Analyses d’association pan-génomiques (GWAS) et architecture génétique des caractères complexes

UE Du gène à la pathologie

Bertrand Servin (INRAe) bertrand.servin@inrae.fr

Architecture génétique des caractères

Définitions

- Les caractères observables (**phénotypes**) : issus de nombreux processus biologiques depuis le développement embryonnaire, la croissance, la vie des individus

- Le **déterminisme** d'un caractère provient des gènes, de l'environnement (au sens large) et de leur interaction

$$P = G + E + G \times E$$

- Les **différences entre individus** (variabilité): différences génétiques (polymorphismes) + différences environnementales

- En médecine : l'**étiologie** est l'étude des déterminismes d'une maladie

- L'étude de l'**architecture génétique des caractères** a pour but d'identifier les polymorphismes (leur nombre, leurs localisations, leurs effets) responsables des différences individuelles

Architecture génétique des caractères

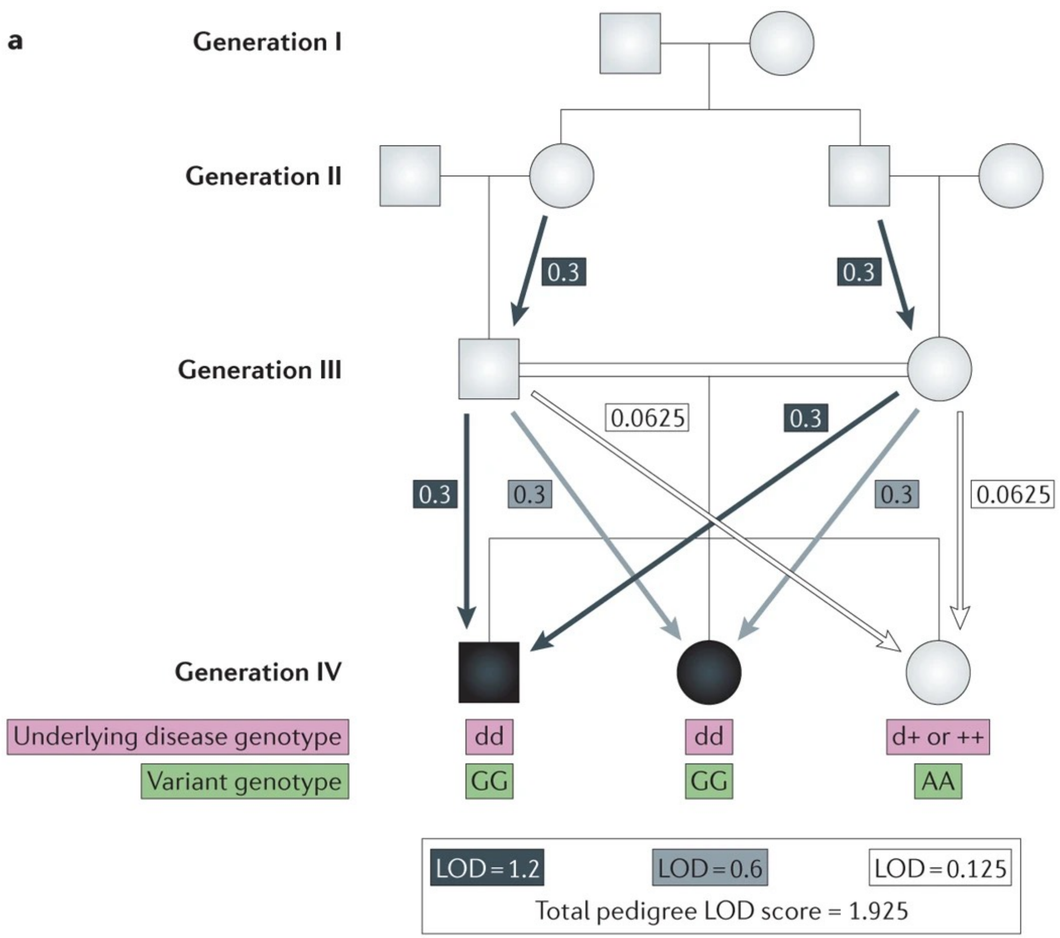
Objectifs

- Comprendre la **biologie des caractères** : identification de nouveaux traitements, design de nouveaux médicaments …
- **Diagnostique génétique**, permet de guider les traitements (*e.g.* pharmacogénétique), avis médical personnalisé …
- Séparer les effets G et E : meilleure **estimation des facteurs de risques non génétiques**
- En agronomie : sélection des individus pour améliorer les performances agronomiques (production, environnement, robustesse …)

Analyses familiales

Analyse de liaison (Linkage Analysis)

- Recueil de pedigrees (familles) dont les membres sont atteints par une maladie
- Hypothèse : 1 gène causal
- Analyse de ségrégation = assignation des génotypes possibles sur la base des phénotypes
- Génotypage des membres du pedigree
- Génotypes aux marqueurs ?==? génotype causal



Exemple d'analyse génétique dans un pedigree simple (tiré de Ott et al. *Nat Rev Genet* **16,** 275–284 (2015))

Analyse de liaison : maladies monofactorielles

- Monogéniques , mendéliennes
- Contrôlées par un faible nombre de locus
- Une mutation, récessive ou co-dominante, cause la maladie. Effet fort
- Maladie rare et souvent très grave
- Exemples : mucoviscidose, anémie falciforme, …

Analyses familiales

Limitations

- Hétérogénéité génétique (différentes familles = différents gènes)
- Caractères / maladies complexes, multigéniques
- Recrutement de familles
- Puissance statistique inhérente à cette approche
- Transition vers les **études d'association**

The Future of Genetic Studies of Complex Human Diseases

Neil Risch and Kathleen Merikangas

1996 - Science 273:1516-1517

		Linkage			Association			
					Singletons		Sib pairs	
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)	Proportion of heterozygous parents (Het)	(N)	(Het)	(N)
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.

Analyses d’association

Principe et critères d’application

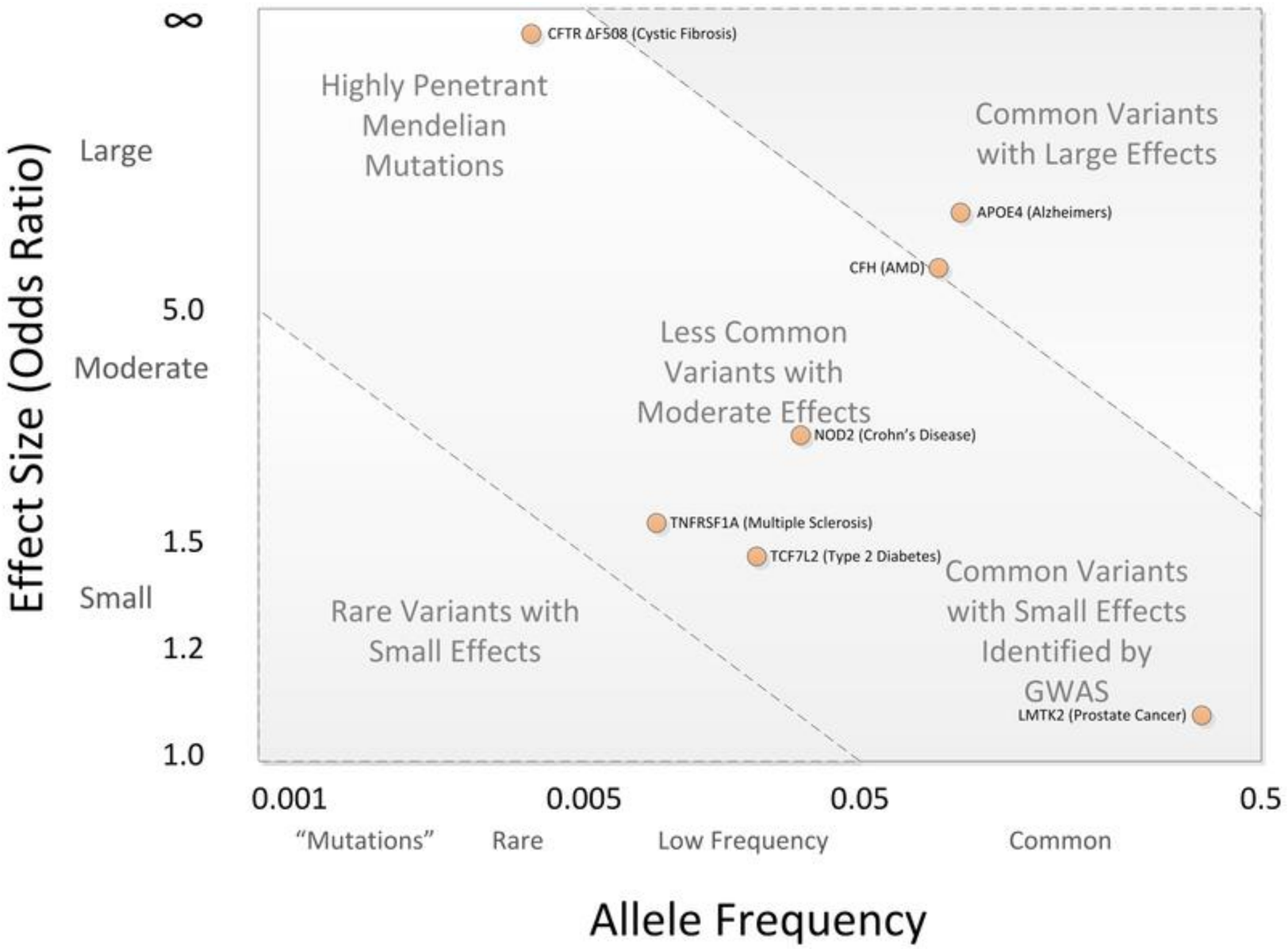
- Recruter un grand nombre d’individus peu ou pas apparentés
- Cas/Contrôles dans le cas de maladies
- Génotyper un (très) grand nombre de marqueurs génétiques
- Tester l’association entre polymorphismes et phénotypes

« La principale limitation des études d’association pan-génomiques [est technologique]. Un grand nombre de (…) polymorphismes doivent d’abord être identifiés et testés.

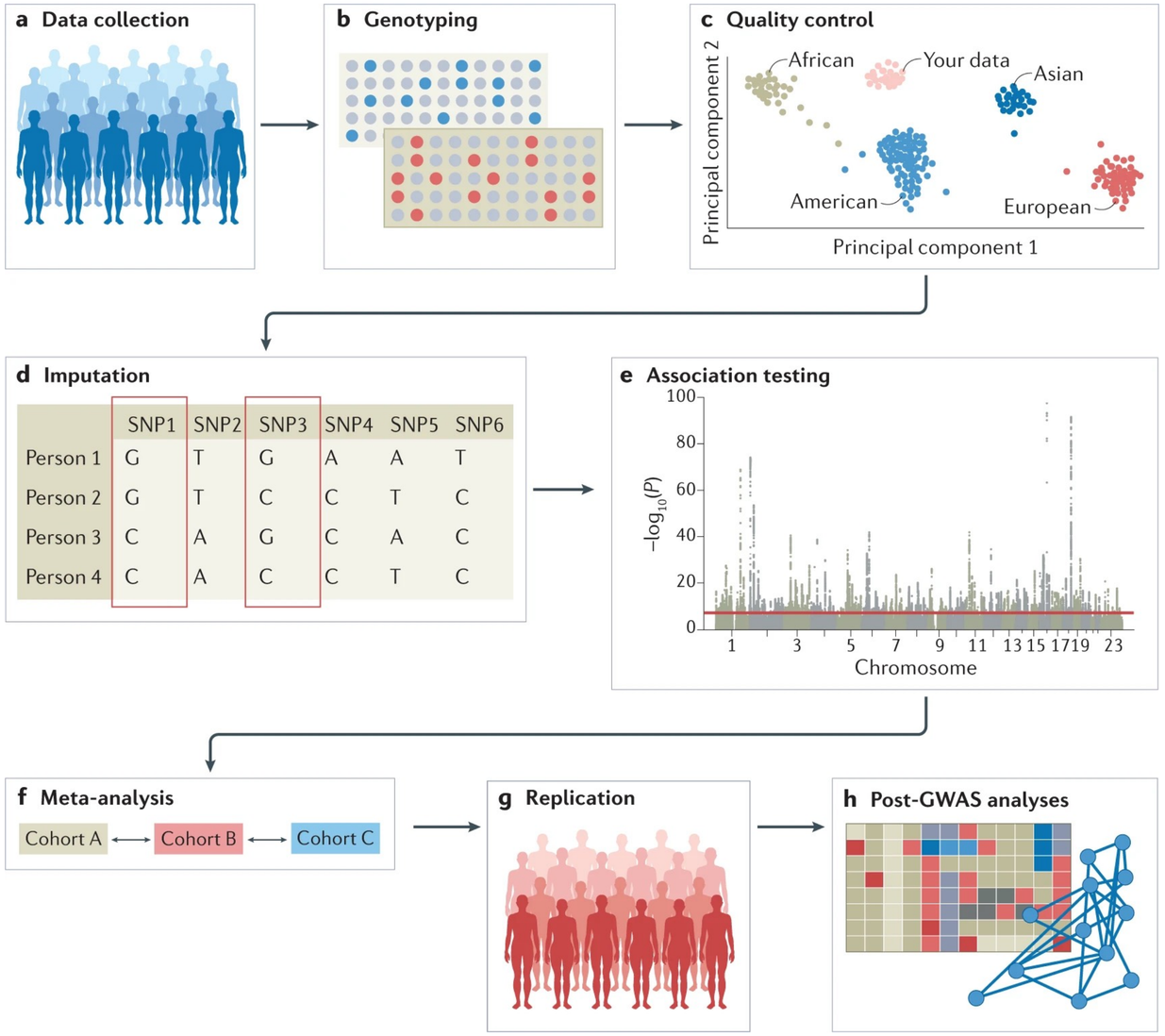
(…)

C’est aux technologies moléculaires de développer les outils pour remporter ce défi et produire l’information nécessaire à l’identification des bases génétiques des caractères complexes. »

Risch et Merikangas Science (1996)


Effect Size (Odds Ratio)
∞
Large
5.0
Moderate
1.5
Small
1.2
1.0
CFTR ΔF508 (Cystic Fibrosis)
Highly Penetrant Mendelian Mutations
Common Variants with Large Effects
APOE4 (Alzheimers)
CFH (AMD)
Less Common Variants with Moderate Effects
NOD2 (Crohn's Disease)
TNFRSF1A (Multiple Sclerosis)
TCF7L2 (Type 2 Diabetes)
Common Variants with Small Effects Identified by GWAS
LMTK2 (Prostate Cancer)
Rare Variants with Small Effects
0.001
0.005
0.05
0.5
"Mutations"
Rare
Low Frequency
Common
Allele Frequency

Etapes d'une analyse GWAS



Tiré de Uffelmann et al. (2021)

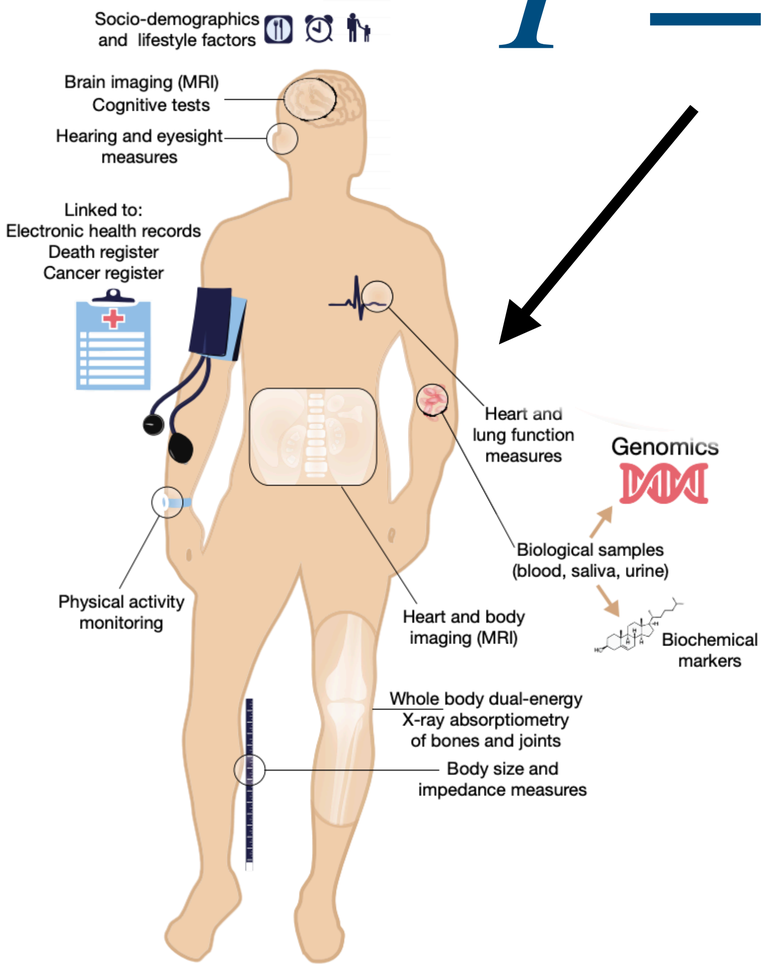
1. Tests statistiques d'association
2. Génotypage et Imputation
3. Structure Génétique
4. Analyses Post-GWAS et implications en génétique humaine

