



# Méthodes d'exploration du génome



Solena LE SCOUARNEC, PhD  
[solena.lescouarnec@univ-nantes.fr](mailto:solena.lescouarnec@univ-nantes.fr)

Equipe I Génétique humaine

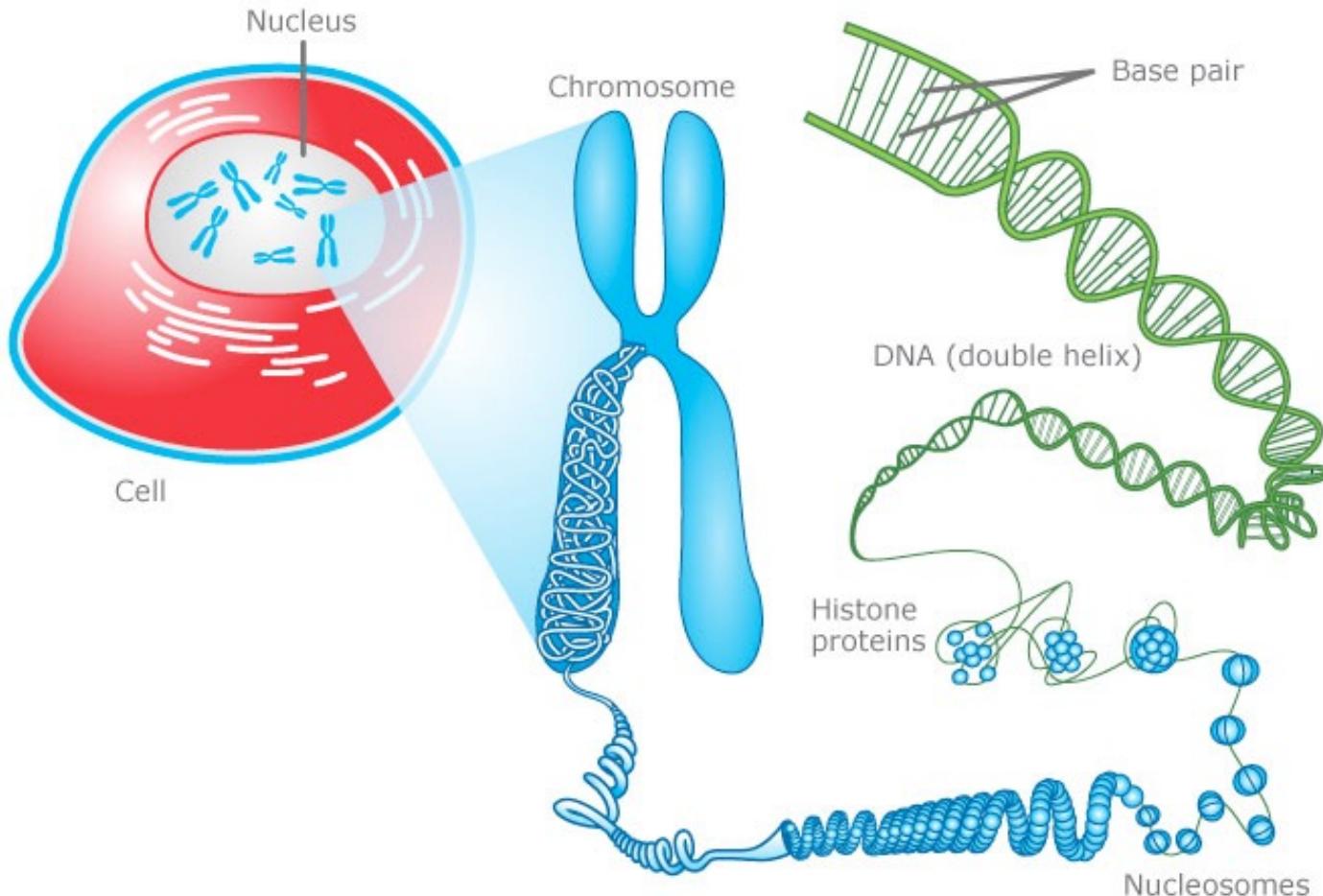
L'unité de recherche de l'institut du thorax  
Inserm UMR 1087 / CNRS UMR 6291  
Nantes, France

# Plan

- Introduction
- Variation du génome humain
- Méthodes d'identification de gènes
  - ✓ Maladies monogéniques
  - ✓ Maladies complexes

# Plan

- Introduction
- Variation du génome humain
- Méthodes d'identification de gènes
  - ✓ Maladies monogéniques
  - ✓ Maladies complexes



- Génome : ensemble du matériel génétique d'un organisme
- Chaque organisme possède son propre génome
- Génome humain : >3 milliards de paires de bases

# Génomique



## Genomics

- The study of an organism's complete set of genetic information.
- The genome includes both genes (coding) and non-coding DNA.
- 'Genome': the complete genetic information of an organism.

VS



## Genetics

- The study of heredity
- The study of the function and composition of single genes.
- 'Gene': specific sequence of DNA that codes for a functional molecule.

# Pourquoi la génomique?

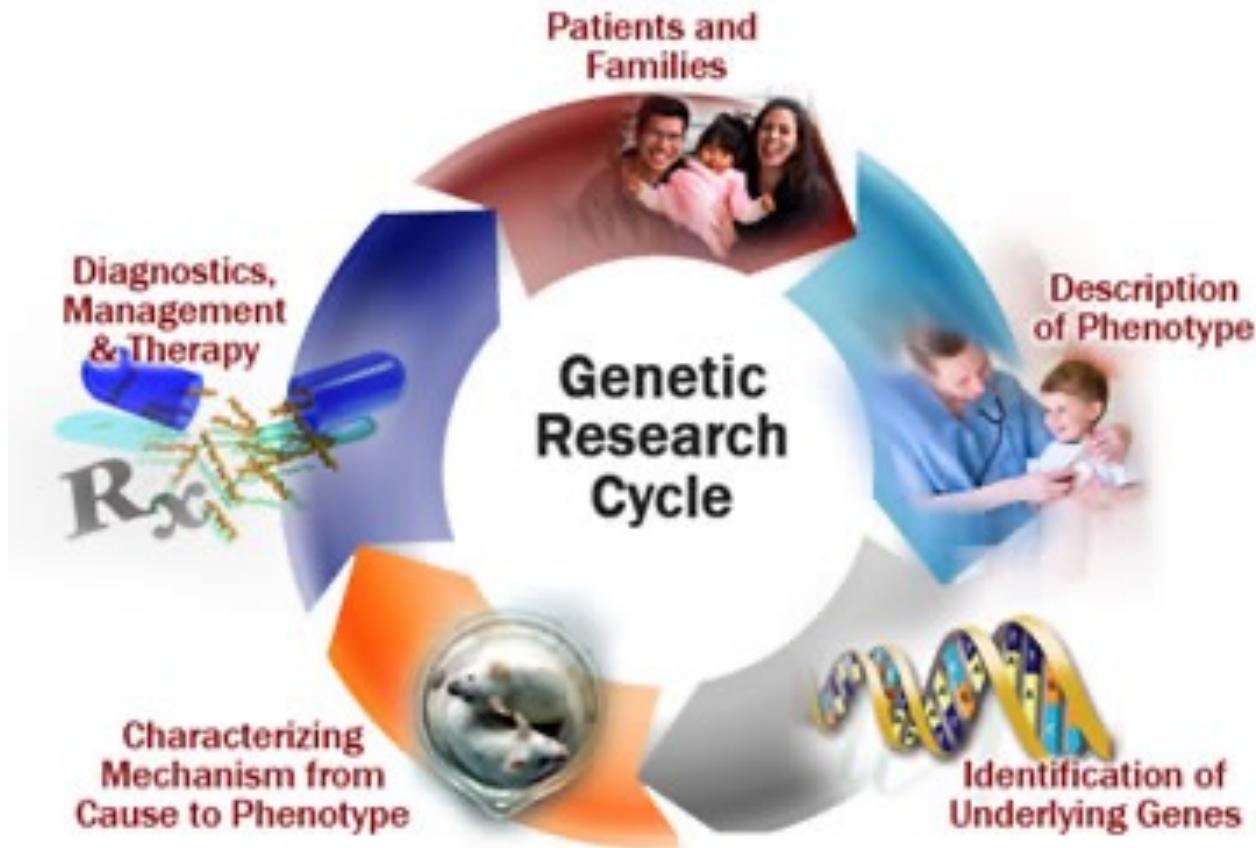
- Découvrir les variations génétiques à l'origine de pathologies
- Nouvelles thérapies
- Comprendre les bases moléculaires de l'évolution
- Découvrir les gènes impliqués dans l'embryogénèse et la physiologie

....

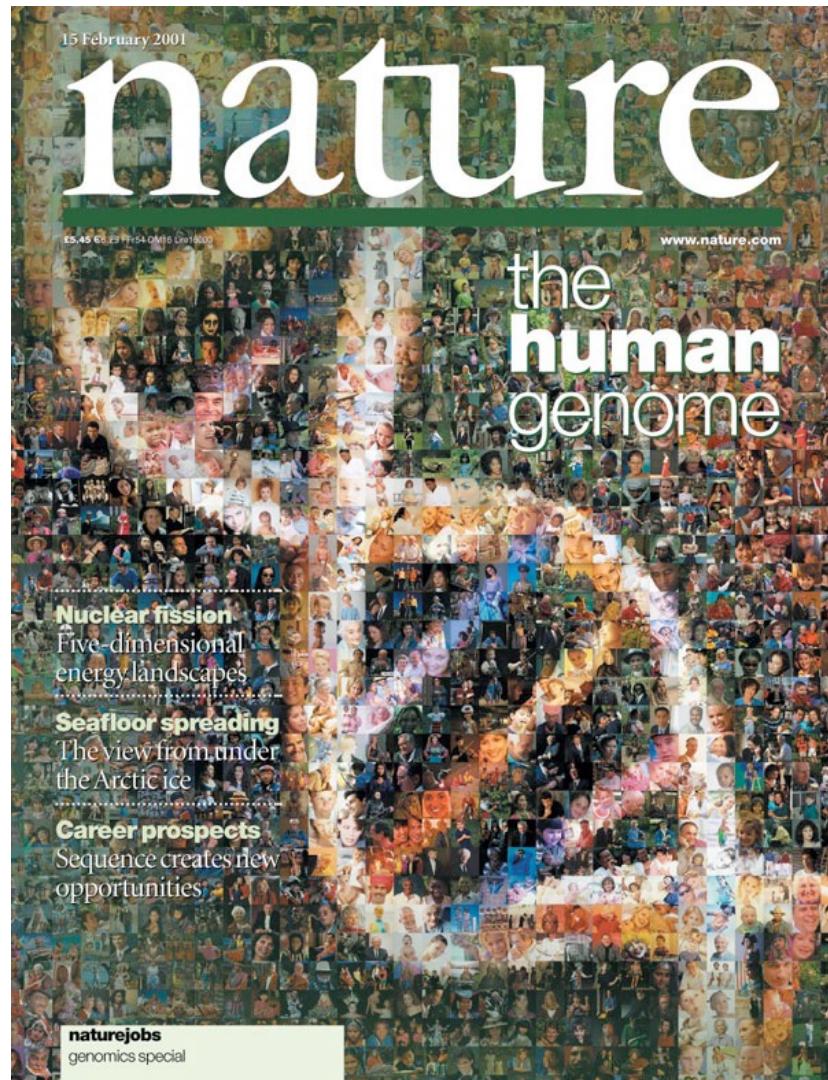
# Pourquoi la génomique?

La génétique moderne a transformé de nombreuses disciplines

Discipline récente, en plein essor, et en constante évolution



# Projet génome humain



## HUMAN GENOMICS

## The complete sequence of a human genome

Sergey Nurk<sup>1†</sup>, Sergey Koren<sup>1†</sup>, Arang Rhee<sup>1†</sup>, Mikko Rautiainen<sup>1†</sup>, Andrey V. Bzikadze<sup>2</sup>, Alla Mikheenko<sup>3</sup>, Mitchell R. Vollger<sup>4</sup>, Nicolas Altemose<sup>5</sup>, Lev Uralsky<sup>6,7</sup>, Ariel Gershman<sup>8</sup>, Sergey Aganezov<sup>9†</sup>, Savannah J. Hoyt<sup>10</sup>, Mark Diekhans<sup>11</sup>, Glennis A. Logsdon<sup>4</sup>, Michael Alonge<sup>9</sup>, Stylianos E. Antonarakis<sup>12</sup>, Matthew Borchers<sup>13</sup>, Gerard G. Bouffard<sup>14</sup>, Shelise Y. Brooks<sup>14</sup>, Gina V. Caldas<sup>15</sup>, Nae-Chyun Chen<sup>9</sup>, Haoyu Cheng<sup>16,17</sup>, Chen-Shan Chin<sup>18</sup>, William Chow<sup>19</sup>, Leonardo G. de Lima<sup>13</sup>, Philip C. Dishuck<sup>4</sup>, Richard Durbin<sup>19,20</sup>, Tatiana Dvorkina<sup>3</sup>, Ian T. Fiddes<sup>21</sup>, Giulio Formenti<sup>22,23</sup>, Robert S. Fulton<sup>24</sup>, Arkarachai Fungtammasan<sup>18</sup>, Erik Garrison<sup>11,25</sup>, Patrick G. S. Grady<sup>10</sup>, Tina A. Graves-Lindsay<sup>26</sup>, Ira M. Hall<sup>27</sup>, Nancy F. Hansen<sup>28</sup>, Gabrielle A. Hartley<sup>10</sup>, Marina Haukness<sup>11</sup>, Kerstin Howe<sup>19</sup>, Michael W. Hunkapiller<sup>29</sup>, Chirag Jain<sup>1,30</sup>, Miten Jain<sup>11</sup>, Erich D. Jarvis<sup>22,23</sup>, Peter Kerpeljiev<sup>31</sup>, Melanie Kirsche<sup>9</sup>, Mikhail Kolmogorov<sup>32</sup>, Jonas Korlach<sup>29</sup>, Milinn Kremitzki<sup>26</sup>, Heng Li<sup>16,17</sup>, Valerie V. Maduro<sup>33</sup>, Tobias Marschall<sup>34</sup>, Ann M. McCartney<sup>1</sup>, Jennifer McDaniel<sup>35</sup>, Danny E. Miller<sup>4,36</sup>, James C. Mullikin<sup>14,28</sup>, Eugene W. Myers<sup>37</sup>, Nathan D. Olson<sup>35</sup>, Benedict Paten<sup>11</sup>, Paul Peluso<sup>29</sup>, Pavel A. Pevzner<sup>32</sup>, David Porubsky<sup>4</sup>, Tamara Potapova<sup>13</sup>, Evgeny I. Rogaev<sup>6,7,38,39</sup>, Jeffrey A. Rosenfeld<sup>40</sup>, Steven L. Salzberg<sup>9,41</sup>, Valerie A. Schneider<sup>42</sup>, Fritz J. Sedlazeck<sup>43</sup>, Kishwar Shafin<sup>11</sup>, Colin J. Shew<sup>44</sup>, Alaina Shumate<sup>41</sup>, Ying Sims<sup>19</sup>, Arian F. A. Smit<sup>45</sup>, Daniela C. Soto<sup>44</sup>, Ivan Sovic<sup>29,46</sup>, Jessica M. Storer<sup>45</sup>, Aaron Streets<sup>5,47</sup>, Beth A. Sullivan<sup>48</sup>, Françoise Thibaud-Nissen<sup>42</sup>, James Torrance<sup>19</sup>, Justin Wagner<sup>35</sup>, Brian P. Walenz<sup>1</sup>, Aaron Wenger<sup>29</sup>, Jonathan M. D. Wood<sup>19</sup>, Chunlin Xiao<sup>42</sup>, Stephanie M. Yan<sup>49</sup>, Alice C. Young<sup>14</sup>, Samantha Zarate<sup>9</sup>, Urvashi Surti<sup>50</sup>, Rajiv C. McCoy<sup>49</sup>, Megan Y. Dennis<sup>44</sup>, Ivan A. Alexandrov<sup>3,7,51</sup>, Jennifer L. Gerton<sup>13,52</sup>, Rachel J. O'Neill<sup>10</sup>, Winston Timp<sup>8,41</sup>, Justin M. Zook<sup>35</sup>, Michael C. Schatz<sup>9,49</sup>, Evan E. Eichler<sup>4,53\*</sup>, Karen H. Miga<sup>11,54\*</sup>, Adam M. Phillippy<sup>1\*</sup>

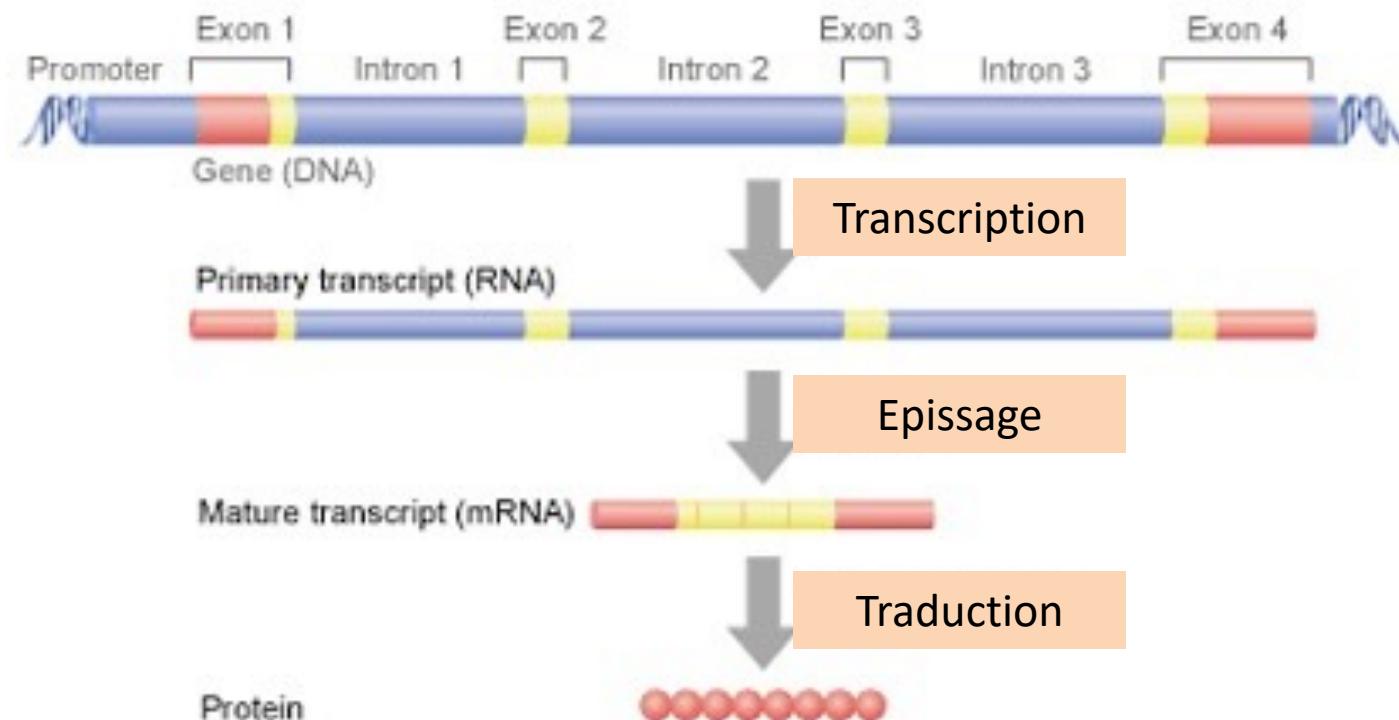
Since its initial release in 2000, the human reference genome has covered only the euchromatic fraction of the genome, leaving important heterochromatic regions unfinished. Addressing the remaining 8% of the genome, the Telomere-to-Telomere (T2T) Consortium presents a complete 3.055 billion-base pair sequence of a human genome, T2T-CHM13, that includes gapless assemblies for all chromosomes except Y, corrects errors in the prior references, and introduces nearly 200 million base pairs of sequence containing 1956 gene predictions, 99 of which are predicted to be protein coding. The completed regions include all centromeric satellite arrays, recent segmental duplications, and the short arms of all five acrocentric chromosomes, unlocking these complex regions of the genome to variational and functional studies.

