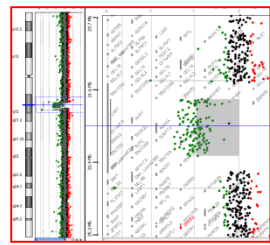
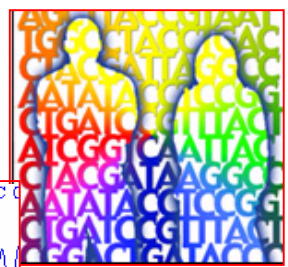
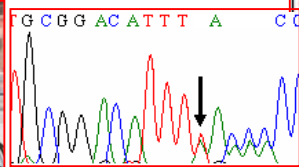
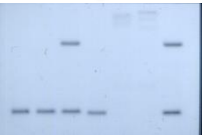
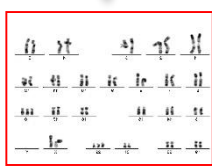
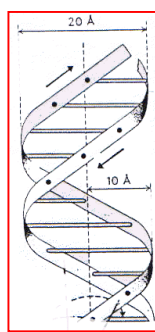
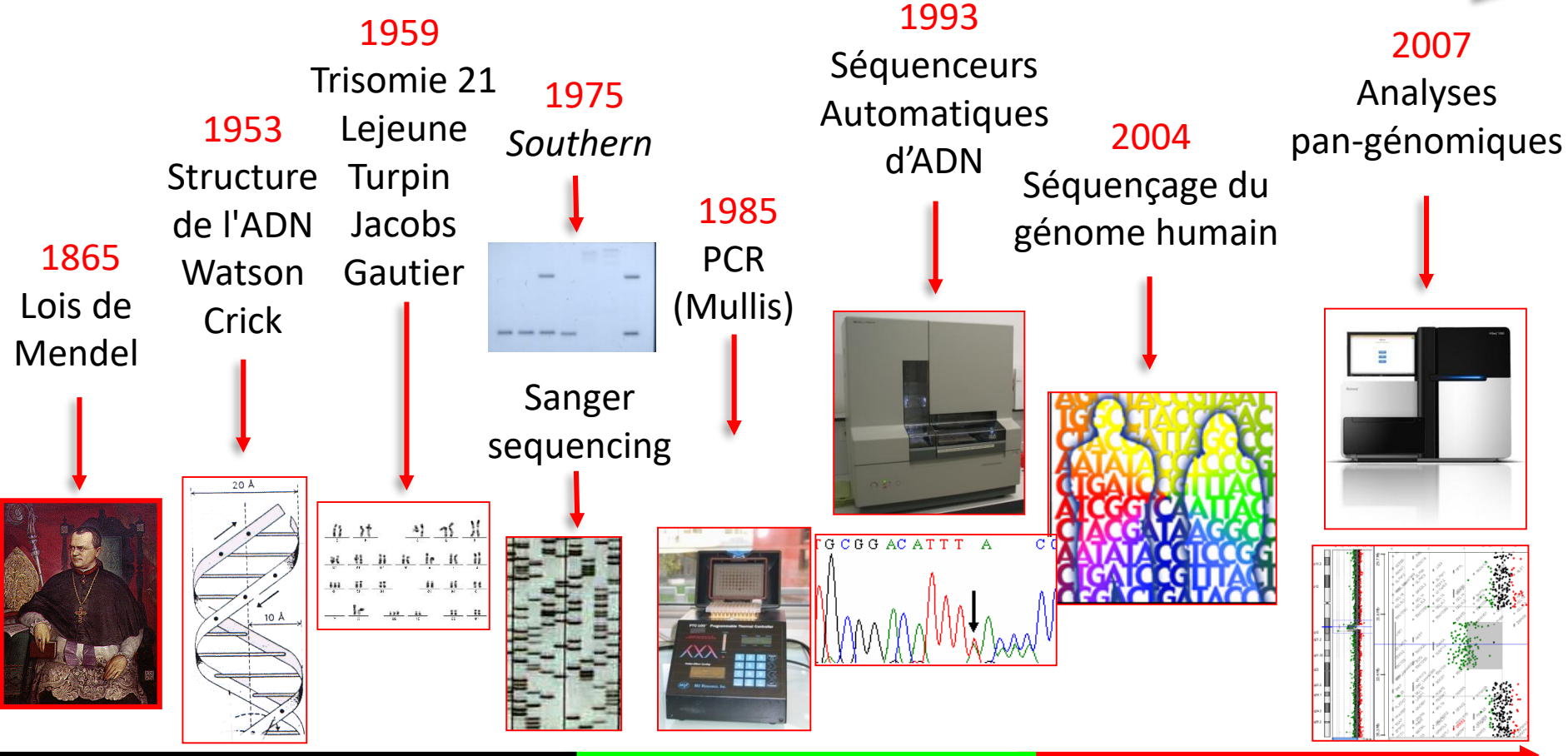
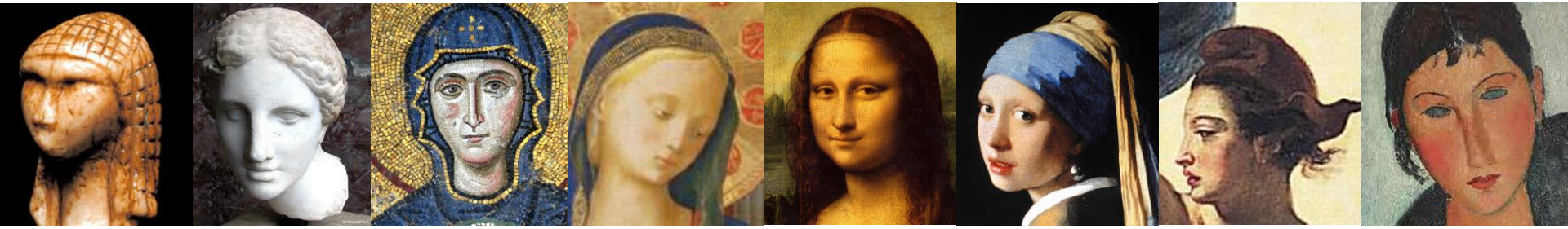


# Organisation du génome

## Variabilité du génome humain

**Formation à la génétique**  
**Académie de Normandie**

Pascale Saugier-Veber  
François Lecoquierre



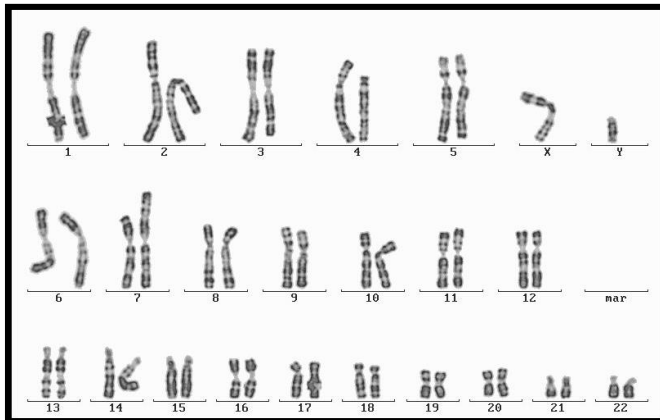
**1865 – 1975**  
Ère pré-moléculaire

**1975 – 2004**  
Ère moléculaire

**2004 –**  
Ère génomique

1975

## Caryotype



Actuellement

## Séquençage du génome entier



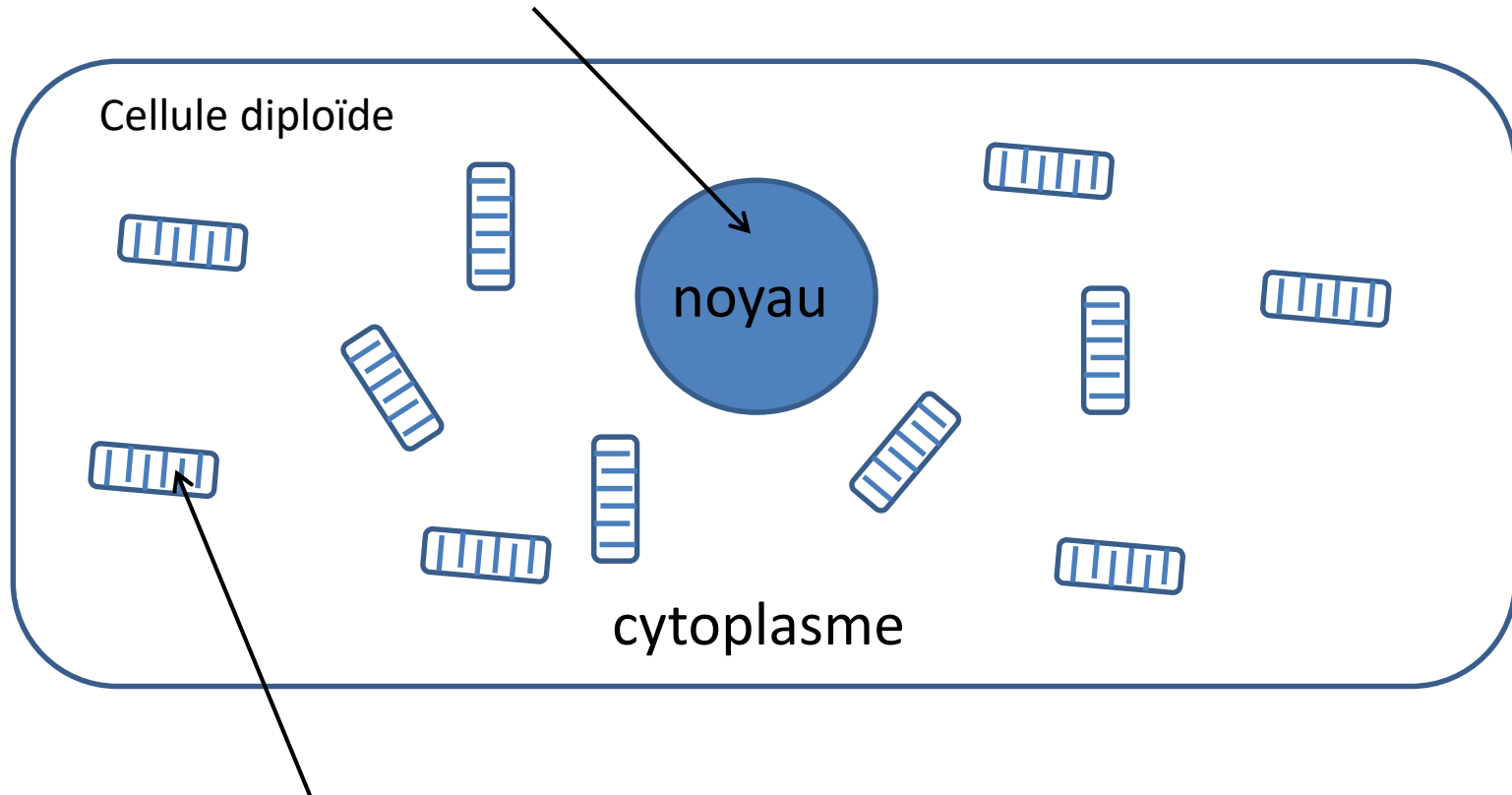
## Plan France Médecine Génomique 2025

Une nouvelle organisation de la génétique médicale



# Génome et génomique

Génome nucléaire (~6 Gb: diploïde ; ~3Gb:haploïde)



Génome mitochondrial (~16 kb)

**Le génome nucléaire et le génome mitochondrial ne partagent pas le même code génétique**



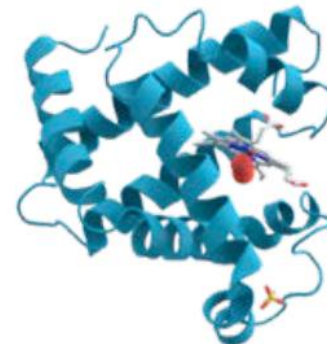
- 0110111001011011100101011111001011...

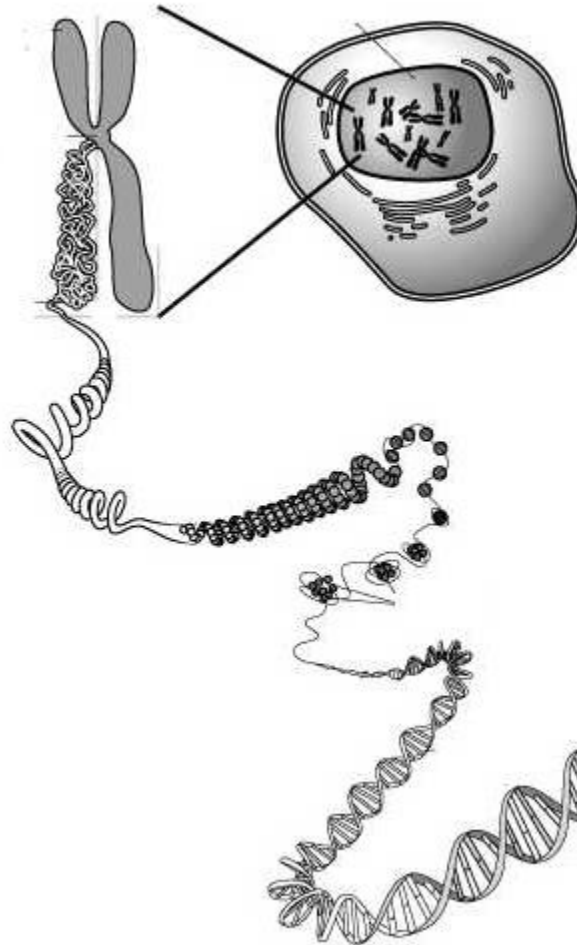
- nombres
- images
- vidéos
- musique



- GAAGGGAGGAAGTGCCTGA....

