



GENOME

Pr Jonathan LOPEZ



Posez vos questions pendant le cours



1 Allez sur
wooclap.com

2 Entrez le code
d'événement dans
le bandeau
supérieur

Code
d'événement
PJINGO

0 Plan

- Organisation du génome humain
- Gènes
- ADN intergénique

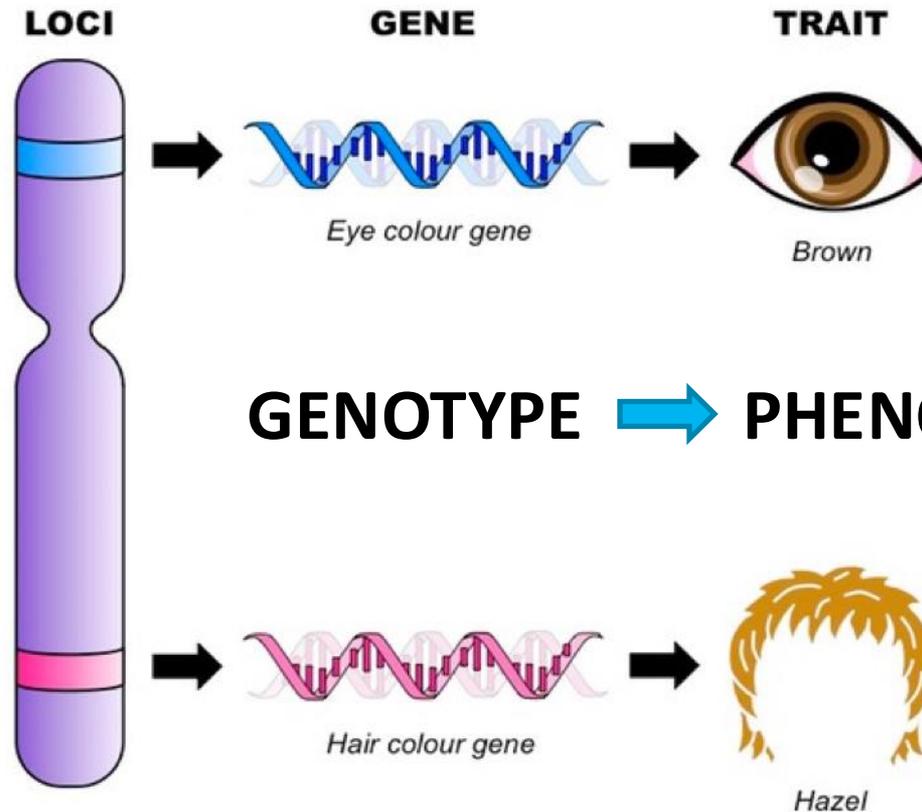
0 Plan

- **Organisation du génome humain**
 - **Génotype vs phénotype**
 - **Polymorphisme génétique**
 - **Génome de référence**
 - **Composition du génome**
- **Gènes**
- **ADN intergénique**

1 Génome : **Génotype vs phénotype**

- **Génotype = information génétique responsable de “traits phénotypiques”**

Locus=1 position sur un chromosome



1 Génome : **Génotype vs phénotype**

- La plupart des gènes existent dans différentes versions ou “allèles”

Gene	Alternative Alleles			
 <i>Eye colour</i>	 <i>Brown</i>	 <i>Blue</i>	 <i>Emerald</i>	 <i>Grey</i>
 <i>Hair colour</i>	 <i>Blonde</i>	 <i>Red</i>	 <i>Brown</i>	 <i>Black</i>

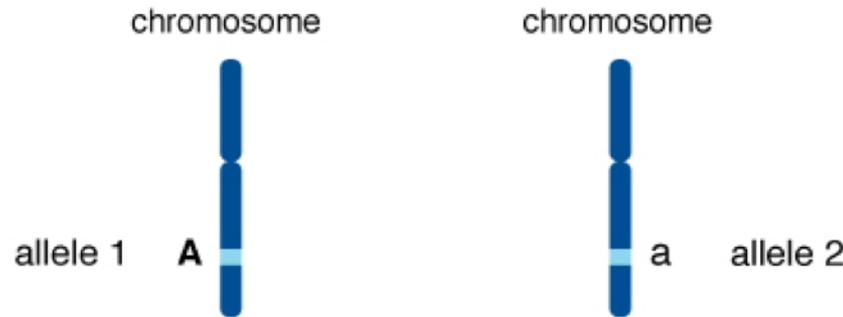
1 Génome : **Génotype vs phénotype**



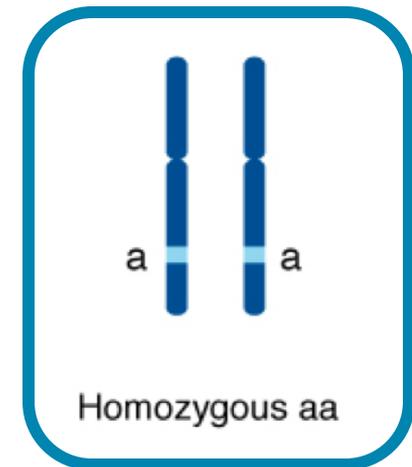
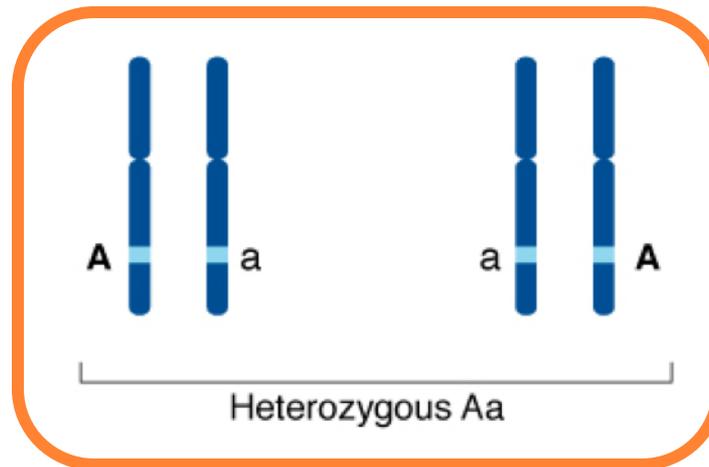
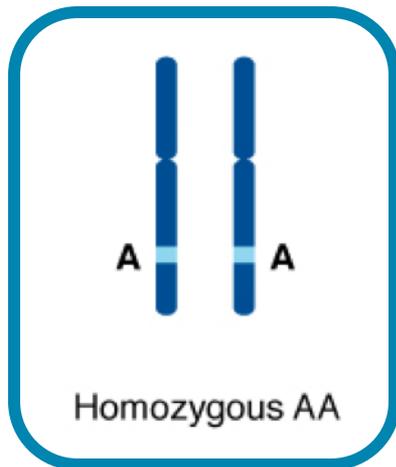
■ **Génome diploïde (2n chromosomes)**

⇒ **2 copies de chaque gène (1 de chaque parent)**

HOMOZYGOTE :
Le même allèle sur
les 2 chromosomes



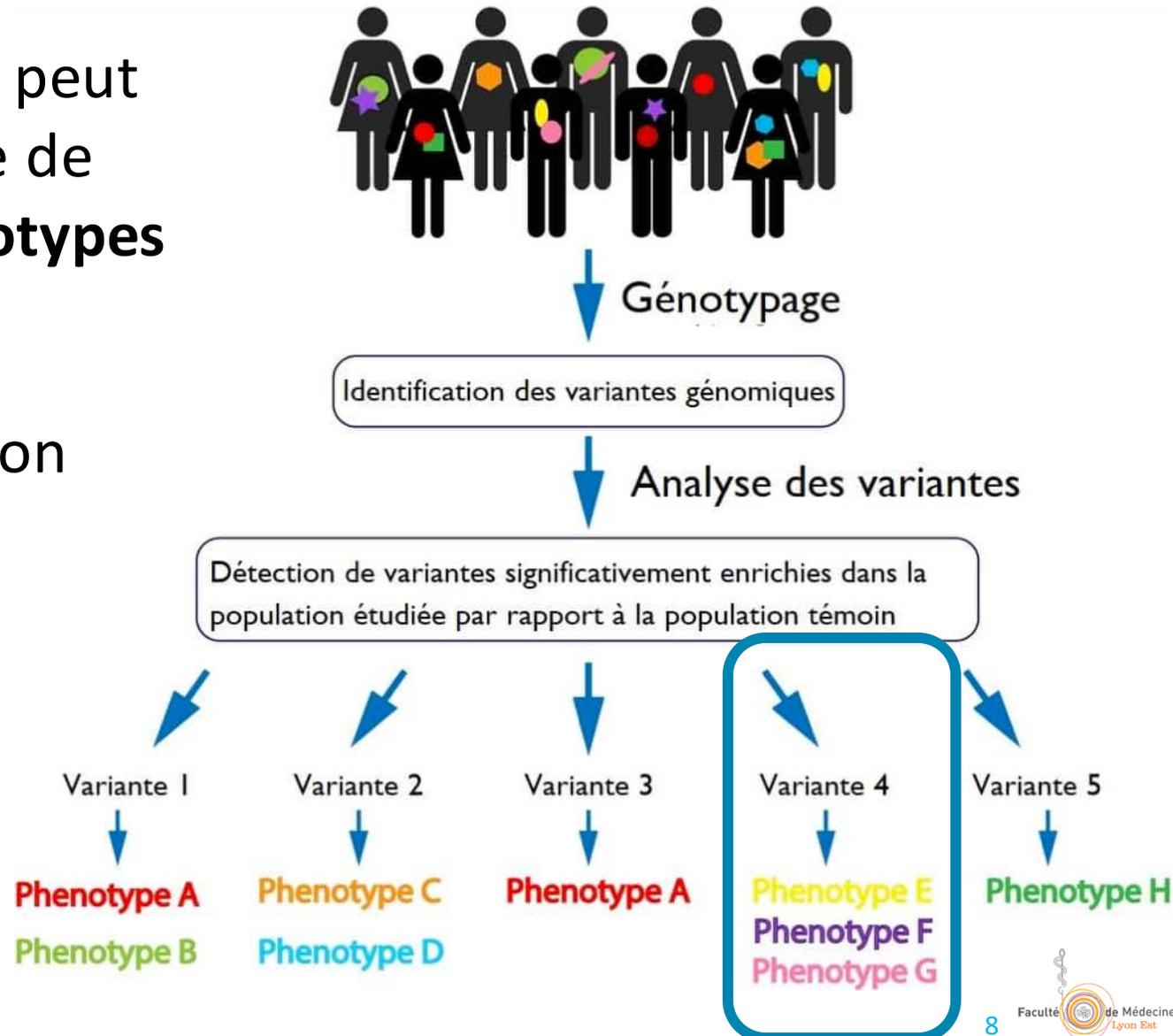
HETEROZYGOTE :
1 allèle différent
sur chaque chrom.



1 Génome : **Génotype vs phénotype**

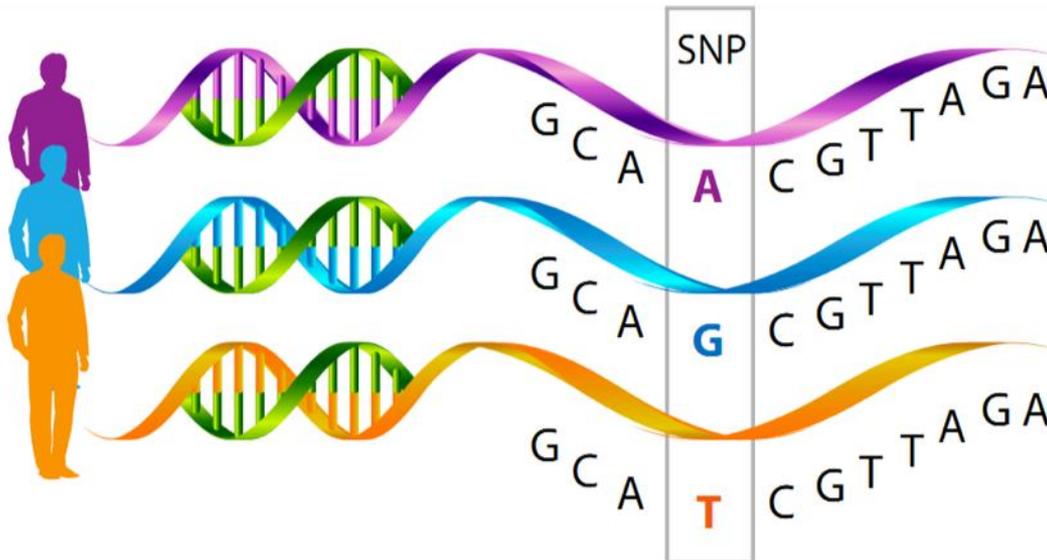
1 **allèle** (variant) peut être responsable de **différents phénotypes**

“**Expressivité**” = degré d’expression du phénotype (force du trait génétique)



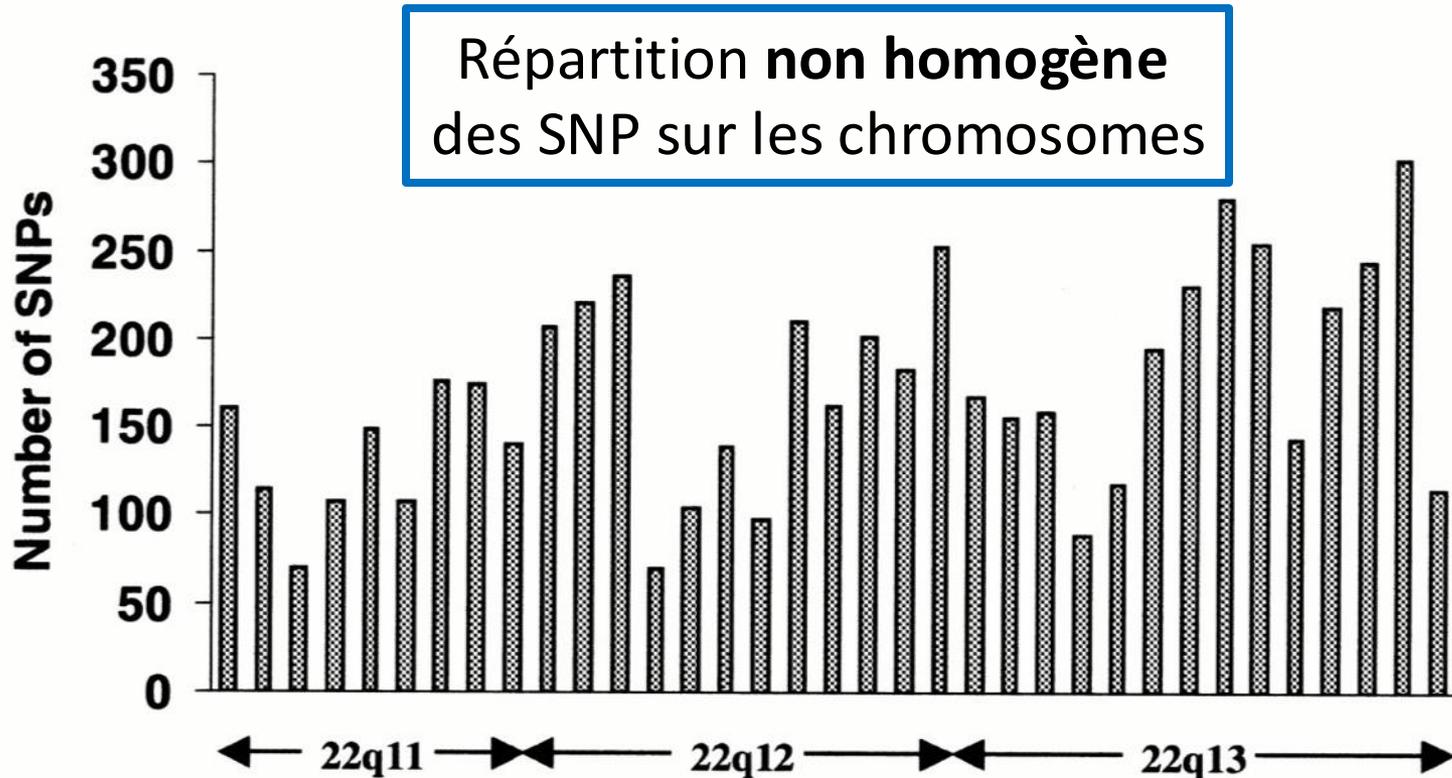
1 Génome : Polymorphisme génétique

- **Single Nucleotide Variant** : variant génétique portant sur 1 nucléotide
- **Single Nucleotide Polymorphism (SNP “snip”)** : SNV >1% dans la population saine
- **1 SNP tous les 100nts** dans le génome



1 Génome : Polymorphisme génétique

■ Single Nucleotide Polymorphism (SNP “snip”) :