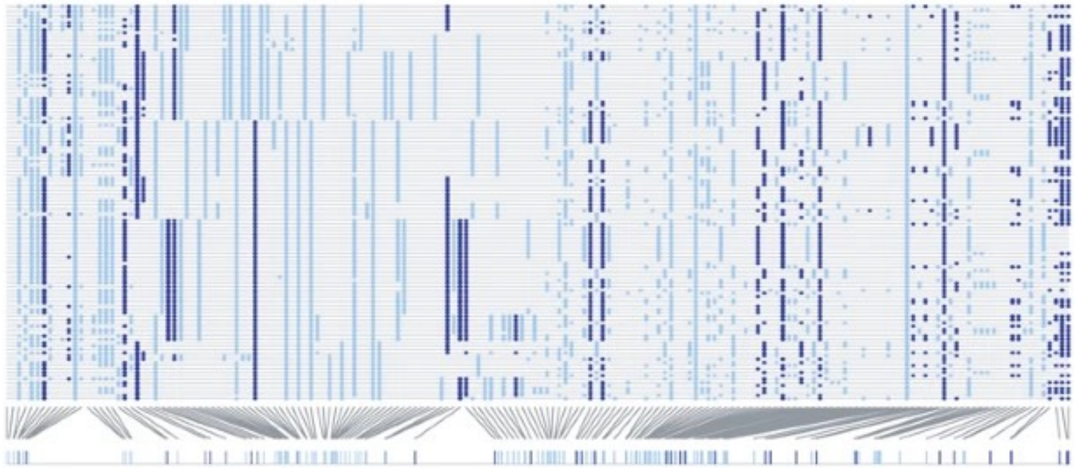
Analyses d’association pan-génomiques

1. Modèles et tests statistiques de l’association marqueur / phénotype
    1. Données génétiques: marqueurs SNPs, projet 1000 génomes, puces de génotypage
    2. Phénotype continu et phénotype binaire
    3. Prise en compte des tests multiples
2. Facteurs influençant la puissance statistique des GWAS
    a. Corrélation marqueur / QTN : déséquilibre de liaison
    b. Imputation de génotypes pour augmenter le nombre de polymorphismes testés
    c. Nombre d’individus : mise en commun d’études multiples et meta-analyse
3. Dangers de l’analyse d’association
    a. Structure de population et apparentement
    b. Approches statistiques pour corriger les effets de structure / apparentement

Principe des GWAS et de l’analyse statistique

$$Y = G + E$$



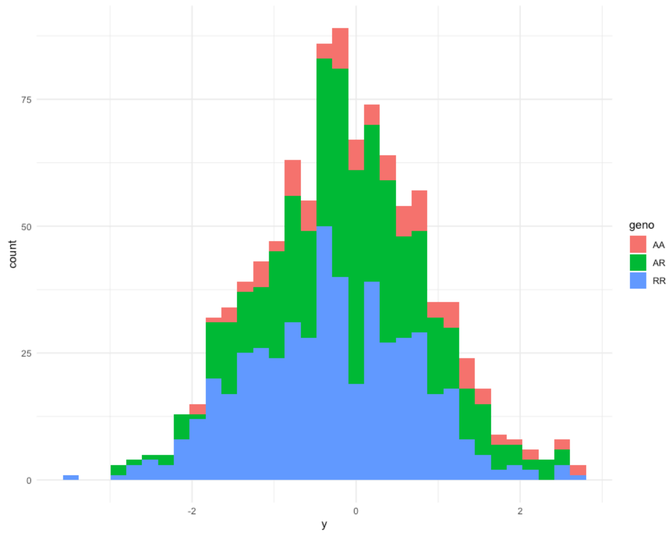


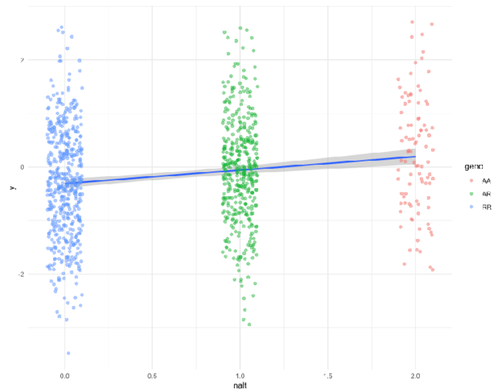
?

D'après Bycroft *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562,** 203–209 (2018)

Caractère continu et mutation causale

- On suppose un caractère mesuré sur une échelle continue
- Distribution Gaussienne (normale)
- Un polymorphisme bi-allélique causal
- Tester association génotype / phénotype = modèle linéaire simple





	term	estimate	std.error	statistic	p.value
1	mean	-0.0588	0.0373	-1.57	1.16e-01
2	effect	0.2521	0.0487	5.18	2.70e-07

$$\mathbf{y} = \mu + \mathbf{g}\beta + \mathbf{e}$$

Caractère binaire et mutation causale

- En génétique humaine, de nombreux caractères d'intérêt sont des maladies
- Phénotype binaire (malade/sain, cas/contrôle)
- Un polymorphisme bi-allélique causal
  - R : allèle de référence
  - A : allèle alternatif

Table de contingence

	AA	AR	RR	Total
CAS	52	194	228	474
CONTROLE	42	218	266	526
Total	94	412	494	1000

Table de fréquences

	AA	AR	RR
CAS	0.110	0.410	0.481
CONTROLE	0.080	0.414	0.506
Total	0.094	0.412	0.494

= ?

Caractère binaire et mutation causale

Tests d'association

Table de contingence

	AA	AR	RR	Total
CAS	52	194	228	474
CONTROLE	42	218	266	526
Total	94	412	494	1000

Table de fréquences

	AA	AR	RR
CAS	0.110	0.410	0.481
CONTROLE	0.080	0.414	0.506
Total	0.094	0.412	0.494

= ?

Score Test

Hypothèse nulle (H0) : les deux distributions sont égales

Effectifs attendus sous H0

	AA	AR	RR
CAS	44.56	195.29	234.15
CONTROLE	49.44	216.71	259.84

$$T = \sum \frac{(O-E)^2}{E} = \frac{(52-44.56)^2}{44.56} + \ldots = 2.6854$$

Caractère binaire et mutation causale

Tests d'association

Table de contingence

		AA	AR	RR	Total
CAS	**n**	52	194	228	474
CONTROLE	**m**	42	218	266	526
Total		94	412	494	1000

Table de fréquences

		AA	AR	RR
CAS	**q**	0.110	0.410	0.481
CONTROLE	**r**	0.080	0.414	0.506
Total	**p**	0.094	0.412	0.494

= ?

Test du rapport de vraisemblance

Hypothèse nulle (H0) : les deux distributions sont égales

$$P(Data\,|\,H_0) = L_0 \propto \prod p_i^{n_i+m_i}$$

Hypothèse alternative (H1) : les deux distributions sont différentes

$$P(Data\,|\,H_1) = L_1 \propto \prod q_i^{n_i} r_i^{m_i}$$

$$T = 2(\ln L_1 - \ln L_0) = 3.631971$$

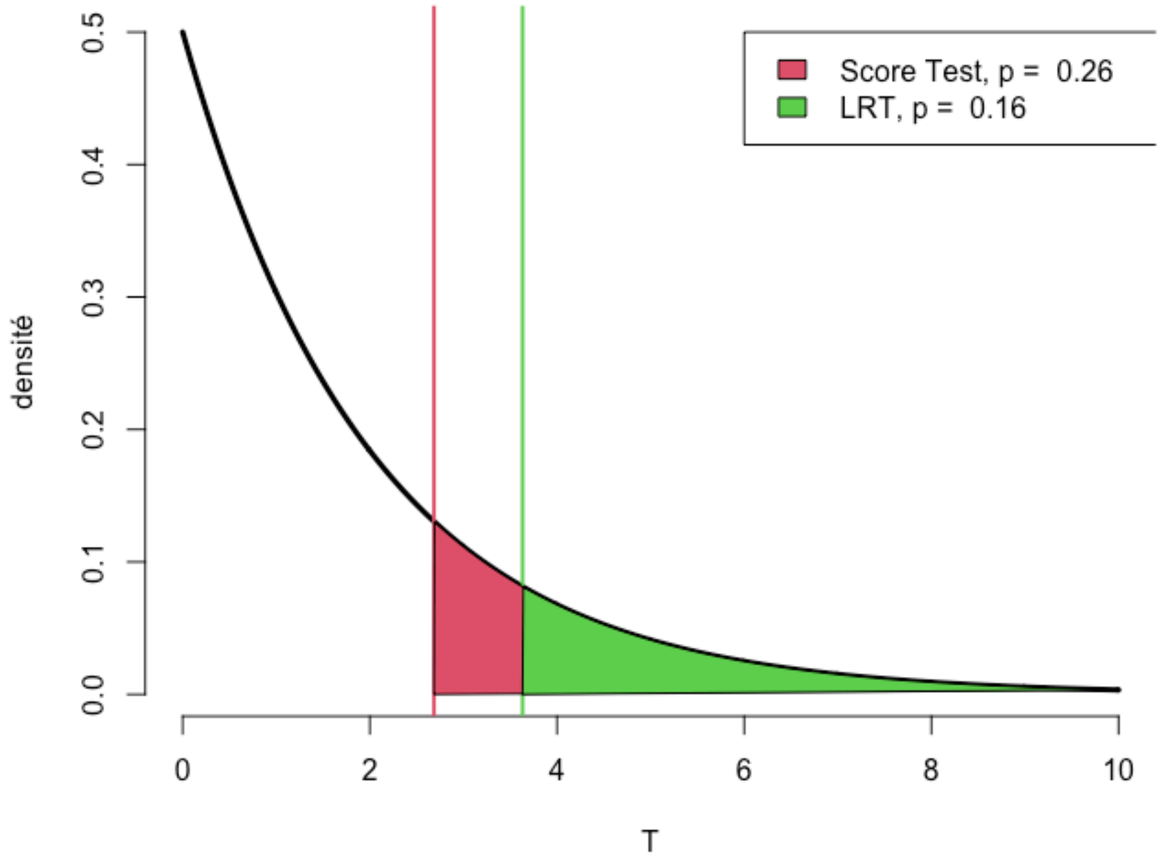
Caractère binaire et mutation causale

Tests d'association

$$T = \sum \frac{(O-E)^2}{E} = \frac{(52-44.56)^2}{44.56} + \ldots = 2.6854$$

$$T = 2(\ln L_1 - \ln L_0) = 3.631971$$

- Les deux tests suivent une loi du $\chi^2(2)$
- Test de l'hypothèse nulle : $P(t > T \mid H_0)$?
- Score Test p = 0.26
- LRT p = 0.16



Caractère binaire et mutation causale

Test allélique

Table de contingence

	A	R	Total
CAS	298	650	948
CONTROLE	302	750	1052
Total	600	1400	2000

Table de fréquences

	A	R
CAS	0.314	0.686
CONTROLE	0.287	0.713
Total	0.3	0.7

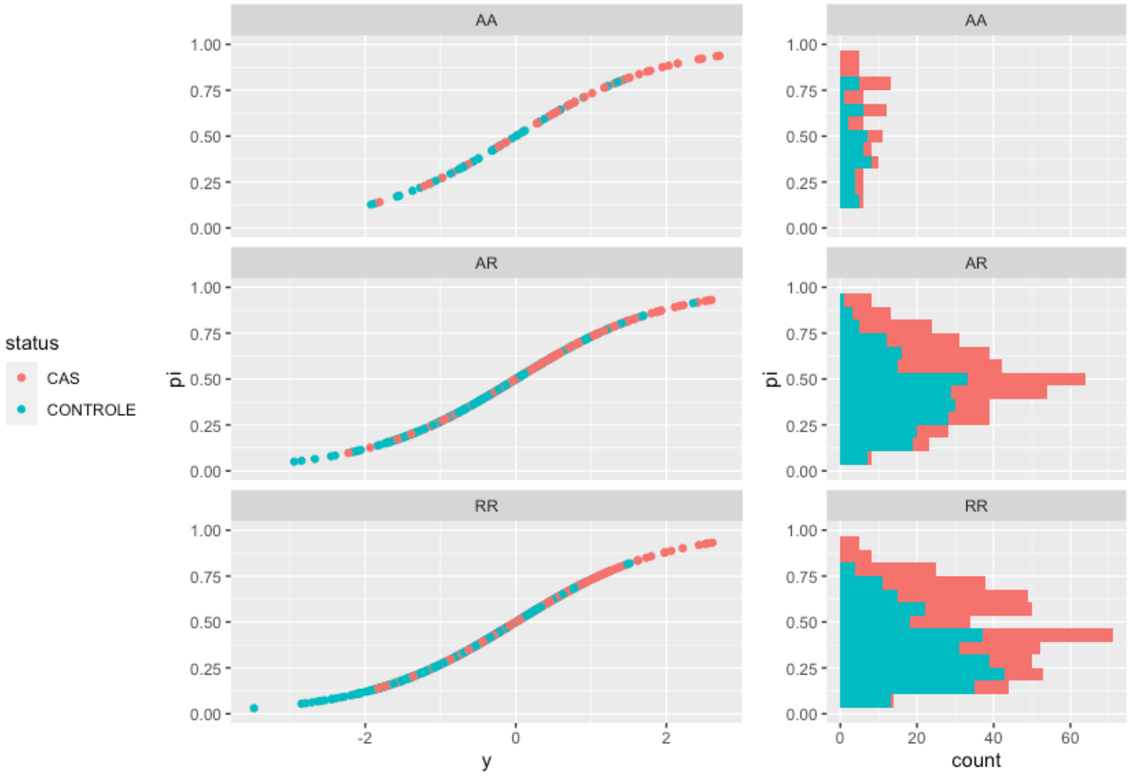
= ?