- protéines
- signaux de régulation de l'expression des gènes





**A : Adénine**

**T : Thymin**

**G : Guanine**

**C : Cytosine**

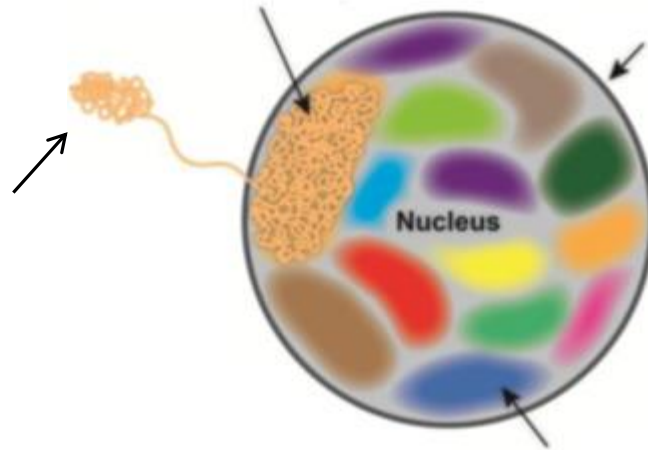
A T G C T C A T  
T A C G A G T A

La métaphase est l'étape du cycle cellulaire où la chromatine est la plus condensée





Chromosome interphasique

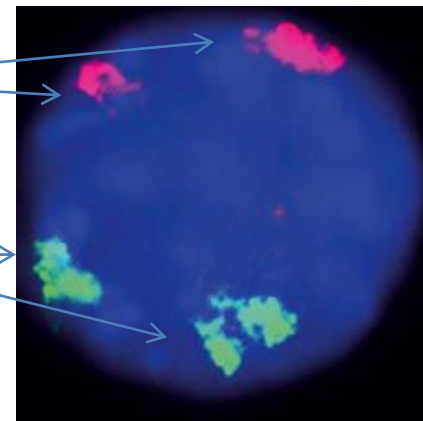


Enveloppe nucléaire

Les chromosomes ne sont pas repartis au hasard mais occupent chacun un volume délimité appelé **territoire chromosomique**

Chromosomes 7

Chromosomes 11



**Chaque chromosome en interphase occupe un territoire dédié**





*Encephalitozoon intestinalis* (2,3 Mb)



*Paris japonica* (150 000 Mb)



La complexité d'un organisme n'est pas liée à la taille de son génome

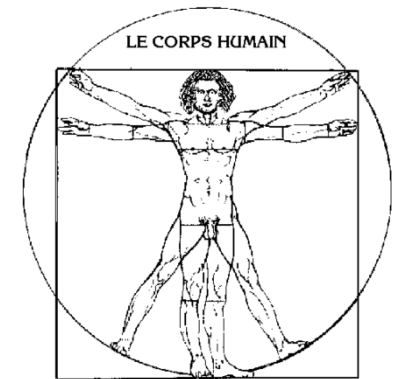
*Arabidopsis thaliana*  
~125 Mb, ~25 000 gènes



*Mus musculus*  
~2,5 Gb, ~ 22 000 gènes



*Homo sapiens*  
~ 3 Gb, ~ 20 000 gènes



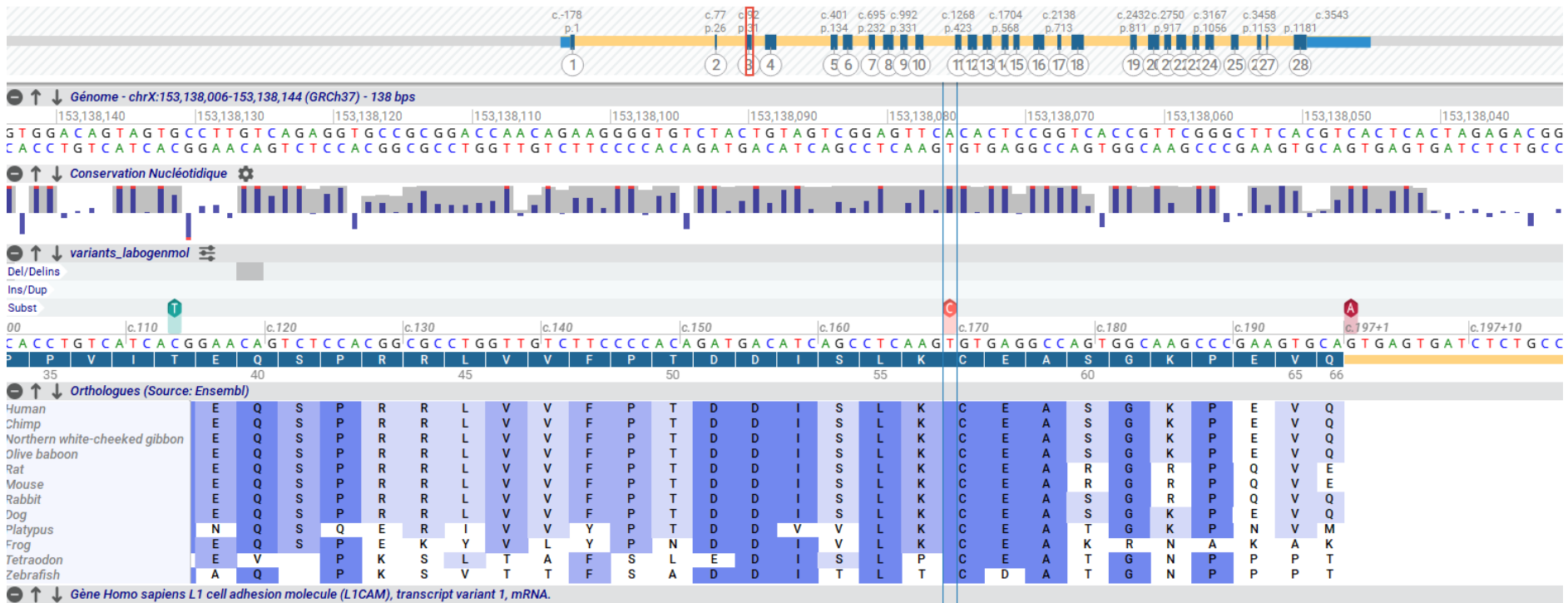
La complexité d'un organisme n'est pas liée à son nombre de gènes

Vue générale du transcrit NM\_000425.5 (L1CAM)

L1CAM - L1 cell adhesion molecule | GRCh37 (Chr X) |

gnomAD SCORES

OMIM®



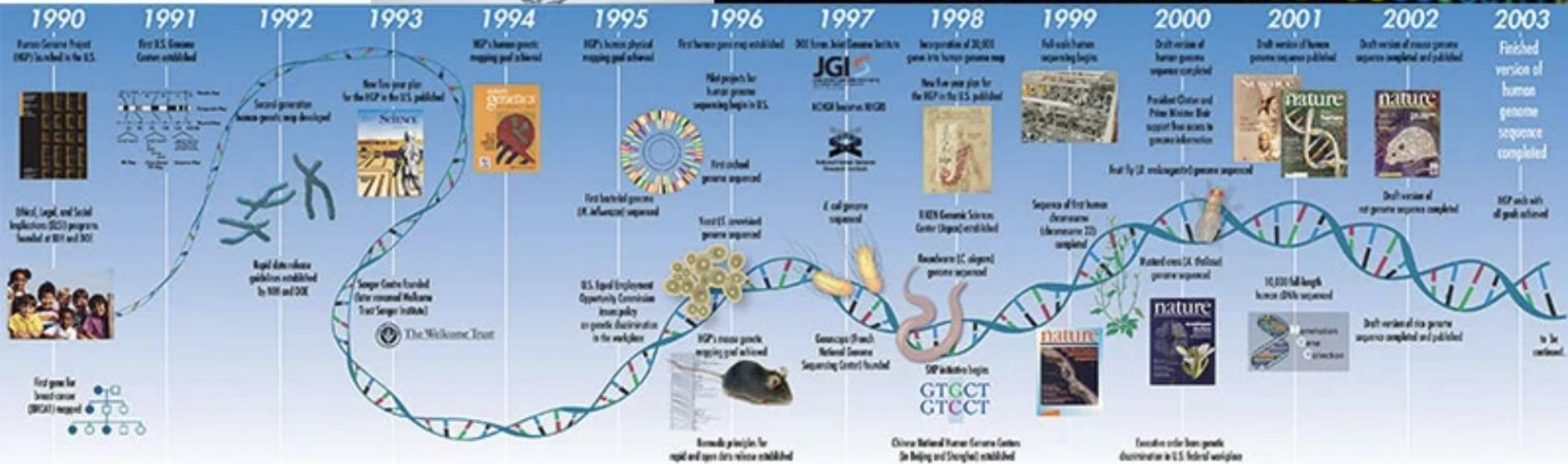
La conservation nucléotidique et la conservation en acides-aminés au cours de l'évolution sont des arguments pour l'interprétation des variants en génétique médicale



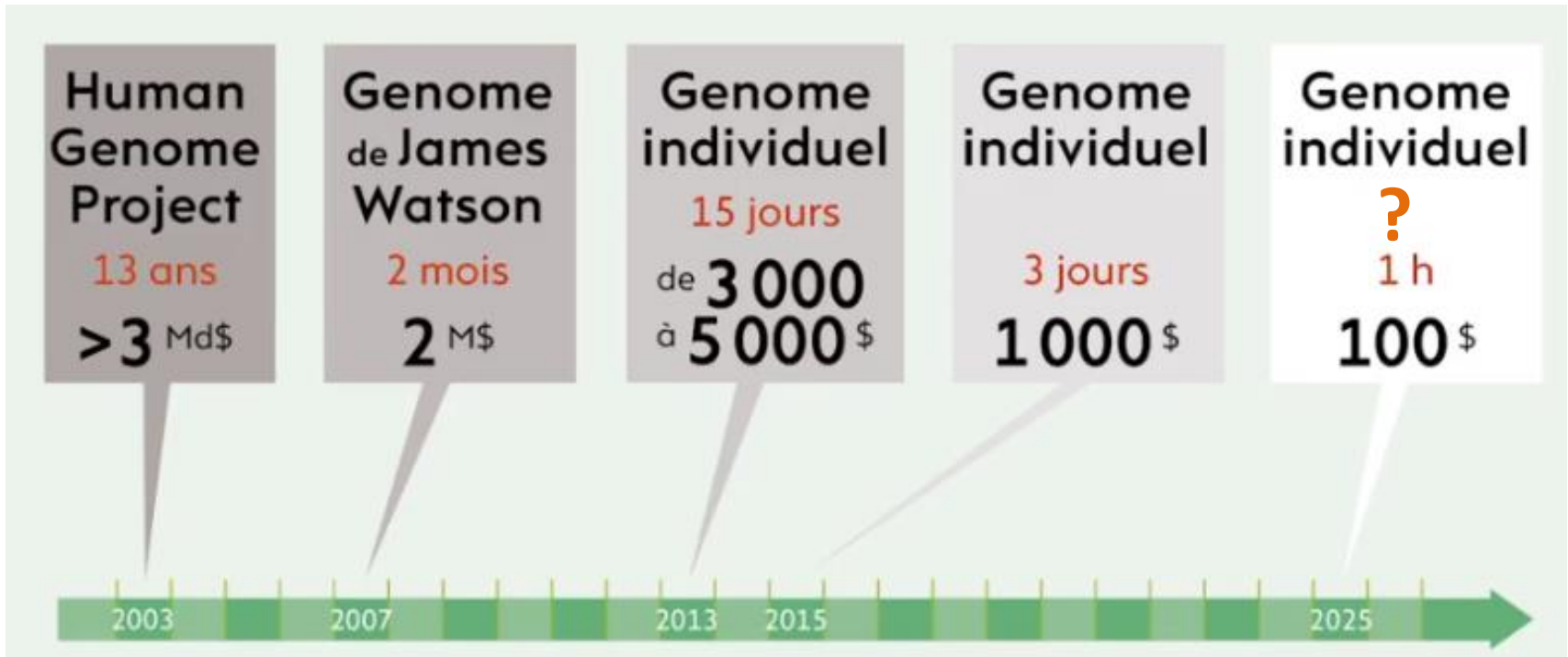
# Séquençage du génome: The Human Genome Project

1990-2003 : 13 ans de travail et 3 milliards de dollars pour parvenir à établir la 1<sup>ère</sup> séquence du génome humain (92%)

NB : séquence d'un pool de plusieurs génomes







**James Watson**



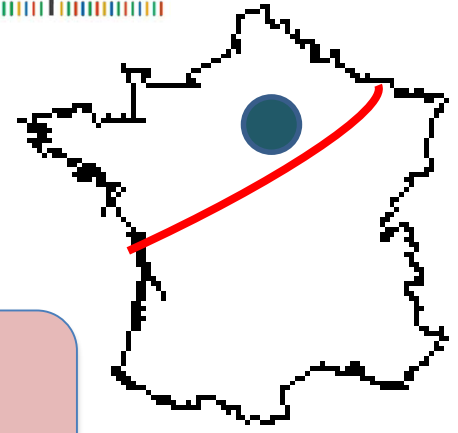
**Craig Venter**



## Plan France Médecine Génomique 2025 (PFMG2025)

Séquençage de génomes de patients au titre du diagnostic

SeqOIA



### Objectif : réduire l'errance et l'impasse diagnostiques

Chaque patient devra avoir un diagnostic confirmé au maximum 1 an après la consultation avec le médecin spécialiste