# Gène

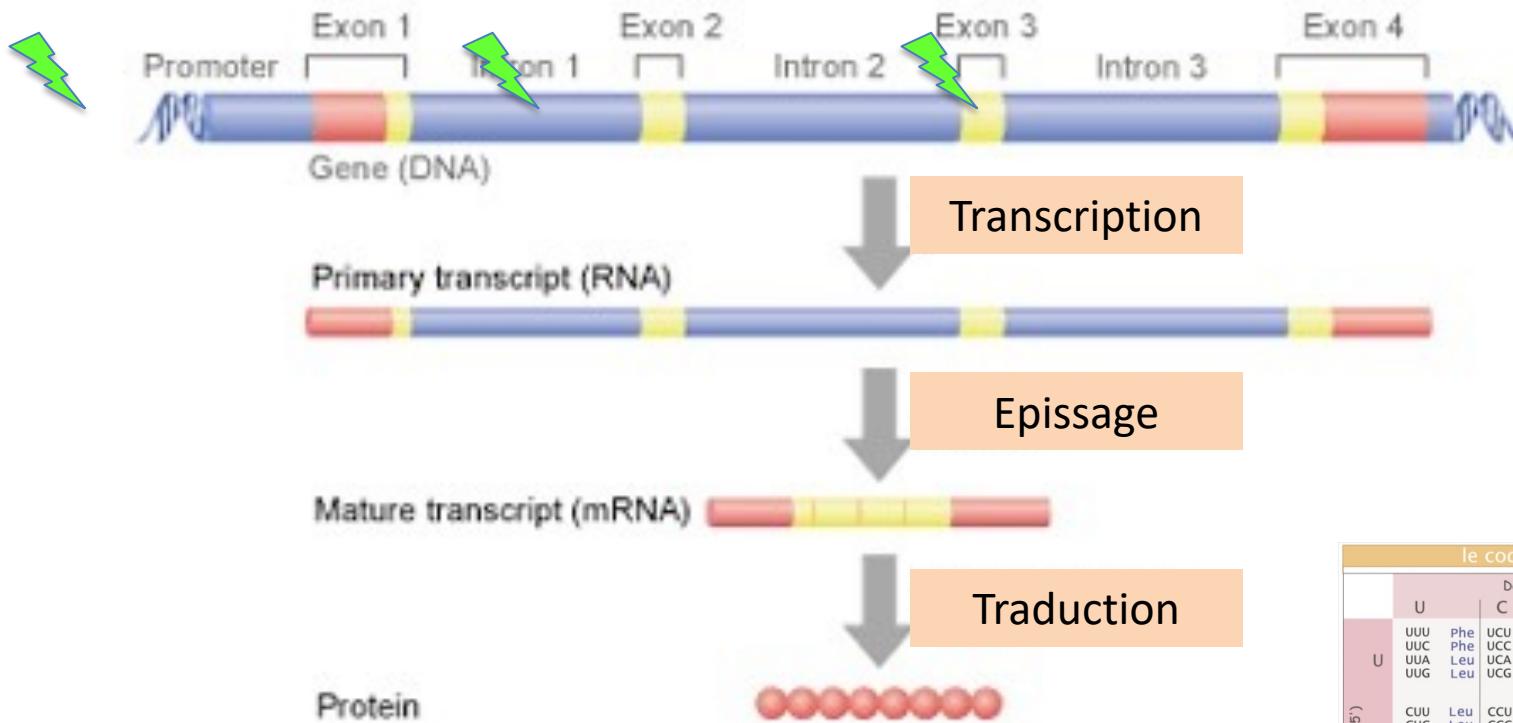
2 copies de chaque gène

20 000 – 25 000 gènes dans le génome humain

Exons / introns



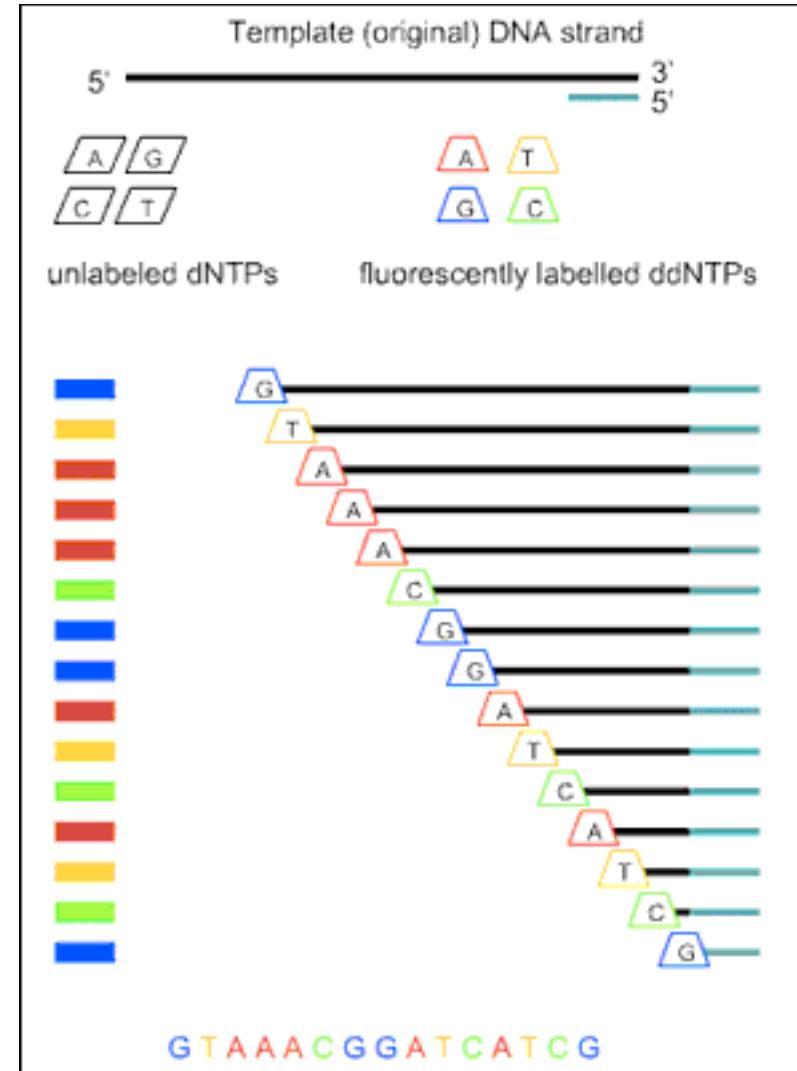
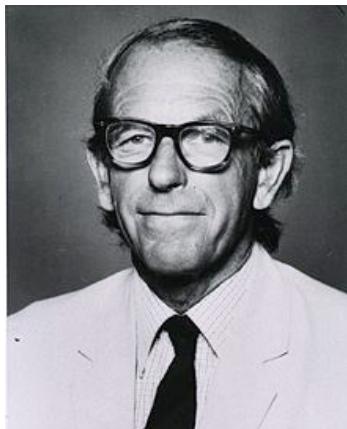
# Mutations génétiques



le code génétique												
	Deuxième lettre				Troisième lettre (côté 3')							
	U	C	A	G	U	C	A	G				
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGC	Cys	U			
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys				
	UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop				
	UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp				
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U			
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg				
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg				
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg				
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U			
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser				
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg				
	AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg				
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U			
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly				
	GUA	Val	GCA	Ala	GAA	Glu	GGG	Gly				
	GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly				

codon d'initiation codon de terminaison

# Séquençage Sanger



# Séquençage nouvelle génération (NGS)

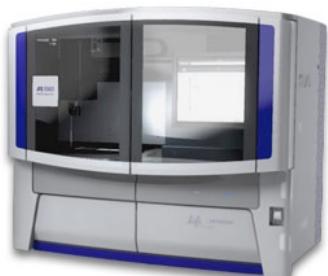
2005-2013



*Pyrosequencing*

life  
technologies™

**ThermoFisher**  
SCIENTIFIC



*Sequencing  
by ligation*

2006-

2006-

illumina®



*Sequencing  
by  
synthesis*

2011-

PB

PACIFIC  
BIOSCIENCES™

*Single molecule  
via dye labels  
Long reads*

2015-

Oxford  
**NANOPORE**  
Technologies™

*Electronic nanopore sensing  
No dye labels*

# Nouvelles technologies de séquençage

Next-generation sequencing (NGS) = Séquençage massif = Séquençage haut-débit

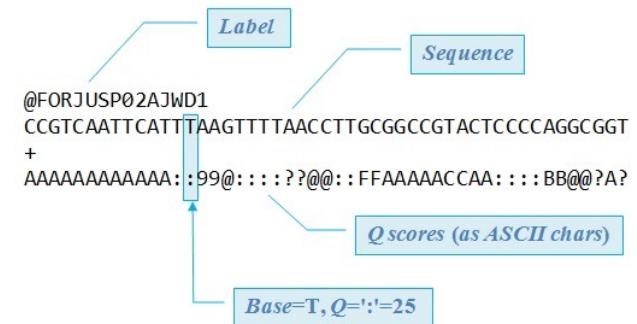
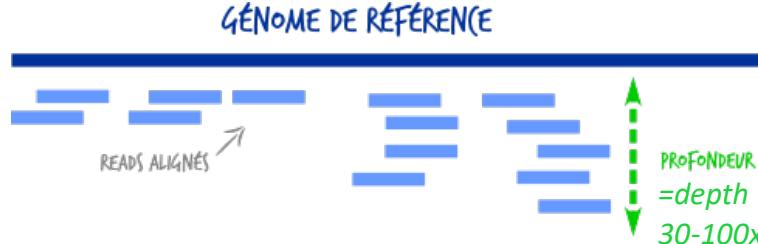
- Séquençage de génomes
- Séquençage ciblé (panels de gènes, exomes)
- Transcriptome
- Epigénétique
- ...



## Technologies

- Short-read (Illumina) ~150 pb
- Long-read (ONT, PacBio) 10-100 kb
- Linked-read (10X Genomics + Illumina)

} Substitutions / indels  
} Variants structuraux / assemblage *de novo*



# Séquençage de génomes personnels

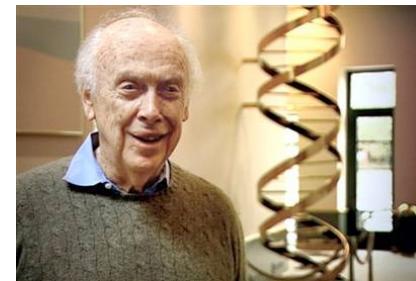
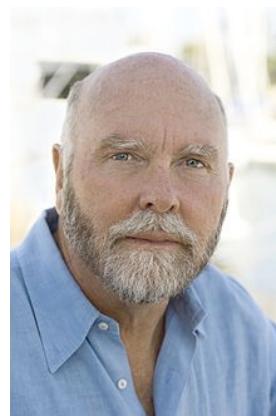
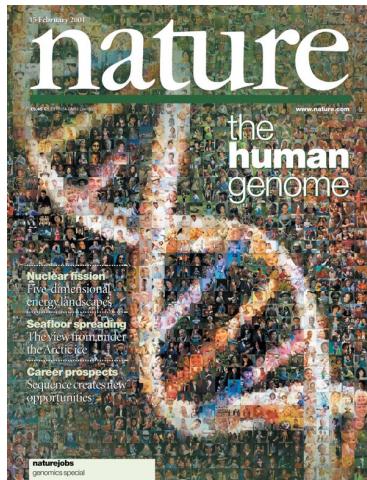
QUICKER, SMALLER, CHEAPER

Sanger

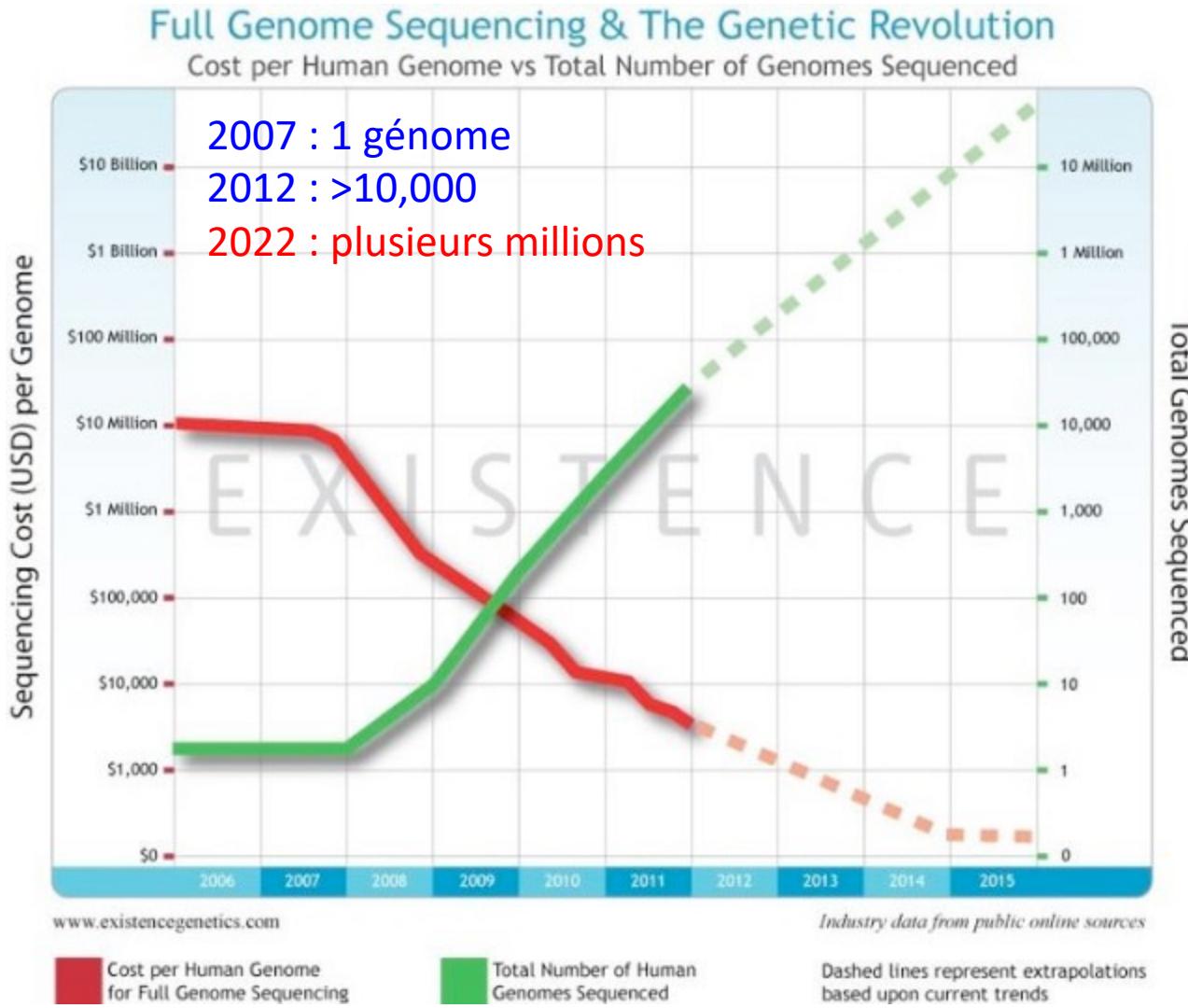
Sanger

454

Genome sequenced (publication year)	HGP (2003)	Venter (2007)	Watson (2008)
Time taken (start to finish)	13 years	4 years	4.5 months
Number of scientists listed as authors	> 2,800	31	27
Cost of sequencing (start to finish)	\$2.7 billion	\$100 million	< \$1.5 million
Coverage	8-10 ×	7.5 ×	7.4 ×
Number of institutes involved	16	5	2
Number of countries involved	6	3	1



# Séquençage de génomes personnels



# Le génome humain

- 3,3 milliards de paires de bases
- 23 paires de chromosomes
- > 20 000 gènes et ~6 isoformes par gène
- ~1-2 % du génome ⇔ exome (régions codantes)



1 génome  
2 jours  
800 euros  
150 Go

If you spoke  
one 'letter' of DNA  
per second

A, T, C, G, T, A, A, C, G, T...