- test allélique : différence de distribution dans les fréquences des deux allèles
- $f_A = \dfrac{2n_{AA} + n_{AR}}{2(n_{AA} + n_{AR} + n_{RR})}$
- Score Test T = 1.7663, $\chi^2(1)$, p=0.1834
- LRT T=1.7646, $\chi^2(1)$, p=0.1840

Caractère binaire et mutation causale

Régression logistique

- En génétique humaine, de nombreux caractères d'intérêt sont des maladies
- Phénotype binaire (malade/ sain, cas/contrôle)
- Un polymorphisme bi-allélique causal
- Tester association génotype / phénotype = **régression logistique**



$$P(\text{CAS}) = \pi$$

$$\mathbf{y} = log(\frac{\pi}{1-\pi}) = \mu + \mathbf{g}\beta$$

Odds-ratio

Une mesure standard de l'effet d'une covariable sur une maladie

Modèle d'association génotypique

$$\log \frac{P(cas\,|\,RR)}{P(controle\,|\,RR)} = \mu$$

$$\log \frac{P(cas\,|\,AR)}{P(controle\,|\,AR)} = \mu + \beta_{AR}$$

$$\log \frac{P(cas\,|\,AA)}{P(controle\,|\,AA)} = \mu + \beta_{AA}$$

$$\log(Odds(AR)) - log(Odds(RR)) = \beta_{AR}$$

$$\log(Odds(AR)/Odds(RR)) = \beta_{AR}$$

$$OddsRatio(AR) = e^{\beta_{AR}}$$

Modèle d'association allélique

$$\log \frac{P(cas\,|\,RR)}{P(controle\,|\,RR)} = \mu$$

$$\log \frac{P(cas\,|\,AR)}{P(controle\,|\,AR)} = \mu + \beta$$

$$\log \frac{P(cas\,|\,AA)}{P(controle\,|\,AA)} = \mu + 2\beta$$

$$OddsRatio(AR) = e^{\beta}$$

$$OddsRatio(AA) = e^{2\beta} = OddsRatio(AR)^2$$

- OR > 1 (logOR>0): le génotype (l'allèle) confère un risque accru par rapport au génotype (l'allèle) de référence
- Les OR se multiplient, les logOR s'additionnent

Caractère binaire et mutation causale

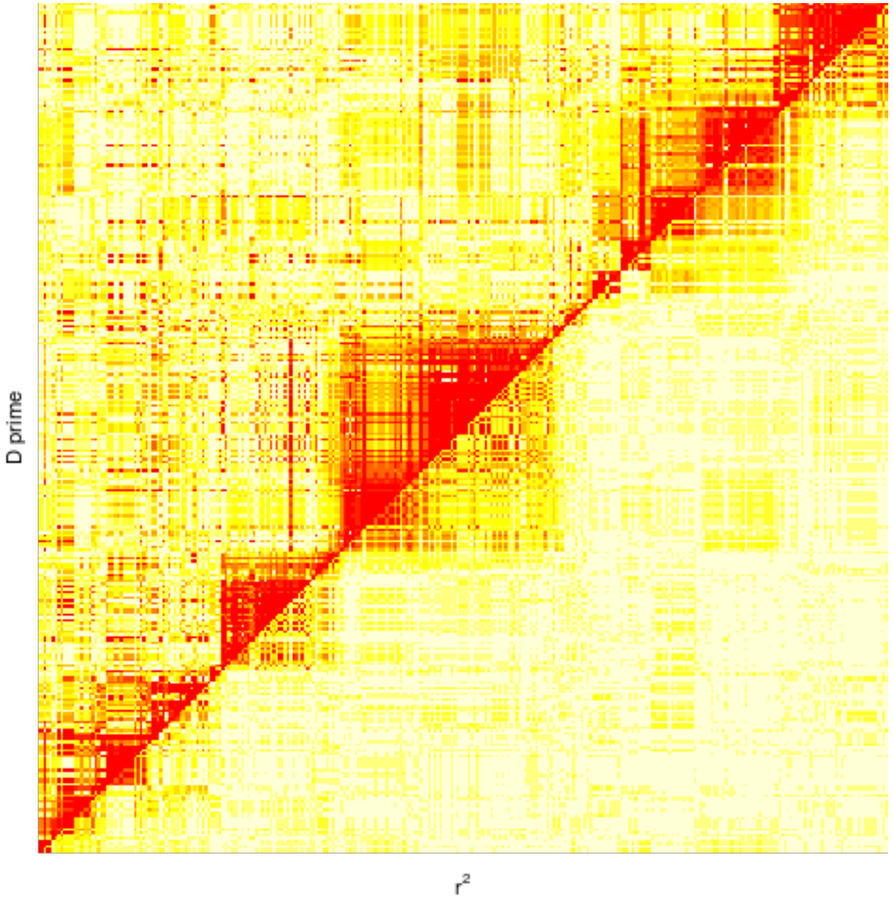
Régression logistique

- Permet d’inclure des effets additionnels pour corriger
  - Sexe, age …
  - Environnement (fumeur, poids, etc…)
  - …
- Plus souple que l’approche par tableau de contingence
- Donne les mêmes résultats que le LRT dans le cas simple

Déséquilibre de liaison

Mutation causale est inconnue

- Balayer le génome pour « ~~toutes~~ » les mutations possibles
- Un nombre suffisant = polymorphismes suffisamment denses pour être « proches » de la mutation causale
- Mesure de proximité = **Déséquilibre de liaison**



Déséquilibre de liaison

Association non aléatoire entre allèles à des locus différents

Caractéristique d’une population

Equilibre de liaison

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

Où pA (pB) est la fréquence de l’allèle A (resp. B) dans la population

Déséquilibre de liaison

	B	b	
A	$p_A p_B + \mathbf{D}$	$p_A(1-p_B) - \mathbf{D}$	p_A
a	$(1-p_A)p_B - \mathbf{D}$	$(1-p_A)(1-p_B) + \mathbf{D}$	$1-p_A$
	p_B	$1-p_B$	1

Où pA (pB) est la fréquence de l'allèle A (resp. B) dans la population

Déséquilibre de liaison

Mesure D

- D : écart à l'attendu aléatoire dans les fréquences gamétiques (haplotypiques)
- $D = f_{AB} - f_A f_B = \ldots = f_{AB} f_{ab} - f_{Ab} f_{aB}$
- D dépend des fréquences alléliques

	B	b	
A	$p_A p_B + \mathbf{D}$	$p_A(1-p_B) - \mathbf{D}$	p_A
a	$(1-p_A)p_B - \mathbf{D}$	$(1-p_A)(1-p_B) + \mathbf{D}$	$1-p_A$
	p_B	$1-p_B$	1

Déséquilibre de liaison

Mesure r^2

	B	b	
A	$p_A p_B + \mathbf{D}$	$p_A(1-p_B) - \mathbf{D}$	p_A
a	$(1-p_A)p_B - \mathbf{D}$	$(1-p_A)(1-p_B) + \mathbf{D}$	$1-p_A$
	p_B	$1-p_B$	1

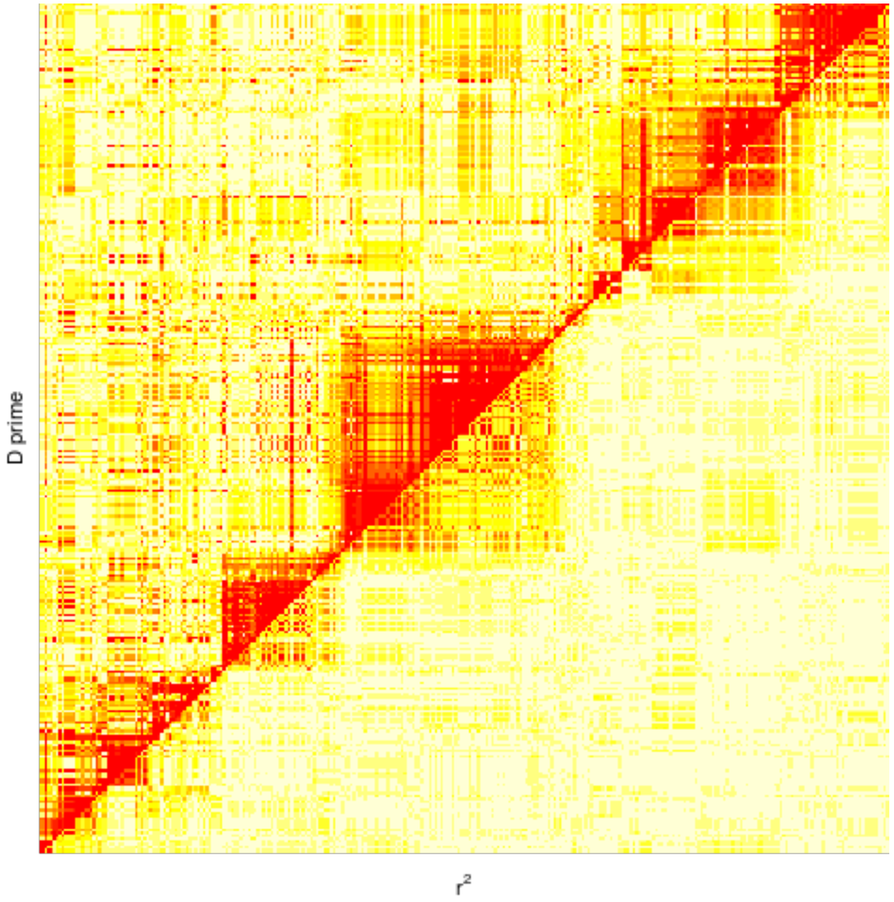
- D : écart à l'attendu aléatoire dans les fréquences gamétiques (haplotypiques)
- $D = f_{AB} - f_A f_B = \ldots = f_{AB} f_{ab} - f_{Ab} f_{aB}$
- D dépend des fréquences alléliques
- $r = \dfrac{D}{\sqrt{p_A(1-p_A)p_B(1-p_B)}}$ est la corrélation entre allèles aux locus A & B
- Très proche de la corrélation entre **génotypes**
- Influence la puissance du test d'association : $\mathscr{P} := n/r^2$, avec n: taille d'échantillon

DL et recombinaison

- Quel est le r^2 moyen à une distance génétique donnée ?
- Approximation pour p_A et p_B ~ 05 :

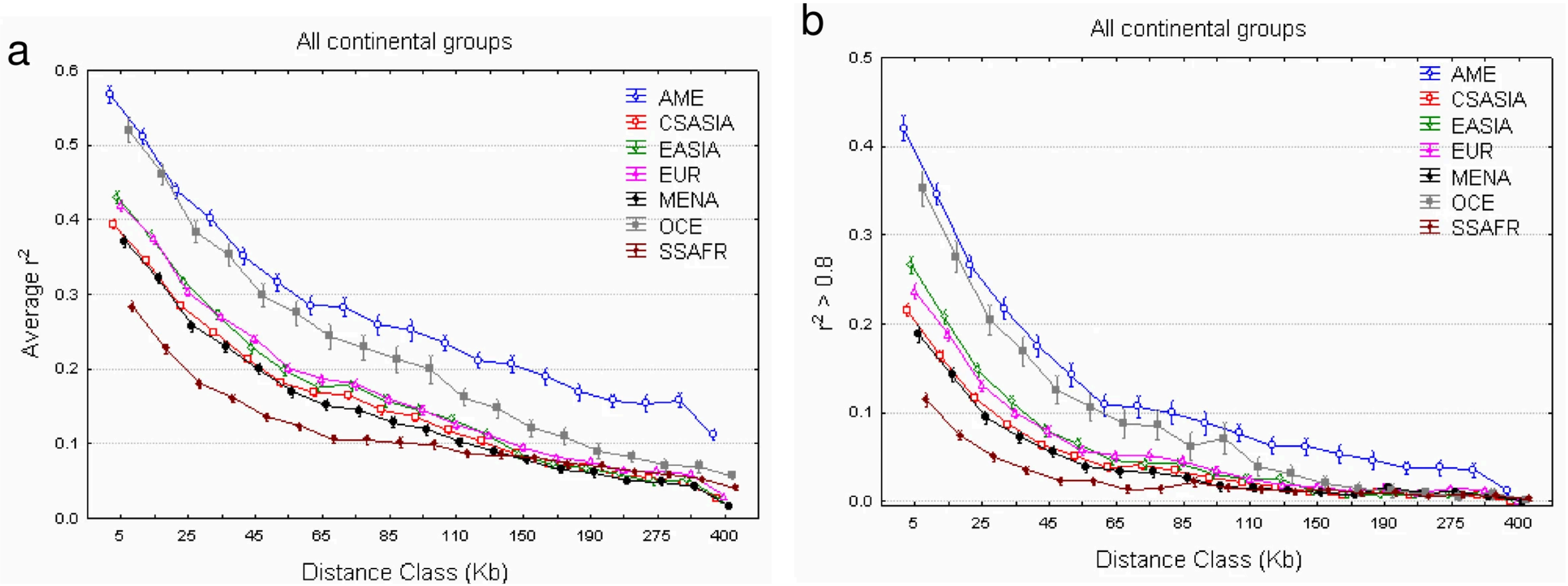
$$E(r^2) = \frac{1}{1 + 4Nc}$$

- N : taille de la population
- c : taux de recombinaison entre locus
- DL décroit en fonction du paramètre $\rho = 4Nc$



DL et recombinaison

Exemple dans des populations humaines



Tiré de Bosch et al. BMC genomics (2009)

1 SNP / 5 Kb ~ 600K SNP sur le génome

Génotypage pour les analyses d'association

- Objectif : détecter des mutations responsables de la variabilité génétique des caractères
- Il faut génotyper des polymorphismes en fort Déséquilibre de Liaison avec la mutation
  - *e.g.* $r^2 = 0.5$ = doubler la taille d'échantillon / test de la mutation causale
- Dans les populations humaines, le DL (r^2) décroit assez vite (~5Kb)
- Il faut génotyper beaucoup de marqueurs (> 600K) !

Analyses d’association

Principe et critères d’application

- Recruter un grand nombre d’individus peu ou pas apparentés
- Cas/Contrôles dans le cas de maladies
- Génotyper un (très) grand nombre de marqueurs génétiques
- Tester l’association entre polymorphismes et phénotypes

« La principale limitation des études d’association pan-génomiques [est technologique]. Un grand nombre de (…) polymorphismes doivent d’abord être identifiés et testés.

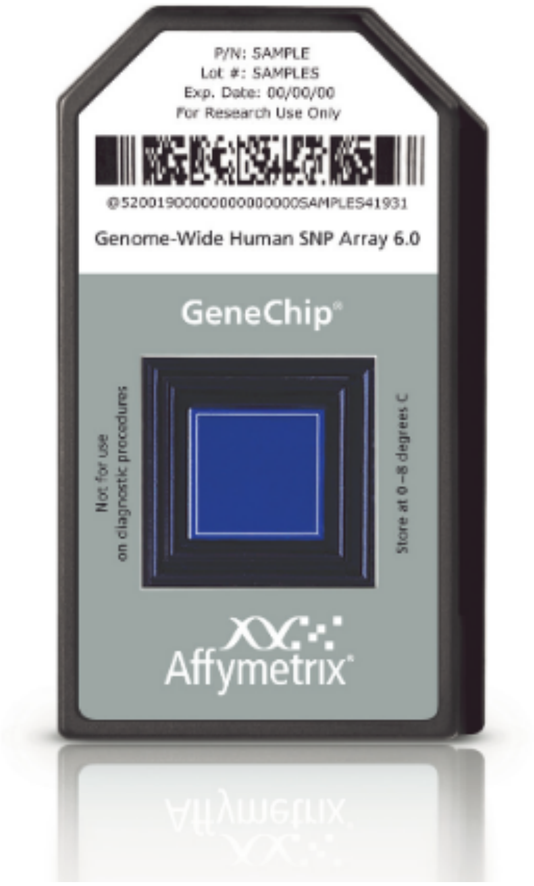
(…)

C’est aux technologies moléculaires de développer les outils pour remporter ce défi et produire l’information nécessaire à l’identification des bases génétiques des caractères complexes. *»*

Risch et Merikangas (1996)