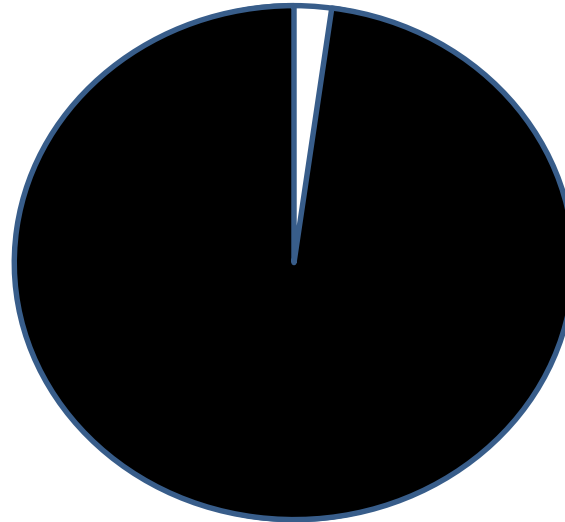


Le tsunami de données générées

Le défi est l'interprétation des données générées

- Importance de la qualité des données cliniques
- Importance de la stratégie : privilégier les approches en trio : le sujet malade et ses parents

**ADN codant pour des protéines: 1 à 2 %**  
(~20 000 gènes codants)



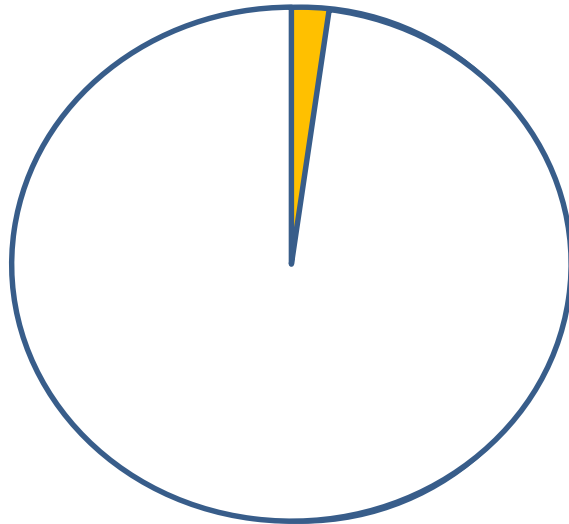
**98 à 99%: ADN non codant. La « matière noire » du génome.**  
(précédemment décrite comme « ADN poubelle » - junk DNA)

## Exome :

ensemble des exons des 20 000 gènes du génome humain

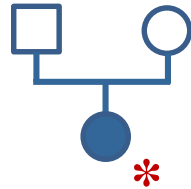
36 Mb

~ **200 000 exons**

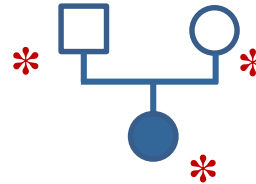




analyse en solo



analyse en trio



ADN



Capture des régions exoniques  
Séquençage

Analyse de 1<sup>ère</sup> intention pour  
le diagnostic des déficiences  
intellectuelles



Filtration bioinformatique

**23000** variations génétiques

**1600** variations rares < 1%

non présentes dans les bases de données de populations contrôles

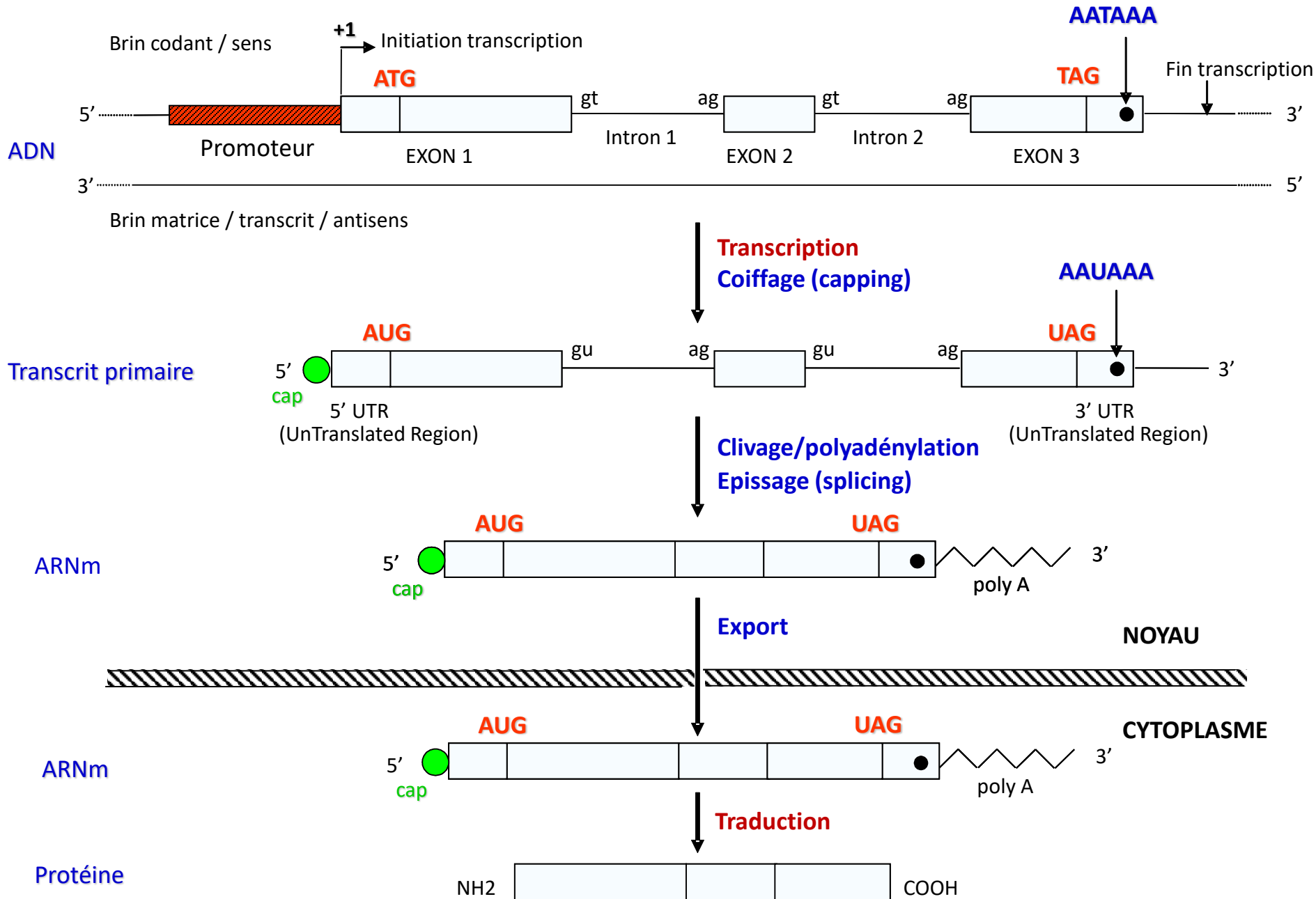
**110-330** variations rares prédites délétères

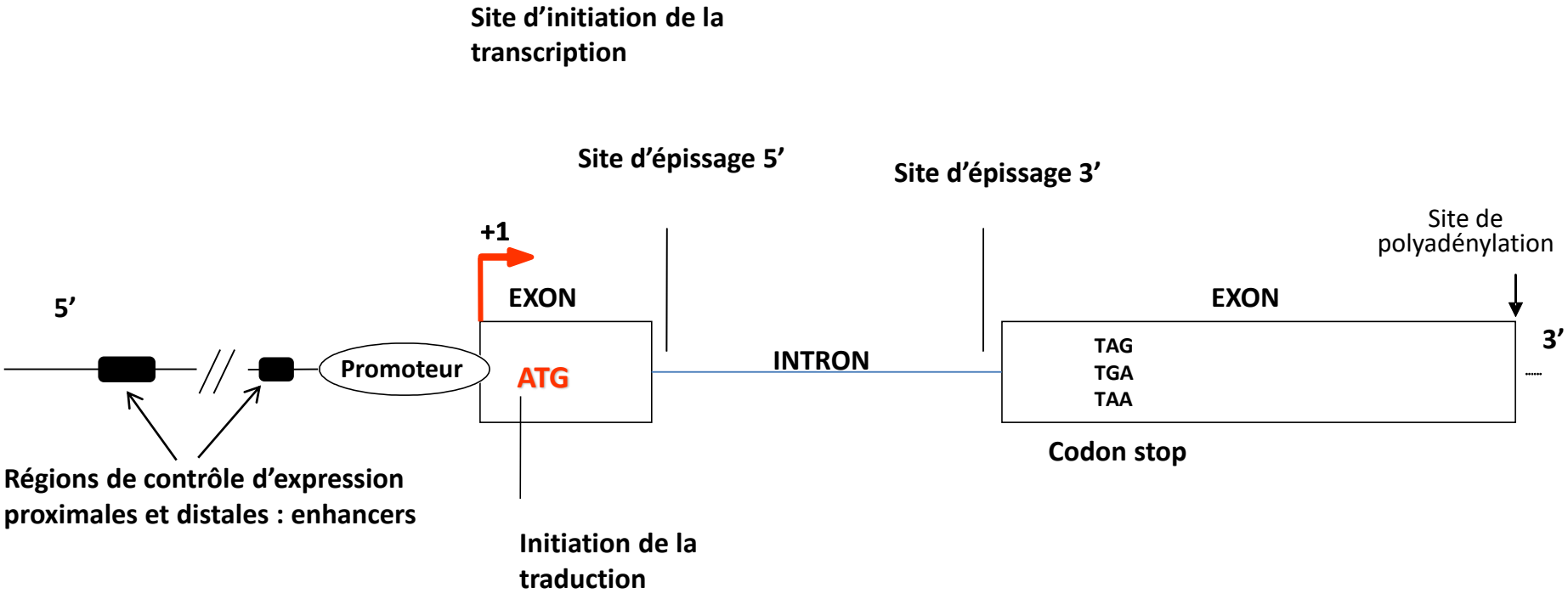


Interprétation médicale

- Identification des **variations causales**
- Identification des variations génétiques **non associées à la maladie**  
**mais d'intérêt médical (Secondary / incidental findings)**







Site d'initiation de la transcription

Site d'épissage 5'

Site d'épissage 3'

Site de polyadénylation

5'

3'

Promoteur

+1

EXON

ATG

INTRON

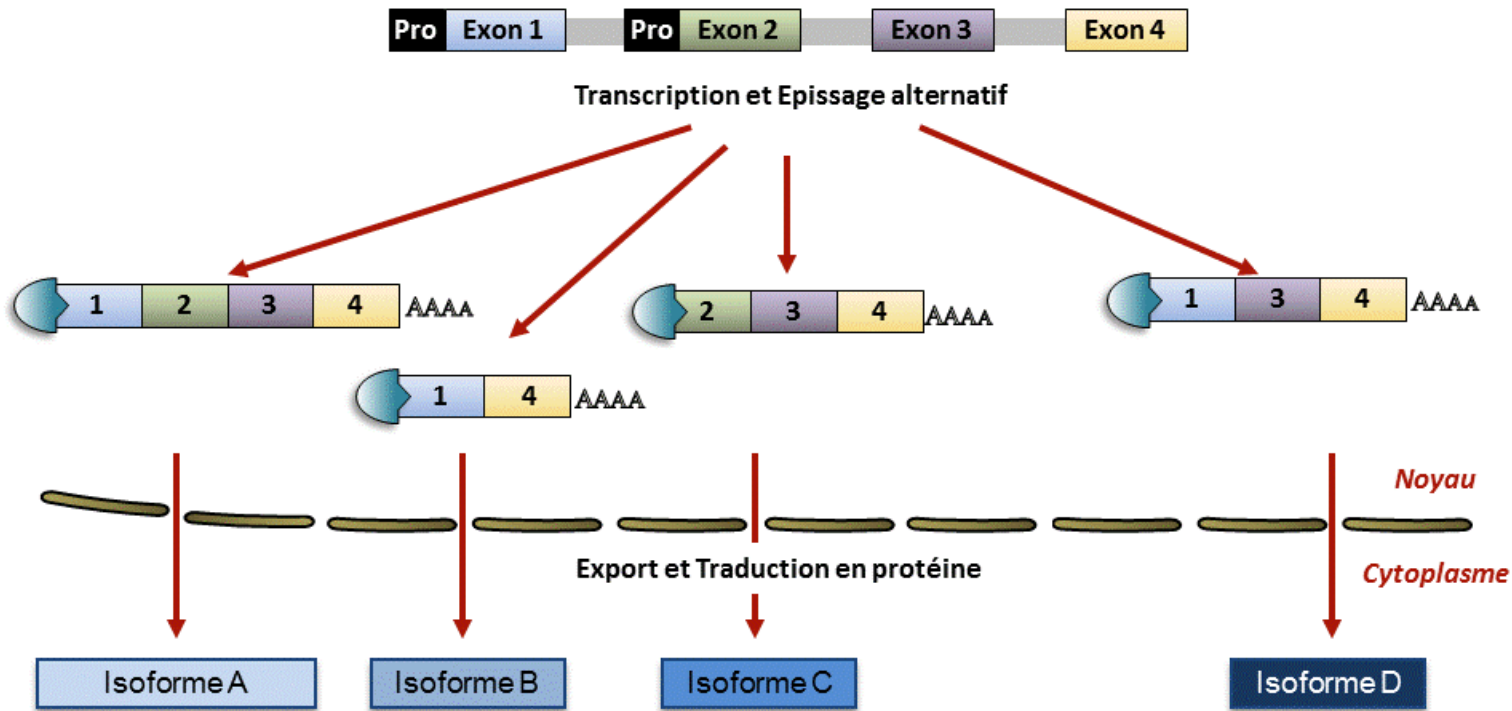
EXON

TAG  
TGA  
TAA

Codon stop

Régions de contrôle d'expression proximales et distales : enhancers

Initiation de la traduction



Grâce à l'inclusion alternative des exons, **un seul pré-ARNm** peut ainsi produire **plusieurs ARNm matures** codant pour **différentes isoformes protéiques** structurellement et fonctionnellement distinctes.