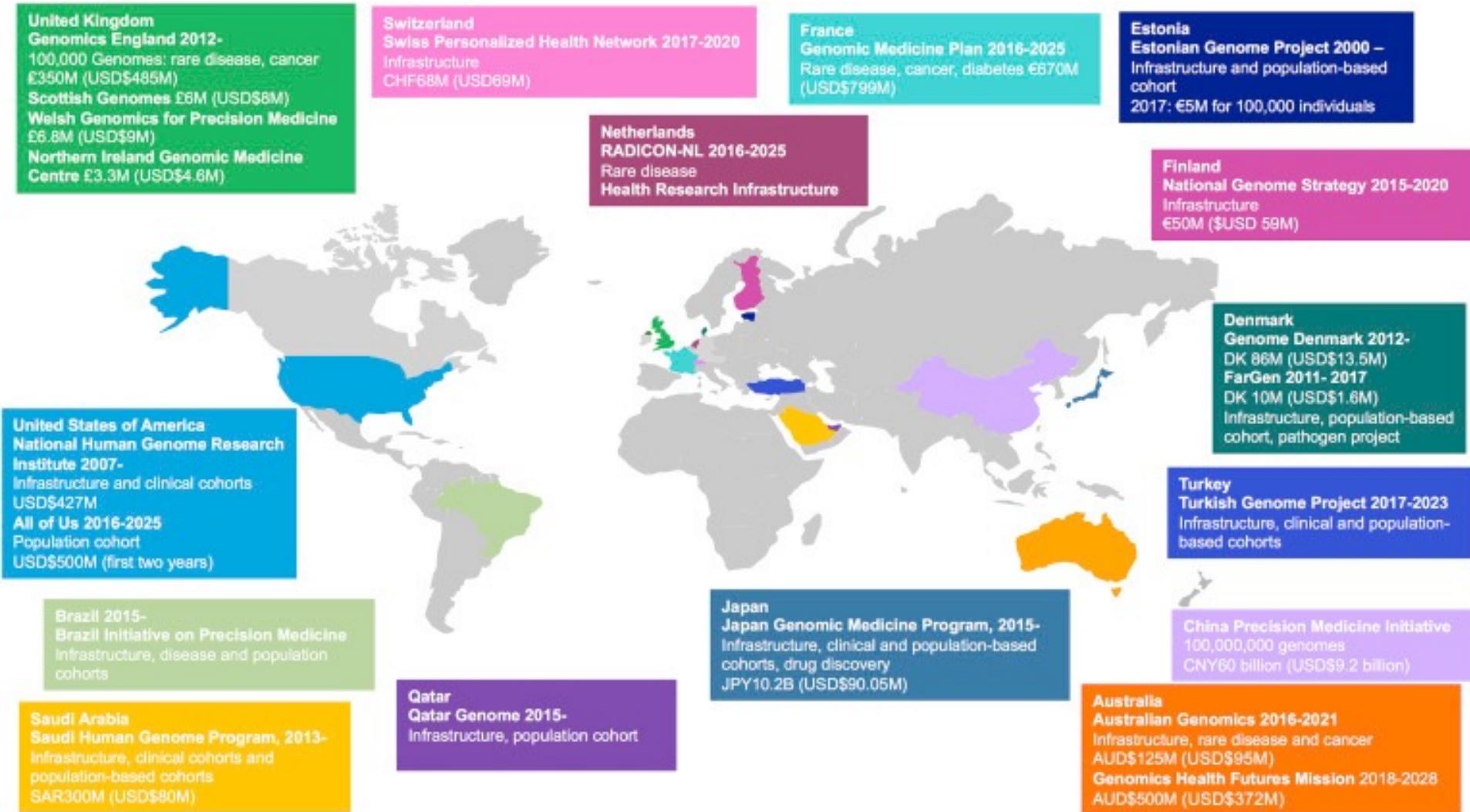
**24 hours a day,**

it would take about

**100 years** to recite the entire  
human genome

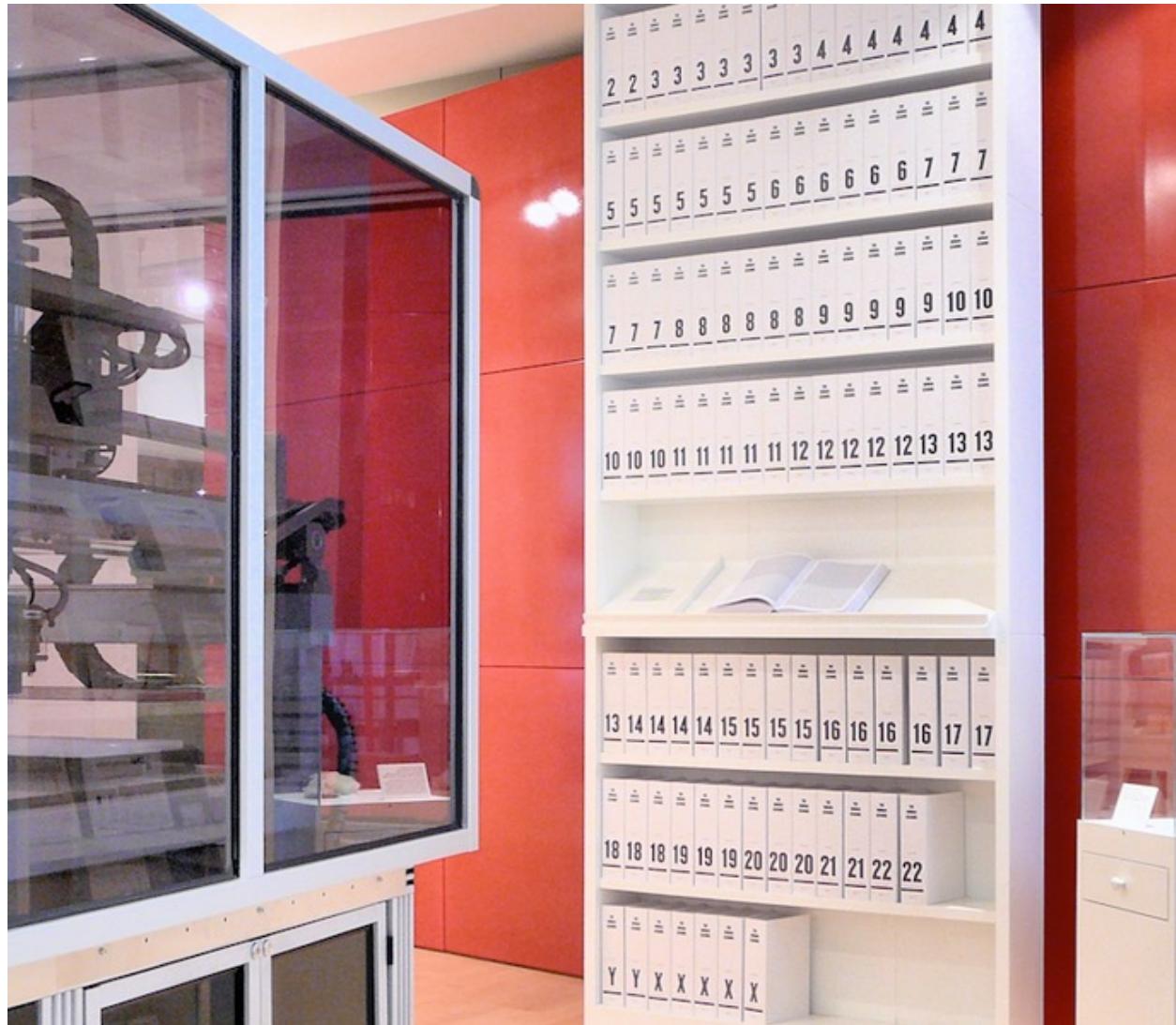


# Séquençage de génomes personnels

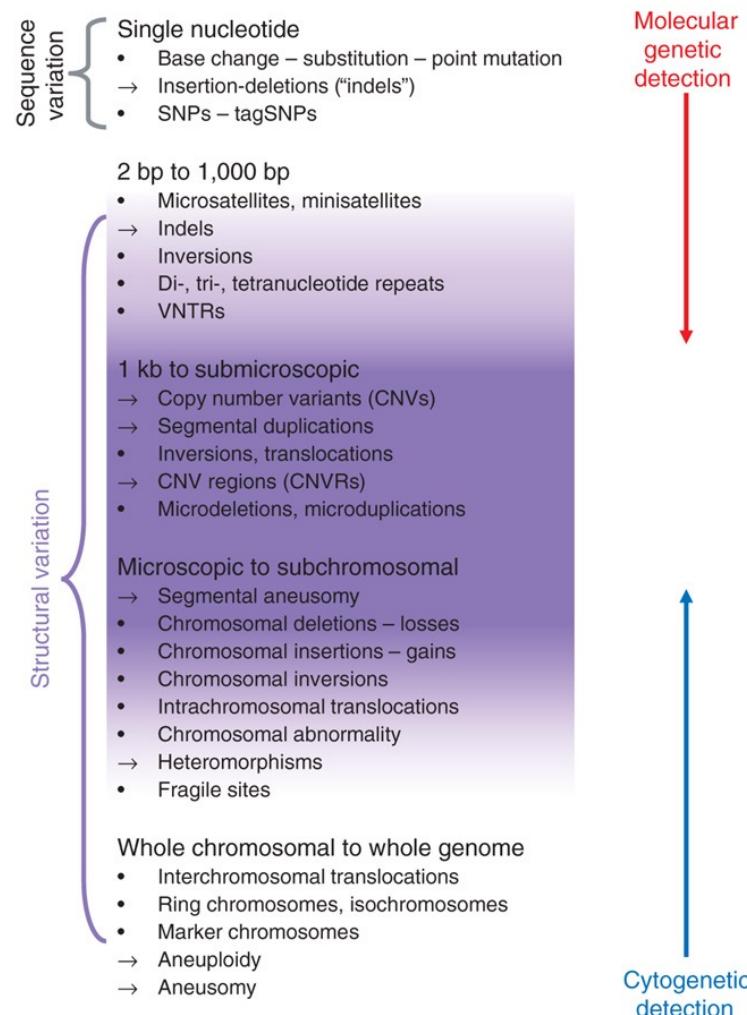


# Plan

- Introduction
- Variation du génome humain
- Méthodes d'identification de gènes
  - ✓ Maladies monogéniques
  - ✓ Maladies complexes



# Types de variations



# Variation du génome

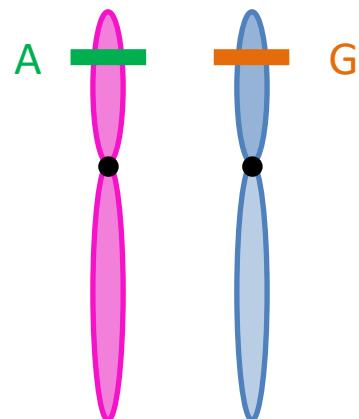
- Deux génomes humains sont identiques à 99,9%
- Différences dans la séquence du génome = variations génétiques
- Chaque variation génétique est caractérisée par  $\geq 2$  allèles

# Allèle

# Version variable d'un même locus génétique

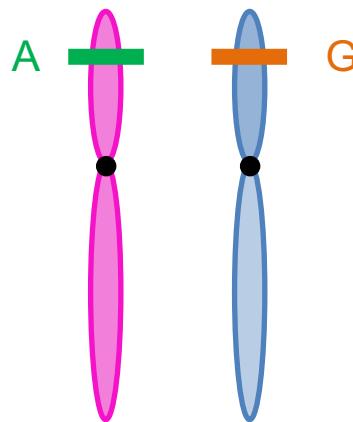
Exemple : à une position donnée, allèle A ou allèle G

CGATGCTA[A/G]GCATGCAT



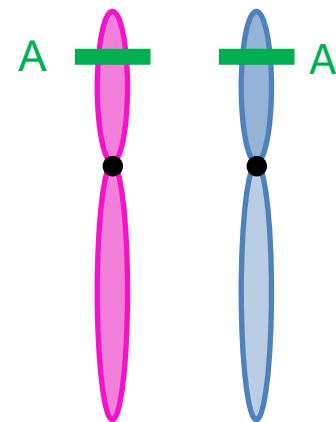
# Génotype

Composition allélique pour une variation génétique donnée



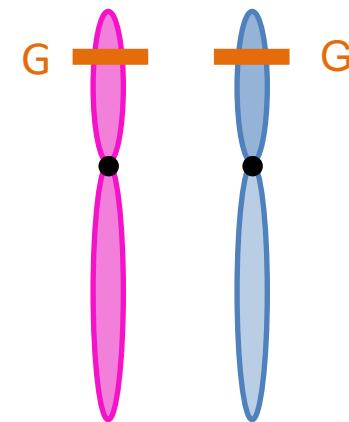
Génotype individu 1 : A/G

Si les 2 allèles sont différents,  
on parle d'**hétérozygotie**.



Génotype individu 2 : A/A

Si les 2 allèles sont identiques,  
on parle d'**homozygotie**.

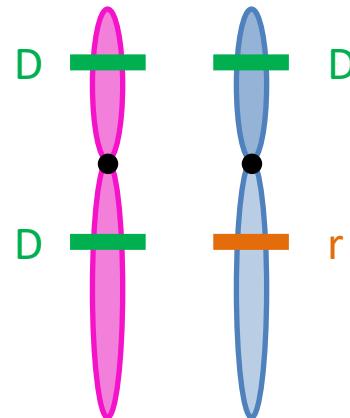


Génotype individu 3 : G/G

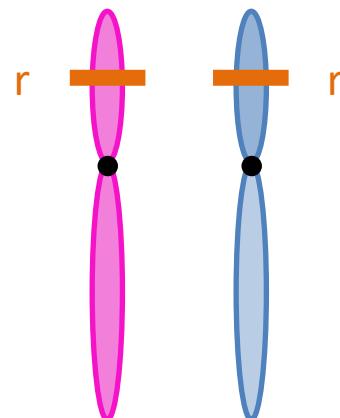
# Phénotype, dominance et récessivité

**Phénotype** : Trait d'un individu observable au niveau anatomique, morphologique, moléculaire, physiologique ou pathologique (dichotomique ou quantitatif).

**Allèle dominant** : Produit son effet qu'il soit à l'état hétérozygote ou homozygote.



**Allèle récessif** : Doit être présent à l'état homozygote pour s'exprimer.



# Variation du génome

- Deux génomes humains sont identiques à 99,9%
- Différences dans la séquence du génome = variations génétiques
- Chaque variation génétique est caractérisée par  $\geq 2$  allèles
- Chaque variation peut être fréquente, rare ou privée  
(notion de **fréquence**)

# Exemple de variant fréquent



Génome de référence : CGATGCTA**A**GCATGCAT

SNP  
Polymorphisme

Variant : CGATGCTA**G**GCATGCAT

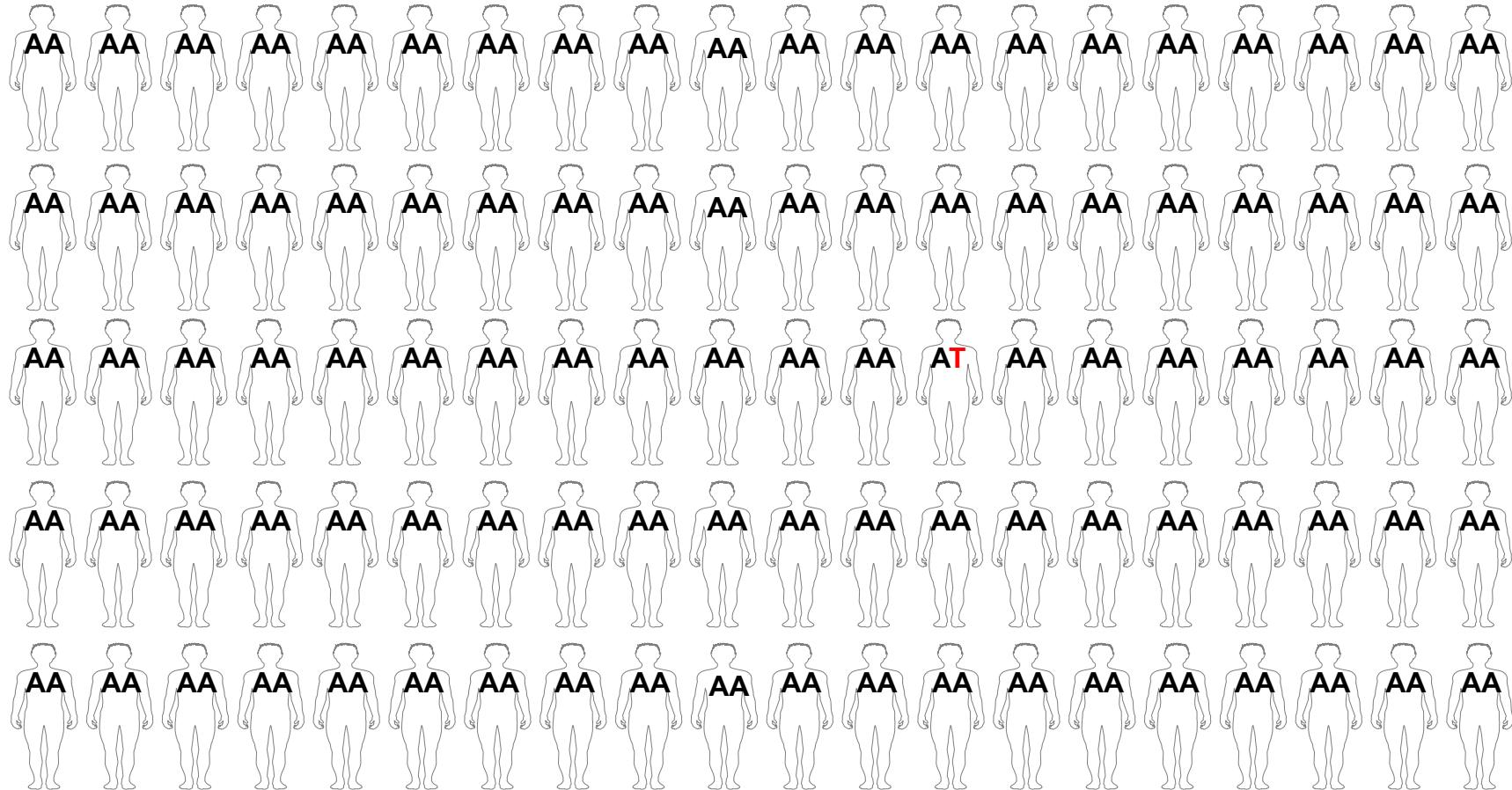


60 **G** / 200 allèles  
=30 %

# Exemple de variant rare (<1%)

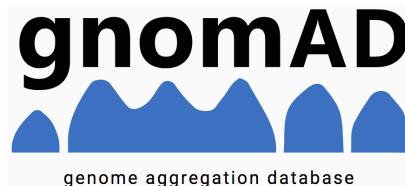
Génome de référence : CGATGCTA**A**GCATGC**A**T

Variant : CGATGCTA**G**GCATGC**T**T



1 **T** / 200 allèles  
=0,5 %

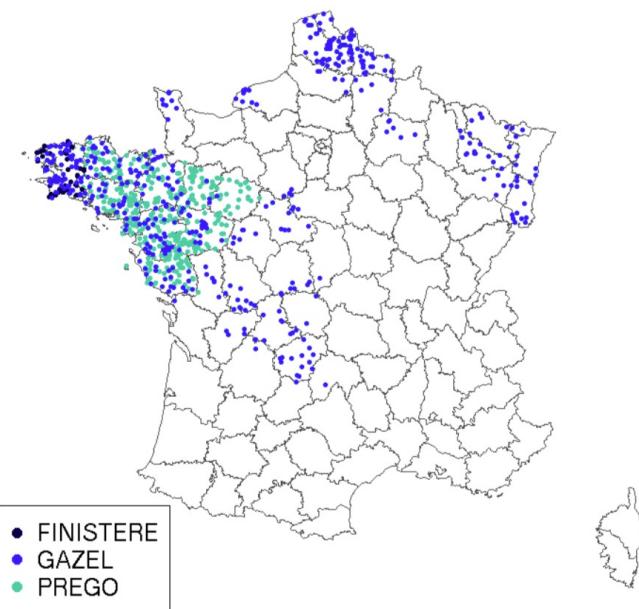
# Fréquence des variants



Exomes : 125 000 individus  
Génomes : 71 000 individus



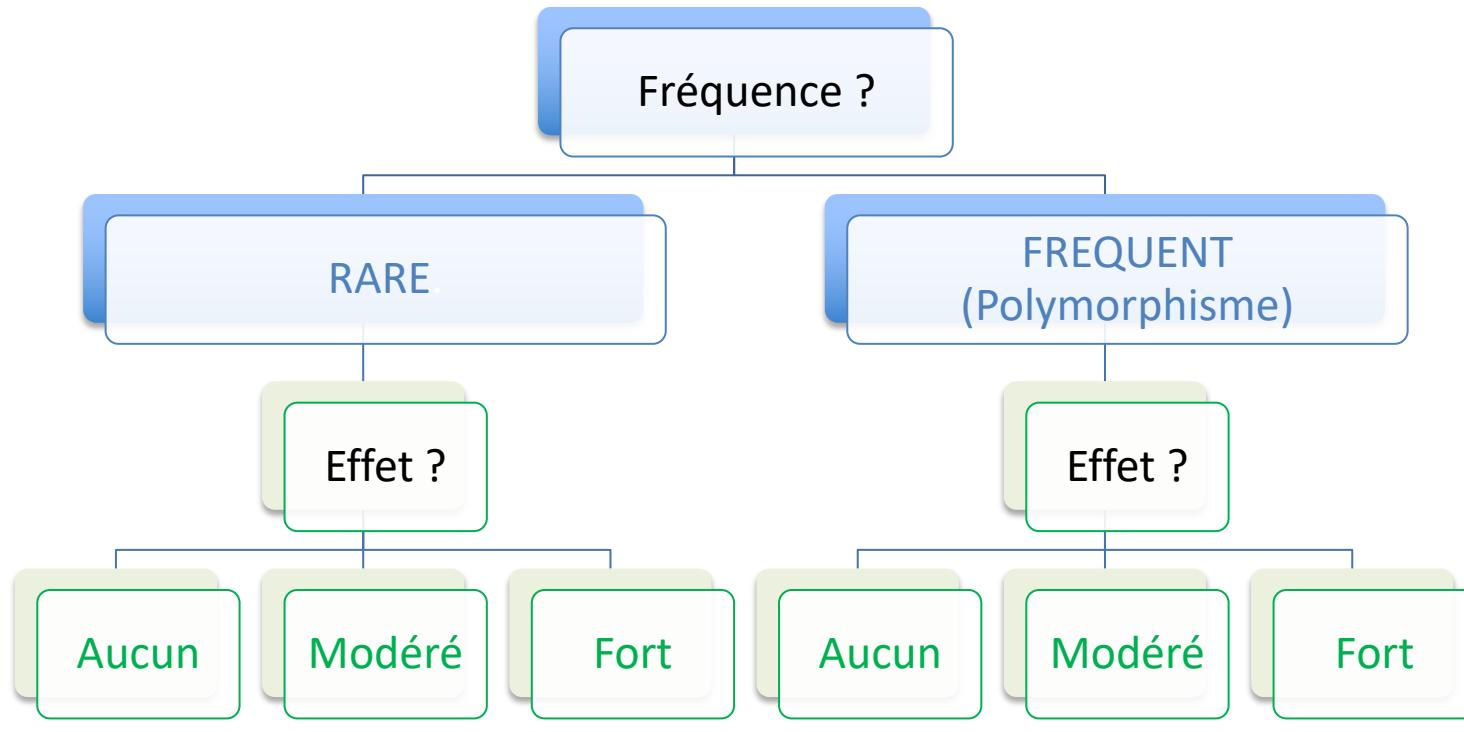
Génomes : 1 000 individus



# Variation du génome

- Deux génomes humains sont identiques à 99,9%
- Différences dans la séquence du génome = variations génétiques
- Chaque variation génétique est caractérisée par  $\geq 2$  allèles
- Chaque variation peut être fréquente, rare ou privée (notion de fréquence)
- Chaque variation, selon sa localisation, peut être bénigne, contribuer à des traits phénotypiques, prédisposer aux maladies, ou être responsable de maladies (notion **d'effet**)

# Effet des variants génétiques



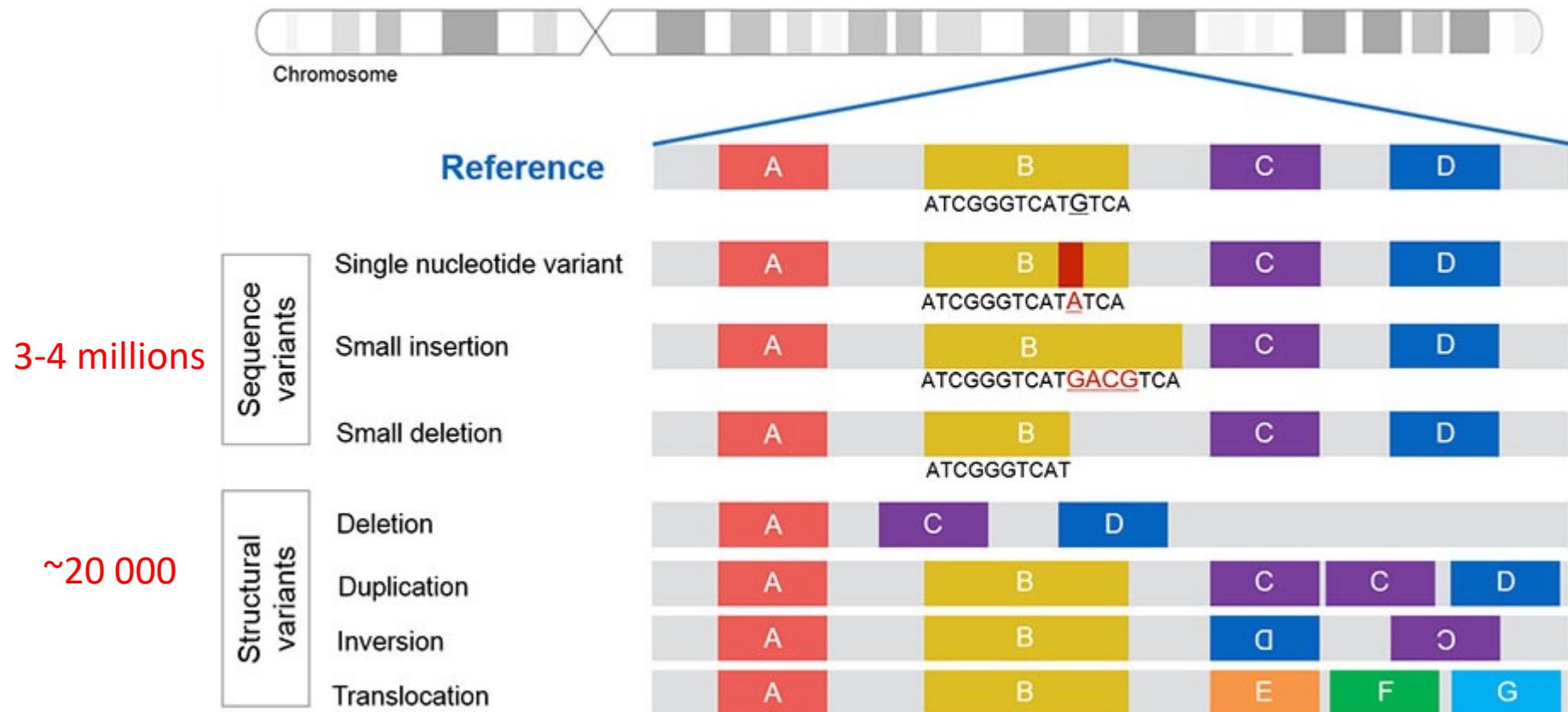
Approche familiale

Exemple :  
*CFTR, BRCA1, HTT, KCNQ1...*

Approche population

↑ risque  
Exemple :  
Diabète, Alzheimer...

