

Chapitre 6. Médecine génomique personnalisée

Introduction à la bioinformatique (SSV3U15, L2 SV AMU)

Jacques van Helden

Aix-Marseille Université

orcid.org/0000-0002-8799-8584

- Maladies génétiques
- Du gène au génome humain aux génomes individuels
- Découverte des gènes d'intérêt pour la santé
- Vers une médecine prédictive
- Utilité clinique
- Enjeux économiques, éthiques et juridiques

Projet 1000 génomes

Projet 2008-2015

- Séquençage génomique + génotypage par biopuces
- En fin de projet, 2500 individus
- Échantillonnage visant à couvrir tous les continents

Nombre total de variations détectées

- 88 millions de SNPs
- 3,6 millions de délétions/insertions courtes
- 60.000 variants structuraux

Variations inter-individuelles moyennes

- ~3 millions de différences entre 2 individus pris au hasard
→ **1 différence / 1000 bp**
- ~4 millions de différences entre un individu et le génome moyen (calculé en retenant pour chaque variation l'allèle majoritaire)

Répartition géographique des variations

- La majorité des variants se retrouvent sur tous les continents (gris foncé), ou dans plusieurs (gris clair)
- Certains variants sont spécifiques d'un continent (couleur claire) ou d'une population (couleur foncée)
- Le nombre de variations par individu est beaucoup plus élevé en Afrique que dans les autres continents. Ceci reflète l'histoire des migrations durant la préhistoire

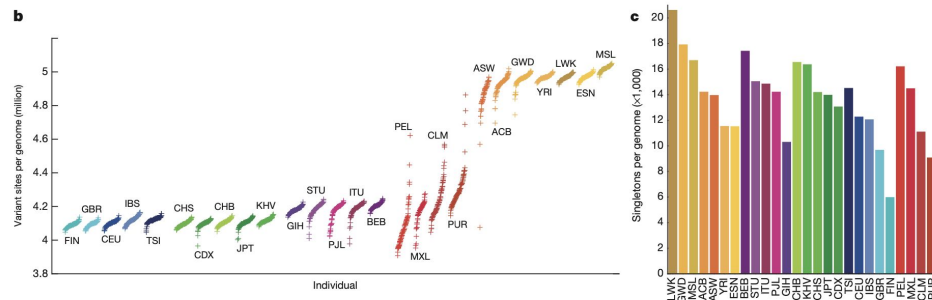
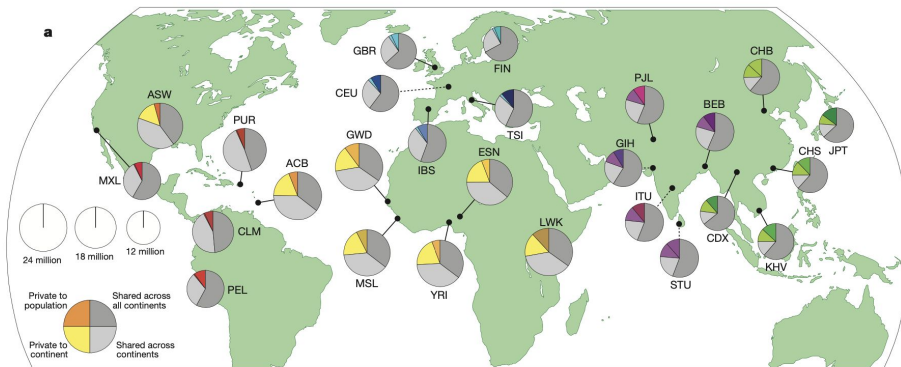


Figure 1 | Population sampling. a, Polymorphic variants within sampled populations. The area of each pie is proportional to the number of polymorphisms within a population. Pies are divided into four slices, representing variants private to a population (darker colour unique to population), private to a continental area (lighter colour shared across continental group), shared across continental areas (light grey), and shared across all continents (dark grey). Dashed lines indicate populations sampled outside of their ancestral continental region. b, The number of variant sites per genome. c, The average number of singletons per genome.

Grands projets de génomique des populations (à des fins de médecine génomique)

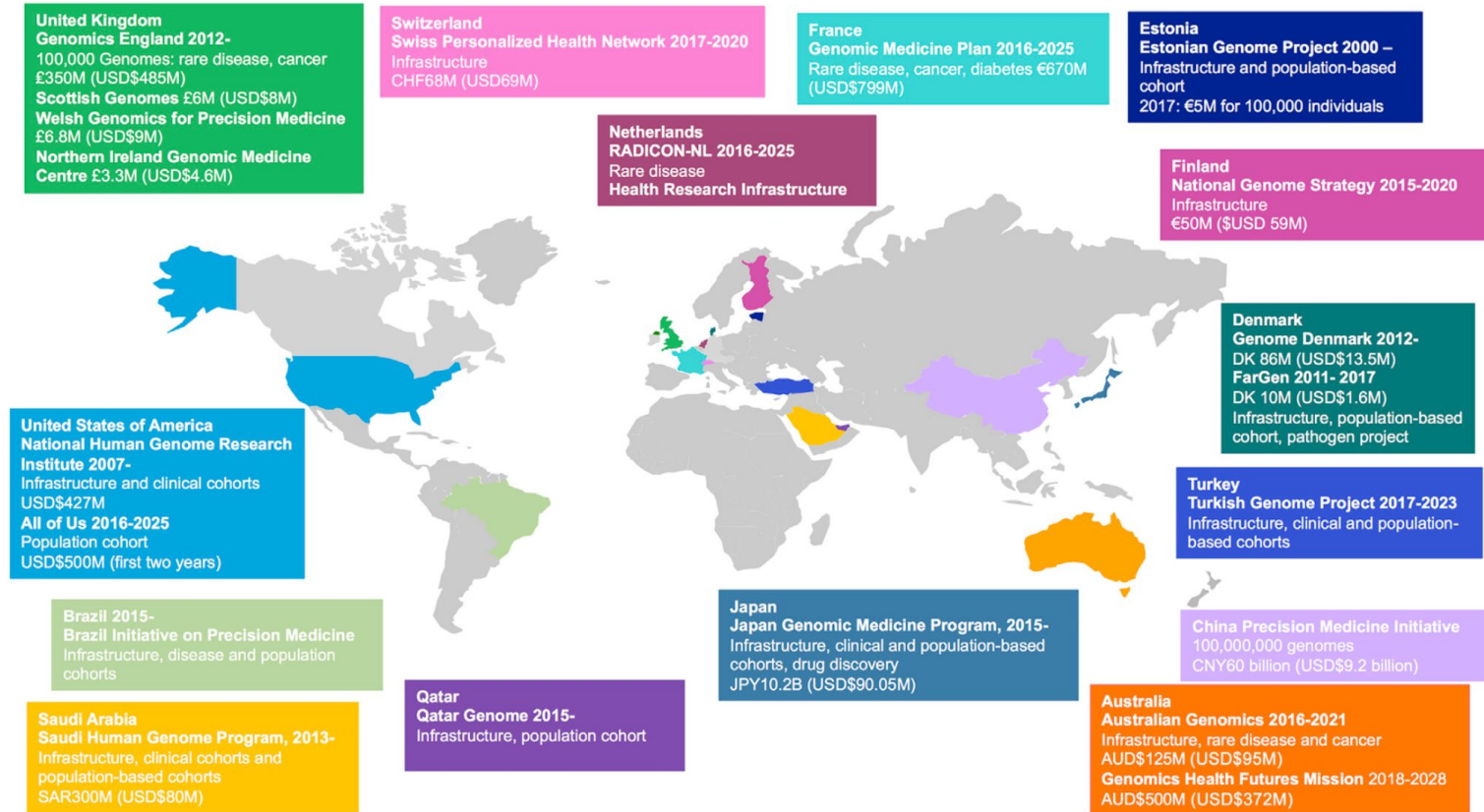


Figure 1. Map of Currently Active Government-Funded National Genomic-Medicine Initiatives

The Genome Aggregation Database (gnomAD)

Short variants

- **Total SNVs: 786,500,648**
- Total InDels: 122,583,462
- Variant type* counts
 - Synonymous: 9,643,254
 - Missense: 16,412,219
 - Nonsense: 726,924
 - Frameshift: 1,186,588
 - Canonical splice site: 542,514

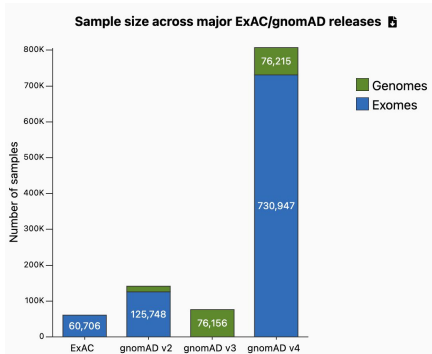
*This is only a subset of commonly asked for variant types from the dataset.

Structural variants

- 1,199,117 genome SVs
 - 627,947 Deletions
 - 258,882 Duplications
 - 711 CNVs
 - 296,184 Insertions
 - 2,185 Inversions
 - 13,116 Complex
 - 92 Canonical reciprocal translocations
- 66,903 rare (<1% site frequency (SF)) exome CNVs
 - 30,877 Deletions
 - 36,026 Duplications

Average number of variants per person

- SNVs per person (*coming soon*)
- 1 rare (<1% SF) coding CNV per individual
- 11,844 SVs per genome



	ExAC	gnomAD v2	gnomAD v3	gnomAD v4*		
	Sample count	Sample count	Sample count	Sample count	%	Increase from v2
Admixed American	5,789	17,720	7,647	30,019	3.72%	1.7x
African/African American	5,203	12,487	20,744	37,545	4.65%	3x
Ashkenazi Jewish	-	5,185	1,736	14,804	1.83%	2.9x
East Asian	4,327	9,977	2,604	22,448	2.78%	2.3x
European^	36,667	77,165	39,345	622,057	77.07%	8.1x
Middle Eastern	-	-	158	3,031	0.38%	19.2x
Remaining^	454	3,614	1,503	31,712	3.93%	8.8x
South Asian	8,256	15,308	2,419	45,546	5.64%	3x
Total	60,706	141,456	76,156	807,162	-	-

*v4 includes all v3 samples.
^ Due to small sample size, Amish are included in remaining individuals, and based on population proximity Finns are included in European totals. Both are presented separately in the v4 browser as before.