**Plus de 95% des gènes sont soumis à transcription, épissage, maturation 3' alternatifs.**  
→ ~ 20 000 gènes mais ~ 155 000 transcrits codant des protéines

Chaque codon est composé de 3 nucléotides.

Chaque codon code pour un acide aminé selon le code génétique.


DISQUESAMESTTONAMI

Cadre de lecture 1      DIS   QUE   SAM   EST   TON   AMI

Cadre de lecture 2      D ISQ UES   AME   STT   ONA   MI

Cadre de lecture 3      DI SQU ESA   MES   TTO   NAM I

Cadre de lecture 1      ATG   CAT   ATA   GAT   AAC

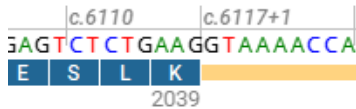
Cadre de lecture 2      A TGC   ATA   TAG   

Cadre de lecture 3      AT GCA   TAT   AGA   TAA   

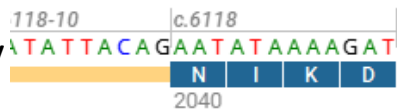
## Gène *DMD*



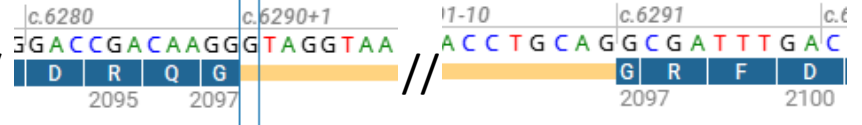
### Exon 42



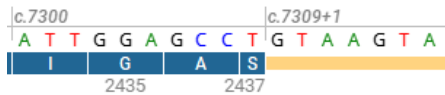
### Exon 43



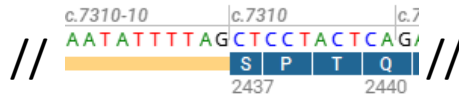
### Exon 44



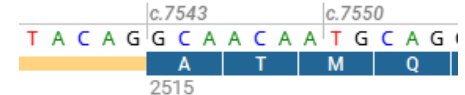
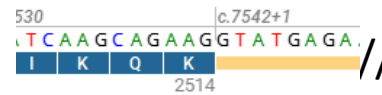
### Exon 50

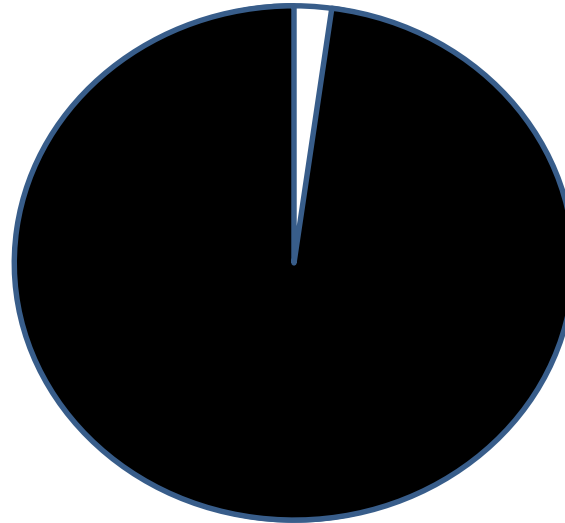


### Exon 51



### Exon 52





**98 à 99%: ADN non codant. La « matière noire » du génome.**  
(précédemment décrite comme « ADN poubelle » - junk DNA)





## THE DARK SIDE OF THE HUMAN GENOME

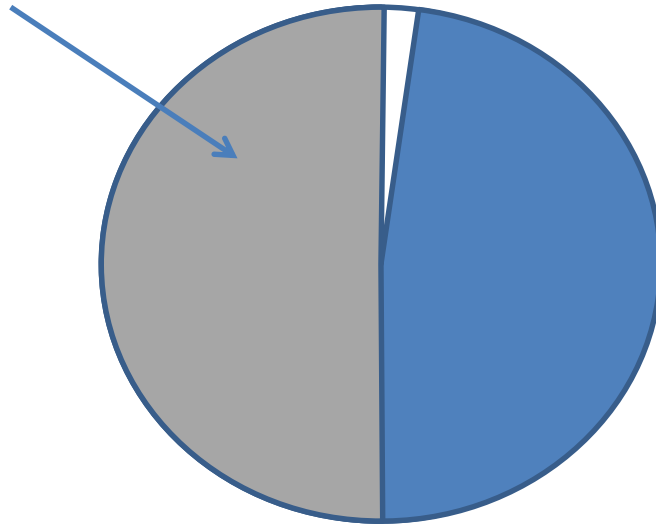
*Scientists are uncovering the hidden switches in our genome that dial gene expression up and down, but much work lies ahead to peel back the many layers of regulation.*

13 OCTOBER 2016 | VOL 538 | NATURE | 275



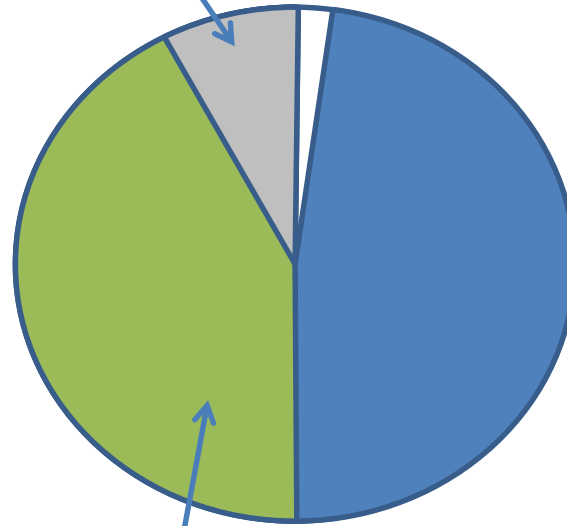
**La compréhension de la complexité du génome passe par la réponse à la question : comment ces 1 à 2% sont-ils régulés par ces 98-99% restants de “matière noire” ?**

**ADN répété: ~50%**



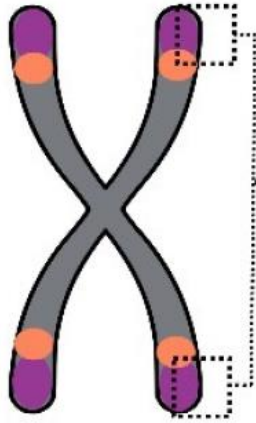


Séquences répétées en tandem (~5%)

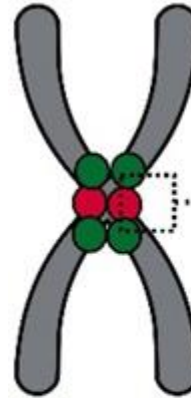


Séquences répétées dispersées (~45%)

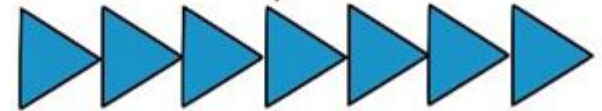
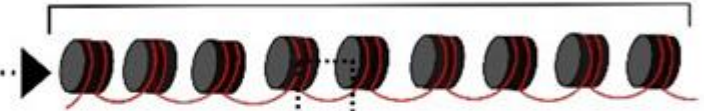
# Séquences répétées en tandem au centromère et aux télomères



ADN répété aux télomères,  
ex: (TTAGGG)<sub>n</sub>



ADN répété au centromère



(alpha satellite)

Les transcrits alpha satellite sont impliqués dans la structure du centromere et jouent un rôle dans la ségrégation chromosomique

Mais du fait de la difficulté de séquençage, ces régions restent peu connues



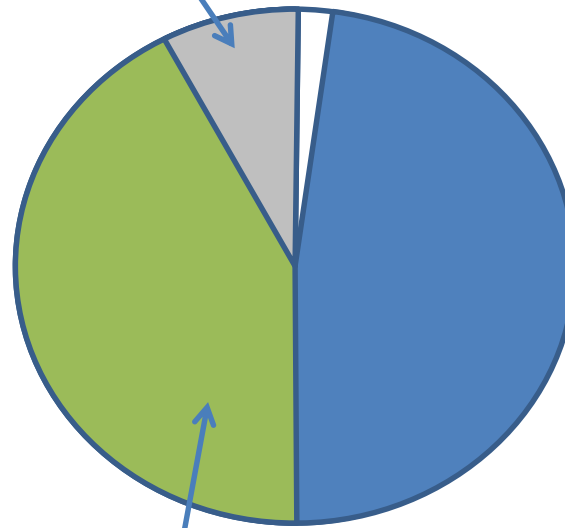
**PATHOLOGIE**

Les anomalies de ségrégation des chromosomes sont la cause d'anomalies chromosomiques

au cours de la méiose → anomalies du développement

au cours de la mitose → cancer

Séquences répétées en tandem (~5%)



**Séquences répétées dispersées (~45%)**

**Transposition/mobilité par  
l'intermédiaire d'un ARN:  
rétro transposons (fréquent)  
Mécanisme de « copier/coller »**

**Transposition/mobilité par  
l'intermédiaire d'un ADN:  
transposons (rare)  
Mécanisme de « couper/coller »**

SINE Short Interspersed Element

LINE Long Interspersed Element

HERV Human Endogenous Retrovirus

REVIEW ARTICLE

## Mobile DNA in Health and Disease

Haig H. Kazazian, Jr., M.D., and John V. Moran, Ph.D.

The insertion of mobile elements into the DNA of gametes or the early embryo can disrupt genes, leading to sporadic cases of disease, and their insertion into the DNA of somatic cells may contribute to cancers and neuropsychiatric disease. Clearly, mobile DNA has been instrumental in shaping the structure, function, and evolution of the human genome.

N Engl J Med 2017;377:361-70.

DOI: 10.1056/NEJMra1510092



**PATHOLOGIE**

## Variabilité du génome humain: un génome « type »

~ 3 Gb

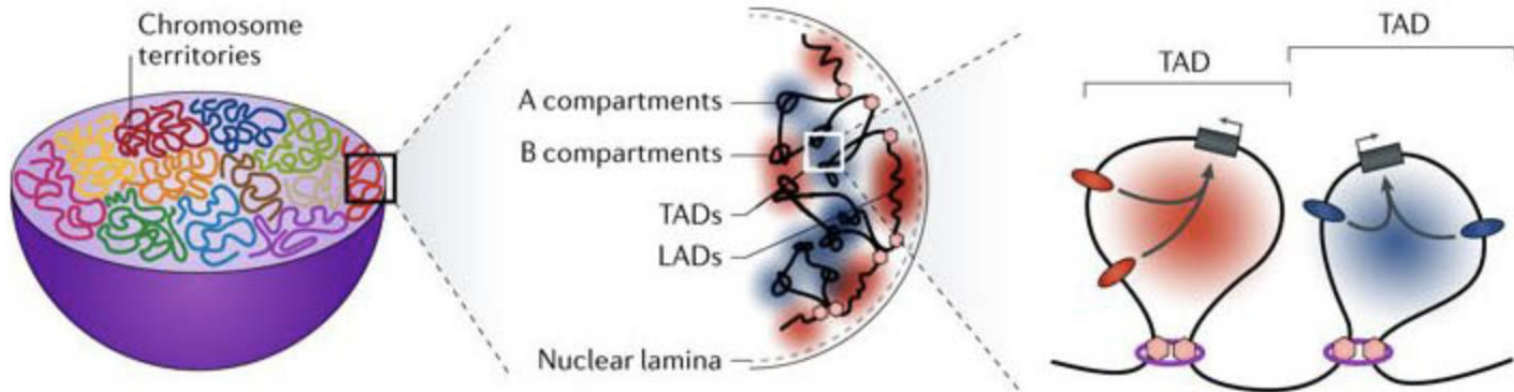
~ 4 à 5 millions de variants  
par comparaison au  
« génome de référence »