Puces de génotypage haute densité

2 à 5 millions de SNPs + CNVs

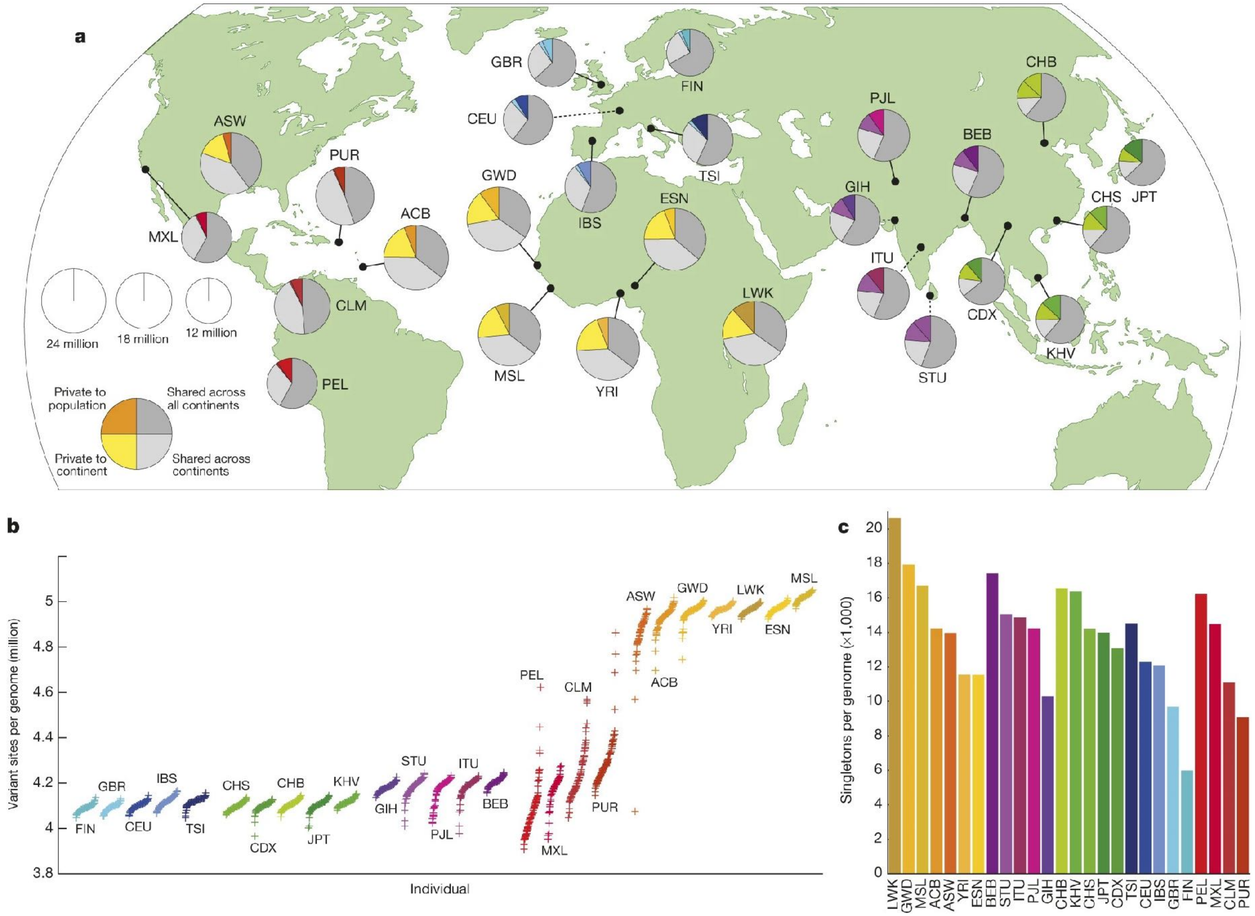


Imputation de génotypes

Panels de diversité génétique

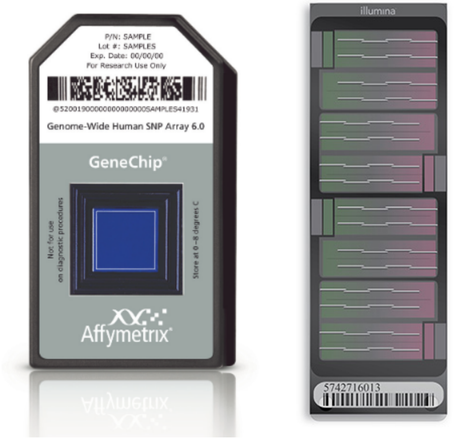
Ex: projet 1000 génomes humain

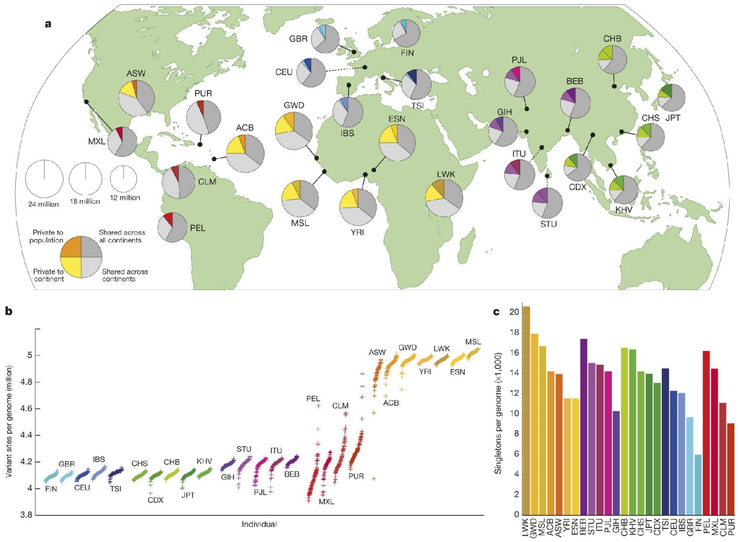
- Reséquençage complete de milliers de génomes
- Découverte de l'ensemble des polymorphismes communs
- Nombreux polymorphismes partagés, quelques spécifiques (continent, population)
- Catalogue de la diversité génétique humaine

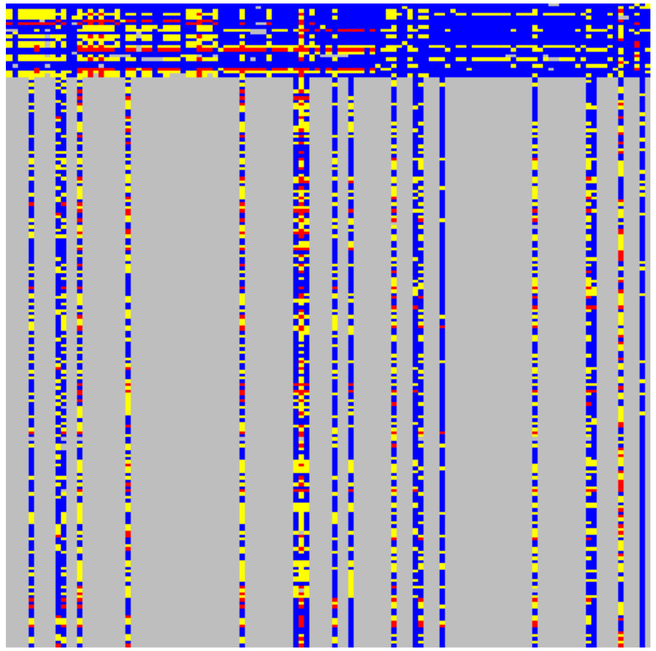


Tiré de The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015)

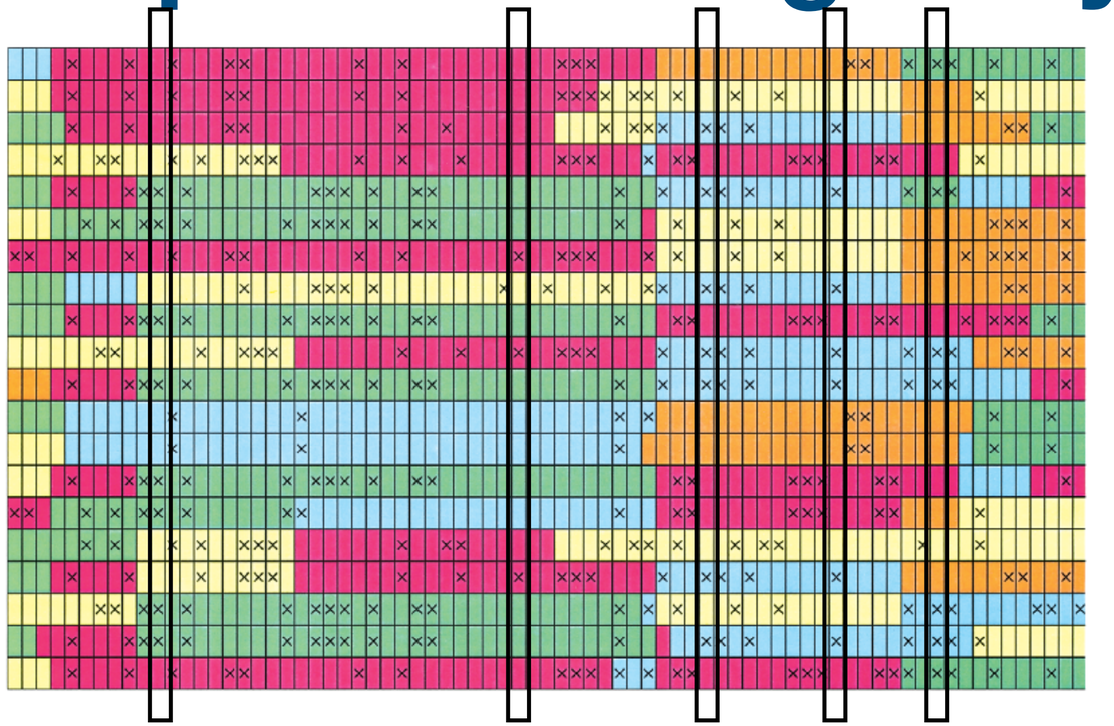
Imputation de génotypes

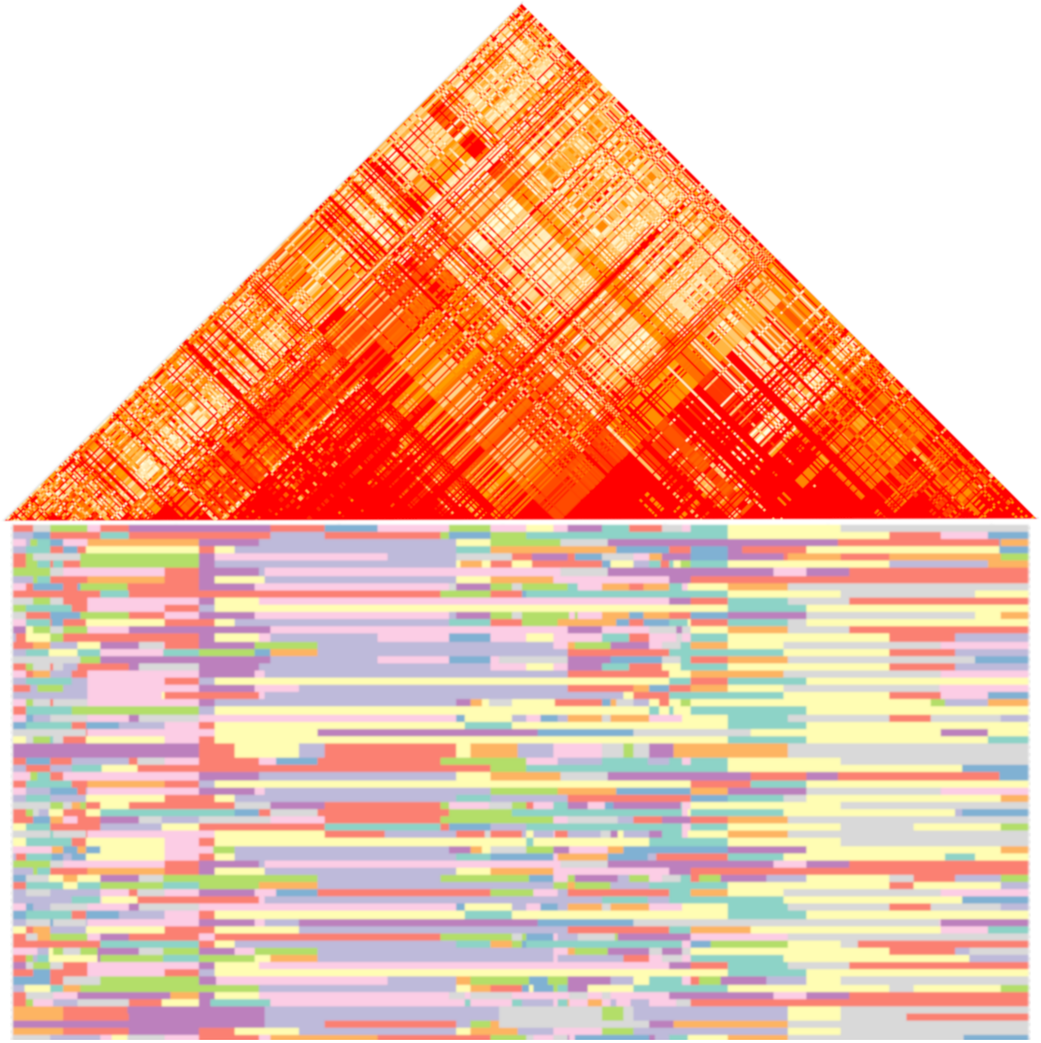






Imputation de génotypes





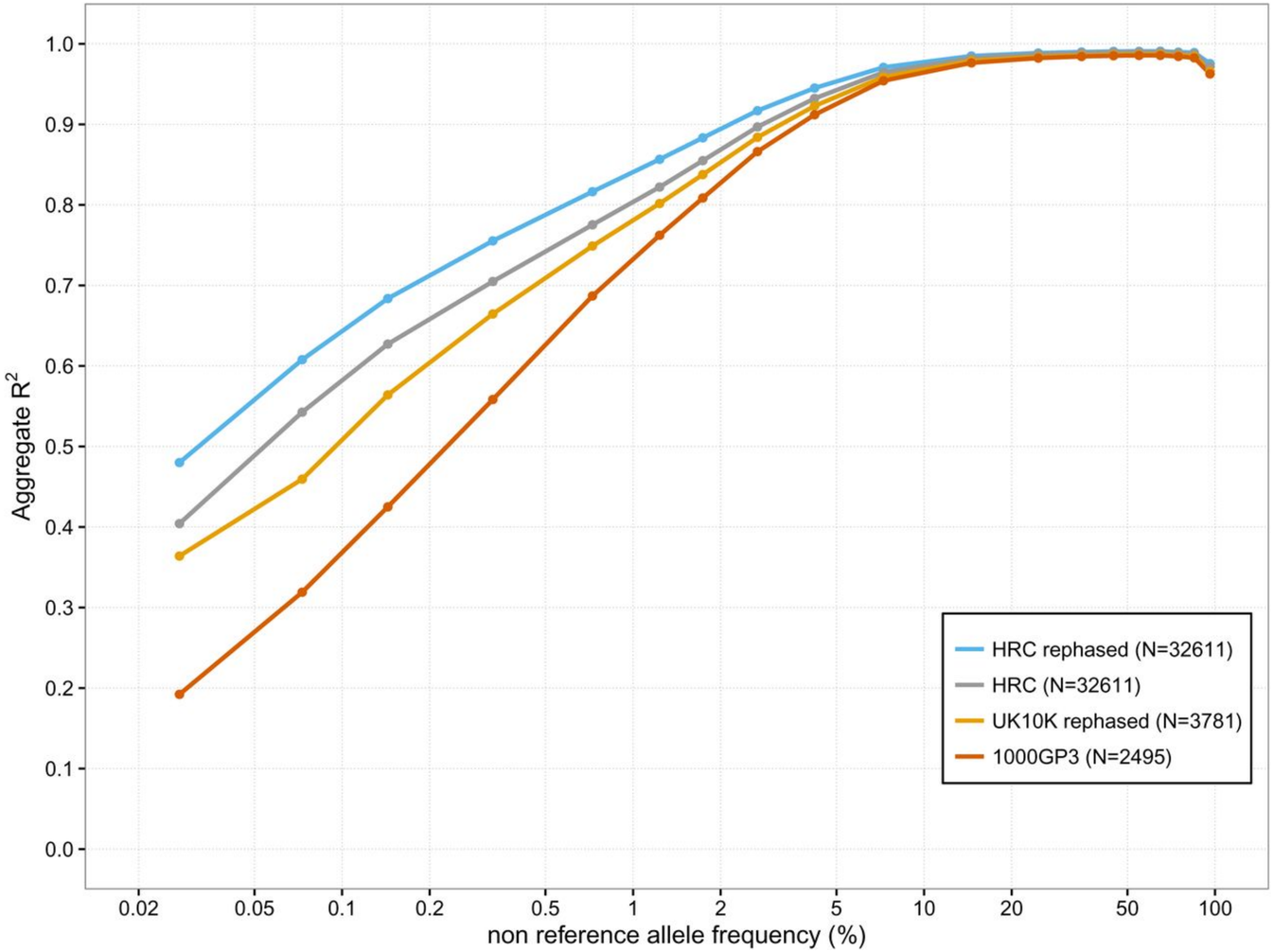
Principe : identifier des patrons d’association alléliques dans une collection d’haplotypes de référence

Puis reconstruire les haplotypes manquants sur la base d’une information partielle