Médecine génomique personnalisée

- Prévention des risques médicaux
 - Détection précoce de facteurs individuels de risques
 - Mesures préventives adaptées
- Diagnostic prénatal
 - Applications actuelles: détection de trisomie du chromosome 21
 - Ponction amniotique, mais risques pour l'embryon
 - Méthode non-invasive: séquençage d'un échantillon de sang maternel, qui contient certaines cellules de l'embryon
- Diagnostic préimplantatoire
 - Application : transmission familiale de facteurs génétiques responsables de handicap ou de maladies graves → fécondation in vitro, sélection des embryons non porteurs
- Traitements adaptés au génotype individuel
 - Insensibilité de certains patients à l'effet curatif d'un médicament
 - Effets secondaires des médicaments particulièrement marqué sur certains individus
- Thérapie génique

Maladies monogéniques

Des mutations dans nos gènes peuvent provoquer des maladies

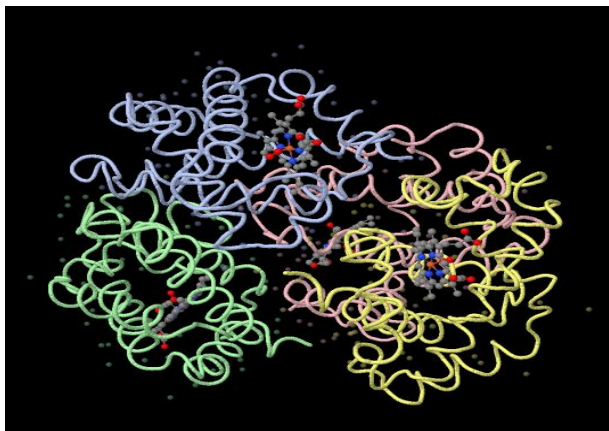
- Des mutations dans les gènes peuvent causer des changements de structure des protéines, qui peuvent provoquer de graves maladies (anémie falciforme, emphysème pulmonaire, chorée de Huntington, mucoviscidose,...).

Hémoglobine

Une mutation ponctuelle provoque

l'anémie falciforme

<http://www.rcsb.org/pdb/explore.do?structureId=1GZX>



Alpha-1-antitrypsine (AAT)

Une mutation provoque la propension

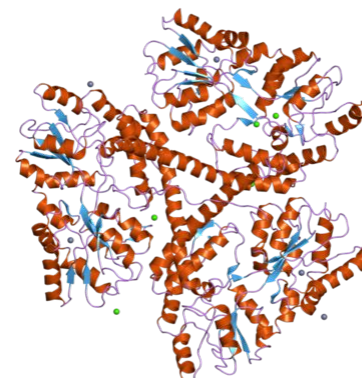
à développer **l'emphysème pulmonaire**

<http://www.rcsb.org/pdb/explore.do?structureId=3CWM>

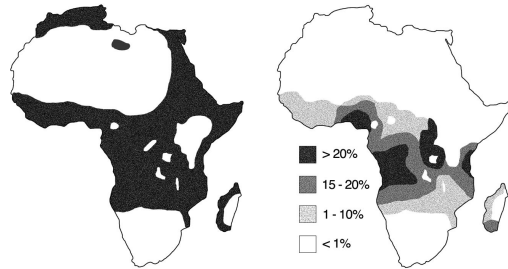
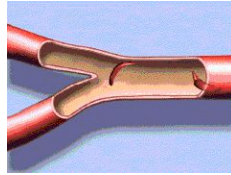
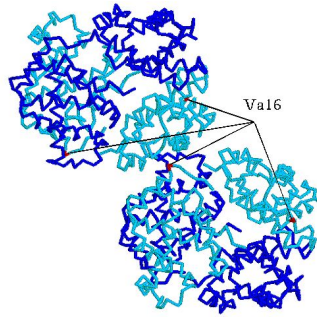
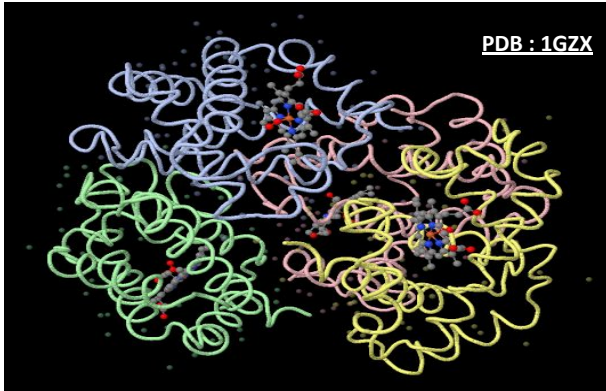


Huntingtine

Région répétée poly-G :6 à 35 glutamines selon les cas. Individus porteurs d'un nombre élevé de répétitions (36 à 250 répétitions) → propension à développer la **Chorée de Huntington**,



Mutations de l'hémoglobine



Les hétérozygote AS ont une sensibilité réduite au paludisme (malaria), ce qui explique la correspondance frappante entre zones endémiques pour le paludisme (gauche) et la drépanocytose (droite).

- L'hémoglobine est une protéine qui transporte l'oxygène dans le sang.
- On connaît >500 formes mutantes, dont >95 contiennent une mutation ponctuelle (substitution d'un seul acide aminé).
- Une substitution particulière d'un acide glutamique en valine au 6ème résidu de l'hémoglobine beta provoque l'**anémie falciforme** (également appelée **drépanocytose**).
- Mécanisme moléculaire : cette mutation modifie complètement la structure de la protéine, en favorisant sa polymérisation, qui entraîne la formation de fibres.
- Ces fibres déforment les globules rouges en leur donnant une forme de faucille (d'où provient le nom de la maladie).
- La mutation a peu d'effet à l'état hétérozygote, mais provoque de gros problèmes sanguins à l'état homozygote. Les globules rouges déformés gênent la circulation sanguine en bloquant les vaisseaux capillaires. Cet exemple montre l'impact potentiel d'une simple mutation ponctuelle sur la structure d'une protéine, et les conséquences fonctionnelles.

Prédisposition au cancer du sein

- BRCA1 pour « Breast cancer 1 », gène associé au cancer du sein n°1, précoce.
- Fonction de la protéine: réparation des dommages à l'ADN, régulation transcriptionnelle.
- Forte association entre certaines mutations et le cancer du sein.

<http://www.rcsb.org/pdb/explore/jmol.do?structureId=1JM7>

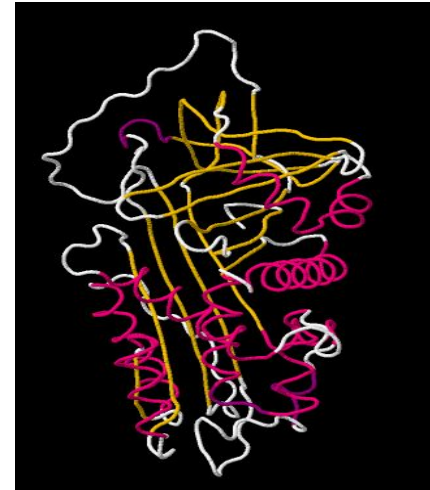


La détection de prédisposition aux maladies

- En caractérisant votre génome, on peut prédire vos prédisposition à développer une maladie.
- Exemple
 - Une mutation dans le gène codant pour l'**alpha-1-antitrypsine** suscite l'**emphysème pulmonaire**
 - La mutation est déjà présente chez l'embryon, mais la maladie peut se déclarer chez l'adulte, ou même ne jamais se déclarer
 - Dans les populations européennes, 1 personne sur 20 est hétérozygote; 1 personne sur 400 est homozygote
- Conséquences positives du test génétique
 - Prévention: éviter à tout prix de fumer, car le tabagisme favorise le déclenchement de la maladie.
 - Suivi: on peut faire des tests réguliers pour détecter les premières phases de la maladie (particulièrement utile dans le cas de cancers).
 - Curation: (perspectives) de thérapie génique dans un futur plus ou moins proche.
- Risques liés au test
 - Protection de la vie privée: ces données doivent-elles être communiquées aux assurances (très demandeuses), qui pourraient vous refuser des assurances-vie ou des garanties hypothécaires ? aux employeurs, qui préfèrent éviter d'embaucher un employé « à risque » ?
 - Gestion psychologique du risque: vivre sa vie dans l'expectative d'une maladie ?
 - Eugénisme: élimination des embryons homozygotes (pour éviter la maladie) ?
Élimination des embryons hétérozygotes (pour éliminer l'allèle mutant de la population humaine) ? Jusqu'où peut-on pousser cela ?

Alpha-1-antitrypsine (AAT)

Une mutation ponctuelle provoque la propension à développer l'emphysème pulmonaire
<http://www.rcsb.org/pdb/explore.do?structureId=3CWM>



Maladies complexes (polygéniques, multifactorielles)

- Maladie complexe (multifactorielle): maladie causée par des variants génétiques multiples (polygénique), des interactions entre ces variants, et des interactions entre variants et facteurs de l'environnement (Hunter et al., 2008).
- Précisions
 - Polygénique: fait référence à l'intervention de facteurs génétiques multiples.
 - Facteurs environnementaux: style de vie (nutrition, consommation de tabac, ...) ou extérieurs (stress, pollution).
 - Complexité: on parle de complexité quand l'effet d'une combinaison est différente de la somme des effets.
- Exemples: maladies cardio-vasculaires, diabète.
- Héritabilité
 - Difficile à établir en cas de maladies complexes:
 - Généralement, chaque facteur individuel joue un rôle modeste dans la susceptibilité à présenter les symptômes cliniques (les tests d'association sont donc faiblement significatifs).
 - L'impact du style de vie et de l'environnement complique la caractérisation des risques transmis génétiquement.
 - Quelques maladies psychiatriques présentent un certain taux d'héritabilité génétique, des études récentes ont permis de détecter des gènes associés à ces maladies. Les mécanismes sous-jacents à la pathologie ne sont néanmoins pas encore élucidés.

