



Méthodes d'exploration du génome



Solena LE SCOUARNEC, PhD
solena.lescouarnec@univ-nantes.fr

Equipe I Génétique humaine

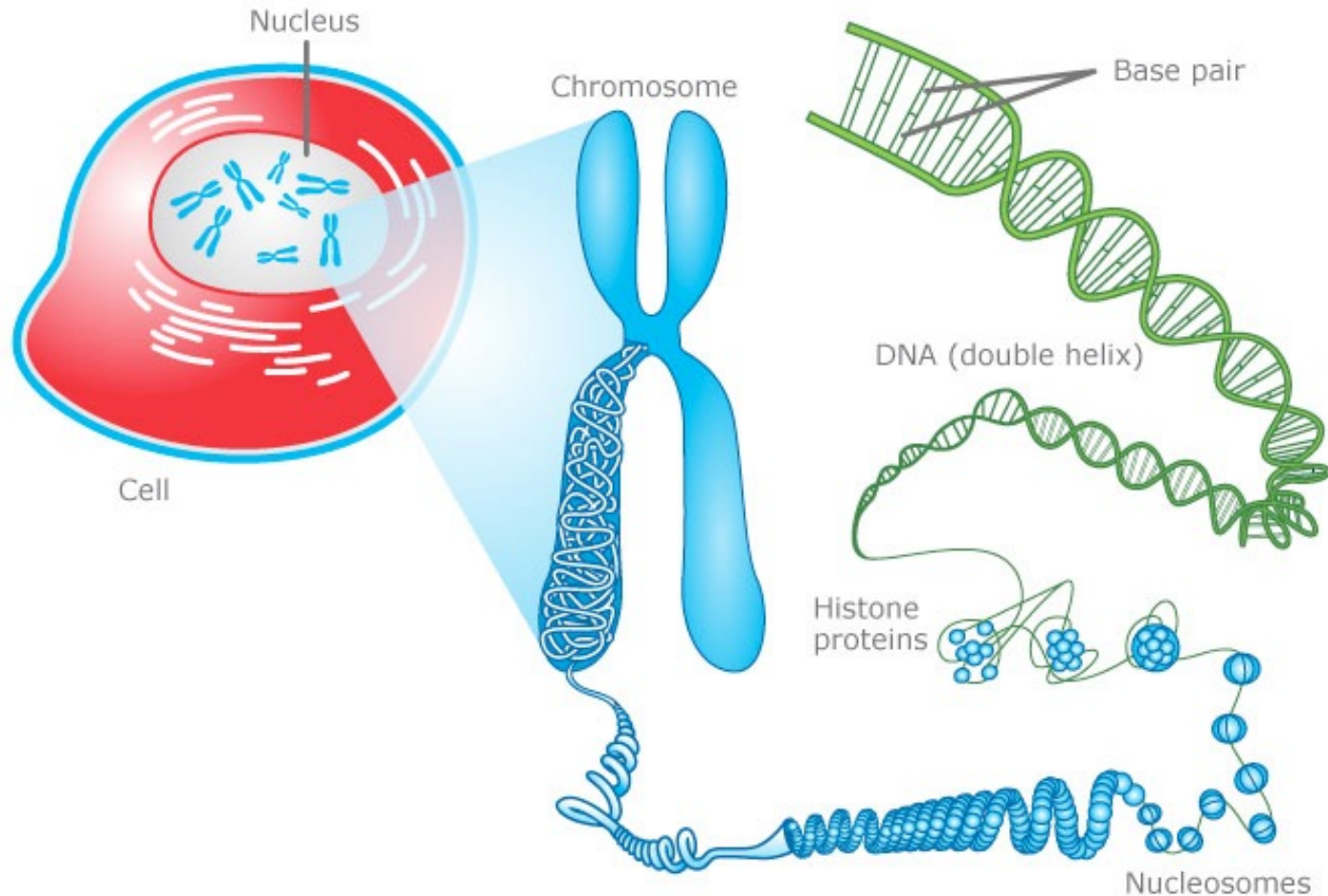
L'unité de recherche de l'institut du thorax
Inserm UMR 1087 / CNRS UMR 6291
Nantes, France

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?

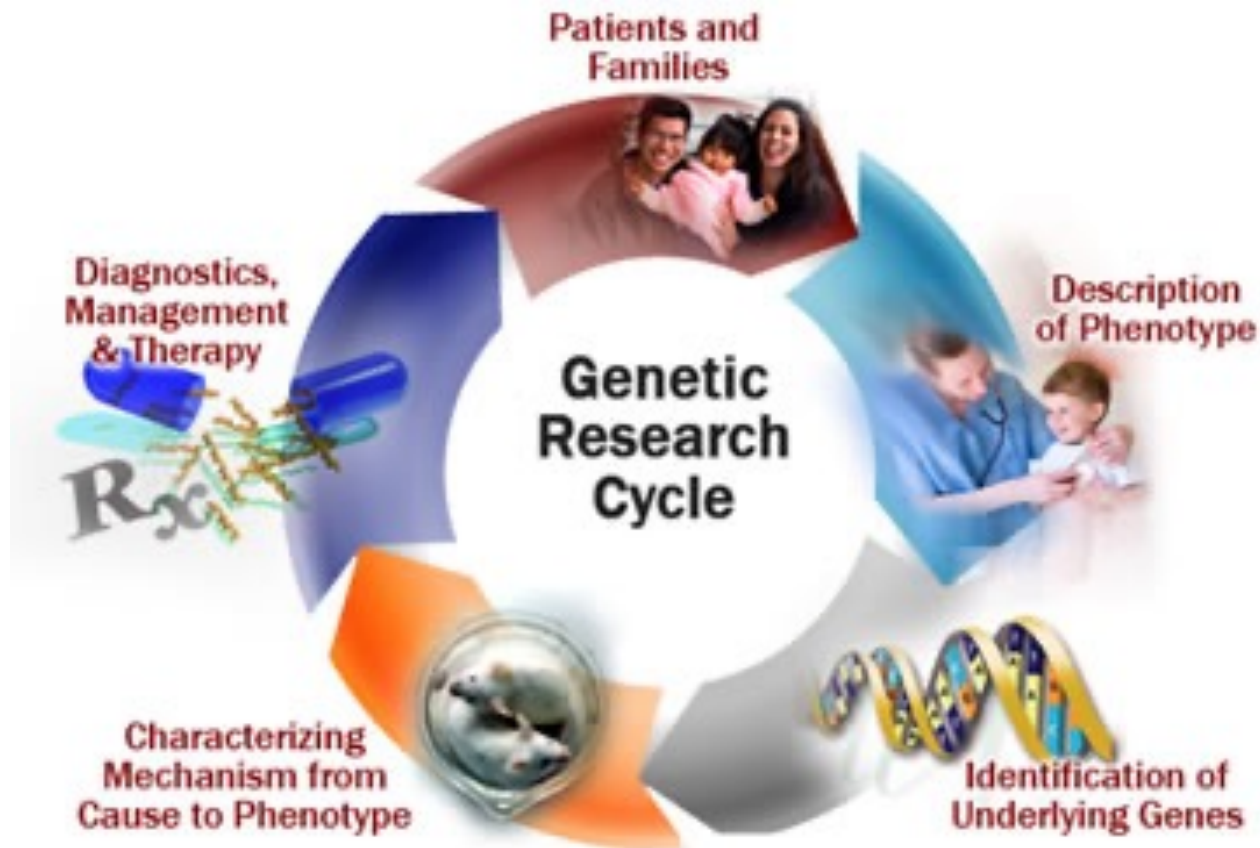
- Découvrir les variations génétiques à l'origine de pathologies
- Nouvelles thérapies
- Comprendre les bases moléculaires de l'évolution
- Découvrir les gènes impliqués dans l'embryogénèse et la physiologie

....

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^{1†}, Sergey Koren^{1†}, Arang Rhie^{1†}, Mikko Rautiainen^{1†}, Andrey V. Bzikadze², Alla Mikheenko³, Mitchell R. Vollger⁴, Nicolas Altemose⁵, Lev Uralsky^{6,7}, Ariel Gershman⁸, Sergey Aganezov^{9†}, Savannah J. Hoyt¹⁰, Mark Diekhans¹¹, Glennis A. Logsdon⁴, Michael Alonge⁹, Stylianos E. Antonarakis¹², Matthew Borchers¹³, Gerard G. Bouffard¹⁴, Shelise Y. Brooks¹⁴, Gina V. Caldas¹⁵, Nae-Chyun Chen⁹, Haoyu Cheng^{16,17}, Chen-Shan Chin¹⁸, William Chow¹⁹, Leonardo G. de Lima¹³, Philip C. Dishuck⁴, Richard Durbin^{19,20}, Tatiana Dvorkina³, Ian T. Fiddes²¹, Giulio Formenti^{22,23}, Robert S. Fulton²⁴, Arkarachai Fungtammasan¹⁸, Erik Garrison^{11,25}, Patrick G. S. Grady¹⁰, Tina A. Graves-Lindsay²⁶, Ira M. Hall²⁷, Nancy F. Hansen²⁸, Gabrielle A. Hartley¹⁰, Marina Haukness¹¹, Kerstin Howe¹⁹, Michael W. Hunkapiller²⁹, Chirag Jain^{1,30}, Miten Jain¹¹, Erich D. Jarvis^{22,23}, Peter Kerpedjiev³¹, Melanie Kirsche⁹, Mikhail Kolmogorov³², Jonas Korlach²⁹, Milinn Kremitzki²⁶, Heng Li^{16,17}, Valerie V. Maduro³³, Tobias Marschall³⁴, Ann M. McCartney¹, Jennifer McDaniel³⁵, Danny E. Miller^{4,36}, James C. Mullikin^{14,28}, Eugene W. Myers³⁷, Nathan D. Olson³⁵, Benedict Paten¹¹, Paul Peluso²⁹, Pavel A. Pevzner³², David Porubsky⁴, Tamara Potapova¹³, Evgeny I. Rogaev^{6,7,38,39}, Jeffrey A. Rosenfeld⁴⁰, Steven L. Salzberg^{9,41}, Valerie A. Schneider⁴², Fritz J. Sedlazeck⁴³, Kishwar Shafin¹¹, Colin J. Shew⁴⁴, Alaina Shumate⁴¹, Ying Sims¹⁹, Arian F. A. Smit⁴⁵, Daniela C. Soto⁴⁴, Ivan Sovic^{29,46}, Jessica M. Storer⁴⁵, Aaron Streets^{5,47}, Beth A. Sullivan⁴⁸, Françoise Thibaud-Nissen⁴², James Torrance¹⁹, Justin Wagner³⁵, Brian P. Walenz¹, Aaron Wenger²⁹, Jonathan M. D. Wood¹⁹, Chunlin Xiao⁴², Stephanie M. Yan⁴⁹, Alice C. Young¹⁴, Samantha Zarate⁹, Urvasi Surti⁵⁰, Rajiv C. McCoy⁴⁹, Megan Y. Dennis⁴⁴, Ivan A. Alexandrov^{3,7,51}, Jennifer L. Gerton^{13,52}, Rachel J. O'Neill¹⁰, Winston Timp^{8,41}, Justin M. Zook³⁵, Michael C. Schatz^{9,49}, Evan E. Eichler^{4,53*}, Karen H. Miga^{11,54*}, Adam M. Phillippy^{1*}

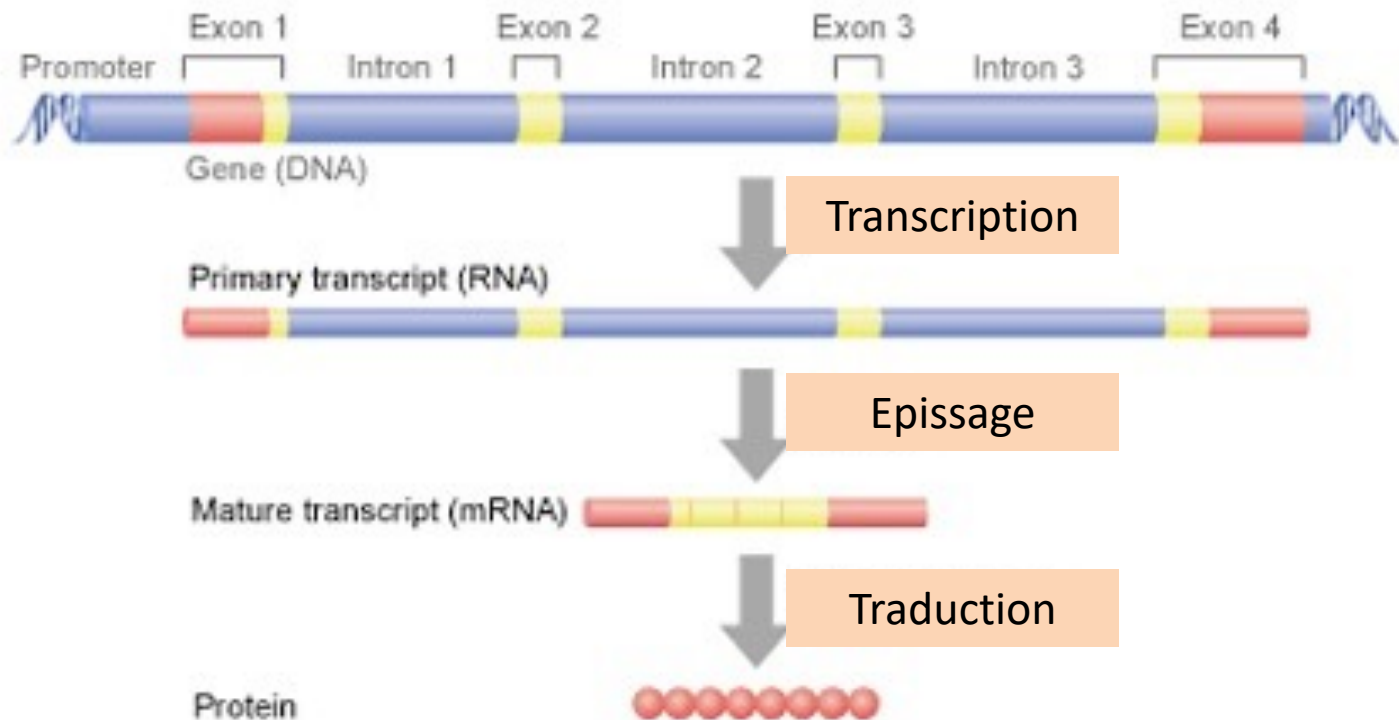
Since its initial release in 2000, the human reference genome has covered only the euchromatic fraction of the genome, leaving important heterochromatic regions unfinished. Addressing the remaining 8% of the genome, the Telomere-to-Telomere (T2T) Consortium presents a complete 3.055 billion-base pair 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

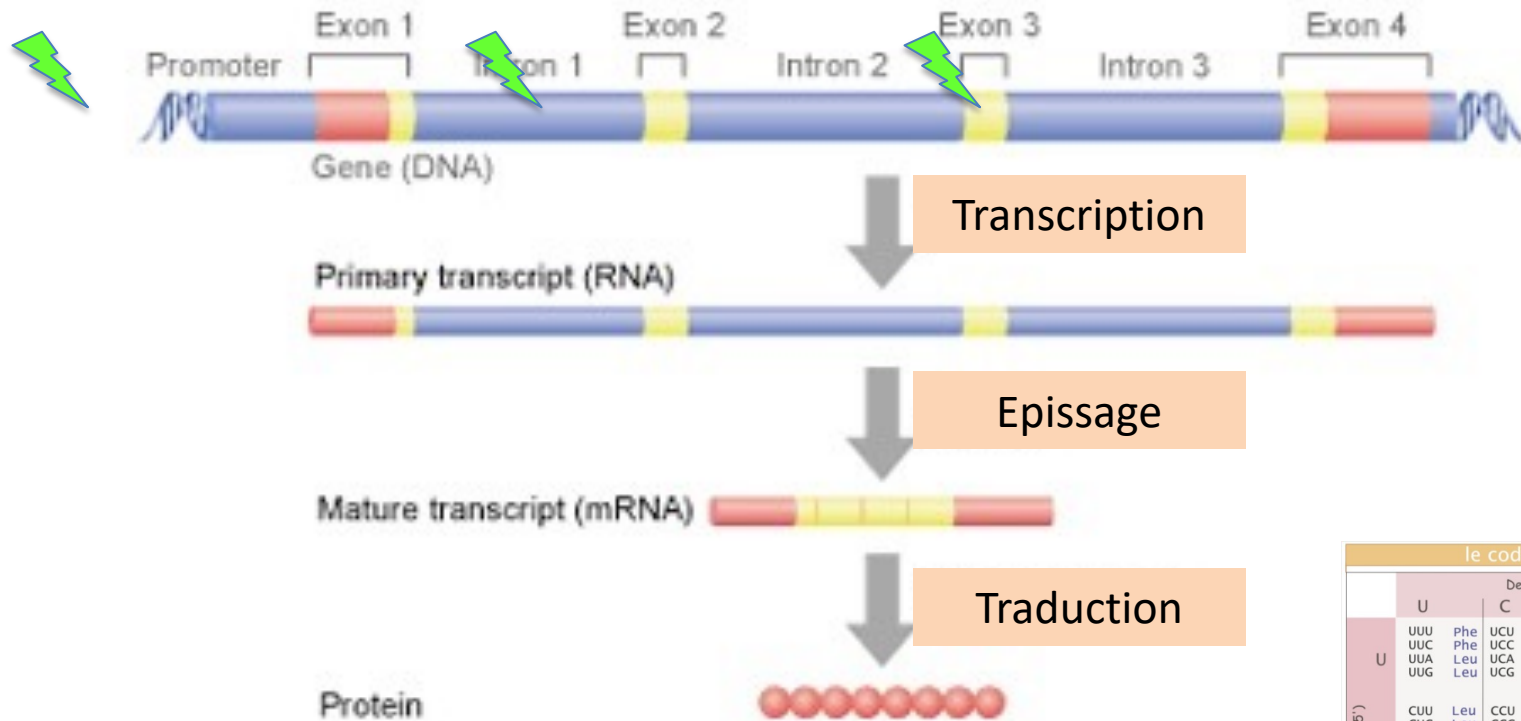
2 copies de chaque gène

20 000 – 25 000 gènes dans le génome humain

Exons / introns

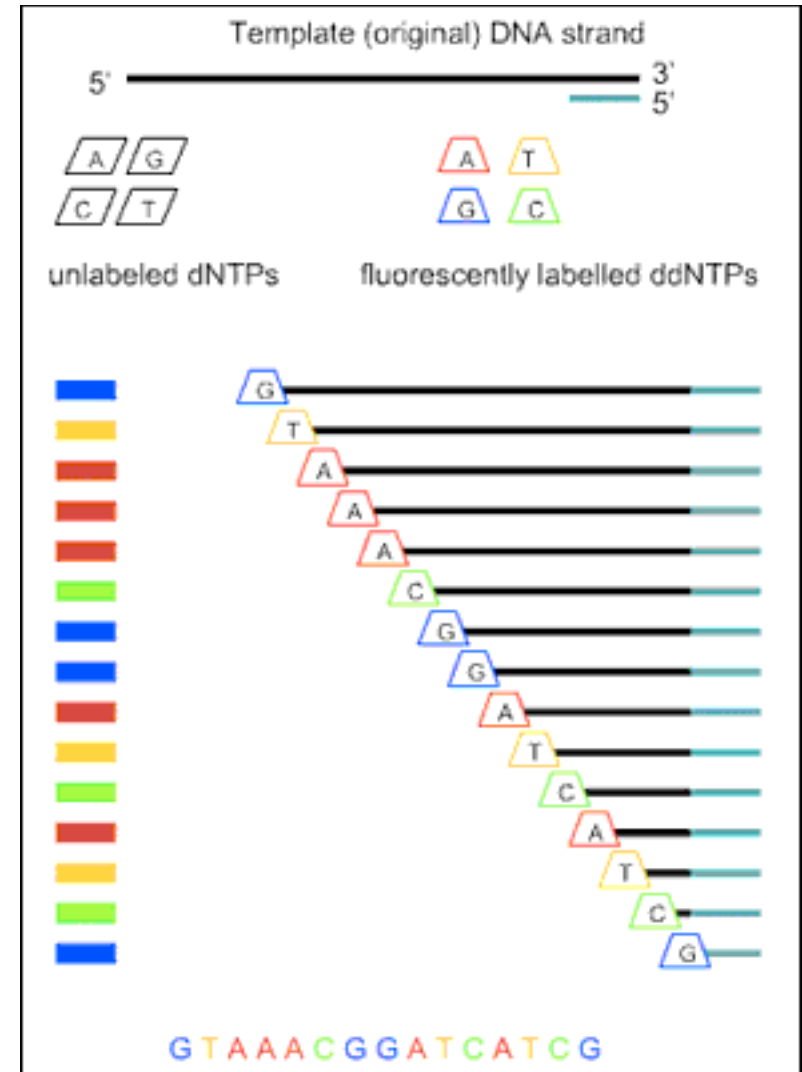
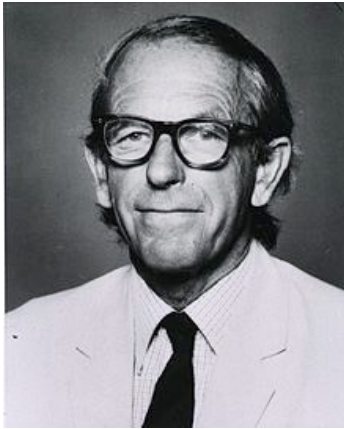