~ 20 000 variants exoniques

~ 155 000 transcrits

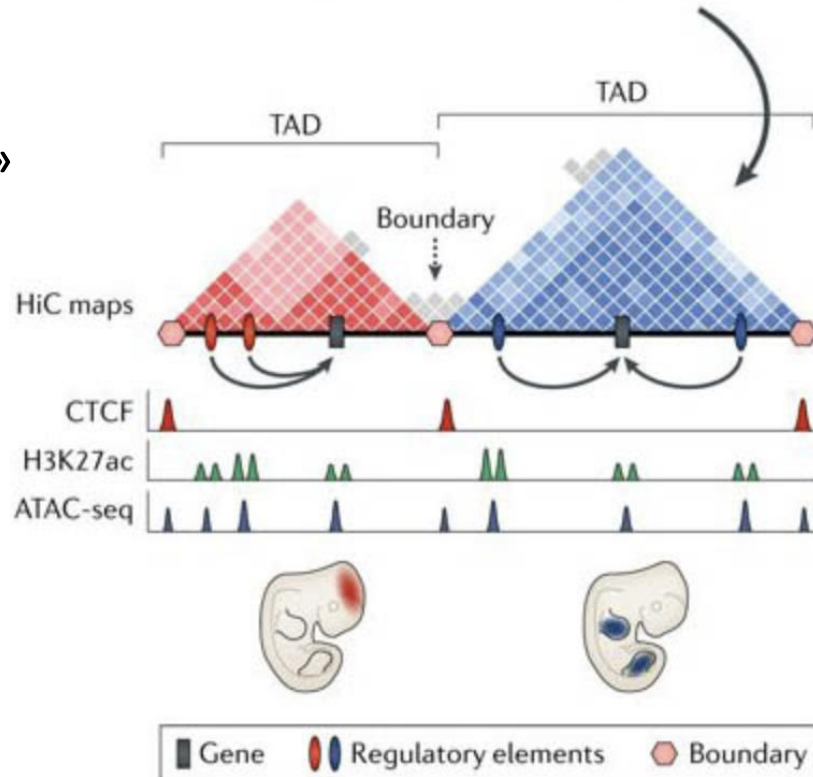


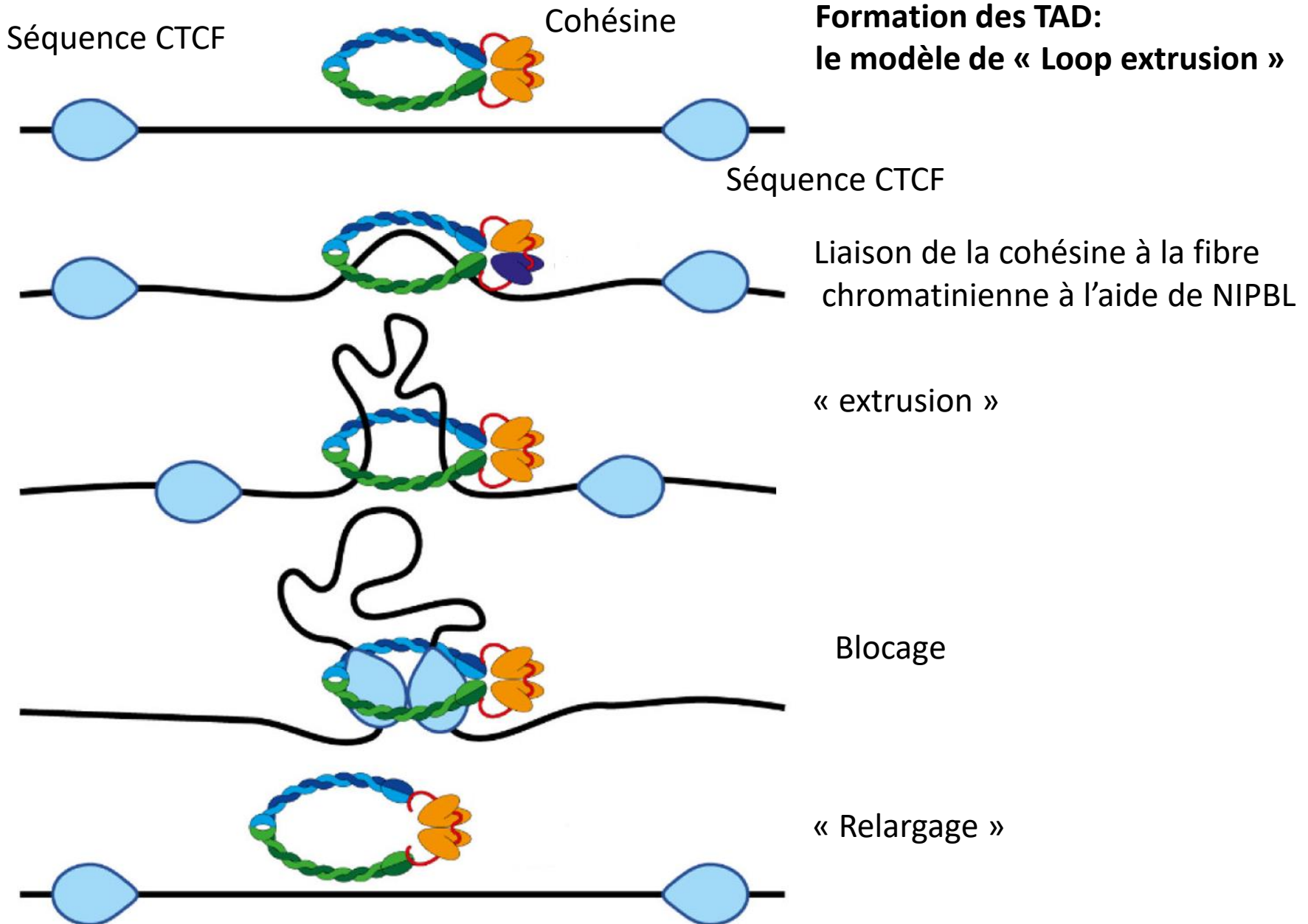


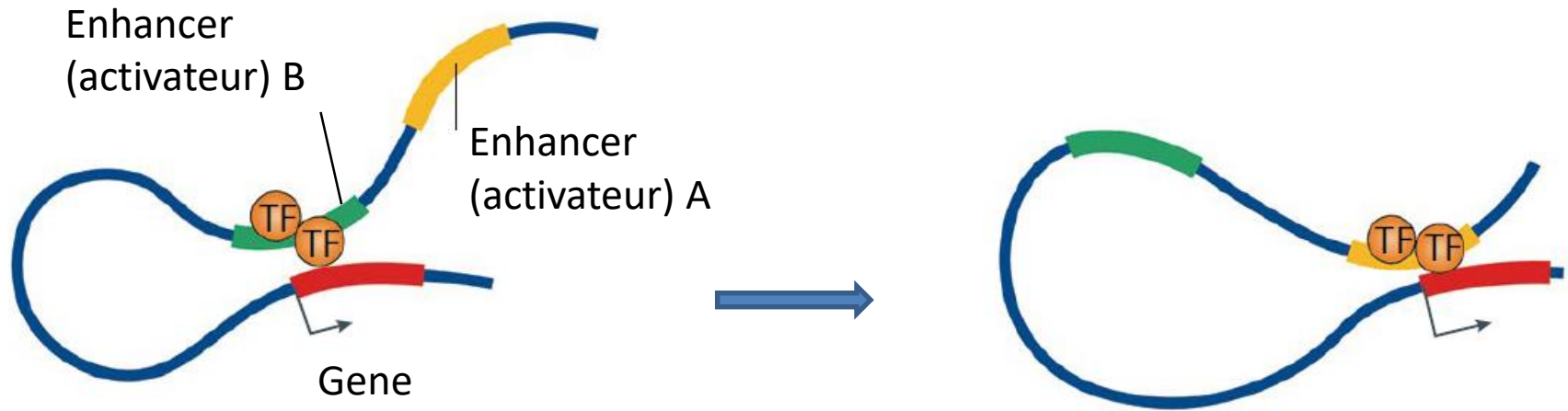


TAD « domaines topologiquement associés »

Frontières de TAD



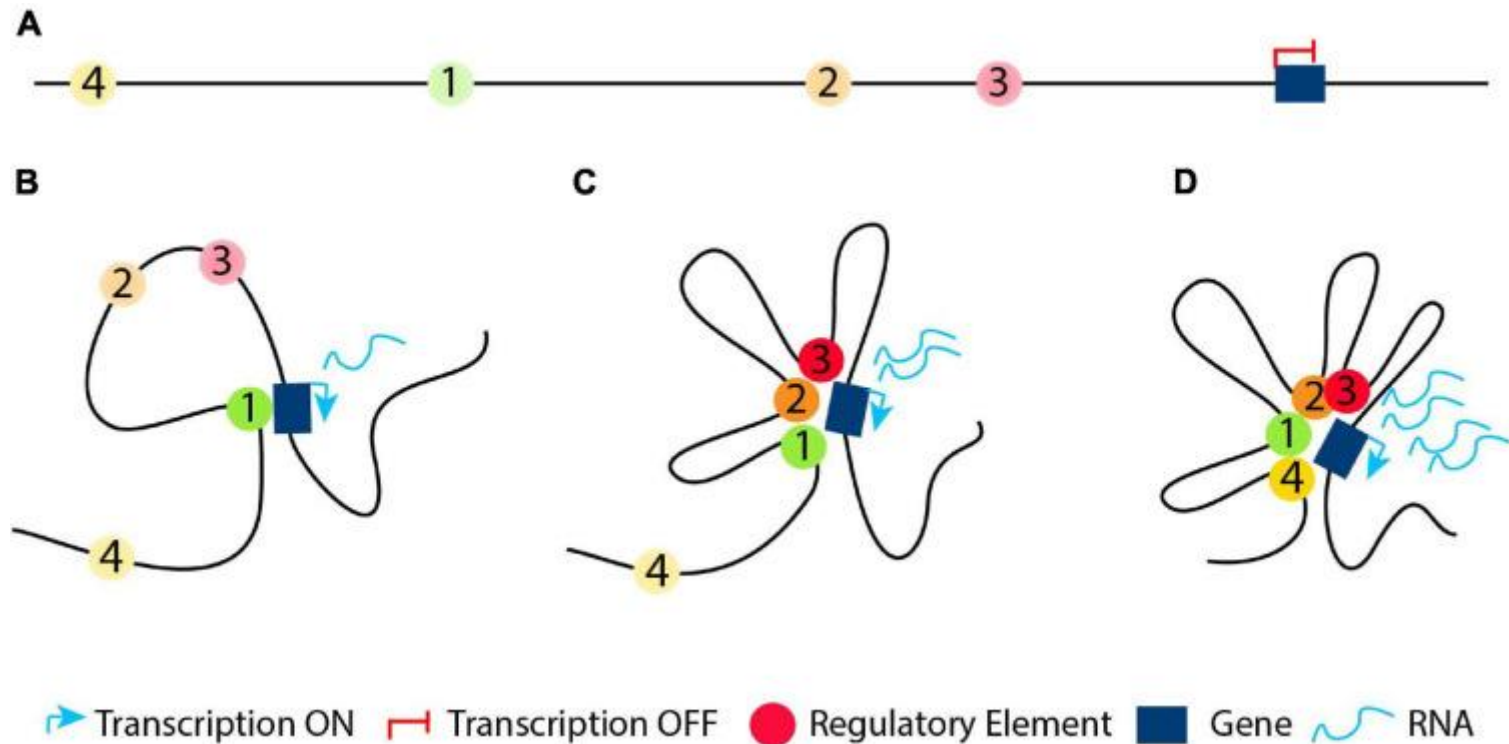




 : facteur de transcription

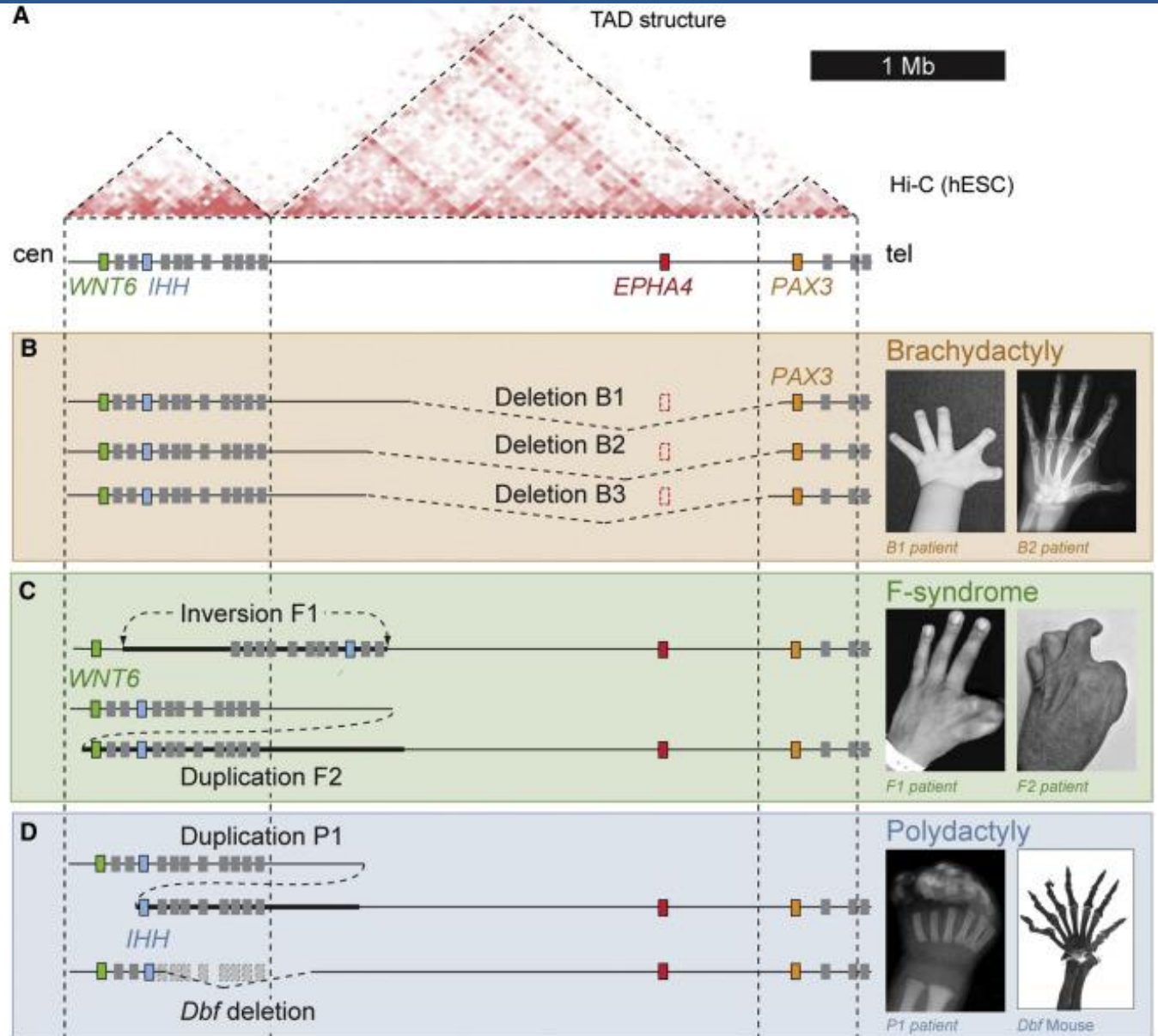
**Un même gène peut s'associer à différents enhancers (activateurs)  
L'activation d'un gène peut être tissu spécifique**

**Un même enhancer (activateur) peut réguler différents gènes, à des stades différents du développement et de façon tissu spécifique**