Études d'association à l'échelle génomique

Genome-wide association studies (GWAS)

Etudes d'association de type « cas - contrôle »

- **Principe** des études d'association

- **Études cas-contrôles** : on caractérise le profil génétique d'un groupe de patients souffrant d'une maladie donnée (les « cas »), et d'un groupe de personnes saines (les « contrôles »).
- On analyse ensuite chaque marqueur, en comparant la fréquence des différents génotypes dans les deux groupes.
- Ceci permet de délimiter des régions génomiques associées à la maladie.

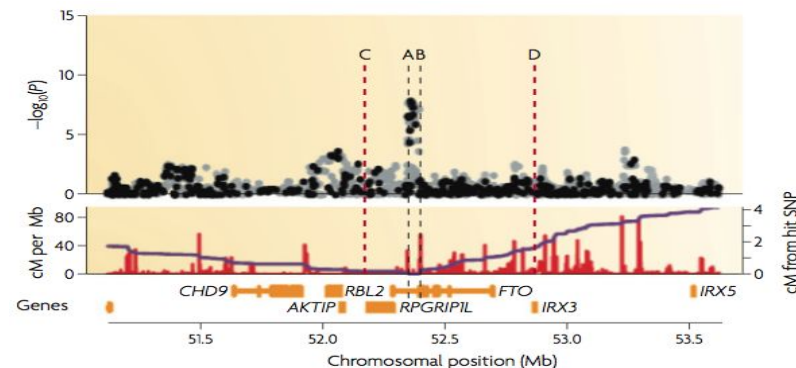
- Attention, **corrélation n'est pas raison**

- Les études d'associations révèlent uniquement la corrélation entre un trait génotypique et une maladie, mais ne constituent pas un élément suffisant pour établir une relation de causalité

- L'association peut provenir d'**effets de stratification**

- Si la maladie est plus fréquente dans une sous-population donnée, et que cette sous-population est caractérisée par un marqueur, on observera une corrélation fortuite.
- On risque donc de confondre les traits génétiques associés aux patients et ceux de la sous-population.
- Il existe des méthodes statistiques pour éviter ce piège de la stratification.

Génotype	aa	aA	AA	Total
Cas (patients)	X1	X2	X3	X
Contrôle	Y1	Y2	Y3	Y
Total	T1	T2	T3	T



McCarthy et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* (2008) vol. 9 (5) pp. 356-69

L'axe horizontal représente un fragment de chromosome. Chaque point noir ou gris représente un marqueur génétique (un site polymorphique), et la hauteur du point reflète la significativité de l'association entre ce site chromosomique et un trait phénotypique particulier. Sur cette figure, on observe une forte association entre le trait phénotypique et les marqueurs génétiques situés dans la région A-B.

Études d'associations à l'échelle du génome complet

- Une étude a été menée sur 17.000 personnes afin d'identifier les régions génomiques associées à 7 maladies (2.000 patients par maladie) par rapport à un groupe de contrôle (3.000 personnes).
- La figure synthétise les résultats, en indiquant (en vert) les SNPs associés de façon significative à l'une des maladies.
- Les zones bleues représentent les chromosomes.
- Chaque point vert représente un SNP, et sa la hauteur indique la significativité.

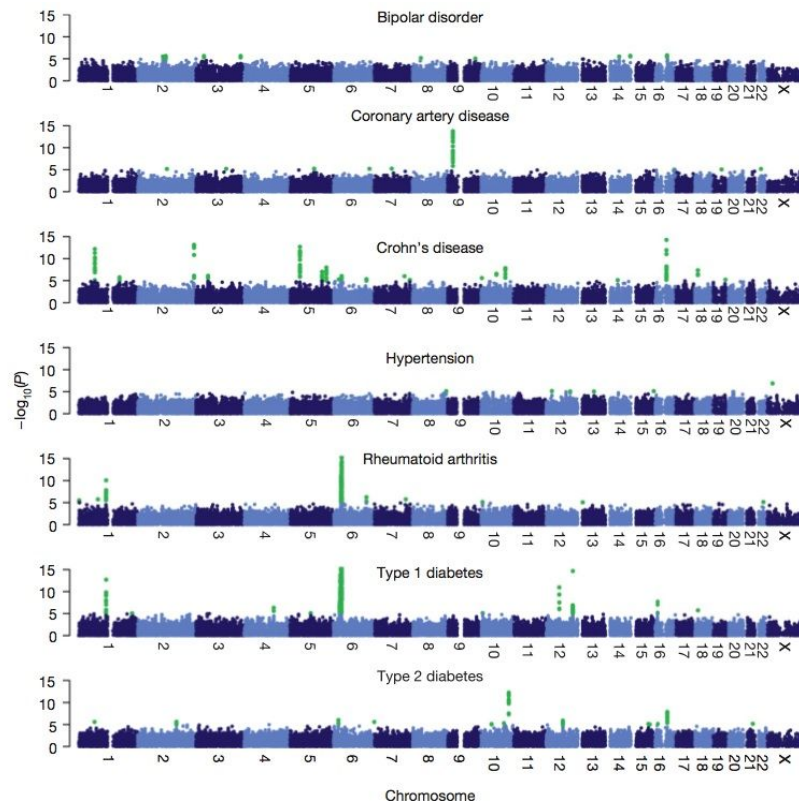
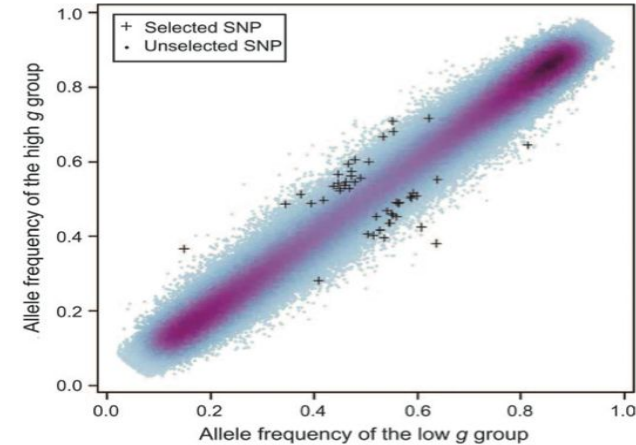


Figure 4 | Genome-wide scan for seven diseases. For each of seven diseases $-\log_{10}$ of the trend test P value for quality-control-positive SNPs, excluding those in each disease that were excluded for having poor clustering after visual inspection, are plotted against position on each chromosome.

Chromosomes are shown in alternating colours for clarity, with P values $< 1 \times 10^{-5}$ highlighted in green. All panels are truncated at $-\log_{10}(P \text{ value}) = 15$, although some markers (for example, in the MHC in T1D and RA) exceed this significance threshold.

Genetics and cognitive ability

- Depuis plusieurs décennies, des chercheurs ont tenté de détecter les gènes associés à des capacités cognitives spécifiques, ou générales (« intelligence »).
- En 2008, l'équipe de Richard Plomin a testé les capacités générales de 7000 enfants en leur faisant passer des tests cognitifs, dont les résultats sont synthétisés au moyen d'un score unique (le facteur « g »).
- Ils ont sélectionné les 1500 enfants ayant le meilleur score et les 1500 ayant le moins bon score, et caractérisé leur génotype au moyen de biopuces à 500.000 SNPs.
- Aucun des SNPs n'a montré d'association significative avec le facteur g.
- Les 47 SNPs les plus significatifs (marqués par des croix vertes) ne passent pas le seuil qu'il faut appliquer quand on teste 500.000 hypothèses simultanées (correction de tests multiples).
- Remarques
- Le facteur g est un outil rudimentaire pour mesurer l'intelligence. Lire La malmesure de l'homme (S.J. Gould, 1981) pour une critique du facteur "g".
- En dépit des résultats négatifs de leur étude, les auteurs sont convaincus qu'une étude plus approfondie (avec plus de SNPs et des cohortes plus grandes) permettrait de prédire les risques génétiques et d'investiguer les effets fonctionnels des gènes sur le cerveau et le comportement.
- Ils n'envisagent à aucun moment que les capacités cognitives pourraient être indépendantes des gènes.



Butcher et al. (2008), Figure 2: A scatter plot showing the 47 top-ranked SNPs (crosses) against the background of unselected SNPs comparing allele frequencies for the low g group (x-axis) and the high g group (y-axis).

- Agrégation de données pour 5 troubles psychiatriques.
- 32.332 cas + 27.888 contrôles.
- Identification de plusieurs loci (positions génomiques) très significatives, qui étaient faiblement significatives dans chacune des maladies étudiées séparément.
- Certains gènes étaient préalablement ressortis séparément des études de schizophrénie et trouble bipolaire.

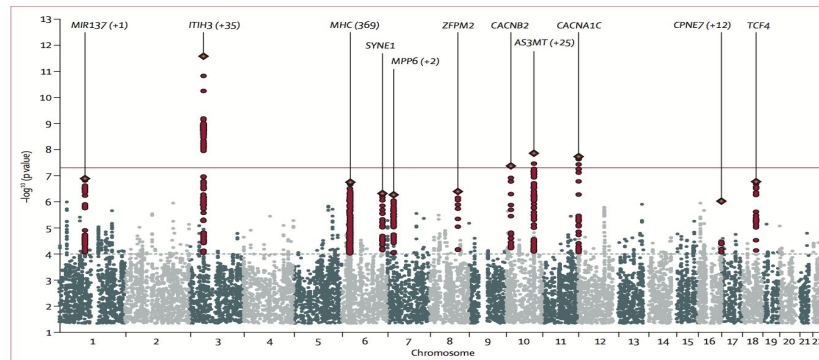


Figure 1: Manhattan plot of primary fixed-effects meta-analysis. Horizontal line shows threshold for genome-wide significance ($p < 5 \times 10^{-8}$).