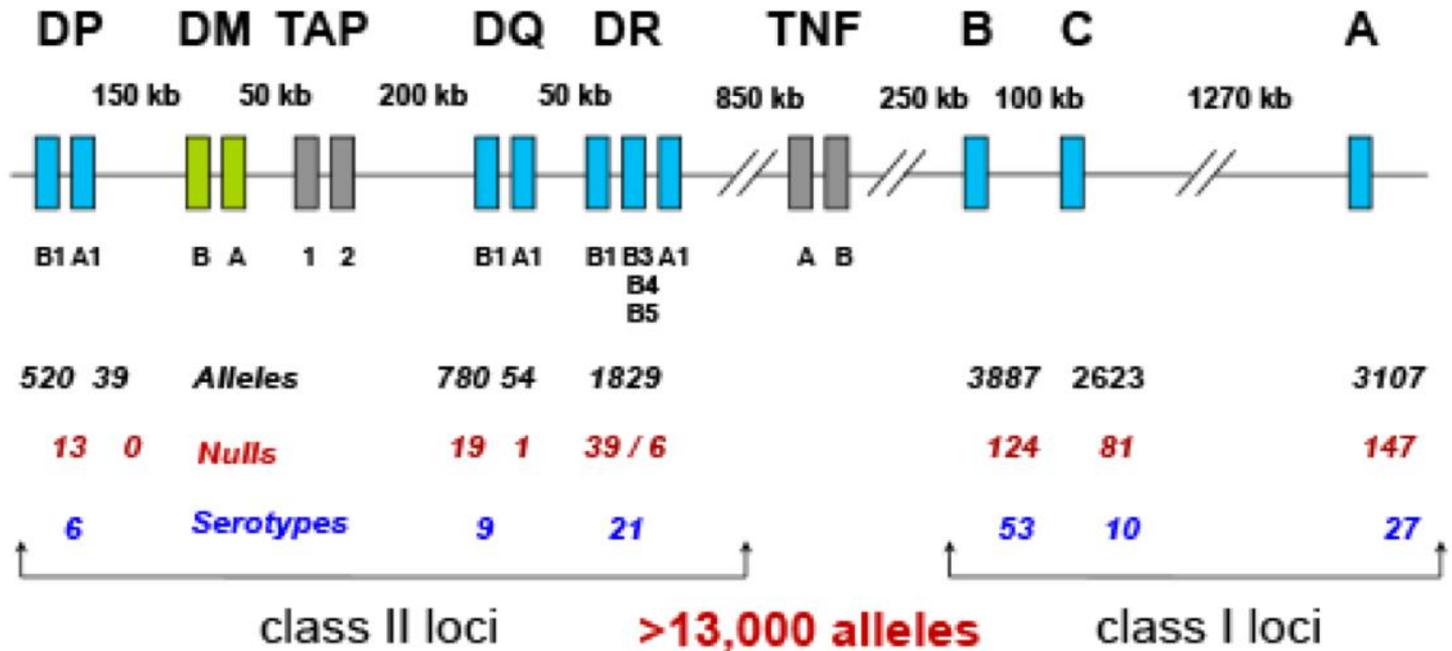


1 Génome : Polymorphisme génétique

■ Single Nucleotide Polymorphism (SNP “snip”) :

Certains loci sont hypervariables

Complex HLA
(6p21.31)

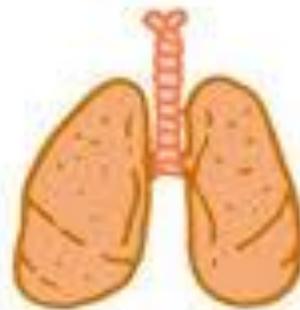


① Génome : Polymorphisme génétique

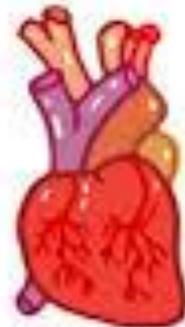
■ Single Nucleotide Polymorphism (SNP “snip”) :

Certains loci sont hypervariables

Complex HLA
(6p21.31)



Poumons



Coeur



Foie



Rein

Compatibilité des greffons

Autogreffe

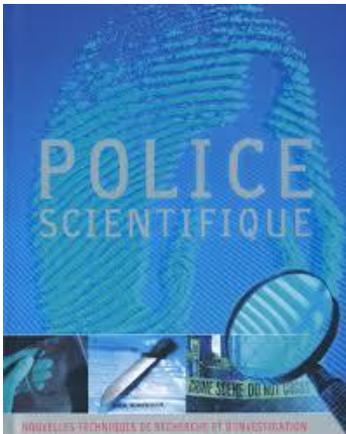
Allogreffe

(apparentés au 1^{er} degré)

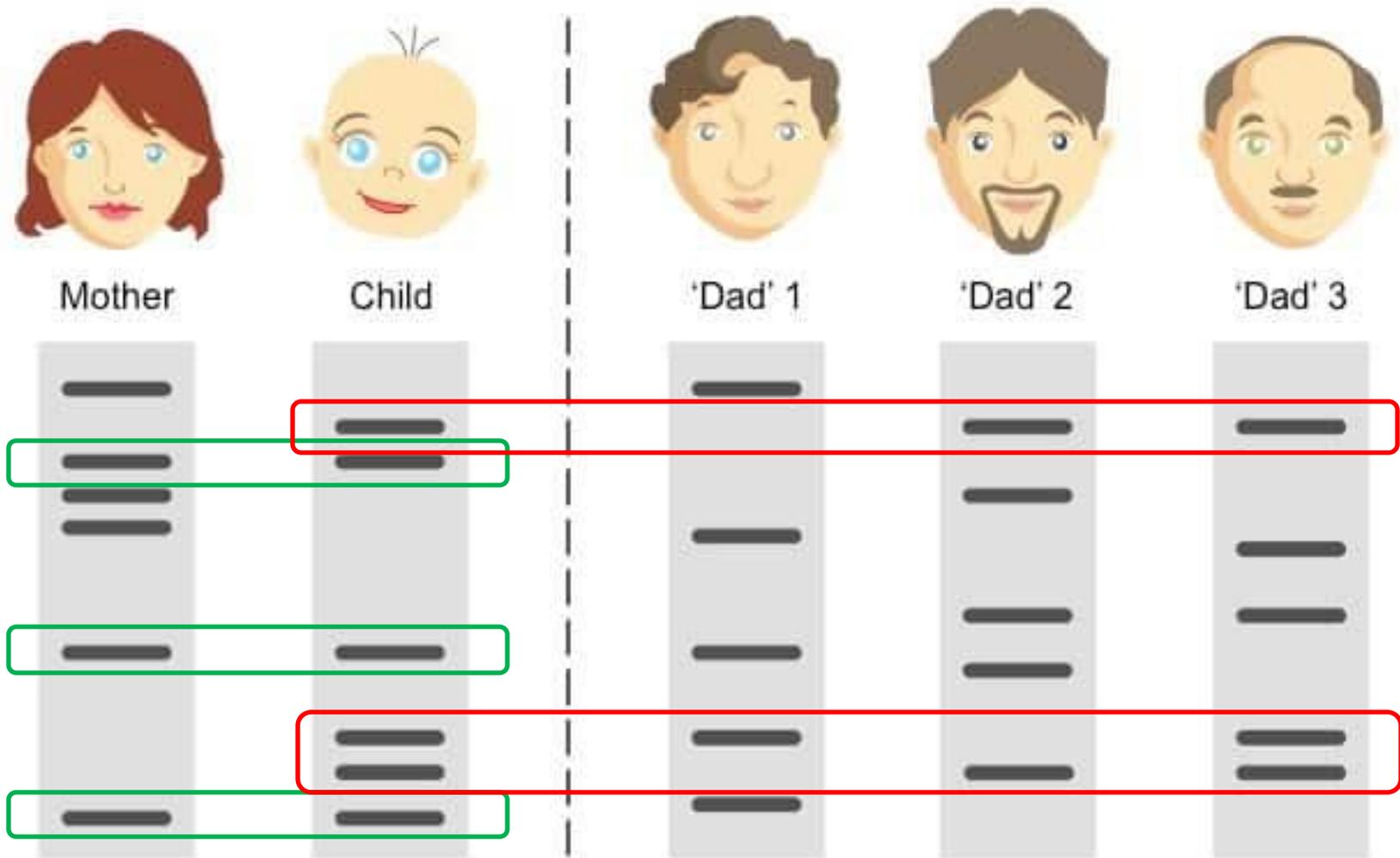
① Génome : Polymorphisme génétique

■ Single Nucleotide Polymorphism (SNP “snip”) :

En comparant 2 individus
⇒ 1 SNP tous les
1000/2000 bases



1 Génome : Polymorphisme génétique



① Génome : Polymorphisme génétique

■ Impact fonctionnel d'un SNV

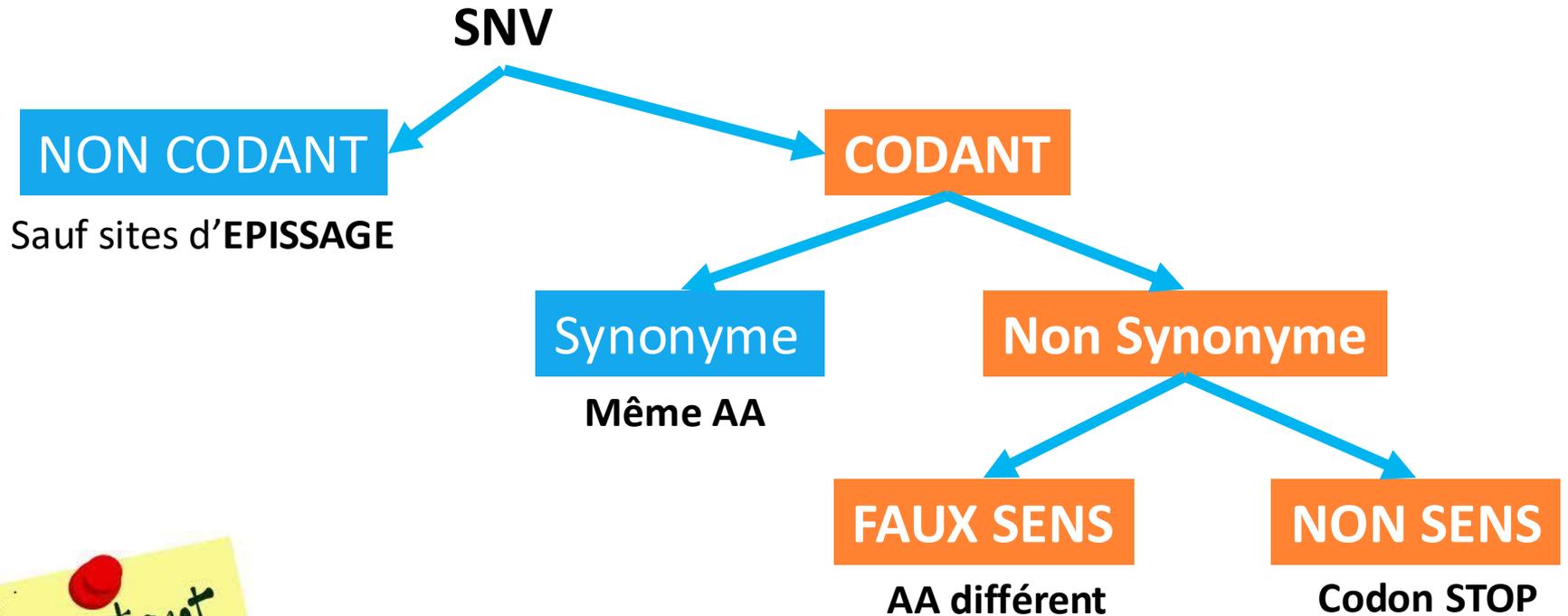
Variant bénin	Variant probablement bénin	Variant de signification incertaine	Variant probablement pathogène	Variant pathogène
Classe 1	Classe 2	Classe 3	Classe 4	Classe 5

La plupart des SNV ne sont pas pathogènes !

(voir cours d'Alexandre Janin en MMOK)

1 Génome : Polymorphisme génétique

■ Impact fonctionnel d'un SNV





1 Génome : Polymorphisme génétique



Chromosome

Reference



Single nucleotide variant



Small insertion <50pb



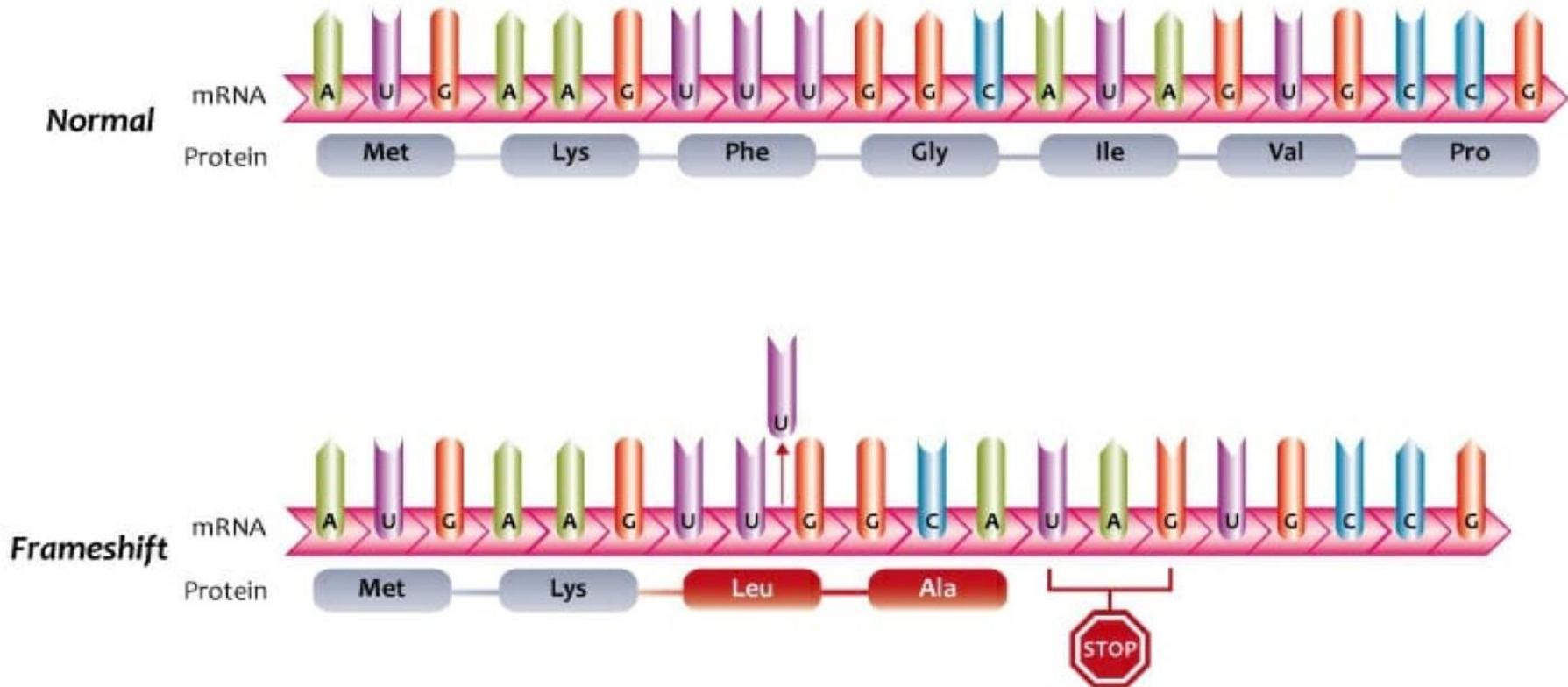
Small deletion



Sequence variants

Important

1 Génome : Polymorphisme génétique



Insertions/délétions (\neq multiple de 3)
⇒ décalage du cadre de lecture (“frameshift”)
⇒ codon STOP prématuré ⇒ protéine tronquée