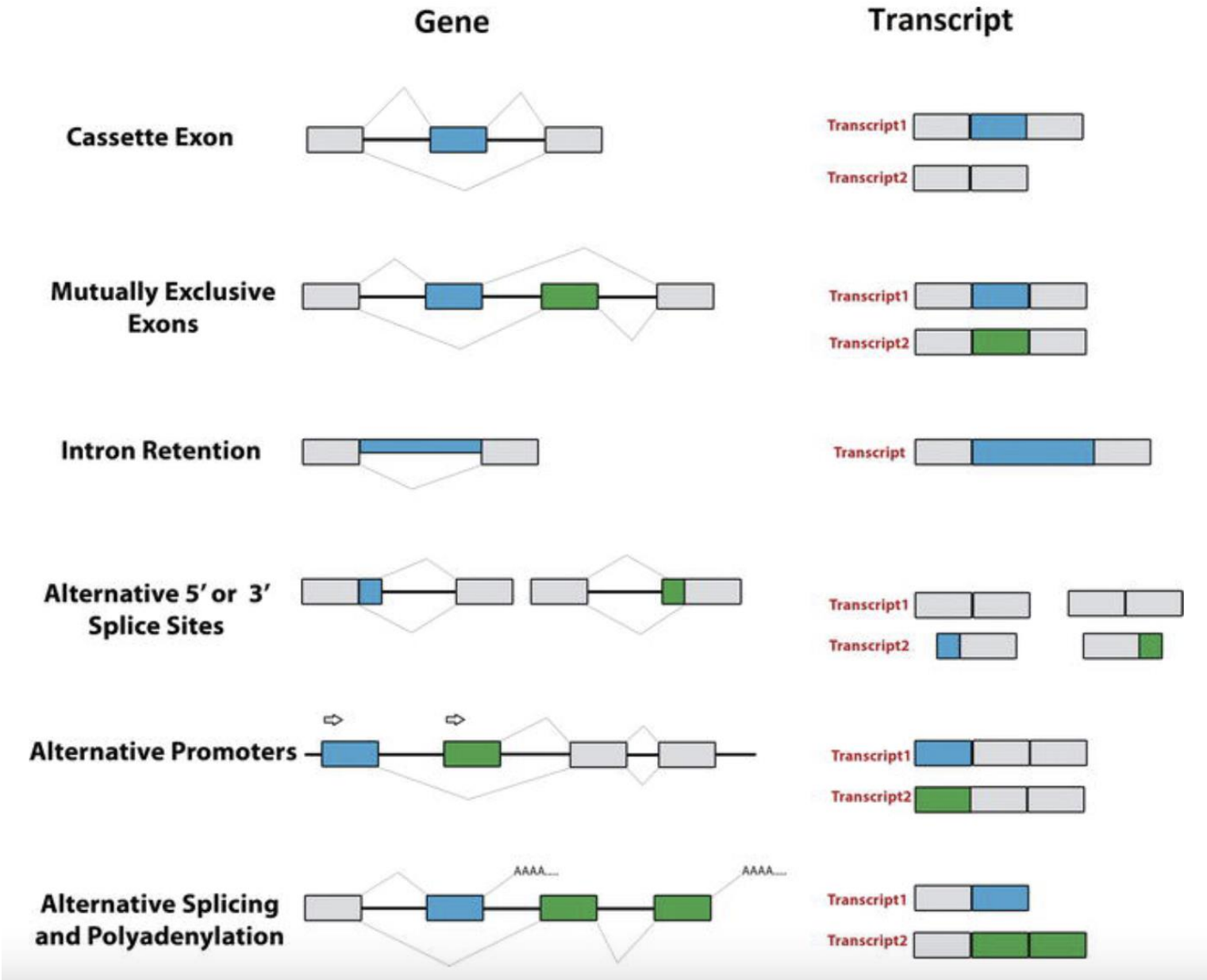
**Les enhancers régulent l'expression durant le développement, mécanisme clé de la différenciation cellulaire**

# Les altérations du génome 3D peuvent être responsables de pathologies



**PATHOLOGIE**

# Le monde fascinant des ARN







## ARN

ARN codant

ARNm (ARN messenger)

ARN non codant

ARN ribosomique  
ARN de transfertPetits ARN non codants  
<200 nucléotides  
(*snc RNA: small non coding RNA*)longs ARN non codants  
> 200 nucléotides  
(*lnc RNA: long non coding RNA*)

miARN (micro ARN) : dégradation de transcrits codants, inhibition de la traduction  
siARN (small interfering RNA) : inhibition de la traduction  
snoARN (petits ARN nucléolaires) : maturation de l'ARNm  
piARN (PIWI-interacting RNA) : inhibition de l'expression des transposons, rôle en épigénétique  
...



**~ 3/4 de notre génome est transcrit**

Le rôle de ces régions non exoniques reste à découvrir

**155 000 transcrits**

Les rôles de ces innombrables transcrits sont à décrypter

**Expression tissu-spécifique**

Importance des projets internationaux visant à estimer l'expression de chaque isoforme dans chaque tissu

**Expression spatio-temporelle**

Importance de nouvelles méthodes visant à détecter ces transcrits sur cellule unique

Analyse de l'ADN



Génomique

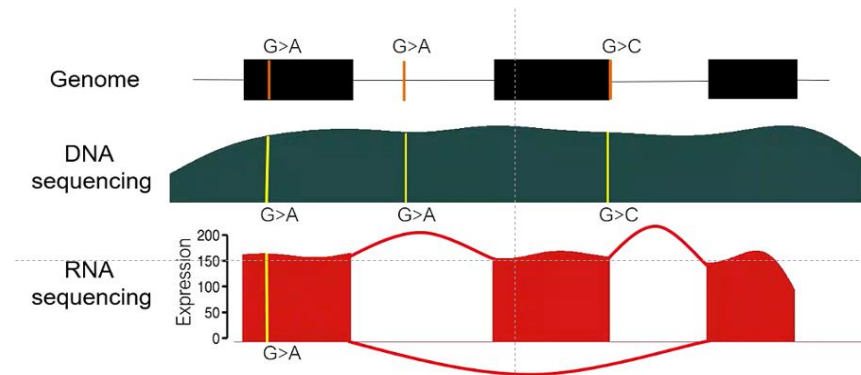
Analyse de l'ARN



Transcriptomique

Analyse de l'ARN – dans quel but ?

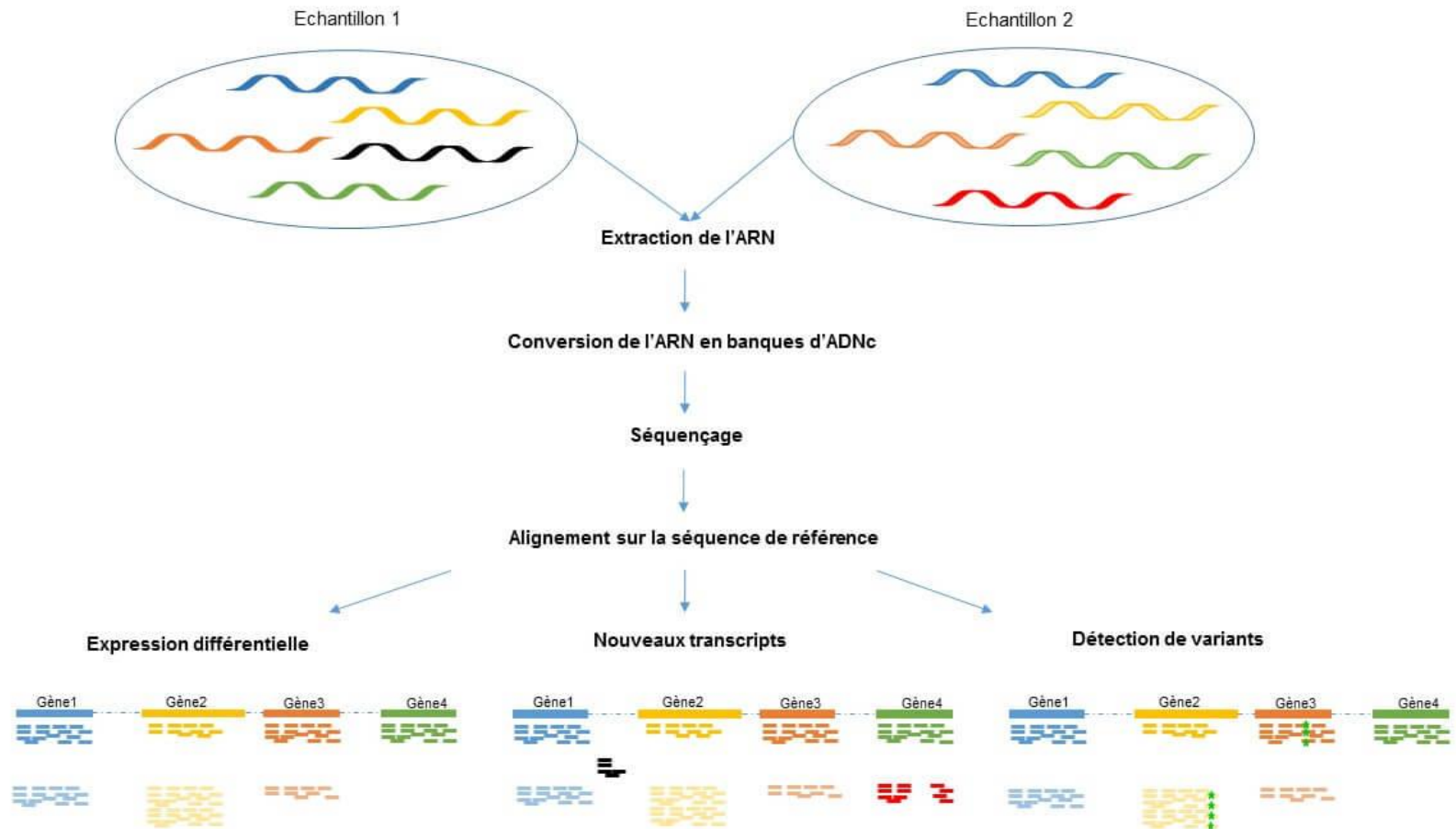
- Pour les variants introniques, caractériser une anomalie d'épissage
- Démontrer une altération de l'expression d'un gène
- Démontrer l'expression mono-allélique (expression d'un seul des deux allèles)



**L'exploration de l'ARN est complémentaire de l'analyse de l'ADN**

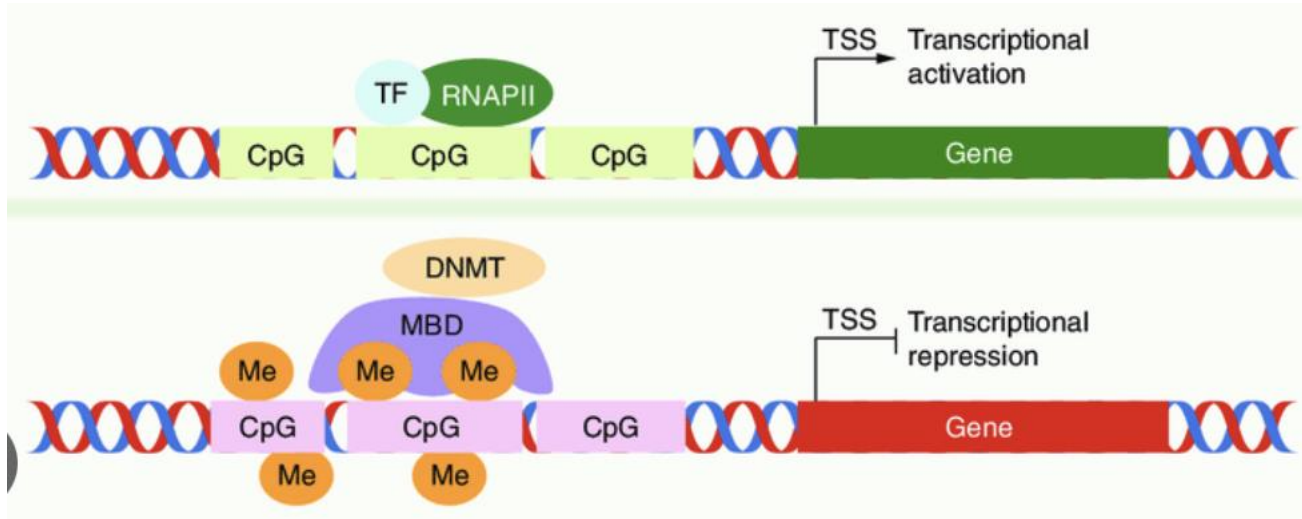
Choix du tissu d'où seront extraits les ARN

Sujets et contrôles appariés en âge et en sexe



# L'épigénétique

La **méthylation** des cytosines dans les promoteurs est associée à une **répression de la transcription** des gènes