	Chromosome	Base-pair position*	Nearest gene	Alleles	Frequency†	Imputation quality score (INFO)	p value	OR (95% CI)‡	Heterogeneity p value	Best-fit model (BIC)§
rs2535629	3	52808259	ITIH3 (+ many)	G/A	0.651	0.942	2.54×10^{-12}	1.10 (1.07-1.12)	0.27	Five disorder¶
rs11191454	10	104649994	AS3MT (+ many)	A/G	0.910	1.01	1.39×10^{-8}	1.13 (1.08-1.18)	0.32	Five disorder¶
rs1024582	12	2272507	CACNA1C	A/G	0.337	0.98	1.87×10^{-8}	1.07 (1.05-1.10)	0.0057	BPD, schizophrenia
rs2799573	10	18641934	CACNB2	T/C	0.715	0.825	4.29×10^{-8}	1.08 (1.05-1.12)	0.57	Five disorder¶

Most strongly associated single-nucleotide polymorphisms (SNP) in associated region after clumping—i.e., grouping SNPs within 250 kb of the index SNP that have $r^2 > 0.2$ with the index SNP as implemented in PLINK. OR=odds ratio. BIC=Bayesian information criteria. BPD=bipolar disorder. *Detected with University of California Santa Cruz Genome Browser (version hg18). †Risk allele frequency in controls. ‡Estimated OR from multinomial logistic regression used in the modelling analysis. §Best-fit multinomial logistic model by BIC criteria; appendix pp 38–45 provide a comparison of BIC and Akaike information criteria across models. ¶Best-fit model supports an effect on all five disorders.

Table 1: Five disorder meta-analysis results for regions with $p < 5 \times 10^{-8}$

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). [Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis](#). Lancet. 2013 Apr 20;381(9875):1371-9. Erratum in: Lancet. 2013 Apr 20;381(9875):1360.

Une étude à très large échelle pour la schizophrénie

2014: une étude génomique d'association basée sur 40.000 cas de schizophrénie et 113.000 contrôles révèle 108 loci fortement significatifs.

25 déjà détectés dans des études précédentes, mais à des niveaux de significativité limite.

83 nouveaux loci

Quelques candidats notables

NOTCH4. Récepteur transmembranaire impliqué dans des voies de signalisation (communications entre cellules voisines par liaison d'une protéine-signal au récepteur).

Protéine à doigt de Zinc (ZNF804A), facteur transcriptionnel exprimé dans le cortex, hippocampe et cervelet, fonction méconnue.

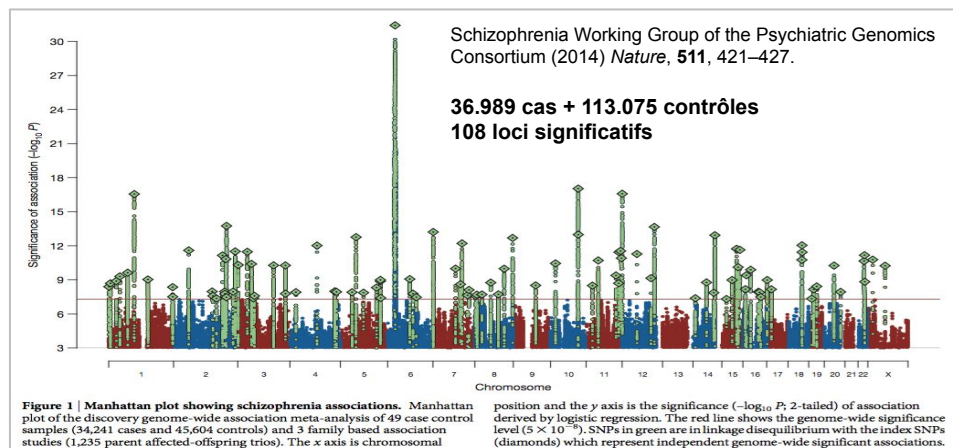
CACNA1C, CACNB2, CACNA11: famille de protéines canaux à ions calcium

Ces associations permettent ensuite de mener des études moléculaires pour comprendre la fonction de ces gènes et leur éventuelle implication dans la pathologie.

Elles ne permettent cependant pas de prédire les risques pour un individu de devenir schizophrène.

Pourcentage de schizophrénie dans la population générale: 1%.

Chez les porteurs d'allèles associés à la maladie : 1.7%, ce qui laisse 98.3% de chances de ne pas être schizophrène si on a le « mauvais » allèle.



Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511, 421–427.

Pouvoir prédictif et utilité clinique de la génomique personnelle

- Pour un bon nombre de maladies humaines on a pu détecter un ou des gènes associés (études d'association).
- **A partir du profil génomique d'un individu, peut-on établir son état de santé et agir dessus ?**
 - déclenchement effectif de la maladie ?
 - susceptibilité à la maladie ?
 - pronostic ?
 - réponse personnelle à un traitement) ?
- **Validité analytique** : validité des résultats techniques (pour la génomique, taux de fiabilité du séquençage)
- **Validité clinique**: capacité d'un test à détecter ou prédire la pathologie associée (Hunter et al., 2008).
 - Peut-on utiliser les profils génomiques pour prédire, sur base du génome d'un individu, la probabilité qu'il a de déclencher une maladie ?
 - L'estimation de la validité clinique repose sur une série de paramètres apportant des informations complémentaires: sensibilité, spécificité, valeur prédictive positive (PPV), valeur prédictive négative (NPV).
- **Utilité clinique**: équilibre entre les risques et les bénéfices si un test doit être mis en pratique.
 - **Par exemple, s'il n'existe aucun traitement, la connaissance d'une prédisposition peut nuire à la qualité de la vie sans apporter de bénéfice évident.**
- Hunter et al. (2008) and Evans et al. (2010) soulignent
 - l'absence actuelle de données à propos de la validité clinique,
 - leur scepticisme concernant l'utilité clinique.

Hunter et al. Letting the genome out of the bottle--will we get our wish?. N Engl J Med (2008) vol. 358 (2) pp. 105-7. doi.org/10.1056/NEJMp0708162

Evans et al. The rules remain the same for genomic medicine: the case against "reverse genetic exceptionalism". Genet Med (2010) vol. 12 (6) pp. 342-3.

doi.org/10.1097/GIM.0b013e3181deb308

Goddard, K. A. B., Lee, K., Buchanan, A. H., Powell, B. C. & Hunter, J. E. Establishing the Medical Actionability of Genomic Variants. Annu. Rev. Genom. Hum. Genet. 23, 173–192 (2022).

doi.org/10.1146/annurev-genom-111021-032401

- La **validité analytique** est elle-même restreinte:
 - Toutes les méthodes (biopuces à SNP, séquençage de génome complet) impliquent un certain taux d'erreurs technologiques.
 - Même si ce taux est faible, des erreurs peuvent occasionnellement affecter les marqueurs utilisés pour prédire un risque, ce qui amène à de fausses prédictions pour l'individu concerné.

- **Prévalence**

- Proportion des individus d'une population donnée qui présentent un attribut particulier (par exemple une maladie ou un trait génotypique). On peut calculer la prévalence ponctuelle (à un moment donné) ou pendant un intervalle de temps (prévalence d'intervalle) ou tout au long de la vie.
- Exemples
 - Prévalence du cancer du sein : 1.7% des femmes en vie aux Etats-Unis en 2012 avaient été diagnostiquée d'un cancer du sein à un moment de leur vie.
 - Prévalence de la mutation BRCA1
 - ~0.1% dans la population américaine
 - ~6% chez les femmes diagnostiquées d'un cancer du sein avant 35 ans

- **Pénétrance**

- Proportion, parmi les individus ayant un type particulier de génotype, de ceux qui exprimeront le phénotype de la maladie associée pendant une période de temps déterminée.
- Exemple: 65% des femmes porteuses de l'allèle de BRCA1 déclarent un cancer du sein avant l'âge de 70 ans

- Hunter et al. (2008) développent une série de raisons pour rester sceptiques quant à l'utilité immédiate des profils génomiques.
- **Valeur prédictive positive (PPV)** : proportion des personnes étant effectivement affectées par la pathologie parmi celles déclarées positives
- **Valeur prédictive négative (NPV)** : proportion de personnes qui ne sont effectivement pas affectées par la pathologie parmi celles déclarées négatives
- La **valeur prédictive** de marqueurs génétiques individuels peut s'avérer très faible.
 - Pour les **maladies polygéniques**, les marqueurs individuels peuvent avoir un intérêt réduit. Quelle est l'utilité de prédire un risque de 1.5%?
 - Pour les maladies complexes, les **facteurs environnementaux** rendent les prédictions encore moins fiables.

- Pour les maladies polygéniques et multifactorielles (polygéniques + facteurs d'environnement), les GWAS détectent des centaines de SNPs significatifs, mais la valeur prédictive de chacun est marginale.
- Une approche pour prédire les risques de maladies polygéniques sur base du génotypage consiste à calculer le **score polygénique de risque (Polygenic Risk Score, PRS)** d'un individu. Ce score prend en compte une série de SNPs ressortis d'une GWAS, avec une pondération qui tient en compte la significativité statistique de l'association de chaque SNP.
- Amélioration de la valeur prédictive par rapport à une approche "SNP unique", mais cette valeur prédictive reste faible.
- La prise en compte de facteurs additionnels (âge, sexe, environnement familial et social) permet d'améliorer la valeur prédictive.