Mutations génétiques



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

Séquençage Sanger



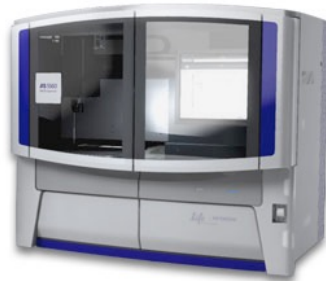
Séquençage nouvelle génération (NGS)

2005-2013



Pyrosequencing

2006-



*Sequencing
by ligation*

2006-



*Sequencing
by
synthesis*

2011-



*Single molecule
via dye labels
Long reads*

2015-



*Electronic nanopore sensing
No dye labels*

Nouvelles technologies de séquençage

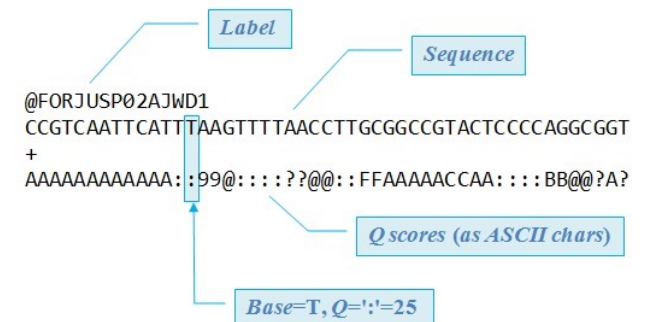
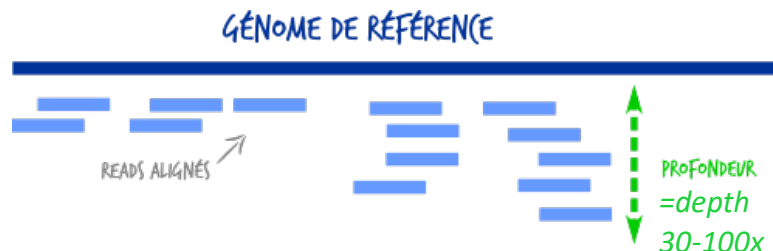
Next-generation sequencing (NGS) = Séquençage massif = Séquençage haut-débit

- 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

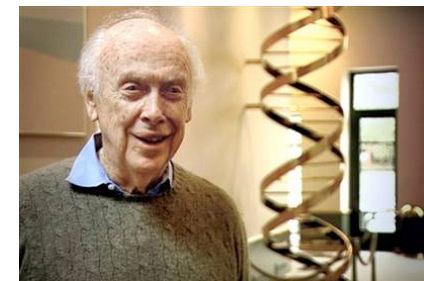
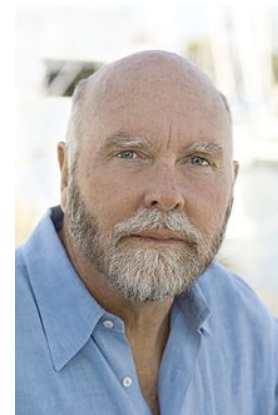
QUICKER, SMALLER, CHEAPER

Sanger

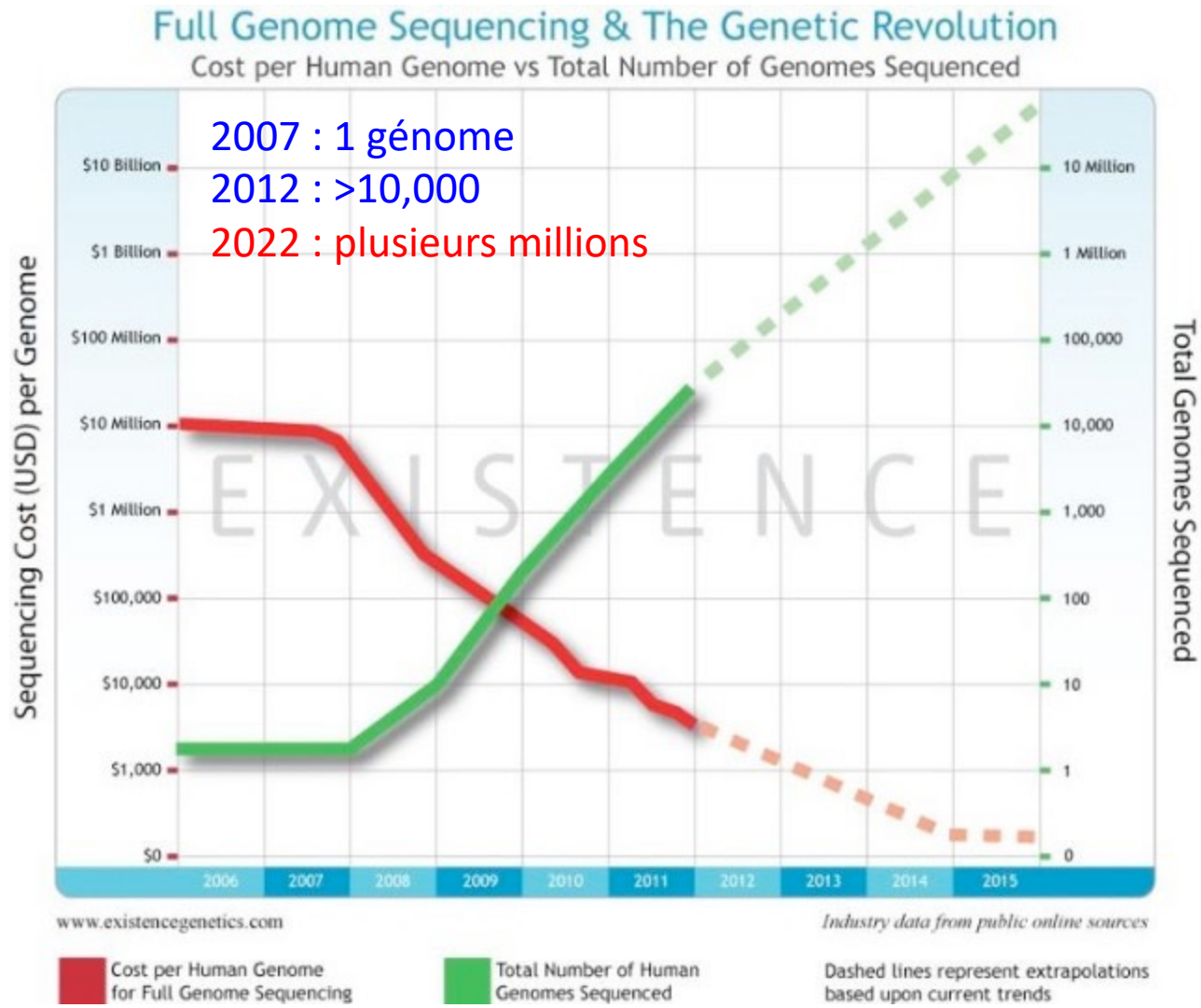
Sanger

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 \Leftrightarrow exome (régions codantes)



1 génome
2 jours
800 euros
150 Go

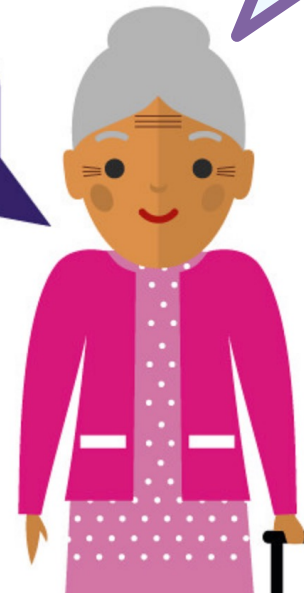
If you spoke
one 'letter' of DNA
per second

24 hours a day,

it would take about

100 years to recite the entire
human genome

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

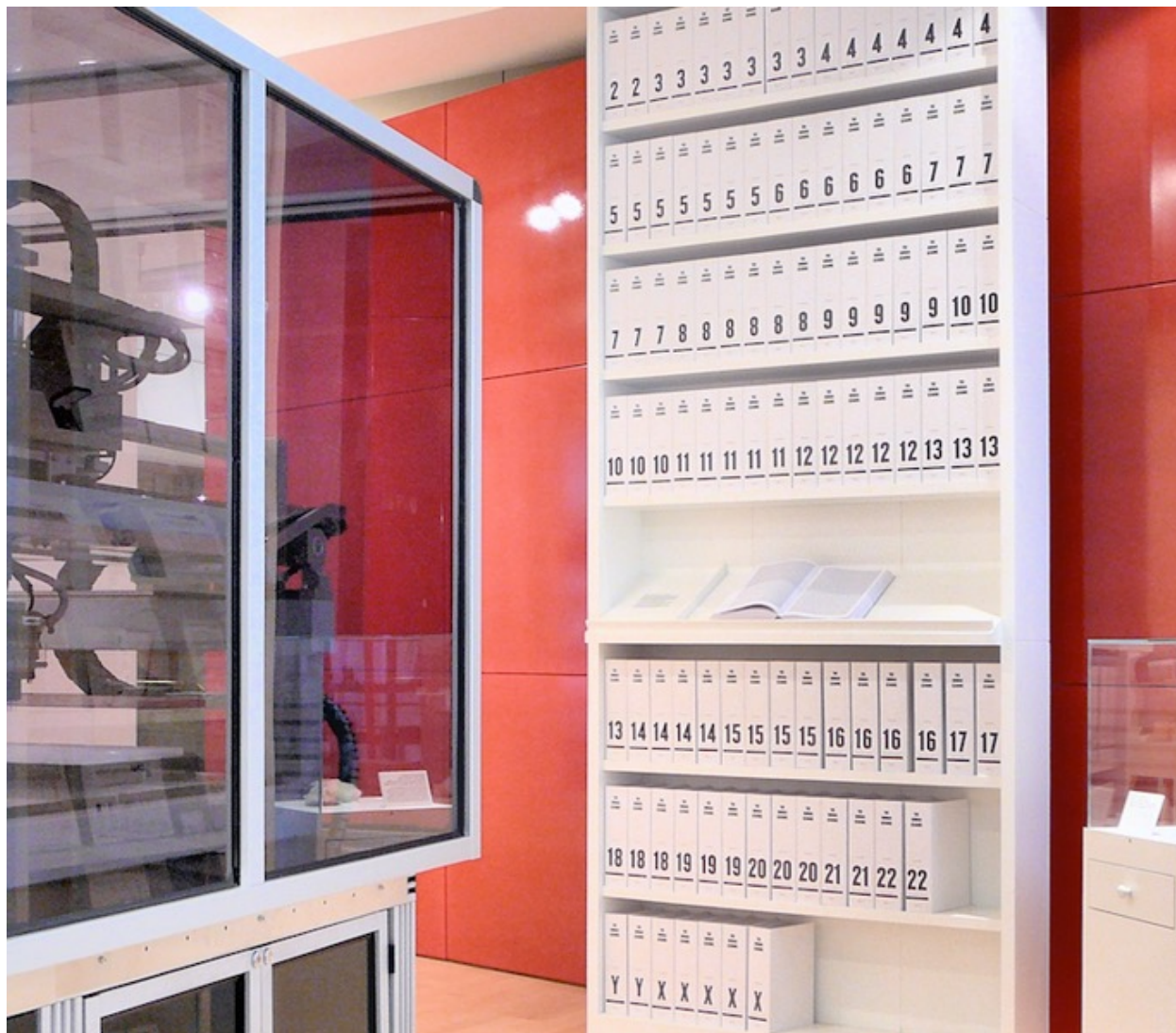


Séquençage de génomes personnels

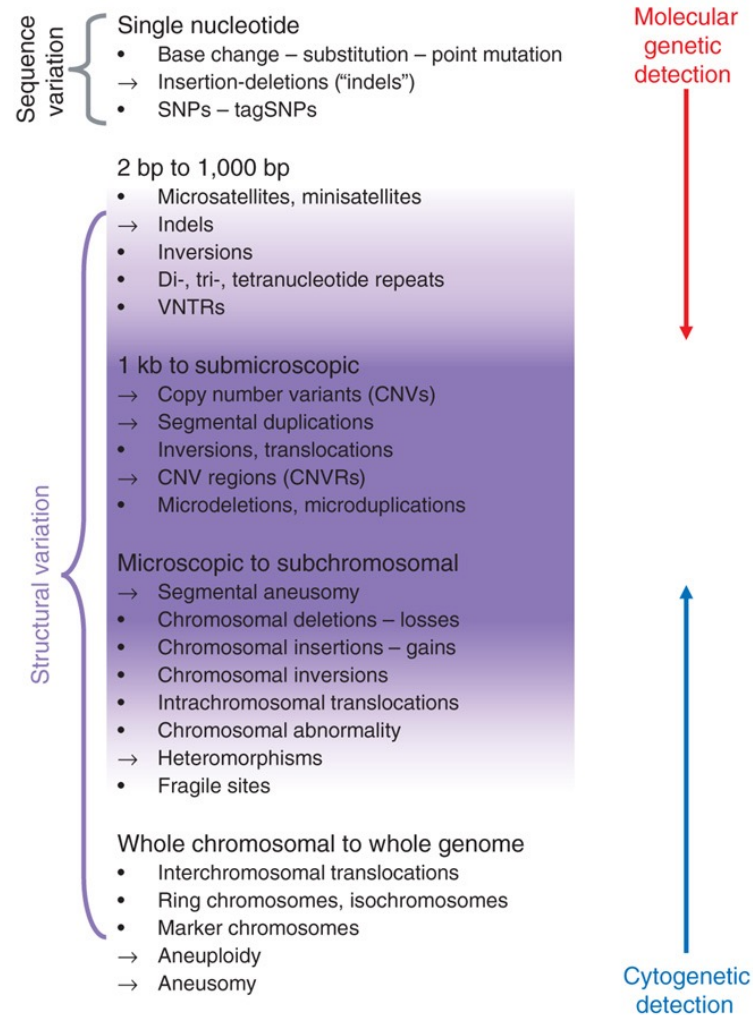


Plan

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



Types de variations



Variation du génome

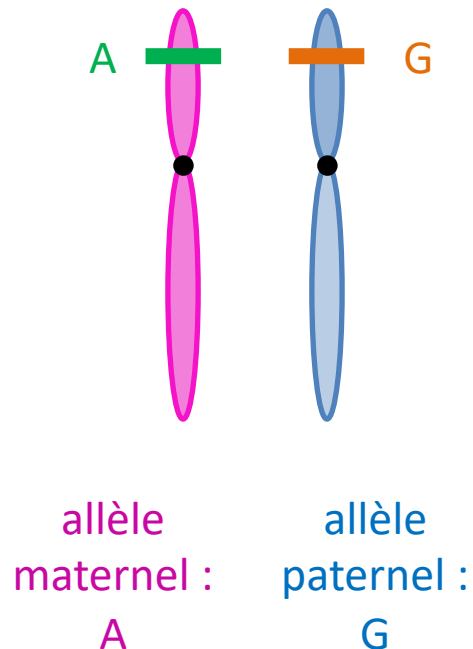
- Deux génomes humains sont identiques à 99,9%
- Différences dans la séquence du génome = variations génétiques
- Chaque variation génétique est caractérisée par ≥ 2 allèles

Allèle

Version variable d'un même locus génétique

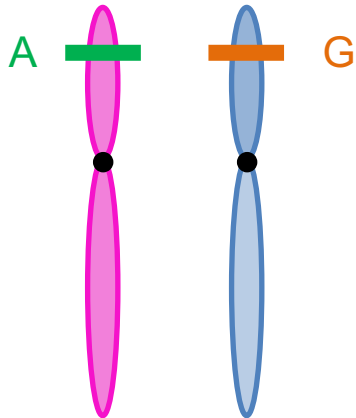
Exemple : à une position donnée, allèle A ou allèle G

CGATGCTA[A/G]GCATGCAT

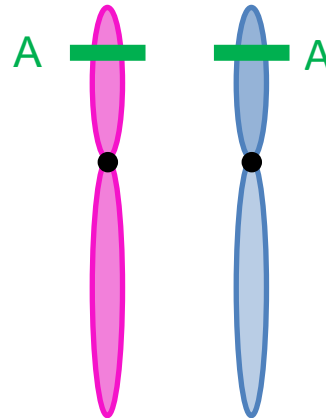


Génotype

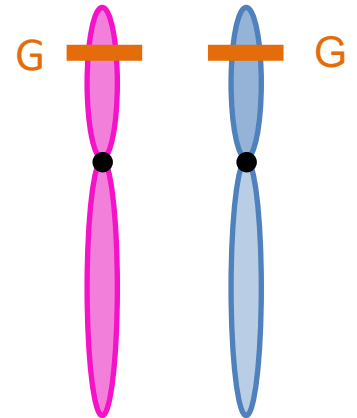
Composition allélique pour une variation génétique donnée



Génotype individu 1 : A/G



Génotype individu 2 : A/A



Génotype individu 3 : G/G

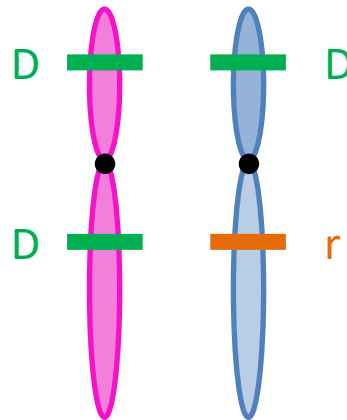
Si les 2 allèles sont différents,
on parle d'hétérozygotie.