1 Génome : Polymorphisme génétique



Reference



Sequence variants

Single nucleotide variant



Small insertion
<50pb



Small deletion



Structural variants

Deletion



Duplication



Inversion



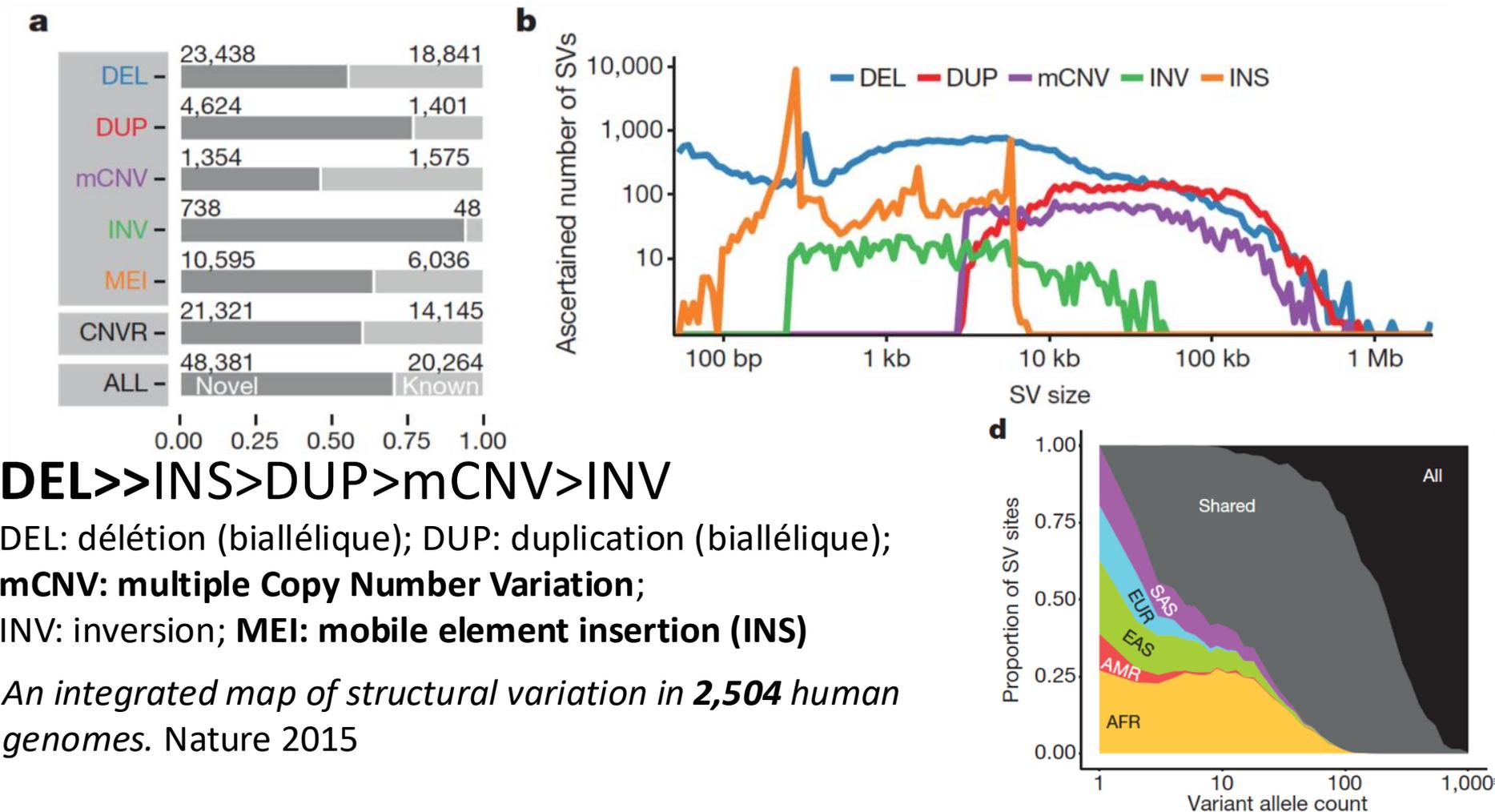
Translocation



variant de structure : >50bp

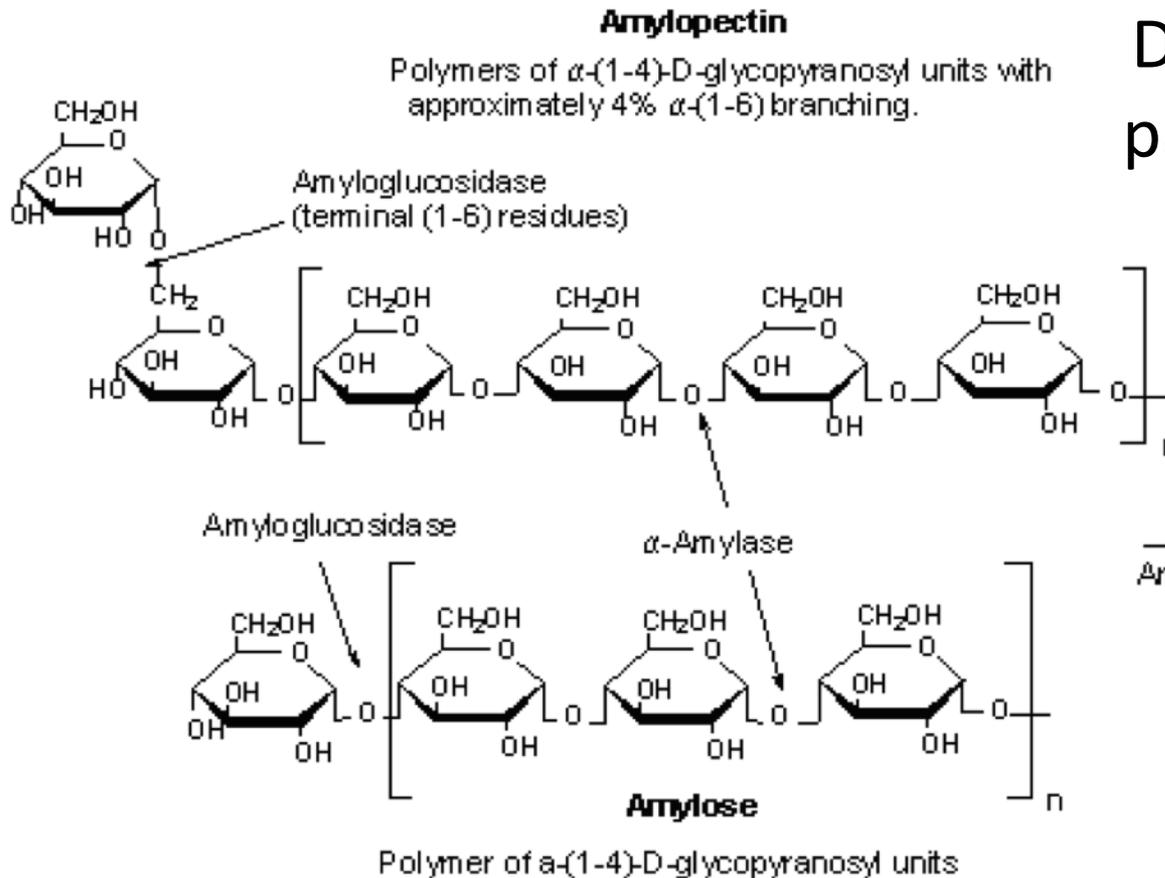
1 Génome : Polymorphisme génétique

Les polymorphismes structuraux sont fréquents :
13% du génome



1 Génome : Polymorphisme génétique

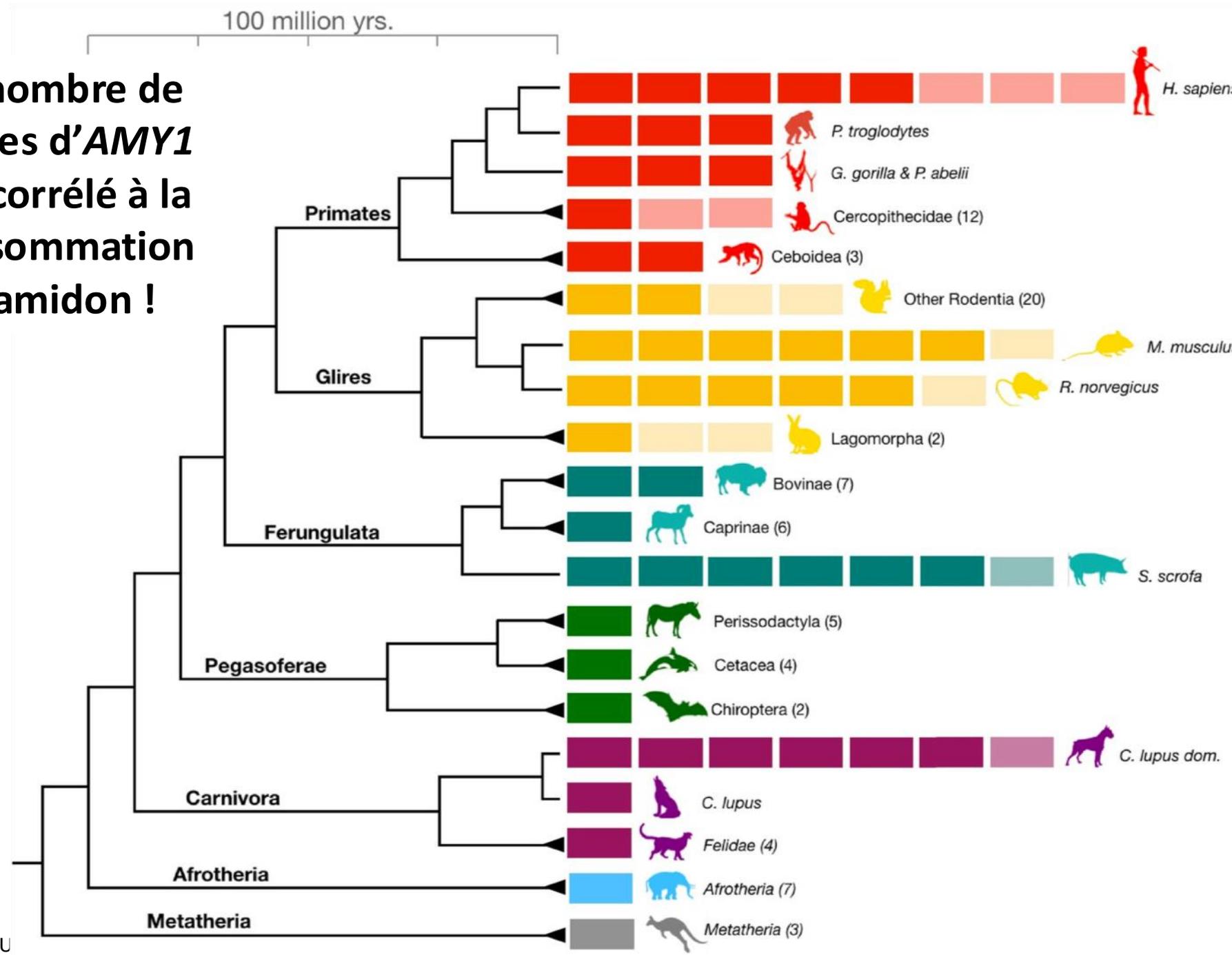
Copy Number Variation



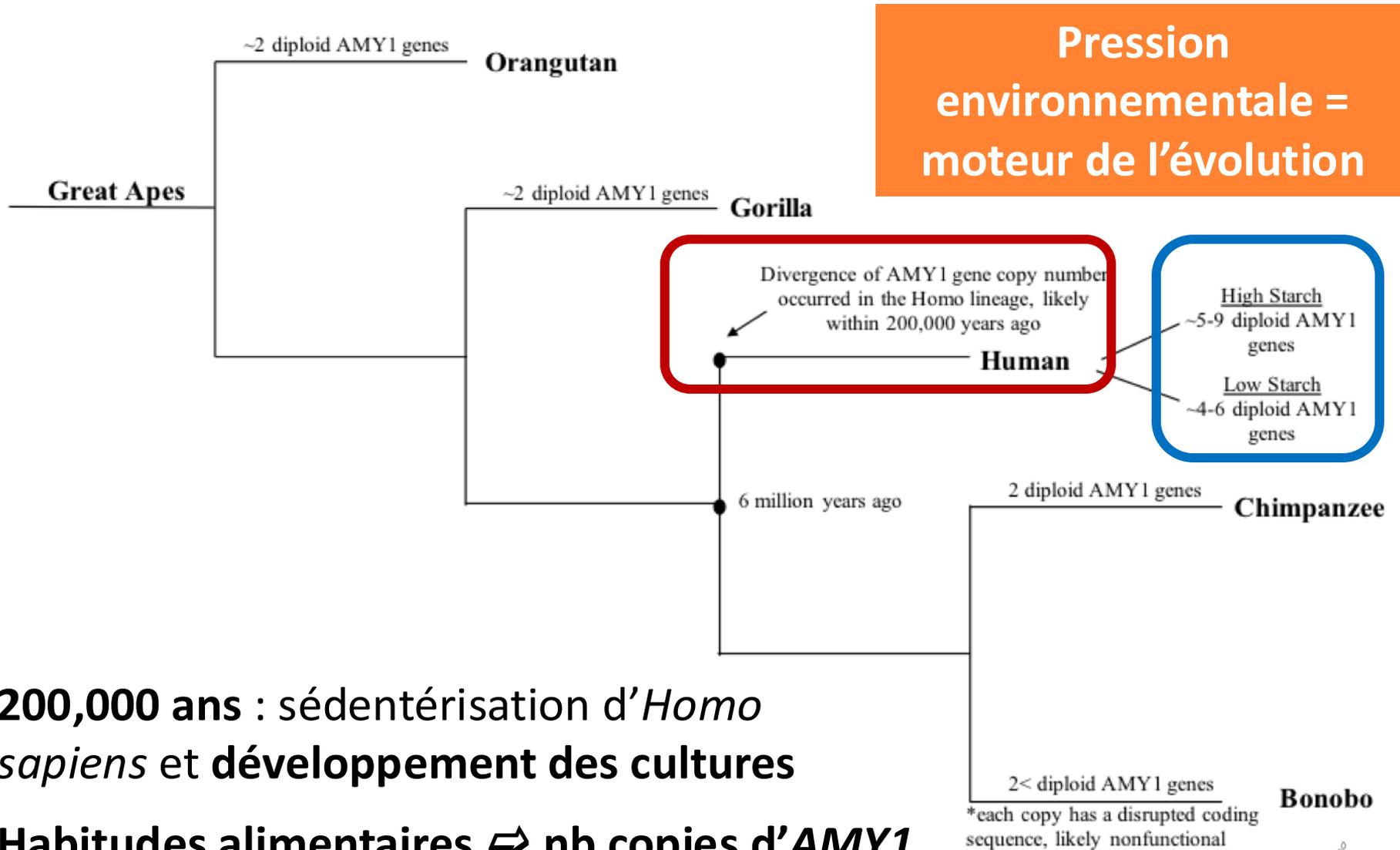
Digestion de l'**AMIDON**
par l' α -**amylase** salivaire
(codée par **AMY1**)



Le nombre de copies d'AMY1 est corrélé à la consommation d'amidon !



1 Génome : Polymorphisme génétique



200,000 ans : sédentérisation d'*Homo sapiens* et développement des cultures

Habitudes alimentaires ⇔ nb copies d'*AMY1*

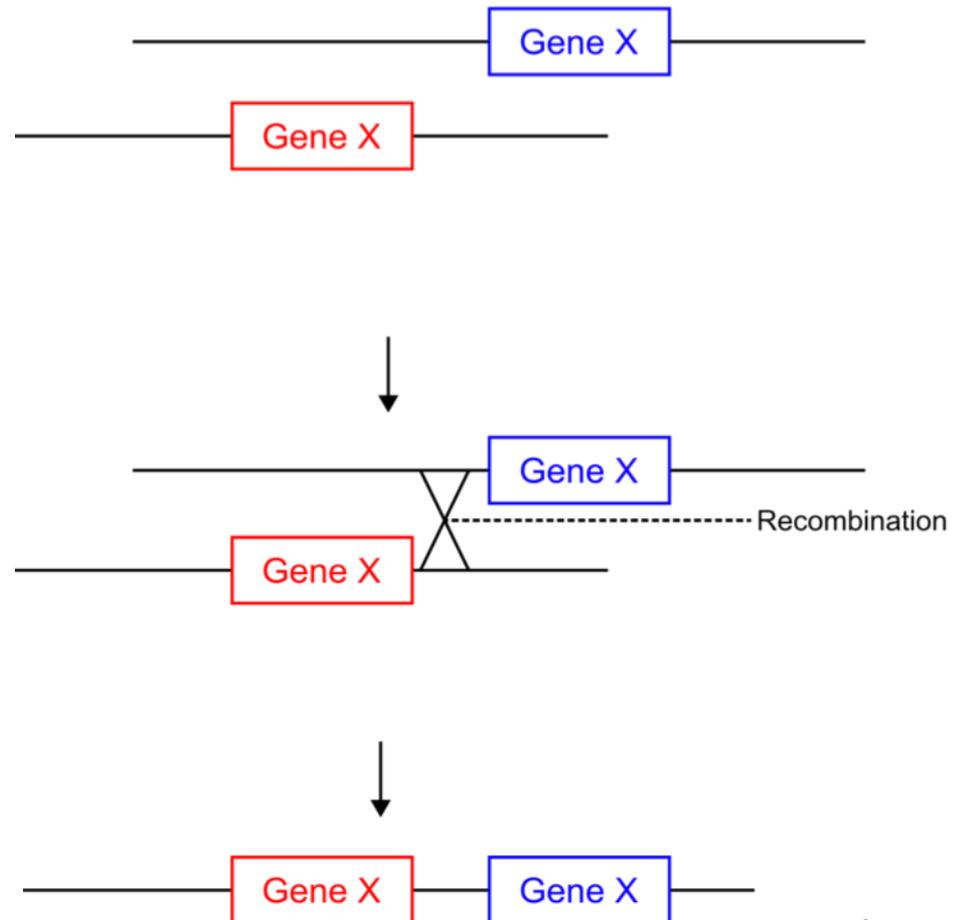
1 Génome : Polymorphisme génétique

Copy Number Variation

Mésalignement des 2 chromosomes portant le gène *AMY1*

Recombinaison homologue “non-allélique”

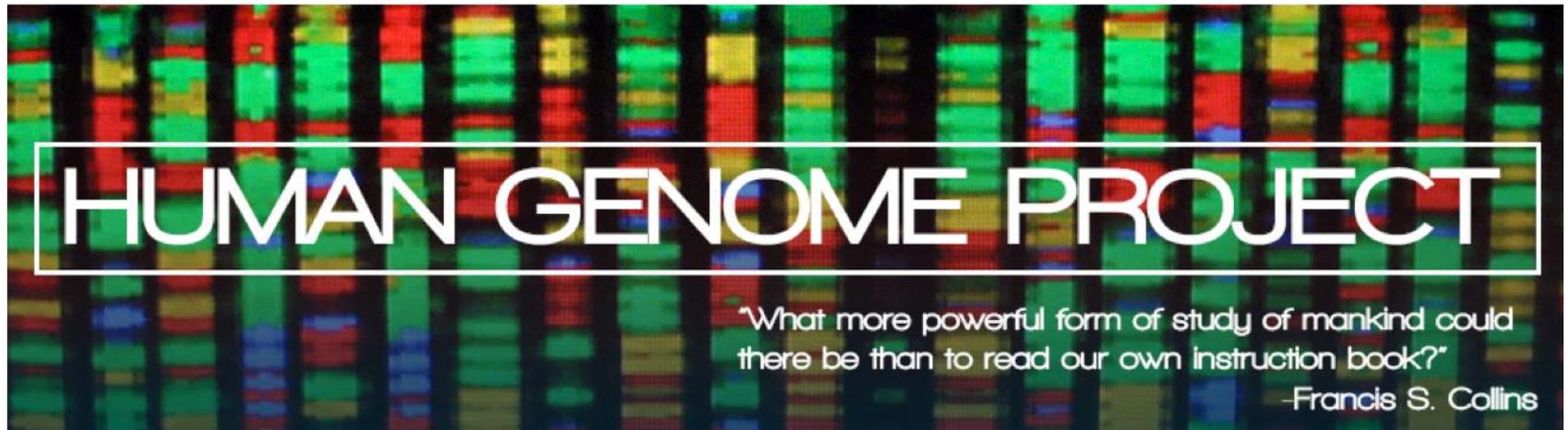
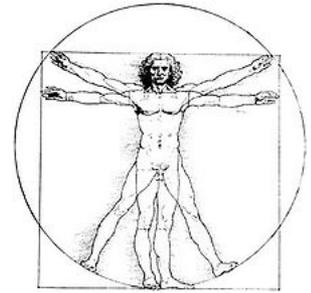
Chromosome avec 2 copies du gène *AMY1*