Imputation de génotypes

Performance

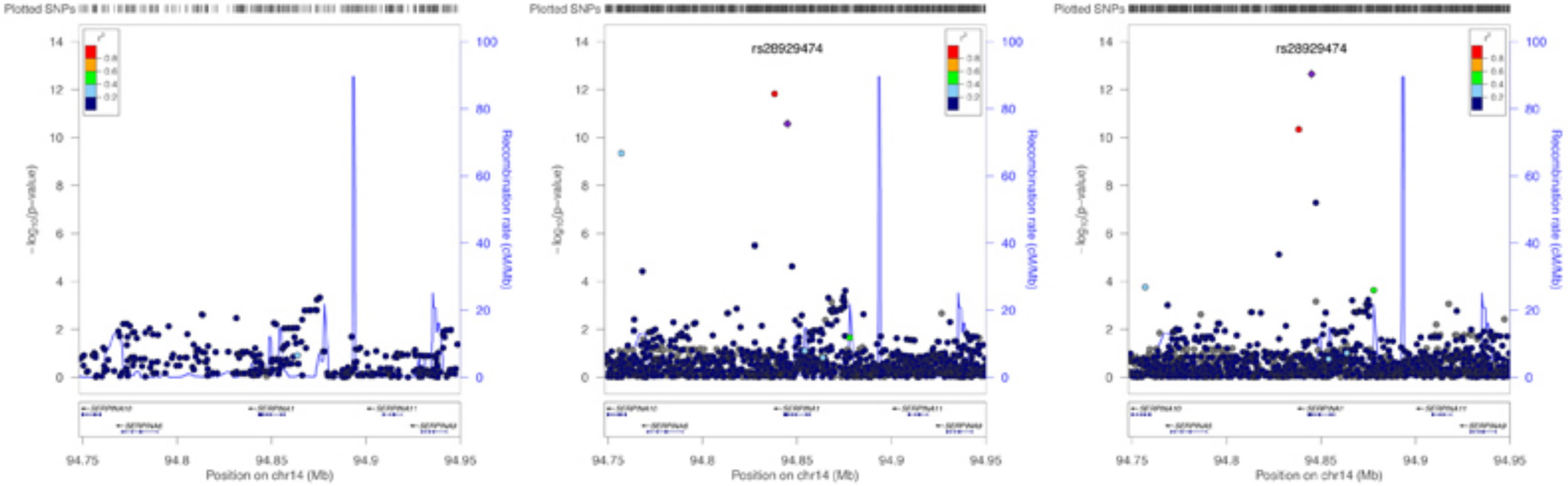
- Panel HRC : 32K haplotypes x 35M de SNPs
- Imputation depuis une puce 1M de SNPs quasi parfaite pour $f > 5\,\%$
- Balayer le génome pour « toutes » les mutations possibles



Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48,** 1279–1283 (2016)

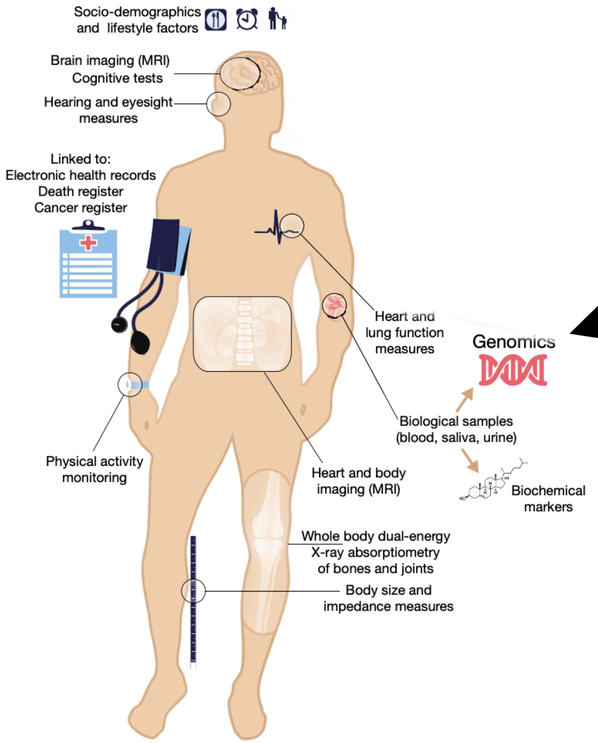
Imputation de génotypes

Exemple



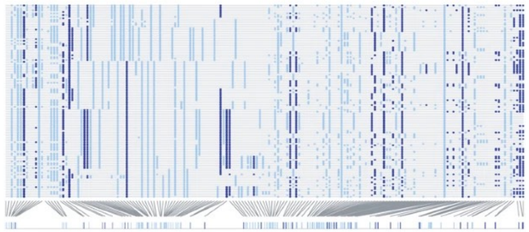
Association signal α_1-antitripsin phenotype at the *SERPINA1* locus.

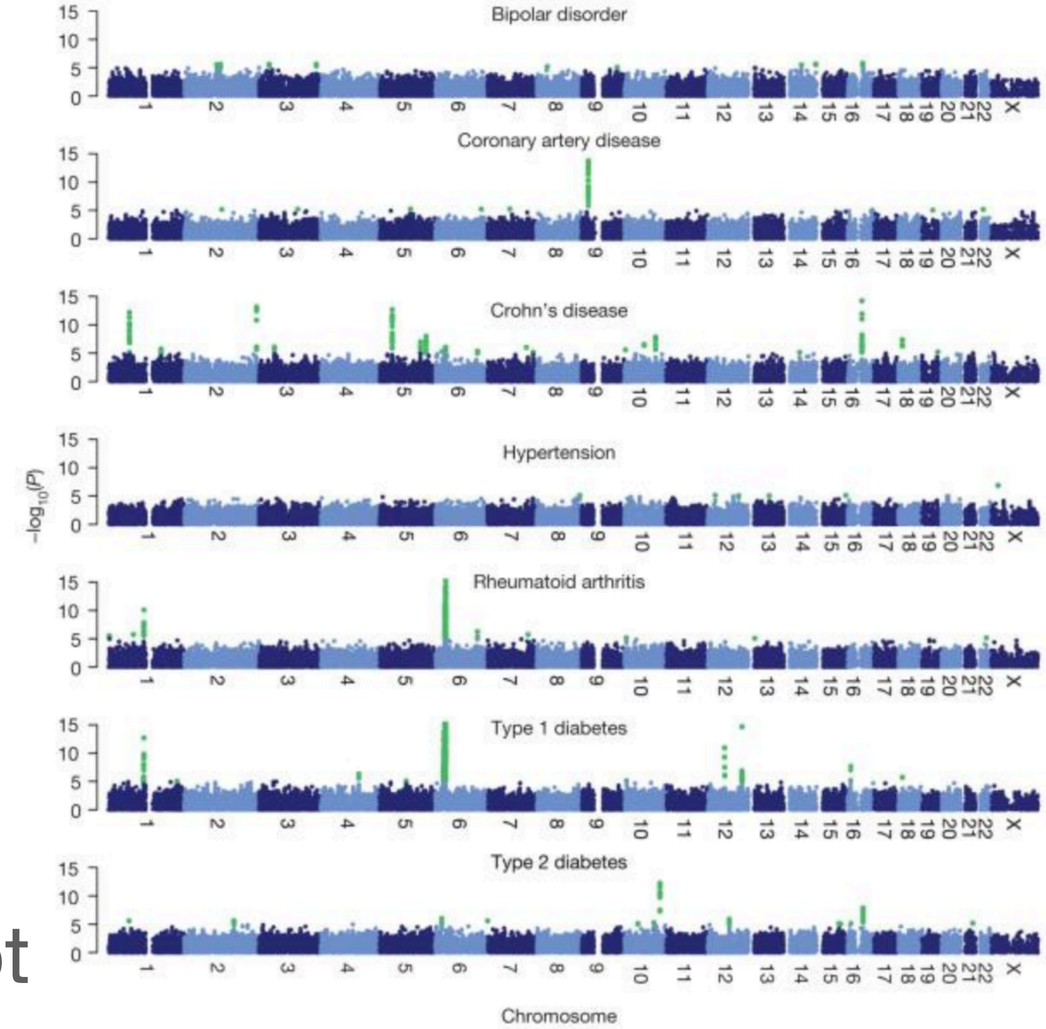
Tiré de McCarthy et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics. 2016



D'après Bycroft *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562,** 203–209 (2018)

$$Y = G + E$$

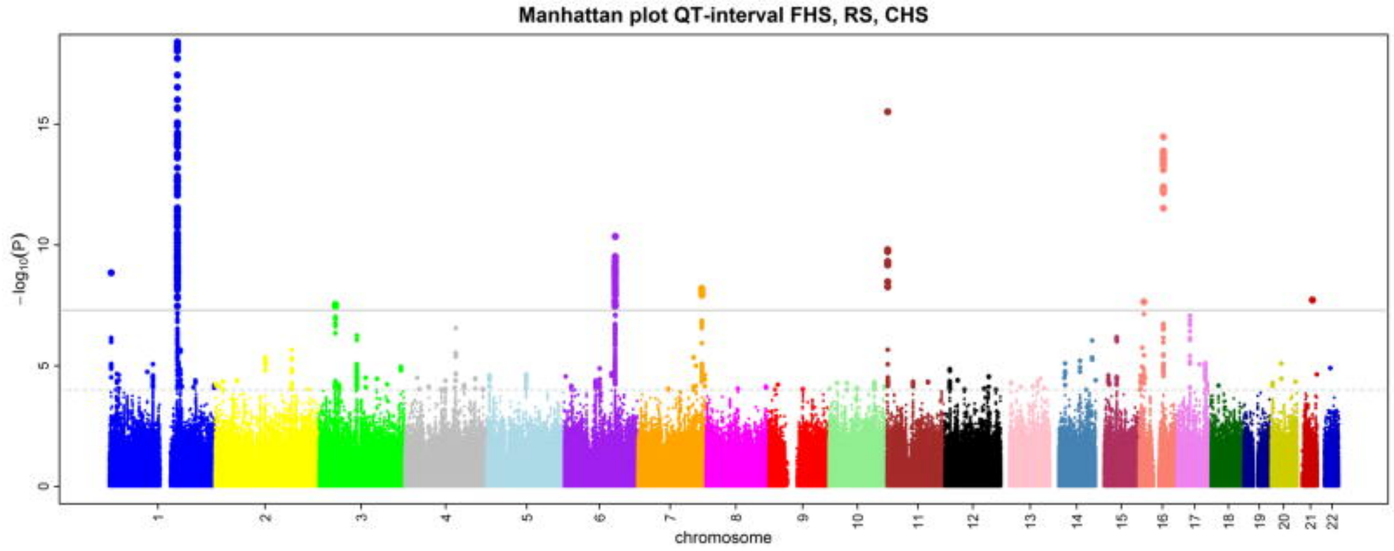




Manhattan Plot

Augmenter la puissance par la méta-analyse

- Mise en commun de résultats de plusieurs GWAS = méta-analyse
- Approche statistique basée sur la combinaison de statistiques d'association
- SNP imputés = communs quelque soit la puce originellement utilisée



Tiré de Newton-Cheh, C.,et al. (2009). Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nature genetics*, *41*(4), 399–406.

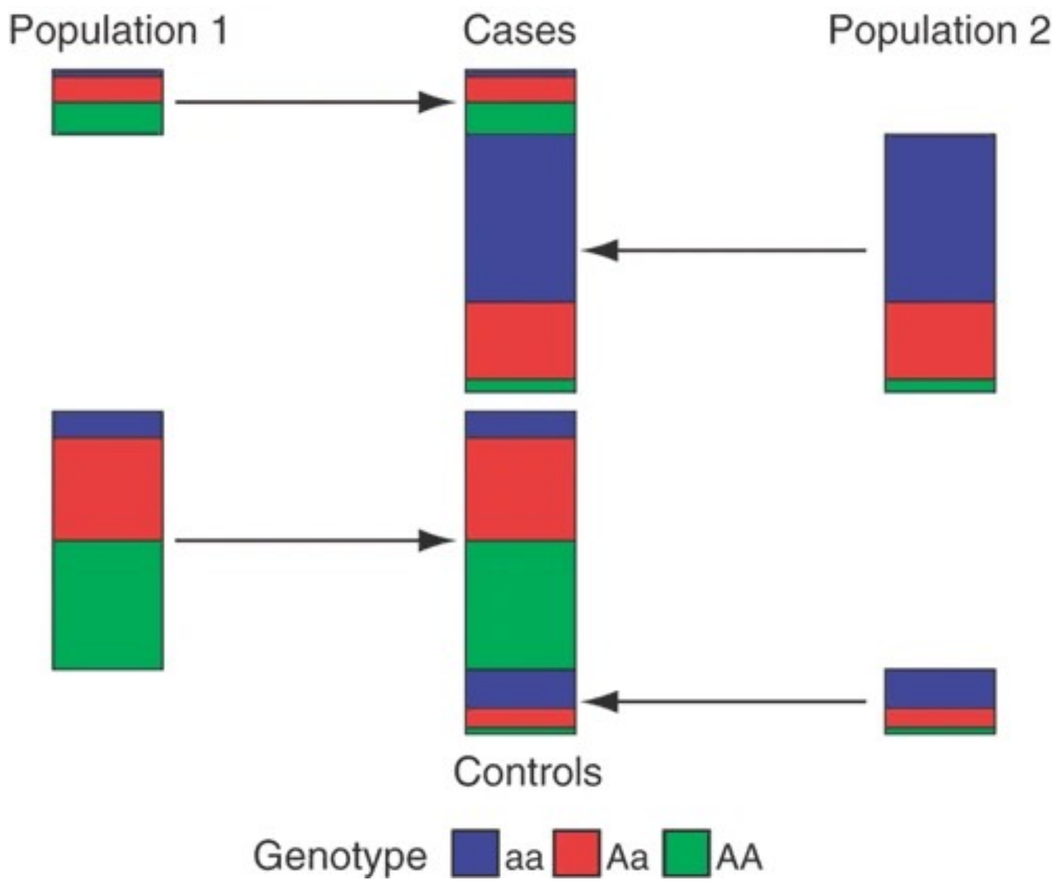
$$z_{meta} = \sum_i z_i w_i$$

$$w_i = \sqrt{\frac{N_i}{N_\bullet}}$$

Structure de population et apparentement

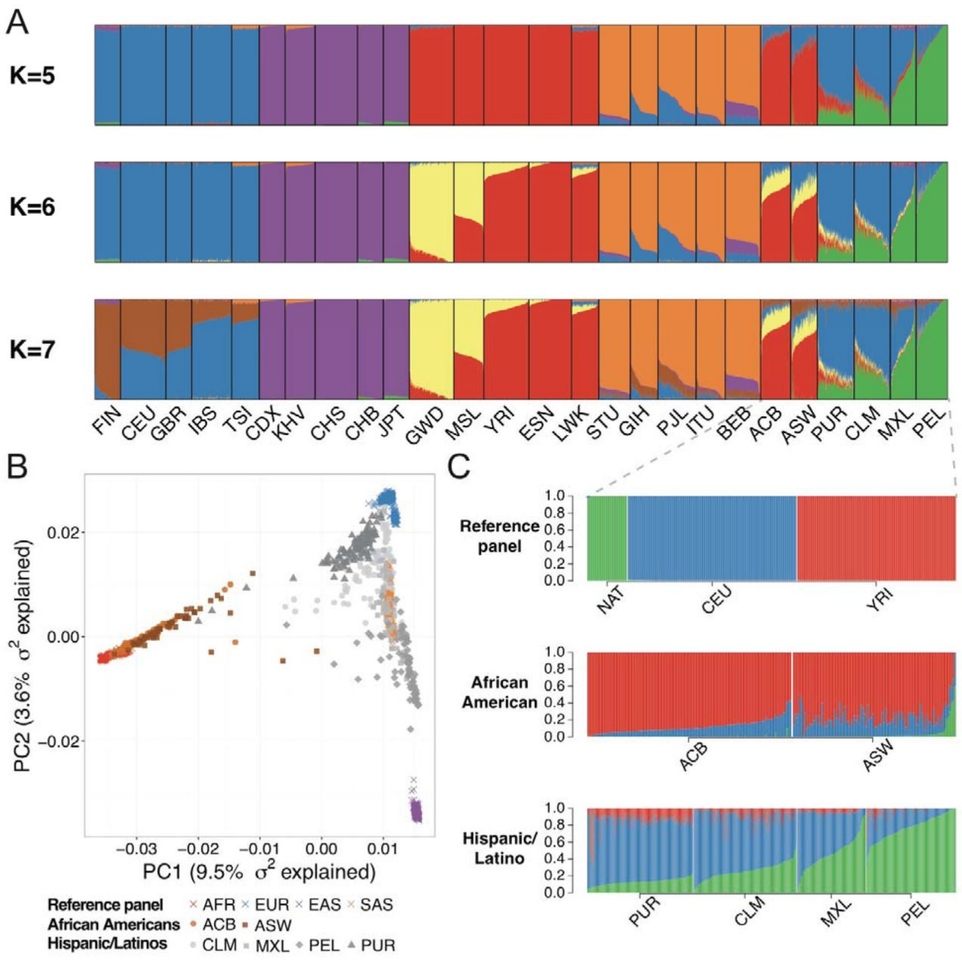
GWAS et structure de population

- Si l'échantillon cas/contrôles est composé de d'individus issus de différentes populations
- Si la prévalence de la maladie est différentes entre les deux populations
- Les marqueurs différenciés entre populations présenteront un faux signal d'association
- D'autant plus important que les tailles d'échantillons sont grandes