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

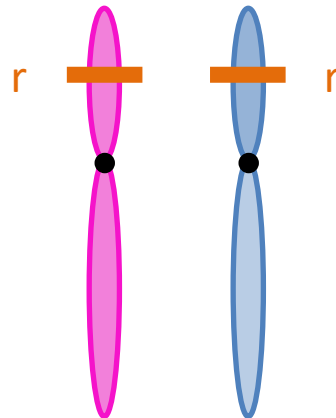
Phénotype, dominance et récessivité

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 ≥ 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



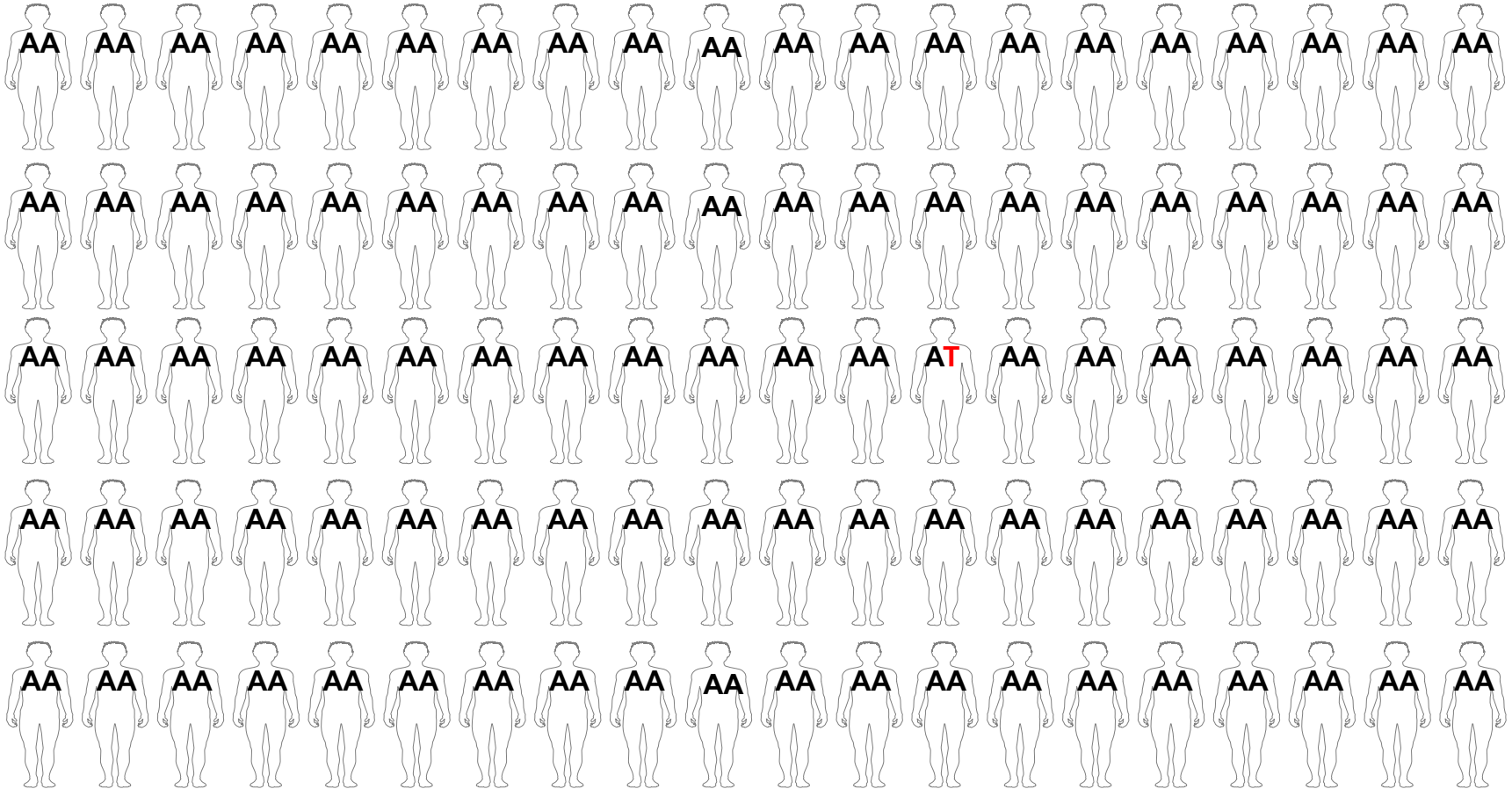
60 **G** / 200 allèles
=30 %

Exemple de variant rare (<1%)



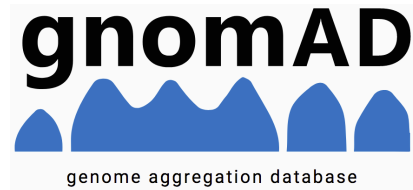
Génome de référence : CGATGCTA**A**GCATGC**A**T

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



1 **T** / 200 allèles
=0,5 %

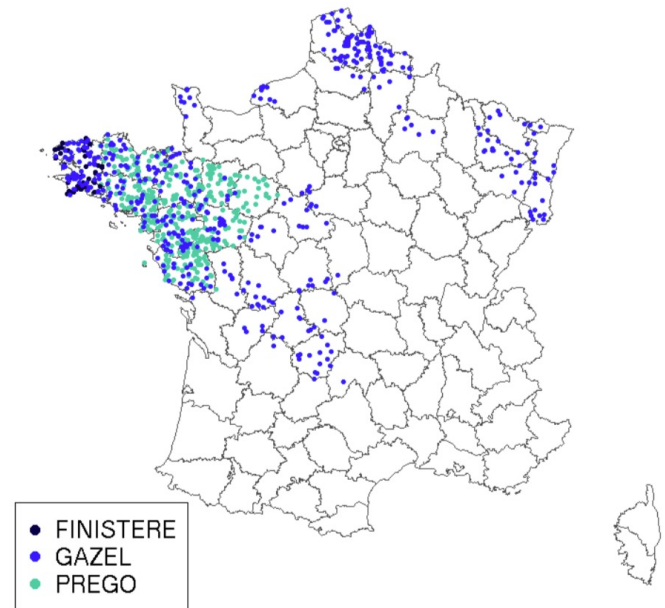
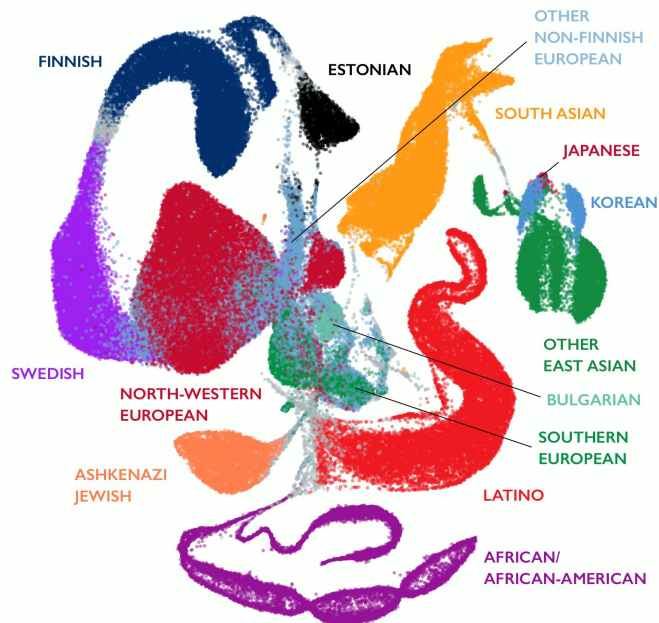
Fréquence des variants



Exomes : 125 000 individus
Génomes : 71 000 individus



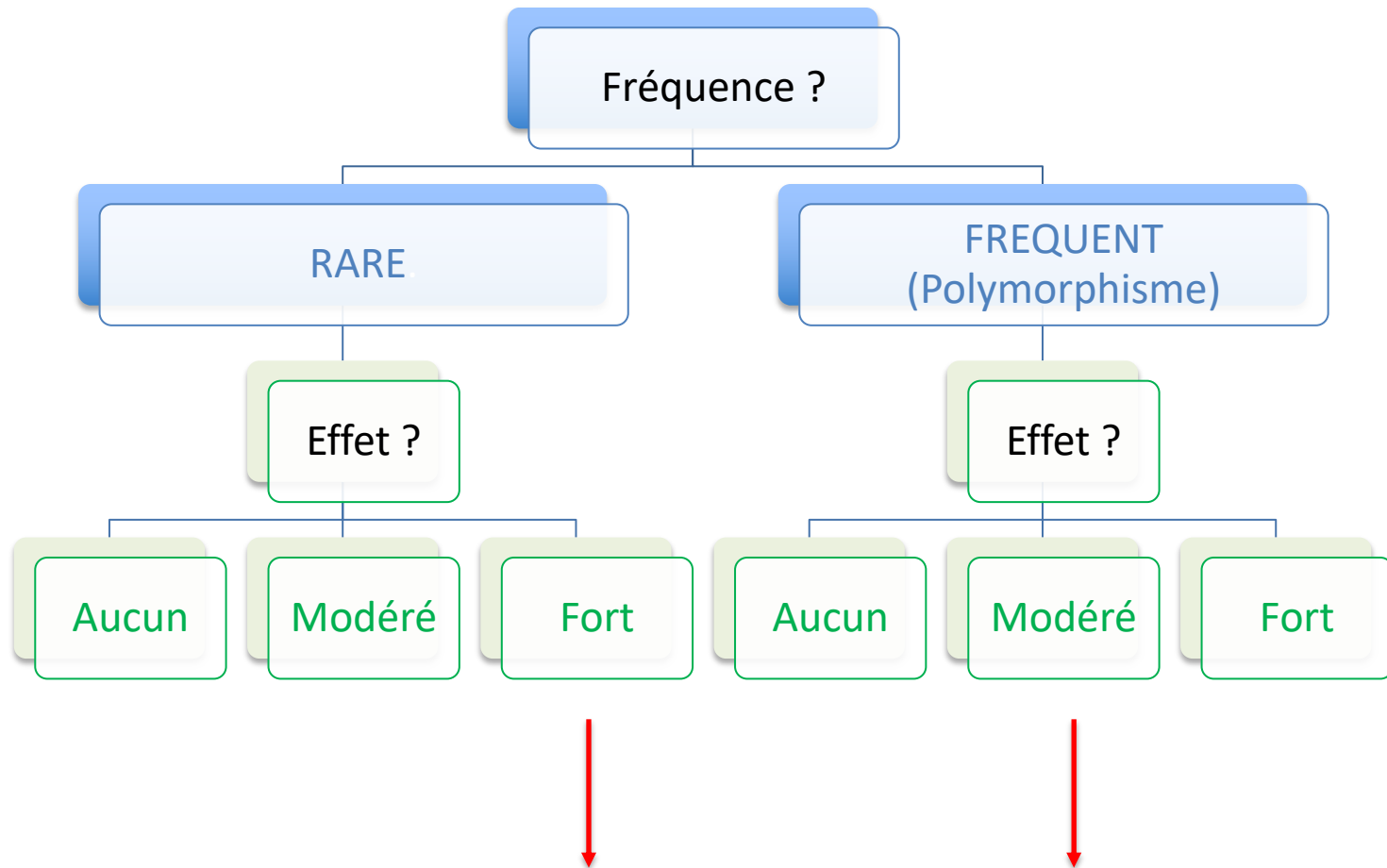
Génomes : 1 000 individus



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 ≥ 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



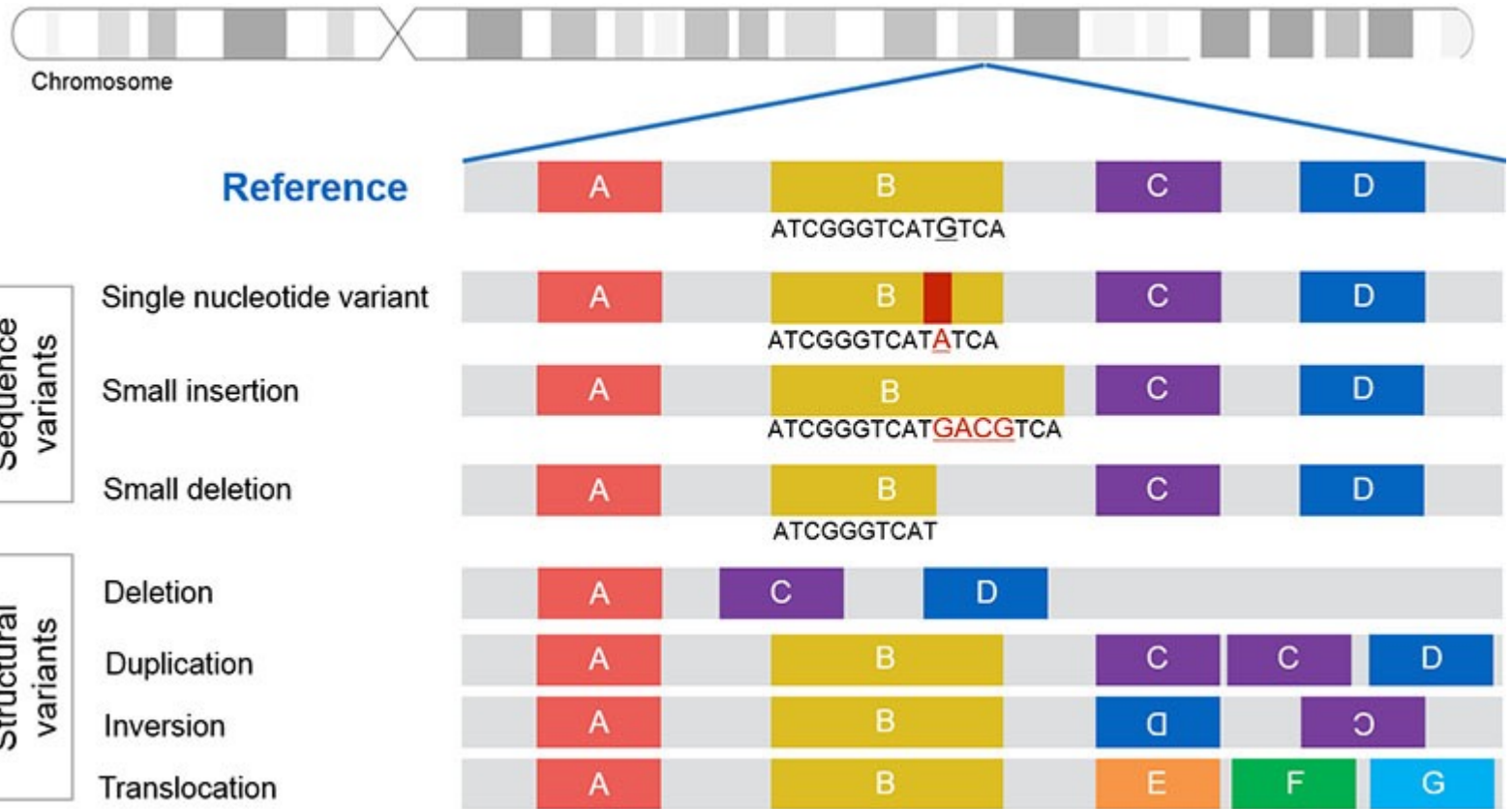
Approche familiale

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

Approche population

↑ risque
Exemple :
Diabète, Alzheimer...

