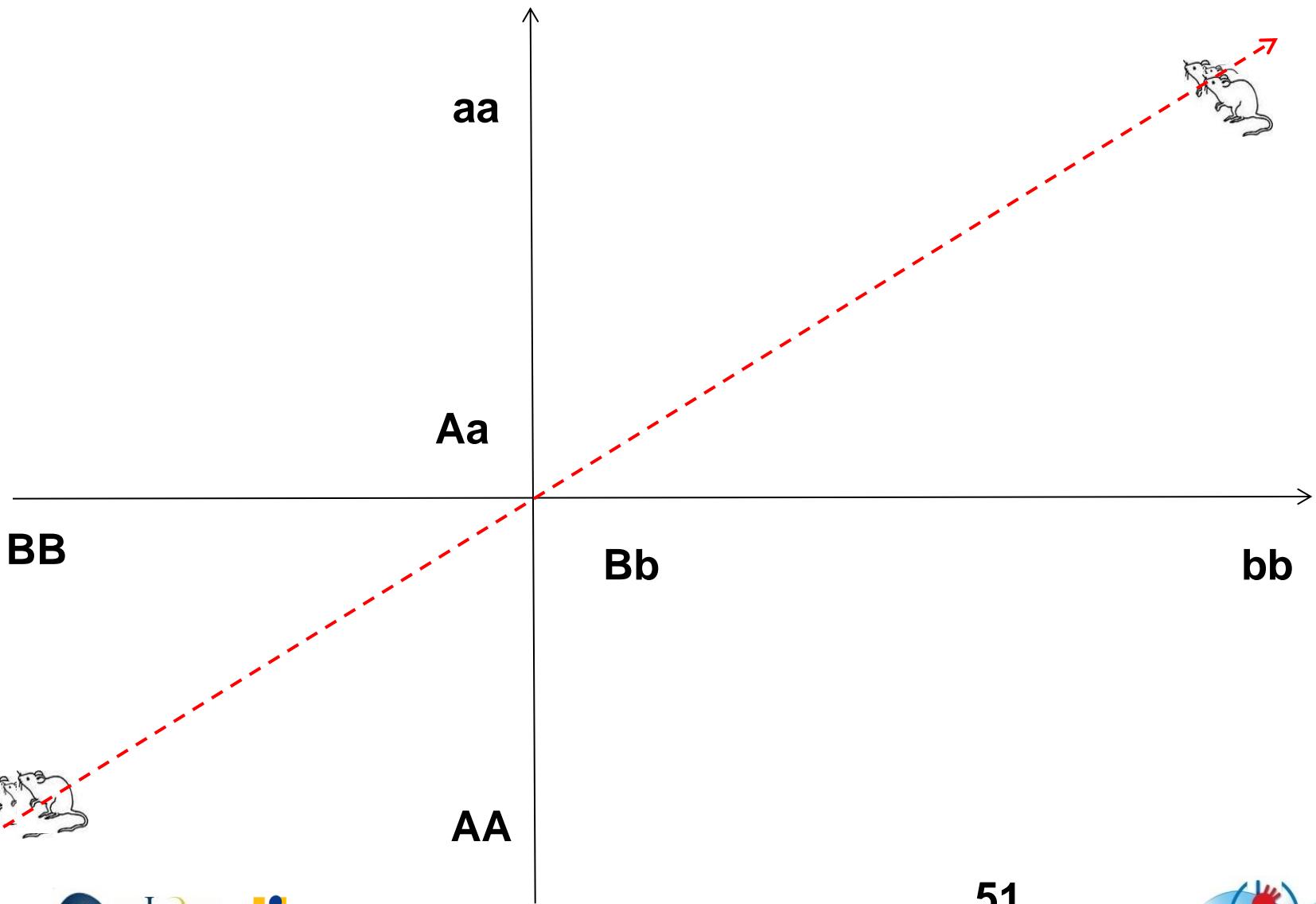
- Appelé Family wise Error Rate (FWER)
- Dans les études d'association pangénomiques :
 - 5×10^{-8} (Prokopenko et al., Aulchenko et al) 1 M
 - 5×10^{-7} (Sabatti et al.) déterminé empiriquement

Principe

n^{eme} génération – absence de Aa



2 loci



Décomposition Spectrale

M SNPs, en DL sont testés pour association.

Les valeurs propres de la matrice de correlation des SNPs sont calculées

$$\lambda_1, \dots, \lambda_M$$

- on calcule $\text{Var}(\lambda) = \sum_{i=1}^M (\lambda_i - 1)^2 / (k-1)$

Cheverud 2001
Nyholt 2004

- Le Nombre de tests efficaces est :

$$M_{\text{eff}} = 1 + (M-1) \left(1 - \frac{\text{var}(\lambda)}{M} \right)$$

Rem : en cas d'indépendance, toutes les valeurs propres sont égales à 1

- on utilise Bonferroni avec M_{eff}

Pour les EAP

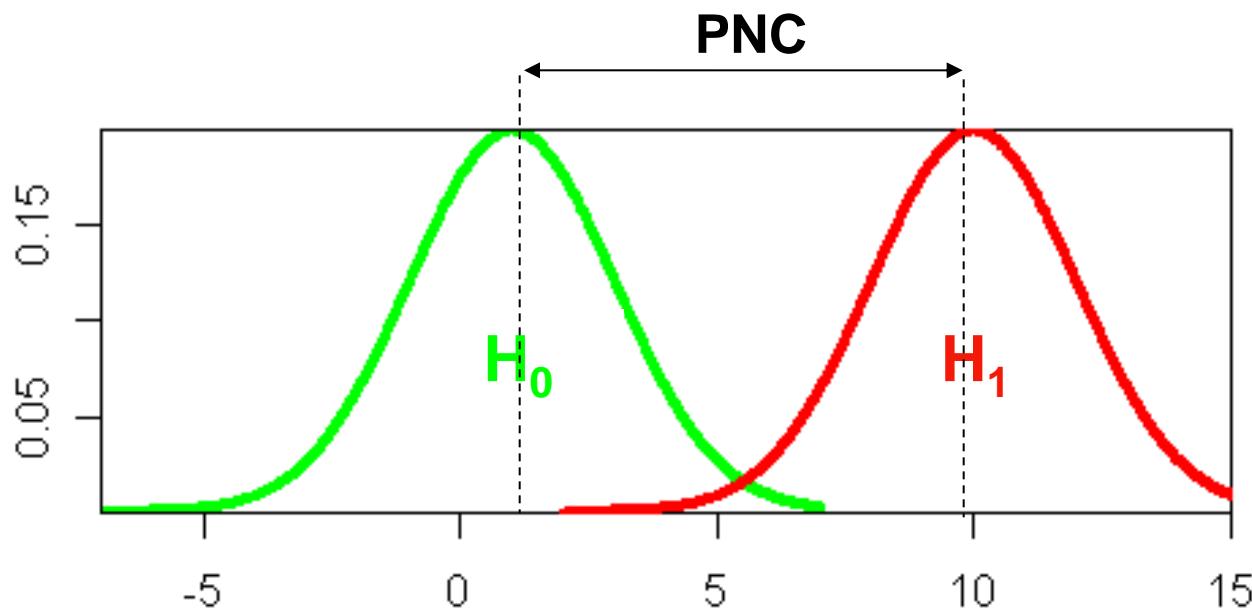
- p-valeur seuil était estimée à $7 \cdot 10^{-8}$ (DUDBRIDGE and GUSNANTO 2008) pour un ensemble infiniment dense de SNPs,
- à 10^{-7} pour les variants fréquents ($MAF > 5\%$) (PE'ER *et al.* 2008) dans une population Européenne.

Mais en final, il y a un consensus pour 1 M de tests indépendents – $p = 5 \cdot 10^{-8}$.

Devrait être population spécifique

Puissance statistique

Paramètre de non centralité



χ^2 centré

χ^2 non centré

Moyenne (μ)

df

df + PCN

Variance (σ^2)

$2*(df)$

$2*(df) + 4*PCN$

Puissance

$$\mathbb{E}[\chi^2] \propto N\gamma^2 p(1 - p)r^2,$$

↓

Effet (RR)

Taille de l'échantillon

↑

Déséquilibre de liaison

Le principe des analyses

Stratégie



3 000 000 000 bases
dans le génome humain



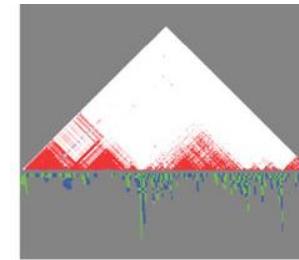
10 000 000 loci à
allèle fréquent



300 000 à 1 000 000
variants testés

SNP
↓
AACACGCCA... TTCTGGGTC... AGTCGACCG...
AACACGCCA... TTCTGGGTC... AGTCACCG...
AACATGCCA... TTCTGGGTC... AGTCACCG...
AACACGCCA... TTCTGGGTC... AGTCGACCG...

Haplotype 1 CTCAAAGTACGGTTCAAGGCA
Haplotype 2 TTGATTGCGCAACAGTAATA
Haplotype 3 CCCGATCTGTGATAACTGGTG
Haplotype 4 TCGATTCCGC GGTTCAAGACA



c Tag SNPs

↓
A / G
↓
T / C
↓
C / G

Séries de cas

Séries de témoins



Tests d'association

Etude d'Association Pangénomique

Patients



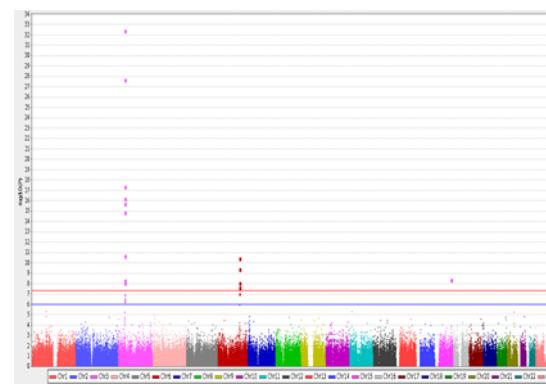
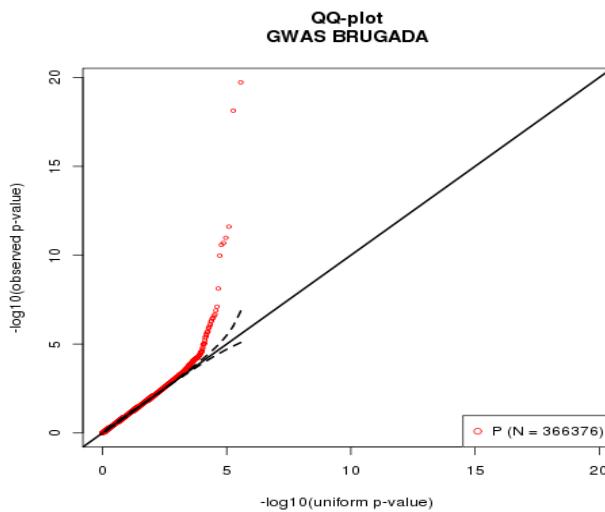
Témoins



$$f(G_{\text{patients}}) - f(G_{\text{témoins}}) = S \sim P(x > S / H_0)$$

force de l'association

Régression
Logistique
(Linéaire)



Etude d'Association Pangénomique

Patients

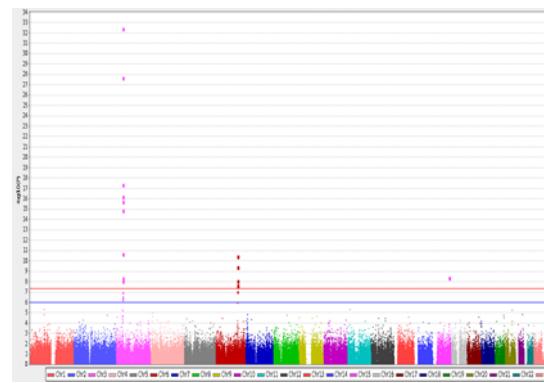
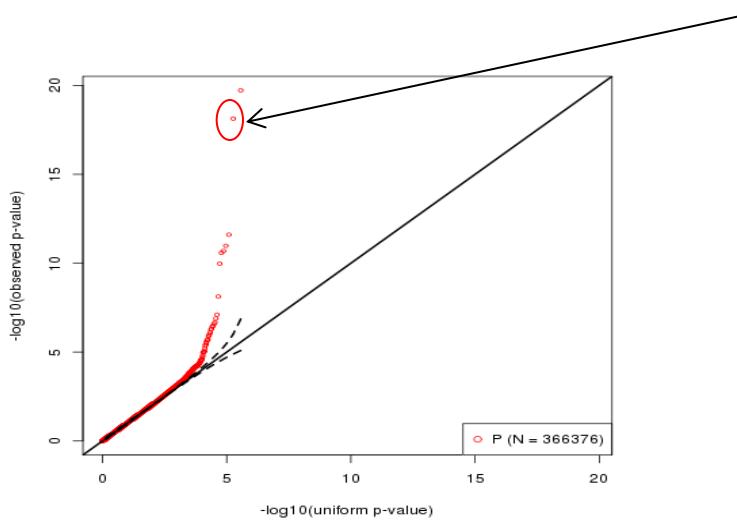


Témoins



$$f(G_{\text{patients}}) - f(G_{\text{témoins}}) = S \sim P(x > S / H_0)$$

force de l'association



Etude d'Association Génome Entier

EAGE

Patients

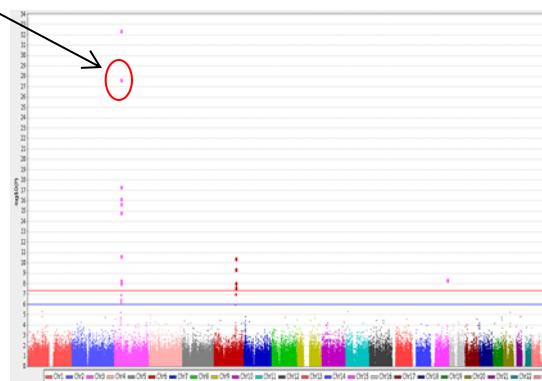
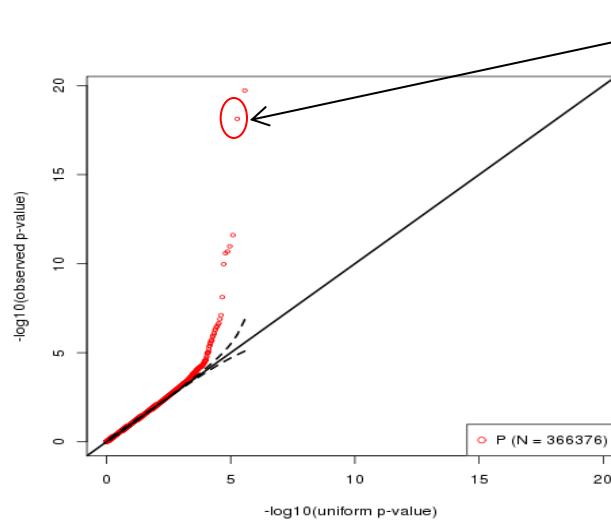