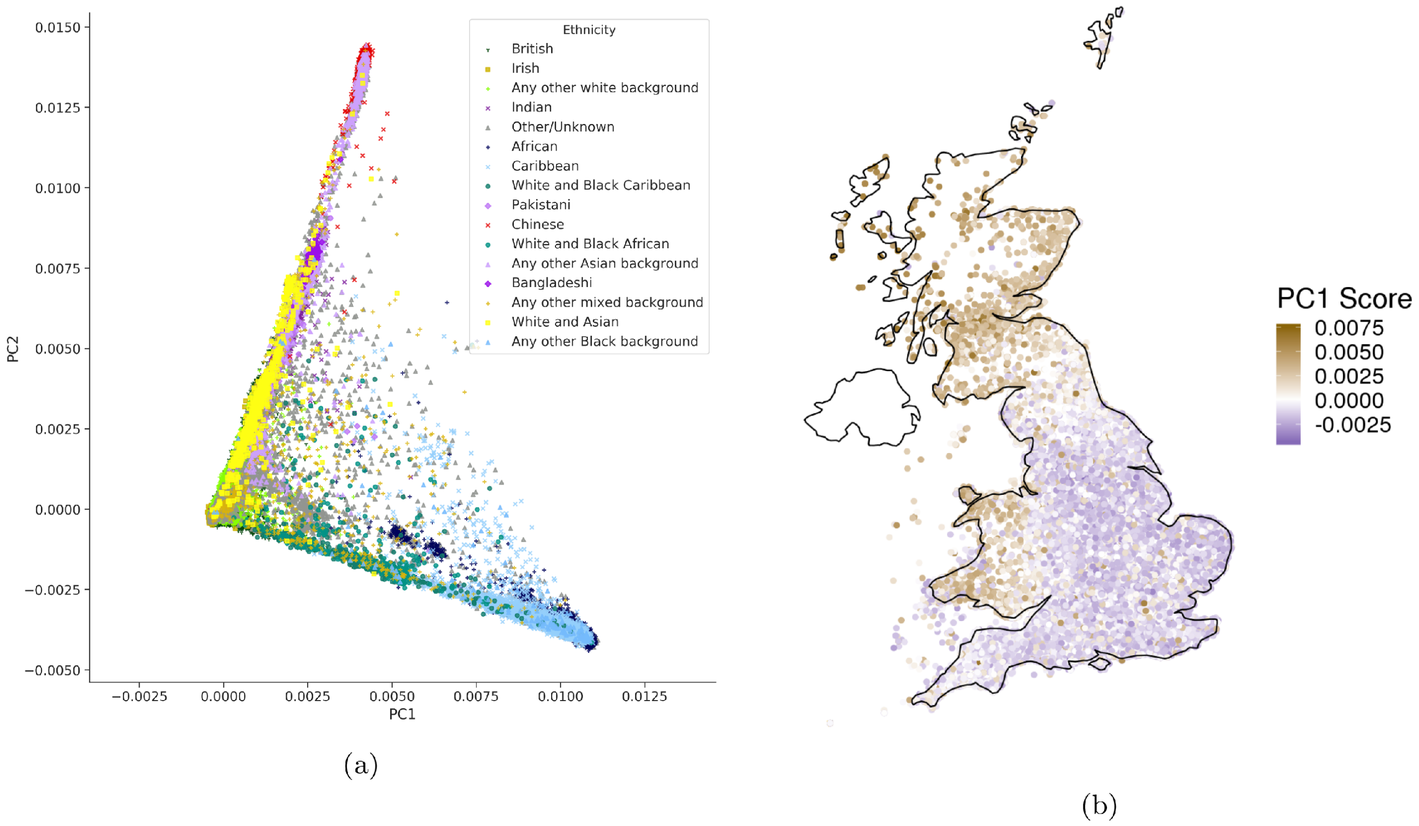


Tiré de Marchini, J., *et al.* The effects of human population structure on large genetic association studies. *Nat Genet* **36,** 512–517 (2004)

Structure génétique des populations humaines



Tiré de Martin, A. R., et al. "Human demographic history impacts genetic risk prediction across diverse populations." *The American Journal of Human Genetics* 100.4 (2017): 635-649.



(a) (b)

Tiré de Agrawal A, Chiu AM, et al. (2020) Scalable probabilistic PCA for large-scale genetic variation data. PLOS Genetics 16(5): e1008773

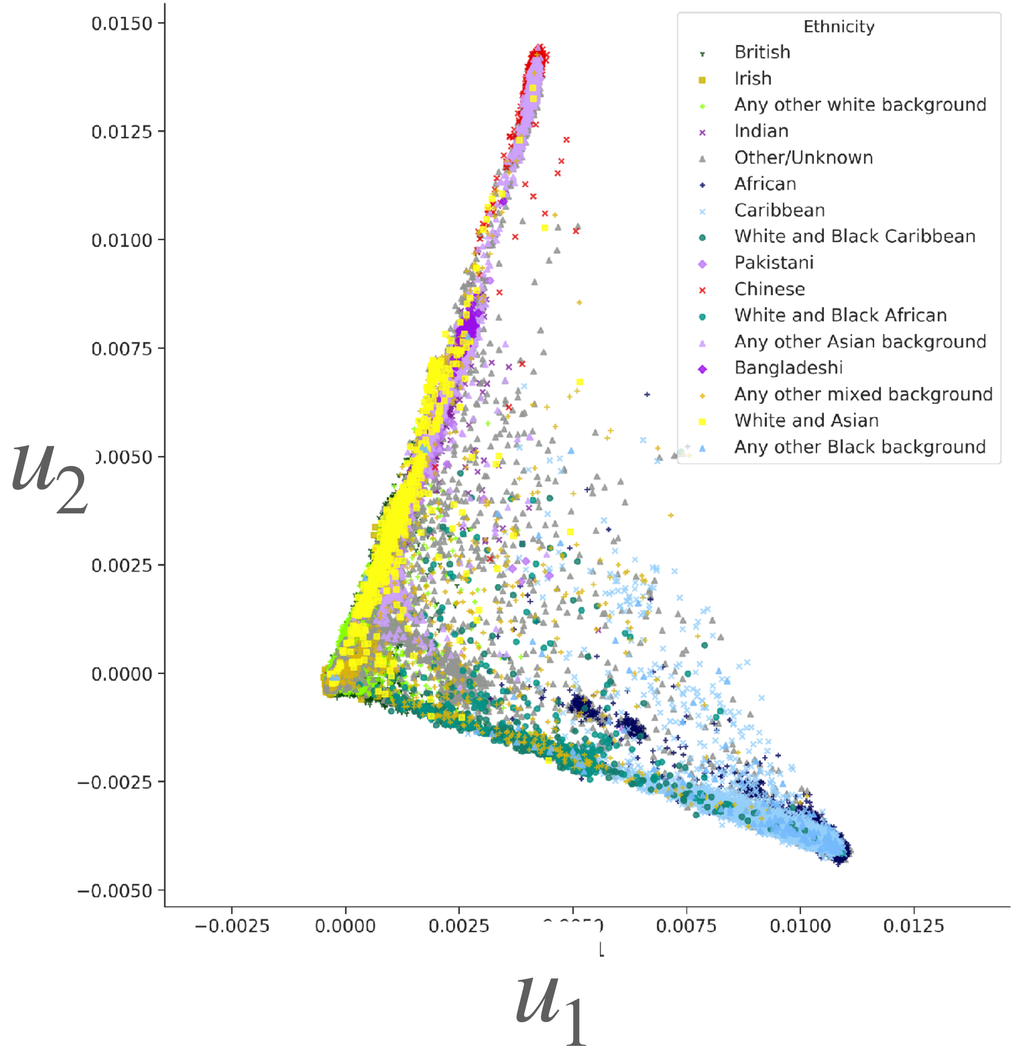
Correction pour la structure

EIGENSTRAT Price, A. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38, 904–909 (2006)

$$\mathbf{y} = \mu + \mathbf{g}\beta + \mathbf{e}$$

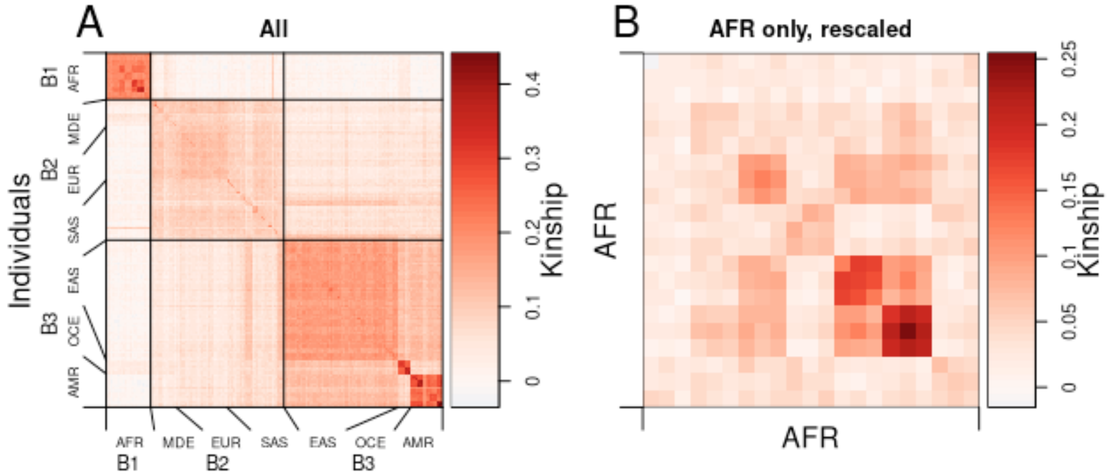
$$\downarrow$$

$$\mathbf{y} = \mu + \mathbf{u}\gamma + \mathbf{g}\beta + \mathbf{e}$$



Correction pour l'apparentement

- Structure de population = forte différentiation génétique
- Apparentement = structure plus subtile due à des relations de parentés distantes (inconnues)
- De plus en plus fréquent avec l'augmentation des tailles d'échantillons

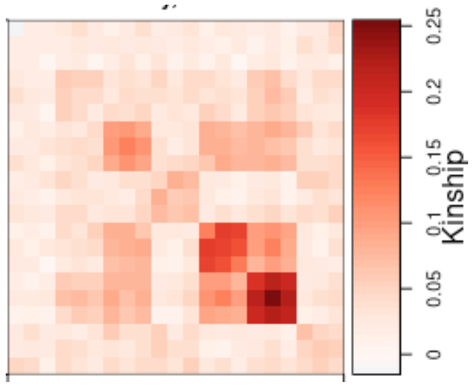


Tiré du package R popkin, Ochoa & Storey, PLoS Genetics (2021)