Variabilité du génome



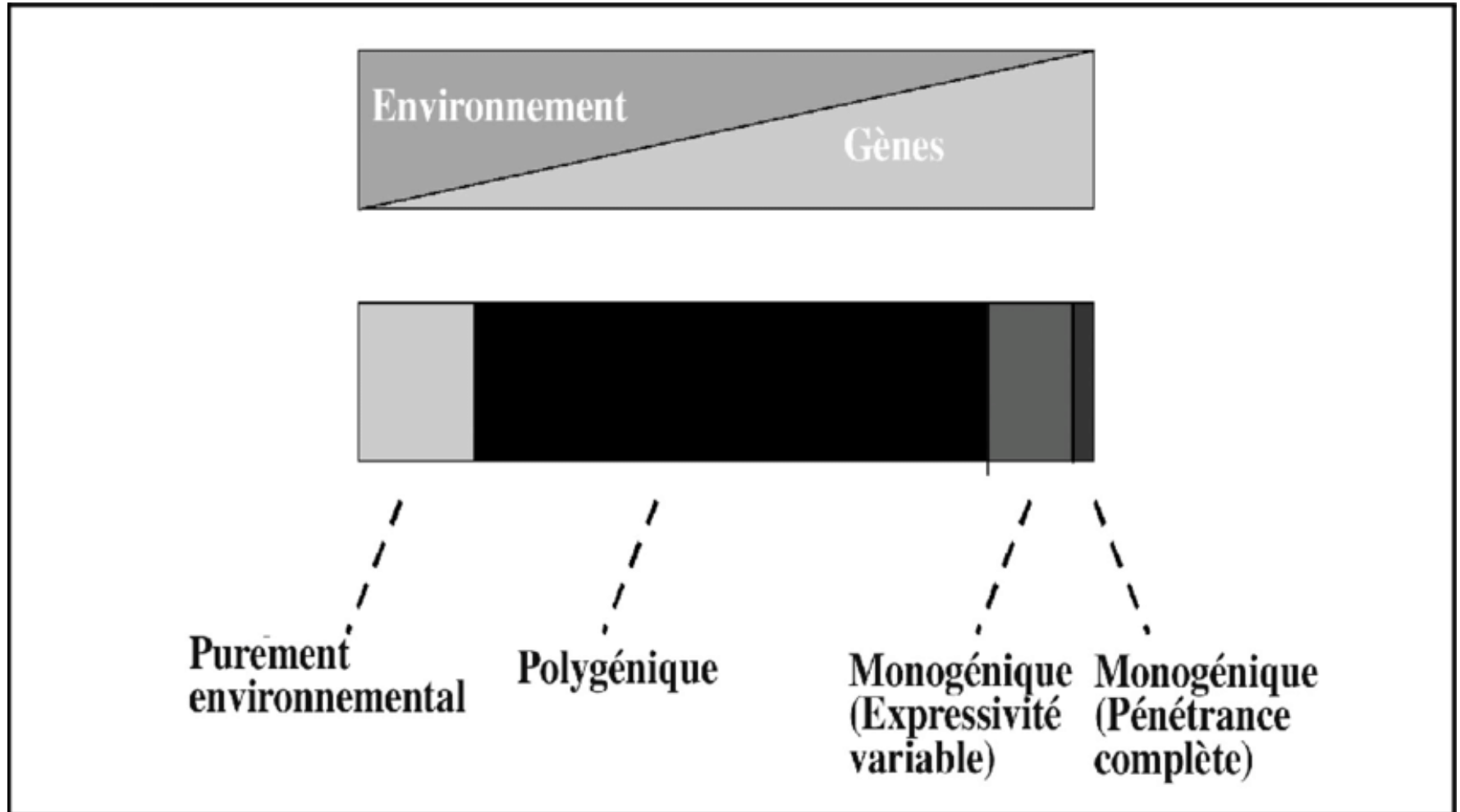
3-4 millions

~20 000

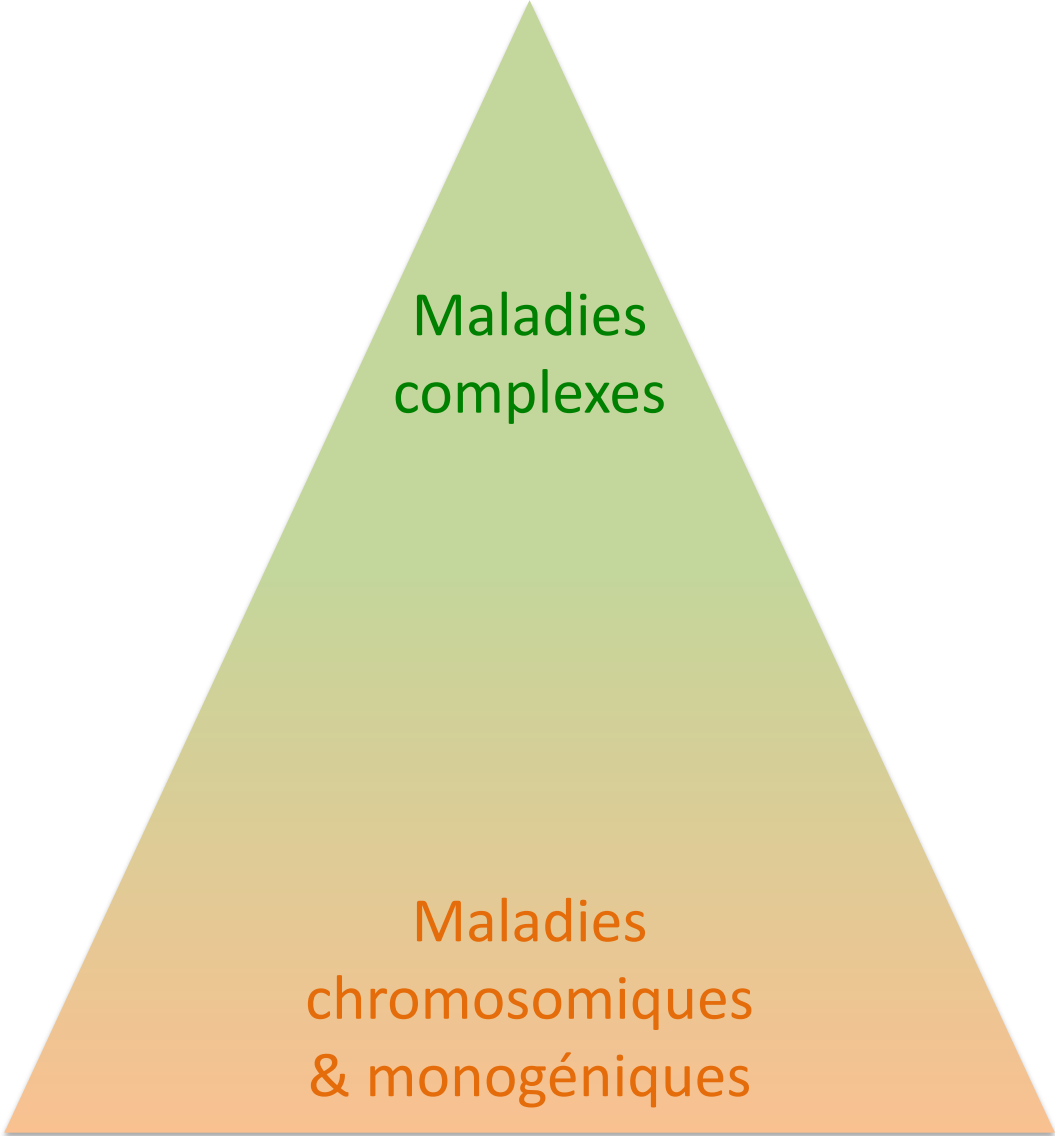
Plan

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

Maladies humaines



Génétique moléculaire



Maladies
complexes

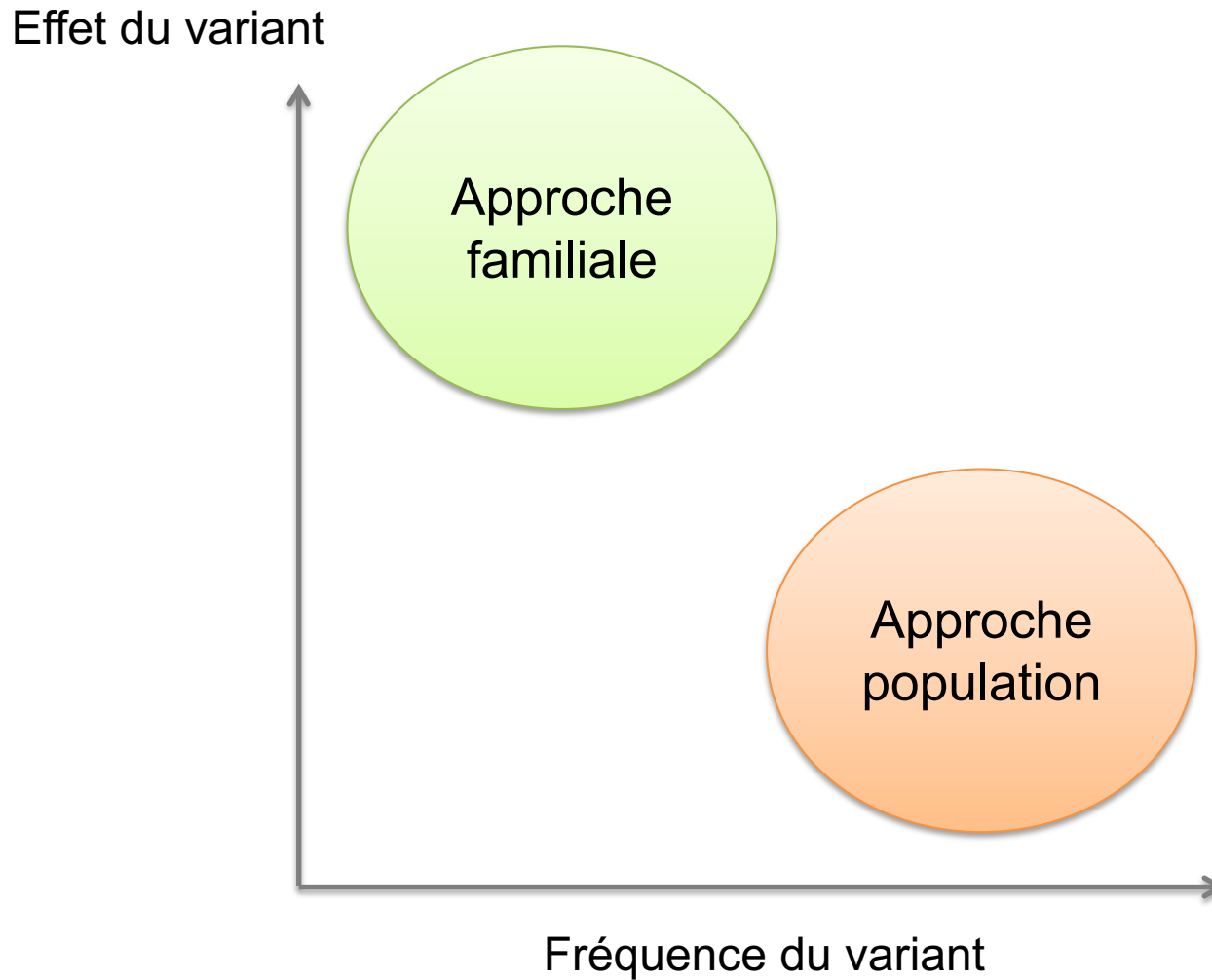
Maladies
chromosomiques
& monogéniques

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

Implications éthiques, légales
et sociales

Ex : 23andMe

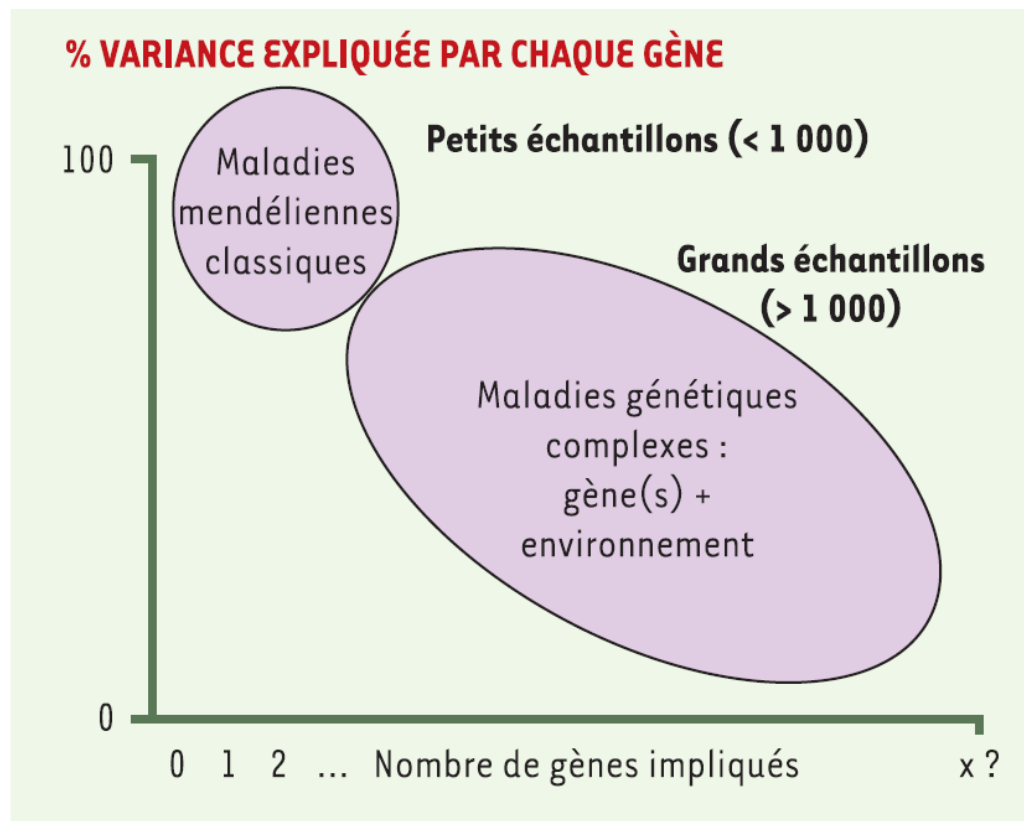
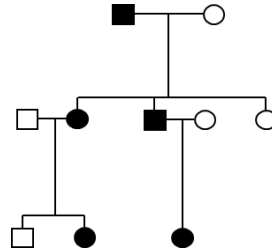
Stratégies génétiques



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

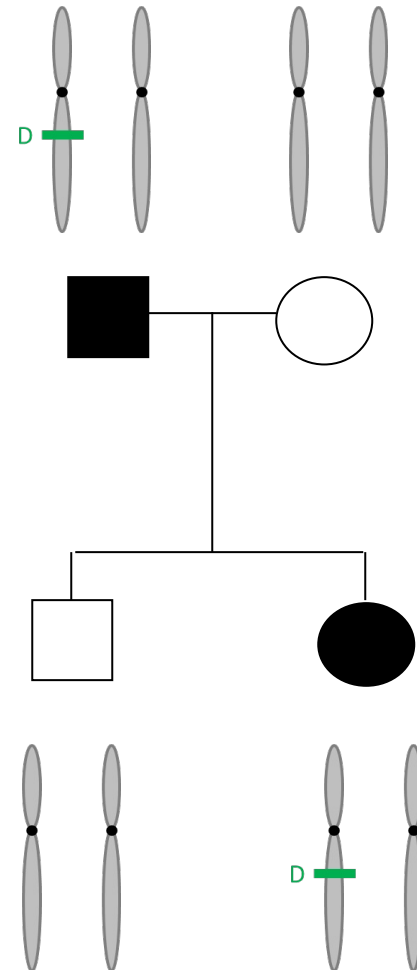
⇒ Un enfant sur 2 est atteint

Mais parfois dans une famille, on observe:

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

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

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



Les modes de transmission: autosomique récessif

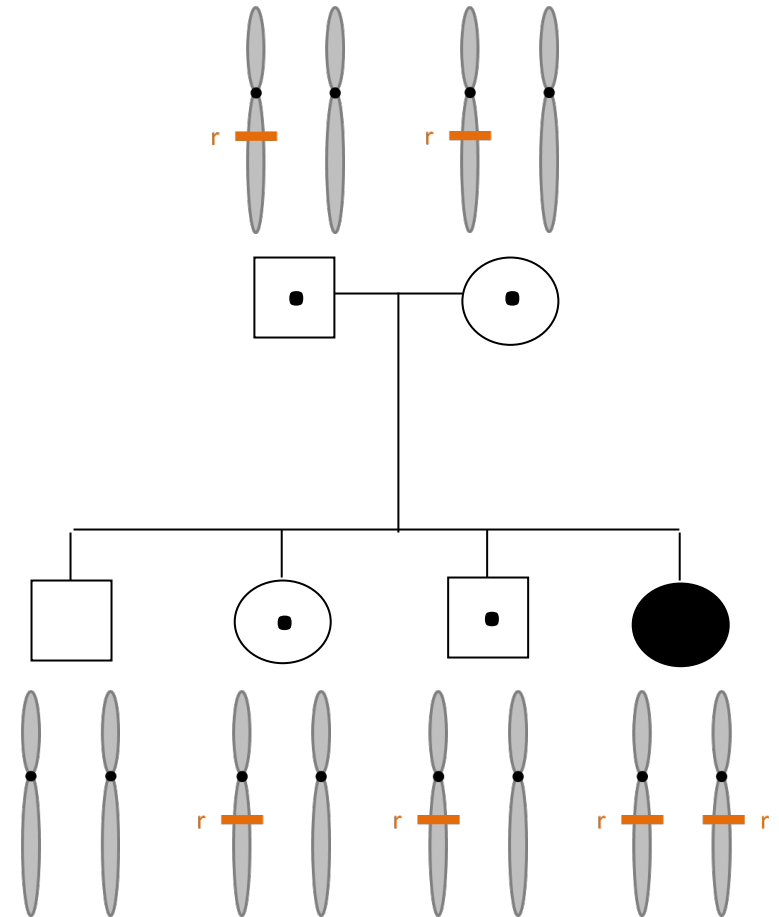
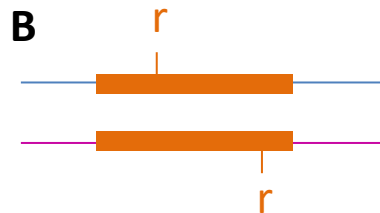
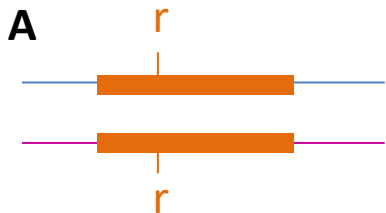
Les parents ne sont pas atteints

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

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

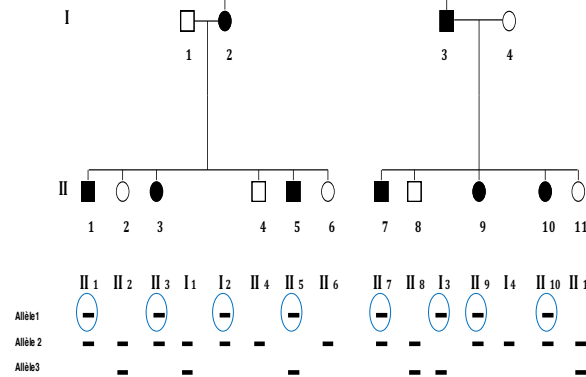
Malade
(cas index)



*Enquête
familiale*



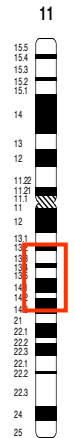
Famille



*Génotypage
Analyse de liaison*



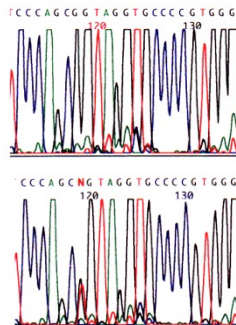
Locus



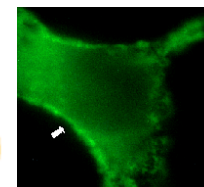
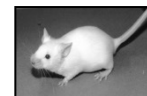
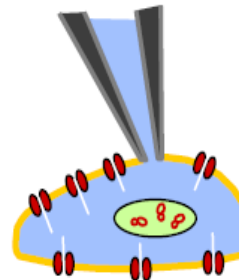
Séquençage



Gène/Mutation



Fonction



Malade
Diagnostic
moléculaire
Prise en charge
clinique

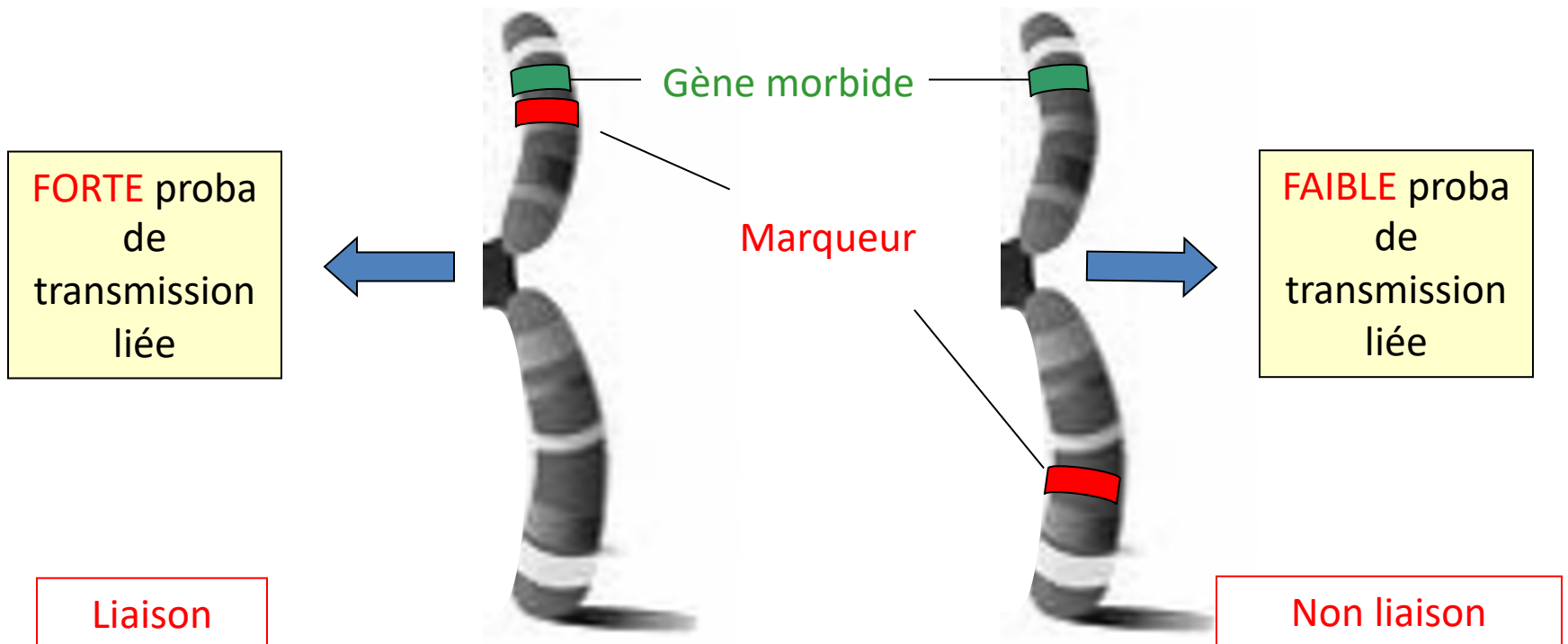
Analyse de liaison

Corréler la transmission d'un génotype avec
un phénotype

L'analyse de liaison génétique

= clonage positionnel

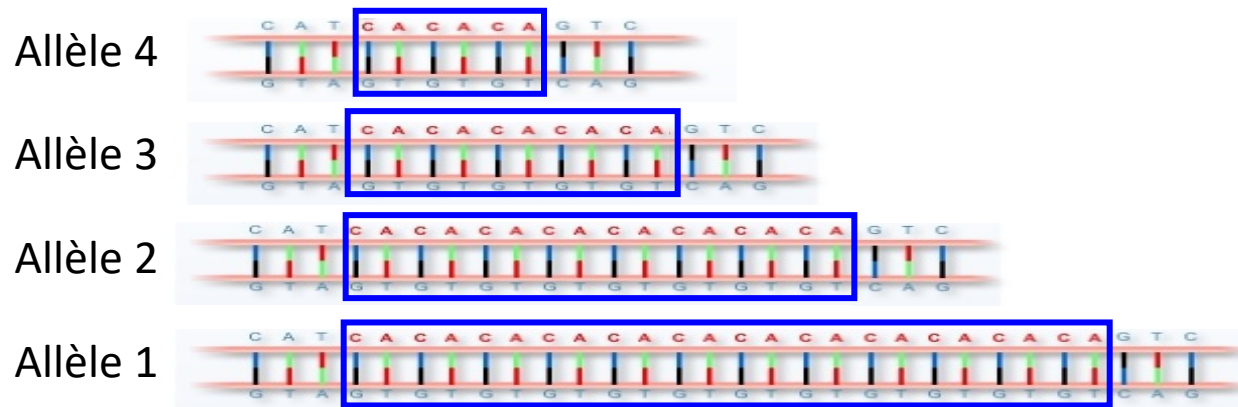
= « linkage »