Témoins



$$f(G_{\text{patients}}) - f(G_{\text{témoins}}) = S \sim P(x > S / H_0)$$

force de l'association

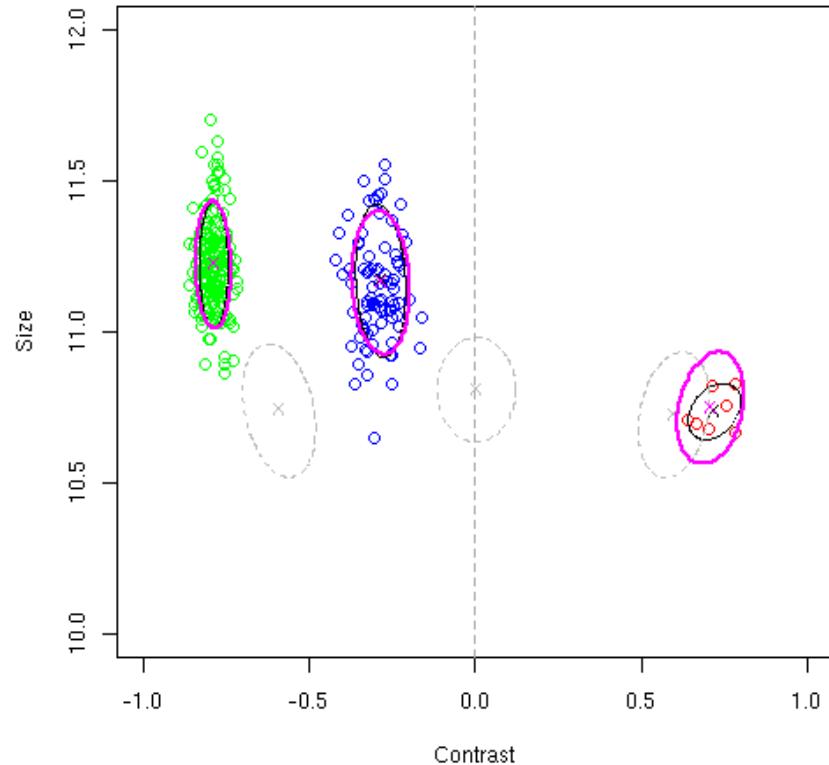


Et le génotypage

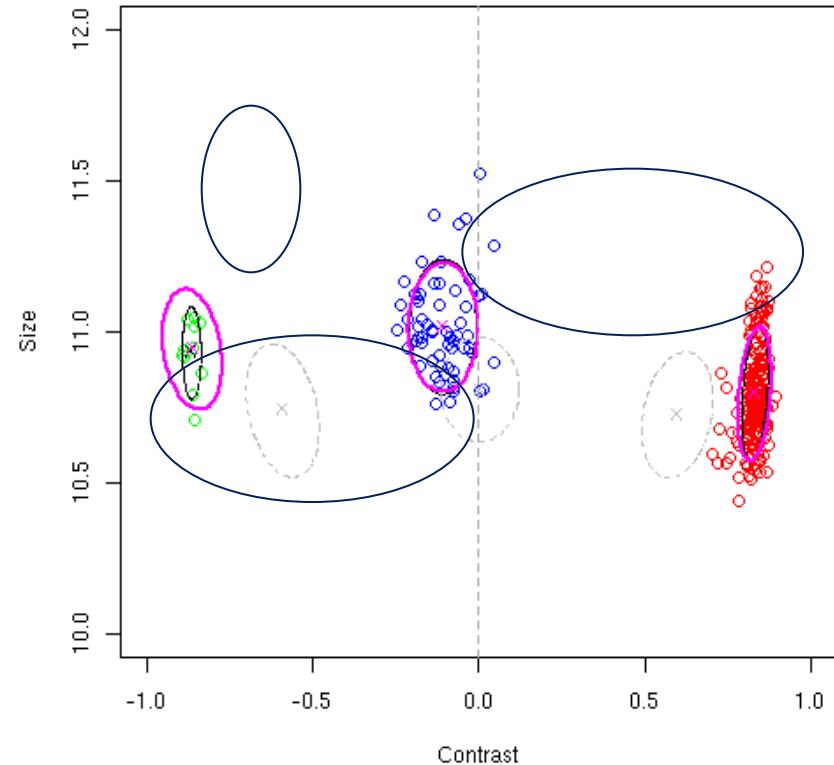
Impact de la chimie sur la statistique

Algorithme

301103



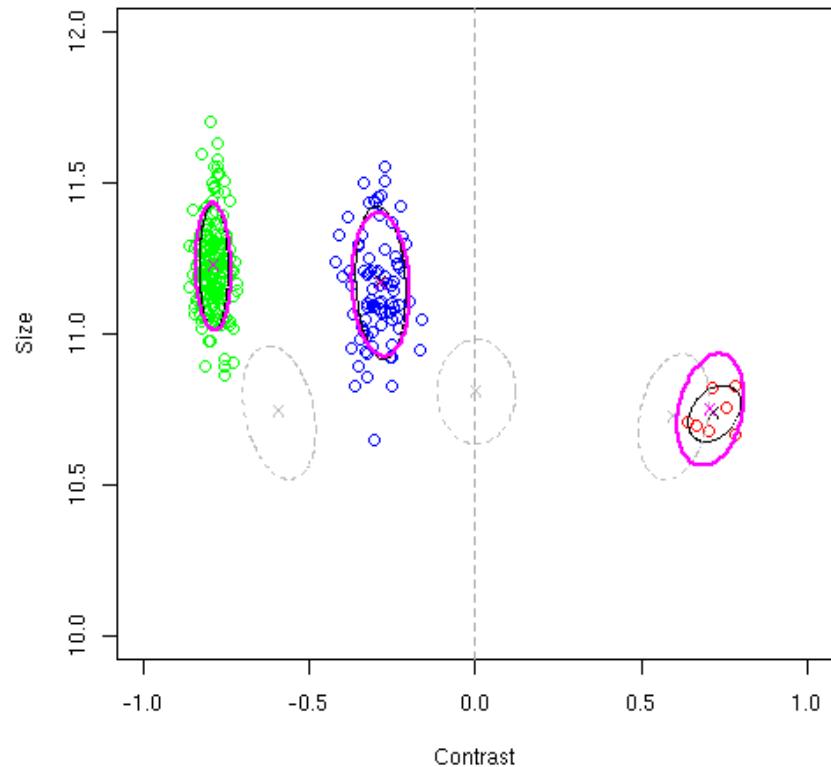
407212



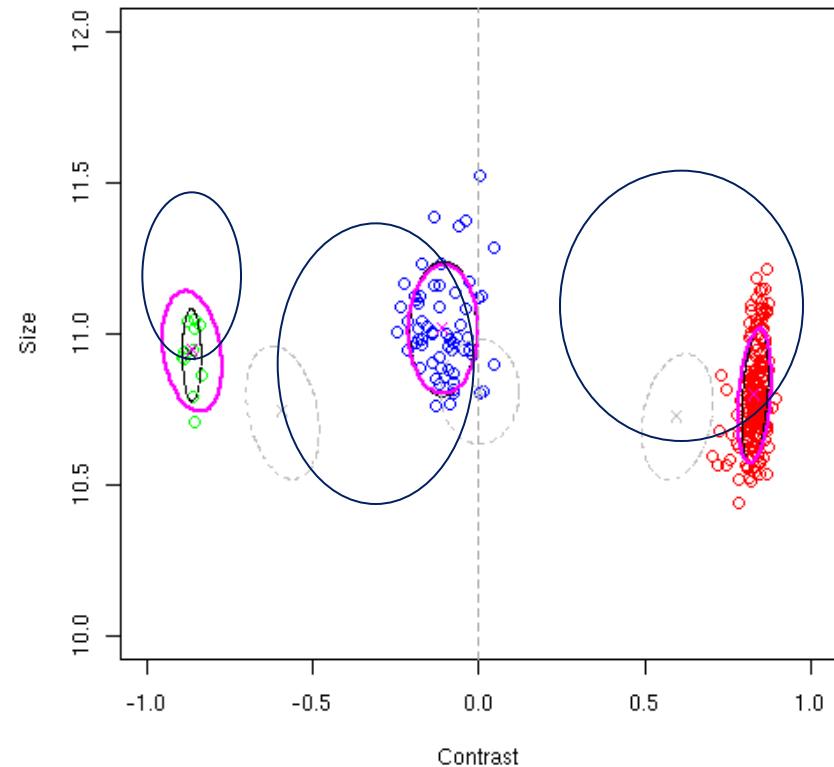
Estimation des trois centres et 3 variances en deux dimensions EM

Algorithme

301103



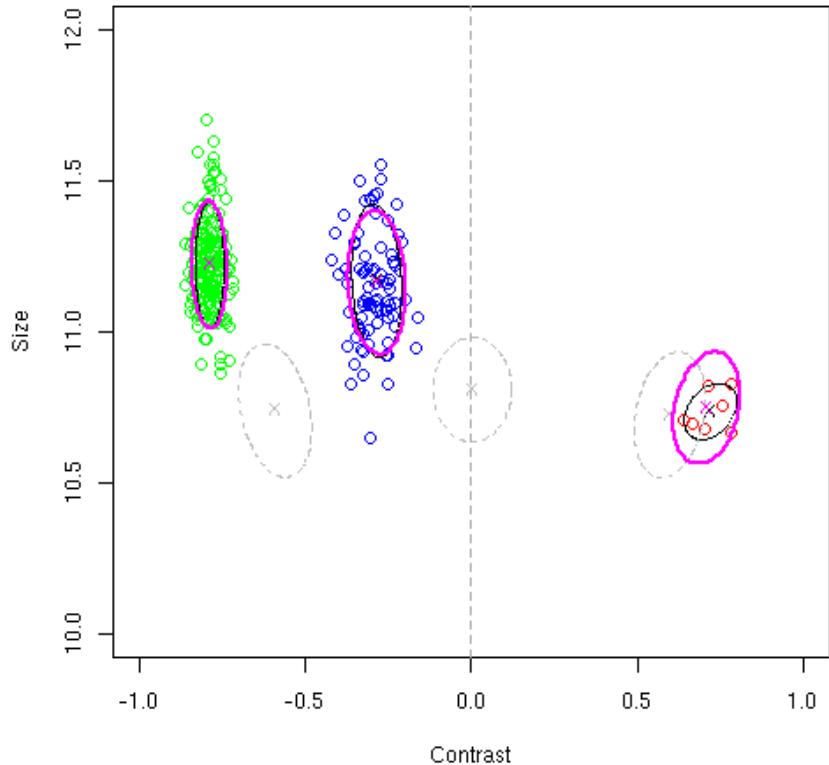
407212



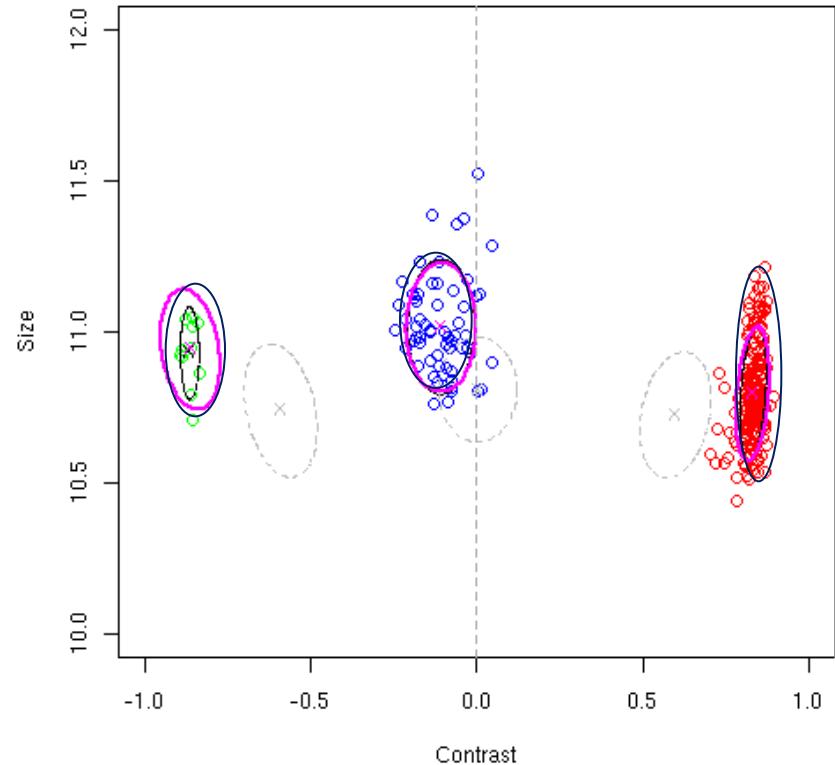
Estimation des trois centres et 3 variances en deux dimensions EM

Algorithme

301103

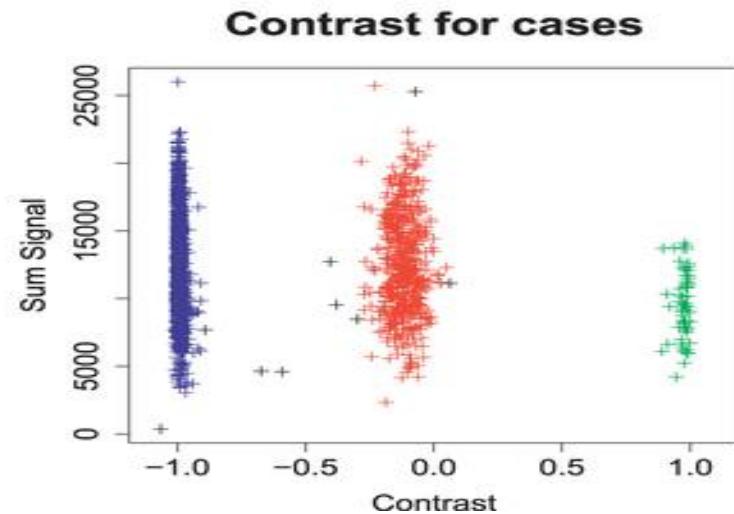
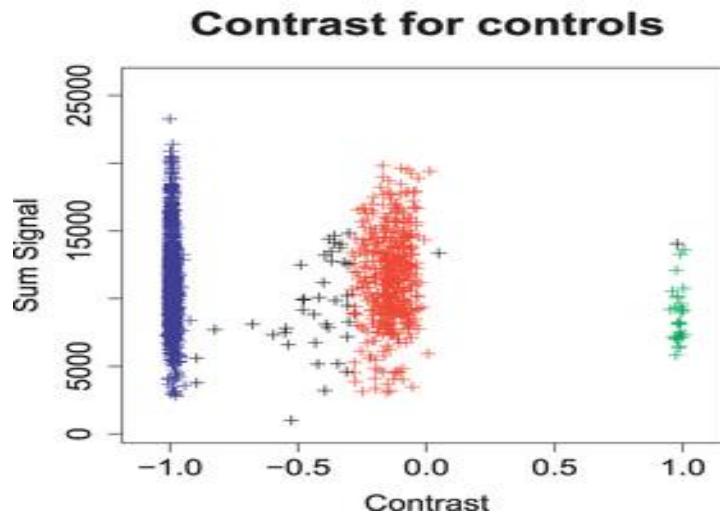
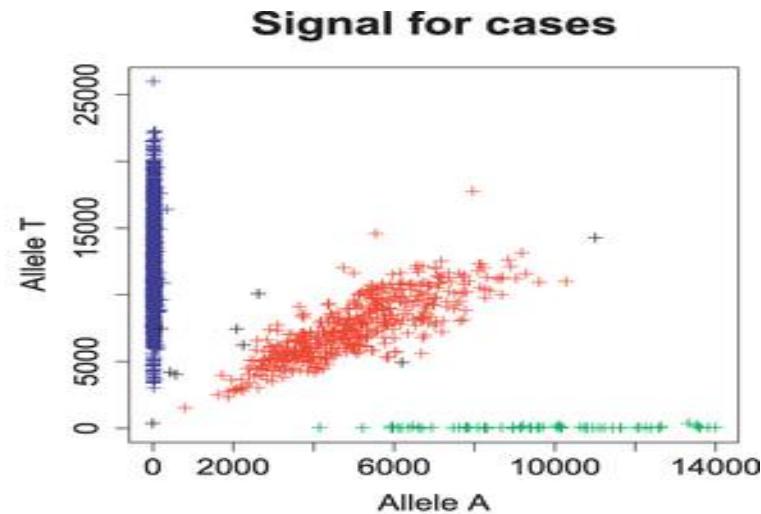
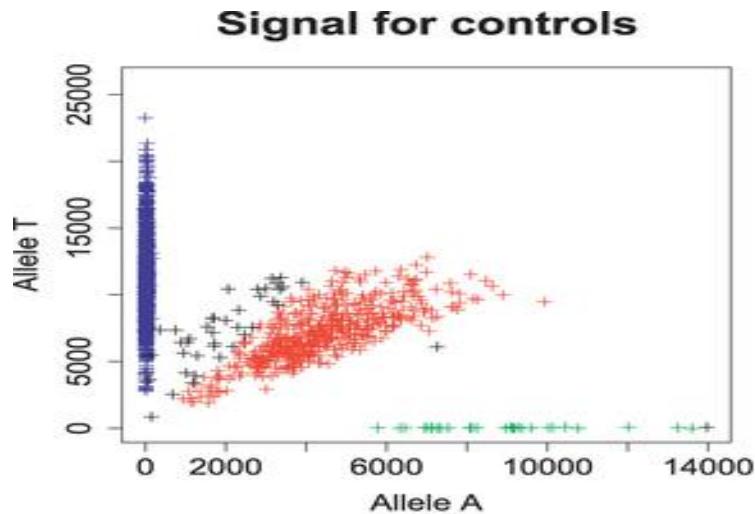


407212

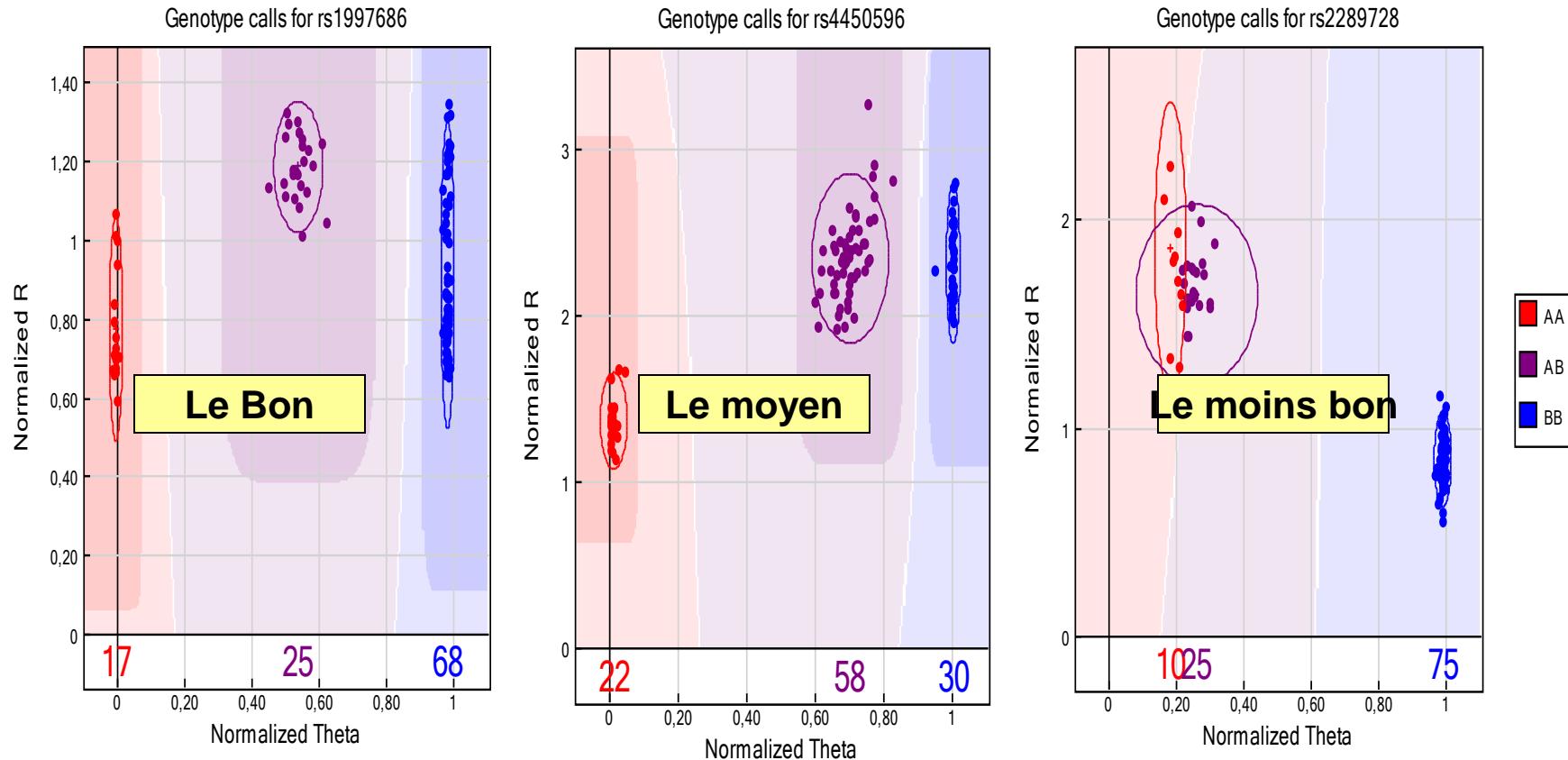


**Estimation des trois centres et 3 variances en deux dimensions EM
Calcul des distances par rapport aux centres des clusters**

Biais différentiel

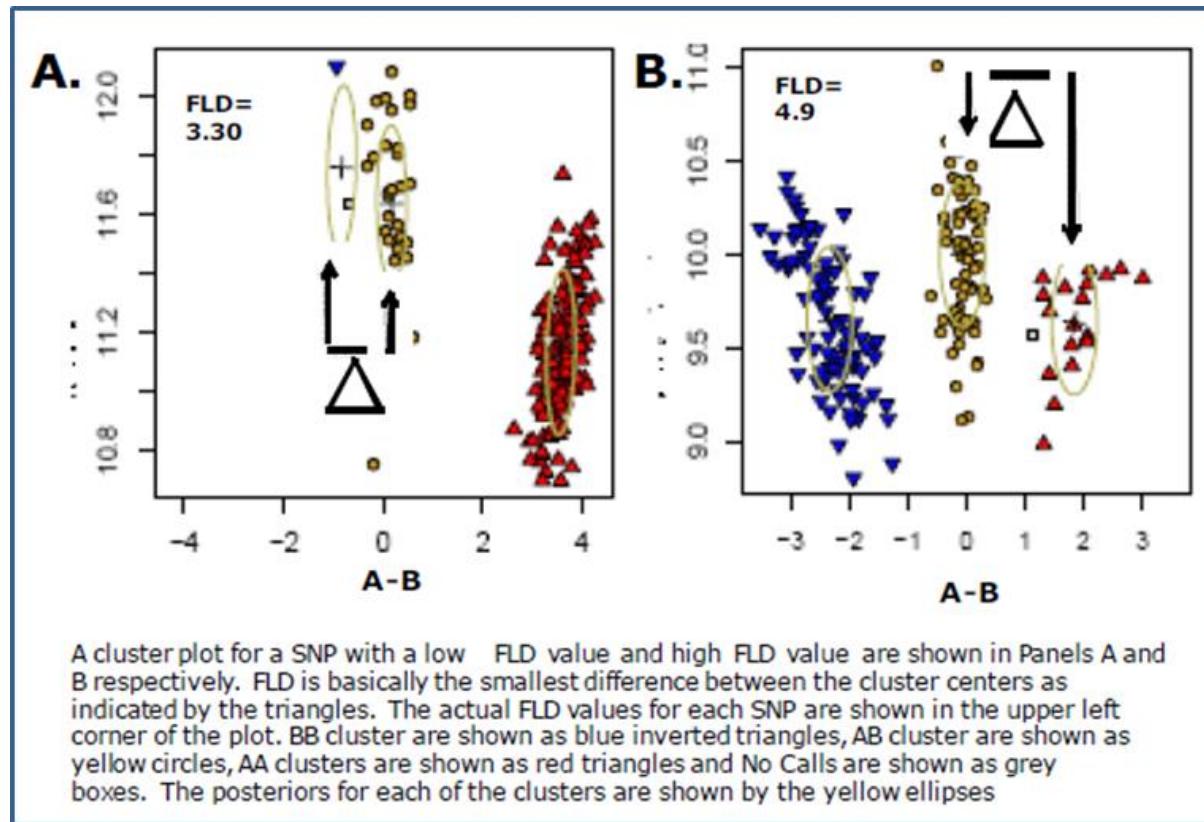


Suis-je AA, AT ou TT?