0 Plan

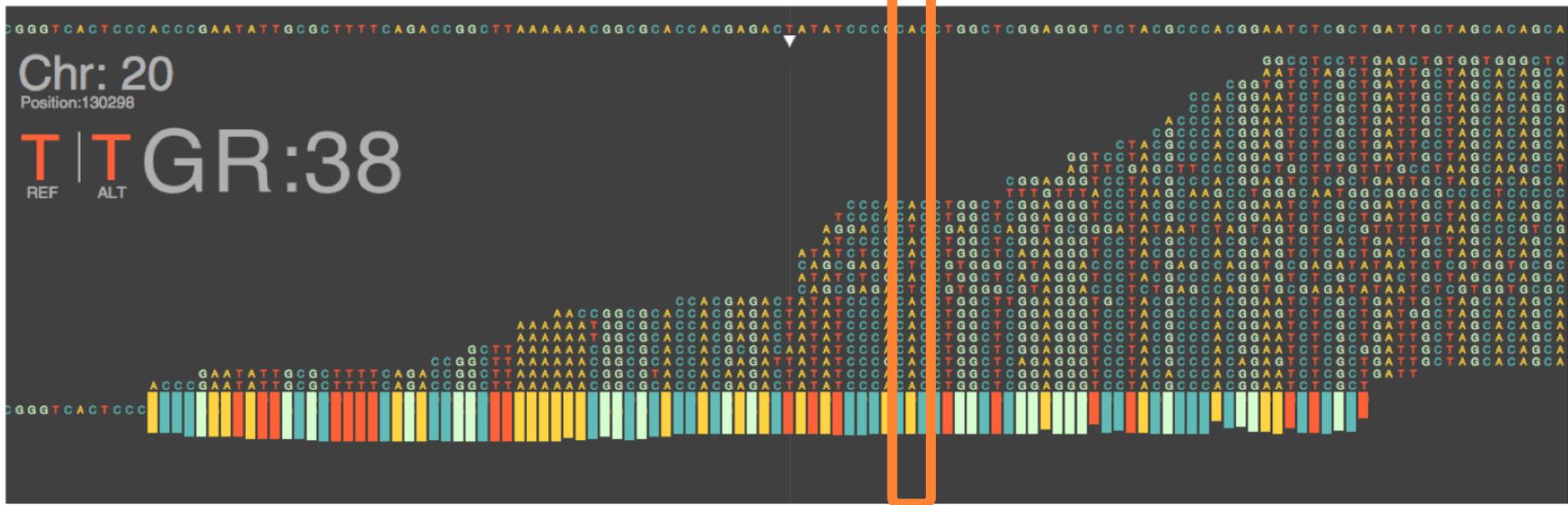
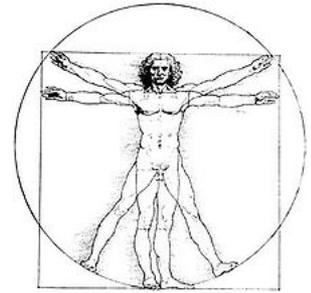
- **Organisation du génome humain**
 - Génotype vs phénotype
 - Polymorphisme génétique
 - **Génome de référence**
 - Composition du génome
- Gènes
- ADN intergénique

1 Génome : **Génome de référence**



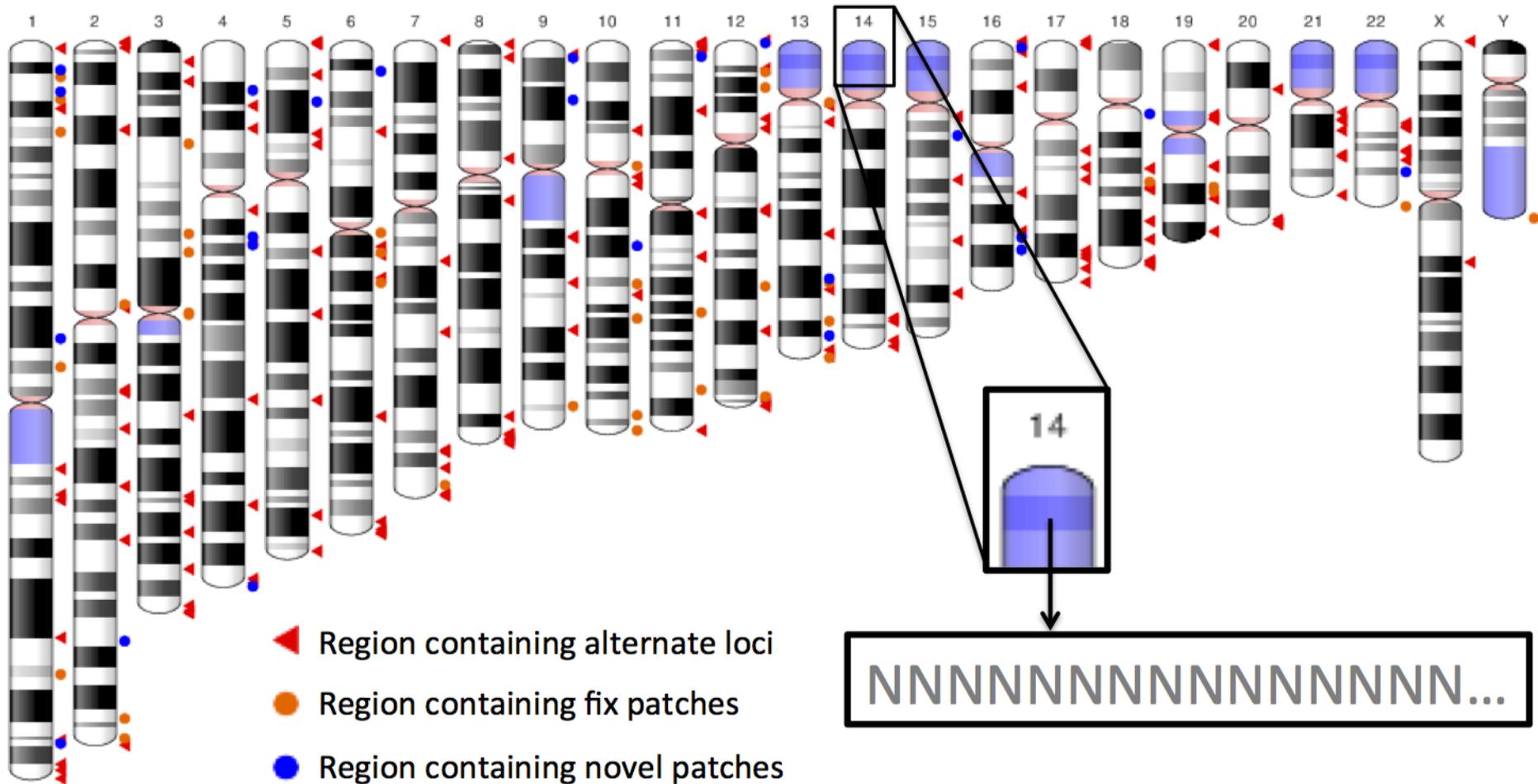
- **13 ans (1990-2003), 3 milliards de \$**
- 20 équipes (USA, UK, Chine, Japon, Allemagne et France)
- **Génome humain haploïde de référence = mosaïque de quelques individus**

1 Génome : Génome de référence



En cas de SNP \Rightarrow l'allèle le plus fréquent est retenu

1 Génome : Génome de référence



Certaines régions (télomères +++) étaient mal séquencées

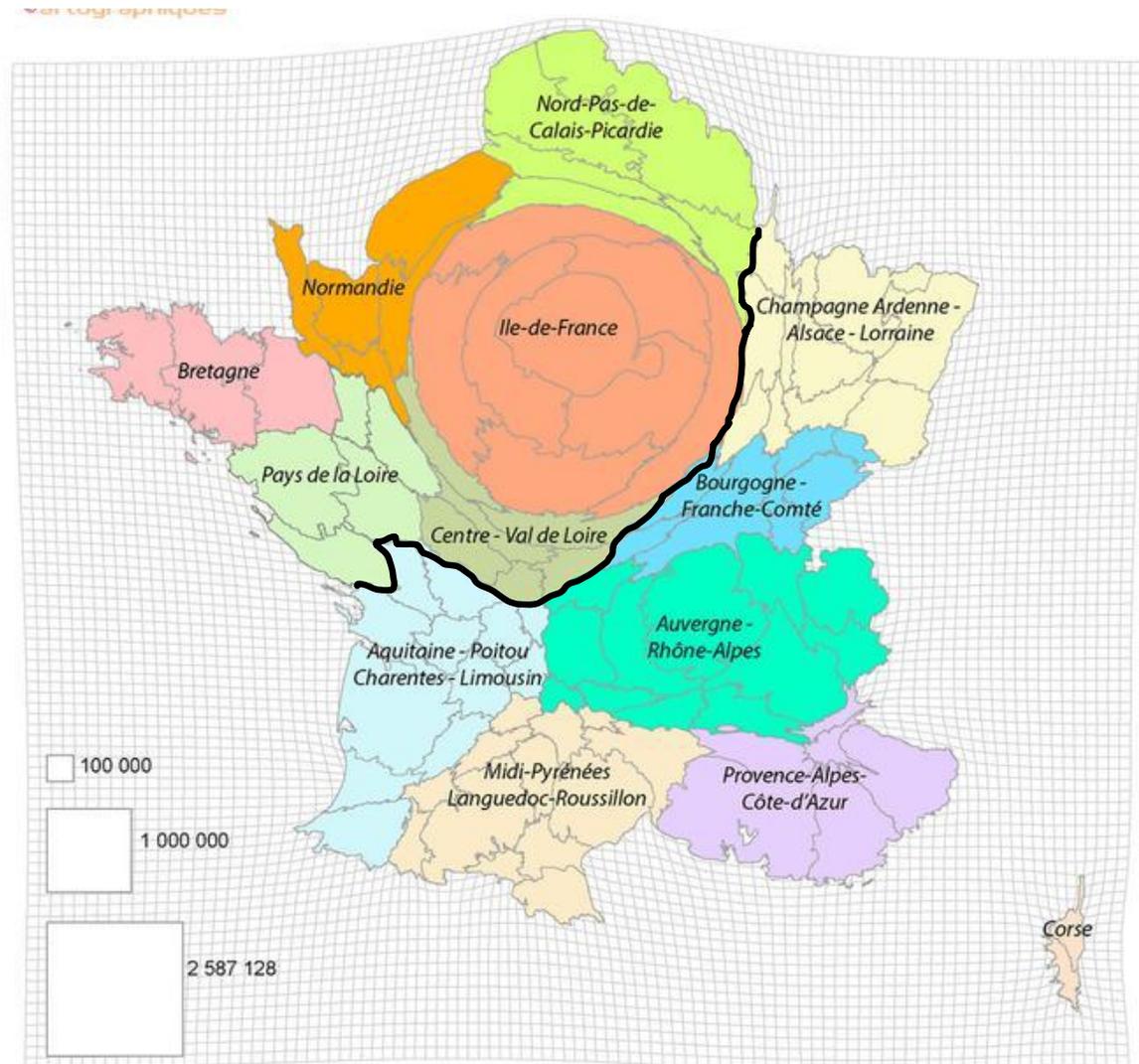
1 Génome : Plan France Médecine Génomique

SeqOIA
LA PLATEFORME GÉNOMIQUE
DE PARIS RÉGION

*Sequencing, Omics,
Information Analysis*

 **AURA
GEN**

Auvergne Rhône Alpes
Genomique



1 Génome : Plan France Médecine Génomique

- “Séquençage h
- 48 WGS simulta
- Environ **1000€**
- Résultat “brut”
<1 semaine



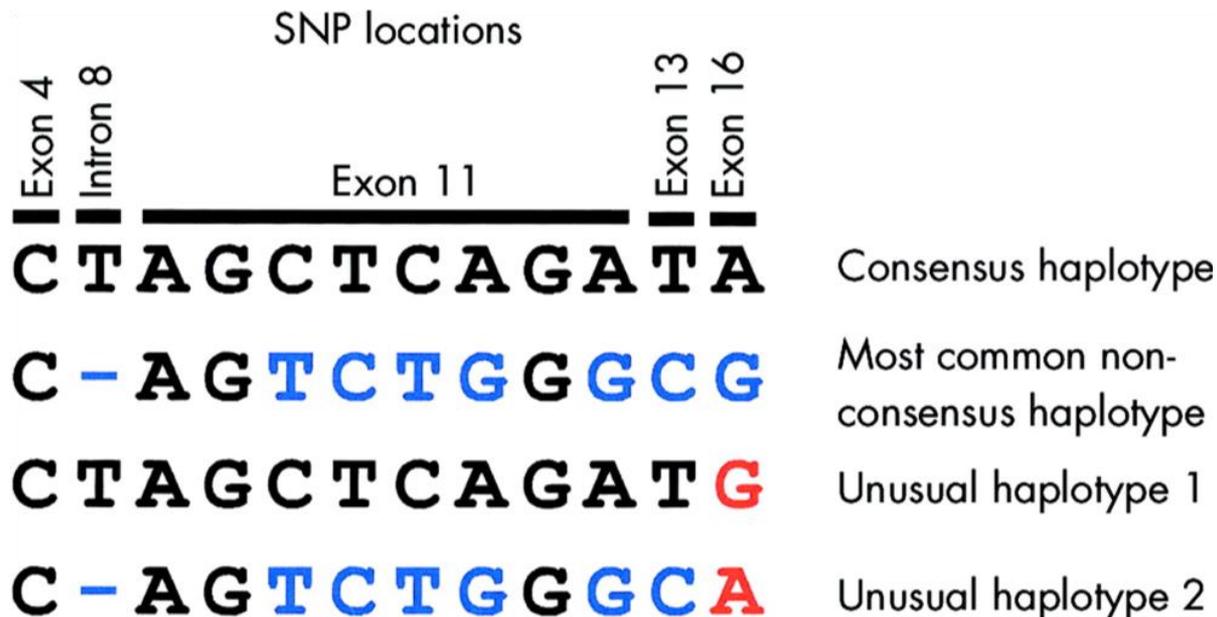
Vue synthétique de l'ensemble des altérations





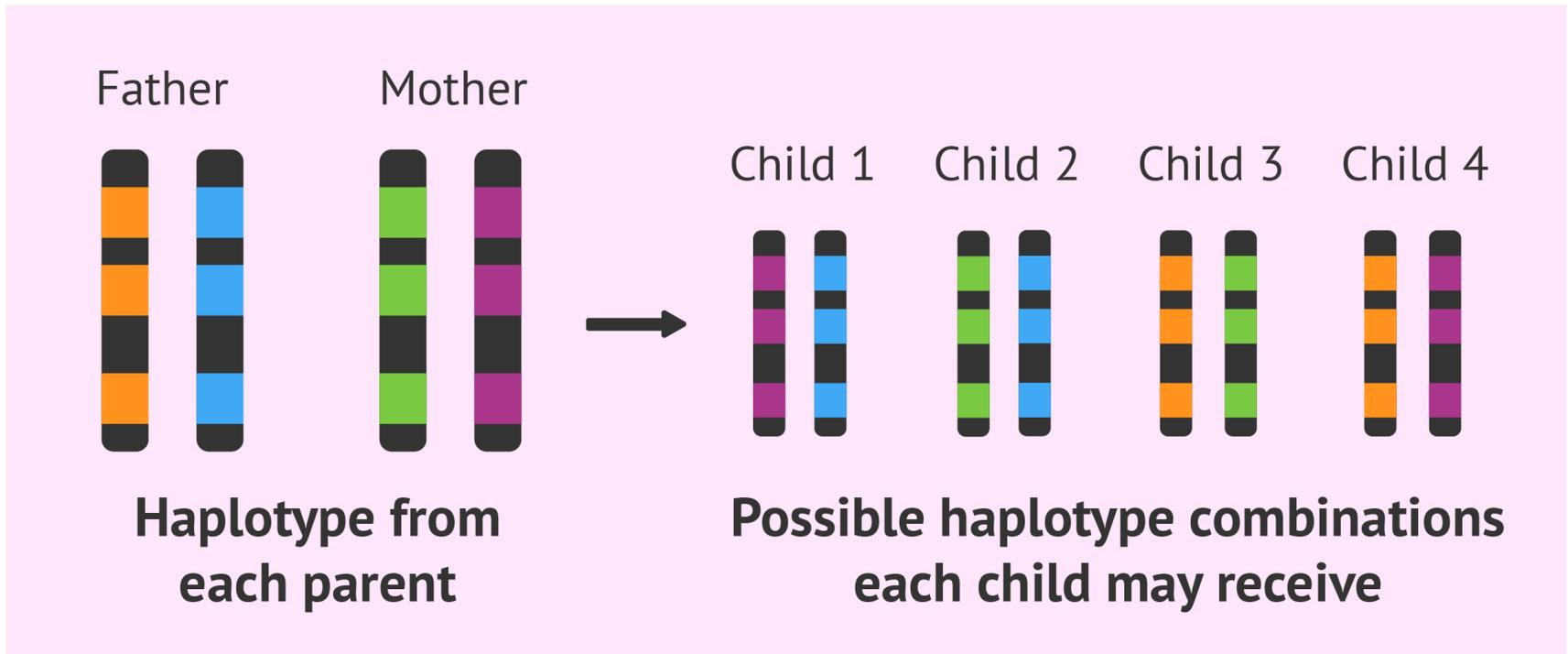
1 Génome : Haplotypes

- Haplotype = groupe de SNPs séquentiels sur un même chromosome
- A l'intérieur de chaque haplotype **forte liaison entre les SNP** (très peu de recombinaisons, transmis en bloc)
⇒ étude de la **transmission** intra-familiale



1 Génome : Haplotypes

- **Haplotypage = identification des chromosomes parentaux transmis**



Posez vos questions pendant le cours



1 Allez sur
wooclap.com

2 Entrez le code
d'événement dans
le bandeau
supérieur

Code
d'événement
PJINGO

0 Plan

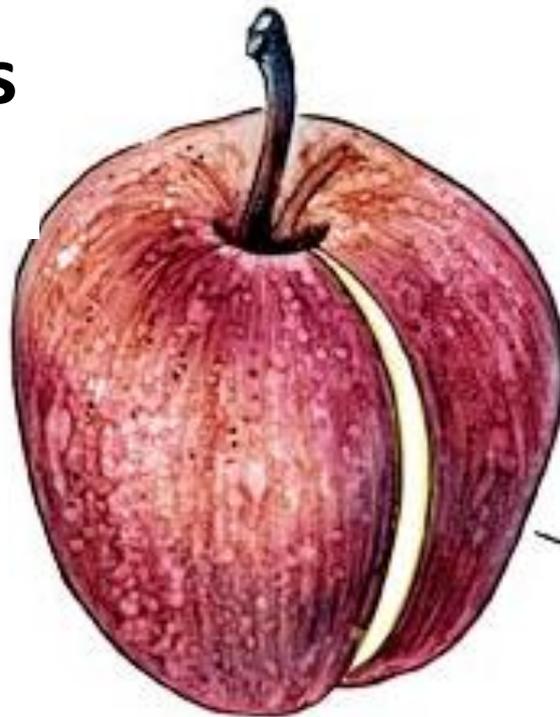
- **Organisation du génome humain**
 - Génotype vs phénotype
 - Polymorphisme génétique
 - Génome de référence
 - **Composition du génome**
- Gènes
- ADN intergénique