**Méthylation** des cytosines dans les promoteurs

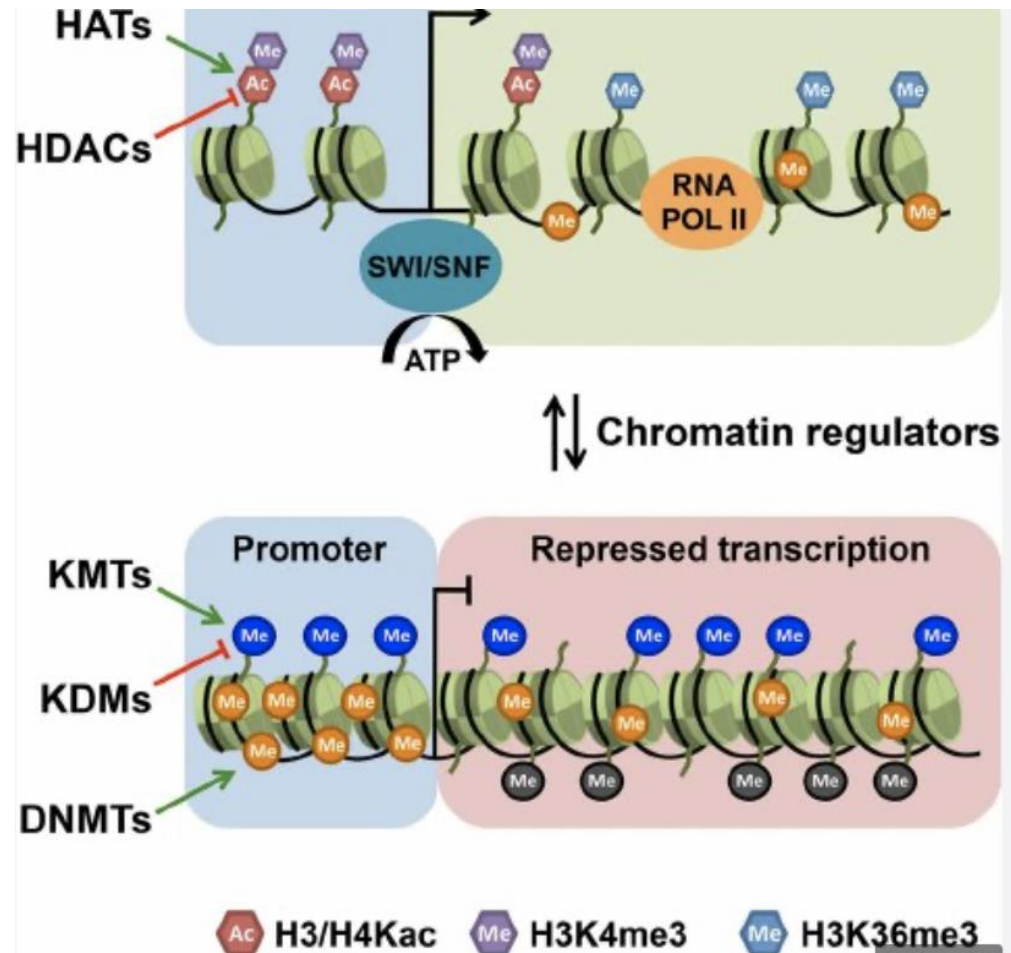
→ **compaction de la chromatine**

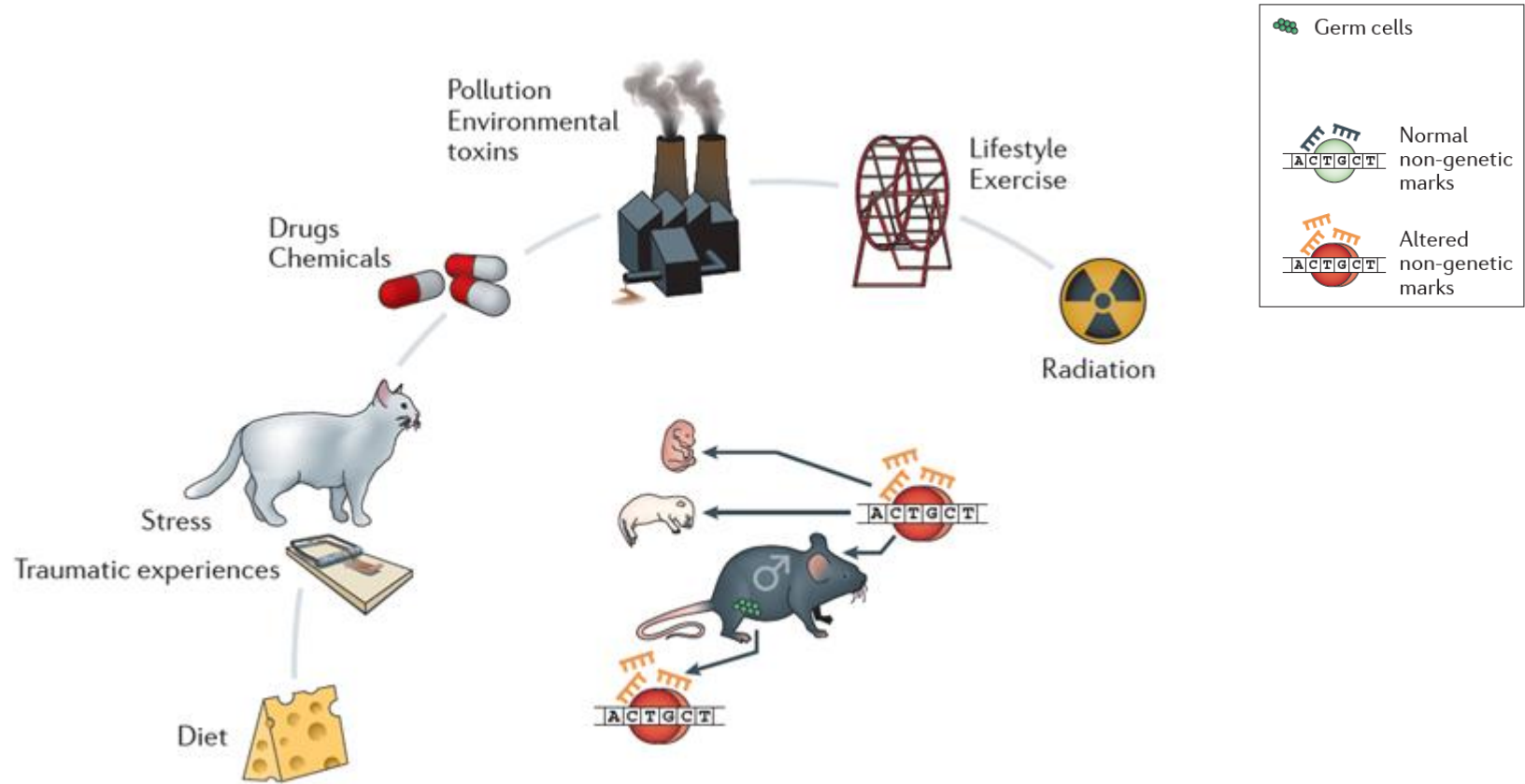
→ inaccessibilité aux facteurs de transcription

→ répression transcriptionnelle

Rôle de la méthylation des promoteurs  
dans la régulation  
de l'expression des gènes

**Epigénétique** : étude des changements  
d'expression des gènes  
indépendamment de changements  
dans la séquence de l'ADN





## Rôle de l'environnement dans la régulation de l'expression de nos gènes.

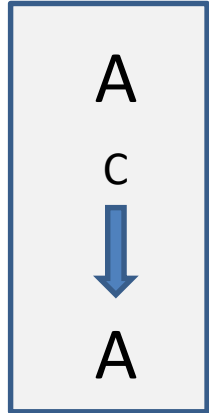
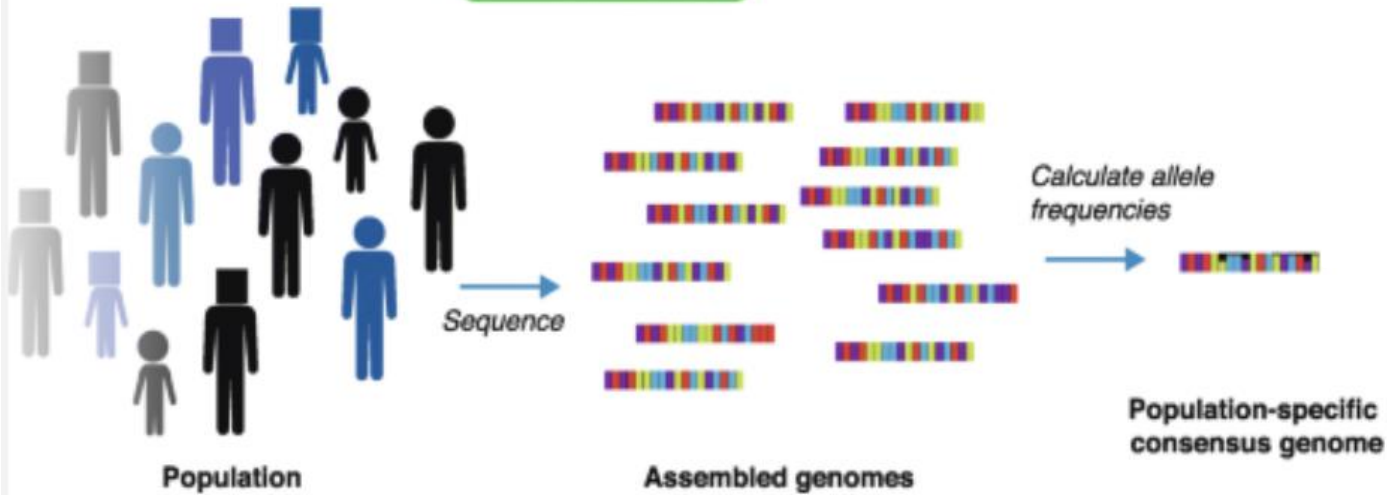
L'expression des gènes de jumeaux monozygotes est identique à 3 ans mais différente à 50 ans.



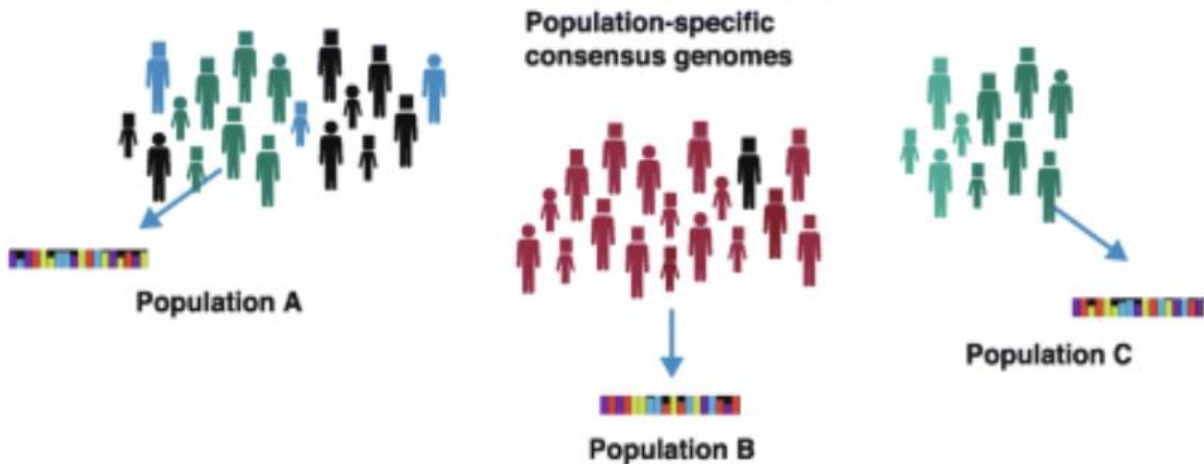
# La variabilité du génome humain



Build



Diversify





## gnomAD browser

## Single nucleotide variant: 5-36985576-G-A(GRCh37)

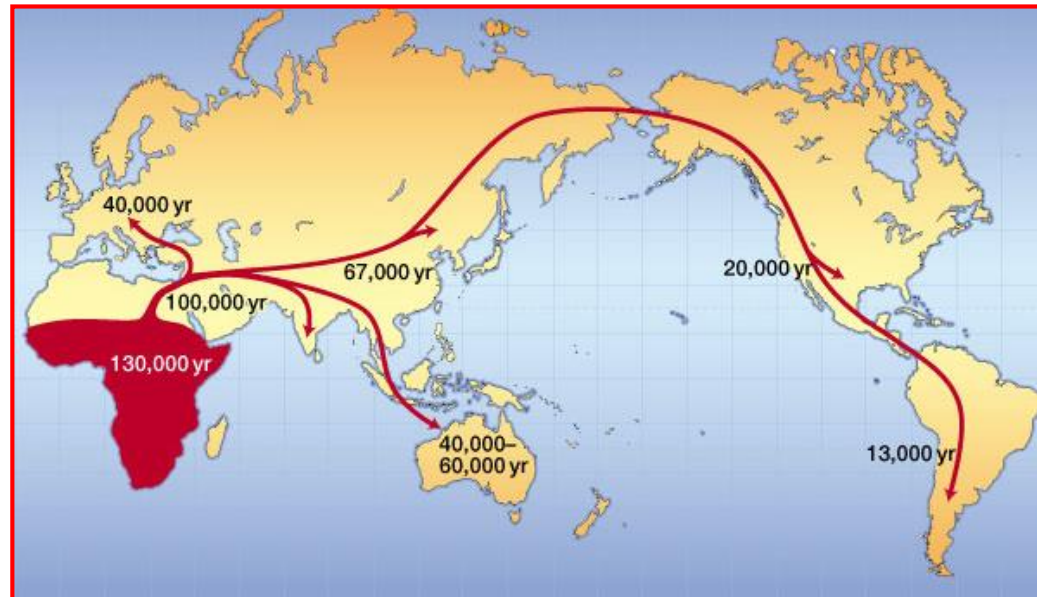
Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
▶ Ashkenazi Jewish	20	10288	0	0.001944
▶ European (non-Finnish)	132	127708	0	0.001034
▶ Other	5	7178	0	0.0006966
▶ European (Finnish)	7	25120	0	0.0002787
▶ Latino/Admixed American	2	35246	0	0.00005674
▶ African/African American	0	24704	0	0.000
▶ East Asian	0	19902	0	0.000
▶ South Asian	0	30546	0	0.000
XX	70	128056	0	0.0005466
XY	96	152636	0	0.0006289
<b>Total</b>	<b>166</b>	<b>280692</b>	<b>0</b>	<b>0.0005914</b>