- **Validité clinique:** capacité d'un test à détecter ou prédire la pathologie associée (Hunter et al., 2008).
 - L'estimation de la validité clinique repose sur une série de paramètres apportant des informations complémentaires: sensibilité, spécificité, valeur prédictive positive (PPV), valeur prédictive négative (NPV).

La validité clinique des études d'association à échelle génomique sera restreinte aux maladies présentant les caractéristiques suivantes.

- **Facteurs causaux agissant indépendamment** ou interagissant de façon simple (effets additifs).
- **Prévalence suffisante:**
 - Les études requièrent une représentation suffisante de l'allèle mineur pour détecter son effet.
- **Pénétrance suffisante**
 - La proportion de personnes présentant les symptômes parmi celles qui portent l'allèle associé au risque.

- Capacité à agir sur les risques associés au génotype, une fois qu'on les a détectés
- Voies envisageables
 - Prévention
 - Traitement médicamenteux
 - Thérapie génique

Thérapie génique

Première tentative de thérapie génique

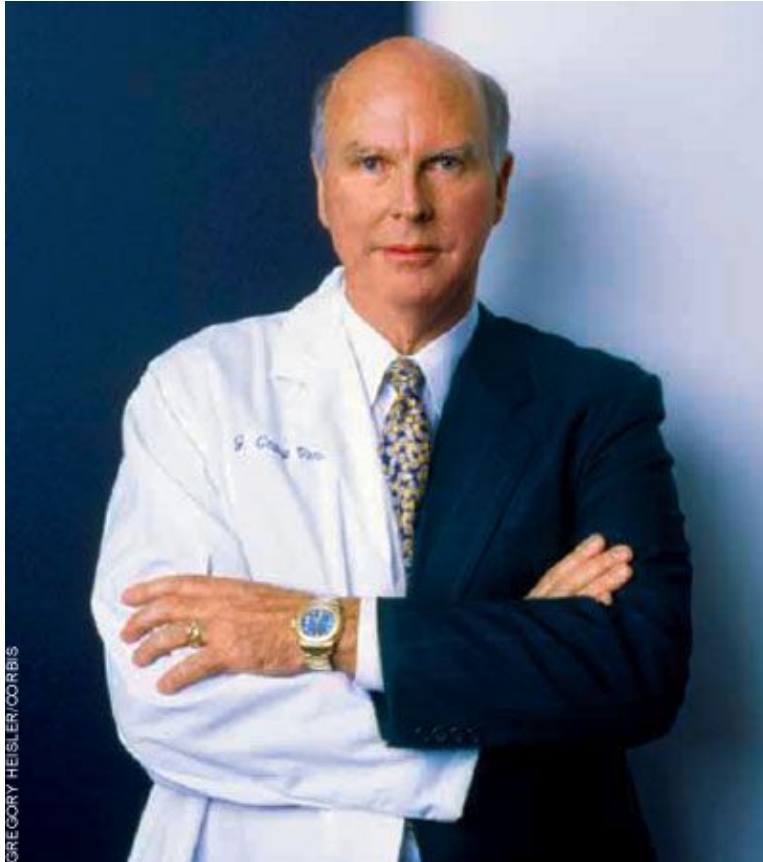
- La thérapie génique consiste à corriger un problème génétique en insérant dans les cellules une copie fonctionnelle du gène.
- La première étude clinique de thérapie génique a porté sur des « **enfants-bulles** ». Cette métaphore désigne des individus qui souffrent d'une déficience innée du système immunitaire, et ne peuvent survivre que dans un milieu complètement aseptisé (une « bulle » isolée du monde des microbes).
- Cette déficience provient d'une mutation d'un seul gène codant pour l'adénosine déaminase.
- Au début des années 2000 on a traité une vingtaine d'enfants-bulles en insérant dans le gène fonctionnel dans des cellules-souches hématopoïétique (ces cellules produisent les globules blancs). La méthode d'insertion reposait sur l'utilisation d'un virus.
- Du point de vue immunitaire, le traitement a donné les résultats escomptés : les enfants traités ont développé une réponse immunitaire.
- Cependant, une partie d'entre eux ont ensuite développé des leucémies, qui résultaient vraisemblablement de perturbations des gènes de contrôle du cycle cellulaire, liées au point d'insertion du gène dans le génome.
- Suite à cet échec, les essais cliniques ont été temporairement suspendus de 2002 à 2004.
- Ils ont ensuite été repris en se basant sur d'autres vecteurs pour transporter l'ADN dans les cellules (par exemple des liposomes).

- La technologie CRISPR-Cas9, développée en 2012, permet de modifier la séquence d'un gène in situ, c'est-à-dire dans le génome même. On parle donc d'édition génomique.
- En principe, elle permet donc non seulement d'apporter une copie fonctionnelle d'un gène pour pallier à une déficience, mais également de corriger des mutations dominantes (par exemple l'anémie falciforme).
- Cette technologie ouvre donc des perspectives sans précédent pour la thérapie génique.
- La principale limitation est le ciblage cellulaire et tissulaire : on peut appliquer la technologie à des cellules isolées du corps (par exemple des cellules-souches) mais on ne dispose pas de moyens pour corriger le génome des cellules déjà différenciées qui composent les tissus d'un organisme.

- La **thérapie génique** est déjà utilisée pour traiter certaines maladies.
- Exemples d'applications récentes
 - drépanocytose
 - mucoviscidose
 - dystrophie musculaire
- Méthodes
 - Vecteur adénovirus
 - Edition génomique via CRISPR-Cas9
- Il s'agit toujours de **maladies monogéniques**.
- L'application à des maladies multifactorielles semble a priori plus complexe, à la fois parce que les gènes potentiellement impliqués sont nombreux, et parce que la maladie dépend également de facteurs non-génétiques (environnement).

- En décembre 2018, le chercheur chinois He Jankui annonce à la presse qu'il a utilisé la technologie CRISPR-Cas9 pour modifier le génome de cellules-souches embryonnaires, et leur conférer la résistance au virus HIV (responsable du SIDA).
- Ceci contrevient à un consensus déontologique selon lequel il est admissible d'effectuer des modifications génétiques sur les cellules somatiques (celles qui composent le corps mais ne participent pas à la transmission héréditaire) mais pas sur les lignées germinales (lignées cellulaires qui donneront des gamètes).
 - Les cellules souches embryonnaires donnent naissance à l'ensemble des cellules de l'individu, et donc ses cellules germinales et somatiques.
- L'affaire suscite un fort émoi médiatique et dans la communauté scientifique.

Enjeux économiques



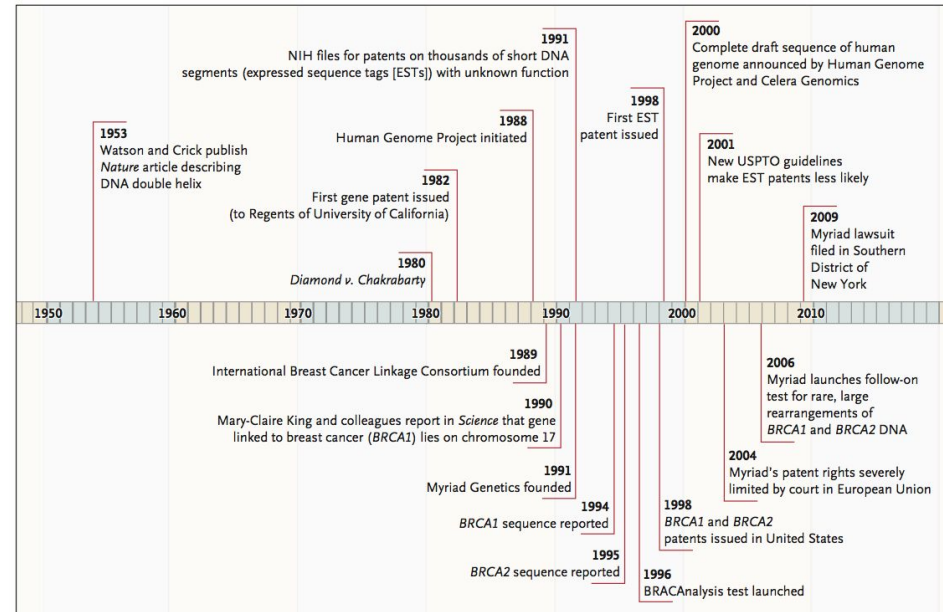
- Parallèlement au projet public de séquençage, une firme privée, appelée Celera et dirigée par Craig Venter, a annoncé qu'elle aurait séquencé l'ensemble du génome humain en 2001, soit trois ans avant la fin prévue pour le projet public.
 - Le but du projet privé de séquençage du génome humain était de séquencer rapidement le plus grand nombre de gènes possible, afin de déposer des brevets sur ces gènes.
- Il existait une différence de qualité de séquençage entre les deux projets
 - La firme Celera avait opté pour une stratégie "shotgun", qui produit rapidement un très grand nombre de petits fragments de séquence. Cette méthode rapide fournit un résultat incomplet et de mauvaise qualité (un brouillon).
 - Le projet public avait opté pour une approche progressive: obtenir une séquence de haute qualité pour le chromosome 22 (le plus petit chromosome du génome humain), afin de pouvoir commencer son interprétation biologique pendant le séquençage des autres chromosomes.

- La directive européenne 98/44/CE du 6 juillet 1998 relative à la protection juridique des inventions biotechnologiques fait l'objet de nombreux débats.
 - <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A31998L0044>
- L'article 5 de cette directive contient une contradiction interne
 - Le corps humain, aux différents stades de sa constitution et de son développement, ainsi que la simple découverte d'un de ses éléments, y compris la séquence ou la séquence partielle d'un gène, ne peuvent constituer des inventions brevetables.
 - Un élément isolé du corps humain ou autrement produit par un procédé technique, y compris la séquence ou la séquence partielle d'un gène, peut constituer une invention brevetable, même si la structure de cet élément est identique à celle d'un élément naturel.
 - L'application industrielle d'une séquence ou d'une séquence partielle d'un gène doit être concrètement exposée dans la demande de brevet.