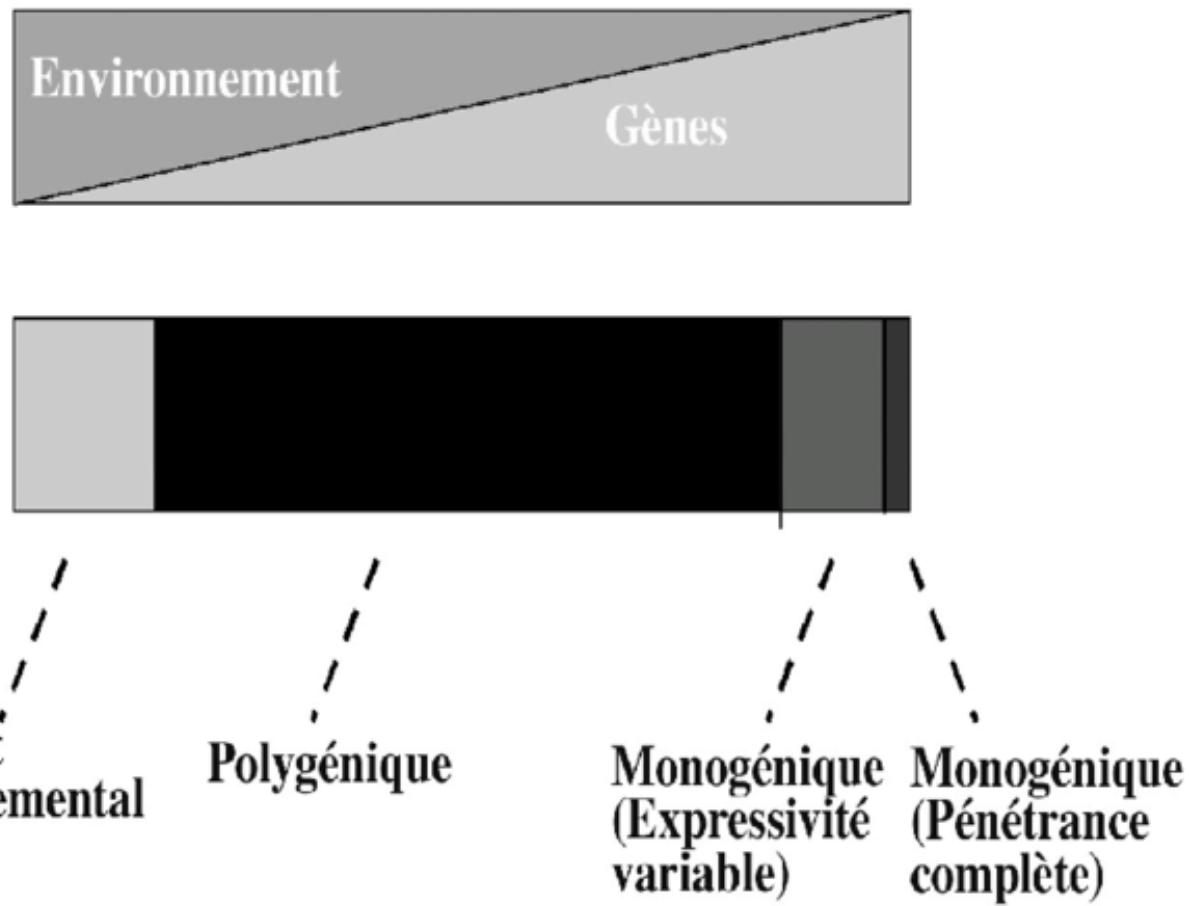
# Variabilité du génome



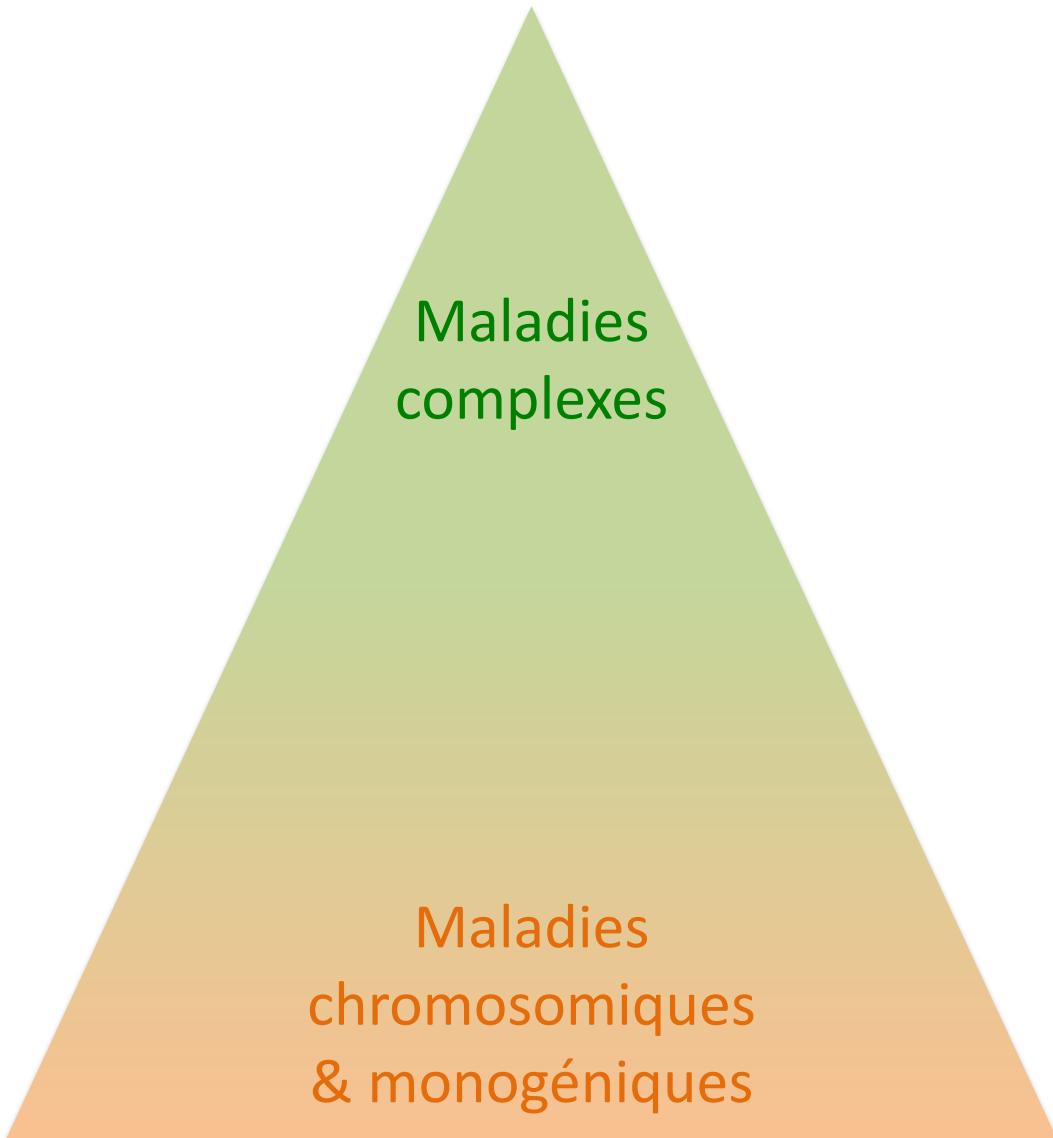
# Plan

- Introduction
- Variation du génome humain
- Méthodes d'identification de gènes
  - ✓ Maladies monogéniques
  - ✓ Maladies complexes

# Maladies humaines



# Génétique moléculaire



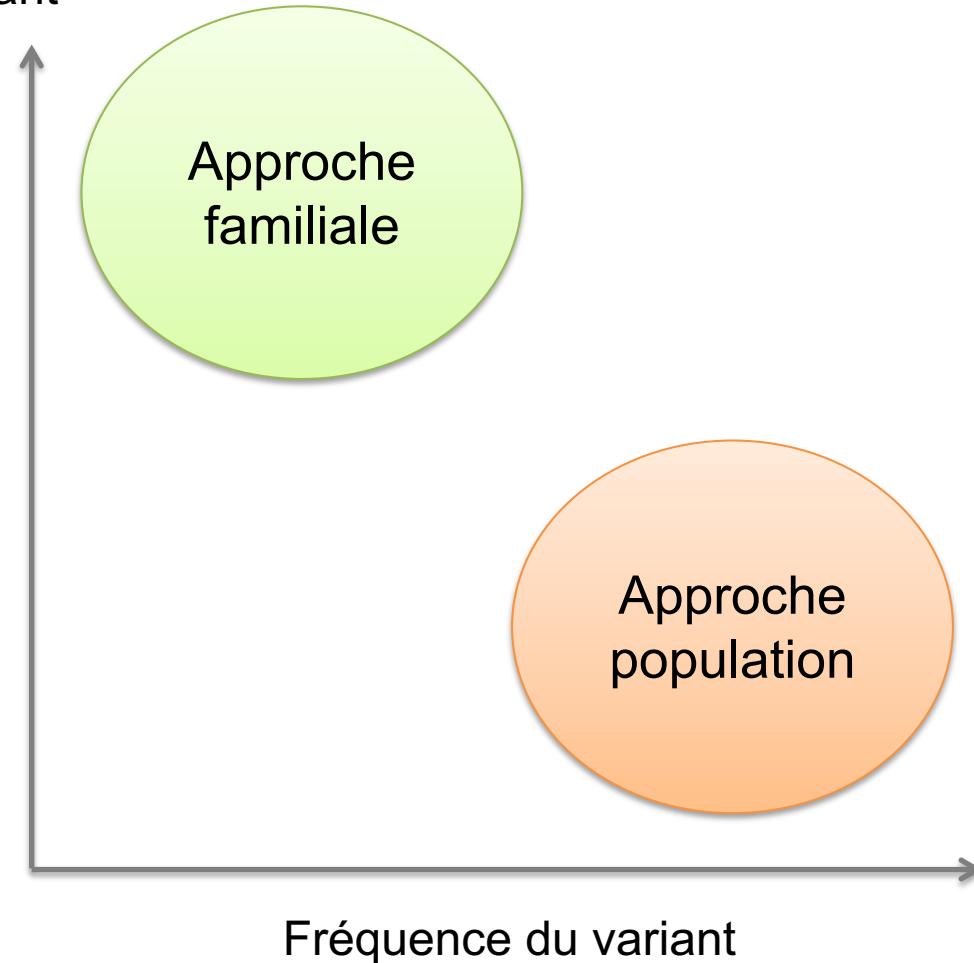
- ⇒ Gènes de susceptibilité
- ⇒ Nouvelles connaissances physiopathologiques
- ⇒ Nouveaux traitements
- ⇒ Médecine personnalisée

Implications éthiques, légales et sociales

Ex : 23andMe

# Stratégies génétiques

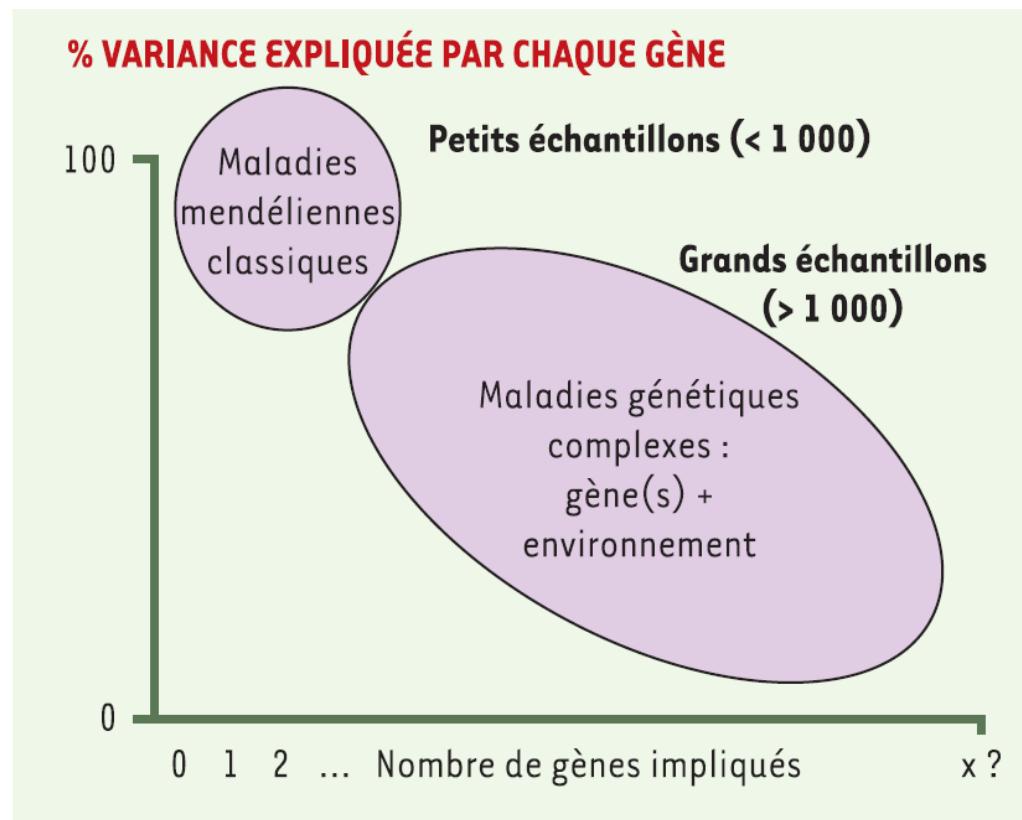
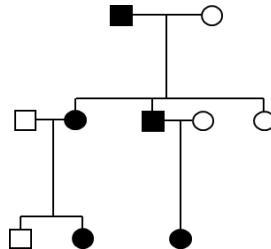
Effet du variant



# Plan

- Introduction
- Variation du génome humain
- Méthodes d'identification de gènes
  - ✓ Maladies monogéniques
  - ✓ Maladies complexes

- Un gène
- Mutations rares avec un effet fort
- Agrégation familiale
- Age d'apparition souvent précoce



# Les modes de transmission: autosomique dominant

Un des 2 parents est atteint

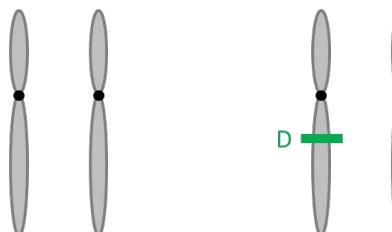
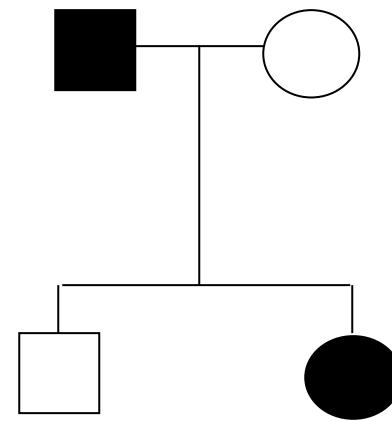
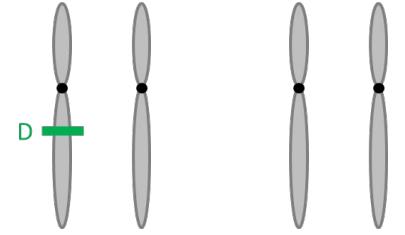
⇒ Un enfant sur 2 est atteint

Mais parfois dans une famille, on observe:

-la mutation chez un individu sain (**« non pénétrance »**)

-la mutation est absente chez un individu atteint (**« phénomécopie »**)

-une variabilité dans les phénotypes observés: **expressivité variable** (fonction d'autres gènes, de l'environnement?!)



# Les modes de transmission: autosomique récessif

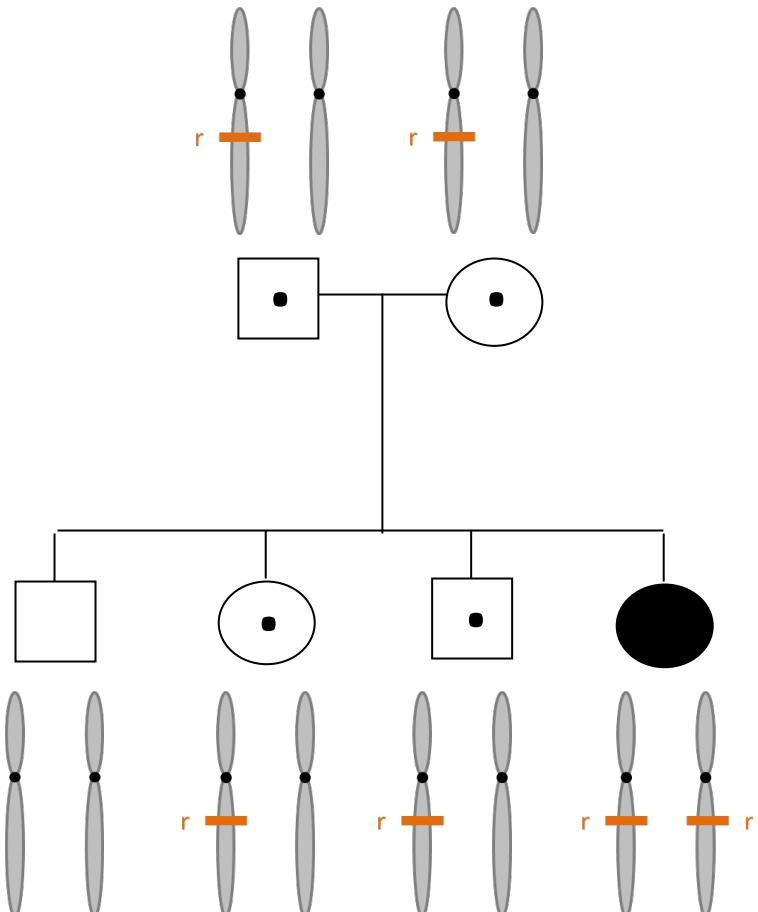
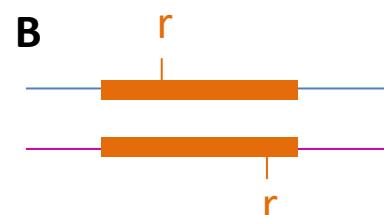
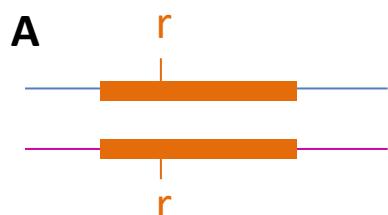
Les parents ne sont pas atteints

⇒ Un enfant sur 4 est atteint

⇒ Un enfant sur 2 est porteur sain

-si les variants paternels/maternels sont identiques, il y a donc homozygotie chez l'enfant atteint (**A**). Risque  $\nearrow$  avec consanguinité.