① Génome : **Composition du génome**

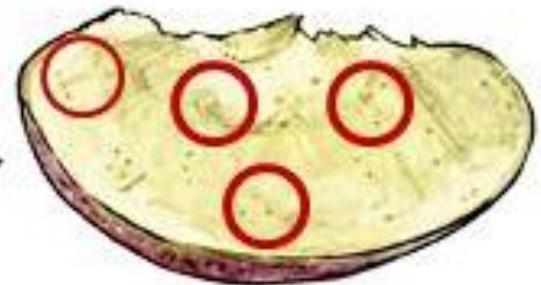
Important

- Seuls 1,5% du génome code pour des protéines (21000 gènes) ⇔ « exome »

Génome
(3,2 milliards de bases)



Exome = 1,5%
(48 millions de bases)



3 Génome : Composition du génome

Taille de génomes haploïdes
(en Millions de bases, Mégabases, Mb)

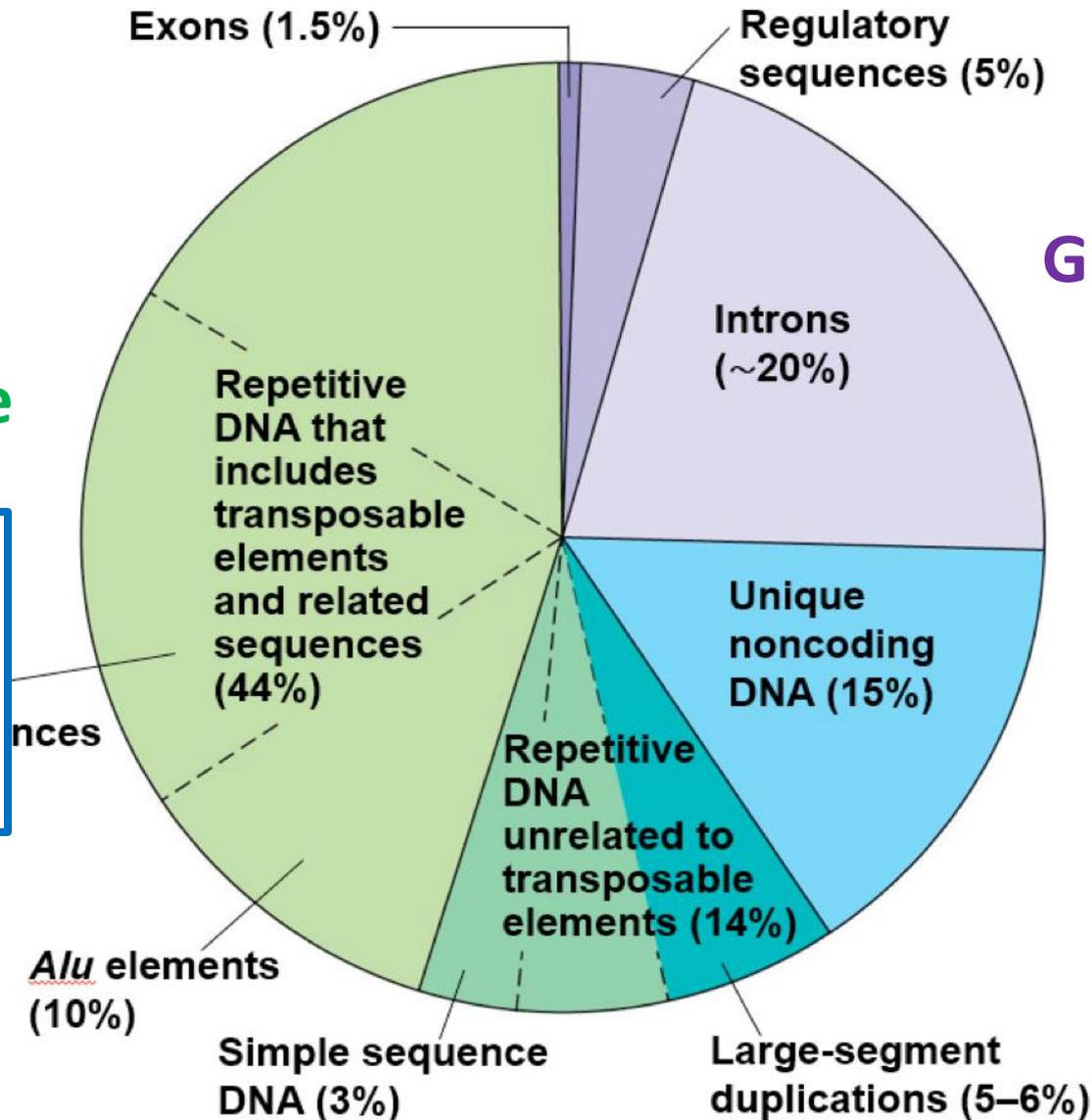
• E. coli	4.5
• S. cerevisiae (levure)	14.4
• C. elegans	100
• D. melanogaster	165
• H. sapiens	3,000
• Arabidopsis thaliana	130
• L. esculentum (tomate)	1,000
• N. tabacum (4x)	4,500
• N. plumbaginifolia	2,300
• B.napus (colza)	1,200
• O. sativa (riz)	420
• Z. mais	2,500
• T. aestivum (blé)	16,000
• Tulipa	30,000

**Génome humain
3 milliards (3 Gb)
de nucléotides**

1 Génome : Composition du génome

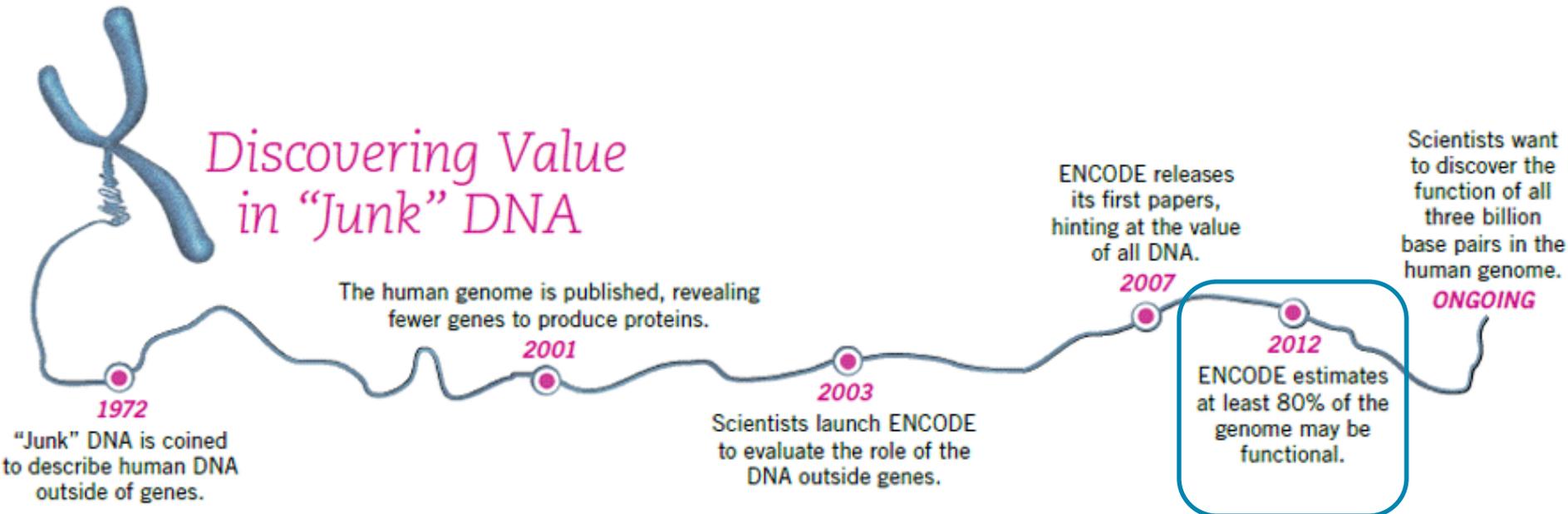
ADN
intergénique

2/3 du
génom est
constitué de
"JUNK DNA"



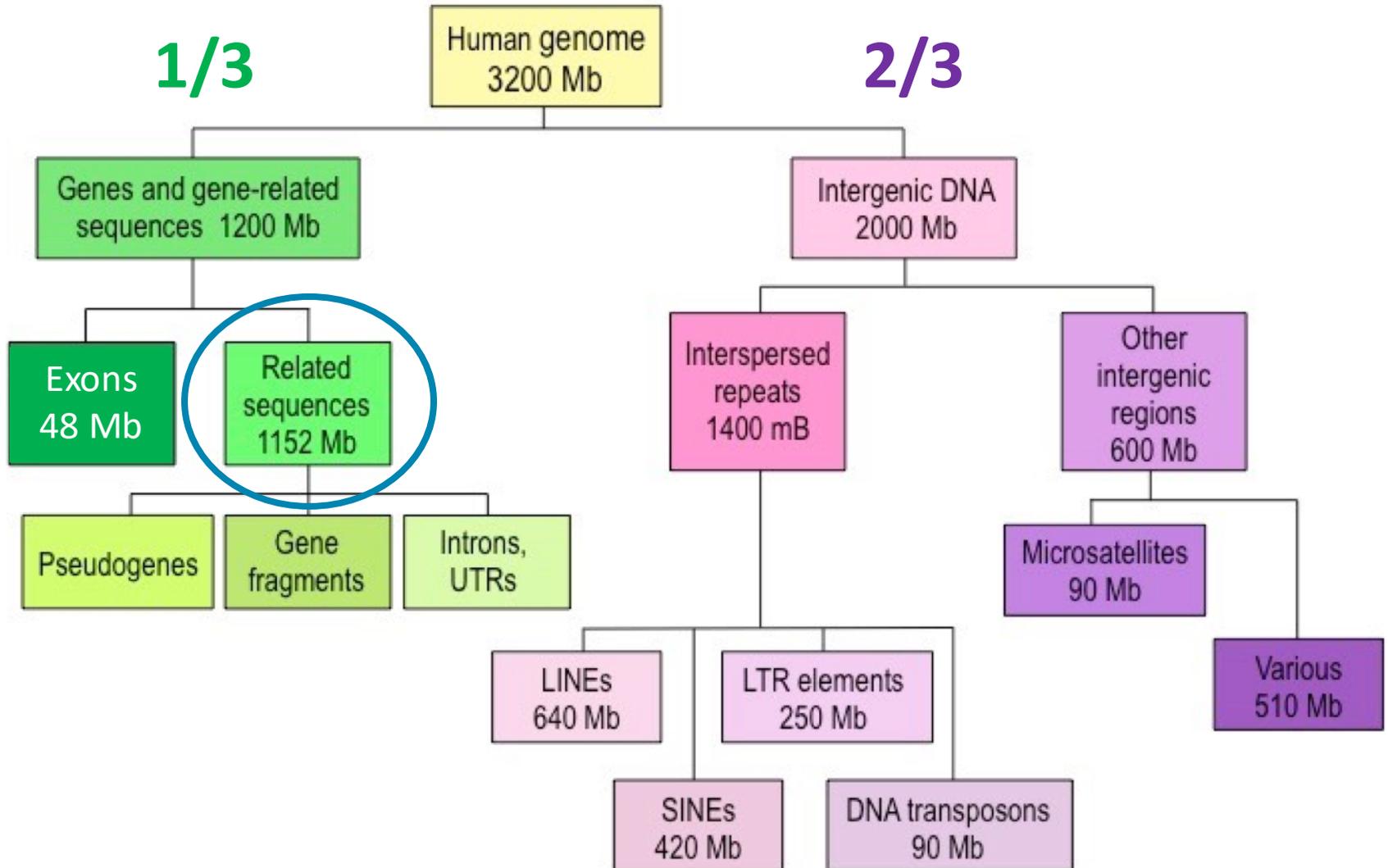
1 Génome : Composition du génome

Discovering Value in "Junk" DNA



Très probablement pas si inutile que ça...

1 Génome : Composition du génome



1 Génome : Composition du génome

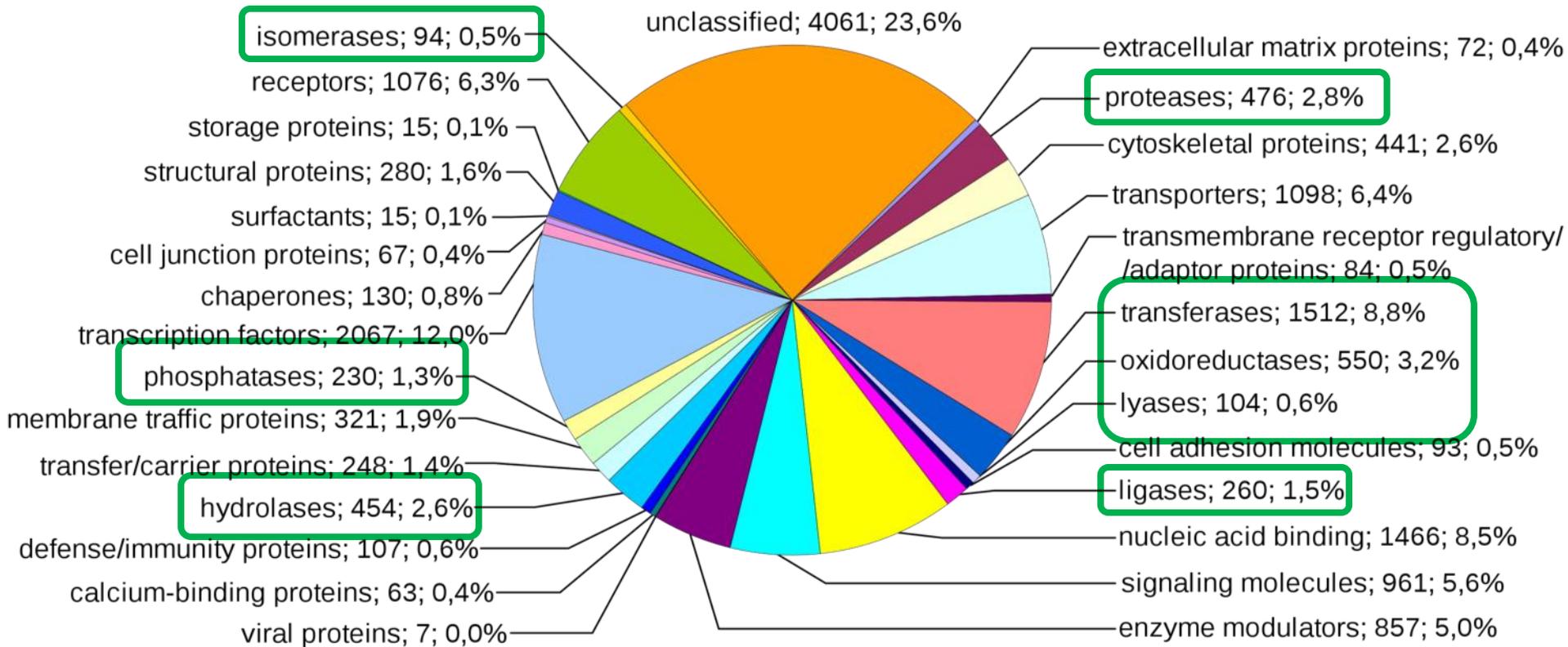
Table 2. Number of human genes in different databases as of July 2018[18]

	Gencode ^[19]	Ensemble ^[20]	Refseq ^[21]	CHESS ^[22]
protein-coding genes	19,901	20,376	20,345	21,306
lncRNA genes	15,779	14,720	17,712	18,484
antisense RNA	5501		28	2694
miscellaneous RNA	2213	2222	13,899	4347
Pseudogenes	14,723	1740	15,952	
total transcripts	203,835	203,903	154,484	328,827

Gènes codants \approx gènes non codants > pseudogènes

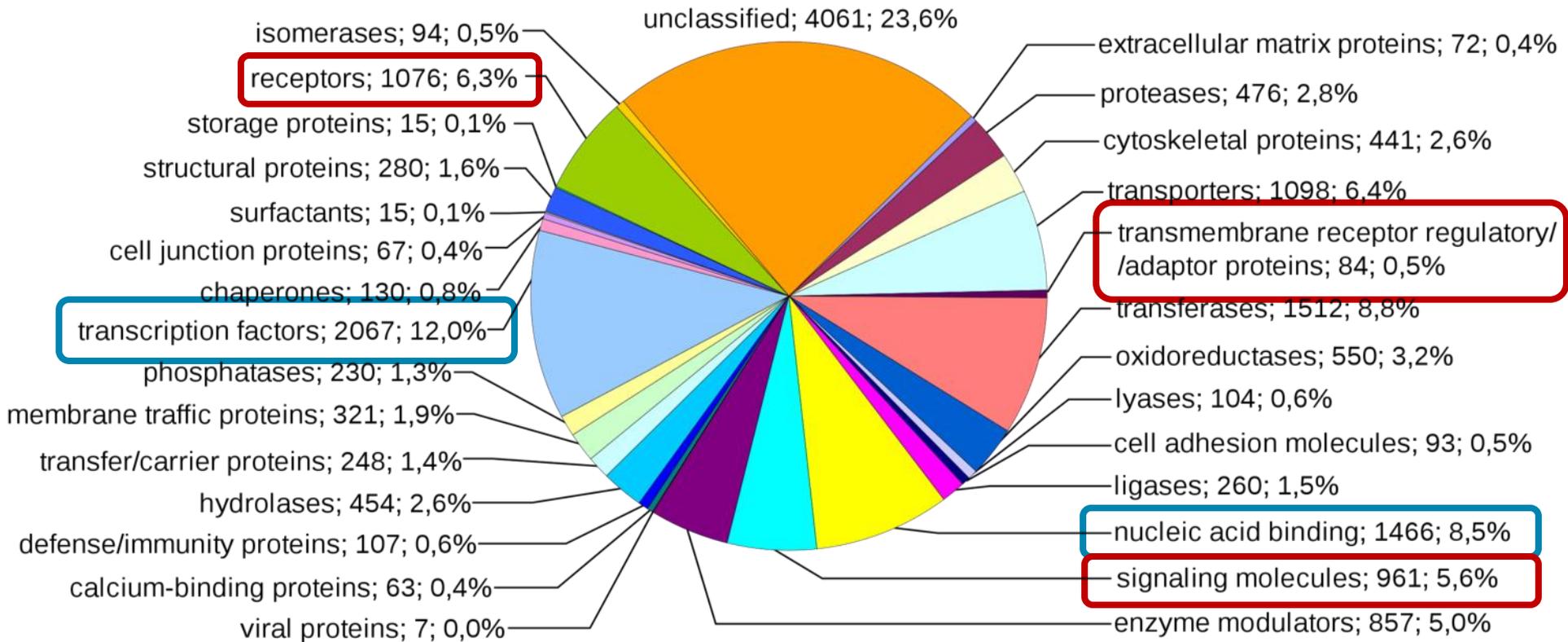
Nb transcrits \gg nb de gènes (1 gène \Leftrightarrow +s transcrits)