- Un des problèmes du brevetage est que les laboratoires académiques hésitent à entamer un travail de caractérisation d'un gène sur lequel existe un brevet, car ils devront ensuite négocier les applications éventuelles de leurs découvertes avec le possesseur du brevet.
- Le brevetage d'une partie des gènes constitue donc un frein à la découverte de leur fonction.
- Chaque gène identifié dans le projet Celera faisait l'objet de demandes de brevets, tandis que ceux du projet public étaient mis en public.
- Les responsables du projet public de séquençage se sont rapidement rendus compte que la plupart des gènes auraient été brevetés par Celera avant d'avoir été identifiés par la stratégie de séquençage de haute qualité du projet public. Ils ont donc complètement changé de stratégie en cours de projet, et décidé d'obtenir un brouillon le plus rapidement possible, pour revenir dans un deuxième temps au séquençage de qualité.
- Une étude de 2005 révèle que 20% des gènes humains font l'objet d'un brevet
 - 63% de ces brevets appartiennent à des firmes privées
 - 28% à des universités
 - Source: Jensen & Murray (2005). Intellectual Property Landscape of the Human Genome. Science, 310: 239 - 240.

Les brevets sur les gènes BRCA1 et BRCA2

- Les gènes BRCA1 et BRCA2 sont associés à certains types de cancer du sein.
- Séquençage: 1994/1995
- 1998: la compagnie Myriad dépose un brevet sur l'utilisation de ces gènes comme indicateurs de risques pour le cancer du sein et de l'ovaire.
- Le test coûtait initialement 1600\$ US. En 2009 son prix avait doublé.
- En 2009, ce brevet représente l'essentiel des revenus annuels (326 M\$/an) de la compagnie.

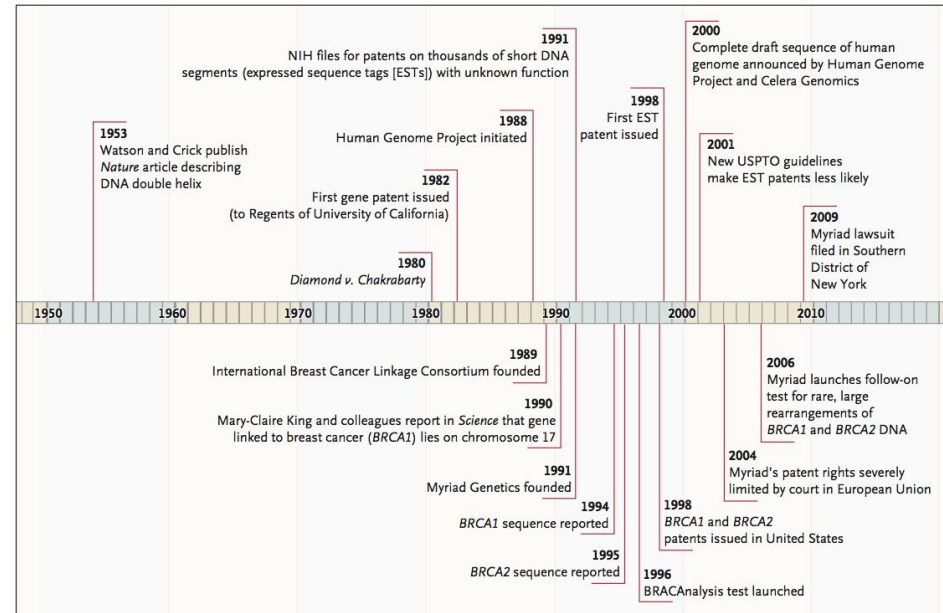


Timeline of Important Events in DNA Patenting (Top) and the Discovery and Use of Genes Conferring Susceptibility to Breast and Ovarian Cancer (Bottom).

NIH denotes National Institutes of Health, and USPTO U.S. Patent and Trademark Office.

Les brevets sur les gènes BRCA1 et BRCA2

- 2010: le brevet de la compagnie Myriad est annulé.
- Le caractère inventif du test est contesté par le bureau européen des brevets.
- Le séquençage ne constitue pas une invention, mais une découverte. D'après la cour, les « produits naturels et propriétés naturelles des objets vivants sont légalement distincts d'objets manufacturés en usant d'une ingéniosité substantielle ».
- La méthode se limite à comparer une séquence de l'échantillon à une séquence de référence, et n'est pas brevetable en soi.
- Le brevet, portant sur la séquence, empêche d'autres compagnies de développer des tests alternatifs utilisant les mêmes gènes.
- En 2013,
 - La cour suprême des Etats-Unis confirme l'annulation du brevet pour Myriad.
 - La cour fédérale d'Australie reconnaît la validité du brevet.
- Octobre 2015: suite à un appel, la cour fédérale d'Australie prend à l'unanimité la décision inverse :
"the high court found that an isolated nucleic acid, coding for a BRCA1 protein, with specific variations from the norm that are indicative of susceptibility to breast cancer and ovarian cancer was not a 'patentable invention.'"



Timeline of Important Events in DNA Patenting (Top) and the Discovery and Use of Genes Conferring Susceptibility to Breast and Ovarian Cancer (Bottom).

NIH denotes National Institutes of Health, and USPTO U.S. Patent and Trademark Office.

Le service direct au “consommateur”

Profils génomiques : 600.000 SNPs pour une poignée de dollars



Find out what your DNA says about you and your family.

- See how your DNA breaks out across 2000+ regions worldwide
- Discover DNA relatives from around the world
- Share reports with family and friends
- Learn how your DNA influences your facial features, taste, smell and other traits

order now

~~USD\$99~~ **USD\$79**


Limit 3 kits. Offer ends 23 Nov.

<http://www.23andme.com/>

- La compagnie 23andMe vend des profils génomiques constitué de 580.000 marqueurs polymorphiques.
- Prix: 79 \$US (2022).
- Financement: Google
- Services offerts
 - Santé: prédiction des risques de développer des maladies.
 - Généalogie : tests de parenté, origines géographiques
 - Réseaux « sociaux » (plutôt génétiques): 23andme propose de créer des communautés d'affinités sur base des gènes (« Facebook » génétique).
 - Recherche: le fait de disposer d'une base de données regroupant les profils de milliers d'individus permet de découvrir les marqueurs génétiques associés à certaines maladies.
- Ces services posent de grandes questions éthiques et juridiques, qui seront discutés au cours.



Carrier Status Disease Risk Drug Response



Find out if your child may be at risk for 40+ inherited conditions
Uncover your carrier status and what it means for your family.

23andMe Personal Genome Service® also includes:

- Your Disease Risk
- Your Drug Response

Order Now »

What mutations do we test for?



Cystic Fibrosis

One in 29 Caucasians carries a mutation that can cause cystic fibrosis. If both parents are carriers of disease-causing mutations, there is a 25% chance their child will be born with the disease. 23andMe tests for the recommended panel of cystic fibrosis mutations, plus several others.



Sickle Cell Anemia

Sickle cell anemia affects about one out of 625 African Americans and one out of 1,000 Hispanic Americans. Knowing your children are at risk for inheriting this disease can help you and your doctor make sure they get the care they need as soon as possible.



Tay-Sachs and More

Approximately one in five Ashkenazi Jews carries a mutation associated with a severe disease that could result in the early death of a child. 23andMe tests for most mutations routinely screened for in this population, including Tay-Sachs disease, Canavan disease, Niemann-Pick disease type A, and Bloom's syndrome.

•See our health reports on 235 diseases and conditions.

Learn if you carry inheritable disease markers.



Alpha-1 Antitrypsin Deficiency
Genetic Report on 2 reported markers.

Your Genetic Data

Who	What It Means
Lily Mendel	ZZ: Has Alpha-1 Antitrypsin Deficiency Increased risk for lung and liver disease.

You could be passing on more than you know.

23andMe can alert you to mutations in your DNA that could cause disease in your children. If you find that your children are at risk for an inherited condition, you can talk to a doctor or healthcare professional about what you can do.

Arm yourself with knowledge.



Counselors

Meet with a genetic counselor near you

- By zipcode and distance
- By name or company
- By name, area or practice, or state

Interactive Tools

View web-based tools and resources:

- Warfarin dosing tool for physicians

23andMe health reports include where to look for more information, support group contacts, and links to genetic counseling resources.


- Estimation des risques pour des maladies à terrain génétique.
- Traitements personnalisés.

Ceci ouvre énormément de questionnements déontologiques et juridiques

- Service direct au “consommateur” (direct to consumer service) pour des diagnostics médicaux
- Interprétation des résultats par les patients
- Protection des données personnelles



Carrier Status **Disease Risk** Drug Response




Take a more active role in managing your health

Your 23andMe results can help you and your doctor make more informed decisions about your healthcare.

23andMe Personal Genome Service® also includes:

- Your Carrier Status
- Your Drug Response

 **Order Now >**

Why should you know your genetic risk?



Make better lifestyle choices.

On average, **one person in five** develops diabetes by age 79. Variations in your DNA tested for by 23andMe might raise your risk to **one in three**, making your lifestyle choices on factors like exercise and weight control even more critical.



Be on the lookout for common conditions.

Age-related macular degeneration is the most common cause of irreversible vision loss in the Western world among people over 60. By learning if your genes put you at increased risk for this condition, you can choose to go in for more frequent eye exams and take other measures to protect your sight.



Prepare for serious diseases.

Some inherited mutations greatly increase your likelihood of developing certain diseases later in life. For example, one rare mutation tested by 23andMe is associated with a nearly 60% lifetime risk of **Parkinson's disease**.

»See our health reports on 235 diseases and conditions.

Your disease risk is impacted by your genetics.



10x increased risk
18.2 out of 100
You have a 10x increased risk of Type 2 Diabetes compared to the average person.

10x increased risk
18.2 out of 100
You have a 10x increased risk of Type 2 Diabetes compared to the average person.