Somme sur les 3 génotypes pondérés par leur probabilité

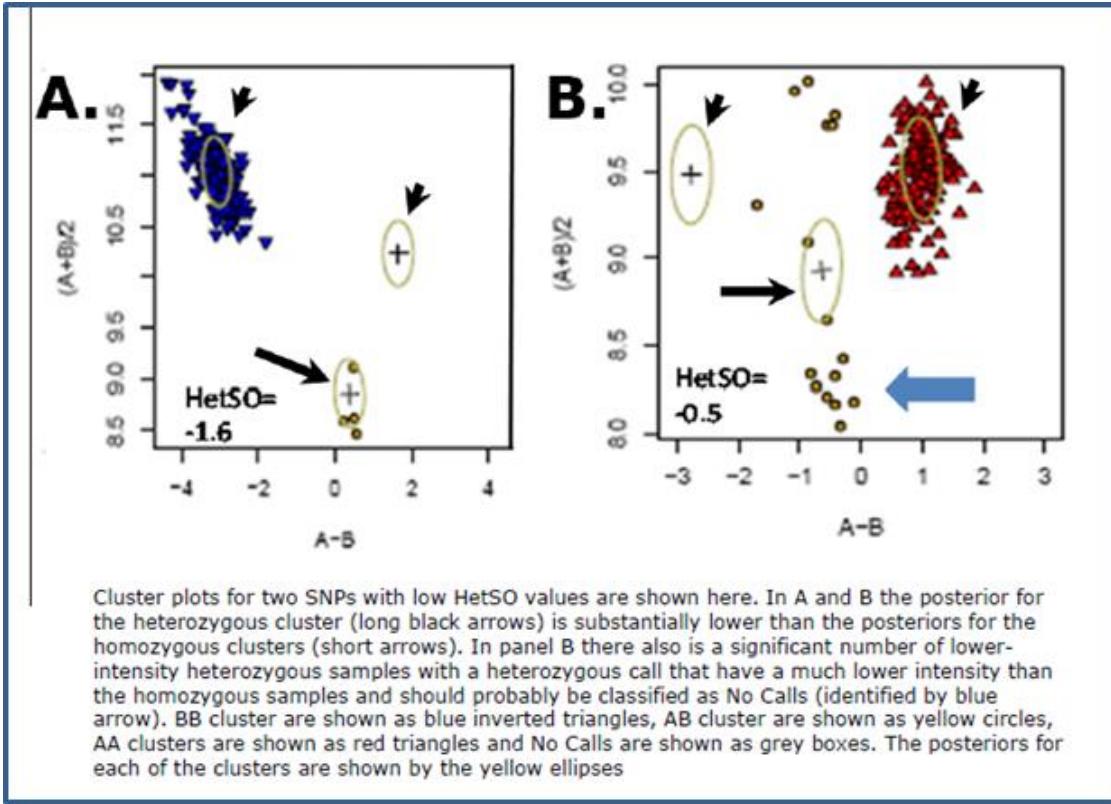
Tests automatiques



Fisher Linear Discriminant

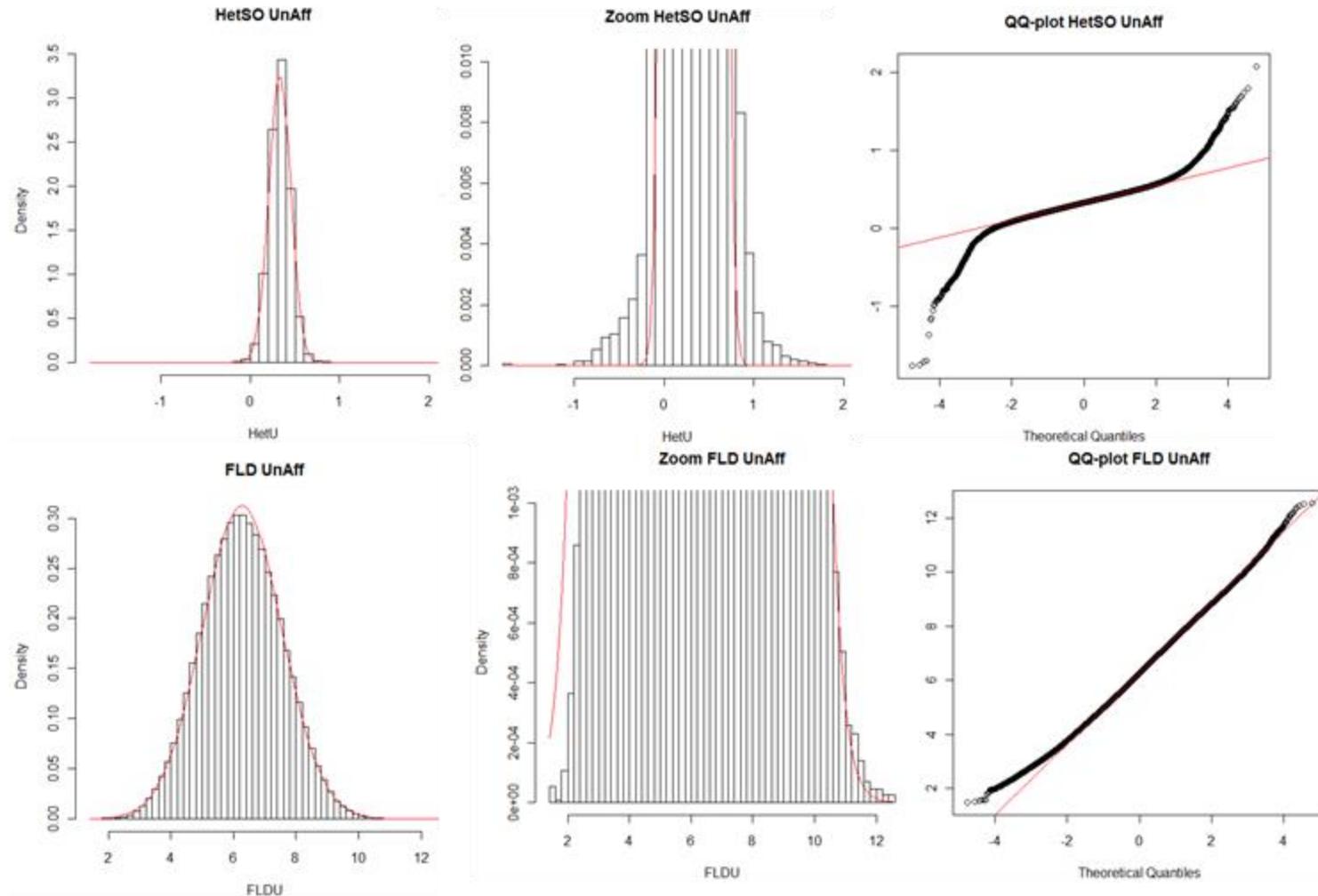
68

Tests automatiques

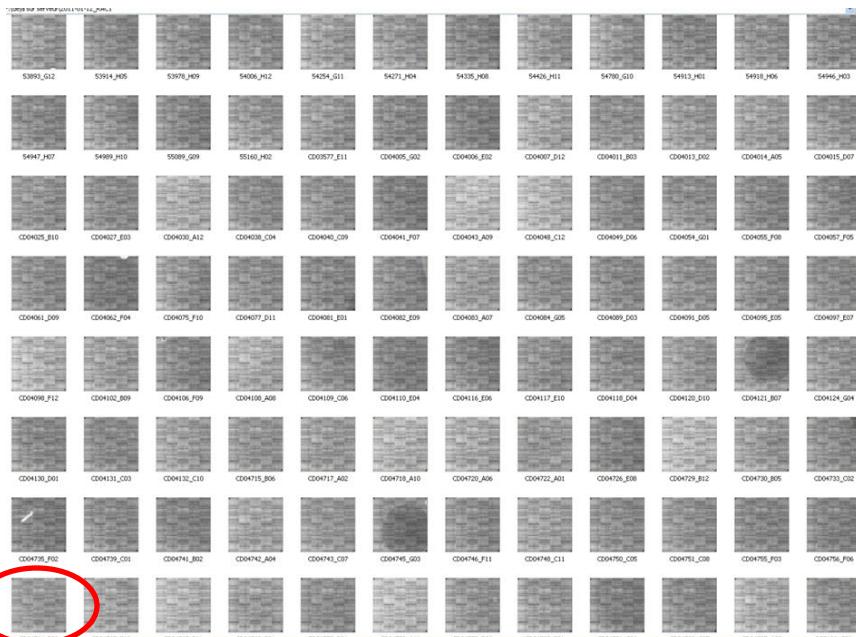


Het SO (heterozygotes Offset)

Evolution des variables



Plaques



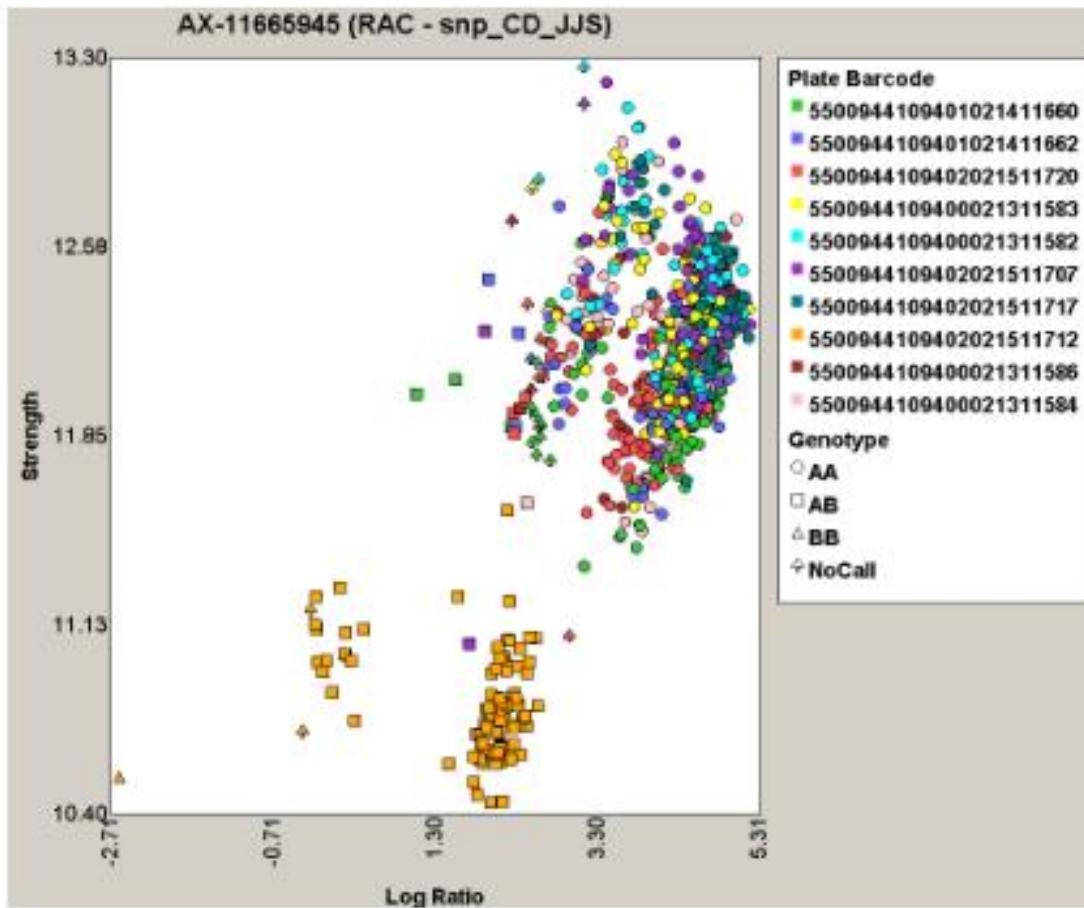
Individu

96 individus par plaques

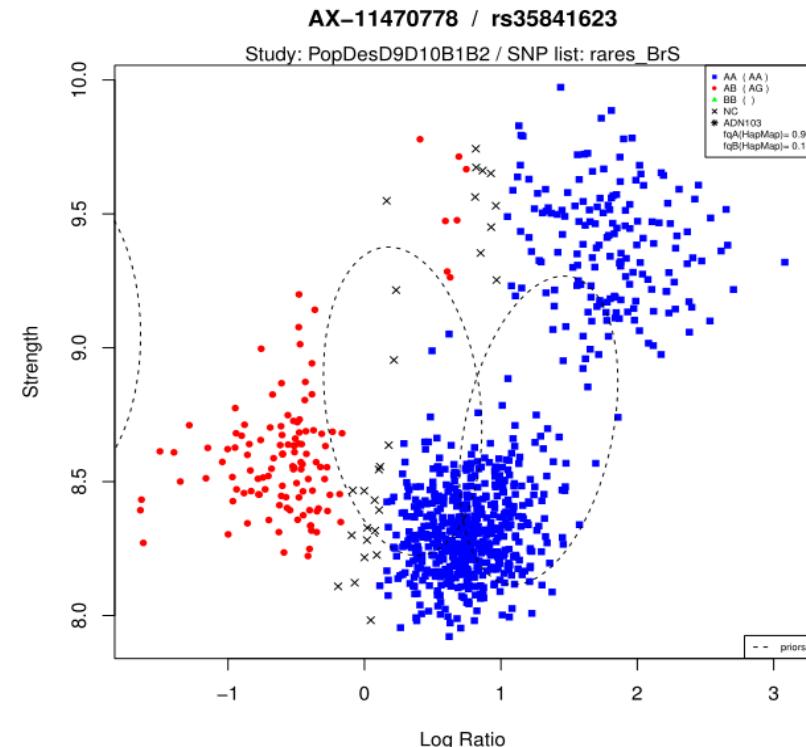
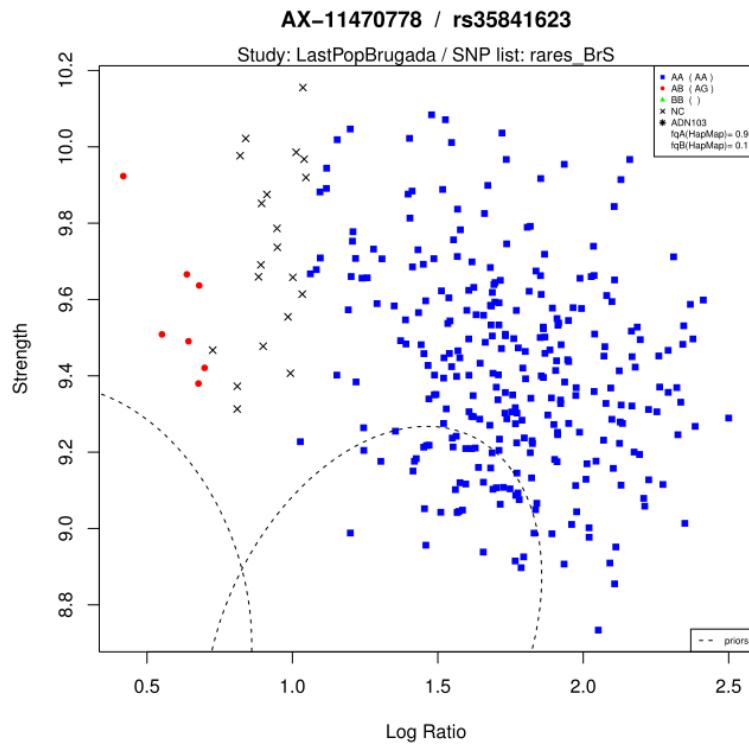
Les plaques sont passées deux par deux.

Conditions expérimentale différente par paires de plaques

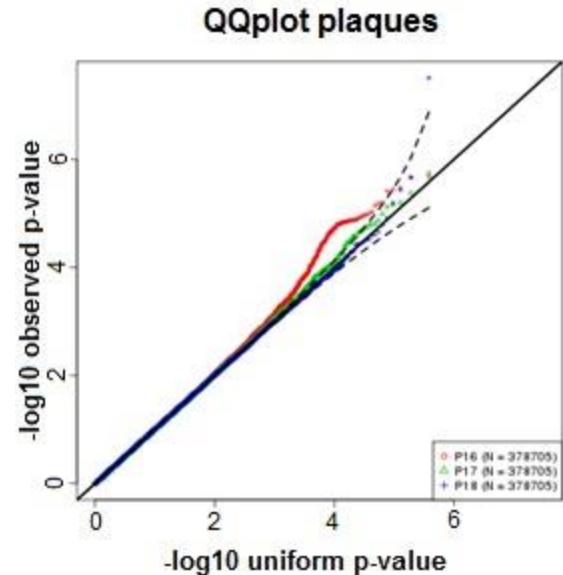
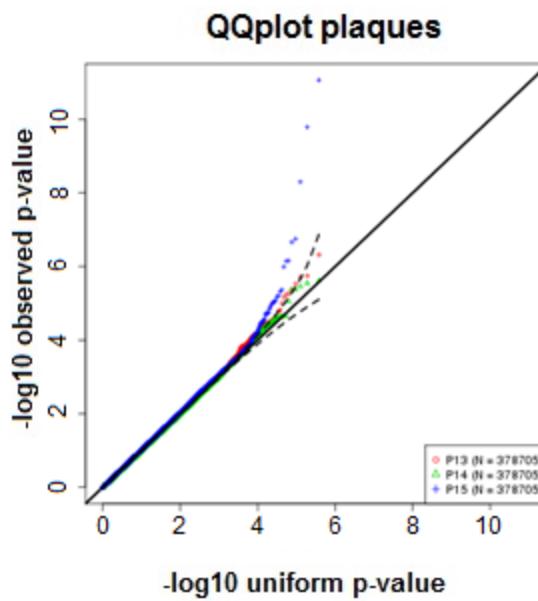
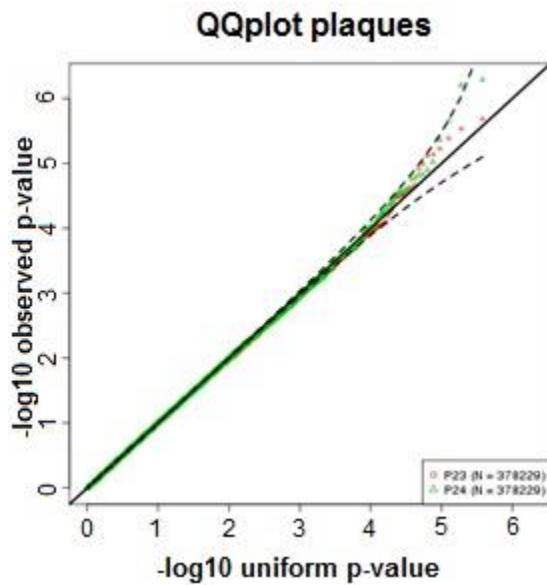
Batch Effect – Effet Plaque