Deux types de marqueurs : microsatellites ou SNPs

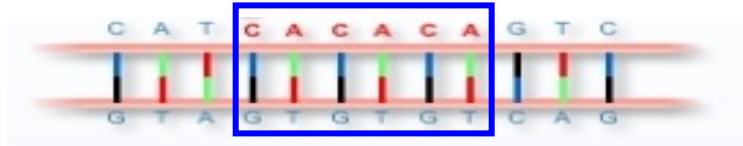
Microsatellites

- 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

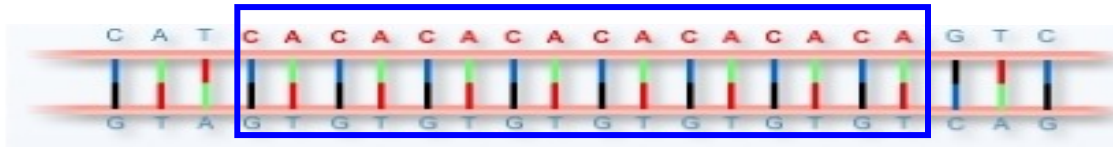
Allèle 4



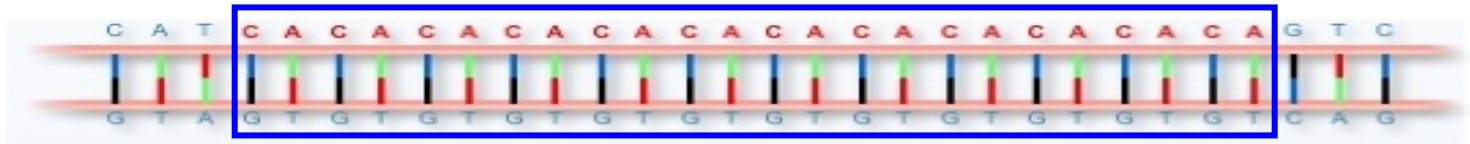
Allèle 3



Allèle 2

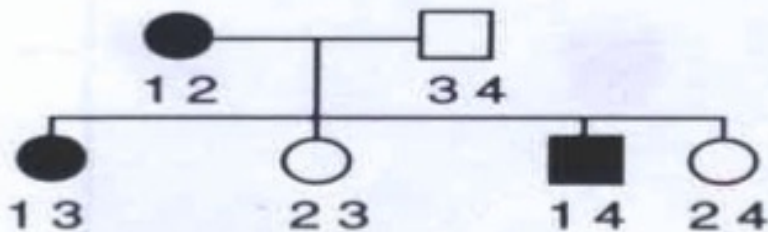


Allèle 1



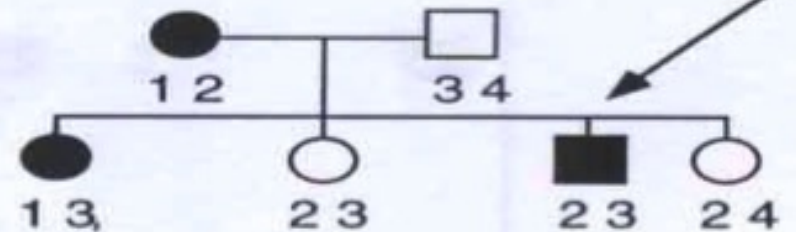
?

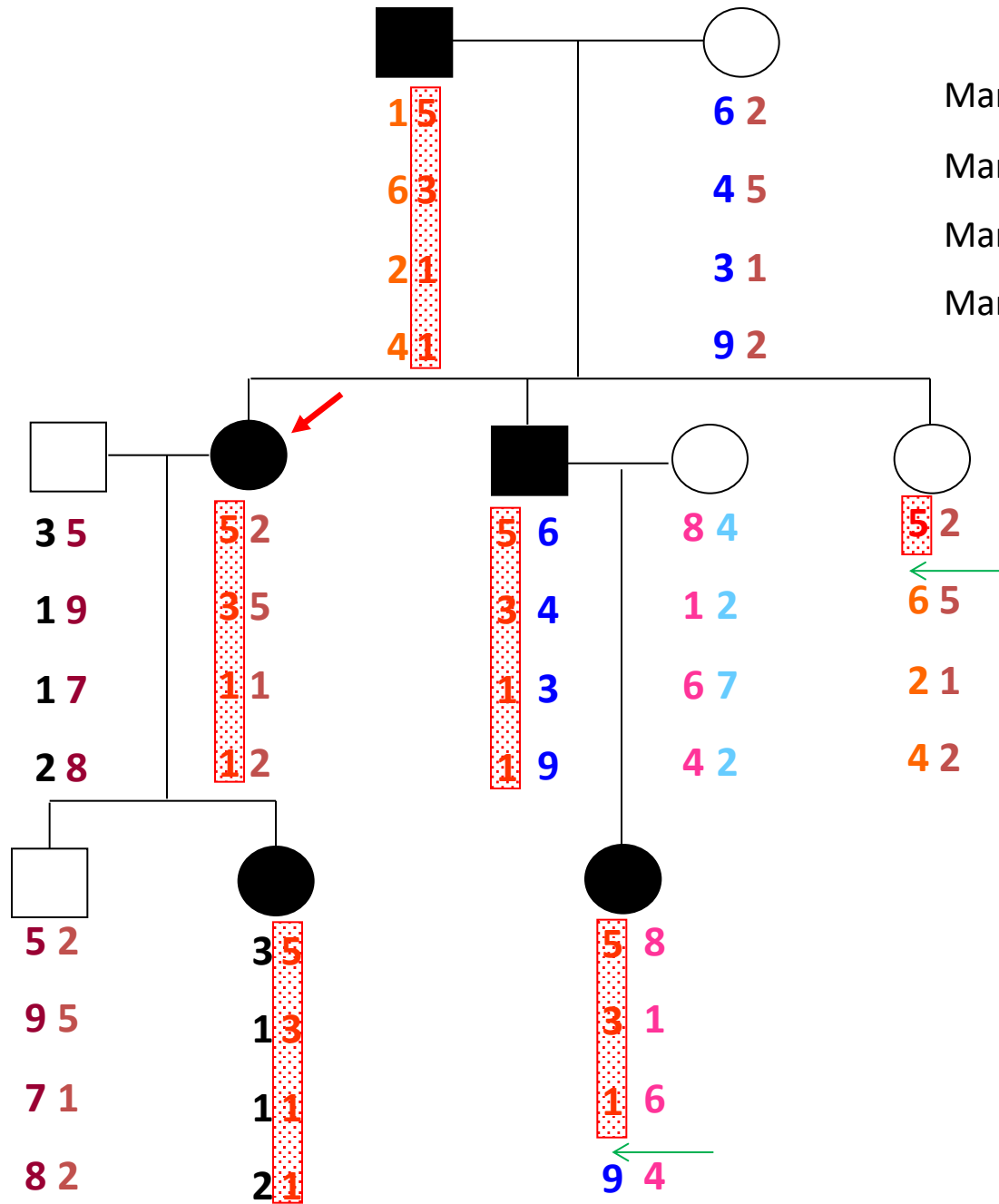
Marqueur 1



?

Marqueur 2



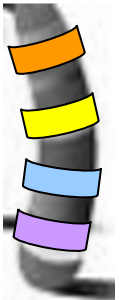


Marqueur A: D5S393

Marqueur B: D5S414

Marqueur C: D5S1979

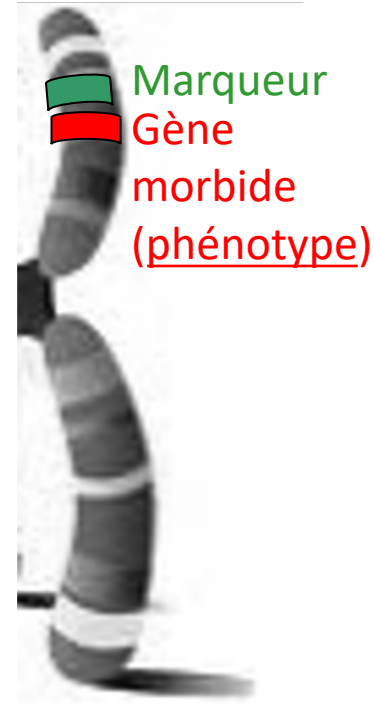
Marqueur D: D5S673



L'analyse de liaison génétique

La méthode des lod scores

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



-2

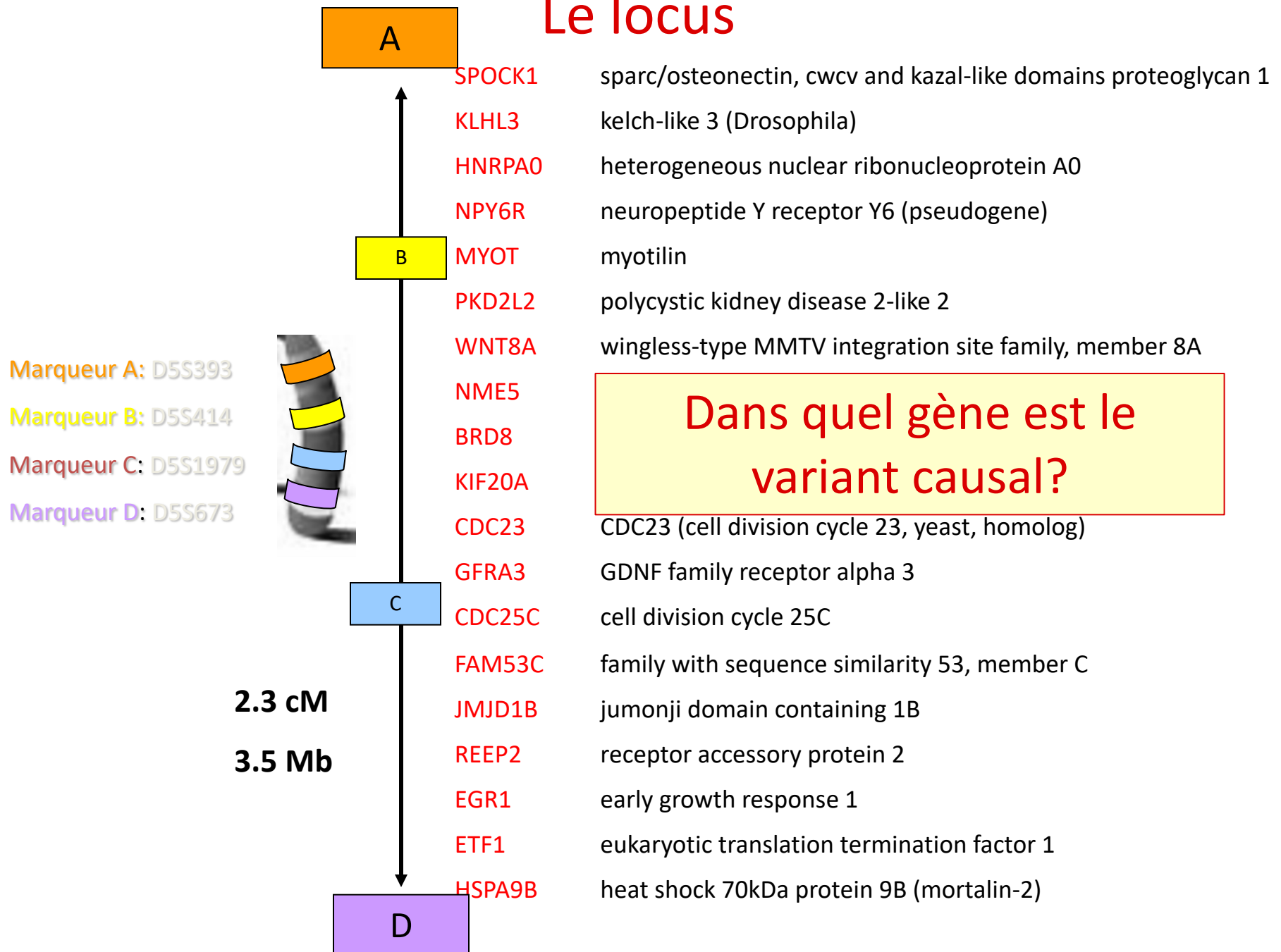
+3

Pas de liaison

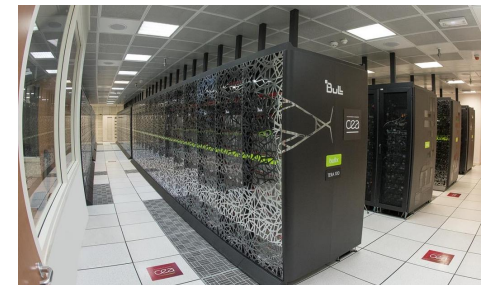
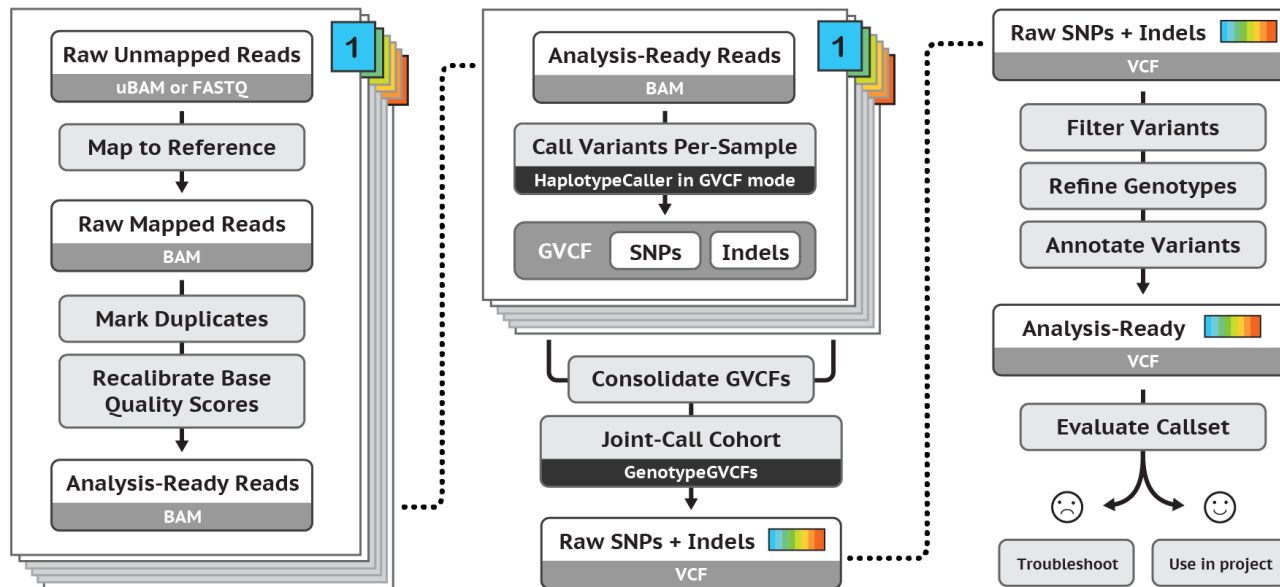
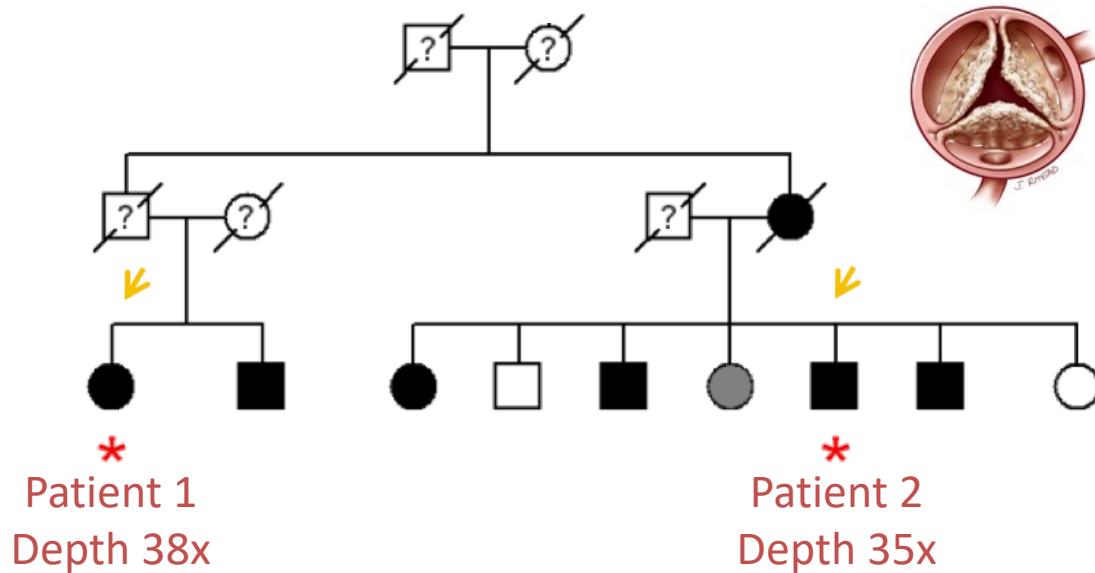
Non informatif