10x increased risk
18.2 out of 100
You have a 10x increased risk of Type 2 Diabetes compared to the average person.

Learning your genetic risk for various diseases and conditions will allow you and your doctor to focus on the lifestyle changes and preventative steps that matter most for you.

State-of-the-art science at your fingertips.



Type 2 Diabetes
Clinical Report on 9 Reported Markers, updated June 05, 2015

About Type 2 Diabetes
The most common type of diabetes, Type 2 diabetes occurs when insulin, the hormone that helps regulate blood sugar, is ineffective at the body's normal response to eating carbs and proteins. Left untreated, Type 2 diabetes can lead to kidney failure, blindness, and neurological problems that increase the risk of heart attack or stroke. In the United States, about 25 million children and adults have diabetes, but the vast majority are Type 2 diabetes.


23andMe scientists pore through the scientific and medical literature, gathering relevant information into one easy to read report.

- La compagnie met également en avant les perspectives d'utiliser la connaissance du génome pour prévenir les risques (notamment pour les maladies multifactorielles).

Choix d'un traitement individualisé



Carrier Status Disease Risk **Drug Response**




Ensure treatment that's personalized to you.

With 23andMe, you and your doctor can determine whether certain medications will be right for you.

23andMe Personal Genome Service® also includes:

- Your Carrier Status
- Your Disease Risk

 **Order Now** >

Why do your genetics matter?



Personalize your medication with your doctor.

Knowledge about variations in your DNA can help your doctor determine if you need more or less of a medication, or whether you might be at increased risk for certain side effects.



Your DNA can impact how effective drugs will be.

Clopidogrel (Plavix®) helps prevent heart attacks by keeping blood cells from sticking together. But a genetic variation that interferes with the drug's metabolism prevents some people from getting the full effect.

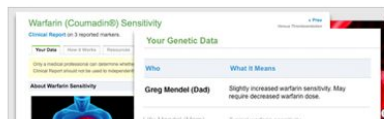


Genetics can determine if you'll have side effects.

Two genetic variations analyzed by 23andMe can increase the chances of experiencing **severe muscular pain and weakness** when taking high doses of cholesterol-lowering statins.

See our health reports on 235 diseases and conditions.

Not all drugs are right for all people.

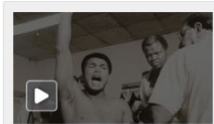
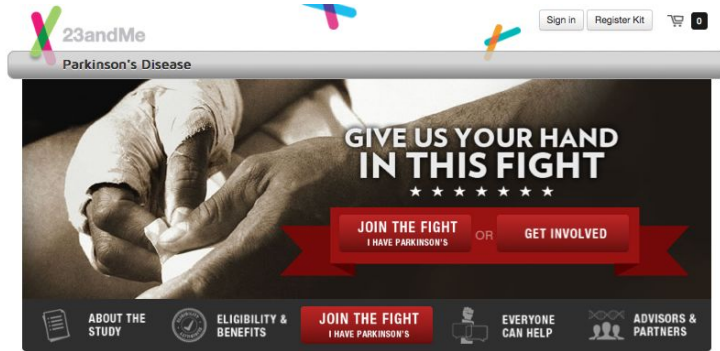


Your genetics can affect your sensitivity to drugs. The blood thinner **Warfarin (Coumadin®)** can help prevent blood clots, but it can also cause excessive bleeding in people whose sensitivity to this drug is increased by two genetic variations covered by 23andMe.

Easy to understand explanations at every step.



Learning about your genetics or your health shouldn't be hard. 23andMe makes things easy with **clear health reports** that tell you what you need to know about conditions that are important to you.



Muhammad Ali Joins 23andMe
Muhammad Ali inspires the world to join 23andMe, a leading personal genetics company that is changing the face of Parkinson's research.



ABOUT

Help revolutionize the way Parkinson's disease is studied and accelerate the search for a cure. Take an active role in groundbreaking research by mailing in your DNA sample and answering surveys online. Over 6,500 people with Parkinson's have come together to form what is now the largest Parkinson's community for genetic research in the world. This group is already powering research breakthroughs - we've found two new genes associated with Parkinson's and a gene that may be protective.

With 10,000 participants, the odds of finding a cure are even better. Join today.

HOW IT WORKS - WE NEED YOUR SPIT!

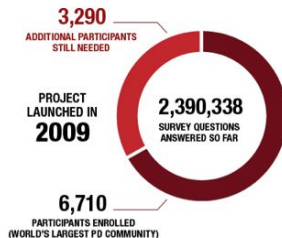


ELIGIBLE INDIVIDUALS

1. Diagnosed with Parkinson's disease by a physician
2. Access to the internet
3. Willing to complete online health surveys

HOW YOU BENEFIT

- FREE lifetime access to 23andMe's Personal Genome Service (valued at \$399)
- Learn about your health and ancestry
- Find out if you carry known genetic mutations associated with Parkinson's (if you choose)
- Receive monthly updates as new health discoveries are made and see how they apply to YOU
- Help advance medical research



- 23andme propose à tous ceux qui le désirent de contribuer à la recherche médicale, en fournissant des données sur leur santé : maladies, conditions générales de santé, traitements médicaux, parcours de soin, ...
- La participation est récompensée par des bonus (accès illimité aux services de génomique personnelle, alertes sur les facteurs de risque trouvés dans son propre génome, information sur les nouvelles découvertes, ...)
- Ces données seront ensuite utilisées pour effectuer des recherches concernant les liens entre données génomiques et santé.

Remarques

- Les bases de données qui appartiennent les séquences génomiques et les données de santé représentent un marché économique phénoménal.
- Le statut juridique de ces données a été mis en cause. En particulier, manque de clarté sur ce qui leur adviendrait en cas de rachat de la compagnie, fusion, faillite, ...
- Il existe aussi des risques concrets de piratage des données, avec un incident important en 2023 (voir diapo suivante)

- Octobre 2023: en piratant les mots de passe de ~14.000 d'utilisateurs, des hackers ont pu obtenir
 - les données complètes de ces 14.000 personnes
 - des informations concernant les profils génomiques, origines ethniques et les liens de parenté concernant 6.9 millions de clients de 23andme (sur un total de 14 millions).
- La propagation de l'impact (de 14.000 à 6.9 millions) a été possible grâce à l'option "DNA relatives", qui permet de partager son profil avec toutes les personnes présentant des caractéristiques génétiques similaires pour certaines parties de chromosomes.
- Suite à cet incident, la compagnie a temporairement inactivé cette option, et ensuite renforcé la sécurité de l'accès (double authentification).

Protection des données à caractère personnel (“privacy protection” en anglais)

Méthodes classiques pour assurer la protection des données à caractère personnel (dé-identification)

- **Pseudonymisation** : remplacer les données identifiantes (nom, prénom, date et lieu de naissance) par un code non-explicite, en assurant la confidentialité et la sécurité du “dictionnaire” (fichier qui établit les correspondances entre les données d’identification et le pseudonyme).
- **Anonymisation** : suppression de toute information permettant directement ou indirectement de ré-identifier la personne à l’origine d’un échantillon.
- **Dégradation des données** : ajout de bruit
- **Agrégation des données** : on ne donne pas accès aux données individuelles mais à des statistiques (tableaux de comptages, moyennes, écarts-types, ...) calculées sur l’ensemble des individus.
- Note : **ces mesures sont inopérantes pour les données génomiques** (voir diapo suivante)

Contrôle de l’accès aux données pour les chercheurs

- Demandes motivées à adresser à un **comité d’accès aux données** (spécifique à chaque ensemble de données)
- Les **autorisation d’accès sont limitées** au sous-ensemble de cas et aux champs d’information nécessaires pour le projet de recherche qui motive la demande
- Avant d’accéder aux données le chercheur doit signer un **accord de confidentialité** (non-rediffusion des données, destruction de sa copie au terme du projet de recherche)
- L’analyse doit être menée dans des **environnements informatiques certifiés pour l’hébergement des données de santé** (HDS), hyper-sécurisés et contrôlés

- Une particularité des données génomiques individuelles est que **la donnée est intrinsèquement identifiante**
 - Police scientifique : un jeu de 30 à 80 marqueurs suffit pour identifier une personne sans ambiguïté
 - Les profils génomiques actuels contiennent des millions de SNPs → même les données dégradées restent identifiantes
- La séquence génomique est **rattachable à des données de santé**
 - gènes de susceptibilité, facteurs de risques pour des maladies
- Le génome d'un individu **concerne également les personnes apparentées**
 - 50% d'allèles partagés avec enfants, parents, frères et soeurs
 - 100% du chromosome Y partagés via la lignée paternelle
 - 100% du chromosome mitochondrial partagés via la lignée maternelle
 - 1/8 d'allèles partagés avec les grands-parents et cousins
- Identification indirecte
 - Traçage des origines géographiques
 - Recoupements avec autres informations associées aux données ("métadonnées") : âge, ville d'origine, statut familial, ...

U.S. Genetic Information Nondiscrimination Act (GINA)

- Loi fédérale aux Etats-Unis sur la non-discrimination en matière d'information génétique
- Protège les citoyens contre la discrimination basée sur l'information génétique, en matière d'assurance maladie et de décisions d'emploi.
- Les **compagnies d'assurance et les plans de santé** n'ont pas le droit
 - de consulter les informations génétiques prédictives avant d'établir un contrat d'assurance
 - de demander ou d'exiger que l'assuré ou les membres de sa famille passent un test génétique
 - de restreindre l'adhésion à une assurance sur la base d'informations génétiques
 - de modifier les primes sur la base d'informations génétiques.
- La GINA interdit également aux **employeurs américains** (y compris les agences pour l'emploi, les organisations syndicales et les programmes de formation)
 - de discriminer les personnes qu'ils embauchent ou le montant de leur salaire sur la base d'informations génétiques ;
 - de demander ou d'exiger que l'employé ou les les membres de sa famille passent un test génétique ;
 - de divulguer les informations génétiques en leur possession, sauf dans des circonstances spécifiques et spécialement contrôlées.