-si les 2 variants sont distincts, mais situés dans un même gène, alors l'enfant atteint est **hétérozygote composite** pour le gène (**B**), ce qui a le même effet qu'un variant récessif homozygote (les 2 copies du gène sont altérées)



# Modes de transmission

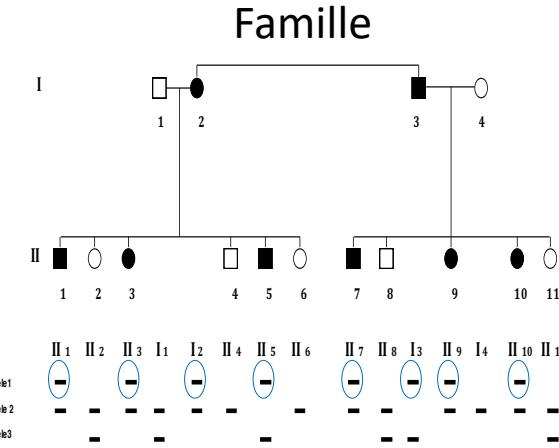
Mode de transmission	Localisation du locus responsable du trait	Principales caractéristiques	Exemples de phénotypes
Autosomique dominant	Chromosome 1-22	<ul style="list-style-type: none"> <li>-hommes et femmes sont atteints avec une fréquence équivalente</li> <li>-les enfants atteints ont un parent atteint</li> </ul>	Huntington, syndrome de Marfan, arythmies, cardiomyopathies
Autosomique récessif	Chromosome 1-22	<ul style="list-style-type: none"> <li>-hommes et femmes sont atteints avec une fréquence équivalente</li> <li>-les enfants atteints n'ont pas de parent atteint</li> <li>-peut « sauter » une ou plusieurs générations</li> </ul>	Mucoviscidose, drépanocytose
Dominant lié à l'X	Chromosome X	<ul style="list-style-type: none"> <li>-les femmes atteintes ont 50% d'enfants atteints</li> <li>-les hommes sont plus sévèrement atteints</li> <li>-toutes les filles de pères atteints sont atteintes</li> </ul>	Syndrome de Rett
Récessif lié à l'X	Chromosome X	<ul style="list-style-type: none"> <li>-beaucoup plus d'hommes atteints que de femmes</li> <li>-les hommes atteints ont une mère saine (porteuse)</li> <li>-le trait n'est jamais transmis de père à fils</li> <li>-peut « sauter » une ou plusieurs générations</li> </ul>	Hémophilie
Lié à l'Y	Chromosome Y	<ul style="list-style-type: none"> <li>-tous les hommes atteints passent le trait à leur fils</li> <li>-les femmes ne sont pas atteintes</li> </ul>	Stérilité
Mitochondrial	Pas chromosomique ADN mitochondrial	<ul style="list-style-type: none"> <li>-tous les enfants d'une femme atteinte sont atteints</li> <li>-les hommes atteints ne transmettent pas le trait</li> </ul>	Neuropathie optique de Leber

# Principe de la génétique inverse: du malade au gène et du gène au malade

Malade  
(cas index)



*Enquête  
familiale*



*Génotypage  
Analyse de liaison*



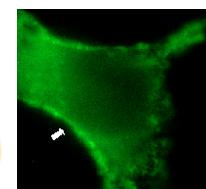
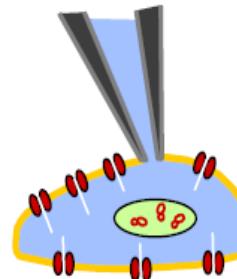
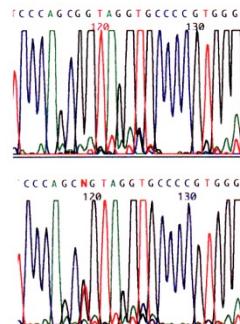
**Locus**



*Séquençage*

**Gène/Mutation**

**Fonction**



Malade  
Diagnostic  
moléculaire  
Prise en charge  
clinique

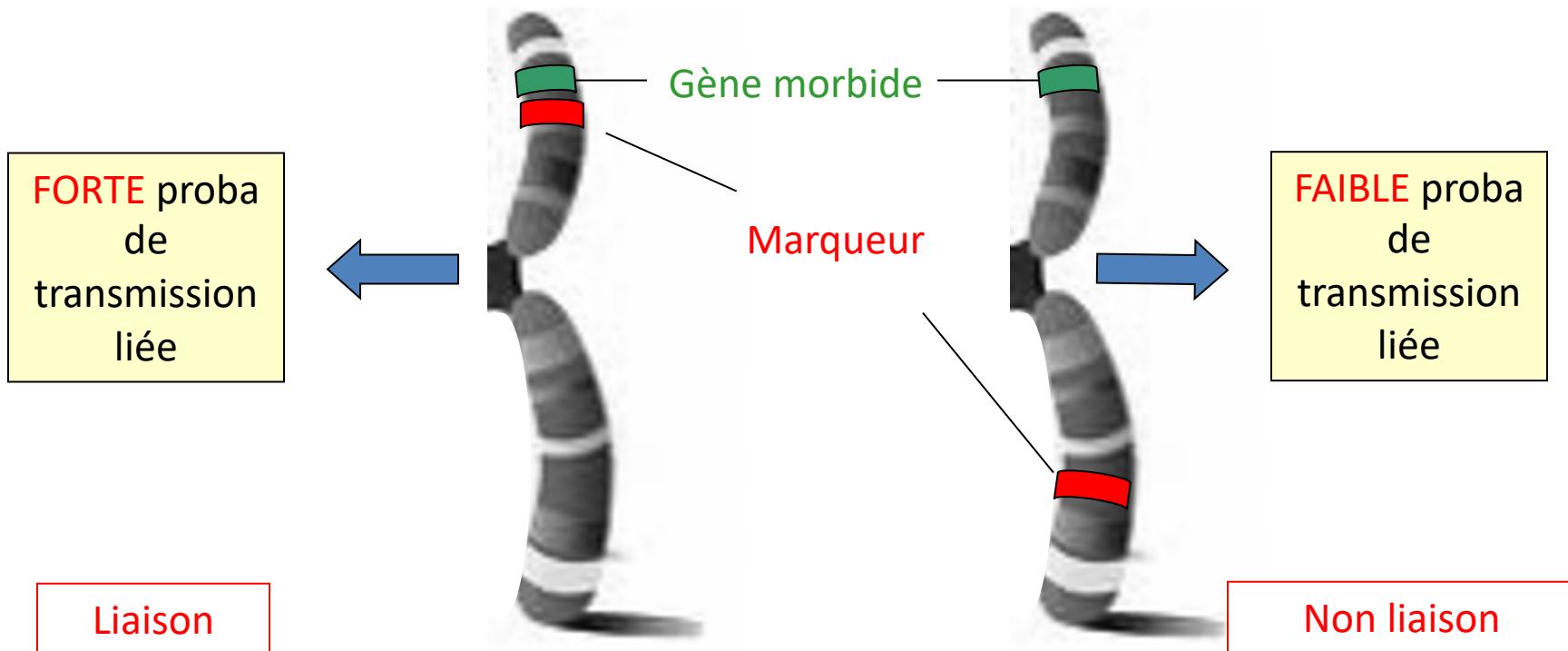
## Analyse de liaison

Corréler la transmission d'un génotype avec  
un phénotype

# L'analyse de liaison génétique

= clonage positionnel

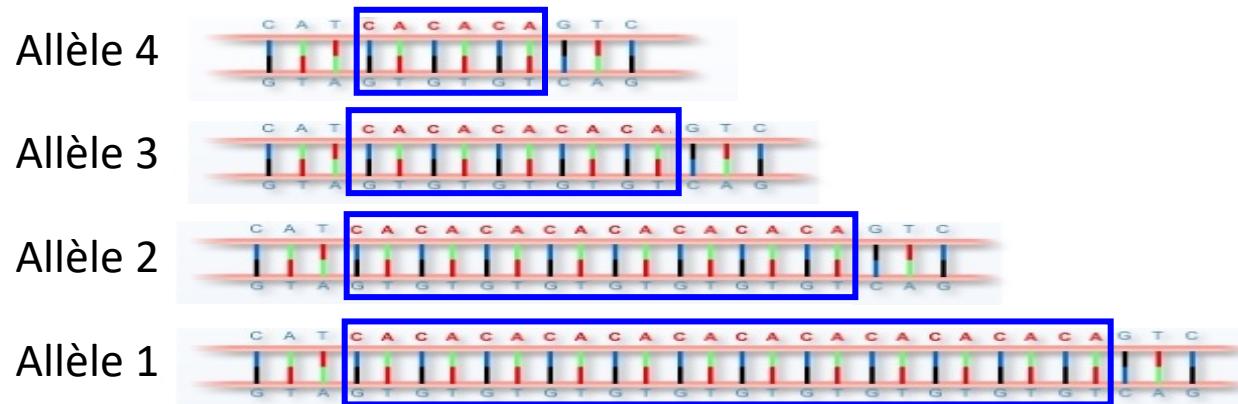
= « linkage »



# Deux types de marqueurs : microsatellites ou SNPs

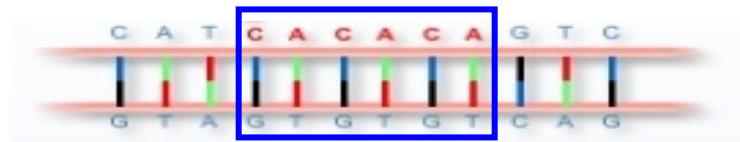
## Microsatellites

- Répétition de motifs courts (2-4 pb): (CA)n ; (CAT)n ; (GATA)n
- Nombre de répétitions variable selon les individus
- Uniformément distribués sur tout le génome → 1 toutes les 30 kb
- Très polymorphes (beaucoup d'allèles)

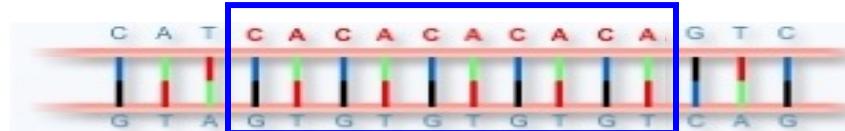


# Génotypage de marqueurs

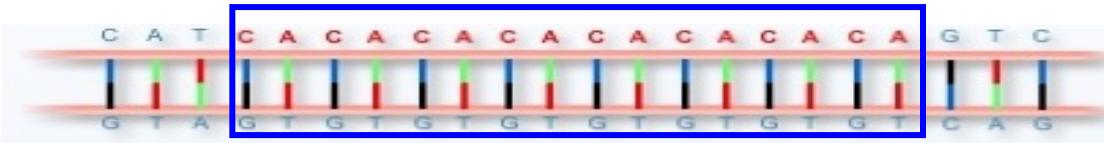
Allèle 4



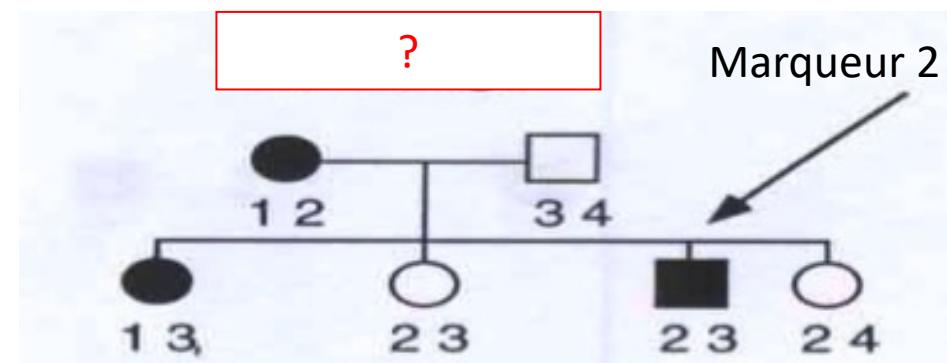
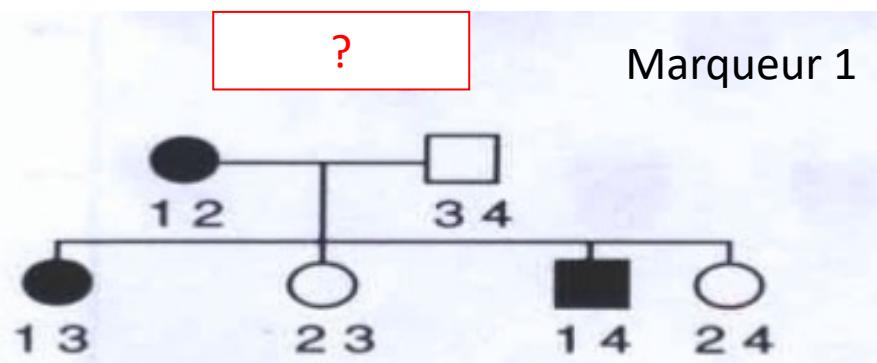
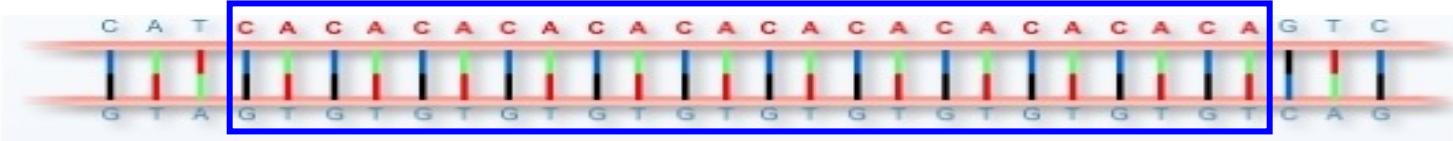
Allèle 3