Modèles intégrant l'apparentement

Intégrer l'apparentement entre individus = **modèles linéaires « mixtes »**

$$\mathbf{y} = \mu + \mathbf{g}\beta + \mathbf{a} + \mathbf{e}$$

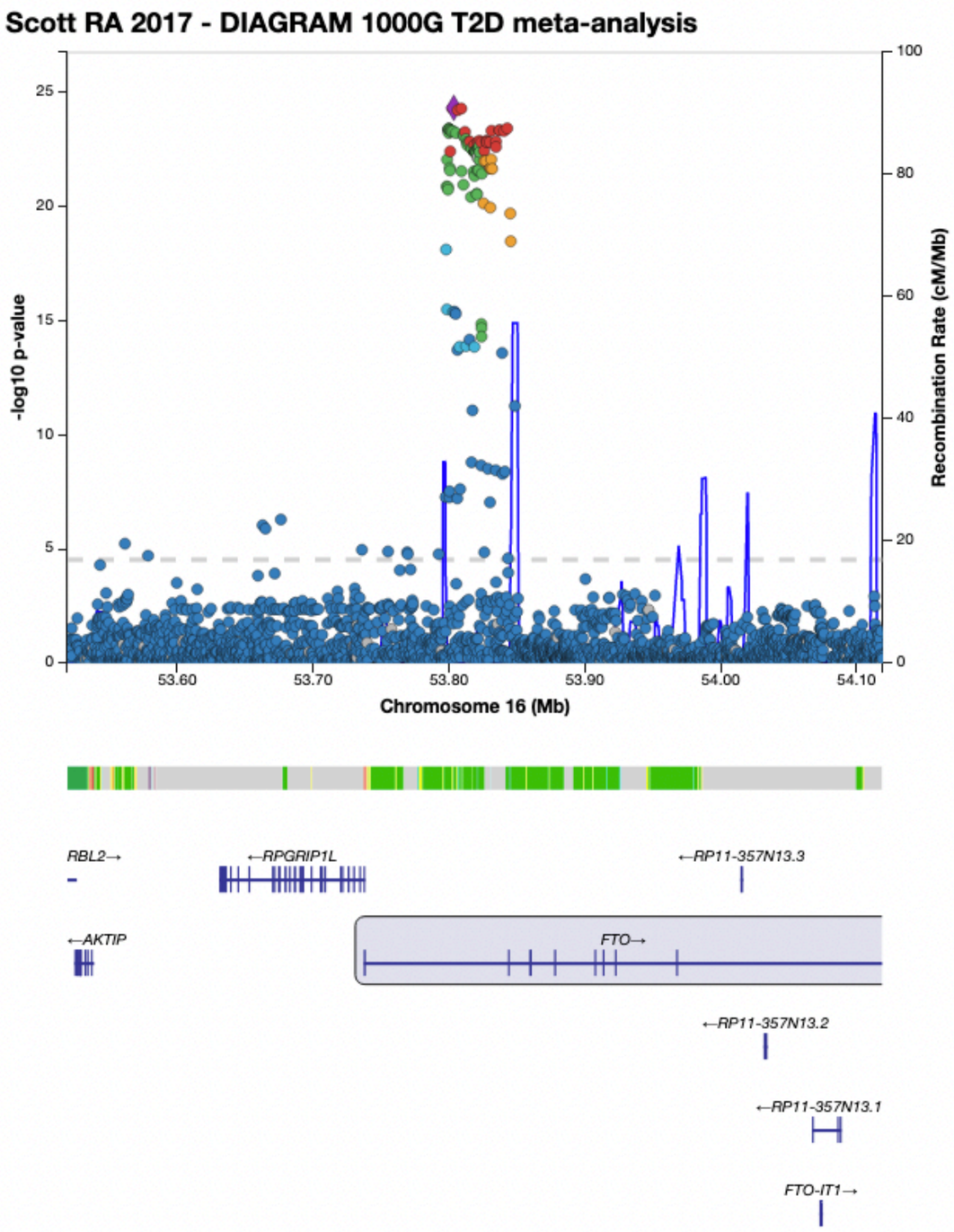


Interprétation biologique des signaux positifs

Annotation de signal

LocusZoom https://locuszoom.org

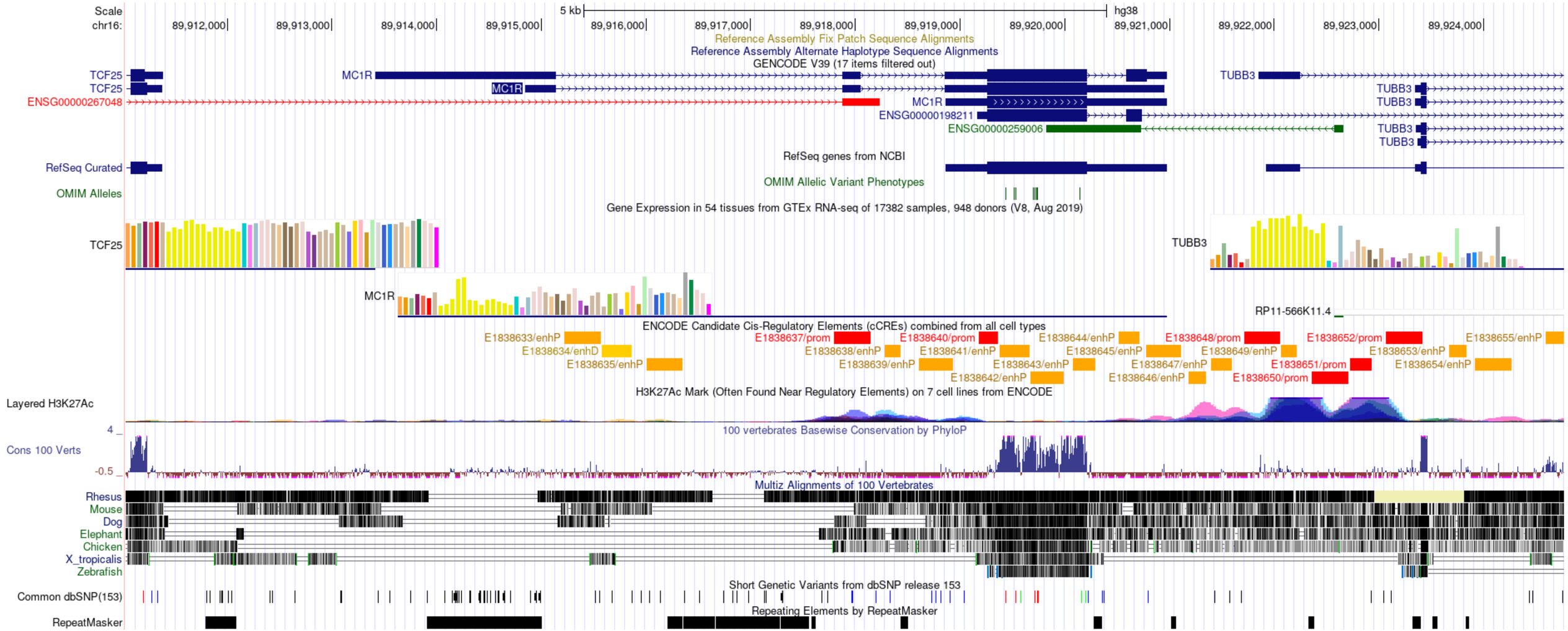
- Outil de représentation locale des résultats
- Intensité du signal : -log10(p-value)
- Couleur = fonction du r^2 avec le variant le plus significatif
- Taux de recombinaison local
- Annotations génomiques variées
- Ici: un polymorphisme du gène FTO est fortement associé au risque de diabète de Type II



Caractéristiques fonctionnelles des variants positifs

- une fois un ensemble de SNP identifiés : annotation fonctionnelle ?
- Variants impliqués dans une modification de la séquence protéique ?
- Pour un codon conservé ? GERP score : conservation des séquences entre espèces
- Régions régulatrices :
  - ENCODE : expériences d’annotation fonctionnelle
  - ChromHMM : Utilise les annotations fonctionnelles pour segmenter le génome en grands types d'éléments

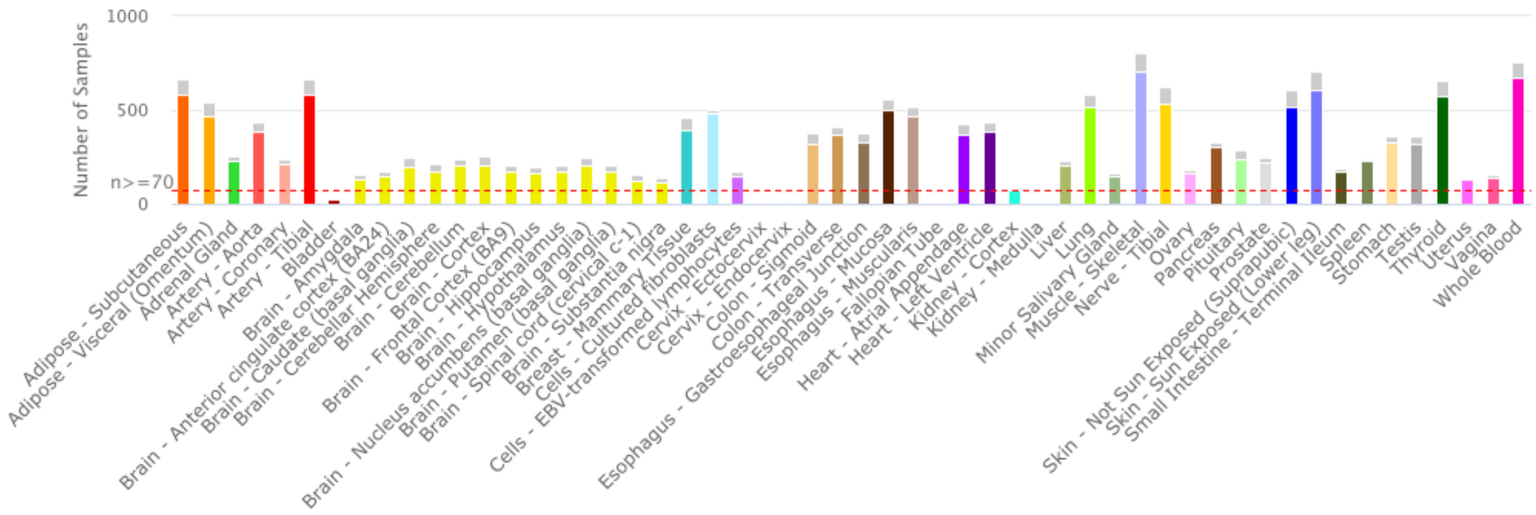
Exemple de région génomique annotée

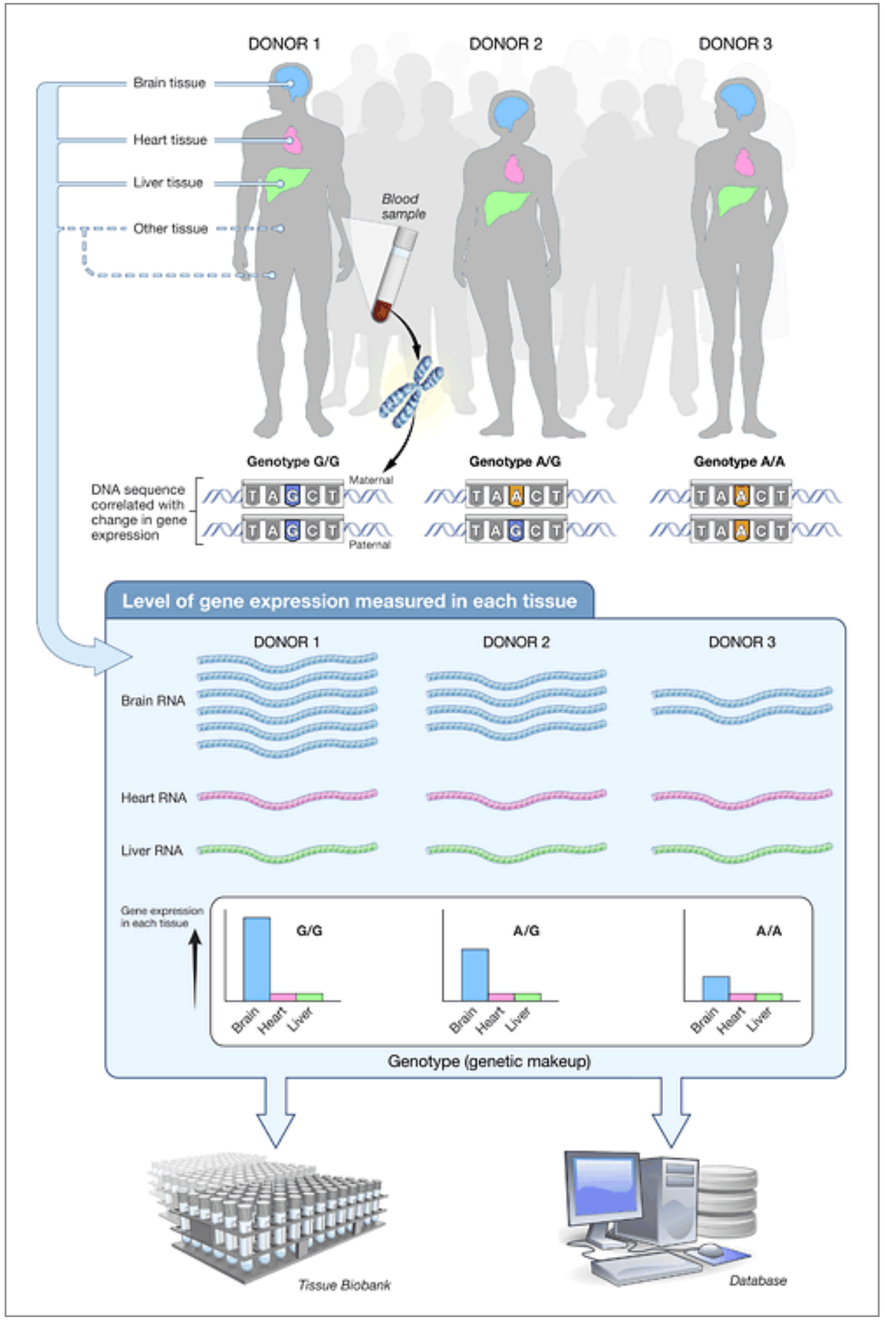


UCSC Genome Browser: https://genome.ucsc.edu

GTeX

- De nombreux hits GWAS sont non-codants
- Difficile de déduire la fonction du contexte génomique (quel gène ?, quel effet sur le gène ?)
- Projet de quantification exhaustive des effets des SNPs sur l'expression des gènes : GTeX
- 838 donneurs dans 54 tissus
- https://gtexportal.org

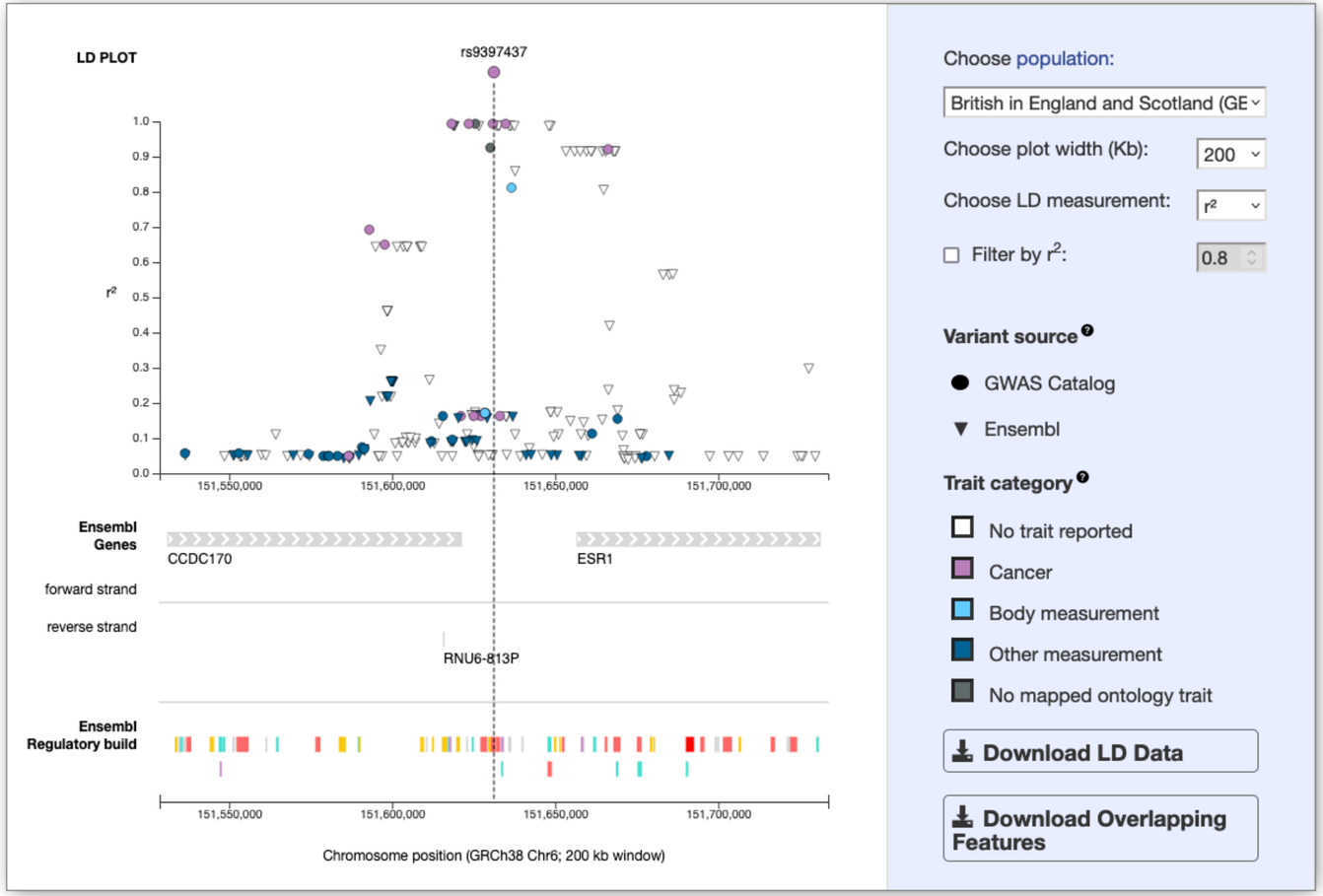




Catalogue des GWAS

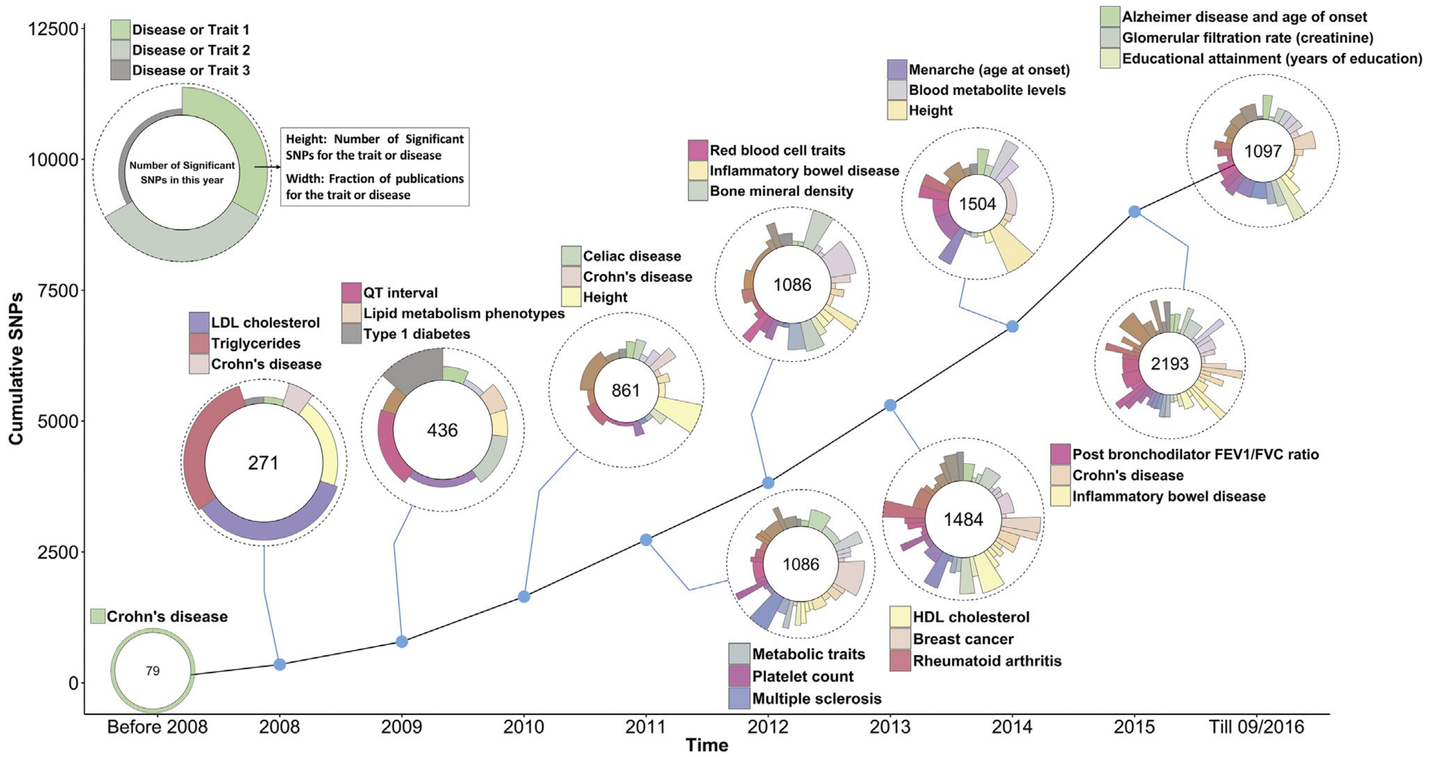
Exemple

- Catalogue exhaustif de l'ensemble des résultats GWAS : https://www.ebi.ac.uk/gwas
- https://www.ebi.ac.uk/gwas/efotraits/EFO_1000650
- https://www.ebi.ac.uk/gwas/variants/rs9397437
- https://www.ebi.ac.uk/gwas/variants/rs67397200
- https://gtexportal.org/home/snp/rs67397200