Liaison

Le locus

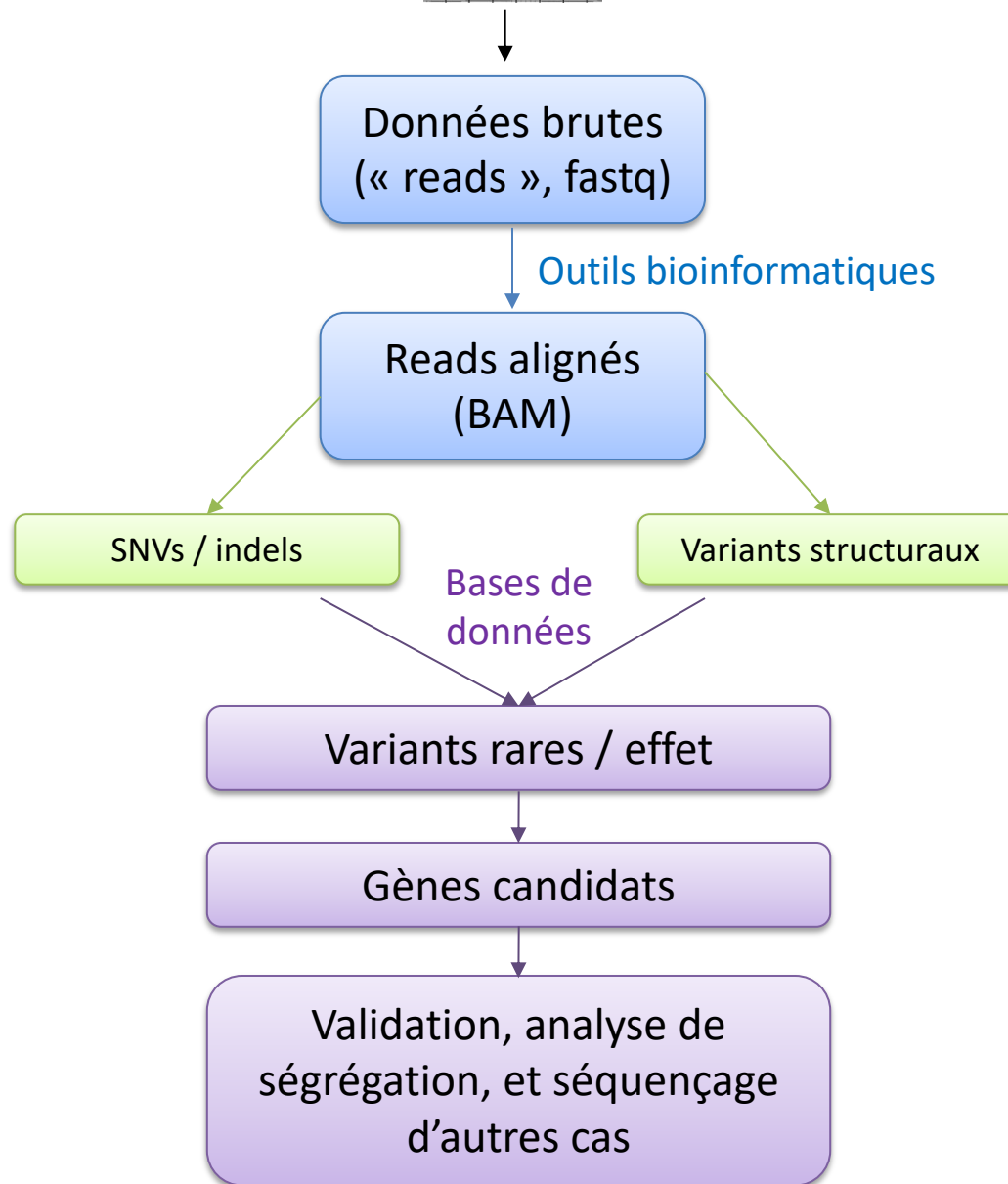
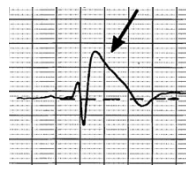


Analyse des données de séquençage

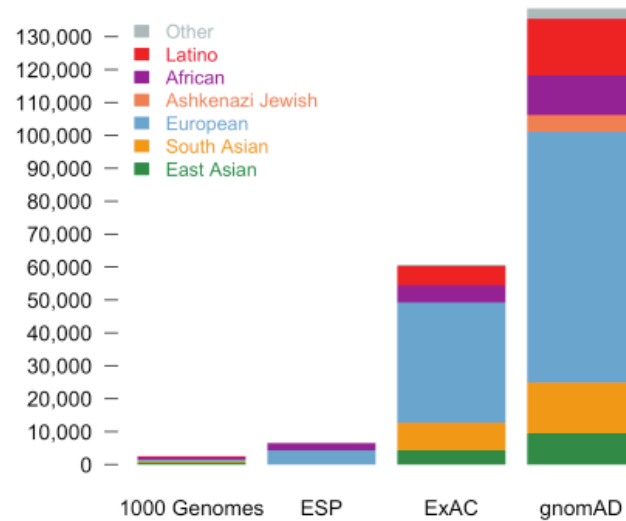


Analyses
bioinformatiques

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
Total		15,496	123,136	138,632

Variants rares

Variant: 3:38591847 G / C

Filter Status

dbSNP

Allele Frequency

Allele Count

UCSC

ClinVar

PASS

[rs45489199](#)

0.001341

146 / 108894

[3-38591847-G-C](#)

[Click to search for variant in Clinvar](#)

Genotype Quality Metrics

Site Quality Metrics

Annotations

This variant falls on 11 transcripts in 1 genes:

missense

• [SCN5A](#)

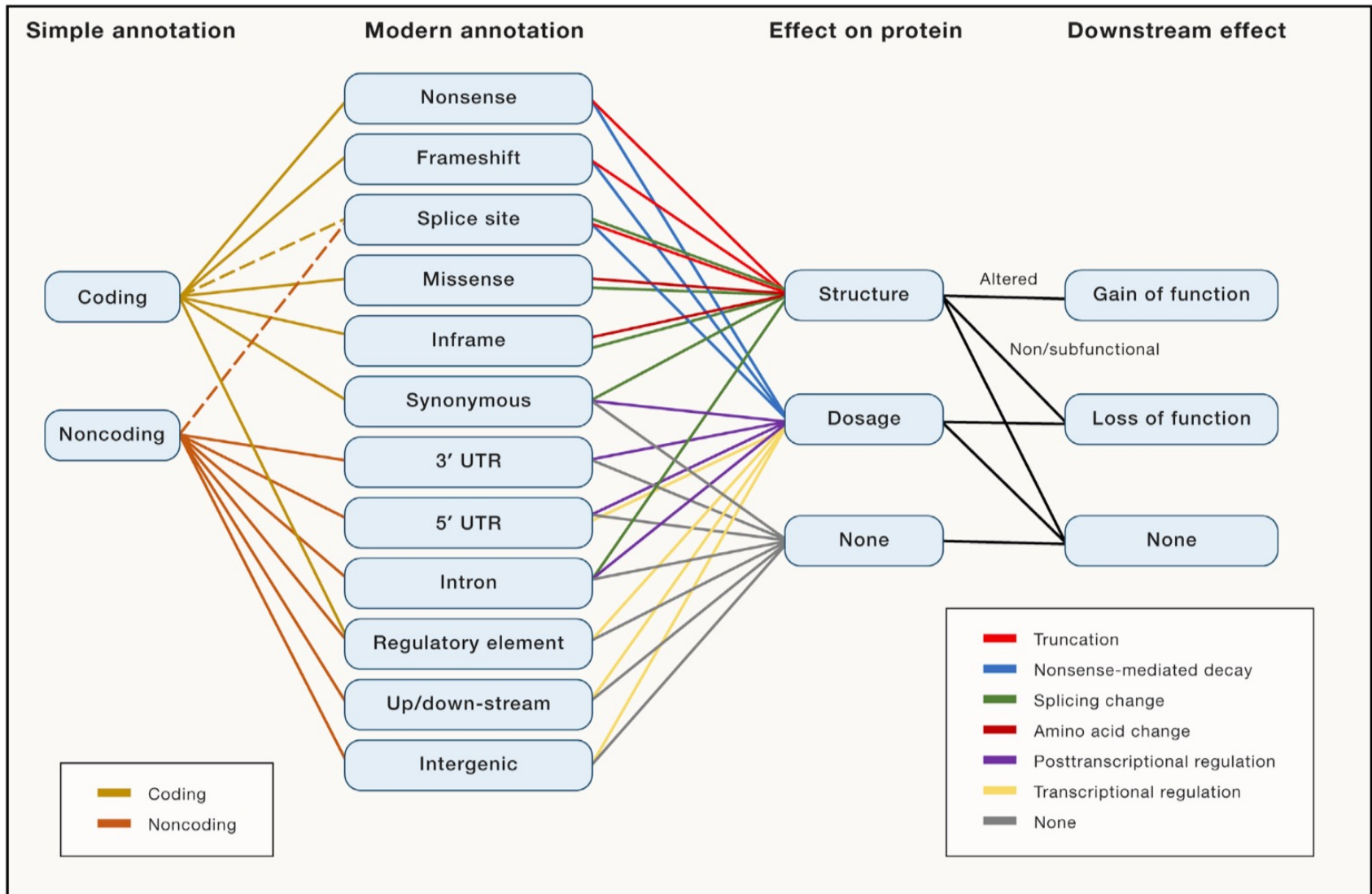
Transcripts ▾

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

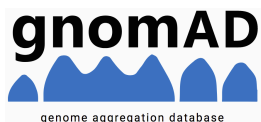
Population Frequencies

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
Total	146	108894	1	0.001341

Déterminer l'effet des variations : un challenge



Interprétation des données de séquençage



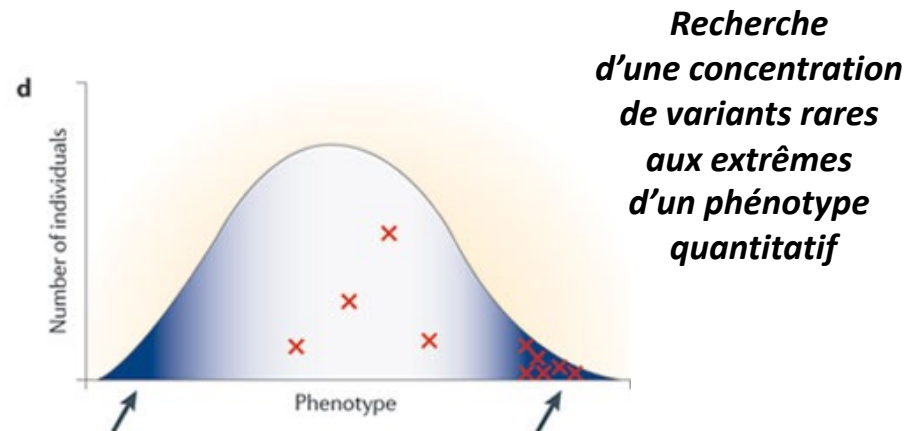
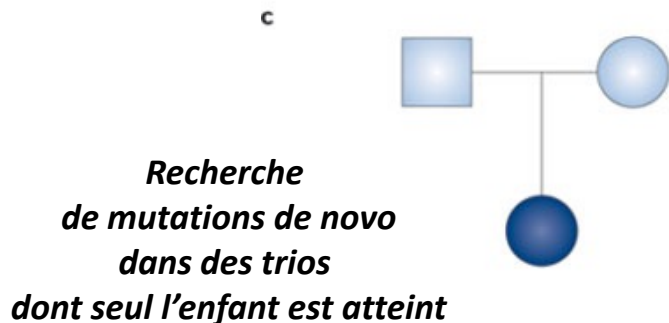
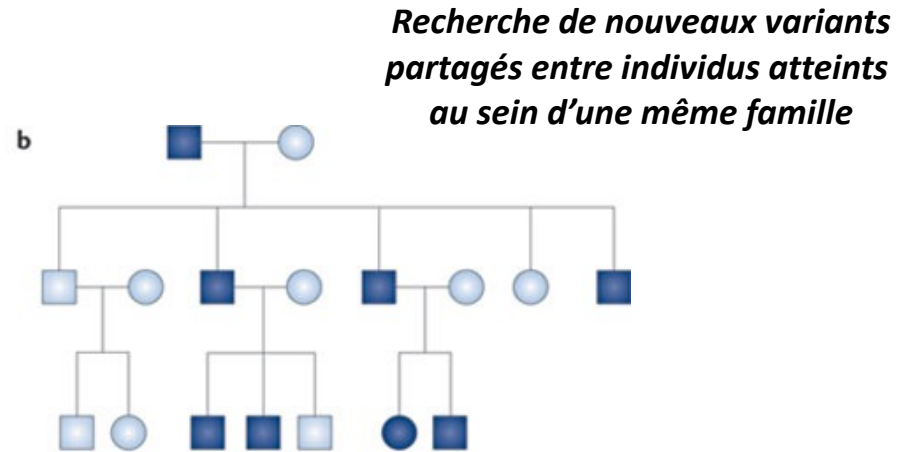
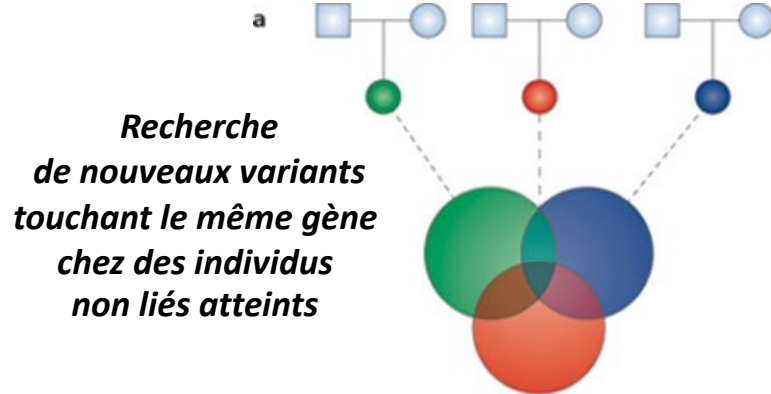
	Patient 1	Patient 2
# total substitutions + indels	4 978 101	4 867 920
# variants rares (<0,1% Non-Finnish Europeans + contrôles internes)	58 468	52 972
# variants rares + csq (exons / épissage) SO:0001818 SO:0001568	450	438
# variants rares + faux-sens SO:0001583	276	266
# variants rares + stop SO:0001587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



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

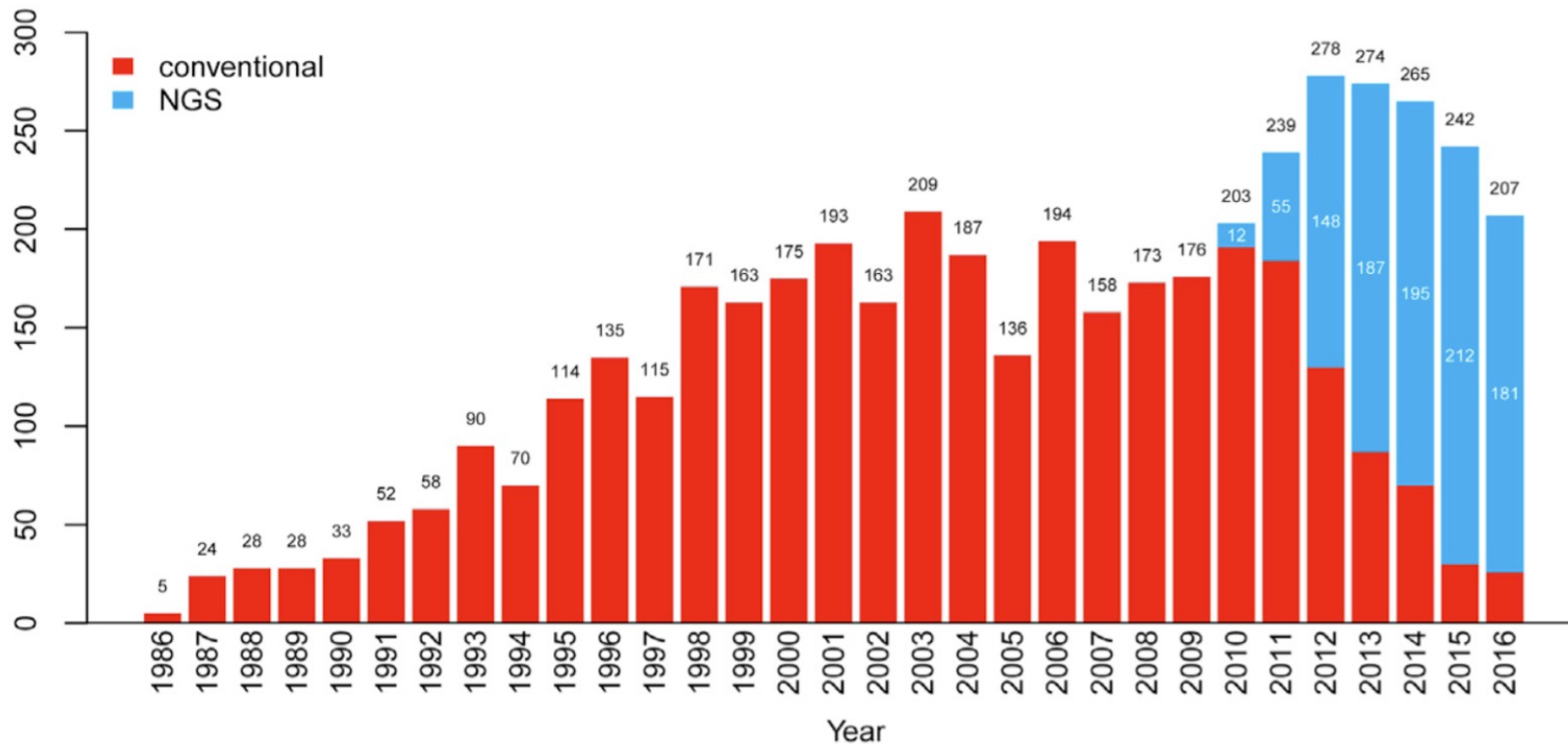
Sarah B. Ng¹, Emily H. Turner¹, Peggy D. Robertson¹, Steven D. Flygare¹, Abigail W. Bigham², Choli Lee¹, Tristan Shaffer¹, Michelle Wong¹, Arindam Bhattacharjee⁴, Evan E. Eichler^{1,3}, Michael Bamshad², Deborah A. Nickerson¹ & Jay Shendure¹

		FSS24895	FSS24895 FSS10208	FSS24895 FSS10208 FSS10066	FSS24895 FSS10208 FSS10066 FSS22194	Any 3 of 4 FSS24895 FSS10208 FSS10066 FSS22194
Number of genes in which each affected has at least one...	Non-synonymous cSNP, splice site variant or coding indel (NS/SS/I)	4,510	3,284	2,765	2,479	3,768
	NS/SS/I not in dbSNP	513	128	71	53	119
	NS/SS/I not in eight HapMap exomes	799	168	53	21	160
	NS/SS/I neither in dbSNP nor eight HapMap exomes	360	38	8	1 (MYH3)	22
	...And predicted to be damaging	160	10	2	1 (MYH3)	3

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

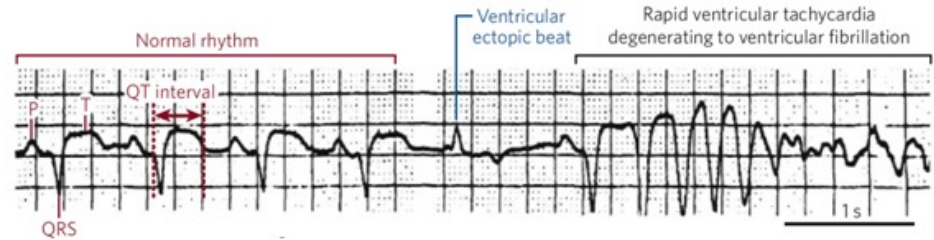
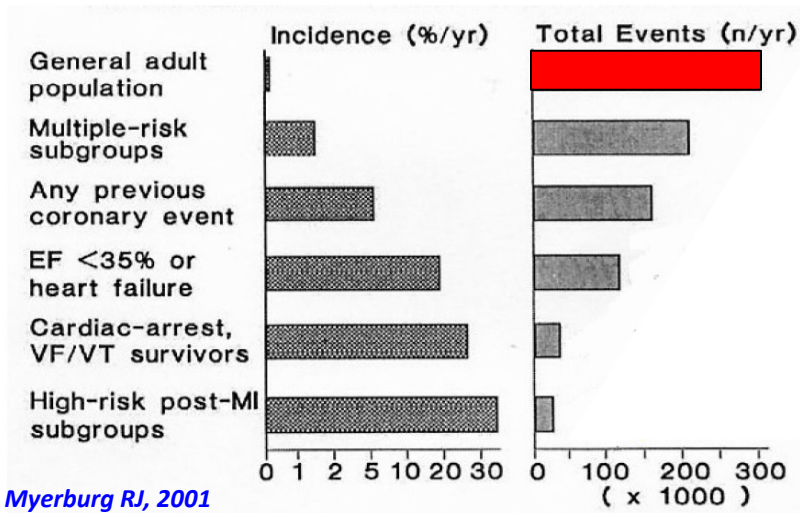
Disorder	Inheritance	Gene Identified	Scope	References
Congenital chloride diarrhea	Recessive	<i>SLC26A3</i>	Exome	Choi <i>et al.</i> [16]
Miller syndrome	Recessive	<i>DHODH</i>	Exome	Ng <i>et al.</i> [14]
Charcot-Marie-Tooth neuropathy	Recessive	<i>SH3TC2</i>	Genome	Lupski <i>et al.</i> [20]
Metachondromatosis	Dominant	<i>PTPN11</i>	Genome	Sobreira <i>et al.</i> [23]
Schinz-Gledion syndrome	Dominant	<i>SETBP1</i>	Exome	Holschen <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	Wortheley <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>POP1</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)