Exemple de fausse association



Analyse des plaques Axiom



Plaques
homogènes

Un groupe SNPs
trop associés

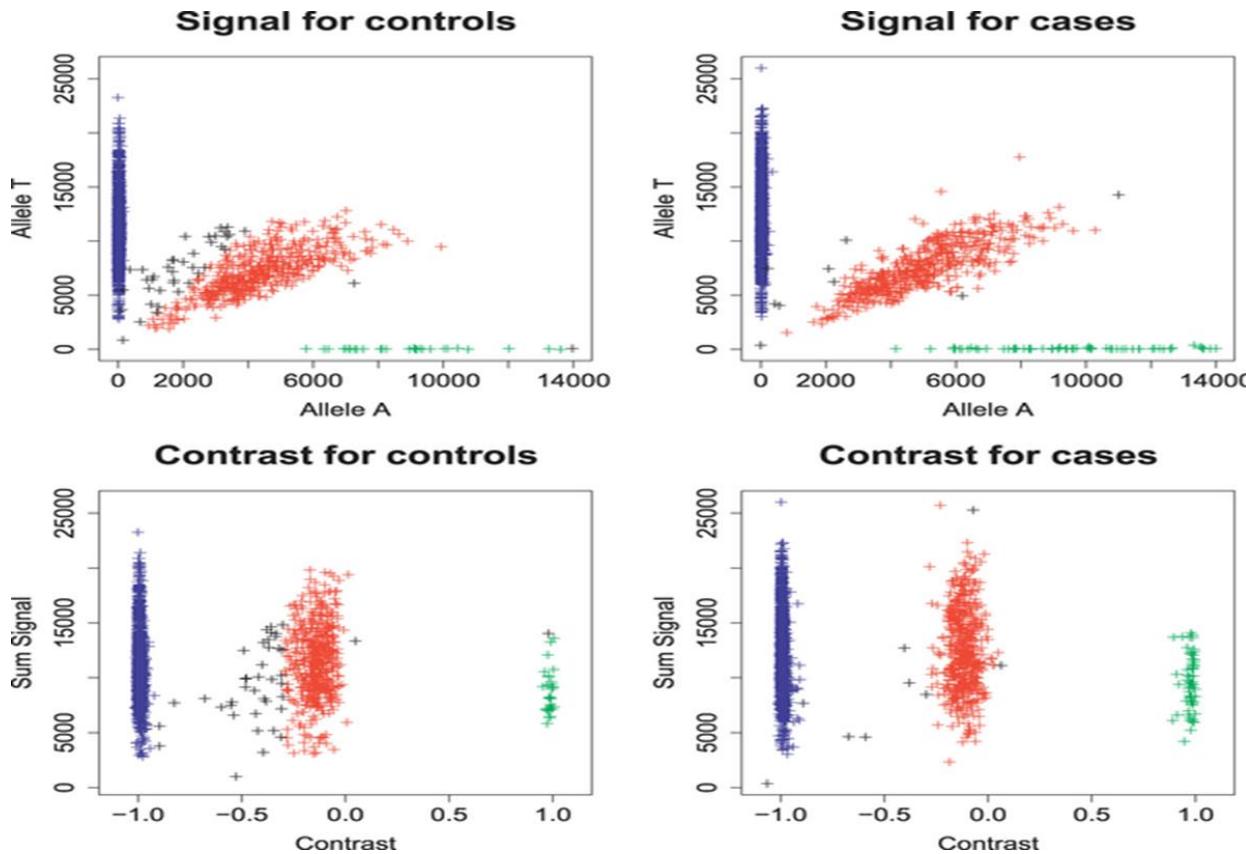
Plaque entière
de mauvaise
qualité

Choix

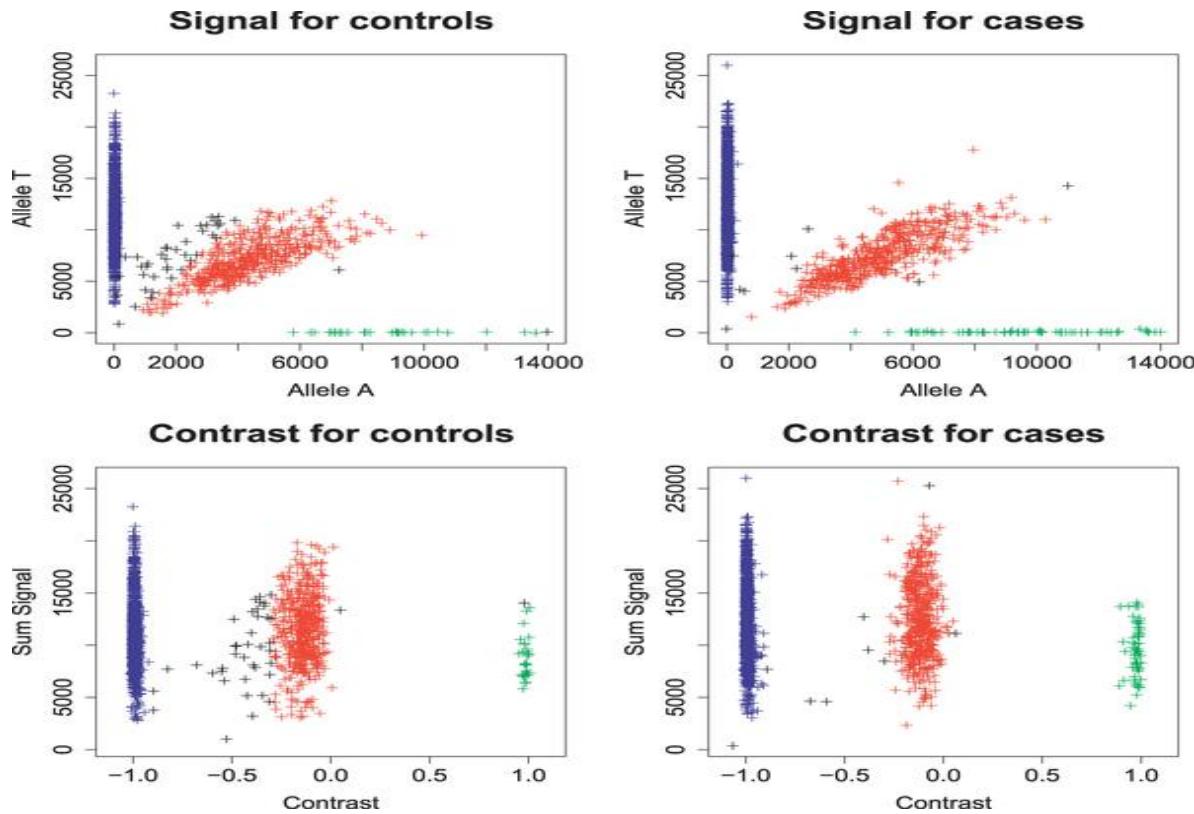
- Faire l'assignation par plaque (homogénéité)
- Faire l'assignation sur toutes les plaques (grand nombre – allèle rare)
- Solution entre les deux : toutes les plaques avec possibilité de moyennes mobiles
- Sujet de stage

Pipe-Line EAP

Proportions de Hardy Weinberg



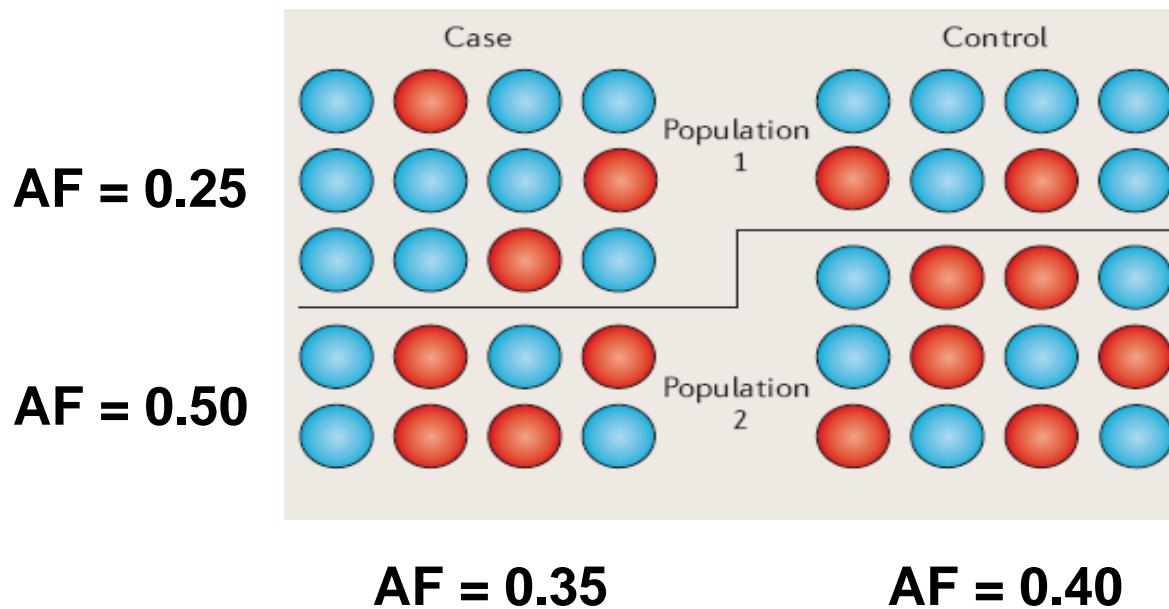
Pourcentage de génotypage



Pourcentage d'individus manquants

78

Stratification

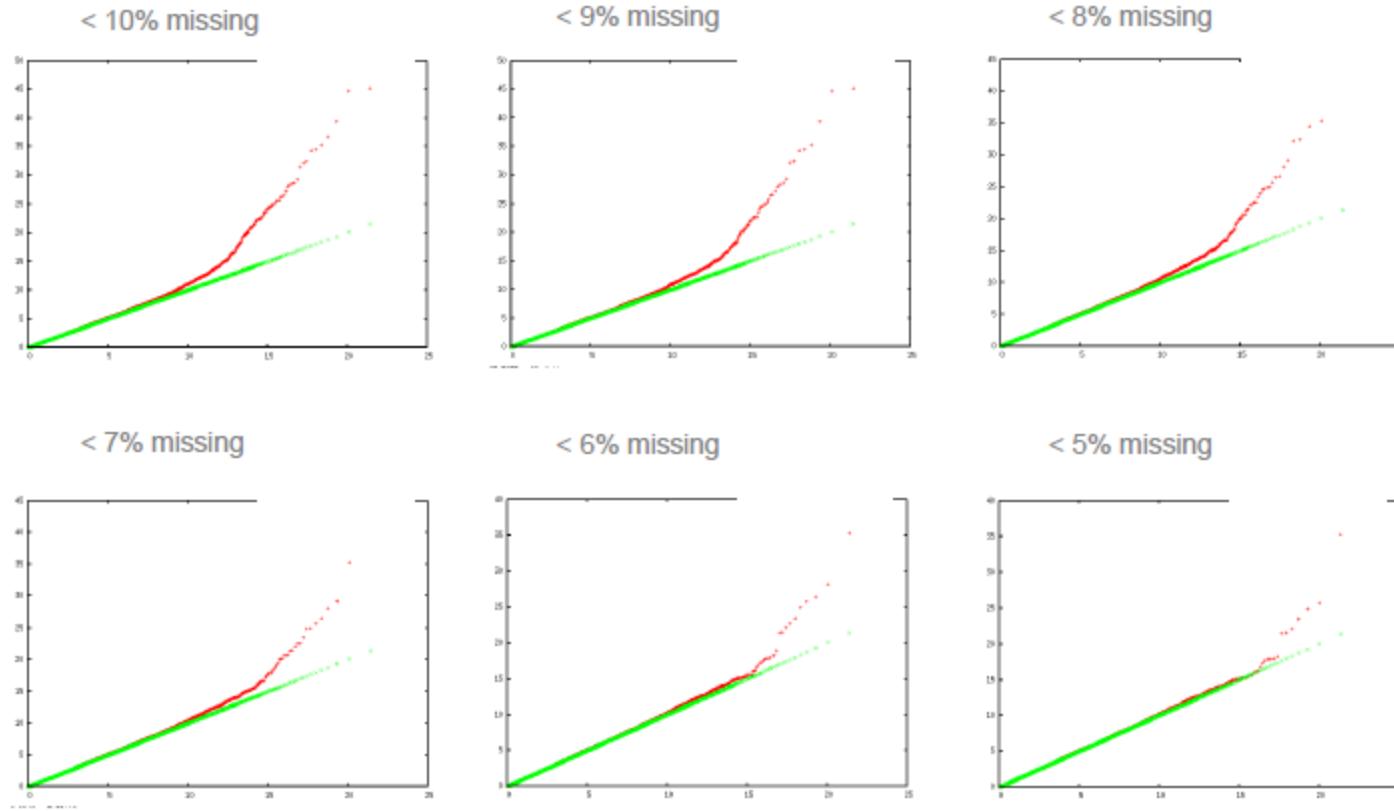


Comment corriger cela ? Voir présentation suivante

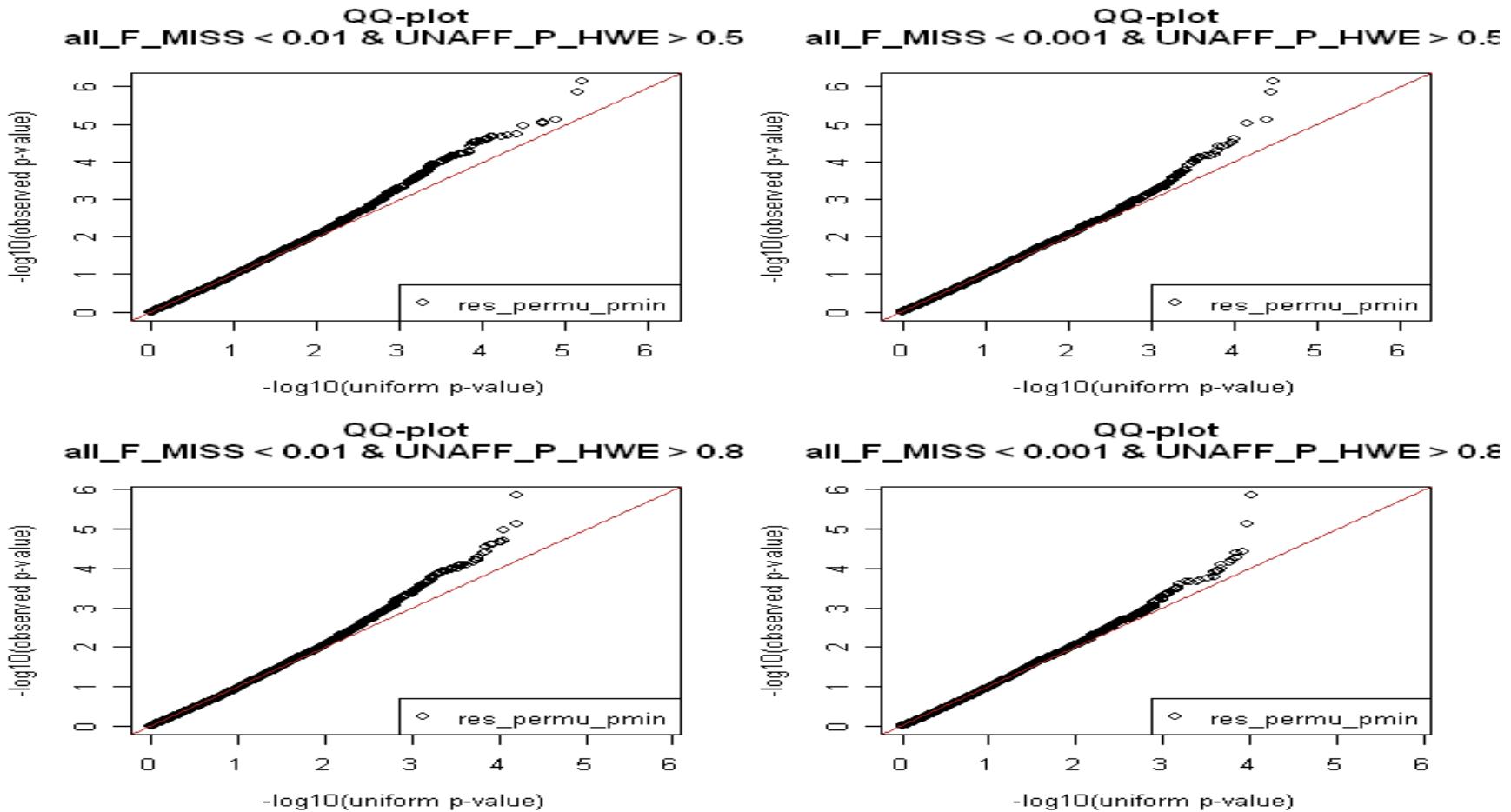
Contrôle Qualité

- % de génotypes manquants/individu
 - Classiquement:
 - Individus génotypés à 95% au moins
 - SNPs génotypés à 95% au moins
- % déséquilibre de Hardy Weinberg
 - Choix arbitraire ($p > 10 - 5$)
- Identification d'apparentement
 - Identité par Etat moyenne sur tout le génome

Qualité en fonction du taux de génotypage



Taux de génotypage et proportions HW



Contrôle génomique

- Estimation d'un facteur d'inflation:
 $\varepsilon = F(\chi^2, 1..M)$ pour M markers
 - F est souvent la moyenne ...
- On ajuste par le facteur d'inflation:

$$\chi^2_{adj} = \chi^2 / \varepsilon$$

$$\varepsilon = 1.12, 1.05, 1.11$$

ε Moyenne observée/moyenne attendue

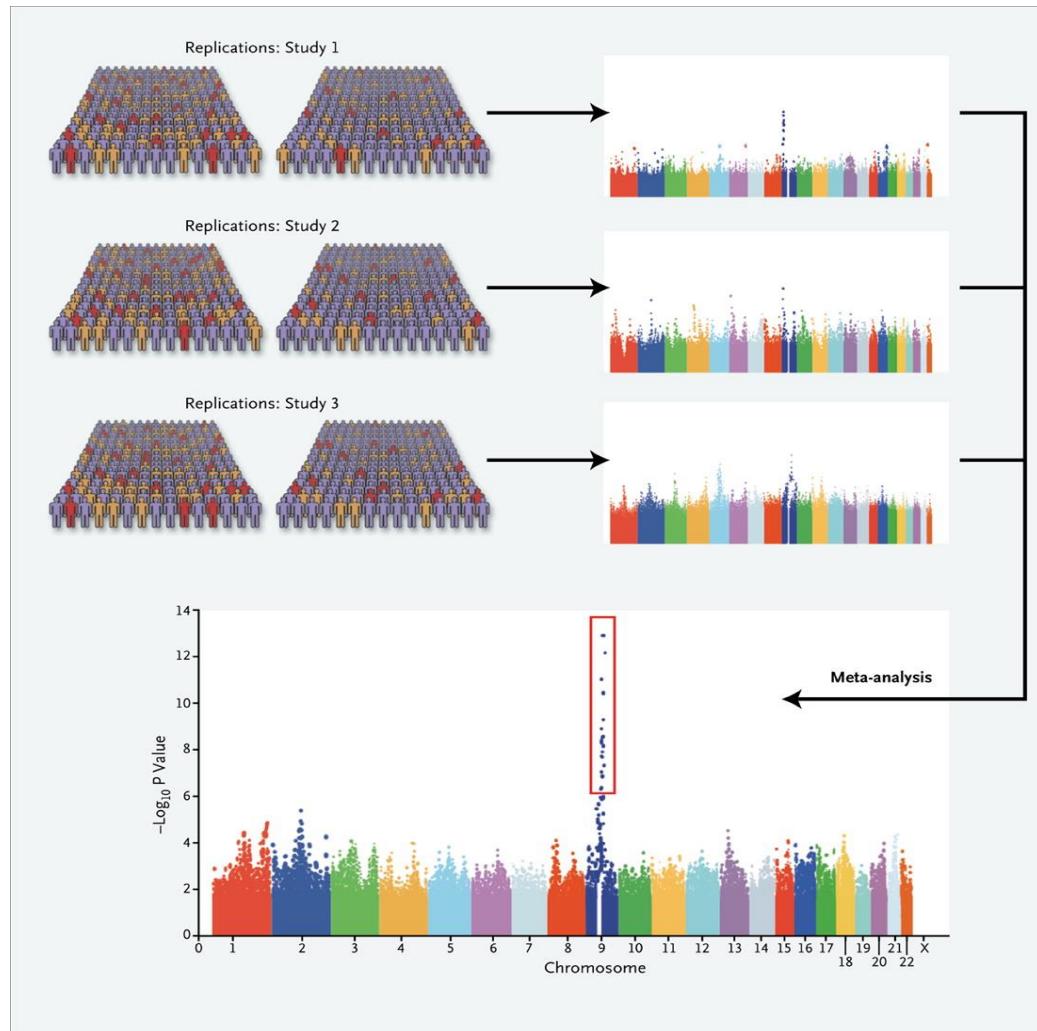
Médiane observée/médiane attendue

Efficace mais peu satisfaisant

Les meta-analyses

Pour la puissance

Meta-Analyse des études d'association génome entier



Meta-analyse

Combinaison, par SNP, des résultats de plusieurs études :

$$f(p_{i.}) = \sum_1^J S_{ij} \times \omega_{ij} \times \emptyset^{-1} (1 - p_i)$$