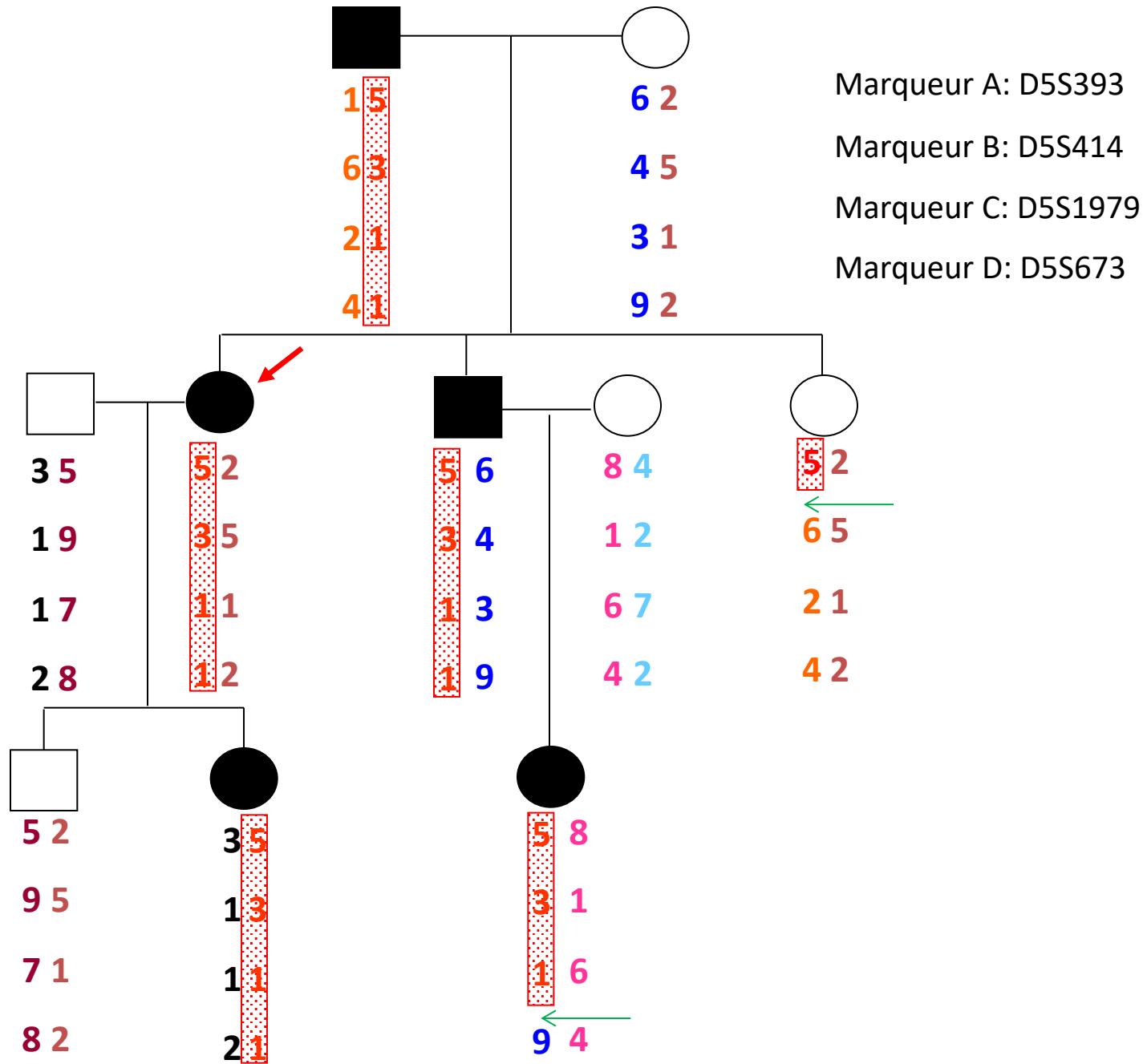


Allèle 2



Allèle 1

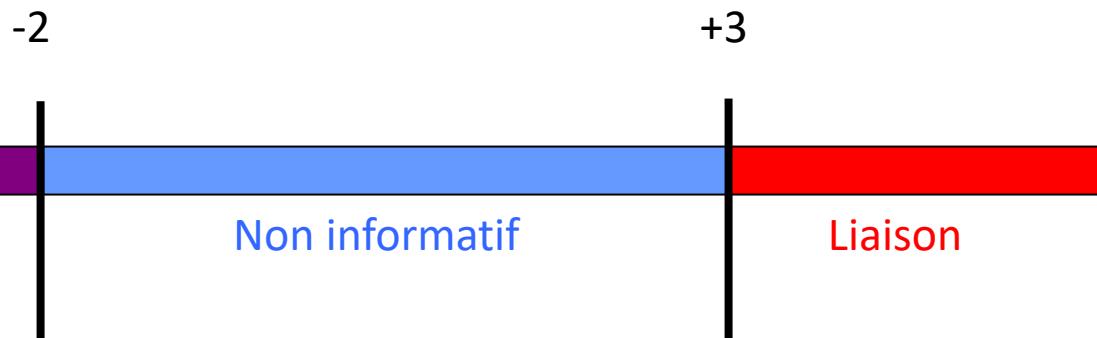




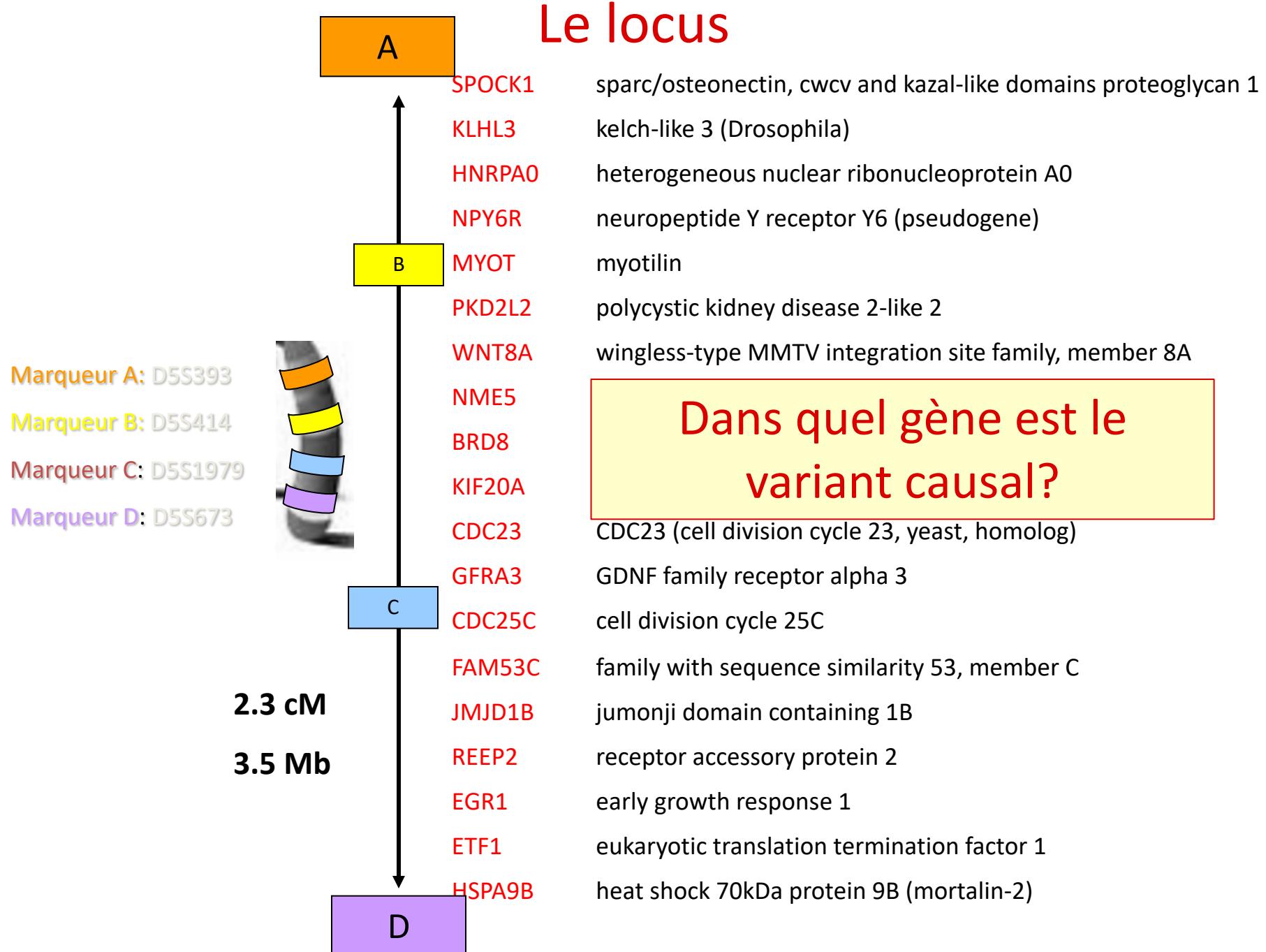
# L'analyse de liaison génétique

## *La méthode des lod scores*

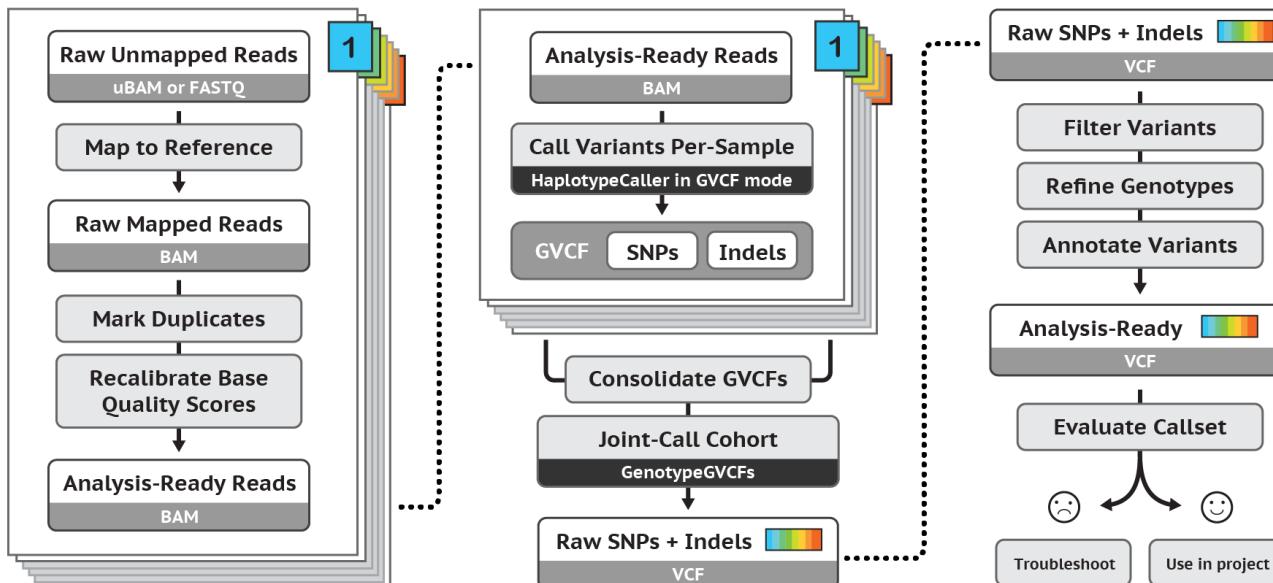
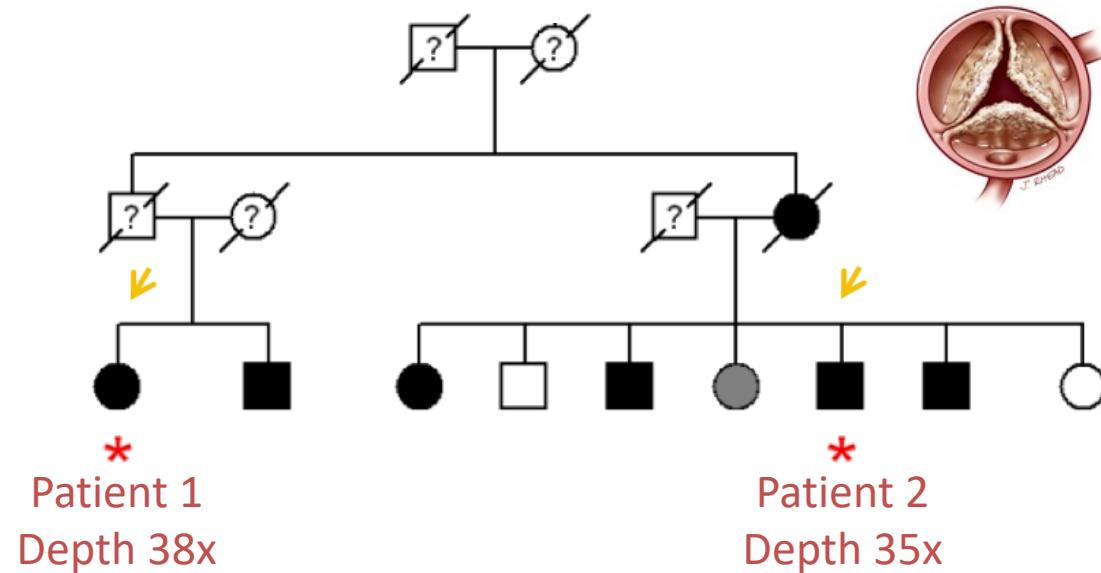
$$Z = \log_{10} \frac{\text{Probabilité de liaison } (\theta < 0.5)}{\text{Probabilité de non liaison } (\theta = 0.5)}$$



# Le locus

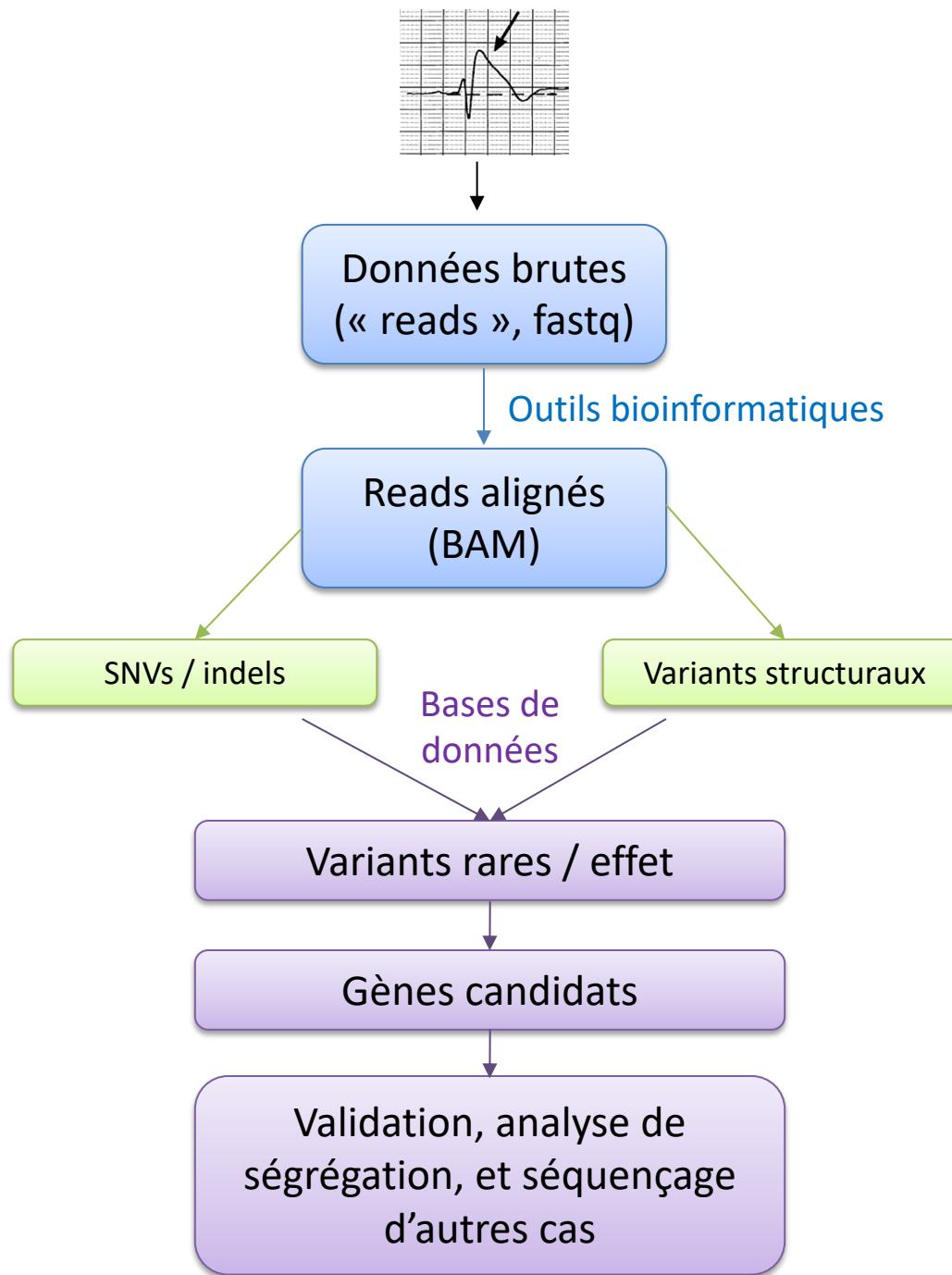


# Analyse des données de séquençage

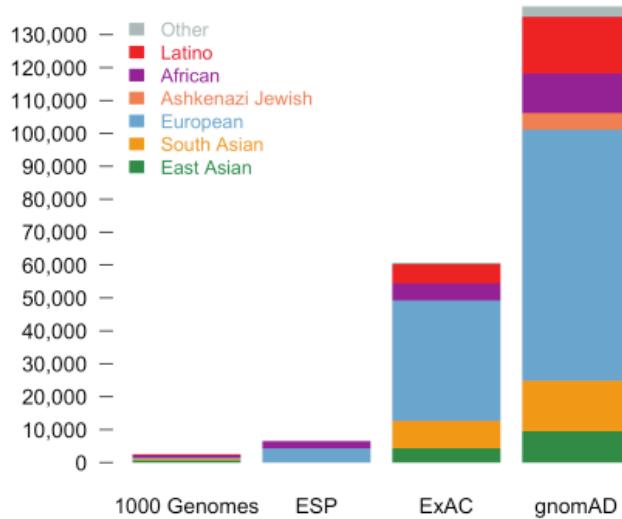


Analyses  
bioinformatiques

GATK Best Practices



# Fréquence des variants en population



gnomAD v3 : 71,702 genomes

POPULATION	DESCRIPTION	GENOMES	EXOMES	TOTAL
AFR	African/African American	4,368	7,652	12,020
AMR	Admixed American	419	16,791	17,210
ASJ	Ashkenazi Jewish	151	4,925	5,076
EAS	East Asian	811	8,624	9,435
FIN	Finnish	1,747	11,150	12,897
NFE	Non-Finnish European	7,509	55,860	63,369
SAS	South Asian	0	15,391	15,391
OTH	Other (population not assigned)	491	2,743	3,234
<b>Total</b>		15,496	123,136	138,632

# Variants rares

## Variant: 3:38591847 G / C

Filter Status	PASS
dbSNP	<a href="#">rs45489199</a>
Allele Frequency	0.001341
Allele Count	146 / 108894
UCSC	<a href="#">3-38591847-G-C</a>
ClinVar	<a href="#">Click to search for variant in Clinvar</a>

### Genotype Quality Metrics

### Site Quality Metrics

## Annotations

This variant falls on 11 transcripts in 1 genes:

missense

- [SCN5A](#)

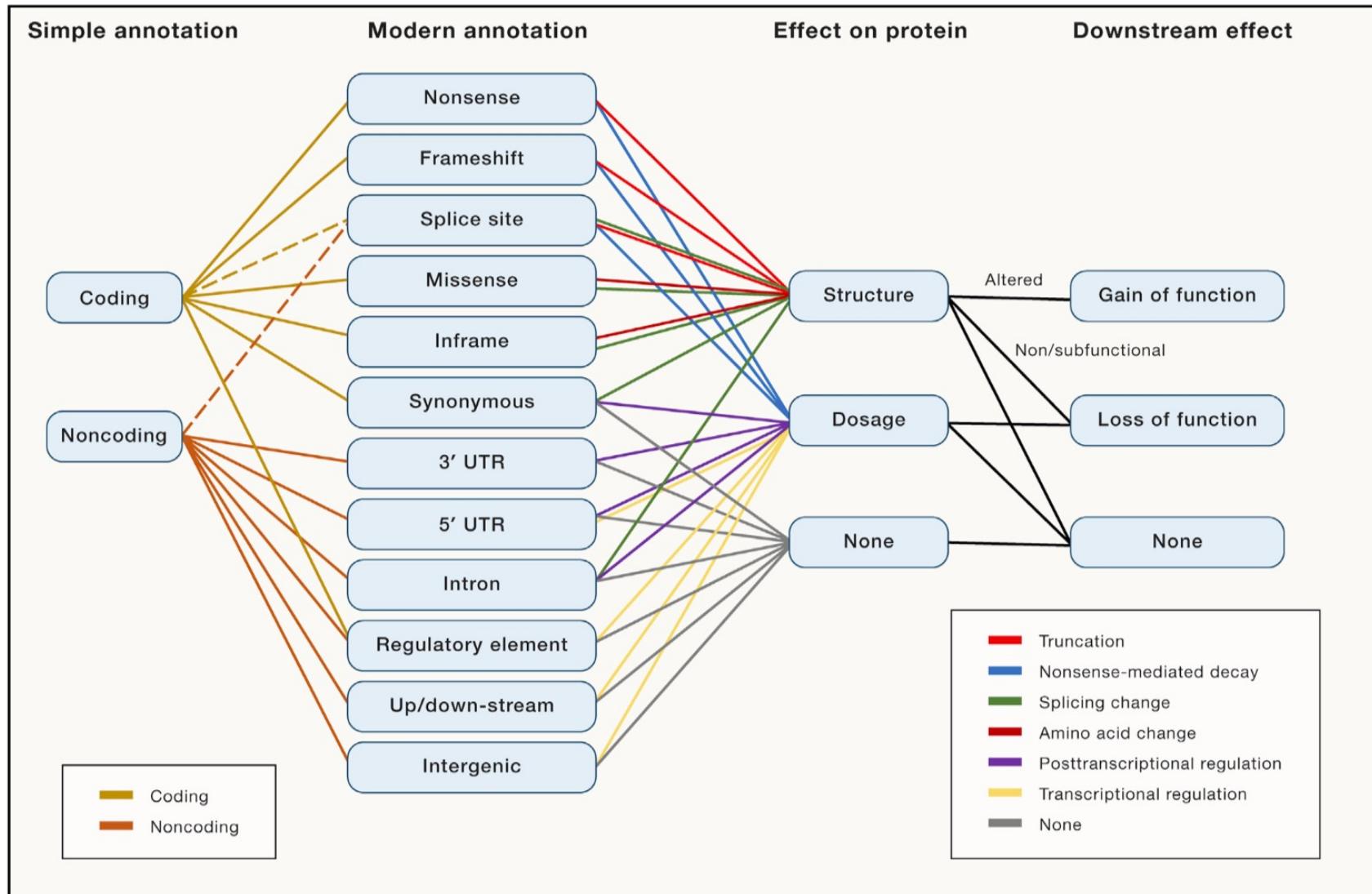
[Transcripts](#) ▾

**Note:** This list may not include additional transcripts in the same gene that the variant does not overlap.

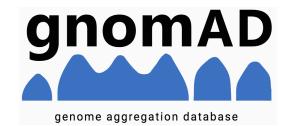
## Population Frequencies

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (Finnish)	15	6614	0	0.002268
European (Non-Finnish)	124	63230	1	0.001961
Other	1	780	0	0.001282
African	4	9314	0	0.0004295
South Asian	1	9006	0	0.000111
Latino	1	11476	0	8.714e-05
East Asian	0	8474	0	0
<b>Total</b>	<b>146</b>	<b>108894</b>	<b>1</b>	<b>0.001341</b>

# Déterminer l'effet des variations : un challenge



# Interprétation des données de séquençage



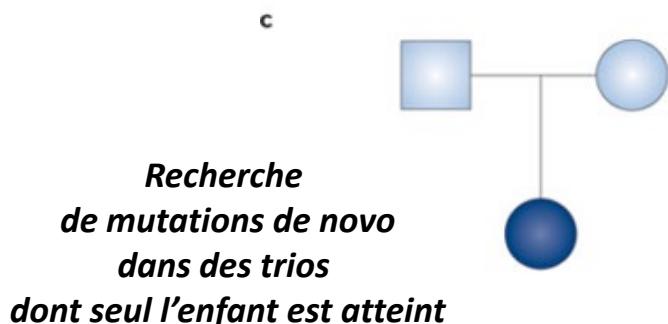
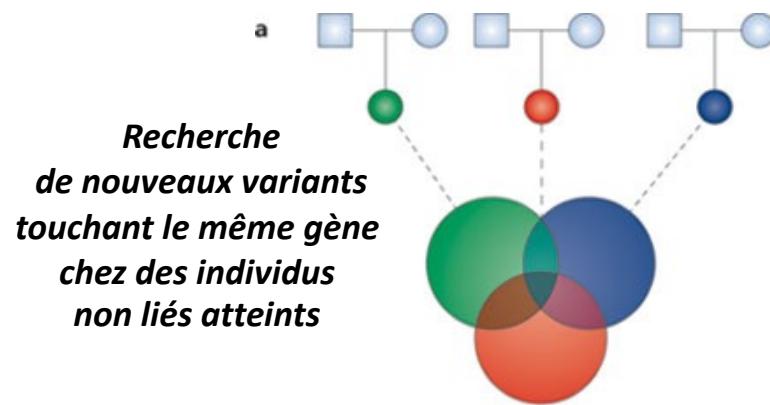
	Patient 1	Patient 2
# total substitutions + indels	4 978 101	4 867 920
# variants rares (<0,1% Non-Finnish Europeans + contrôles internes)	58 468	52 972
# variants rares + csq (exons / épissage)	450	438
SO:0001818		
SO:0001568		
# variants rares + faux-sens	276	266
SO:0001583		
# variants rares + stop	8	12
SO:0001587		
# variants rares + frameshift	44	38
SO:0001589		
# variants rares + csq partagés	55 (dont 8 stop/frameshift)	



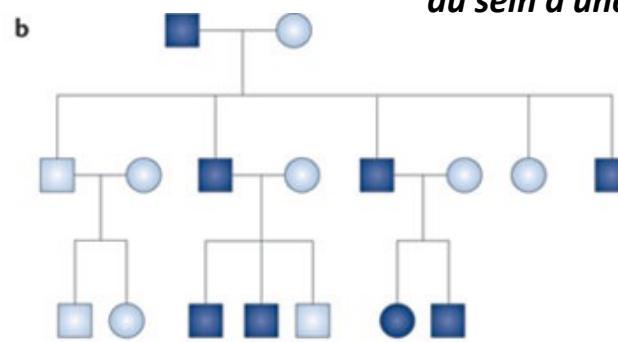
« Grands » gènes (exemple de la *TTN* : un variant rare faux-sens chez un individu sur 2)

**Take-home message :** la plupart des variants rares sont bénins

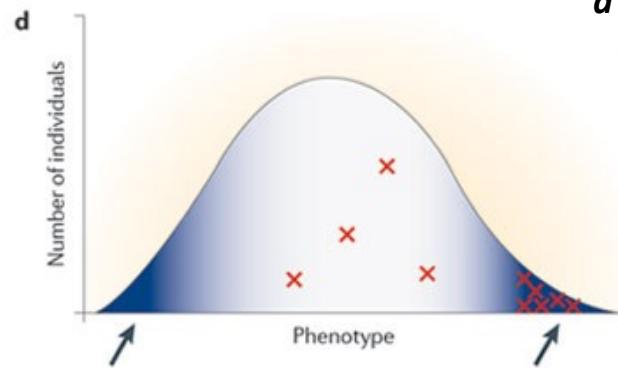
# Découverte par séquençage de variants rares causant des maladies



*Recherche de nouveaux variants partagés entre individus atteints au sein d'une même famille*



*Recherche d'une concentration de variants rares aux extrêmes d'un phénotype quantitatif*