1 Génome : Composition du génome



Les gènes codent pour des protéines de fonctions différentes

1 Génome : Composition du génome

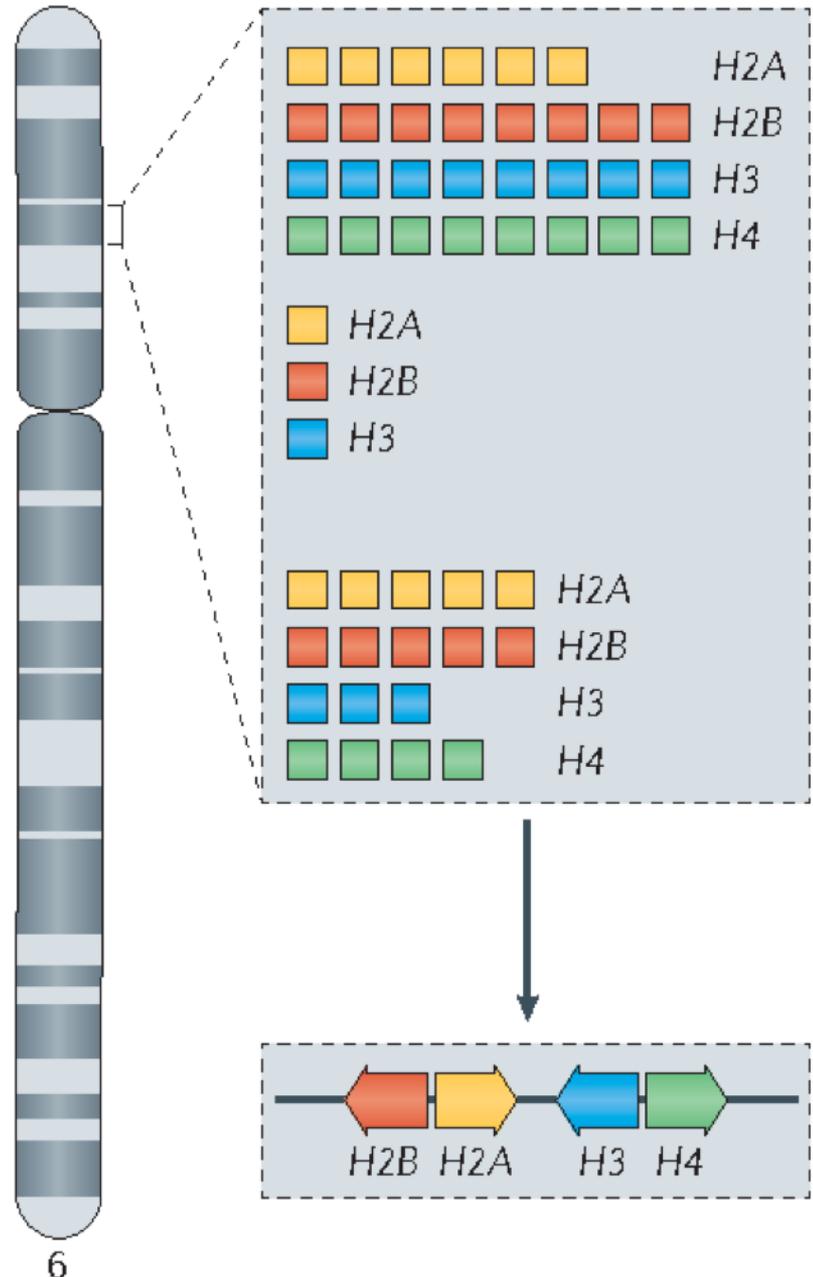


Les gènes codent pour des protéines de fonctions différentes

1 Génome : Composition du génome

Les gènes sont souvent regroupés par homologie de séquence et fonction biologique

⇒ **“CLUSTER”**

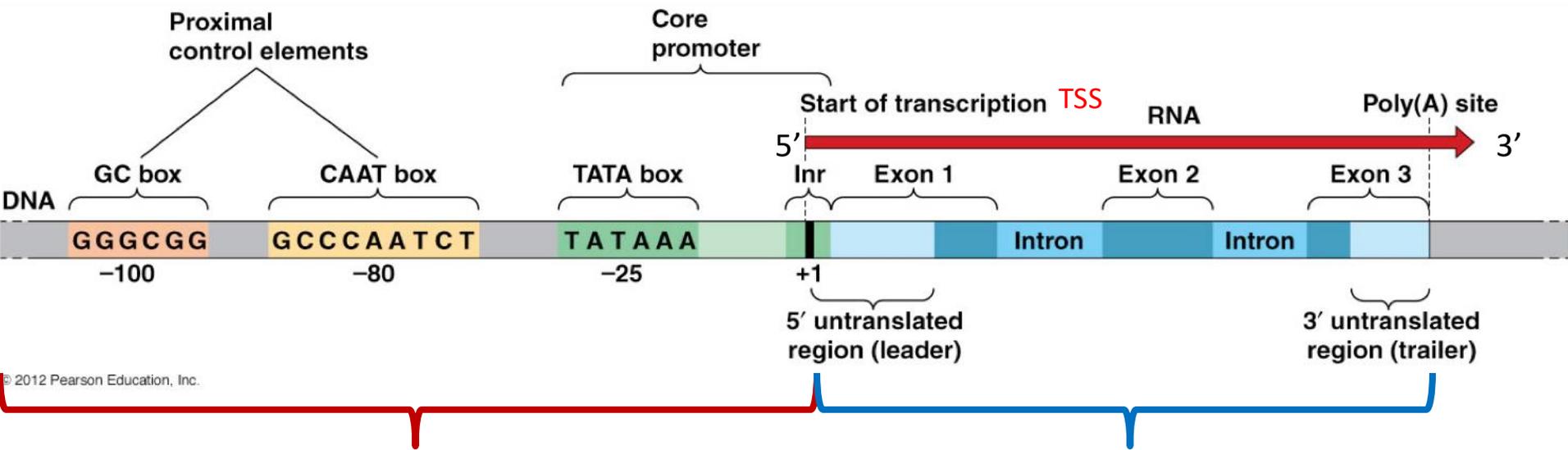


0 Plan

- Organisation du génome humain
- **Gènes**
 - **Structure d'un gène eucaryote codant**
 - **Pseudogènes**
- ADN intergénique

Important

2 Génome : Structure d'un gène eucaryote codant



© 2012 Pearson Education, Inc.

Régions "cis" régulatrices

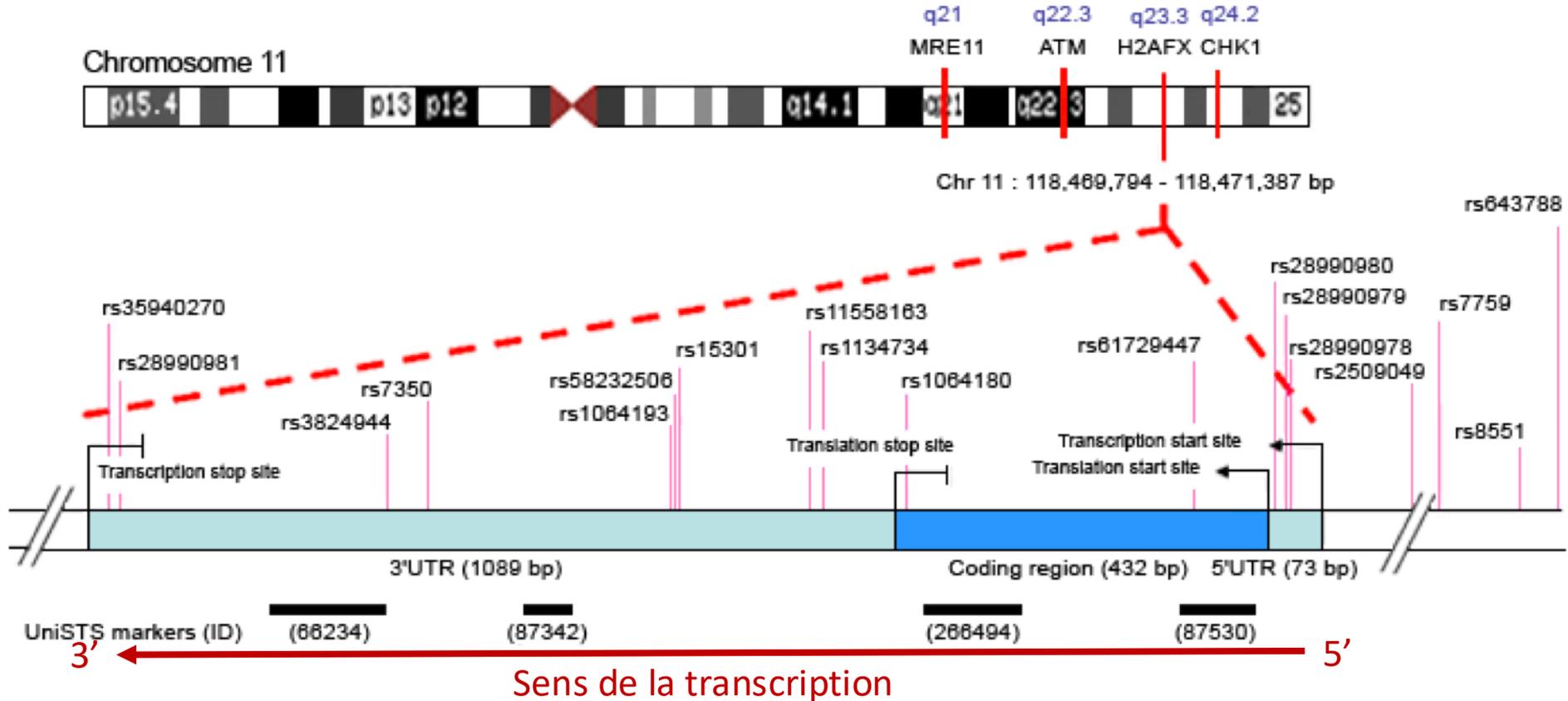
⇒ contrôle de la transcription
(voir cours d'Alexandre Janin + cours "régulation")

Région transcrite

⇒ pré-ARNm

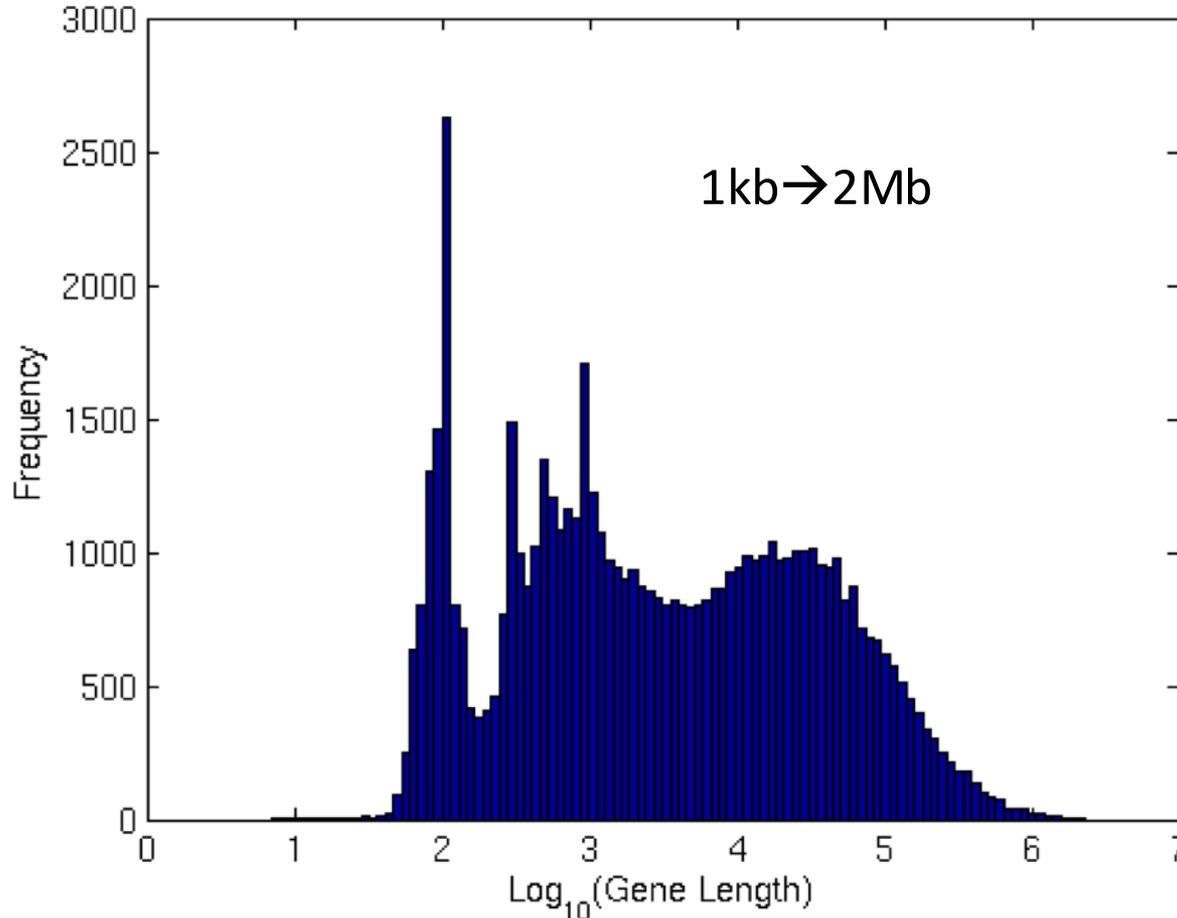
Morcelée en Exons (codants) et Introns (non codants)

2 Génome : Structure d'un gène eucaryote codant



**1 seul brin est transcrit en ARN
= brin sens OU antisens**

2 Génome : Structure d'un gène eucaryote codant

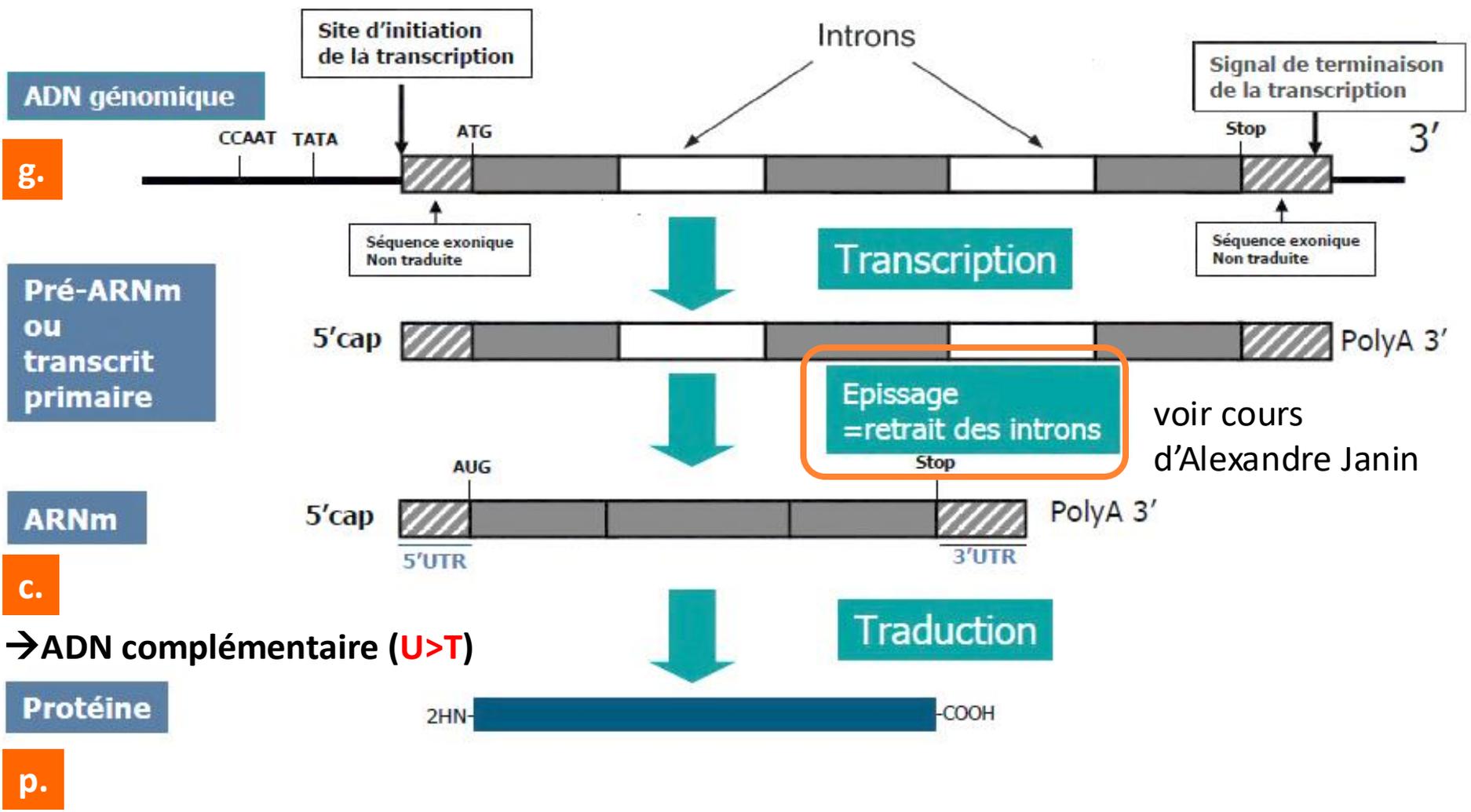


1 des + longs: *DMD*
codant pour la
dystrophine, 2,2Mb

Le + d'exons: *TTN*
codant pour la **titine**,
350 exons

2 Structure d'un gène eucaryote

Important



→ ADN complémentaire (U>T)