S_{ij} = -1 ou +1

$$\sum_1^J \omega_{ij}^2 = 1$$

Direction de l'effet

Poids de chaque étude

Objectif : PUISSANCE

86

Exemple d'étude

Diabète de Type 2

87

Analyse d'association génome entier pour le diabète de Type 2

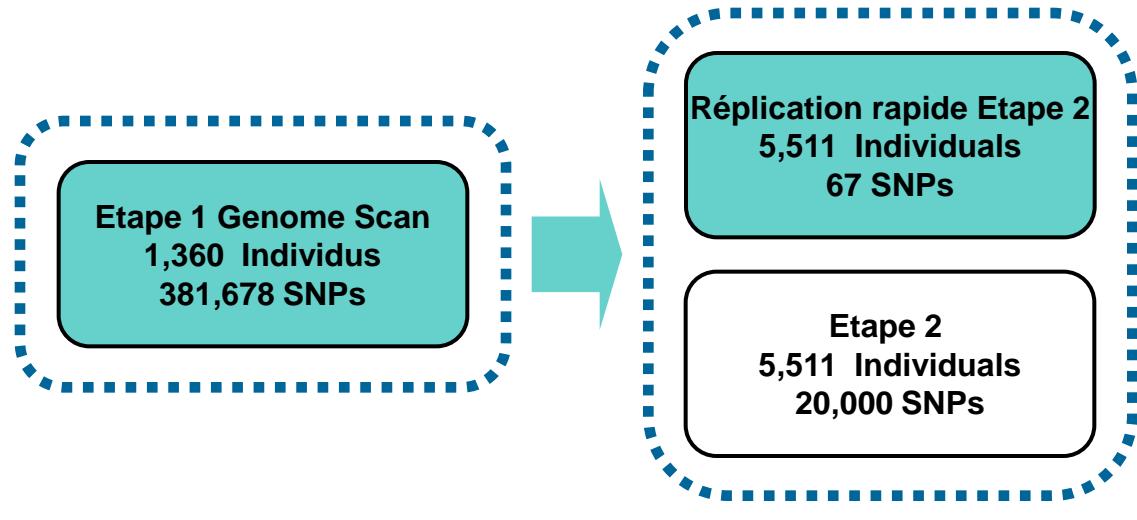
① A genome-wide association study identifies novel risk loci for type 2 diabetes

Robert Sladek^{1,2,4}, Ghislain Rocheleau^{1*}, Johan Rung^{4*}, Christian Dina^{5*}, Lishuang Shen¹, David Serre¹, Philippe Boutin⁵, Daniel Vincent⁴, Alexandre Belisle⁴, Samy Hadjadj⁶, Beverley Balkau⁷, Barbara Heude⁷, Guillaume Charpentier⁸, Thomas J. Hudson^{4,9}, Alexandre Montpetit⁴, Alexey V. Pshezhetsky¹⁰, Marc Prentki^{10,11}, Barry I. Posner^{2,12}, David J. Balding¹³, David Meyre⁵, Constantin Polychronakos^{1,3} & Philippe Froguel^{5,14}

Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis

Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes

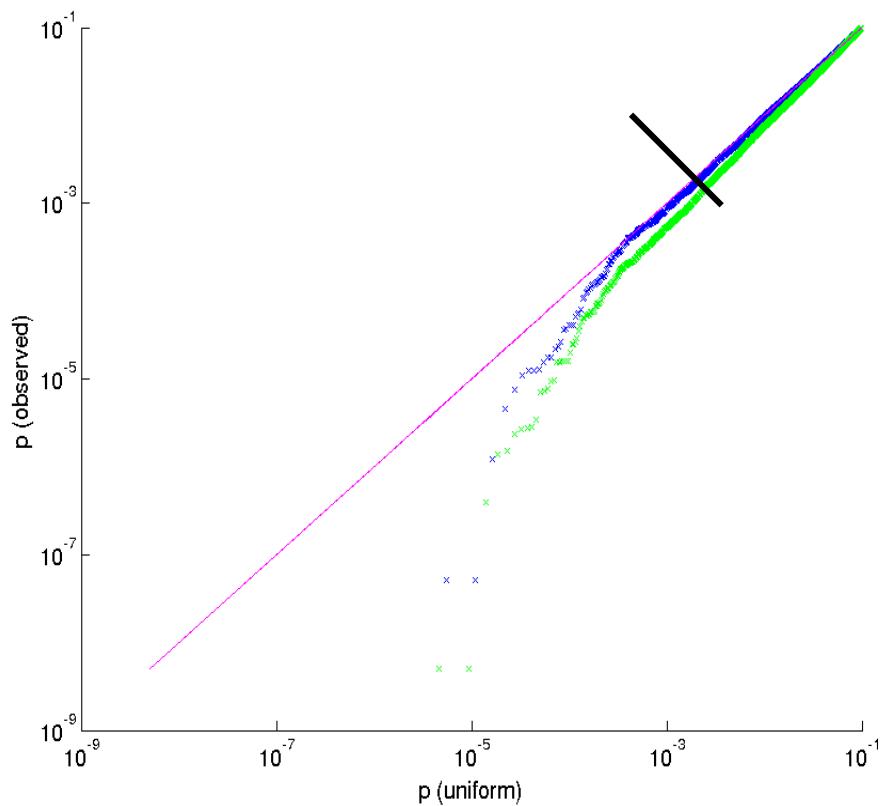
Plan d'Analyse



Identification de SNPs candidats

**Confirmation
RéPLICATION**

Sélection de réplication rapide en phase 2



Les bons résultats

| SNP | Chr | Position | Gene | pMAX (Corrected) | SNP | Chr | Position | Gene | pMAX (Corrected) |
|------------|-----|-----------|------------|----------------------|------------|-----|-----------|-----------|-----------------------|
| rs7900150 | 10 | 114783813 | TCF7L2 | 5.1×10^{-8} | rs7903146 | 10 | 114748339 | TCF7L2 | $< 1 \times 10^{-16}$ |
| rs1193179 | 1 | 7503868 | CAMTA1 | 1.3×10^{-6} | rs932206 | 2 | 136659004 | CXCR2DARS | 6.4×10^{-7} |
| rs1635250 | 23 | 3235152 | MXRA5 | 2.3×10^{-6} | rs35666 | 12 | 91036838 | BTG1 | 2.2×10^{-6} |
| rs932206 | 2 | 136659004 | CXCR4,DARS | 4.6×10^{-6} | rs7950175 | 11 | 126033245 | KIRREL3 | 2.7×10^{-6} |
| rs1978717 | 19 | 57189062 | ZNF615 | 7.5×10^{-6} | rs1037386 | 3 | 1453453 | | 4.2×10^{-6} |
| rs1111875 | 10 | 94452862 | HHEX | 1.2×10^{-5} | rs1193179 | 1 | 7503868 | CAMTA1 | 4.8×10^{-6} |
| rs282705 | 4 | 59343615 | | 1.3×10^{-5} | rs1256526 | 14 | 64809658 | | 6.4×10^{-6} |
| rs3740878 | 11 | 44214378 | EXT2 | 1.8×10^{-5} | rs6894954 | 5 | 144294556 | | 6.5×10^{-6} |
| rs2499953 | 11 | 4967481 | MMP26 | 2.4×10^{-5} | rs2317948 | 1 | 55146464 | | 7.4×10^{-6} |
| rs6670163 | 1 | 233862625 | RYR2 | 2.7×10^{-5} | rs2327112 | 6 | 8944645 | | 8.9×10^{-6} |
| rs945384 | 9 | 136892579 | FAM69B | 3.7×10^{-5} | rs1111875 | 10 | 94452862 | HHEX | 9.1×10^{-6} |
| rs7651936 | 3 | 163505661 | | 4.2×10^{-5} | rs859101 | 1 | 95036805 | SLC44A3 | 9.1×10^{-6} |
| rs10211998 | 22 | 35569377 | FLJ90680 | 4.2×10^{-5} | rs2589001 | 16 | 53776363 | | 9.5×10^{-6} |
| rs13064991 | 3 | 45809815 | SLC6A20 | 5.6×10^{-5} | rs9290240 | 3 | 165780254 | | 1.0×10^{-5} |
| rs1256517 | 14 | 64805437 | | 5.7×10^{-5} | rs282705 | 4 | 59343615 | | 1.2×10^{-5} |
| rs6541240 | 1 | 227397313 | TTC13 | 6.2×10^{-5} | rs2866016 | 4 | 99861413 | TSPAN5 | 1.2×10^{-5} |
| rs6413504 | 19 | 11102915 | LDLR | 8.4×10^{-5} | rs7949067 | 11 | 44248060 | ALX4 | 1.4×10^{-5} |
| rs2050831 | 9 | 77117950 | VPS13A | 8.5×10^{-5} | rs1978717 | 19 | 57189062 | ZNF615 | 1.4×10^{-5} |
| rs11078674 | 17 | 7251197 | NLGN2 | 9.0×10^{-5} | rs12629276 | 3 | 16403485 | RAFTLIN | 1.6×10^{-5} |
| rs873492 | 22 | 44818409 | FLJ27365 | 9.9×10^{-5} | rs1005316 | 17 | 66501964 | | 1.7×10^{-5} |
| | | | | | rs1888533 | 21 | 45825267 | | 1.8×10^{-5} |
| | | | | | rs1293143 | 20 | 52351866 | | 1.9×10^{-5} |
| | | | | | rs375694 | 21 | 42907029 | CR592664 | 1.9×10^{-5} |
| | | | | | rs13266634 | 8 | 118253964 | SLC30A8 | 2.1×10^{-5} |
| | | | | | rs11249433 | 1 | 120892655 | LOC642123 | 2.6×10^{-5} |
| | | | | | rs2876711 | 13 | 76314505 | | 2.7×10^{-5} |
| | | | | | rs231461 | 17 | 39388569 | PYY | 2.9×10^{-5} |
| | | | | | rs6823091 | 4 | 153427388 | DB127798 | 3.1×10^{-5} |
| | | | | | rs10823406 | 10 | 70982029 | | 3.2×10^{-5} |
| | | | | | rs10483096 | 22 | 16926334 | PEX26 | 3.9×10^{-5} |
| | | | | | rs10503677 | 8 | 20248324 | | 4.5×10^{-5} |

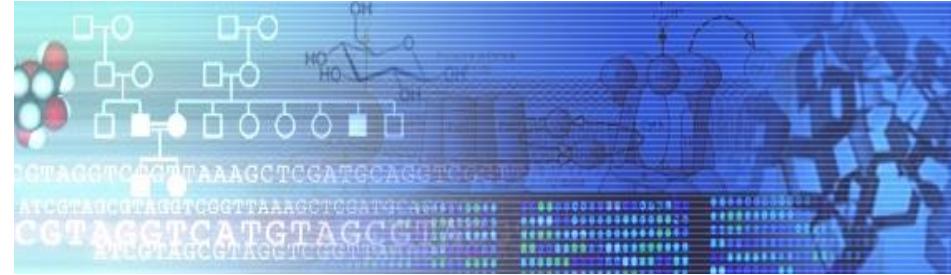
Réduit après ACF

Validation rapide confirme certaines associations

| SNP | Chr | Position | Gene | Stage I | Stage II | Combined |
|------------|-----|-----------|----------|--------------------------|--------------------------|--------------------------|
| rs7903146 | 10 | 114748339 | TCF7L2 | 3.94 X 10 ⁻¹⁹ | 1.51 X 10 ⁻³⁴ | 9.70 X 10 ⁻⁵² |
| rs13266634 | 8 | 118253964 | SLC30A8 | 7.18 X 10 ⁻⁶ | 6.15 X 10 ⁻⁸ | 2.14 X 10 ⁻¹¹ |
| rs1111875 | 10 | 94452862 | HHEX | 2.58 X 10 ⁻⁶ | 2.96 X 10 ⁻⁶ | 3.74 X 10 ⁻¹⁰ |
| rs7923837 | 10 | 94471897 | | 8.58 X 10 ⁻⁷ | 7.53 X 10 ⁻⁶ | 4.57 X 10 ⁻¹⁰ |
| rs2876711 | 13 | 76314505 | | 9.34 X 10 ⁻⁶ | 1.41 X 10 ⁻³ | 1.93 X 10 ⁻⁶ |
| rs3740878 | 11 | 44214378 | EXT2 | 7.18 X 10 ⁻⁶ | 1.17 X 10 ⁻⁴ | 2.42 X 10 ⁻⁶ |
| rs10823406 | 10 | 70982029 | | 1.12 X 10 ⁻⁵ | 2.65 X 10 ⁻³ | 6.36 X 10 ⁻⁶ |
| rs6413504 | 19 | 11102915 | LDLR | 2.65 X 10 ⁻⁵ | 2.78 X 10 ⁻³ | 9.14 X 10 ⁻⁶ |
| rs1193179 | 1 | 7503868 | CAMTA1 | 3.96 X 10 ⁻⁷ | 9.06 X 10 ⁻³ | 1.24 X 10 ⁻⁵ |
| rs729287 | 11 | 44236666 | | 9.01 X 10 ⁻⁶ | 6.71 X 10 ⁻⁴ | 1.64 X 10 ⁻⁵ |
| rs2499953 | 11 | 4967481 | MMP26 | 9.41 X 10 ⁻⁶ | 7.04 X 10 ⁻³ | 1.83 X 10 ⁻⁵ |
| rs1256517 | 14 | 64805437 | | 2.51 X 10 ⁻⁵ | 5.12 X 10 ⁻³ | 2.40 X 10 ⁻⁵ |
| rs1005316 | 17 | 66501964 | | 9.57 X 10 ⁻⁶ | 8.31 X 10 ⁻⁴ | 2.13 X 10 ⁻⁴ |
| rs282705 | 4 | 59343615 | | 2.82 X 10 ⁻⁶ | 1.46 X 10 ⁻¹ | 5.35 X 10 ⁻⁴ |
| rs8101509 | 19 | 57100638 | FLJ12644 | 7.76 X 10 ⁻⁶ | 1.00 X 10 ⁻¹ | 6.23 X 10 ⁻⁴ |
| rs2589001 | 16 | 53776363 | | 2.73 X 10 ⁻⁶ | 9.85 X 10 ⁻² | 6.51 X 10 ⁻⁴ |
| rs10483096 | 22 | 16926334 | PEX26 | 1.40 X 10 ⁻⁵ | 1.02 X 10 ⁻¹ | 7.74 X 10 ⁻⁴ |

OR max = 1.3

92

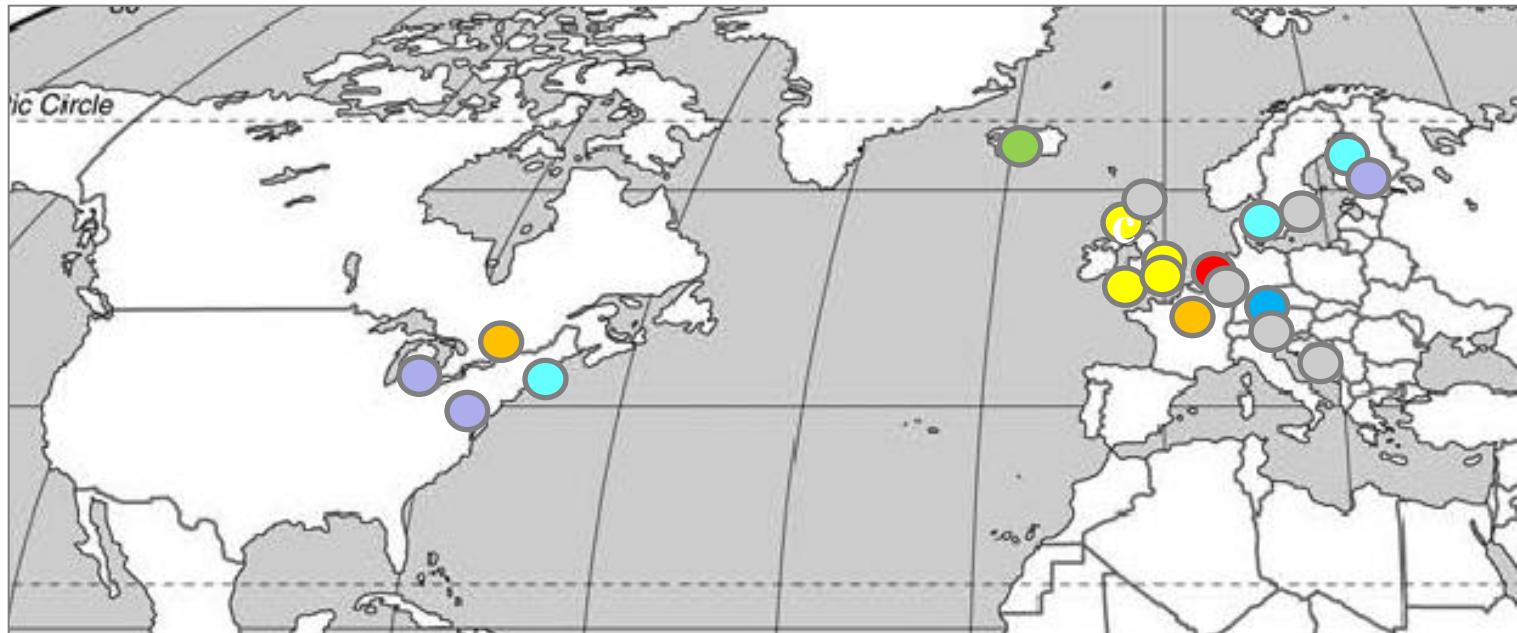


Twelve Type-2 Diabetes Loci Identified Through Large Scale Association Analysis

Benjamin F. Voight, Laura J. Scott, Valgerdur Steinthorsdottir, Andrew P. Morris, and Christian Dina on behalf of the DIAGRAM Consortium

Slides from B. Voight

The DIabetes GReplication And Meta-analysis [**DIAGRAM**] Consortium



- UK
- FUSION (US/Finland)
- DGI (US/Sweden/Finland)
- DeCODE
- KORA
- Rotterdam
- DGDG (France/Canada)
- EUROSPAN