Qui lira notre avenir dans nos gènes ?

Enjeux médicaux, sociaux et économiques de la génomique personnelle

■ Applications médicales

- ❑ **Études d'association à échelle génomique (GWAS):** identification d'allèles (formes spécifiques d'un gène) associés à des maladies, prédisposition à certains types de cancer, maladie de Huntington, ...
- ❑ **Diagnostic :** détection chez l'individu des causes d'une maladie
- ❑ **Pronostic :** évaluation des risques individuels, notamment pronostic anténatal (discuter la relation avec l'eugénisme)
- ❑ **Médecine personnalisée :** choix d'un traitement médical en tenant compte des spécificités génétiques.
- ❑ **Thérapie génique :** modification (« réparation ») des gènes responsables d'une maladie.

■ Applications privées

- ❑ **Assurances et banques :** estimation de risques médicaux avant d'établir des contrats d'assurance ou des prêts.
- ❑ **Employeurs :** estimation des risques médicaux à l'embauche, prévention d'accidents du travail.

■ Tests de parenté : établissement des liens de parenté entre individus (test de paternité, de maternité, liens plus éloignés).

- ❑ Usage privé (tests de paternité).
- ❑ Usage juridique : preuves de filiation (abandons d'enfants, contestation de paternité).
- ❑ Usage étatique : exigence de tests de parenté pour le regroupement familial d'immigrés

■ Identification génétique en criminalistique (empreintes génétiques)

- ❑ Identification de criminels à partir de traces corporelles.
- ❑ Identification des victimes par comparaison d'échantillons de cadavres avec les profils génétiques de membres de la famille.

■ Origines géographiques

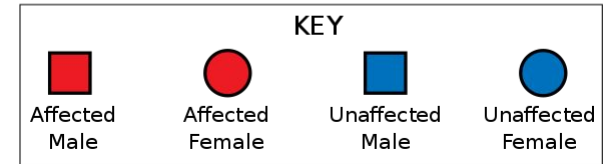
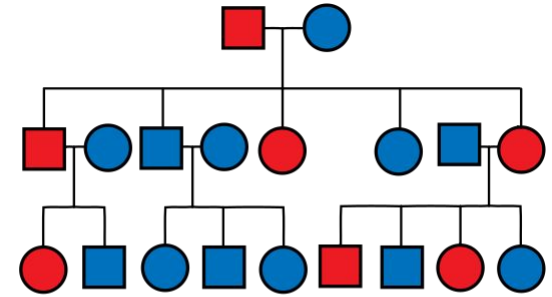
- ❑ Au niveau des individus: retraçage des origines ethniques.
- ❑ Au niveau des populations: hypothèses concernant les migrations de populations humaines. .

- Demain, le séquençage du génome sera incontournable dans les parcours de soin
 - Avantages ?
 - Risques ?
- Questions
 - Protection des données
 - Partage des données
 - Découvertes incidentes
 - Consentement éclairé
 - Eugénisme individuel ou d'Etat
 - Utilisation à des fins non-médicales
 - ...

Compléments d'information (pas vus au cours)

Maladies dominantes

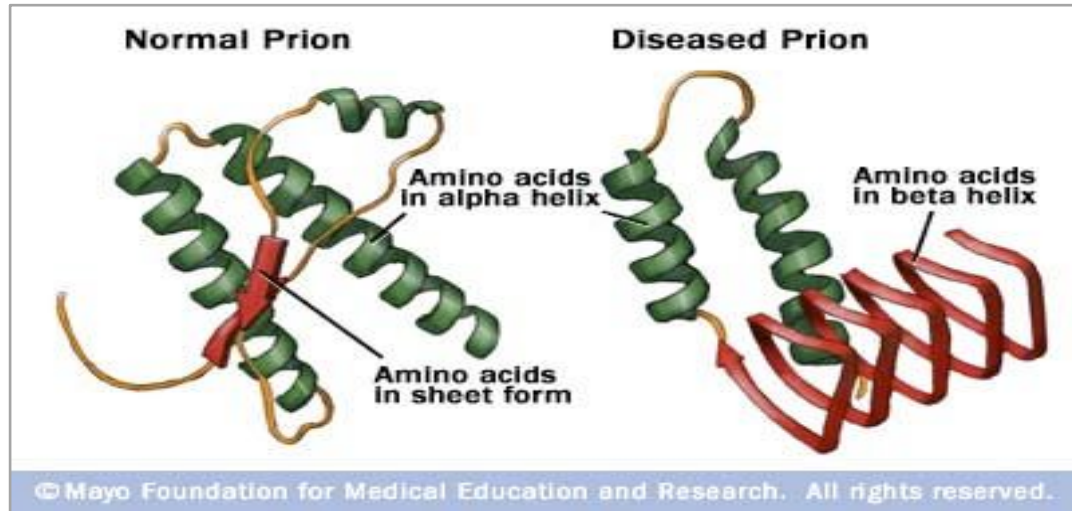
- Certaines maladies se transmettent au sein des familles selon un mode de transmission dominant (par exemple maladie de Huntington).
- Dominance: les individus qui héritent d'une copie mutée manifestent le phénotype, indépendamment du génotype de l'autre chromosome.
- Une transmission génétique est dite
 - **autosomique** si elle est liée à un chromosome non-sexuel (chromosomes 1 à 22 chez l'humain, pour lequel tous les individus sont diploïdes) -> la transmission ne dépend pas du genre.
 - **hétérosomique** (ou **allosomique**) si elle est liée à un chromosome sexuel (X ou Y). La transmission dépend alors du genre.
- Le schéma ci-dessous illustre l'hérédité



Une maladie liée à un changement de conformation protéique

- Un cas plus complexe: le syndrome de Kreuzfeld-Jacob est un cas très particulier, car la maladie ne provient pas d'une modification du gène, mais d'un changement dans la conformation (la forme tri-dimensionnelle) de la protéine.
- Cependant, certains allèles prédisposent au développement d'un syndrome de Kreuzfeld-Jacob.

Prion: une mauvaise conformation cause le syndrome de Kreuzfeld-Jacob
http://www.valleyhealth.com/images/image_popup/r7_prion.jpg



Exemples de loci associés à des maladies, détectés par des études à échelle génomique

Table. Pertinent Details of Findings of Recent Whole-Genome Association Studies (All From 2007)

Disease	Source	Gene/Locus ^a	No. of SNPs	Primary Study Cases/Controls	Replication Study Cases/Controls	OR _{het}	OR _{hom}	P Value	PAR, %
Breast cancer	Easton et al ¹	<i>FGFR2</i>	528 000	1145/1142	1776/2072	1.20	1.64	1 × 10 ⁻¹⁰	16
	Hunter et al ²	<i>FGFR2</i>	228 000	4398/4316	21 860/22 578	1.23	1.63	2 × 10 ⁻⁷⁶	NR
		<i>TNRC9</i>				1.23	1.39	1 × 10 ⁻³⁶	NR
		<i>MAP3K1</i>				1.13	1.27	2 × 10 ⁻²⁰	NR
		<i>LSP1</i>				1.06	1.17	3 × 10 ⁻⁹	NR
	Stacey et al ³	8q24				1.06	1.18	5 × 10 ⁻¹²	NR
		2q35	311 000	1600/11 536	4533/17 513	1.11	1.44	5 × 10 ⁻¹⁴	14
<i>TNRC9</i>					1.27	1.64	6 × 10 ⁻¹⁹	13	
CAD	McPherson et al ⁴	Chromosome 9p21	73 000	322/312	3989/18 805	1.20	NR	4 × 10 ⁻⁶	13
MI	Helgadóttir et al ⁵	Chromosome 9p21	306 000	1607/6728	4587/12 769	1.25	1.64	1 × 10 ⁻²⁰	21
Obesity	Frayling et al ⁶	<i>FTO</i>	490 000	1924/2938	3757/5346	1.32	1.67	3 × 10 ⁻³⁵	20
Diabetes	Sladek et al ⁷	<i>TCF7L2</i>	393 000	1380/1323	2617/2894	1.65	2.77	1 × 10 ⁻³⁴	28
		<i>SLC30A8</i>				1.18	1.53	6 × 10 ⁻⁸	24
		<i>HHEX</i>				1.19	1.44	3 × 10 ⁻⁶	19
		<i>EXT2</i>				1.25	1.50	1 × 10 ⁻⁴	16
	Steinthorsdóttir et al ⁸	<i>CDKAL1</i>	339 000	1399/5275	4739/9379	1.25	1.50	8 × 10 ⁻⁹	16
	Scott et al ⁹	<i>IGF2BP1</i>	315 000	1161/1174	1215/1258	1.14	NR	9 × 10 ⁻¹⁶	NR
		<i>CDKN2A/B</i>				1.20	NR	8 × 10 ⁻¹⁵	NR
		11p12				1.25	NR	4 × 10 ⁻⁷	NR
	Zeggini et al ¹⁰	<i>KCNJ11</i>	490 000	1924/2938	3757/5346	1.14	NR	5 × 10 ⁻¹¹	NR
	Saxena et al ¹¹	<i>PPARG</i>				1.14	NR	2 × 10 ⁻¹⁴	NR
Prostate cancer	Yeager et al ¹²	8q24	550 000	1772/1157	4290/4299	1.26	1.58	9 × 10 ⁻¹³	21
		8q24	316 000	1453/3064	1583/2817	1.71	NR	2 × 10 ⁻¹⁴	13 ^b

Abbreviations: CAD, coronary artery disease; het, heterozygotes; hom, homozygotes; MI, myocardial infarction; NR, not reported; OR, odds ratio; PAR, population-attributable risk; SNPs, single nucleotide polymorphisms.

^aFor information on genes listed here, see the National Center for Biotechnology Information gene database at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=Gene>.

^bThe PAR in individuals of African ancestry was 24%.