Evolution des résultats de GWAS

- Augmentation des tailles d'échantillons
- Augmentation du nombre de caractères
- Diversité génétique croissante



Tiré de Visscher et al. (2017) *American Journal of Human Genetics (101):5-22*

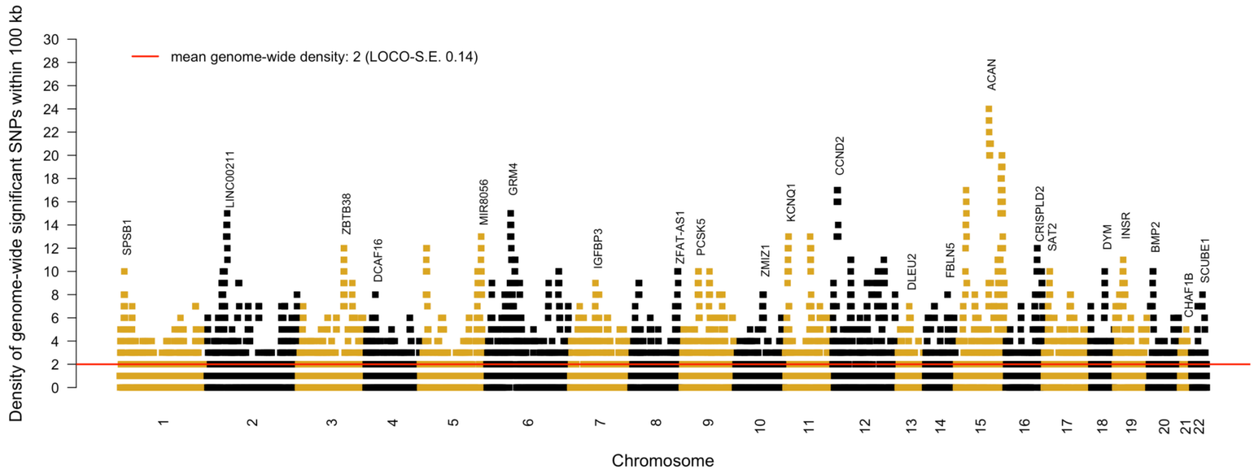
Composante polygénique

- GWAS : identification des SNPs à effets « significatifs"
- Puissance dépend de la taille d’échantillon
- Avec l'augmentation de la taille d’échantillon, augmentation régulière des SNPs significatifs
- = beaucoup de SNP d’effets très faibles : caractères hautement **polygéniques**
- **Très forte pléiotropie**

A Saturated Map of Common Genetic Variants Associated with Human Height from 5.4 Million Individuals of Diverse Ancestries

ABSTRACT

Common SNPs are predicted to collectively explain 40-50% of phenotypic variation in human height, but identifying the specific variants and associated regions requires huge sample sizes. Here we show, using GWAS data from 5.4 million individuals of diverse ancestries, that 12,111 independent SNPs that are significantly associated with height account for nearly all of the common SNP-based heritability. These SNPs are clustered within 7,209 non-overlapping genomic segments with a median size of ~90 kb, covering ~21% of the genome. The density of independent associations varies across the genome and the regions of elevated density are enriched for biologically relevant genes. In out-of-



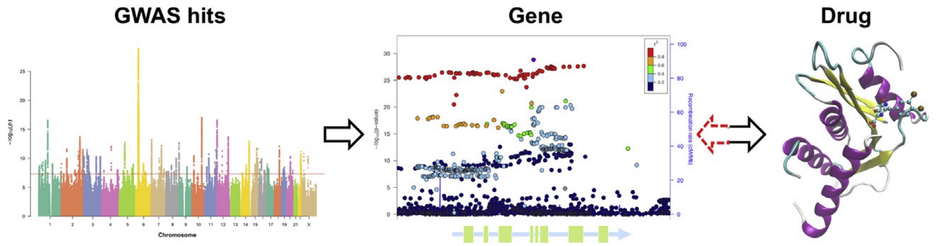
Retombées des résultats de GWAS

Biologie des caractères

- Identification de gènes clefs pour le déterminisme des maladies : cibles pour la validation fonctionnelle
- Mise en évidence de corrélation entre caractères, de gènes partagés (pléiotropie) :
- Identification de voies métaboliques spécifiques
- Concourt à une meilleure compréhension de l’établissement de la maladie

Pharmacologie

- Variants responsables de l'efficacité de traitement pharmaceutiques:
  - VIH et abacavir
  - Ajustement de la dose d'anti-coagulant (Warfarin)
- Optimiser le screening de nouvelles molécules sur la base de leur effet sur des gènes associés à la maladie
- Repositionner des molécules vers de nouvelles maladies

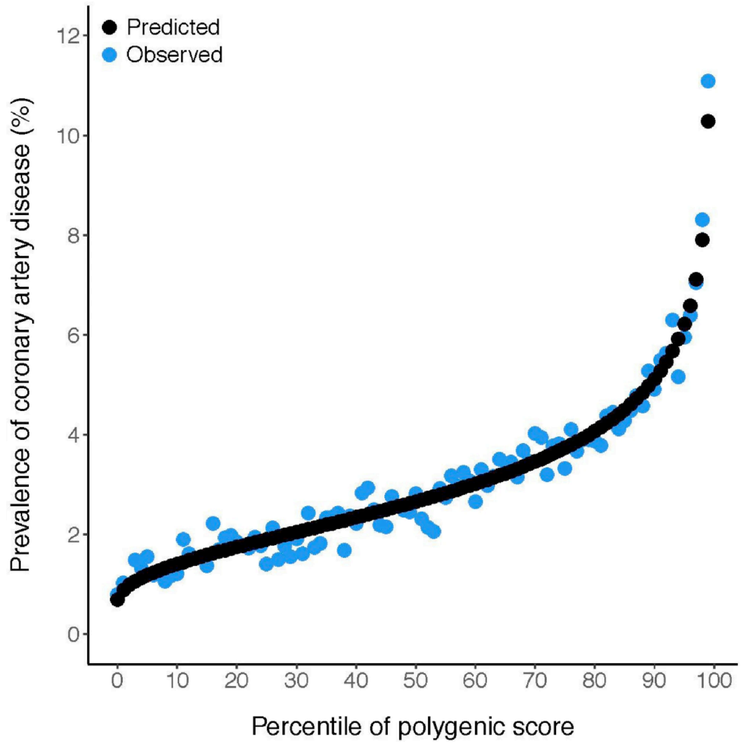


Trait	Gene with GWAS hits	Known or candidate drug
Type 2 Diabetes	*SLC30A8/KCNJ11*	ZnT-8 antagonists/Glyburide
Rheumatoid Arthritis	*PADI4/IL6R*	BB-Cl-amidine/Tocilizumab
Ankylosing Spondylitis(AS)	*TNFR1/PTGER4/TYK2*	TNF-inhibitors/NSAIDs/fostamatinib
Psoriasis(Ps)	*IL23A*	Risankizumab
Osteoporosis	*RANKL/ESR1*	Denosumab/Raloxifene and HRT
Schizophrenia	*DRD2*	Anti-psychotics
LDL cholesterol	*HMGCR*	Pravastatin
AS, Ps, Psoriatic Arthritis	*IL12B*	Ustekinumab

Epidémiologie

Scores de risque polygénique

- Sur la base des résultats de GWAS + un génotypage individuel
- "prédiction" d’un phénotype à partir du génotype d’un individu : « score »
- Participer au suivi personnalisé des patients et recommandations de traitement
- Démontré comme efficace dans certains cas

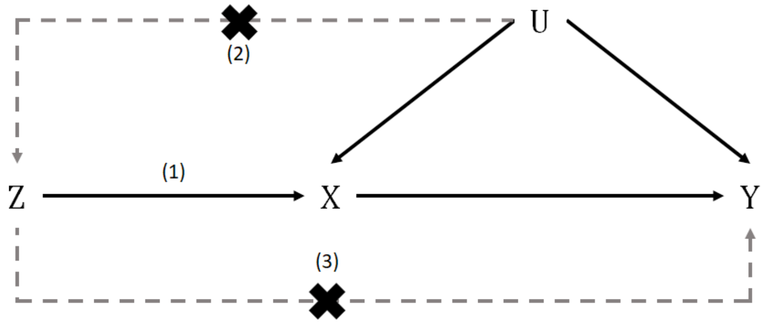


Tiré de Khera et al. *Nat Genet* **50,** 1219–1224 (2018)

Epidémiologie

Randomisation Mendélienne (Mendelian Randomization)

- Effet d'un cofacteur (traitement, comportement …) sur le risque de maladie ?
- Etude observationelle : corrélation, causalité ?
- Essais contrôlés randomisés (Randomized Control Trial), oui mais pas toujours possible (ex. : tabagisme et cancer …)
- Utiliser des SNPs comme *variable instrumentale* pour inférer des relations de causalité

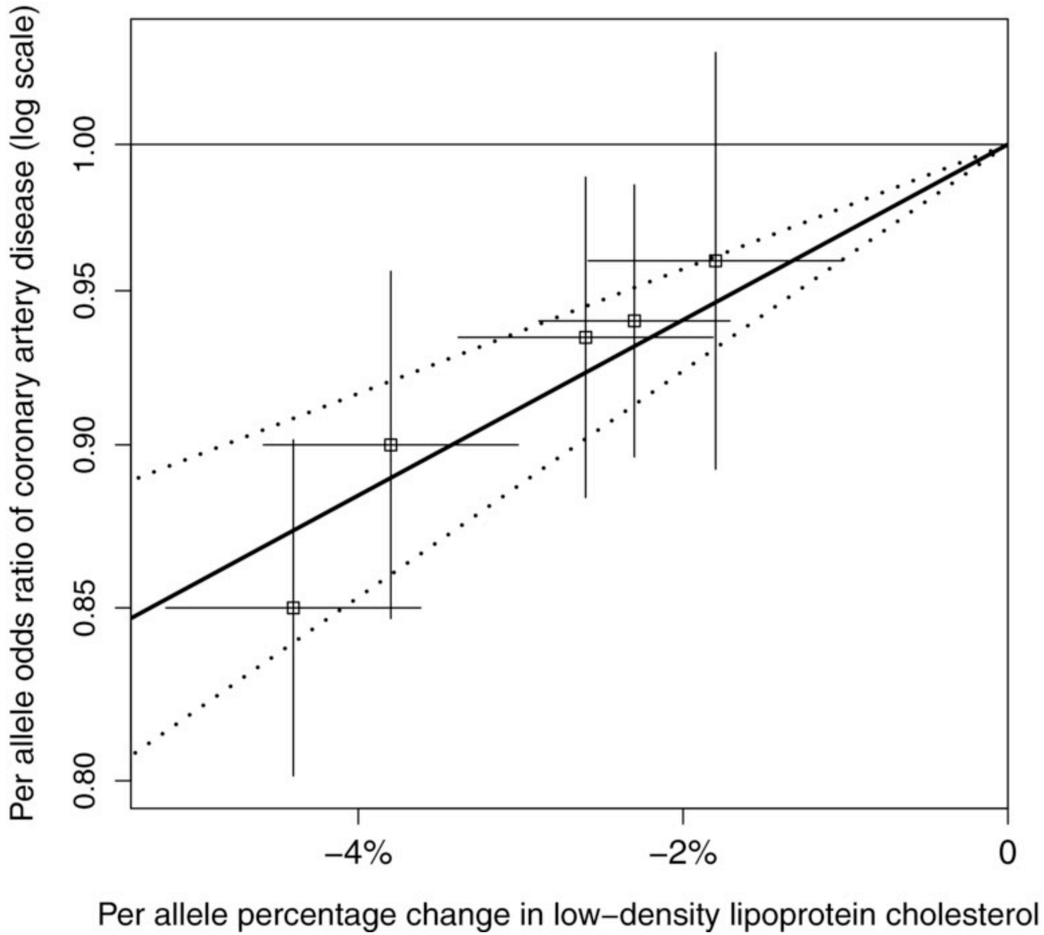


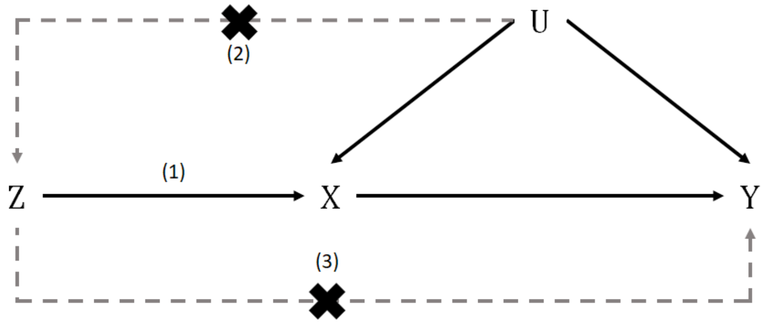
- Z = SNP, associé à X (le cofacteur), Y est la maladie d'intérêt
- hypothèses:

1. association Z -> X (GWAS)
2. Pas d'effet indirect du SNP à travers une stratification des facteurs confondants
3. Pas d'effet direct du SNP sur Y (passe par X)

Epidémiologie

Randomisation Mendélienne (Mendelian Randomization)





- Z = SNP, associé à X (le cofacteur), Y est la maladie d'intérêt
- hypothèses:

  1. association Z -> X (GWAS)
  2. Pas d’effet indirect du SNP à travers une stratification des facteurs confondants
  3. Pas d’effet direct du SNP sur Y (passe par X)

GWAS : conclusions

- Depuis les années 2010 les GWAS sont devenues l’approche standard pour les études de génétique humaine
- Permis par le développement
  - d’outils moléculaires de génotypage haute densité (puces, séquençage aujourd’hui)
  - de méthodes d’analyse statistique pour les tests d’association, l'imputation de génotype, la correction pour la structure génétique …
- Multitudes d’analyse mettent en évidence l’importance de nombreux variants d’effets très faibles = la polygénicité des caractères semble la norme
- Changement de paradigme pour la médecine personnalisée
- Une revue complète récente : **Uffelmann, E., Huang, Q.Q., Munung, N.S. *et al.* Genome-wide association studies. *Nat Rev Methods Primers* 1, 59 (2021). https://doi.org/10.1038/s43586-021-00056-9**