Je parlerai spécifiquement de meta-analyse la prochaine fois

Gènes ou loci associés

Sladek et al, Nature 2007
Frayling et al, Science 2007
Dina et al. Nature Genetics 2007
Zeggini et al, Science 2007
DGI, Science 2007
Scott et al, Science 2007
WTCCC, Nature 2007
Steinthorsdottir, Nature Genetics 2007
Sandhu et al, Nature Genetics 2007
Zeggini et al, Nature Genetics 2008
Yasuda et al, Nature Genetics 2008
Unoki et al, Nature Genetics 2008
Rung et al, Nature Genetics 2009
Kong et al, Nature, 2009
Dupuis et al, Nature Genetics 2010
Saxena et al, Nature Genetics 2010
Qi et al, 2010

Études Gènes candidats

FTO *MTNR1B*

SLC30A8 *KCNQ1*

HHEX/IDE *TSPAN8*

CDKAL1 *ADAMTS9*

IGF2BP2 *NOTCH2*

CDKN2A *CAMK1D*

WFS1 *THADA*

TCF7L2 *HNF1B* *JAZF1* *IRS1*

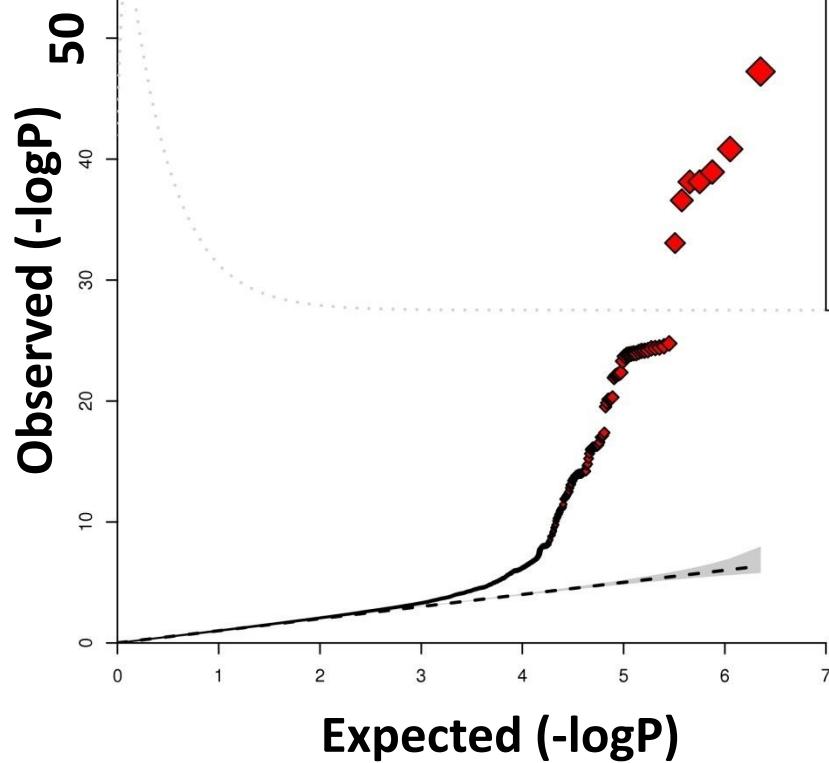
pré-EAGE – gènes candidats

PPARG

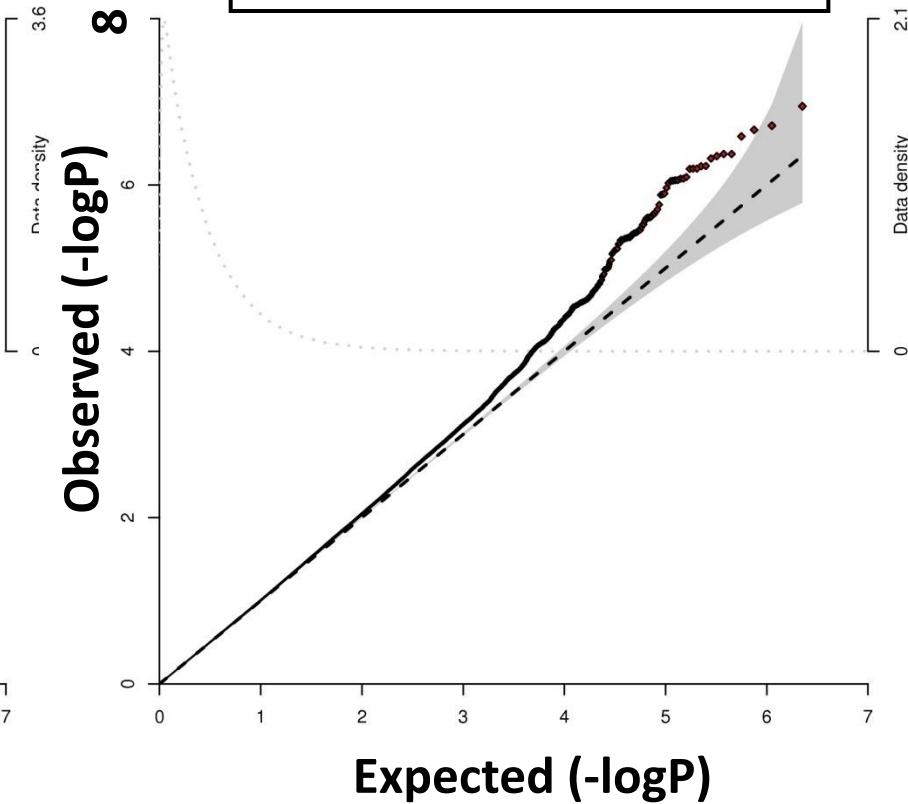
KCNJ11

| Study | Origin | Cases | Controls | Effective size ^a |
|-------------------------|-----------------------------------|--------------|---------------|-----------------------------|
| DCDG | France | 679 | 697 | 1,376 |
| DeCODE | Iceland | 1,465 | 23,194 | 4,218 |
| DGI | Sweden/Finland | 1,022 | 1,075 | 2,098 |
| ERGO/Rotterdam | Netherlands | 1,178 | 4,761 | 3,908 |
| EUROSPAN | UK, Croatia, Netherlands, Sweden, | 268 | 3,710 | 1,050 |
| FUSION | Finland | 1,161 | 1,174 | 2,335 |
| KORA | Germany | 433 | 1,438 | 1,372 |
| WTCCC | UK | 1,924 | 2,938 | 4,706 |
| Combined Stage 1 | | 8,130 | 38,987 | 21,063 |

Stage 1



Excluding Previously Known



Stage 2

2,255,857 SNPs



24 SNPs
Independent
 $P < 1 \times 10^{-5}$



Stage 2
replication

| Stage 2 | | | | | |
|------------------------------|-------------|--|--------|---------|---------|
| <i>In silico replication</i> | | | | | |
| Nurses Health Study | US | | 1,532 | 1,754 | 3,283 |
| Framingham Heart Study | US | | 674 | 7,664 | 2,600 |
| ARIC (European subset) | US | | 691 | 6425 | 2,612 |
| | | | 2,897 | 15,843 | 8,495 |
| <i>De novo genotyping</i> | | | | | |
| UK T2D Genetics Consortium | UK | | 5,107 | 6,612 | 11,610 |
| UK T2D Genetics Consortium | UK | | 632 | 1,648 | 1,874 |
| DGI: Malmo Diabetes | Sweden | | 2,449 | 3,363 | 5,718 |
| DGI (GCI, US set) | US | | 1,190 | 1,168 | 2,356 |
| DGI (GCI, Polish set) | Poland | | 790 | 803 | 1,594 |
| FUSION2 | Finland | | 1,211 | 1,266 | 2,478 |
| METSIM | Finland | | 898 | 3,594 | 2,974 |
| DCDG | France | | 3,554 | 5,894 | 8,992 |
| KORA | Germany | | 1,056 | 1,458 | 2,470 |
| Diagene/ERGO+ | Netherlands | | 1,699 | 1,665 | 3,350 |
| Ulm | Germany | | 944 | 1,949 | 2,592 |
| Steno | Denmark | | 4,089 | 5,043 | 9,086 |
| HUNT | Norway | | 1,235 | 3,292 | 3,686 |
| Cambridgeshire Case Control | UK | | 541 | 527 | 1,066 |
| NDCCS2(EPIC/Norfolk) | UK | | 6056 | 6428 | 12,492 |
| ADDITION/Ely | UK | | 892 | 1,610 | 2,332 |
| | | | 32,343 | 46,320 | 74,670 |
| Combined Stage 2 | | | | | |
| | | | 35,240 | 62,163 | 83,165 |
| Total Overall | | | | | |
| | | | 43,370 | 101,150 | 104,228 |

Combined Analysis (Stages One + Two) Identified 12 New Loci

| Chr | f(Risk Allele) | Nearby Gene | Stage One | | Stage Two <i>p</i> -value | Stage One + Two | | | |
|--------------------------------------|----------------|-----------------|------------------|---------------------------------------|------------------------------|-------------------------|-----------------------|--|--|
| | | | OR (95%CI) | <i>p</i> -value | | OR (95%CI) | <i>p</i> -value | | |
| | | | | up to 8,130 cases and 38,987 controls | | up to 35,240 and 62,163 | | | |
| Previously Known | | | | | | | | | |
| 2 | 0.64 | <i>IRS1</i> | 1.12 (1.07-1.17) | 8.7×10^{-7} | 2.2×10^{-15} | 1.11 (1.08-1.13) | 5.4×10^{-20} | | |
| 11 | 0.28 | <i>MTNR1B</i> | 1.12 (1.07-1.17) | 1.0×10^{-6} | 4.4×10^{-10} | 1.09 (1.06-1.11) | 7.8×10^{-15} | | |
| Novel T2D-susceptibility loci | | | | | | | | | |
| 11 | 0.88 | <i>CENTD2</i> | 1.13 (1.07-1.19) | 7.0×10^{-6} | 3.2×10^{-18} | 1.14 (1.11-1.17) | 1.4×10^{-22} | | |
| 2 | 0.46 | <i>BCL11A</i> | 1.09 (1.05-1.13) | 8.1×10^{-6} | 6.2×10^{-11} | 1.08 (1.06-1.10) | 2.9×10^{-15} | | |
| 11 | 0.52 | <i>KCNQ1</i> | 1.11 (1.06-1.16) | 6.4×10^{-6} | 3.2×10^{-9} | 1.08 (1.06-1.10) | 2.8×10^{-13} | | |
| 5 | 0.26 | <i>ZBED3</i> | 1.16 (1.10-1.23) | 4.2×10^{-8} | 2.7×10^{-7} | 1.08 (1.06-1.11) | 2.8×10^{-12} | | |
| 7 | 0.55 | <i>KLF14</i> | 1.10 (1.06-1.15) | 1.8×10^{-6} | 6.4×10^{-6} | 1.07 (1.05-1.10) | 2.2×10^{-10} | | |
| 15 | 0.22 | <i>PRC1</i> | 1.10 (1.06-1.15) | 8.2×10^{-6} | 1.6×10^{-6} | 1.07 (1.05-1.09) | 2.4×10^{-10} | | |
| 8 | 0.48 | <i>TP53INP1</i> | 1.10 (1.06-1.15) | 1.2×10^{-6} | 2.2×10^{-5} | 1.06 (1.04-1.09) | 9.9×10^{-10} | | |
| 15 | 0.60 | <i>ZFAND6</i> | 1.11 (1.06-1.16) | 5.1×10^{-6} | 1.2×10^{-5} | 1.06 (1.04-1.08) | 2.4×10^{-9} | | |
| 12 | 0.10 | <i>HMGAA2</i> | 1.20 (1.12-1.29) | 1.7×10^{-7} | 1.1×10^{-4} | 1.10 (1.07-1.14) | 3.6×10^{-9} | | |
| 12 | 0.85 | <i>HNF1A</i> | 1.14 (1.08-1.19) | 4.6×10^{-7} | 4.6×10^{-4} | 1.07 (1.05-1.10) | 2.4×10^{-8} | | |
| 9 | 0.93 | <i>CHCHD9</i> | 1.20 (1.11-1.29) | 1.5×10^{-6} | 2.4×10^{-4} | 1.11 (1.07-1.15) | 2.8×10^{-8} | | |
| X | 0.79 | <i>DUSP9**</i> | 1.12 (1.07-1.17) | 2.3×10^{-6} | 3.5×10^{-6} | 1.14 (1.09-1.18) | 9.5×10^{-11} | | |

Current count of 39 confirmed T2D-susceptibility loci

Sladek et al, Nature 2007
 Frayling et al, Science 2007
 Zeggini et al, Science 2007
 DGI, Science 2007
 Scott et al, Science 2007
 WTCCC, Nature 2007
 Steinhorsdottir, Nature Genetics 2007
 Sandhu et al, Nature Genetics 2007
 Zeggini et al, Nature Genetics 2008
 Yasuda et al, Nature Genetics 2008
 Unoki et al, Nature Genetics 2008
 Rung et al, Nature Genetics 2009
 Kong et al, Nature, 2009
 Dupuis et al, Nature Genetics 2010
 Saxena et al, Nature Genetics 2010
 Qi et al, 2010
Voight et al, Accepted in Principal, Nature Genetics 2010

For purposes of presentation, loci are named according to a nearby gene of interest. In only a few cases is the causal gene yet proven.

| | | | |
|-----------------|----------------|-----------------|------------------|
| <i>ADCY5</i> | <i>GCK</i> | <i>GCKR</i> | <i>PROX1</i> |
| <i>FTO</i> | <i>MTNR1B</i> | <i>TP53INP1</i> | <i>DGKB</i> |
| <i>SLC30A8</i> | <i>KCNQ1</i> | <i>KLF14</i> | <i>HCCA2</i> |
| <i>HHEX/IDE</i> | <i>TSPAN8</i> | <i>ZBED3</i> | <i>RBMS1</i> |
| <i>CDKAL1</i> | <i>ADAMTS9</i> | <i>BCL11A</i> | <i>DUSP9</i> |
| <i>IGF2BP2</i> | <i>NOTCH2</i> | <i>CHCHD9</i> | <i>KCNQ1 [2]</i> |
| <i>CDKN2A</i> | <i>CAMK1D</i> | <i>HNF1A</i> | <i>ZFAND6</i> |
| <i>WFS1</i> | <i>THADA</i> | <i>HMGA2</i> | <i>PRC1</i> |
| <i>TCF7L2</i> | <i>HNF1B</i> | <i>JAZF1</i> | <i>IRS1</i> |
| <i>KCNJ11</i> | | | <i>CENTD2</i> |

Candidate Gene Studies

GWAS for T2D

GWAS of Related Traits

Large-scale association efforts

PPARG

KCNJ11

TCF7L2

HNF1B

JAZF1

IRS1

CENTD2

2000

2003

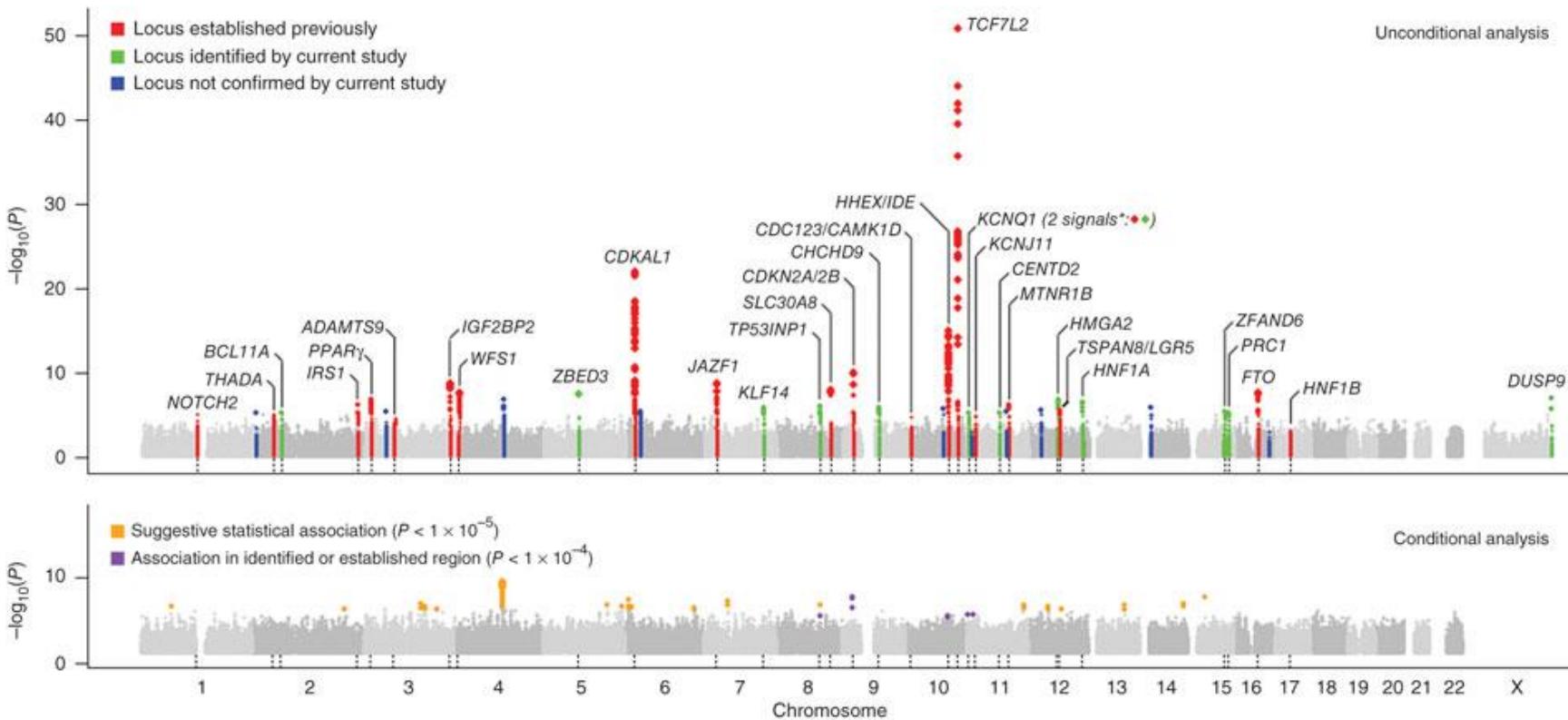
2006

2007

2008

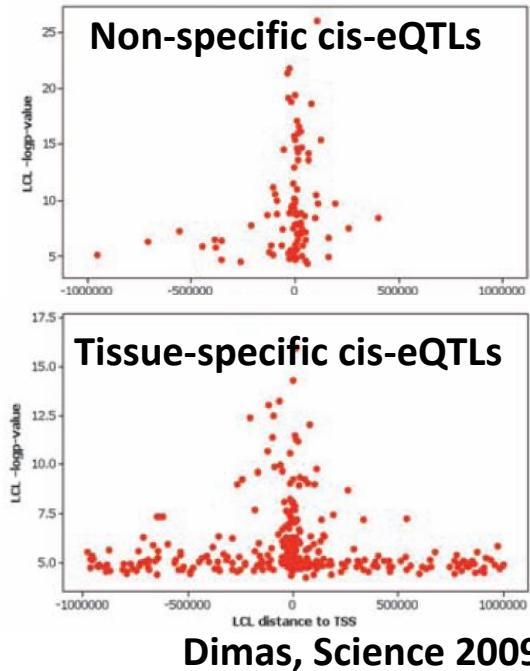
2009/2010

Encore des association (Diabète de Type 2)



Same Region, Multiple Traits

13 of 30 T2D loci tested have an independent association to another phenotype within 500kb (by simulation, $p < 2.4 \times 10^{-3}$)