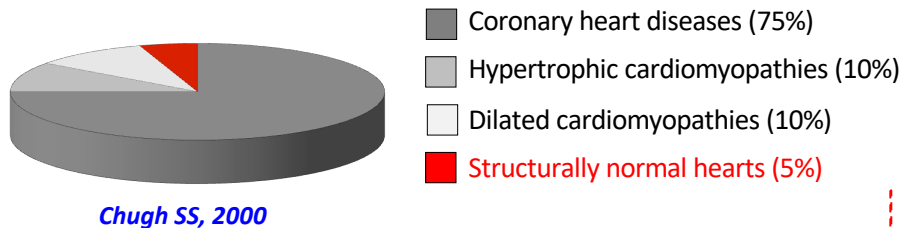


Arythmies cardiaques

Sudden Cardiac Death



Ventricular Fibrillation: common mechanism in 80% of SCD cases

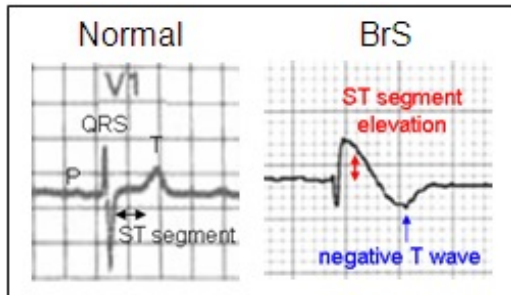


Idiopathic ventricular fibrillation

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

Syndrome de Brugada

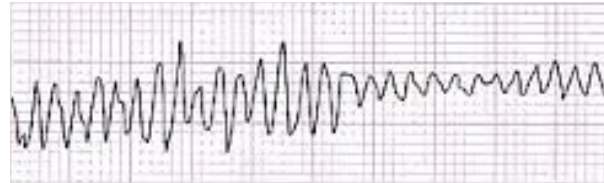
ST-segment elevation (> 2 mm) in the right precordial ECG leads



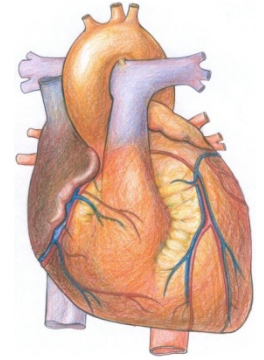
Brugada et al, 1992

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

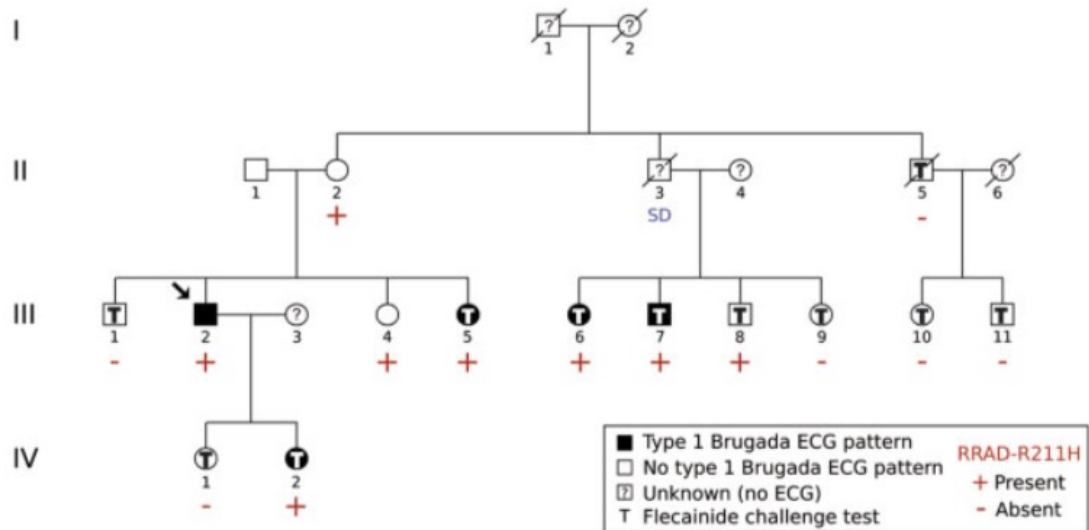
➤ **La majorité des cas sont encore inexpliqués**

Séquençage d'exome – syndrome de Brugada

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

Belbachir*, Portero* et al
European Heart Journal, 2019

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

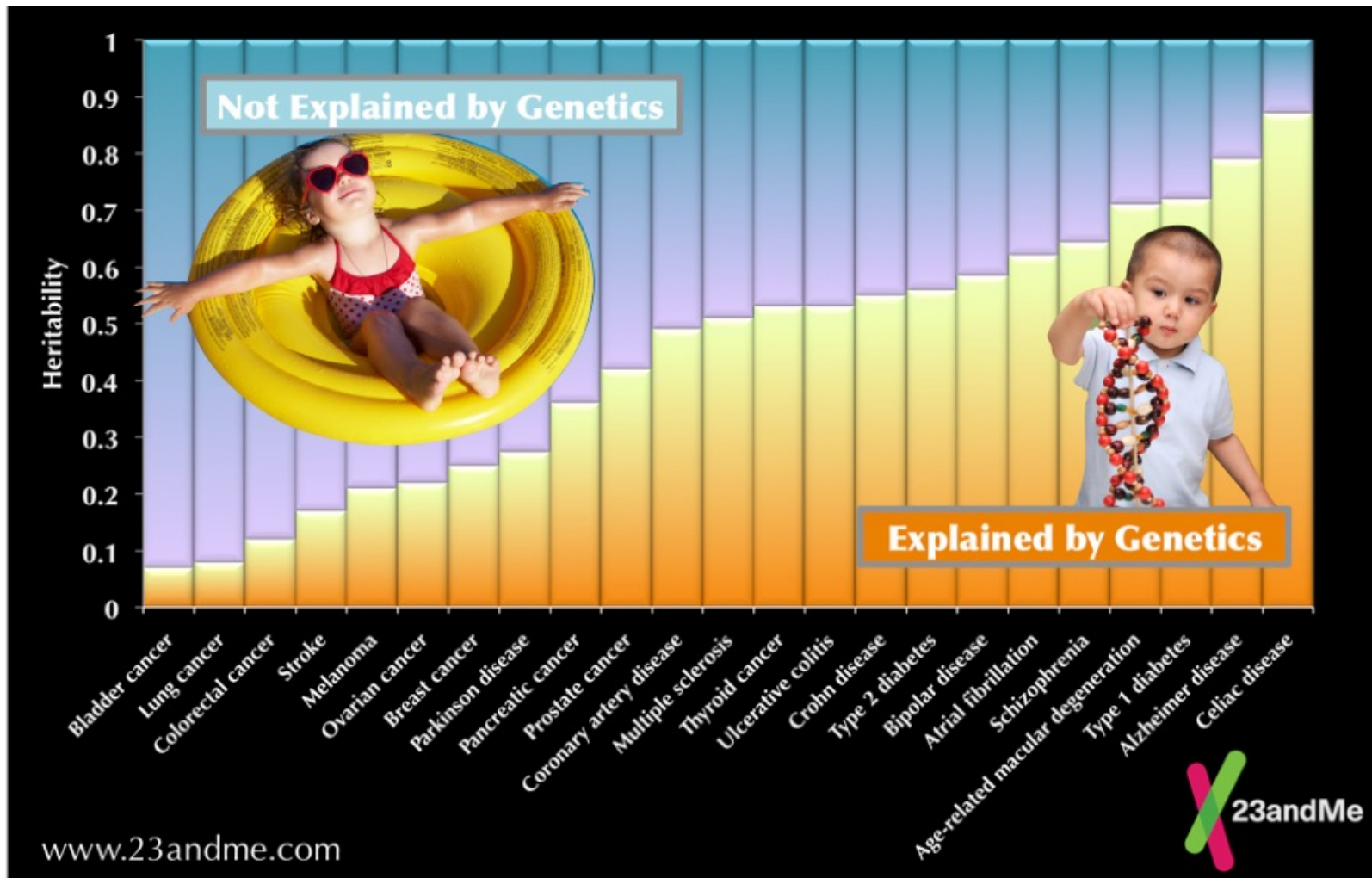


Exome sequencing (patient III:2)	
Coding +/- 8 bp*	22,562
Functional**	10,412
Rare in ExAC***	217
In shared genomic regions	9
Validated and shared by the 5 BrS patients	6 missense variants (<i>RRAD</i> , <i>FHOD1</i> , <i>H6PD</i> , <i>SNAP23</i> , two in <i>SPTBN5</i>)
Predicted deleterious (SIFT and PolyPhen)	<i>RRAD</i> (R211H)

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

Cas 35 %

Témoins 15 %

Association

Marqueur 2

Cas 20 %

Témoins 20 %

Pas d'association

Etudes d'association chez l'homme

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

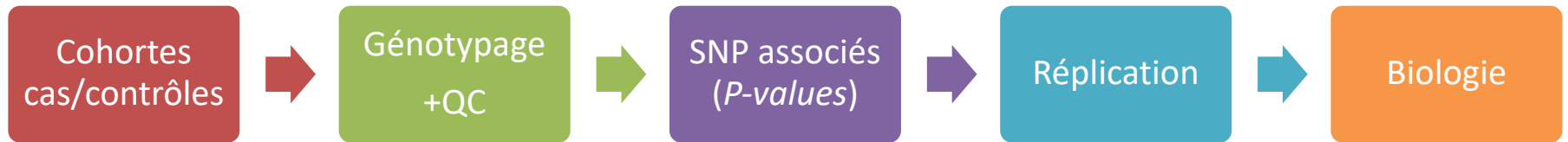
Avancées dans les
technologies de génotypage
(SNP)

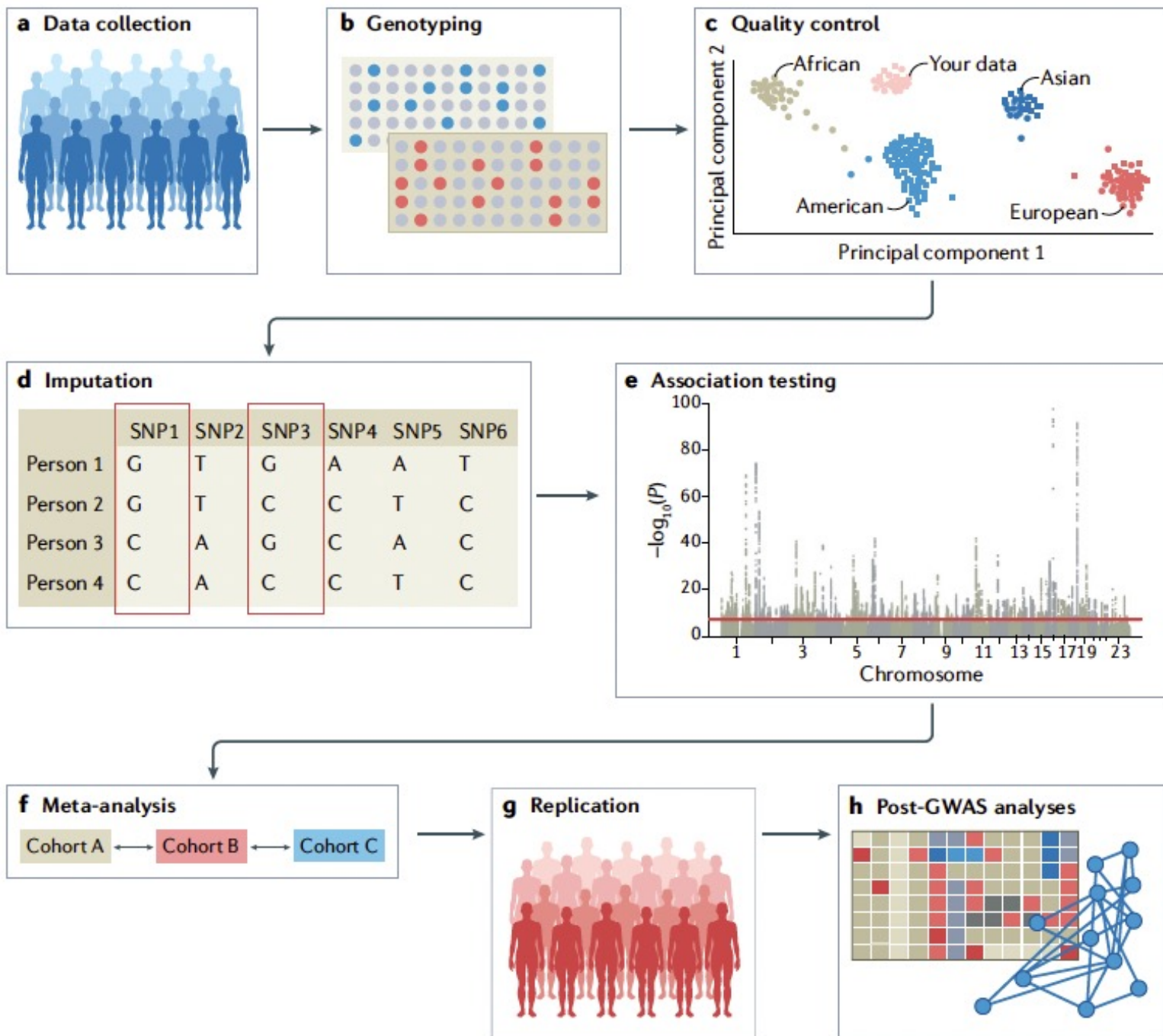
Grandes bio-collections
de patients

Genome-wide association studies
= (GWAS)

```
graph LR; A[Meilleure compréhension de la variation du génome humain (HapMap, 1000G)] --> D[Genome-wide association studies = (GWAS)]; B[Avancées dans les technologies de génotypage (SNP)] --> D; C[Grandes bio-collections de patients] --> D;
```