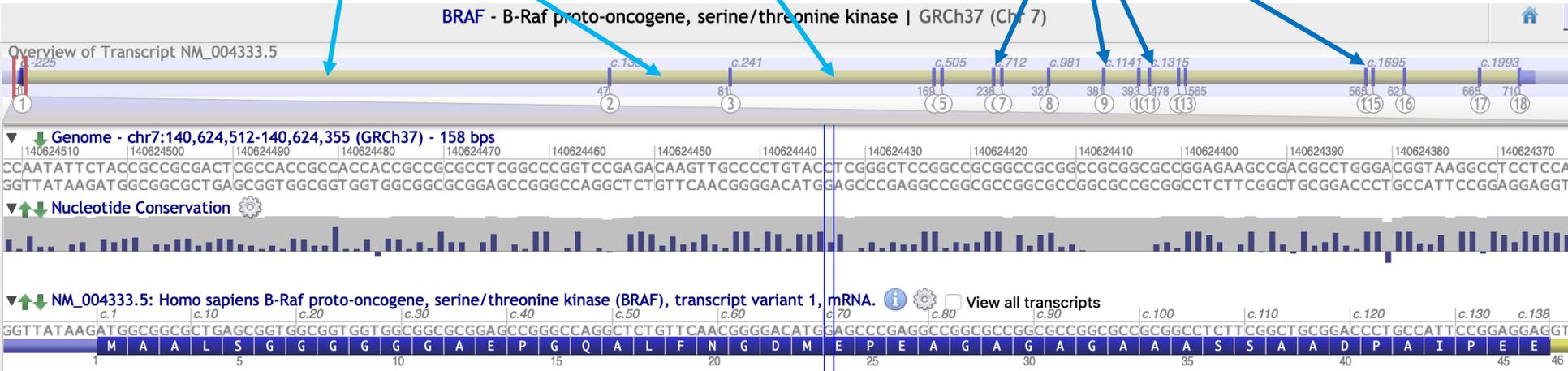
voir cours d'Alexandre Janin

2 Structure d'un gène eucaryote



Introns

Exons



Gène →

ADNc →

Protéine →

Genome - chr7:140,453,196-140,453,039 (GRCh37)

140453190 | 140453180 | 140453170 | 140453160

5TCTATATAAAGAAGTACTTCTGGAGTGTCAATTTTATCC/

3CAGATATATTTCTTCATGAAGACCTCACAGTAAAAATAGGT

Nucleotide Conservation

NM_004333.5: Homo sapiens B-Raf proto-oncogene,

c.1742 | c.1750 | c.1760 | c.1770

3CAGATATATTTCTTCATGAAGACCTCACAGTAAAAATAGGT

581 | 585 | 590

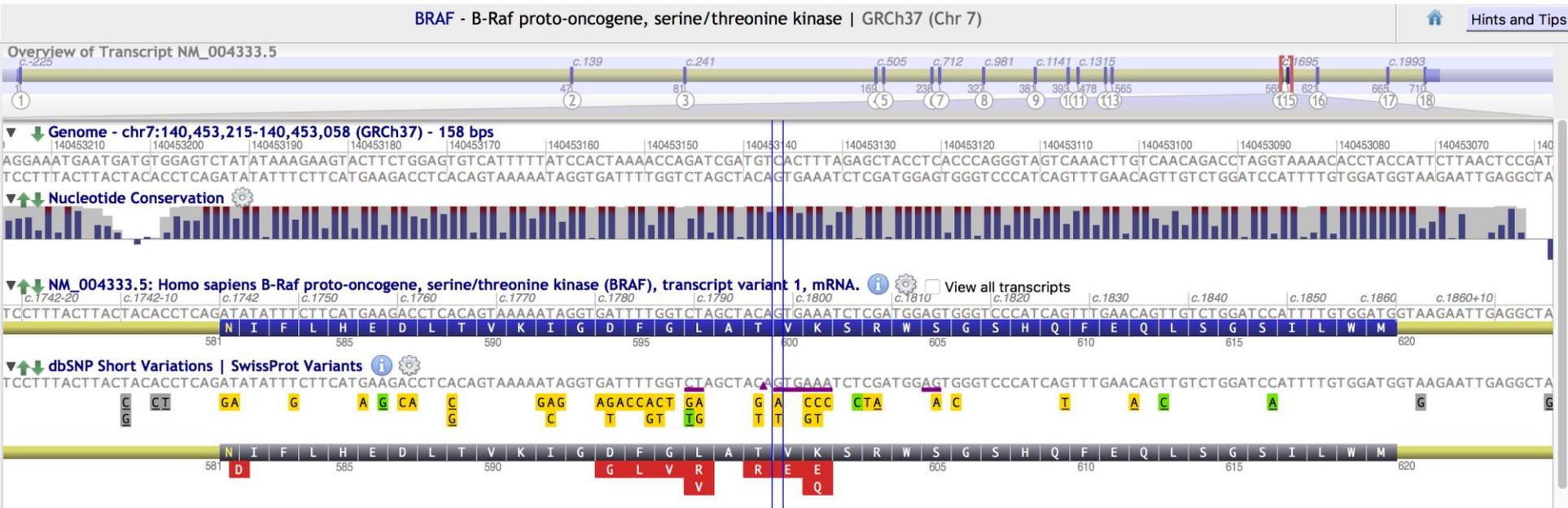
N I F L H E D L T V K I G

Position chromosomique

Position sur le transcrit

Position sur la protéine

2 Structure d'un gène eucaryote



Chr7:g.1404531136T>A : substitution d'1 nucléotide sur le gène
(nomenclature g. rarement utilisée en routine)



c.1799T>A : modification d'1 nucléotide sur l'ADN complémentaire
(essentiel d'indiquer quel transcrit est utilisé comme référence)



p.600Val>Glu ou **p.600V>E** : "La valine en position 600 de la protéine BRAF est remplacée par un glutamate"

2 Structure d'un gène eucaryote

Gène

Génome de référence

Version du transcrit

Transcrits alternatifs

Select a transcript of **EGFR**

Build	Transcript	Exons	AAs	Description
	NM_201284.1	16	705	This variant (4) uses a different 3' terminal exon when ...
	NM_201282.1	16	628	This variant (2) uses a different 3' terminal exon when ...
	NM_201283.1	10	405	This variant (3, also known as ErbB1-S) uses a different...
GRCh38 (hg...	NM_005228.3	28	1210	This variant (1) encodes the longest isoform (a).
	NM_005228.4	28	1210	This variant (1) encodes the longest isoform (a).
	NM_001346899.1	27	1165	This variant (7) lacks an in-frame exon in the 5' coding ...
	NM_001346900.1	28	1157	This variant (8) uses a novel 5' terminal exon compared...
	NM_001346898.1	27	1136	This variant (6) has a 3' terminal exon that extends pas...
	NM_001346897.1	26	1091	This variant (5) lacks an in-frame exon in the 5' coding ...
	NM_001346941.1	22	943	This variant (EGFRvIII, also known as delta-EGFR and d...
	NM_201284.1	16	705	This variant (4) uses a different 3' terminal exon when ...
	NM_201282.1	16	628	This variant (2) uses a different 3' terminal exon when ...
	NM_201283.1	10	405	This variant (3, also known as ErbB1-S) uses a different...
LRG_304	t1 (NM_005228....	28	1210	Epidermal growth factor receptor, transcript variant 1

Transcript Details

NM_005228.4 (on GRCh38) - 28 exons - 1210 AAs Preferred Transcript

This variant (1) encodes the longest isoform (a).



Cancel OK

2 Les séquences du concours...

IMPORTANT : L'**acide aminé** correspondant à chacun des codons est indiqué **sous la 1^{ère} base** de celui-ci. Pour faciliter le repérage, les numérotations du 1^{er} et du dernier nucléotide ou acide aminé sont indiquées pour chacune des lignes. Dans la séquence 2, les exons sont en majuscules et les bordures introniques en minuscules. (...) indique que certaines bases ne sont pas indiquées.

Séquence 1

```

1   GTCCGATGGGGCTCTGGTGGCGTGATCTGCGCGCCCCAGGGCTCAAGCACCCACACCCTA 60
61  GAAGGTTTCCGCAGCGACGTCGAGGCGCTCATGGTTGCAGGCGGGCGCCGCCGTTTCAGTT 120
121 CA                               ; 180
181 CG ADNc/ARNm mature : pas d'intron ; 240
241 CT                               ; 300
301 GCGTGGACCGCGCACGGCCTCTAGGTCTCCTCGCCAGGACAGCAACCTCTCCCCTGGCCC 360
361 TCATGGGCACCGTCAGCTCCAGGCGGTCTGGTGGCCGCTGCCACTGCTGCTGCTGCTGC 420
1   M G T V S S R R S W W P L P L L L L L L L 20
421 TGCTGCTCCTGGGTCCC CGGGCGCCCGTGC GCAGGAGGACGAGGACGGCGACTACGAGG 480
21   L L L G P A G A R A Q E D E D G D Y E E 40
481 AGCTGGTGCTAGCCTTGC GTTCCGAGGAGGACGGCCTGGCCGAAGCACCCGAGCACGGAA 540
41   L V L A L R S E E D G L A E A P E H G T 60
541 CCACAGCCACCTTCCACCGCTGCGCCAAGGATCCGTGGAGGTTGCCTGGCACCTACGTGG 600
61   T A T F H R C A K D P W R L P G T Y V V 80
  
```

Séquence 2

```

GTCCGATGGGGCTCTGGTGGCGTGATCTGCGCGCCCCAGGGCTCAAGCACCCACACCCTA
GAAGGTTTCCGCAGCGACGTCGAGGCGCTCATGGTTGCAGGCGGGCGCCGCCGTTTCAGTT
                                     ;AGACTGGCTCGGGCGGGCCGGGACG
                                     ;GATTCCGCGCGCCCCCTTCACGCGCC
                                     ;CACCGCAAGGCTCAAGGCGCCGCCG
GCGTGGACCGCGCACGGCCTCTAGGTCTCCTCGCCAGGACAGCAACCTCTCCCCTGGCCC
TCATGGGCACCGTCAGCTCCAGGCGGTCTGGTGGCCGCTGCCACTGCTGCTGCTGCTGC
1   M G T V S S R R S W W P L P L L L L L L L 20
TGCTGCTCCTGGGTCCC CGGGCGCCCGTGC GCAGGAGGACGAGGACGGCGACTACGAGG
21   L L L G P A G A R A Q E D E D G D Y E E 40
AGCTGGTGCTAGCCTTGC GTTCCGAGGAGGACGGCCTGGCCGAAGCACCCGAGCACGGAA
41   L V L A L R S E E D G L A E A P E H G T 60
CCACAGCCACCTTCCACCGCTGCGCCAAGgtgcggggtgtagggatgggagggccggggcga
61   T A T F H R C A K                               69
(...) tcatgttccctccttgc atggggccagGATCCGTGGAGGTTGCCTGGCACCTACGTGG
70   D P W R L P G T Y V V 80
  
```

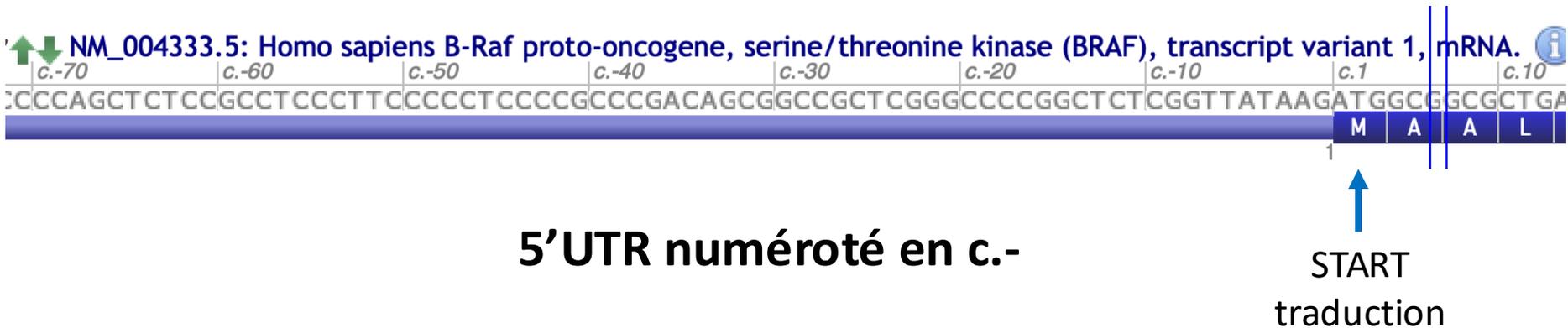
ADN/gène : présence d'introns



2 Structure d'un gène eucaryote

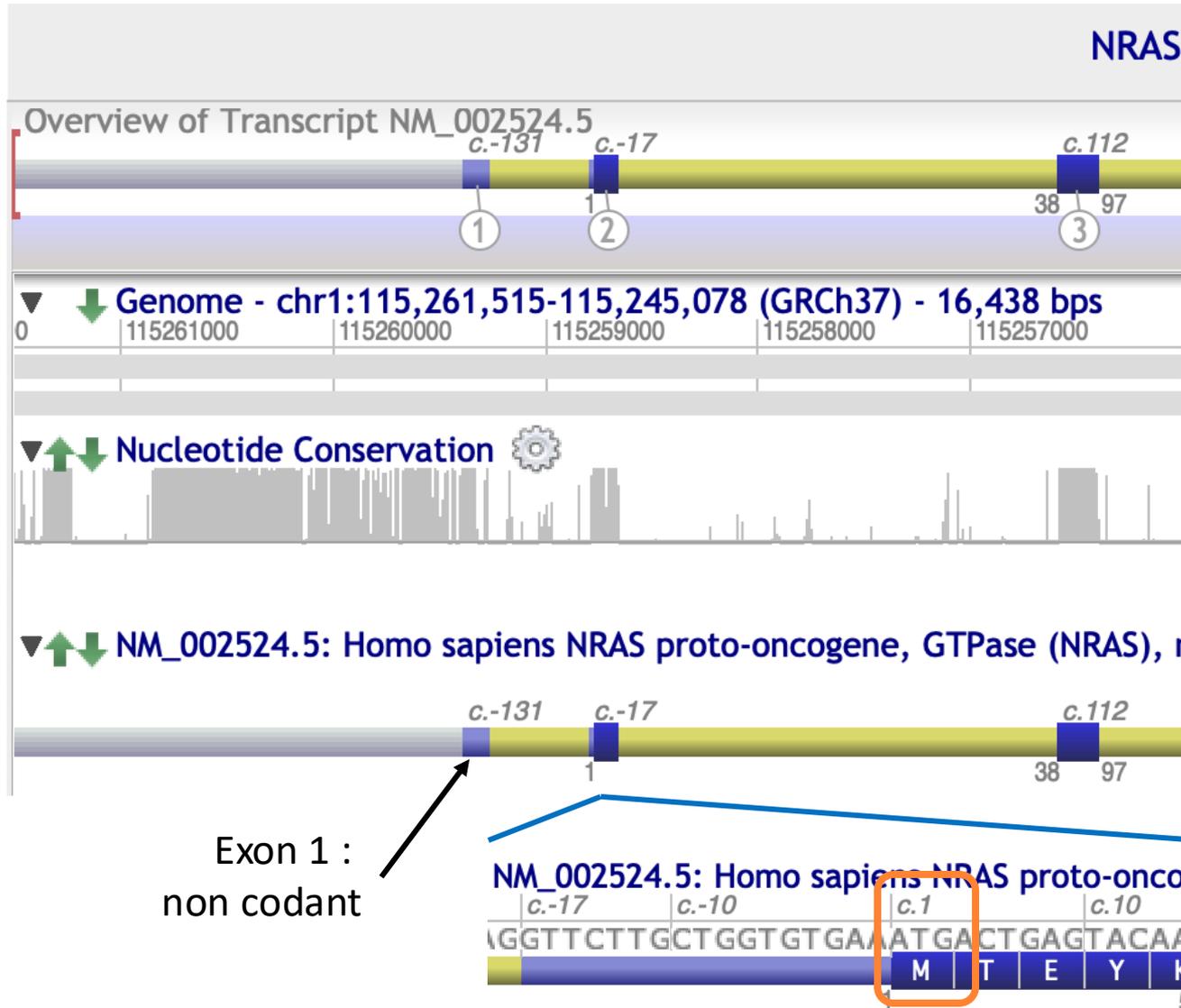


Attention! la numérotation de l'ADNc (c.) commence à l'ATG
= ne comprend pas le 5'UTR...



➔ Ajouter la longueur du 5'UTR à la numérotation c. du variant pour retrouver sa position

2 Structure d'un gène eucaryote



Attention!
Le codon **START (ATG)** n'est pas toujours localisé dans le premier exon

2 Les séquences du concours...

IMPORTANT : L'**acide aminé** correspondant à chacun des codons est indiqué **sous la 1^{ère} base** de celui-ci. Pour faciliter le repérage, les numérotations du 1^{er} et du dernier nucléotide ou acide aminé sont indiquées pour chacune des lignes. Dans la séquence 2, les exons sont en majuscules et les bordures introniques en minuscules. (...) indique que certaines bases ne sont pas indiquées.