Hypothesis: Tissue specific use of the same machinery

| Locus | Other trait | R2 | Distance apart (kb) |
|----------------------------|--------------------------|-------|---------------------|
| <i>THADA (ABCG5/ABCG8)</i> | Plasma LDL | 0.007 | 455 |
| <i>BCL11A</i> | Fetal Hemoglobin Level | 0.021 | 135 |
| <i>IRS1</i> | Coronary disease | 0.929 | 47 |
| | Thyroid function | 0.034 | 93 |
| <i>ZBED3</i> | Thyroid function | 0.028 | 105 |
| | Thyroid function | 0.028 | 111 |
| <i>CDKAL1</i> | Crohn's | 0.055 | 41 |
| | Psoriasis | 0.055 | 41 |
| <i>JAZF1</i> | Prostate Cancer | 0.025 | 220 |
| | Stature | 1 | 7 |
| <i>KLF14</i> | Basal cell carcinoma | 0.019 | 119 |
| | CAD, Aortic Aneurysm | 0 | 9 |
| <i>CDKN2A/B</i> | Melanoma | 0.003 | 317 |
| | Basal cell carcinoma | 0 | 99 |
| | Glioma (Brain Cancer) | 0.001 | 65 |
| <i>CHCHD9 (TLE4)</i> | Asthma | 0.163 | 87 |
| <i>KCNQ1</i> | QT interval | 0.006 | 207 |
| <i>KCNJ11</i> | Diastolic blood pressure | 0.002 | 491 |
| <i>HMGA2</i> | Aortic Root Size | 0.002 | 220 |
| | Stature | 0 | 183 |
| | CRP | 0.11 | 21 |
| <i>HNF1A</i> | Liver enzymes | 0.11 | 18 |
| | CAD | 0.11 | 25 |
| | Plasma LDL | 0.05 | 72 |
| <i>PRC1</i> | ADHD | 0.357 | 71 |
| | Serum metabolites | 0.011 | 196 |
| <i>HNF1B</i> | Prostate Cancer | 0.766 | 3 |

39 loci confirmés

Sladek et al, Nature 2007
Frayling et al, Science 2007
Zeggini et al, Science 2007
DGI, Science 2007
Scott et al, Science 2007
WTCCC, Nature 2007
Steinthorsdottir, Nature Genetics 2007
Sandhu et al, Nature Genetics 2007
Zeggini et al, Nature Genetics 2008
Yasuda et al, Nature Genetics 2008
Unoki et al, Nature Genetics 2008
Rung et al, Nature Genetics 2009
Kong et al, Nature, 2009
Dupuis et al, Nature Genetics 2010
Saxena et al, Nature Genetics 2010
Qi et al, 2010
Voight et al, Nature Genetics 2010

12 nouveaux identifiés dans cette étude

Études Gènes candidats

Études Association Génome Entier D2T

EAGE traits quantitatifs

pré-EAGE – gènes candidats

PPARG

KCNJ11

| | | |
|-----------------|----------------|------------------|
| <i>FTO</i> | <i>MTNR1B</i> | <i>TP53INP1</i> |
| <i>SLC30A8</i> | <i>KCNQ1</i> | <i>KLF14</i> |
| <i>HHEX/IDE</i> | <i>TSPAN8</i> | <i>ZBED3</i> |
| <i>CDKAL1</i> | <i>ADAMTS9</i> | <i>BCL11A</i> |
| <i>IGF2BP2</i> | <i>NOTCH2</i> | <i>DUSP9</i> |
| <i>CDKN2A</i> | <i>CAMK1D</i> | <i>CHCHD9</i> |
| <i>WFS1</i> | <i>THADA</i> | <i>KCNQ1 [2]</i> |
| | | <i>HNF1A</i> |
| | | <i>ZFAND6</i> |
| | | <i>HMGA2</i> |
| | | <i>PRC1</i> |
| | | <i>IRS1</i> |
| | | <i>CENTD2</i> |

2000

2003

2006

2007

2008

2009/2010

39 loci confirmés

Sladek et al, Nature 2007
 Frayling et al, Science 2007
 Zeggini et al, Science 2007
 DGI, Science 2007
 Scott et al, Science 2007
 WTCCC, Nature 2007
 Steinthorsdottir, Nature Genetics 2007
 Sandhu et al, Nature Genetics 2007
 Zeggini et al, Nature Genetics 2008
 Yasuda et al, Nature Genetics 2008
 Unoki et al, Nature Genetics 2008
 Rung et al, Nature Genetics 2009
 Kong et al, Nature, 2009
 Dupuis et al, Nature Genetics 2010
 Saxena et al, Nature Genetics 2010
 Qi et al, 2010
Voight et al, Nature Genetics 2010

12 nouveaux identifiés dans cette étude

| | |
|-----------------|------------------|
| <i>ADCY5</i> | <i>GCK</i> |
| <i>GCKR</i> | <i>PROX1</i> |
| <i>PROX1</i> | <i>DGKB</i> |
| <i>TP53INP1</i> | <i>HCCA2</i> |
| <i>KLF14</i> | <i>RBMS1</i> |
| <i>ZBED3</i> | <i>DUSP9</i> |
| <i>BCL11A</i> | <i>KCNQ1 [2]</i> |
| <i>CHCHD9</i> | <i>ZFAND6</i> |
| <i>HNF1A</i> | <i>PRC1</i> |
| <i>HMGA2</i> | <i>CENTD2</i> |

Études Gènes candidats

Études Association Génome Entier D2T

EAGE traits quantitatifs

pré-EAGE – gènes candidats

PPARG

KCNJ11

TCF7L2

HNF1B

JAZF1

IRS1

2000

2003

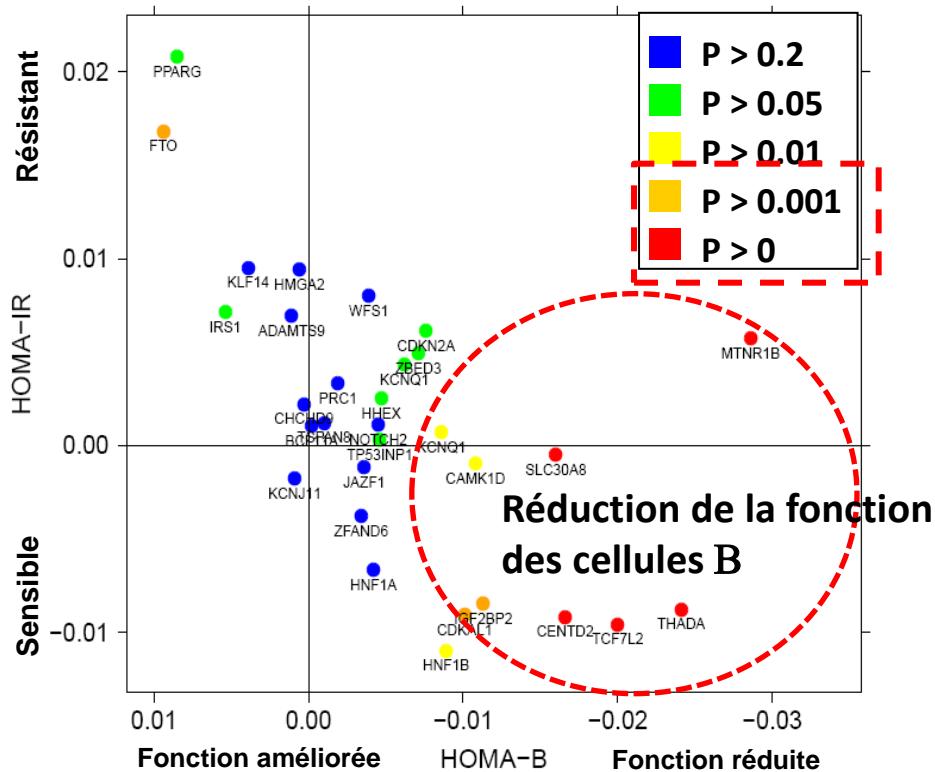
2006

2007

2008

2009/2010

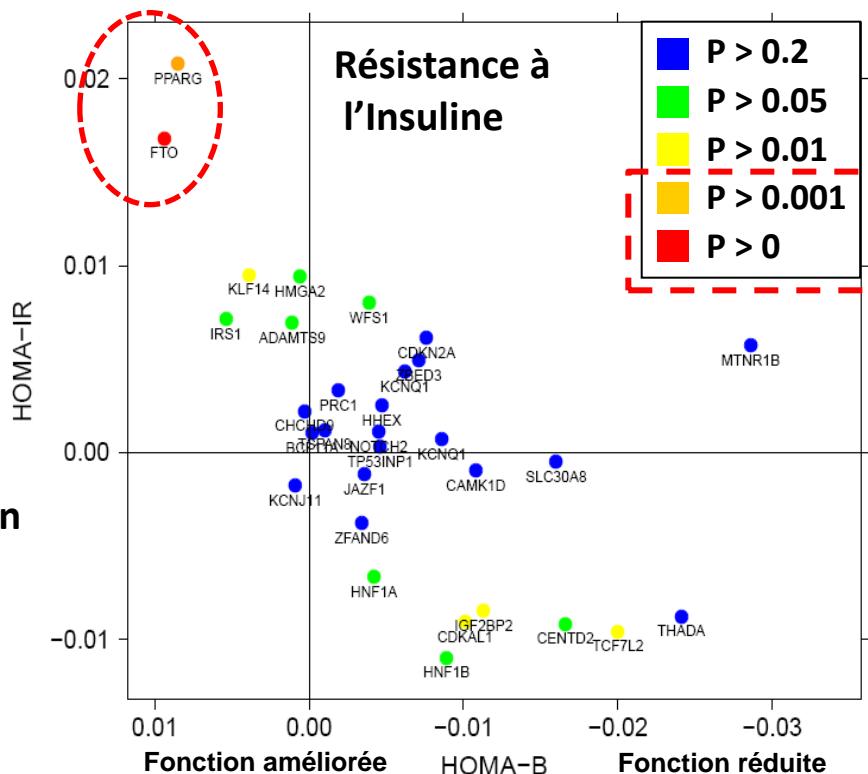
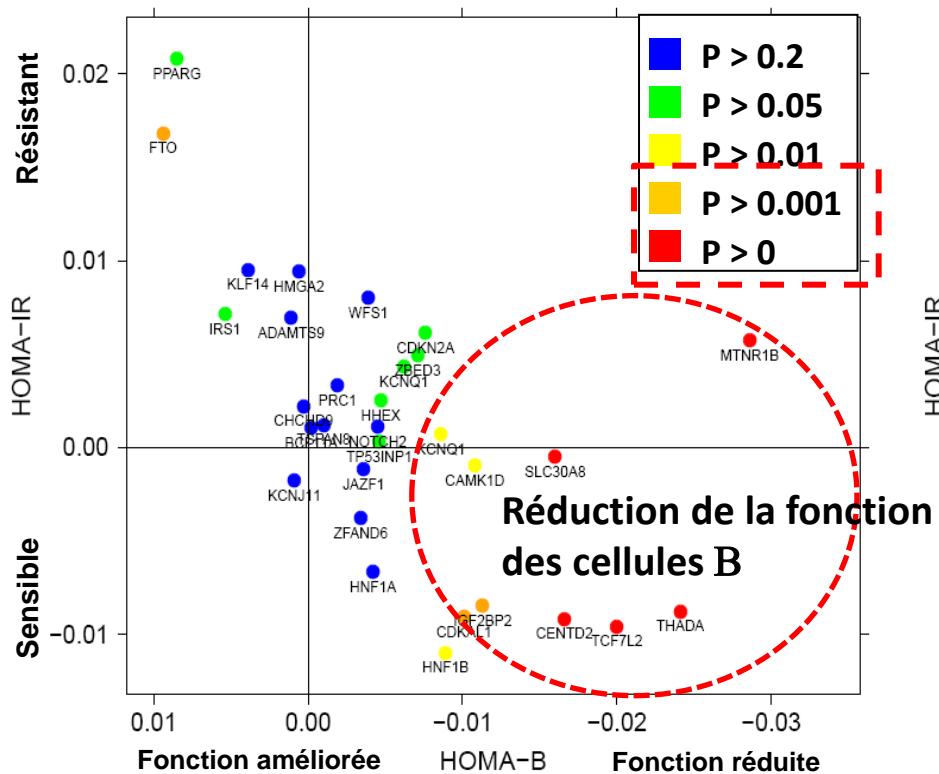
Gènes et physiologie



37,000 individus Non-diabétiques Glycémie à Jeun <7mg

Effets sur le fonction des cellules Beta

Gènes et physiologie



37,000 individus Non-diabétiques Glycémie à Jeun <7mg

Peu d'Effets sur la fonction d'Insulino - Résistance

Effet en population

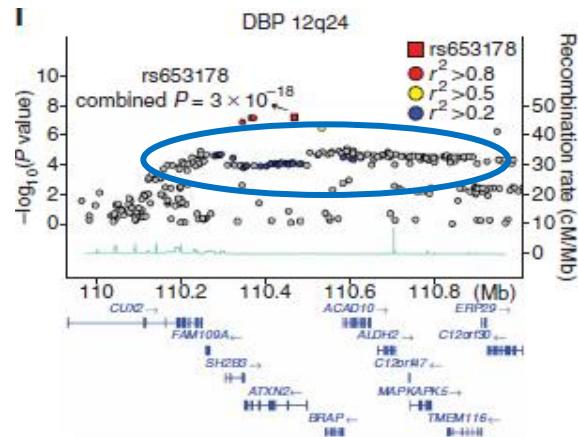
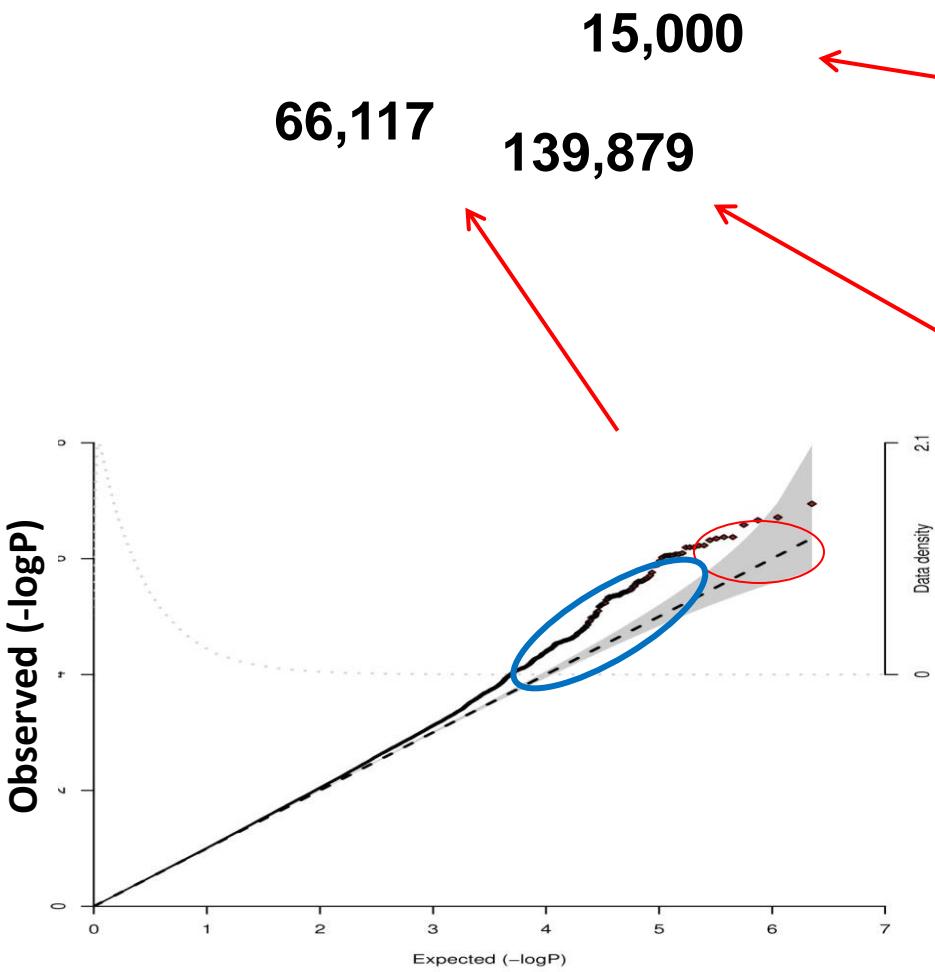
En population 39 variants

Héritabilité ~30% ~ 5%

λ_s 3 1.16

**Variants associé : Faible proportion de la composante génétique
Tous variants : 10 % de la composante génétique (héritabilité)**

METABOCHIP ~ 200 k SNPs



Dernière phase

Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes

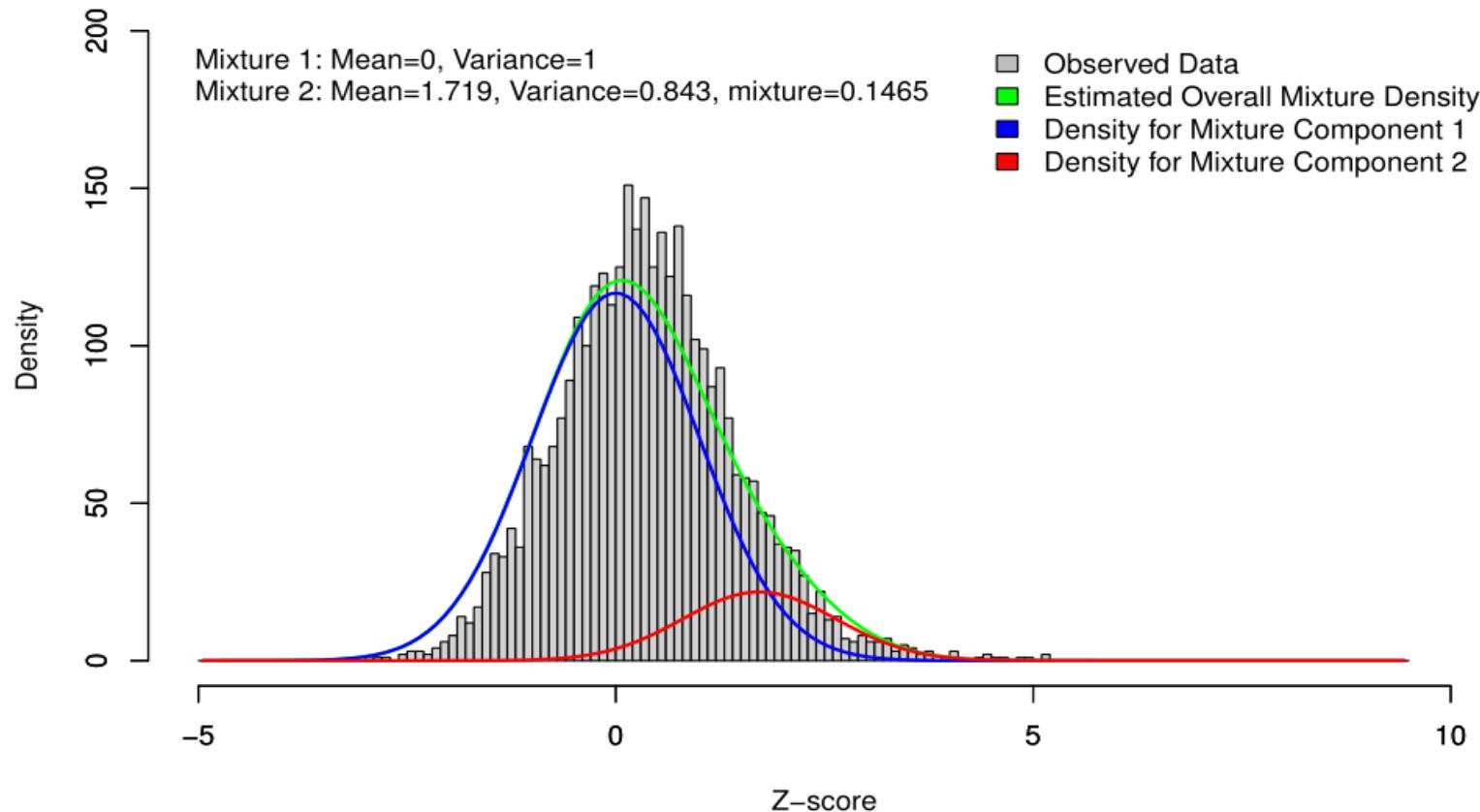
To extend understanding of the genetic architecture and molecular basis of type 2 diabetes (T2D), we conducted a meta-analysis of genetic variants on the Metabochip, including 34,840 cases and 114,981 controls, overwhelmingly of European descent. We identified ten previously unreported T2D susceptibility loci, including two showing sex-differentiated association. Genome-wide analyses of these data are consistent with a long tail of additional common variant loci explaining much of the variation in susceptibility to T2D. Exploration of the enlarged set of susceptibility loci implicates several processes, including CREBBP-related transcription, adipocytokine signaling and cell cycle regulation, in diabetes pathogenesis.