## *Single Nucleotide Variation (SNV) et Copy Number Variation (CNV) :*

**Mécanisme clef de l'adaptabilité d'*Homo sapiens* à l'environnement**

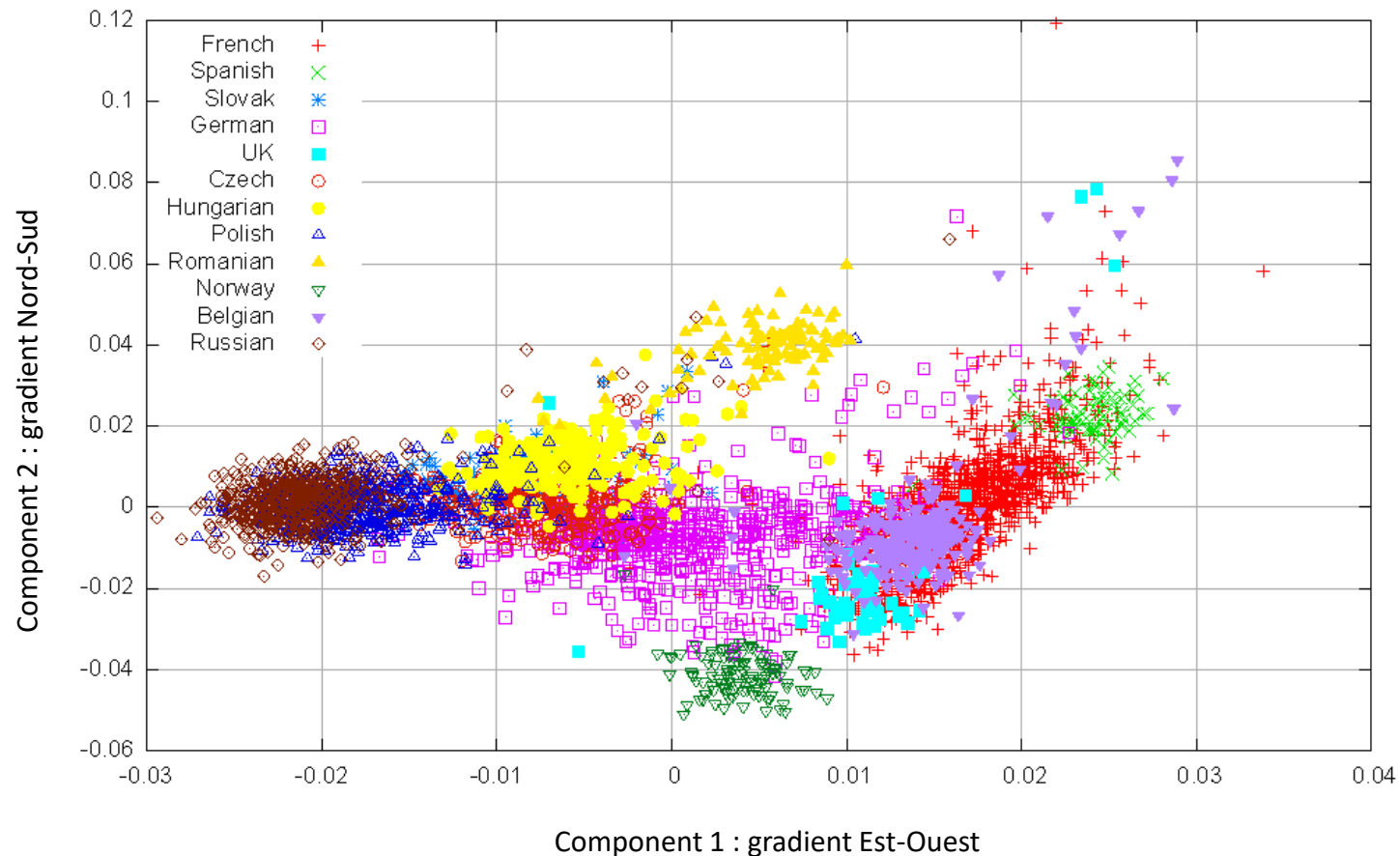
- Infections (10 000 ans +++)
- Modification des comportements alimentaires...

La plupart des variations sont survenues dans les 30 000 dernières années  
La majorité des SNVs délétères sont récents (<10 000 ans)

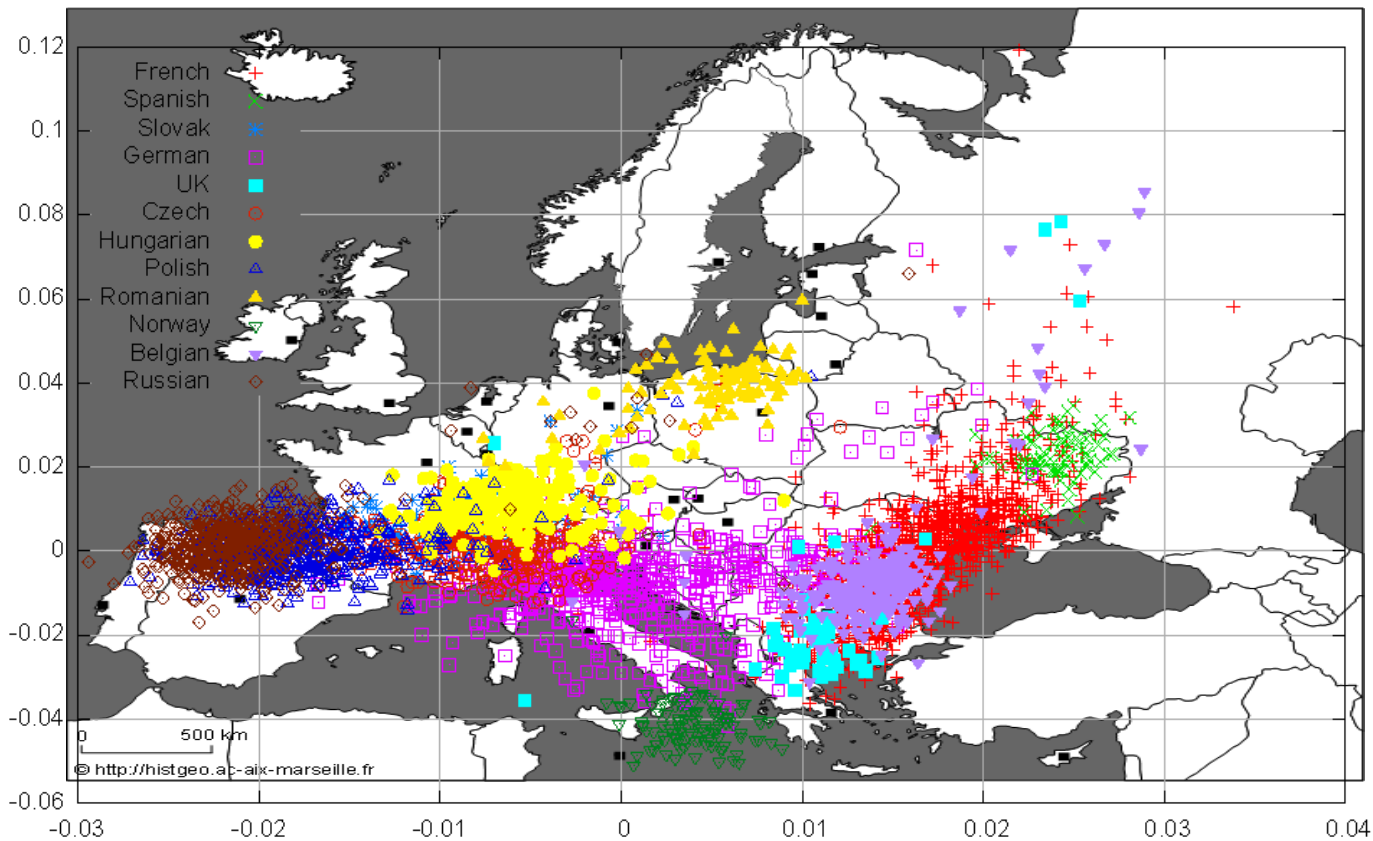


~**6000** individus de **12** populations européennes

~**120.000** SNPs parmi >**300.000** avec  $MAF > 0.10$  et  $r^2 < 0.5$

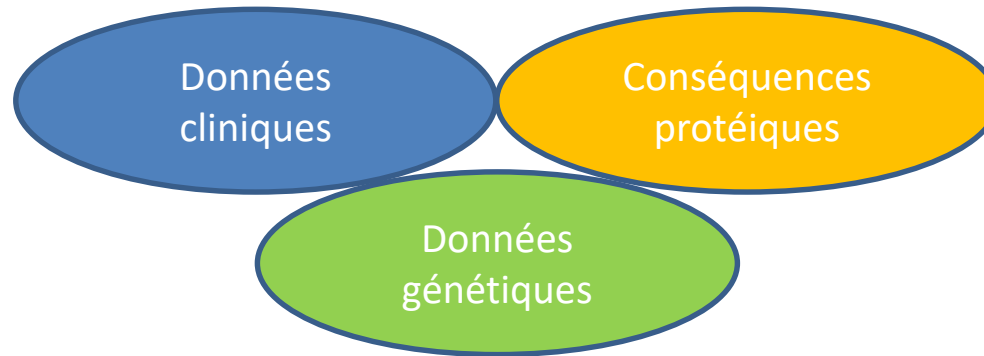


# Différences de fréquences alléliques de SNP communs : Miroir de la géographie européenne



## L'interprétation d'un variant repose sur plusieurs familles d'arguments

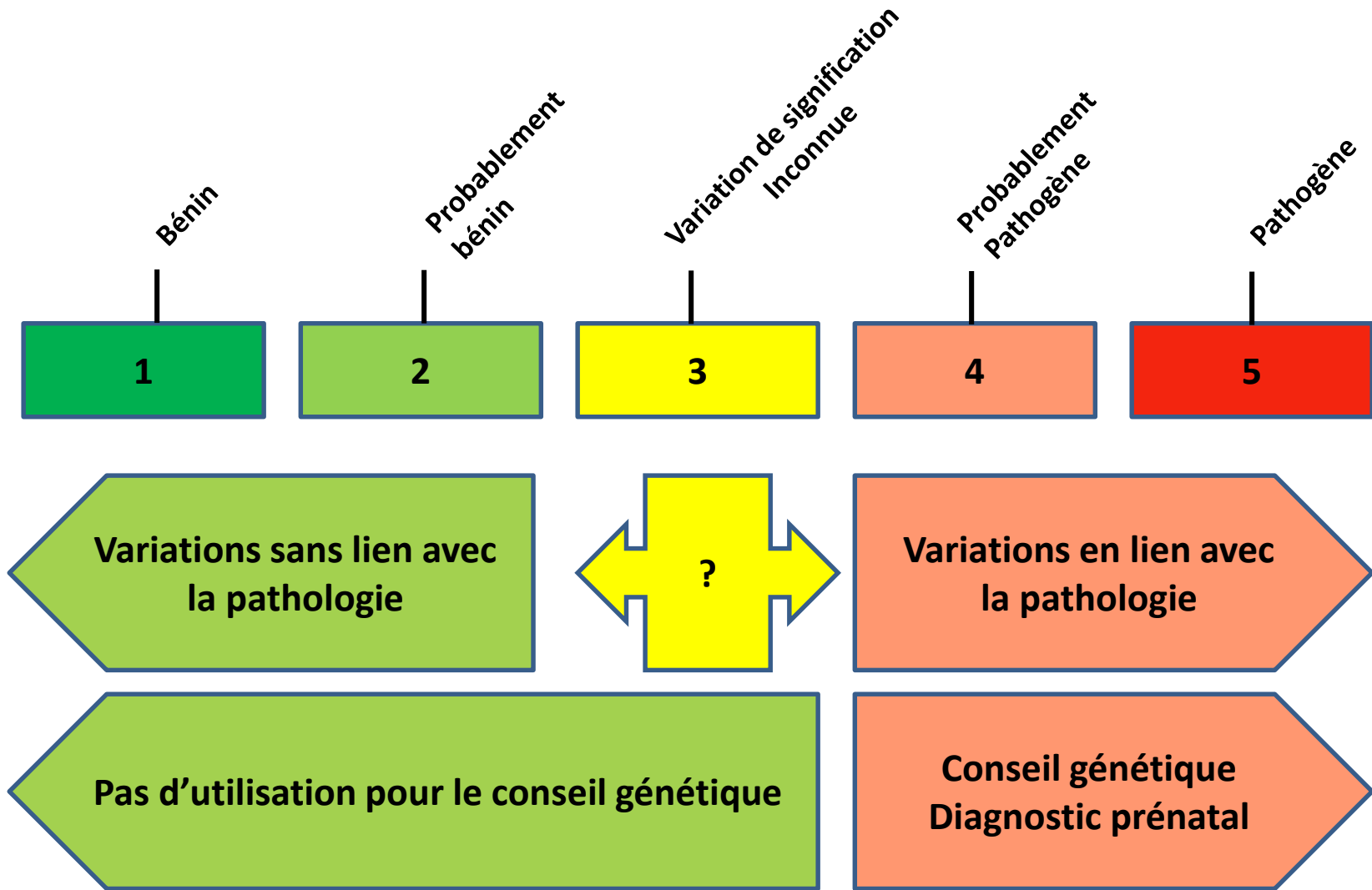
Concordance  
avec le  
phénotype



- **Fréquence du variant**
- Analyse de la ségrégation

- Effet prédit du variant sur la fonction de la protéine
- Résultats de tests fonctionnels

**Plus un variant est fréquent en population générale, plus la probabilité qu'il soit pathogène est faible**







- **Conjonction des progrès phénoménaux en nanotechnologies et en informatique**
  - Développement de nouveaux outils
  - Explosion des connaissances
  - Des résultats inattendus ouvrent de nouvelles questions !
- **Le séquençage du génome ouvre des possibilités de diagnostic des maladies rares**
  - Saut quantique dans la médecine : médecine génomique
  - L'étape d'après : les maladies communes : médecine de précision
- **Le génome est loin d'avoir livré tous ses secrets**
  - Rigueur dans l'utilisation des données