Nature Reviews | Genetics

# 2009 : Preuve de concept

Freeman–Sheldon syndrome

doi:10.1038/nature08250

nature

## Targeted capture and massively parallel sequencing of 12 human exomes

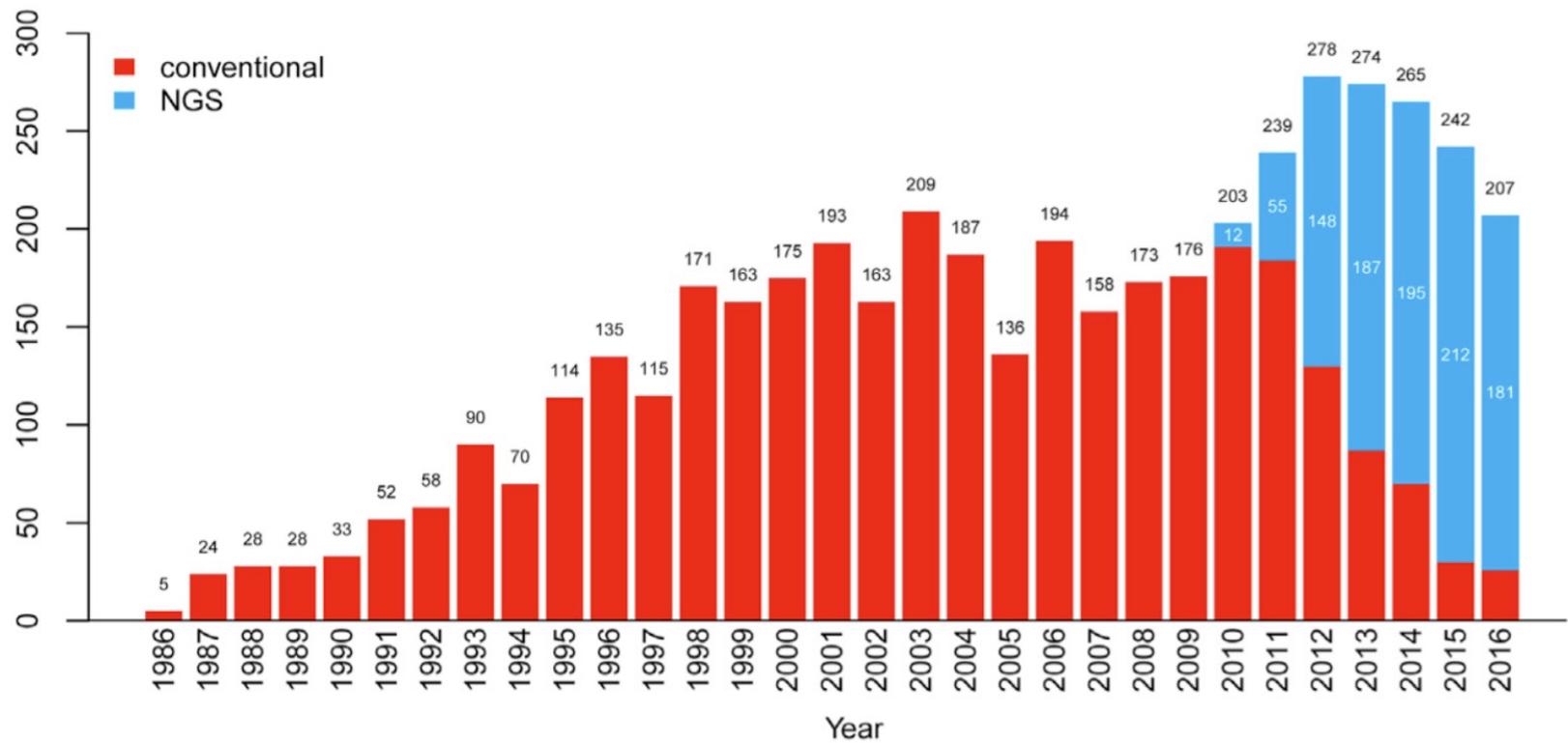
Sarah B. Ng<sup>1</sup>, Emily H. Turner<sup>1</sup>, Peggy D. Robertson<sup>1</sup>, Steven D. Flygare<sup>1</sup>, Abigail W. Bigham<sup>2</sup>, Choli Lee<sup>1</sup>,  
Tristan Shaffer<sup>1</sup>, Michelle Wong<sup>1</sup>, Arindam Bhattacharjee<sup>4</sup>, Evan E. Eichler<sup>1,3</sup>, Michael Bamshad<sup>2</sup>,  
Deborah A. Nickerson<sup>1</sup> & Jay Shendure<sup>1</sup>

Number of genes in which each affected has at least one...	Non-synonymous cSNP, splice site variant or coding indel (NS/SS/I)	Any 3 of 4					
		FSS24895	FSS10208	FSS24895	FSS10208	FSS10066	FSS22194
	4,510	3,284	2,765	2,479			3,768
	513	128	71	53			119
	799	168	53	21			160
	360	38	8	1 (MYH3)			22
	160	10	2	1 (MYH3)			3
...And predicted to be damaging							

**Table 2. Mendelian disease gene identifications by exome or genome sequencing**

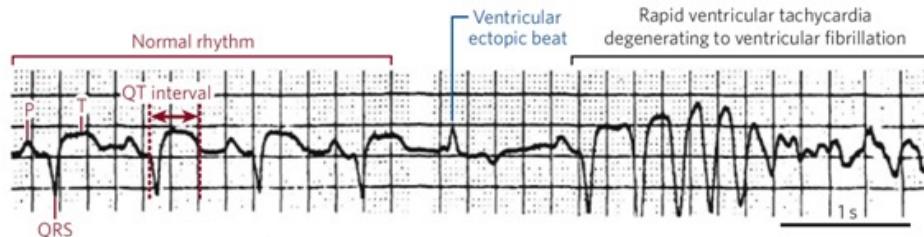
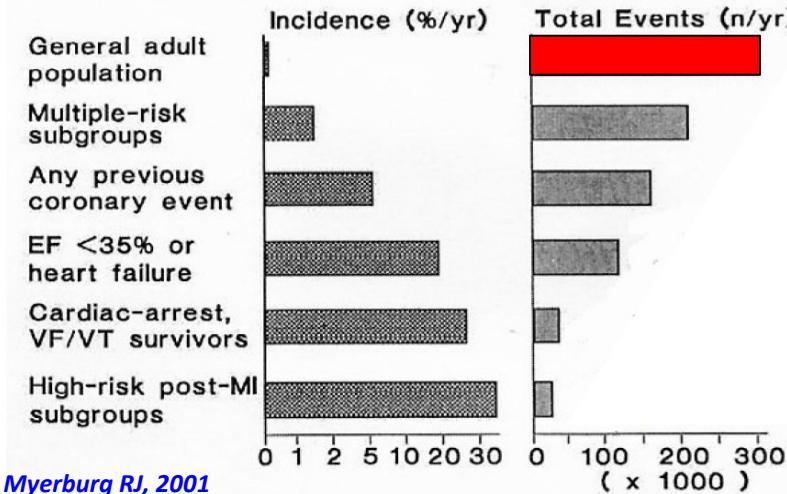
Disorder	Inheritance	Gene Identified	Scope	References
Congenital chloride diarrhea	Recessive	<i>SLC26A3</i>	Exome	Choi <i>et al.</i> [16]
Miller syndrome	Recessive	<i>DHODH</i>	Exome	Ng <i>et al.</i> [14]
Charcot-Marie-Tooth neuropathy	Recessive	<i>SH3TC2</i>	Genome	Lupski <i>et al.</i> [20]
Metachondromatosis	Dominant	<i>PTPN11</i>	Genome	Sobreira <i>et al.</i> [23]
Schinzel-Giedion syndrome	Dominant	<i>SETBP1</i>	Exome	Hoischen <i>et al.</i> [29]
Nonsyndromic hearing loss	Recessive	<i>GPSM2</i>	Exome	Walsh <i>et al.</i> [69]
Perrault syndrome	Recessive	<i>HSD17B4</i>	Exome	Pierce <i>et al.</i> [25]
Hyperphosphatasia mental retardation syndrome	Recessive	<i>PIGV</i>	Exome	Krawitz <i>et al.</i> [68]
Sensenbrenner syndrome	Recessive	<i>WDR35</i>	Exome	Gilissen <i>et al.</i> [26]
Cerebral cortical malformations	Recessive	<i>WDR62</i>	Exome	Bilguvar <i>et al.</i> [70]
Kaposi sarcoma	Recessive	<i>STIM1</i>	Exome	Byun <i>et al.</i> [71]
Spinocerebellar ataxia	Dominant	<i>TGM6</i>	Exome	Wang <i>et al.</i> [72]
Combined hypolipidemia	Recessive	<i>ANGPTL3</i>	Exome	Musunuru <i>et al.</i> [40]
Complex I deficiency	Recessive	<i>ACAD9</i>	Exome	Haack <i>et al.</i> [52]
Autoimmune lymphoproliferative syndrome	Recessive	<i>FADD</i>	Exome	Bolze <i>et al.</i> [73]
Amyotrophic lateral sclerosis	Dominant	<i>VCP</i>	Exome	Johnson <i>et al.</i> [74]
Nonsyndromic mental retardation	Dominant	Various	Exome	Vissers <i>et al.</i> [31]
Kabuki syndrome	Dominant	<i>MLL2</i>	Exome	Ng <i>et al.</i> [30]
Inflammatory bowel disease	Dominant	<i>XIAP</i>	Exome	Worthey <i>et al.</i> [18]
Nonsyndromic mental retardation	Recessive	<i>TECR</i>	Exome	Caliskan <i>et al.</i> [75]
Retinitis pigmentosa	Recessive	<i>DHDDS</i>	Exome	Züchner <i>et al.</i> [56]
Osteogenesis imperfecta	Recessive	<i>SERPINF1</i>	Exome	Becker <i>et al.</i> [53]
Dilated cardiomyopathy	Dominant	<i>BAG3</i>	Exome	Norton <i>et al.</i> [24]
Hajdu-Cheney syndrome	Dominant	<i>NOTCH2</i>	Exome	Simpson <i>et al.</i> [76]
Hajdu-Cheney syndrome	Dominant	<i>NOTCH2</i>	Exome	Isidor <i>et al.</i> [77]
Skeletal dysplasia	Recessive	<i>POPI</i>	Exome	Glazov <i>et al.</i> [78]
Amelogenesis	Recessive	<i>FAM20A</i>	Exome	O'Sullivan <i>et al.</i> [80]
Chondrodysplasia and abnormal joint development	Recessive	<i>IMPAD1</i>	Exome	Vissers <i>et al.</i> [80]
Progeroid syndrome	Recessive	<i>BANF1</i>	Exome	Puente <i>et al.</i> [81]
Infantile mitochondrial cardiomyopathy	Recessive	<i>AARS2</i>	Exome	Götz <i>et al.</i> [82]
Sensory neuropathy with dementia and hearing loss	Dominant	<i>DNMT1</i>	Exome	Klein <i>et al.</i> [49]
Autism	Dominant	Various	Exome	O'Roak <i>et al.</i> [32]

# Identification de nouveaux gènes (maladies rares)

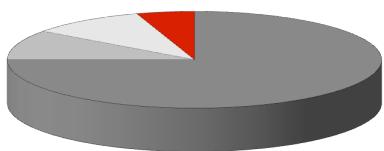


# Arythmies cardiaques

## Sudden Cardiac Death



Ventricular Fibrillation: common mechanism in 80% of SCD cases



Chugh SS, 2000

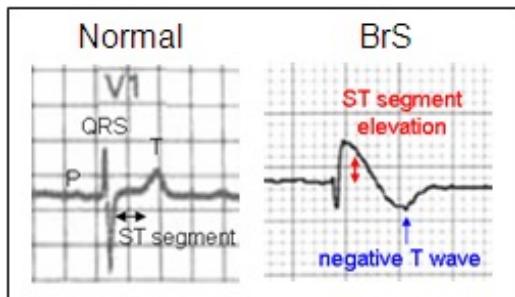
- Coronary heart diseases (75%)
- Hypertrophic cardiomyopathies (10%)
- Dilated cardiomyopathies (10%)
- Structurally normal hearts (5%)

## Idiopathic ventricular fibrillation

- ⇒ Brugada syndrome
- ⇒ Early repolarization syndrome
- ⇒ Long QT Syndrome
- ⇒ Cardiac conduction defects

# Syndrome de Brugada

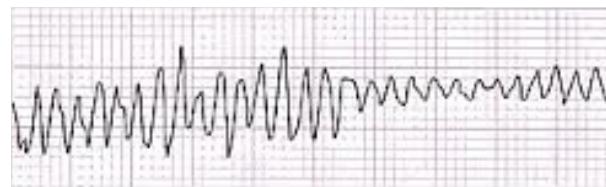
ST-segment elevation (> 2 mm) in the right precordial ECG leads



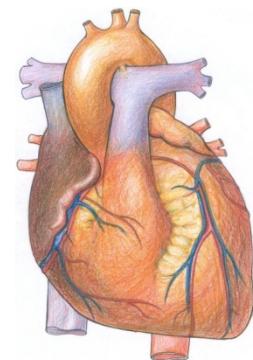
Brugada et al, 1992

Chen et al, 1998

Personal or familial susceptibility to (fatal) ventricular arrhythmias



Structurally normal heart



- Model: autosomal dominant transmission
- SCN5A mutations in ~ 15-25% cases
- Low penetrance (~ 20-30%) and overlap syndromes
- Mutations in >20 other susceptibility genes

➤ La majorité des cas sont encore inexpliqués

# Séquençage d'exome – syndrome de Brugada

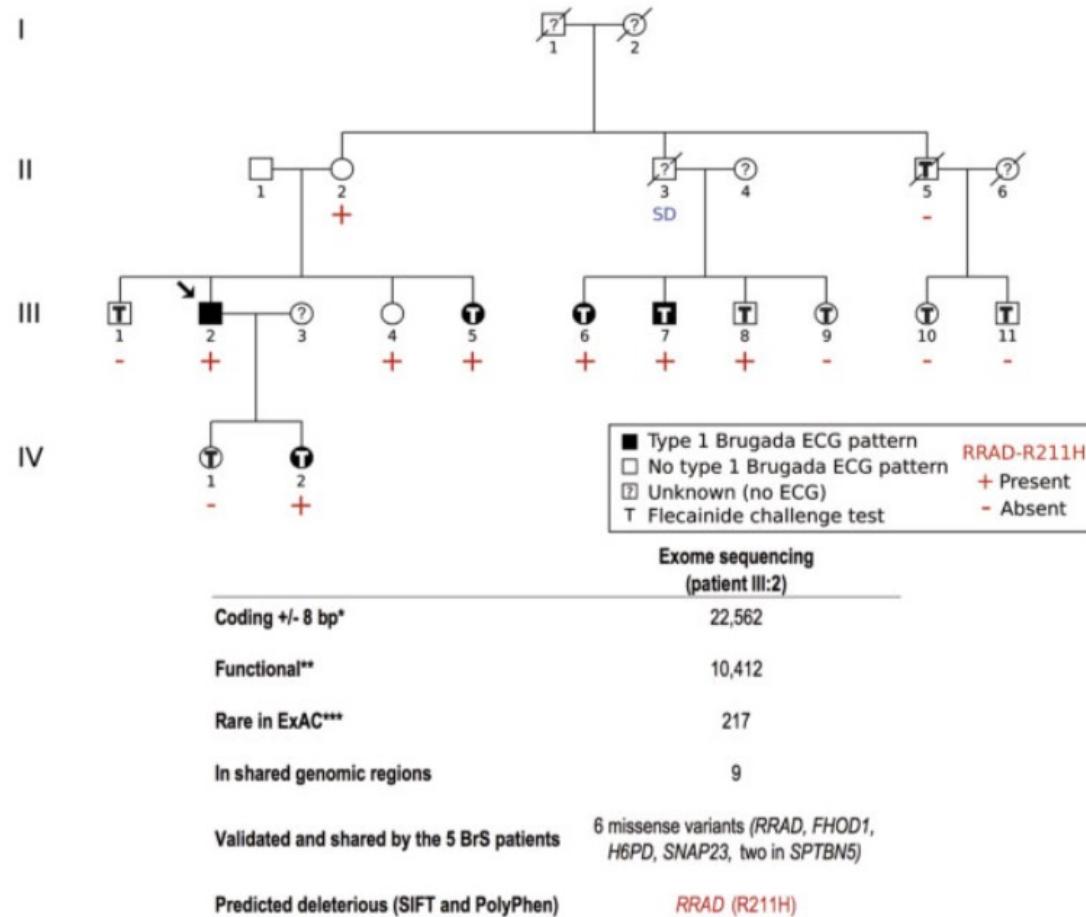


European Heart Journal (2019) 0, 1–14  
doi:10.1093/eurheart/ehz308

**BASIC SCIENCE**  
Arrhythmia/electrophysiology

## RRAD mutation causes electrical and cytoskeletal defects in cardiomyocytes derived from a familial case of Brugada syndrome

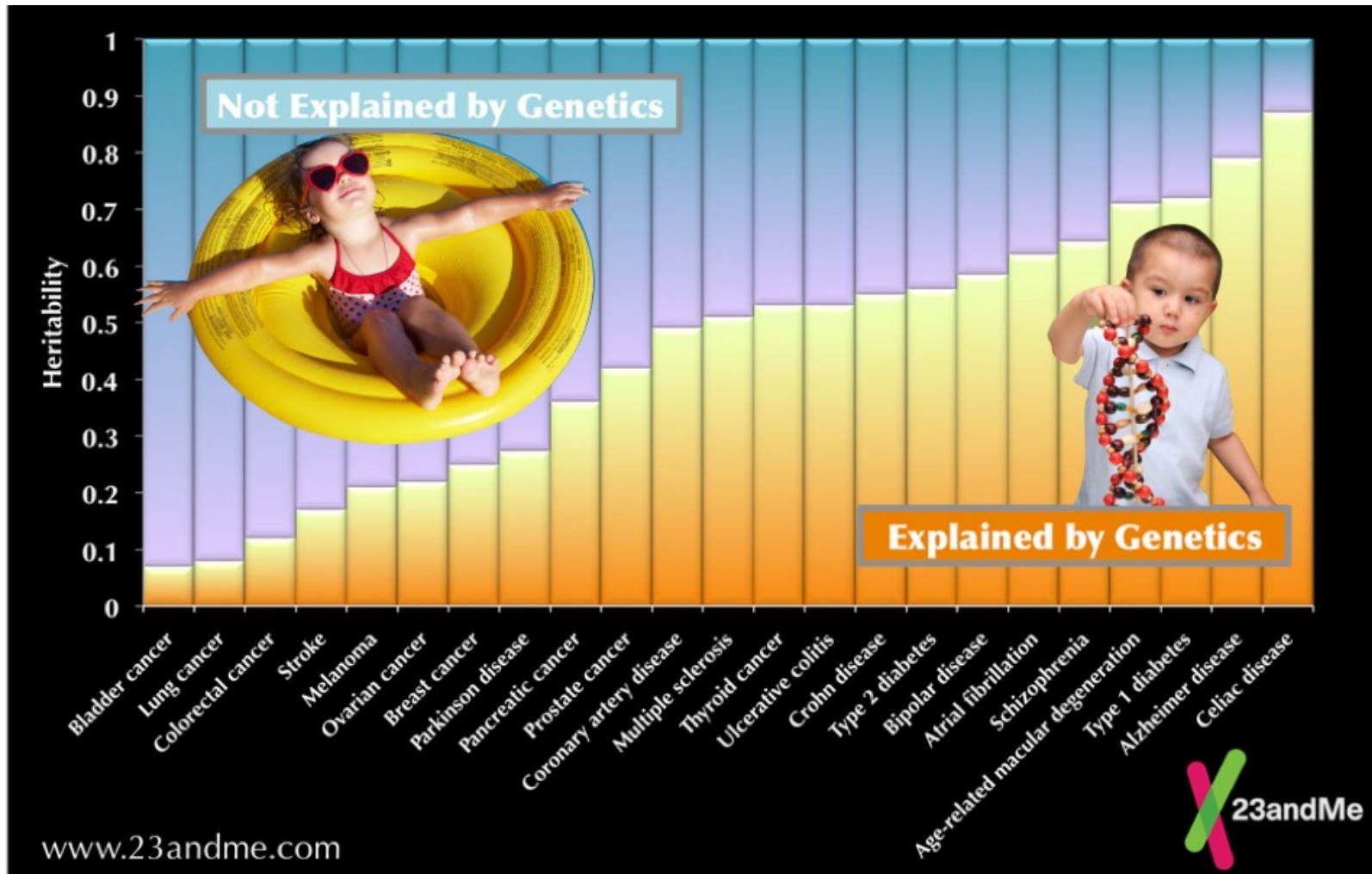
- Identification de RRAD-R211H dans une forme familiale
- 3 autres variants identifiés chez des cas sporadiques
- Investigations fonctionnelles iPSC-CMs



# Plan

- Introduction
- Variation du génome humain
- Méthodes d'identification de gènes
  - ✓ Maladies monogéniques
  - ✓ Maladies complexes

# Héritabilité des maladies communes



Maladies multifactorielles : GENES + ENVIRONNEMENT

Héritabilité : donnée statistique évaluant la part des facteurs génétiques dans la probabilité d'apparition d'un trait phénotypique donné au sein d'une population donnée

## Etude d'association

Corréler la présence d'un génotype avec un phénotype

Hypothèse : l'héritabilité des maladies communes est liée à des variants génétiques fréquents

# Principe de l'étude d'association

Marqueur 1

Marqueur 2

Cas 35 %

Cas 20 %

Témoins 15 %

Témoins 20 %

**Association**

**Pas d'association**

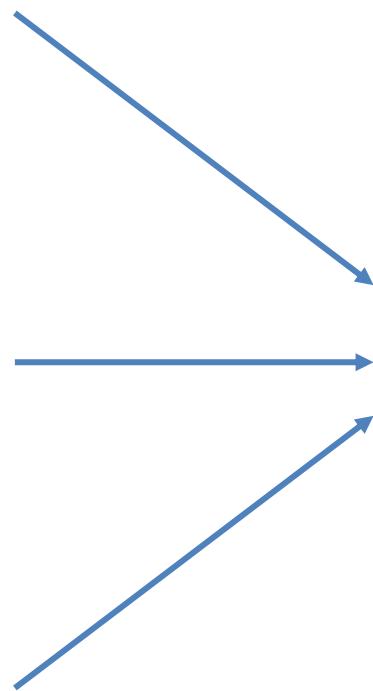
# Etudes d'association chez l'homme

Meilleure compréhension de  
la variation du génome  
humain (HapMap, 1000G)