Nouveaux loci - GWAs

| Locus | Lead SNP | Chr | Position (Build 36 bp) | Combined meta-analysis p-value | Stage 2 risk allele frequency | Stage 2 OR (95% CI) | Sibling relative risk ^a | Explained liability-scale variance (%) ^b |
|--|------------|-----|---------------------------|--------------------------------|-------------------------------|---------------------|------------------------------------|---|
| Novel susceptibility loci achieving genome-wide significance in combined meta-analysis | | | | | | | | |
| ZMZ1 | rs12571751 | 10 | 80 612 637 | 1,0E-10 | 0,52 | 1.07 (1.04-1.10) | 1,001 | 0,066 |
| ANK1 | rs516946 | 8 | 41 638 405 | 2,5E-10 | 0,76 | 1.08 (1.05-1.12) | 1,001 | 0,059 |
| KLHDC5 | rs10842994 | 12 | 27 856 417 | 6,1E-10 | 0,80 | 1.10 (1.07-1.14) | 1,001 | 0,079 |
| TLE1 | rs2796441 | 9 | 83 498 768 | 5,4E-09 | 0,57 | 1.07 (1.04-1.10) | 1,001 | 0,064 |
| ANKRD55 | rs459193 | 5 | 55 842 508 | 6,0E-09 | 0,70 | 1.10 (1.06-1.13) | 1,002 | 0,106 |
| CILP2 | rs10401969 | 19 | 19 268 718 | 7,0E-09 | 0,08 | 1.14 (1.08-1.20) | 1,001 | 0,082 |
| MC4R | rs12970134 | 18 | 56 035 730 | 1,2E-08 | 0,27 | 1.08 (1.05-1.11) | 1,001 | 0,070 |
| BCAR1 | rs7202877 | 16 | 73 804 746 | 3,5E-08 | 0,89 | 1.10 (1.05-1.15) | 1,001 | 0,047 |
| Total | | | | | | | 1,010 | 0,574 |
| Combined total | | | | | | | 1,104 | 5,730 |

Proportion de la variance très petite

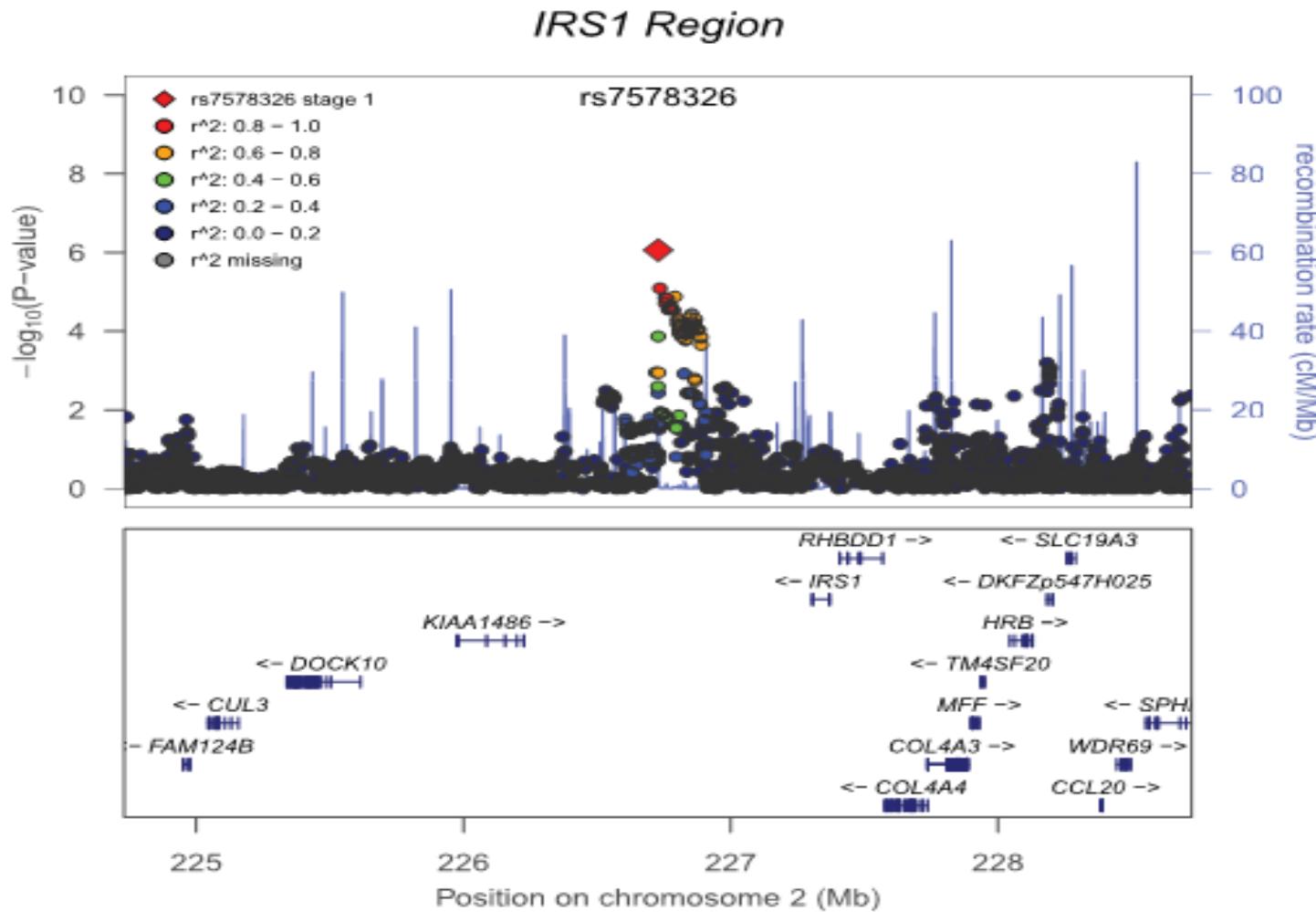
Estimation des vrais positifs



488 nouveaux loci ([456-521])

110

Résultats – fine mapping



Conclusion - temporaire

- Analyse d'association permet d'identifier des variants associés.
- Fait ses preuves dans le cas de pathologies complexes.
- Mais ne pointent pas sur un gène
- Expliquent une toute petite partie de la variance
- Effets individuels faibles

Etude pan-génomiques pour la taille

113

Taille est un phénotype modèle

- Peu de variabilité environnementale
- Pas de variation culturelle
- Mesure phénotypique parfaite
- Pas d'interaction gène x environnement évidente

Resultats sur 63 000 individus

| <u>Etude</u> | | <u>Indiv R2</u> | <u>Total R2</u> |
|--|---------|-----------------|-----------------|
| Weedon et al. | 20 SNPs | 0.01-0.3% | 2.9% |
| Lettre et al | 10 SNPs | 0.1-0.08% | 2.0% |
| Gudbjartsson et al | 21 SNPs | 0.1-0.17 | 3.7% |
| Variants identifiés dans les 3 papiers | | 6 | |



The case of the missing heritability

When scientists opened up the human genome, they expected to find the genetic components of common traits and diseases. But they were nowhere to be seen. **Brendan Maher** shines a light on six places where the missing loot could be stashed away.

Maher, B. (2008). Personal genomes: The case of the missing heritability. *Nature*, 456(7218), 18-21.

Héritabilité perdue

Table 1 | Estimates of heritability and number of loci for several complex traits

| Disease | Number of loci | Proportion of heritability explained | Heritability measure |
|---|----------------|--------------------------------------|-------------------------------|
| Age-related macular degeneration ⁷² | 5 | 50% | Sibling recurrence risk |
| Crohn's disease ²¹ | 32 | 20% | Genetic risk (liability) |
| Systemic lupus erythematosus ⁷³ | 6 | 15% | Sibling recurrence risk |
| Type 2 diabetes ⁷⁴ | 18 | 6% | Sibling recurrence risk |
| HDL cholesterol ⁷⁵ | 7 | 5.2% | Residual* phenotypic variance |
| Height ¹⁵ 180 000 | 40 | 5% | Phenotypic variance |
| Early onset myocardial infarction ⁷⁶ | 9 | 2.8% | Phenotypic variance |
| Fasting glucose ⁷⁷ | 4 | 1.5% | Phenotypic variance |

* Residual is after adjustment for age, gender, diabetes.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747-753.

Méthodes multi-SNPs

Modèles polygéniques

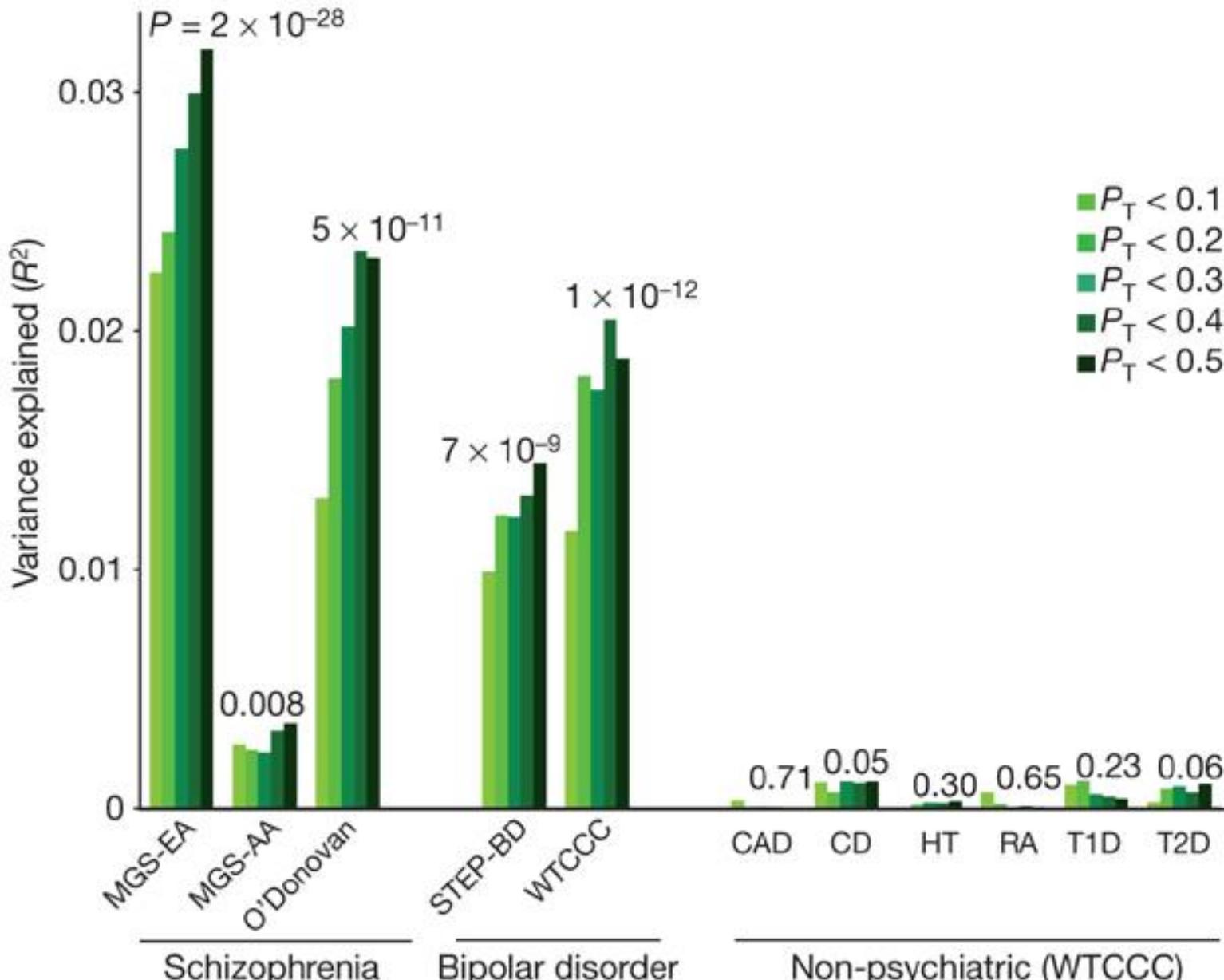
- Effet simultané de variants à effet faible?
- Crédit d'un score:

$$x_j = \frac{\sum_{i=1}^m \ln(OR_i) \times SNP_{ij}}{m}$$

où

- $\ln(OR_i)$ = OR pour le SNP_i dans l'échantillon initial
- SNP_{ij} = # d'alleles (0,1,2) pour SNP_i, individu j dans l'échantillon de validation.
- Réalisé sur un grand nombre de SNPs (m)
- x_j est-il associé à la maladie ?

Application of Model



Héritabilité sur tout le génome

Régression linéaire sur tous les SNPs du génome

Prise en compte du degré d'apparentement en utilisant les SNPs
(Estimation de du degré d'apparentement – proche des méthodes de recherche de stratification qui seront présentées)

Où est l'héritabilité?

Table 1. Population Variation Explained by GWAS for a Selected Number of Complex Traits

| Trait or Disease | h^2 Pedigree Studies | h^2 GWAS Hits ^a | h^2 All GWAS SNPs ^b |
|------------------------|----------------------------|------------------------------|----------------------------------|
| Type 1 diabetes | 0.9 ⁹⁸ | 0.6 ^{99, c} | 0.3 ¹² |
| Type 2 diabetes | 0.3–0.6 ¹⁰⁰ | 0.05–0.10 ³⁴ | |
| Obesity (BMI) | 0.4–0.6 ^{101,102} | 0.01–0.02 ³⁶ | 0.2 ¹⁴ |
| Crohn's disease | 0.6–0.8 ¹⁰³ | 0.1 ¹¹ | 0.4 ¹² |
| Ulcerative colitis | 0.5 ¹⁰³ | 0.05 ¹² | |
| Multiple sclerosis | 0.3–0.8 ¹⁰⁴ | 0.1 ⁴⁵ | |
| Ankylosing spondylitis | >0.90 ¹⁰⁵ | 0.2 ¹⁰⁶ | |
| Rheumatoid arthritis | 0.6 ¹⁰⁷ | | |
| Schizophrenia | 0.7–0.8 ¹⁰⁸ | 0.01 ⁷⁹ | 0.3 ¹⁰⁹ |
| Bipolar disorder | 0.6–0.7 ¹⁰⁸ | 0.02 ⁷⁹ | 0.4 ¹² |
| Breast cancer | 0.3 ¹¹⁰ | 0.08 ¹¹¹ | |

| | | | |
|-----------------------|------------------------------|------------------------|----------------------|
| Von Willebrand factor | 0.66–0.75 ^{112,113} | 0.13 ¹¹⁴ | 0.25 ¹⁴ |
| Height | 0.8 ^{115,116} | 0.1 ¹³ | 0.5 ^{13,14} |
| Bone mineral density | 0.6–0.8 ¹¹⁷ | 0.05 ¹¹⁸ | |
| QT interval | 0.37–0.60 ^{119,120} | 0.07 ¹²¹ | 0.2 ¹⁴ |
| HDL cholesterol | 0.5 ¹²² | 0.1 ⁵⁷ | |
| Platelet count | 0.8 ¹²³ | 0.05–0.1 ⁵⁸ | |

^a Proportion of phenotypic variance or variance in liability explained by genome-wide-significant and validated SNPs. For a number of diseases, other parameters were reported, and these were converted and approximated to the scale of total variation explained. Blank cells indicate that these parameters have not been reported in the literature.

^b Proportion of phenotypic variance or variance in liability explained when all GWAS SNPs are considered simultaneously. Blank cell indicate that these parameters have not been reported in the literature.

^c Includes pre-GWAS loci with large effects.

Héritabilité manquante

... moins que prévu

Variants rares

Genetic Heterogeneity in Human Disease

Jon McClellan^{1,*} and Mary-Claire King^{2,*}

¹Department of Psychiatry

²Departments of Medicine and Genome Sciences

University of Washington, Seattle, WA 98195-7720, USA

*Correspondence: drjack@uw.edu (J.M.), mcking@uw.edu (M.-C.K.)

DOI 10.1016/j.cell.2010.03.032

Strong evidence suggests that rare mutations of severe effect are responsible for a substantial portion of complex human disease. Evolutionary forces generate vast genetic heterogeneity in human illness by introducing many new variants in each generation. Current sequencing technologies offer the possibility of finding rare disease-causing mutations and the genes that harbor them.

Héritabilité manquante / variants rares

OPEN  ACCESS Freely available online

PLOS GENETICS

Fine Mapping of Five Loci Associated with Low-Density Lipoprotein Cholesterol Detects Variants That Double the Explained Heritability

Serena Sanna^{1*}, Bingshan Li^{2*}, Antonella Mulas¹, Carlo Sidore^{1,2,3}, Hyun M. Kang², Anne U. Jackson², Maria Grazia Piras¹, Gianluca Usala¹, Giuseppe Maninchetta⁴, Alessandro Sassi⁴, Fabrizio Serra⁴, Maria Antonietta Palmas⁴, William H. Wood III⁵, Inger Njølstad⁶, Markku Laakso⁷, Kristian Hveem⁸, Jaakko Tuomilehto^{9,10,11}, Timo A. Lakka^{12,13}, Rainer Rauramaa¹³, Michael Boehnke², Francesco Cucca^{1,3}, Manuela Uda¹, David Schlessinger^{14†}, Ramaiah Nagaraja^{14†}, Gonçalo R. Abecasis^{2†*}

Table 2. Heritability estimates in all study samples.

| Study | N samples | Variance explained by 5 GWAS SNPs | Variance explained by 8 SNPs |
|-------------------------------|-----------|-----------------------------------|------------------------------|
| SardiNIA | 5,382 | 3.1% | 6.5% |
| Norwegian T2D | 1,171 | 5.8% | 9.3% |
| Norwegian controls | 1,436 | 3.1% | 8.5% |
| Finnish T2D | 1,742 | 2.1% | 5.0% |
| Finnish controls | 5,678 | 3.4% | 7.0% |
| Average Finnish and Norwegian | 10,027 | 3.5% | 7.1% |

The table shows the LDL-C variance accounted for by the 5 GWAS SNPs and the 8 SNPs here described in all studies. A sample size weighted average estimate is given for the Finnish and Norwegian samples.

doi:10.1371/journal.pgen.1002198.t002

Conclusion générale

- Résultats décevants quant à la force individuelle des variants
- Héritabilité manquante – mais moins que prévu
- Utilité du modèle polygénique ?
- Il reste de la place pour des variants rares