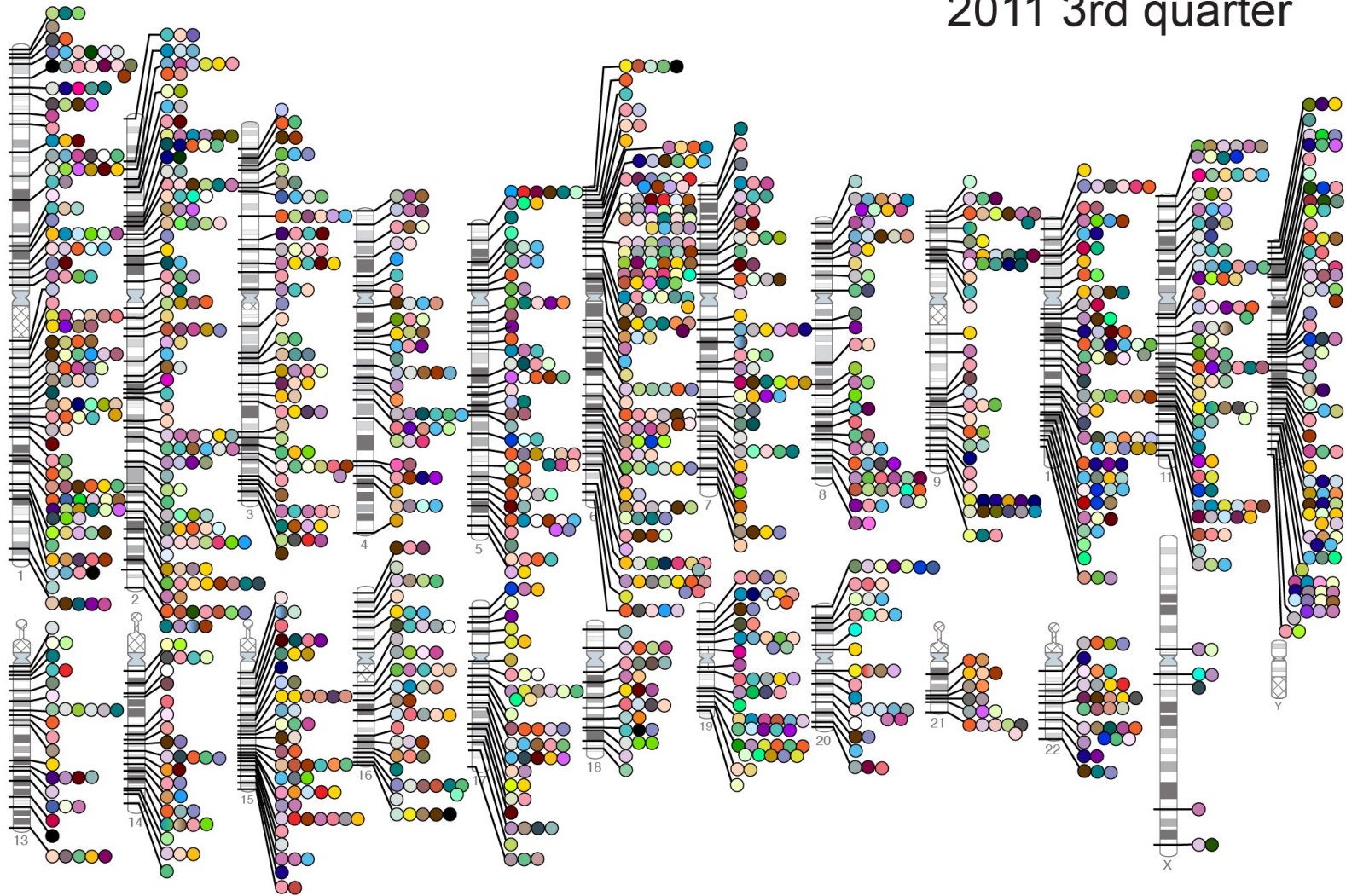
Aperçu de l'approche GWAS





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

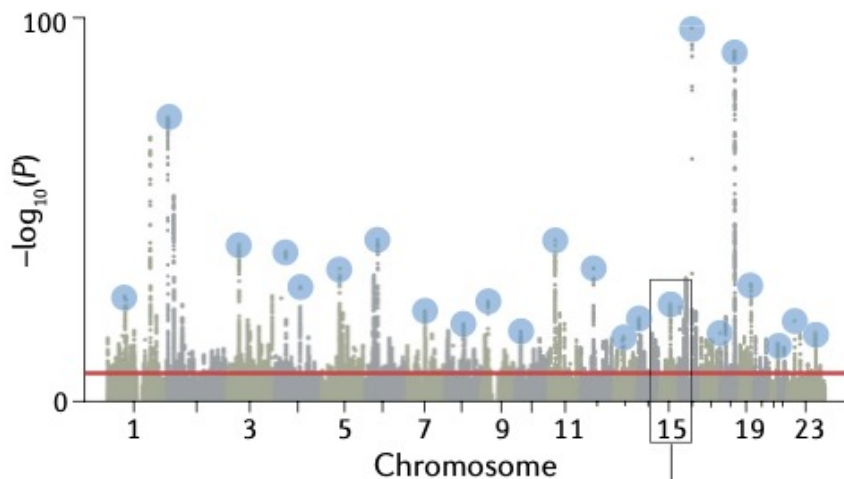
2011 3rd quarter



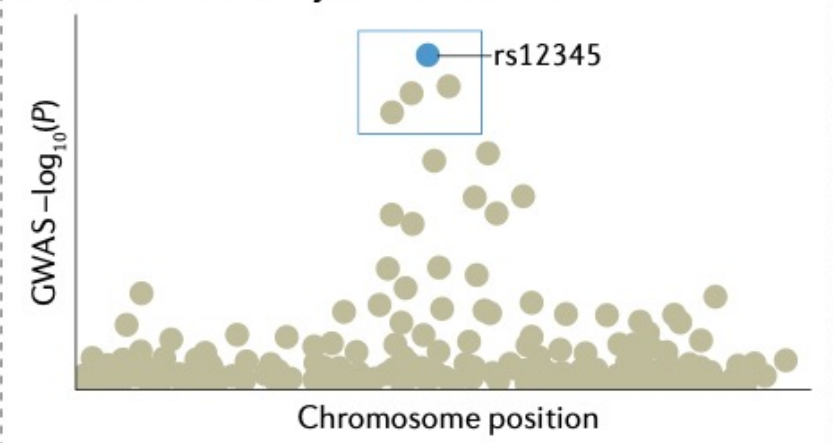
Abdominal aortic aneurysm	Coffee consumption	Hepatitis B vaccine response	Neuroblastoma	Response to metaformin
Acute lymphoblastic leukemia	Cognitive function	Hepatocellular carcinoma	Nicotine dependence	Response to statin therapy
Adhesion molecules	Conduct disorder	Hirschsprung's disease	Obesity	Restless legs syndrome
Adiponectin levels	Colorectal cancer	HIV-1 control	Open angle glaucoma	Retinal vascular caliber
Age-related macular degeneration	Corneal thickness	Hodgkin's lymphoma	Open personality	Retinol levels
AIDS progression	Coronary disease	Homocysteine levels	Optic disc parameters	Rheumatoid arthritis
Alcohol dependence	Cortical thickness	HPV seropositivity	Osteoarthritis	Ribavirin-induced anemia
Alopecia areata	Creutzfeldt-Jakob disease	Hypospadias	Osteoporosis	Schizophrenia
Alzheimer disease	Crohn's disease	Idiopathic pulmonary fibrosis	Otosclerosis	Serum metabolites
Amyloid A levels	Crohn's disease and celiac disease	IFN-related cytopeni	Other metabolic traits	Skin pigmentation
Amyotrophic lateral sclerosis	Cutaneous nevi	IgA levels	Ovarian cancer	Smoking behavior
Angiotensin-converting enzyme activity	Cystic fibrosis severity	IgE levels	Pancreatic cancer	Speech perception
Ankylosing spondylitis	Dermatitis	Inflammatory bowel disease	Pain	Sphingolipid levels
Arterial stiffness	DHEA-s levels	Insulin-like growth factors	Paget's disease	Statin-induced myopathy
Asparagus anosmia	Diabetic retinopathy	Intracranial aneurysm	Panic disorder	Stevens-Johnson syndrome
Asthma	Dilated cardiomyopathy	Iris color	Parkinson's disease	Stroke
Atherosclerosis in HIV	Drug-induced liver injury	Iron status markers	Periodontitis	Sudden cardiac arrest
Atrial fibrillation	Drug-induced liver injury (amoxicillin-clavulanate)	Ischemic stroke	Peripheral arterial disease	Suicide attempts
Attention deficit hyperactivity disorder	Endometrial cancer	Juvenile idiopathic arthritis	Personality dimensions	Systemic lupus erythematosus
Autism	Endometriosis	Keloid	Phosphatidylcholine levels	Systemic sclerosis
Basal cell cancer	Eosinophil count	Kidney stones	Phosphorus levels	T-tau levels
Behcet's disease	Eosinophilic esophagitis	LDL cholesterol	Photic sneeze	Tau AB1-42 levels
Bipolar disorder	Epirubicin-induced leukopenia	Leprosy	Phytosterol levels	Telomere length
Biliary atresia	Erectile dysfunction and prostate cancer treatment	Leptin receptor levels	Platelet count	Testicular germ cell tumor
Bilirubin	Erythrocyte parameters	Liver enzymes	Polycystic ovary syndrome	Thyroid cancer
Bitter taste response	Esophageal cancer	Longevity	Primary biliary cirrhosis	Thyroid volume
Birth weight	Essential tremor	LP (a) levels	Primary sclerosing cholangitis	Tooth development
Bladder cancer	Exfoliation glaucoma	LpPLA(2) activity and mass	PR interval	Total cholesterol
Bleomycin sensitivity	Eye color traits	Lung cancer	Progranulin levels	Triglycerides
Blond or brown hair	F cell distribution	Magnesium levels	Progressive supranuclear palsy	Tuberculosis
Blood pressure	Fibrinogen levels	Malaria	Prostate cancer	Type 1 diabetes
Blue or green eyes	Folate pathway vitamins	Male pattern baldness	Protein levels	Type 2 diabetes
BMI, waist circumference	Follicular lymphoma	Mammographic density	PSA levels	Ulcerative colitis
Bone density	Fuch's corneal dystrophy	Matrix metalloproteinase levels	Psoriasis	Urate
Breast cancer	Freckles and burning	MCP-1	Psoriatic arthritis	Urinary albumin excretion
Butyrylcholinesterase levels	Gallstones	Melanoma	Pulmonary funct. COPD	Urinary metabolites
C-reactive protein	Gastric cancer	Menarche & menopause	QRS interval	Uterine fibroids
Calcium levels	Glioma	Meningioma	QT interval	Venous thromboembolism
Cardiac structure/function	Glycemic traits	Meningococcal disease	Quantitative traits	Ventricular conduction
Cardiovascular risk factors	Graves disease	Metabolic syndrome	Recombination rate	VEGF levels
Carnitine levels	Hair color	Migraine	Red vs.non-red hair	Vertical cup-disc ratio
Carotenoid/tocopherol levels	Hair morphology	Moyamoya disease	Refractive error	Vitamin B12 levels
Carotid atherosclerosis	Handedness in dyslexia	Multiple sclerosis	Renal cell carcinoma	Vitamin D insufficiency
Celiac disease	HDL cholesterol	Myeloproliferative neoplasms	Renal function	Vitamin E levels
Celiac disease and rheumatoid arthritis	Heart failure	Myopia (pathological)	Response to antidepressants	Vitiligo
Cerebral atrophy measures	Heart rate	N-glycan levels	Response to antipsychotic therapy	Warfarin dose
Chronic lymphocytic leukemia	Height	Narcolepsy	Response to carbamazepine	Weight
Chronic myeloid leukemia	Hemostasis parameters	Nasopharyngeal cancer	Response to clopidogrel therapy	White cell count
Cleft lip/palate	Hepatic steatosis	Natriuretic peptide levels	Response to hepatitis C treat	White matter hyperintensity
	Hepatitis		Response to interferon beta therapy	YKL-40 levels

Post-GWAS

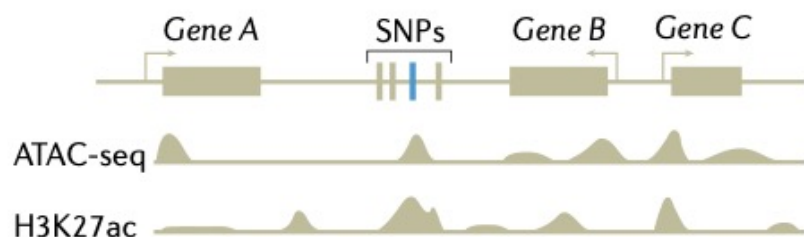
a What are the associated loci?



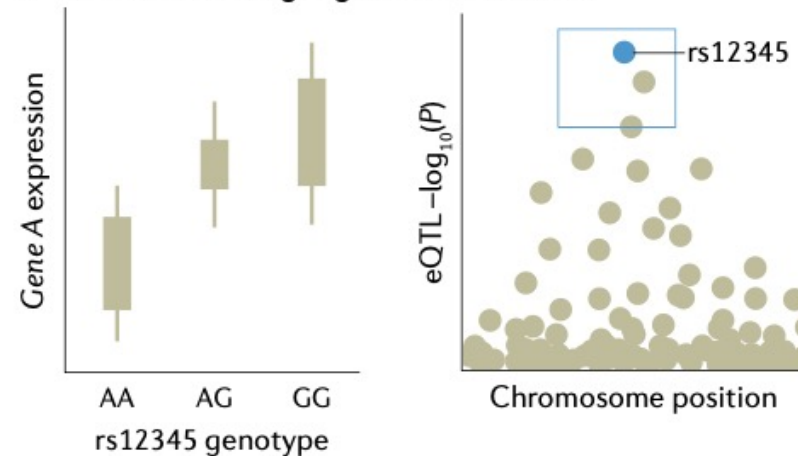
b What are the likely causal variants?



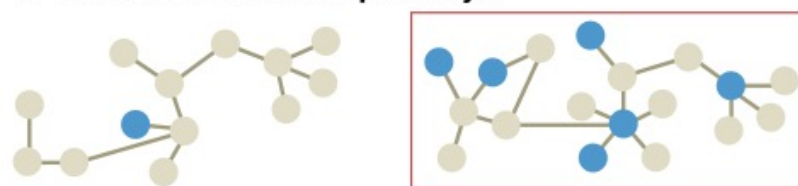
c What are the epigenomic effects of variants?



d What are the target genes in the locus?

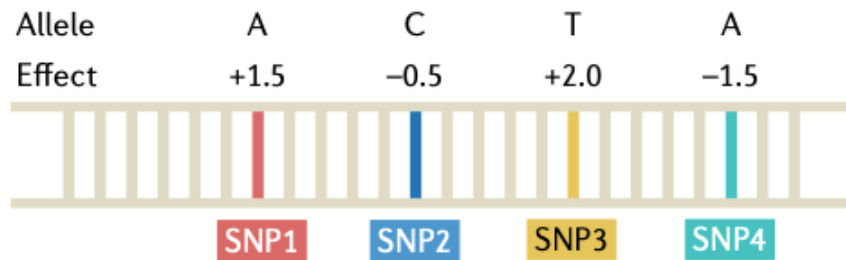


e What are the affected pathways?



Polygenic Risk Scores

① GWAS summary statistics



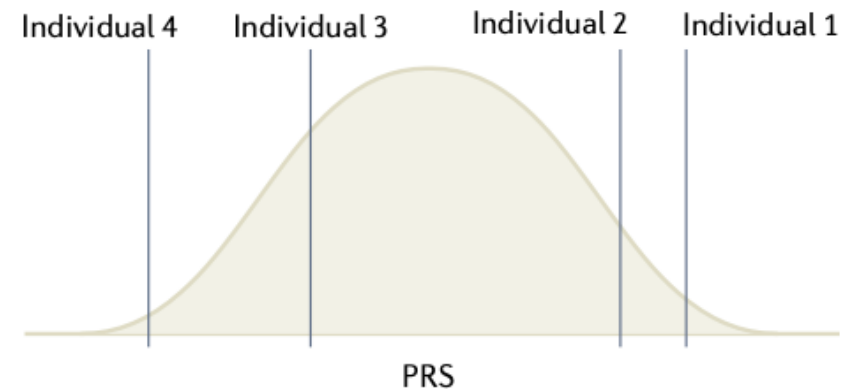
② Genotype data

	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

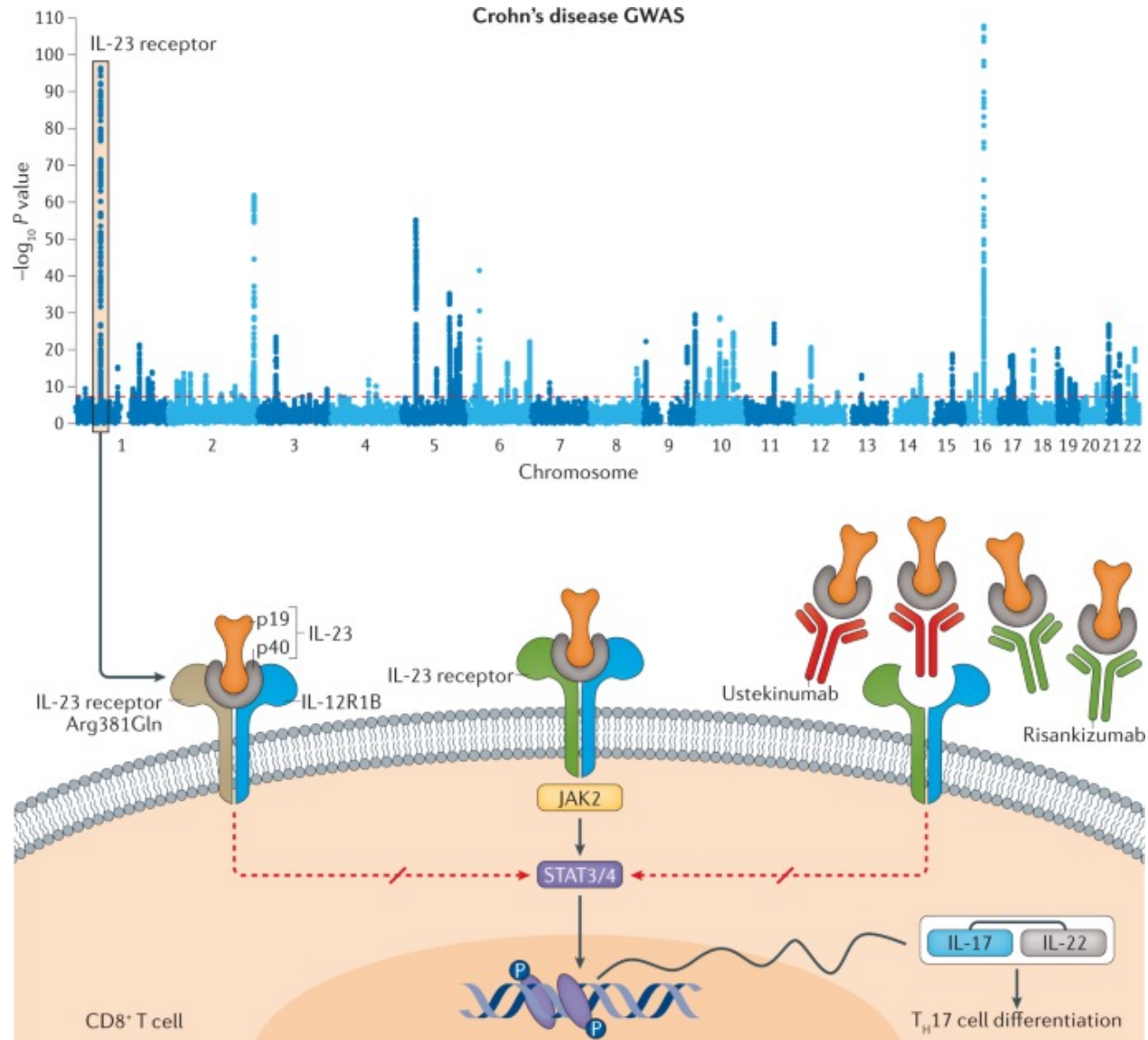
③ Polygenic risk score

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

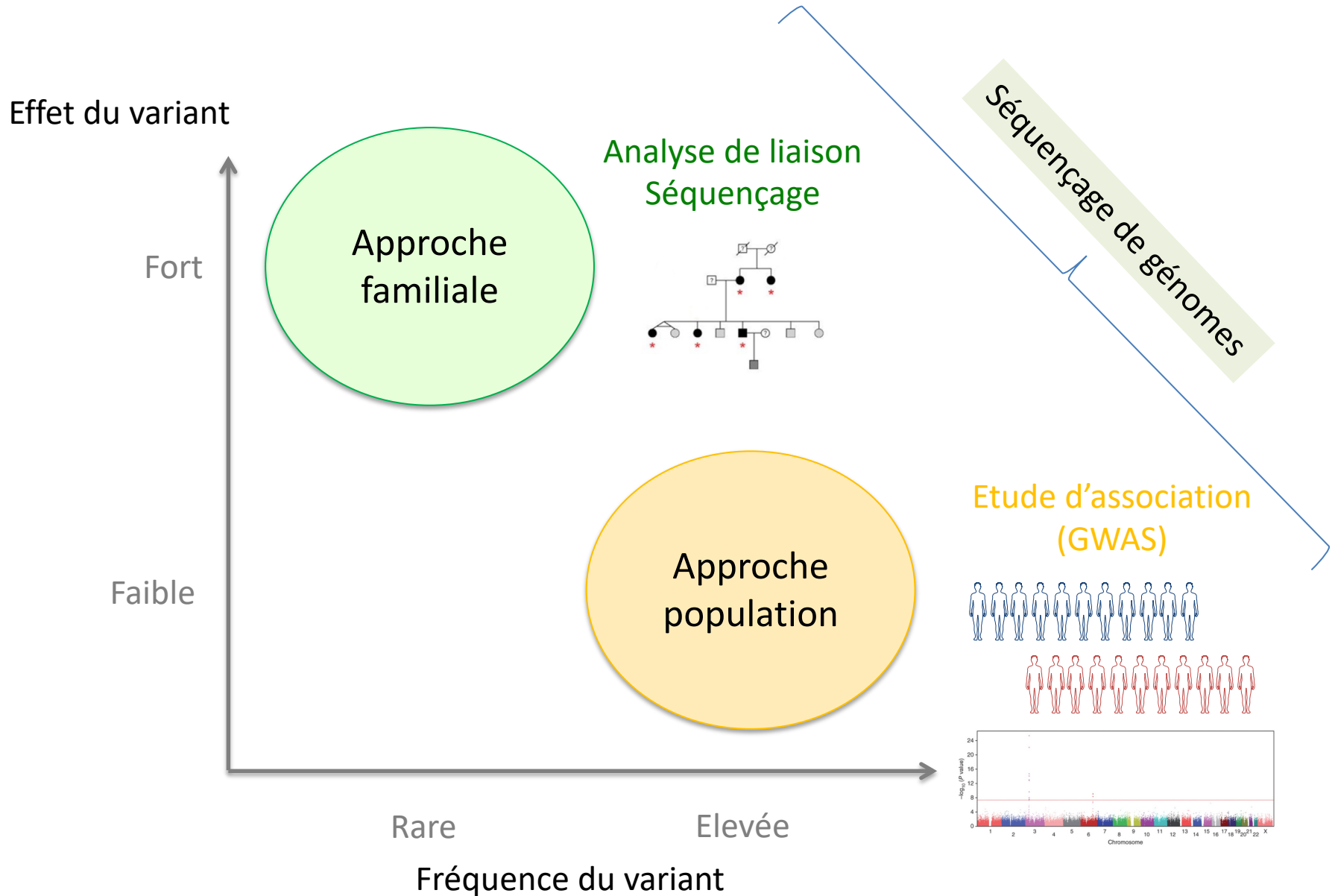
④ PRS distribution



Traitements



Stratégies génétiques



Questions