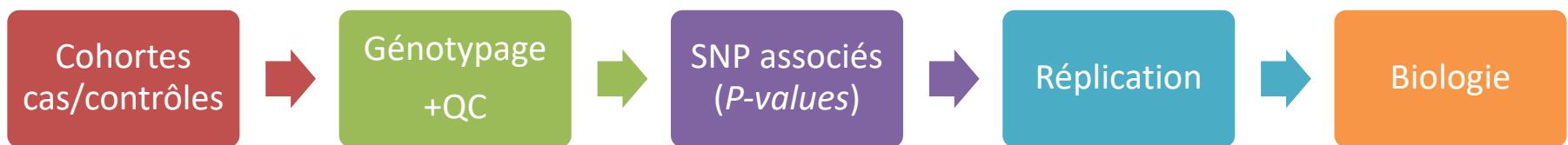
Avancées dans les  
technologies de génotypage  
(SNP)

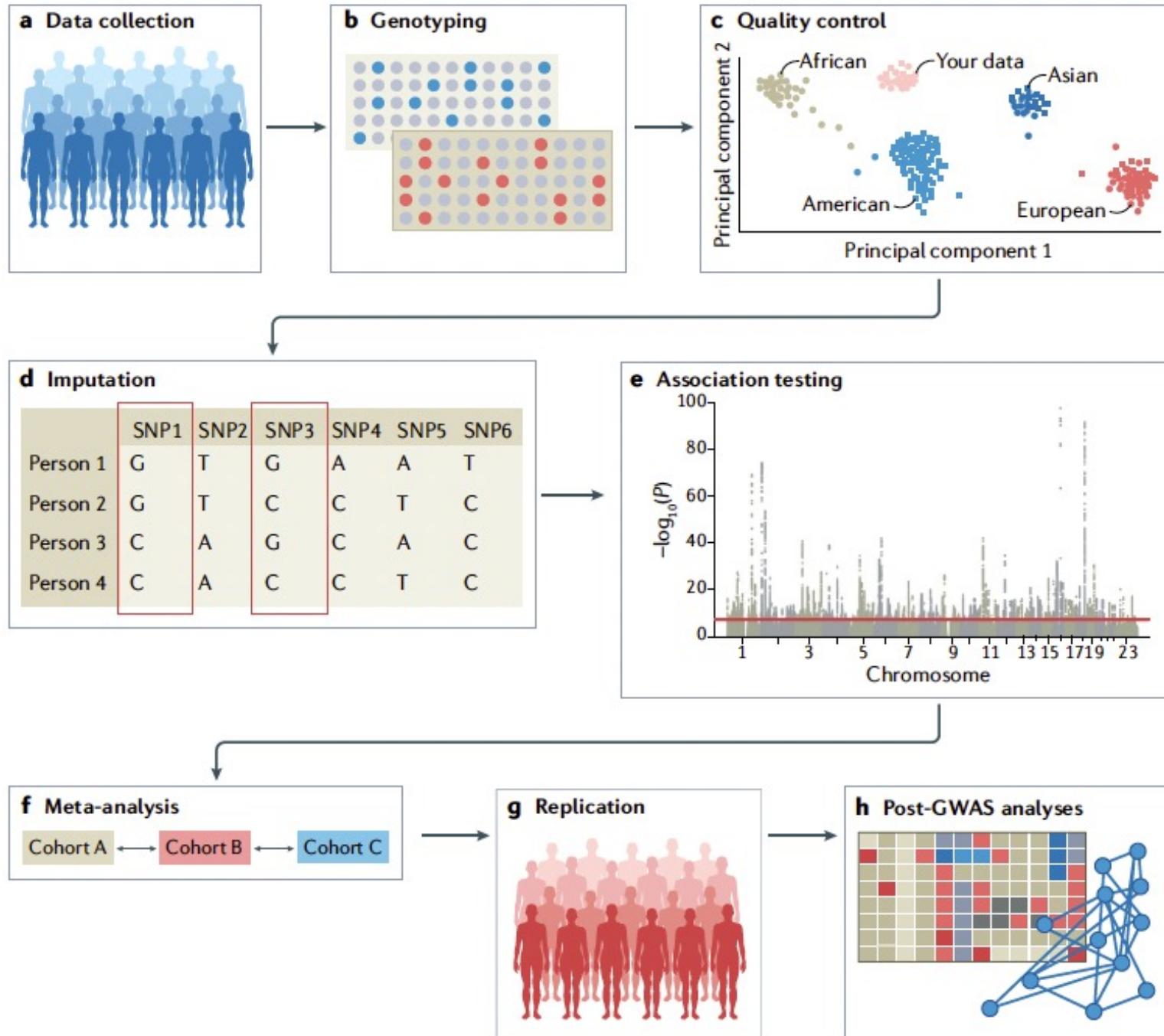
Grandes bio-collections  
de patients

Genome-wide association studies  
= (GWAS)



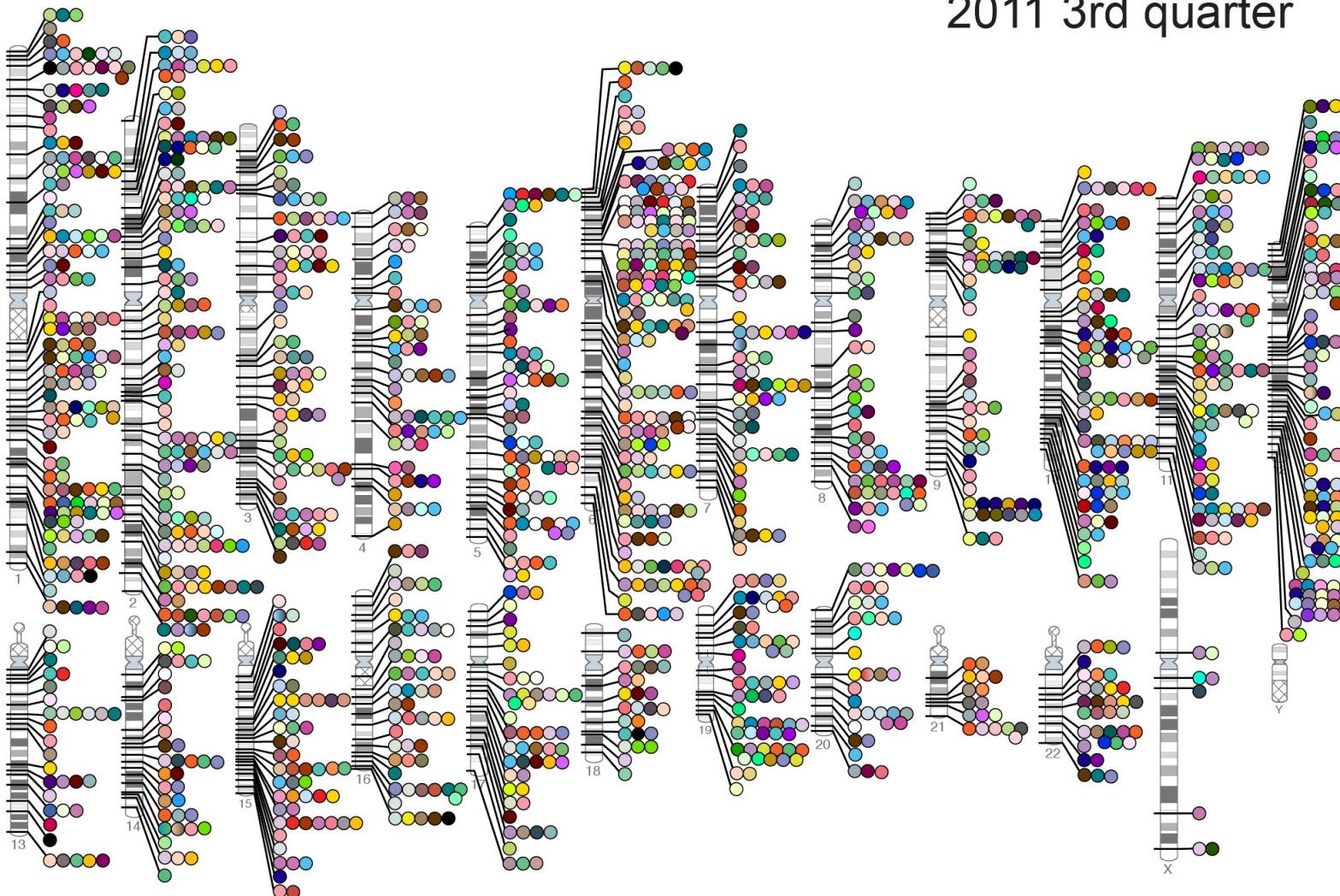
# Aperçu de l'approche GWAS





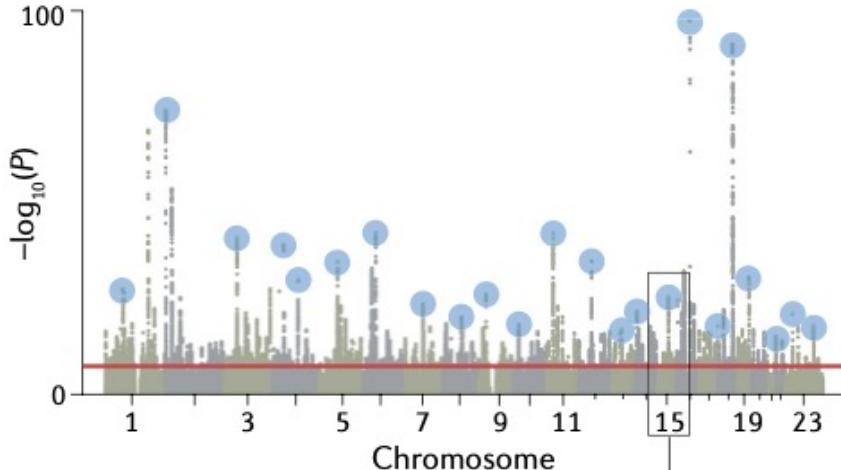
1,617 associations ( $p \leq 5 \times 10^{-8}$ ) publiées pour  
249 traits (09/2011)

2011 3rd quarter

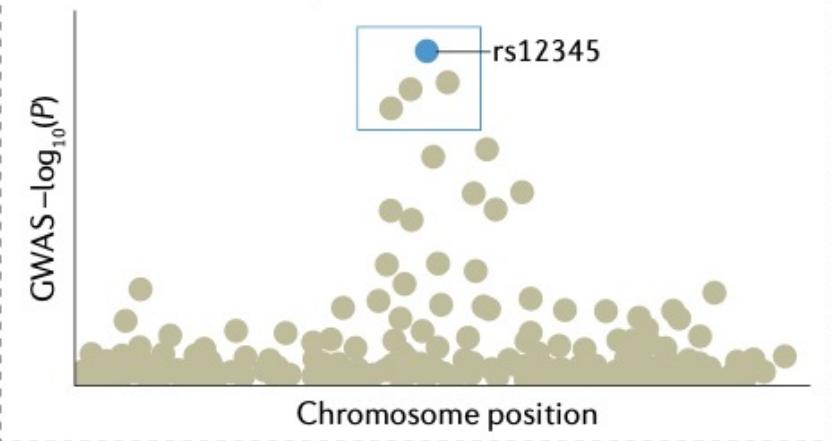




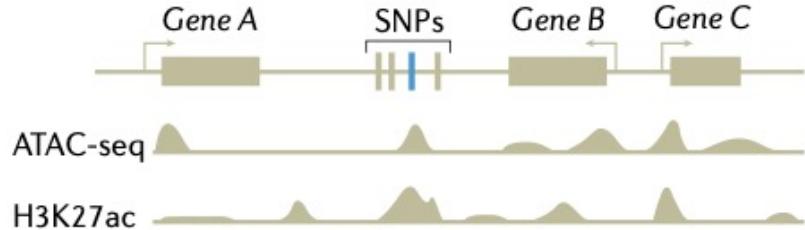
## a What are the associated loci?



## b What are the likely causal variants?

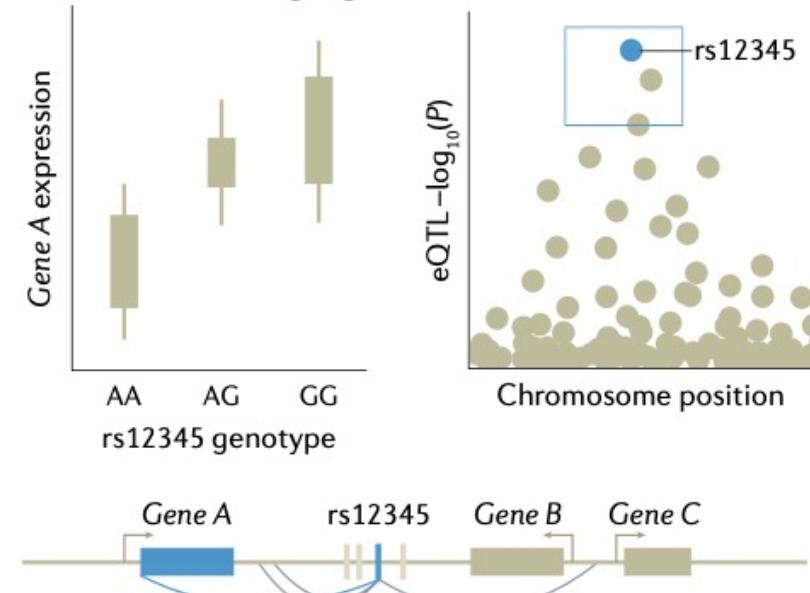


## c What are the epigenomic effects of variants?

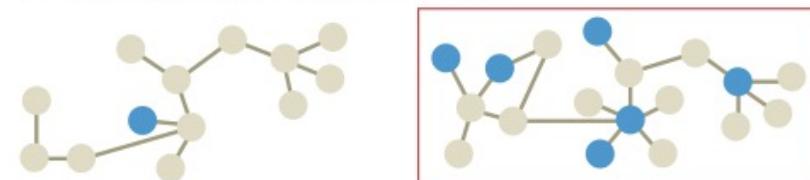


# Post-GWAS

## d What are the target genes in the locus?

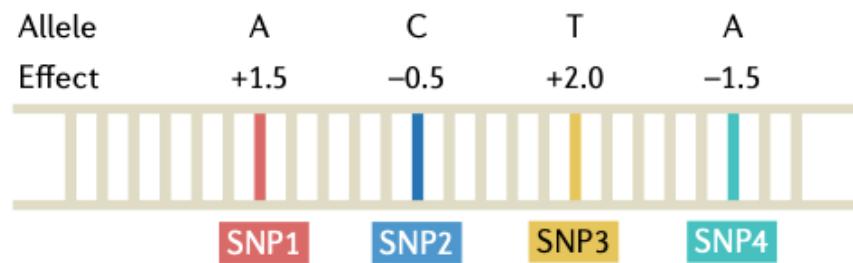


## e What are the affected pathways?



# Polygenic Risk Scores

## ① GWAS summary statistics



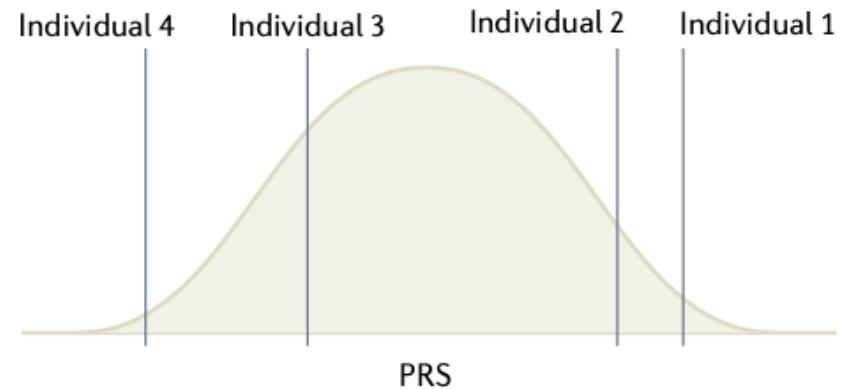
## ② Genotype data

	SNP1	SNP2	SNP3	SNP4
Individual 1	AT	CG	TT	CC
Individual 2	TA	GG	GT	CA
Individual 3	TT	CC	GT	CA
Individual 4	TT	CC	GG	AA

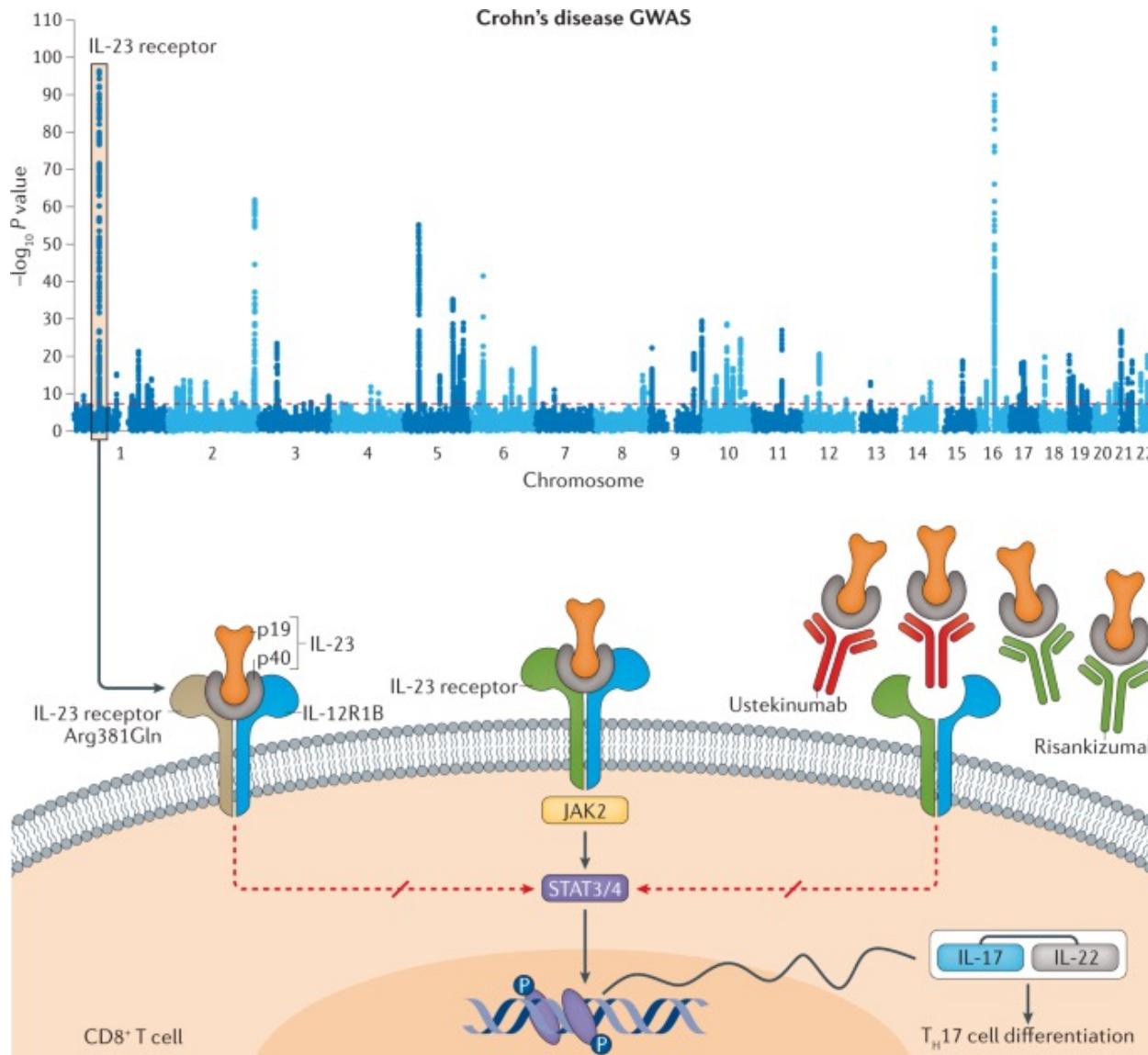
## ③ Polygenic risk score

Individual 1	1.5	-	0.5	+	4.0	-	0.0	=	5.0
Individual 2	1.5	-	0.0	+	2.0	-	1.5	=	2.0
Individual 3	0.0	-	1.0	+	2.0	-	1.5	=	-0.5
Individual 4	0.0	-	1.0	+	0.0	-	3.0	=	-4.0

## ④ PRS distribution



# Traitements



# Stratégies génétiques

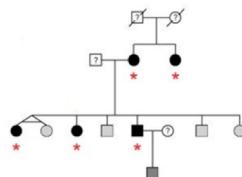
Effet du variant

Fort

Faible

Approche familiale

Analyse de liaison  
Séquençage



Rare

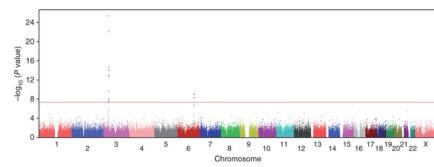
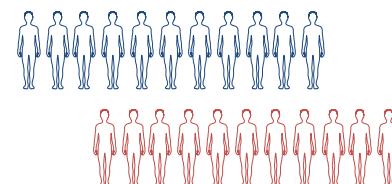
Elevée

Fréquence du variant

Approche population

Séquençage de génomes

Etude d'association  
(GWAS)



# Questions