Séquence 1

```

1   GTCCGATGGGGCTCTGGTGGCGTGATCTGCGCGCCCCAGGCGTCAAGCACCCACACCCTA 60
61  GAAGGTTT CGCCGCCGTTTCAGTT 120
121 CAGGGTCT Début de la traduction = c.1 GGGCGGGCCGGGACG 180
181 CGTCGTTG GCCCCTTCACGCGCC 240
241 CTGCTCCTGAACTTCAGTCTCTGCACAGTCTCCCCACCGCAAGGCTCAAGGCGCCGCCG 300
301 GCGTGGACCGCGCACGGCCTCTAGGTCTCCTCGCCAGGACAGCAACCTCTCCCCTGGCCC 360
361 TCATGGGCACCGTCAGTCCAGGCGGTCTGGTGGCCGCTGCCACTGCTGCTGCTGCTGC 420
1   M G T V S S R R S W W P L P L L L L L L L 20
421 TGCTGCTCCTGGGTCCCGCGGGCGCCCGTGCAGGAGGACGAGGACGGCGACTACGAGG 480
21  L L L G P A G A R A Q E D E D G D Y E E 40
481 AGCTGGTGCTAGCCTTGCCTCCGAGGAGGACGGCCTGGCCGAAGCACCCGAGCACGGAA 540
41  L V L A L R S E E D G L A E A P E H G T 60
541 CCACAGCCACCTTCCACCGCTGCGCCAAGGATCCGTGGAGGTTGCCTGGCACCTACGTGG 600
61  T A T F H R C A K D P W R L P G T Y V V 80
  
```



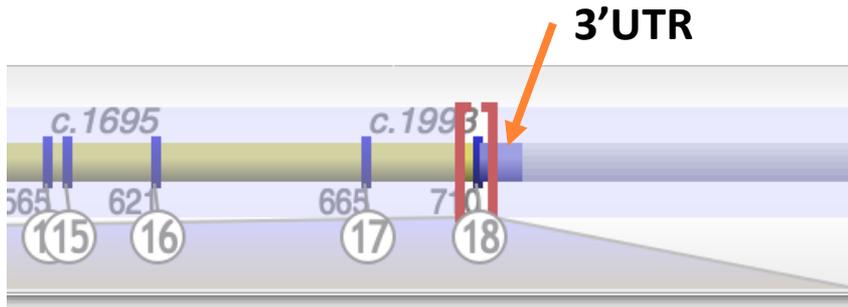
Séquence 2

```

GTCCGATGGGGCTCTGGTGGCGTGATCTGCGCGCCCCAGGCGTCAAGCACCCACACCCTA
GAAGGTTTCCGCAGCGACGTCGAGGCGTTCATGGTTGCAGGCGGGCGCCGCGTTTCAGTT
CAGGGTCTGAGCCTGGAGGAGTGAGCCAGGCAGTGAGACTGGCTCGGGCGGGCCGGGACG
CGTCGTTGCAGCAGCGGCTCCCAGCTCCCAGCCAGGATCCGCGCGCCCCCTTCACGCGCC
CTGCTCCTGAACTTCAGTCTCTGCACAGTCTCCCCACCGCAAGGCTCAAGGCGCCGCCG
GCGTGGACCGCGCACGGCCTCTAGGTCTCCTCGCCAGGACAGCAACCTCTCCCCTGGCCC
TCATGGGCACCGTCAGTCCAGGCGGTCTGGTGGCCGCTGCCACTGCTGCTGCTGCTGC
1   M G T V S S R R S W W P L P L L L L L L L 20
TGCTGCTCCTGGGTCCCGCGGGCGCCCGTGCAGGAGGACGAGGACGGCGACTACGAGG
21  L L L G P A G A R A Q E D E D G D Y E E 40
AGCTGGTGCTAGCCTTGCCTCCGAGGAGGACGGCCTGGCCGAAGCACCCGAGCACGGAA
41  L V L A L R S E E D G L A E A P E H G T 60
CCACAGCCACCTTCCACCGCTGCGCCAAGGATCCGTGGAGGTTGCCTGGCACCTACGTGG
61  T A T F H R C A K D P W R L P G T Y V V 69
(...) tcatgttcctccttgcattggggccagcATCCGTGGAGGTTGCCTGGCACCTACGTGG
70  D P W R L P G T Y V V 80
  
```

Introns en minuscule
(non traduits)

2 Structure d'un gène eucaryote



```
140434400 | 140434390 | 140434380
GGTGACTTTGTTTACTCACTCTCTCAAGT
CCACTGAAACAAATGAGTGAGAGAGTTC/
```



Codon STOP

Attention!

Le codon STOP (TAA, TAG ou TGA) est toujours localisé dans le dernier exon qui code aussi pour le 3'UTR (site polyA)

2 Structure d'un gène eucaryote



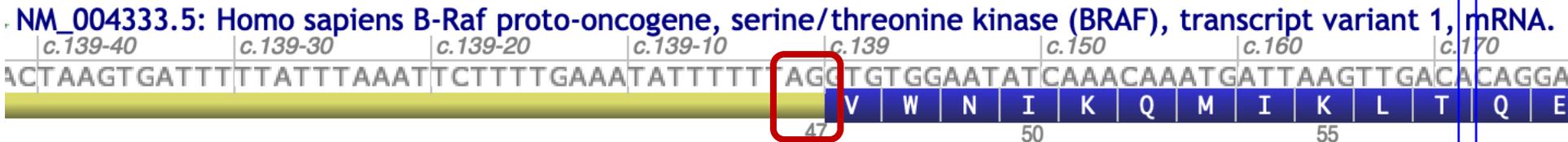
Fin de l'exon | Début de l'intron



c.138+1/+2

GT=site donneur d'épissage

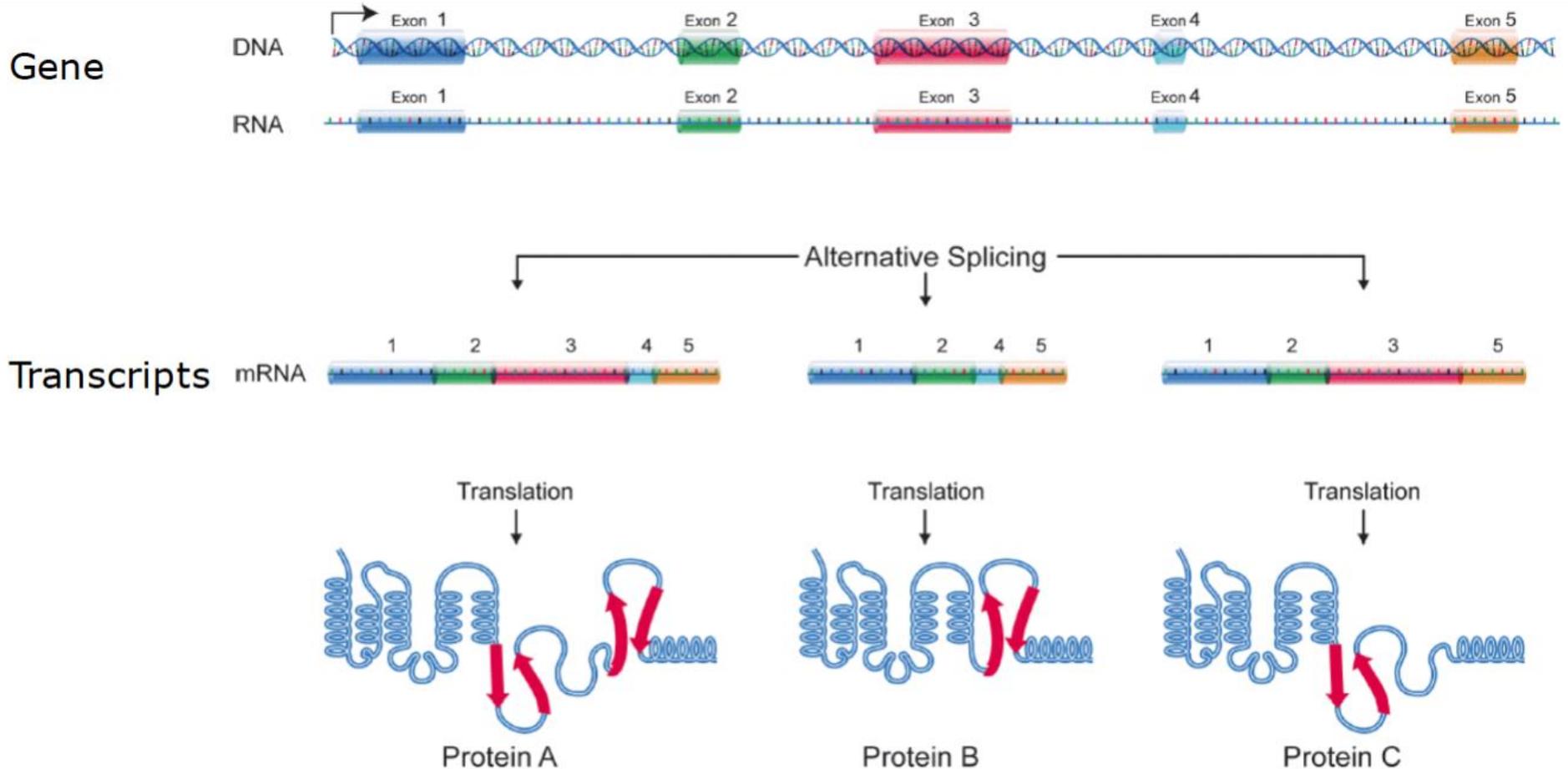
Fin de l'intron | Début de l'exon



c.139-2/-1

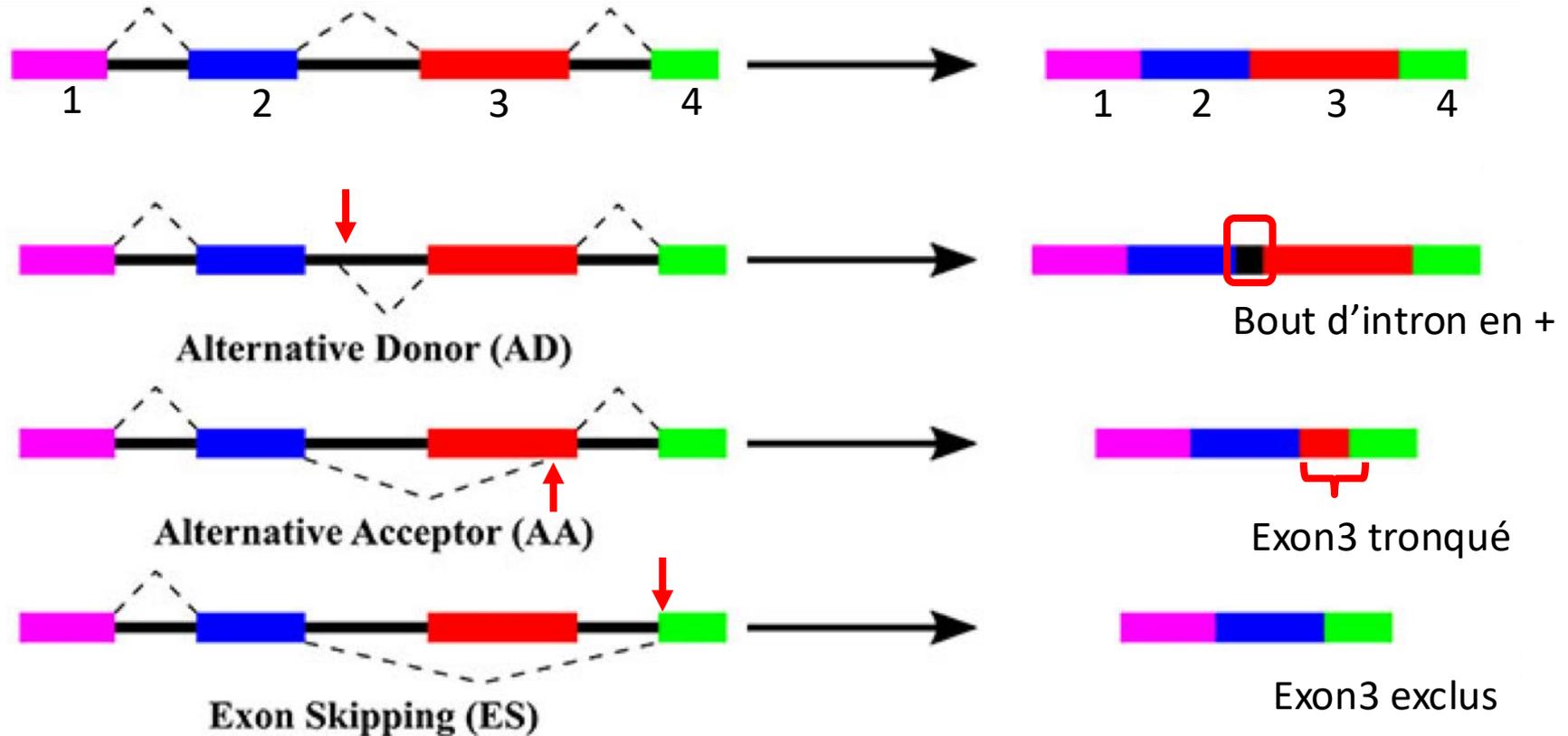
AG=site accepteur d'épissage

2 Epissage alternatif des pré-ARN



1 gène ⇒ plusieurs transcrits : Epissage alternatif +++

② Structure d'un gène eucaryote



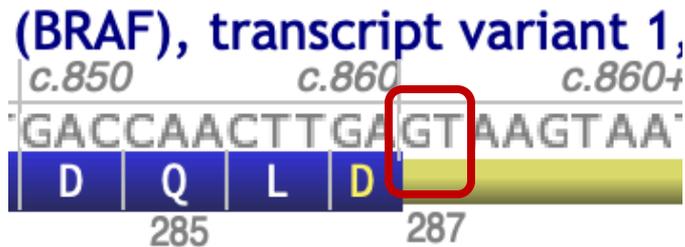
Physiologique ou causé par des mutations sur les sites D ou A

1 gène \Rightarrow plusieurs transcrits : Epissage alternatif +++

2 Structure d'un gène eucaryote

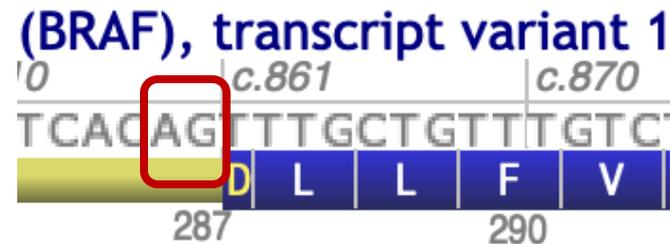


Attention! parfois un codon chevauche la jonction exon-exon



Fin **exon 6**

Site donneur



Site accepteur

Début **exon 7**

GA/.../T \Leftrightarrow GAT \Leftrightarrow D (Asp)

2 Structure d'un gène eucaryote



Attention! parfois un codon chevauche la jonction exon-exon si épissage alternatif \Rightarrow possible décalage du cadre de lecture...



Fin **exon 6**

Site donneur

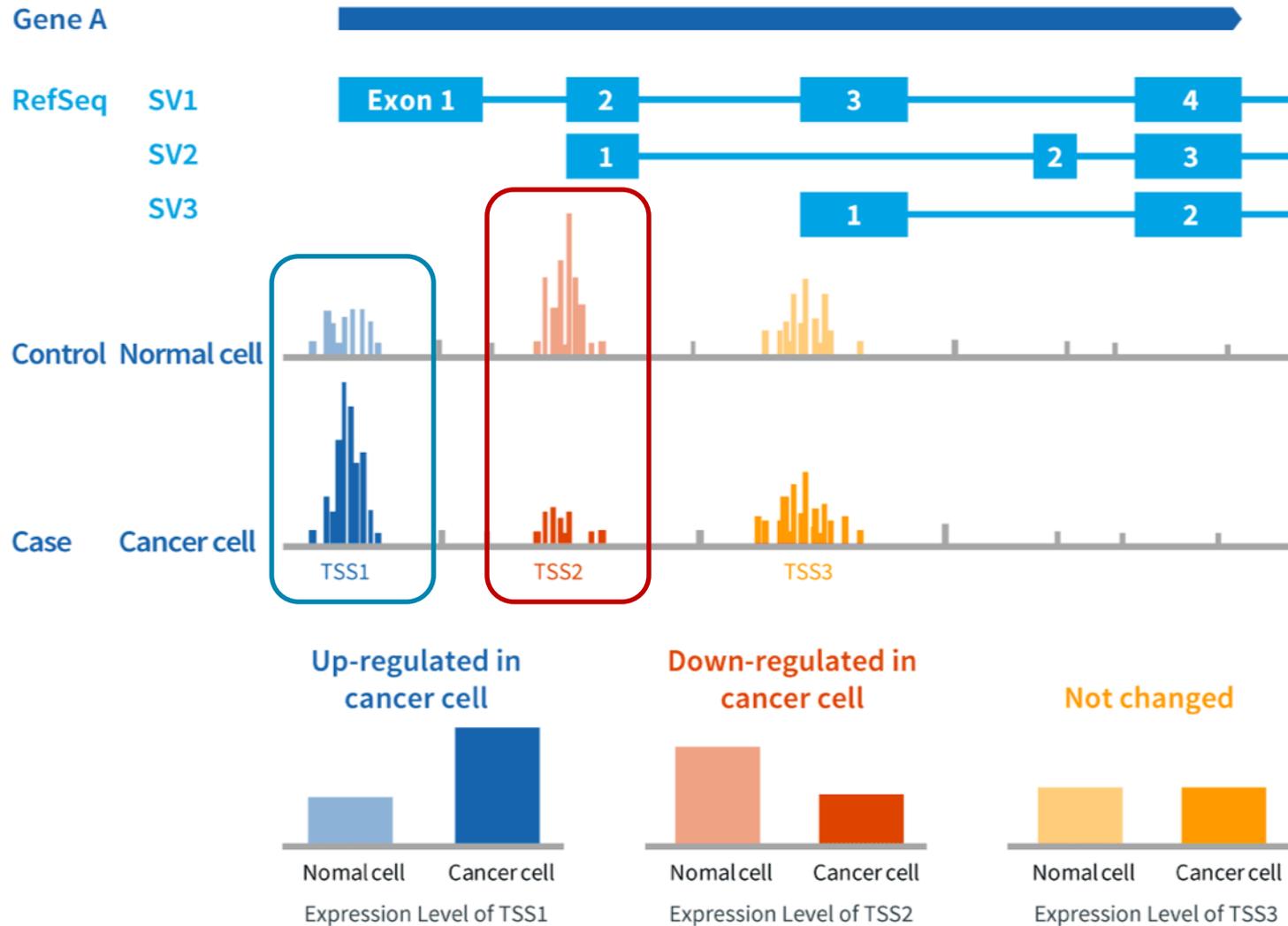


Site accepteur

Début **exon 9**

GA/.../G \Rightarrow GAG \Rightarrow E (Glu)
puis ACT (Thr)
puis TGA (**STOP**)

2 Structure d'un gène eucaryote

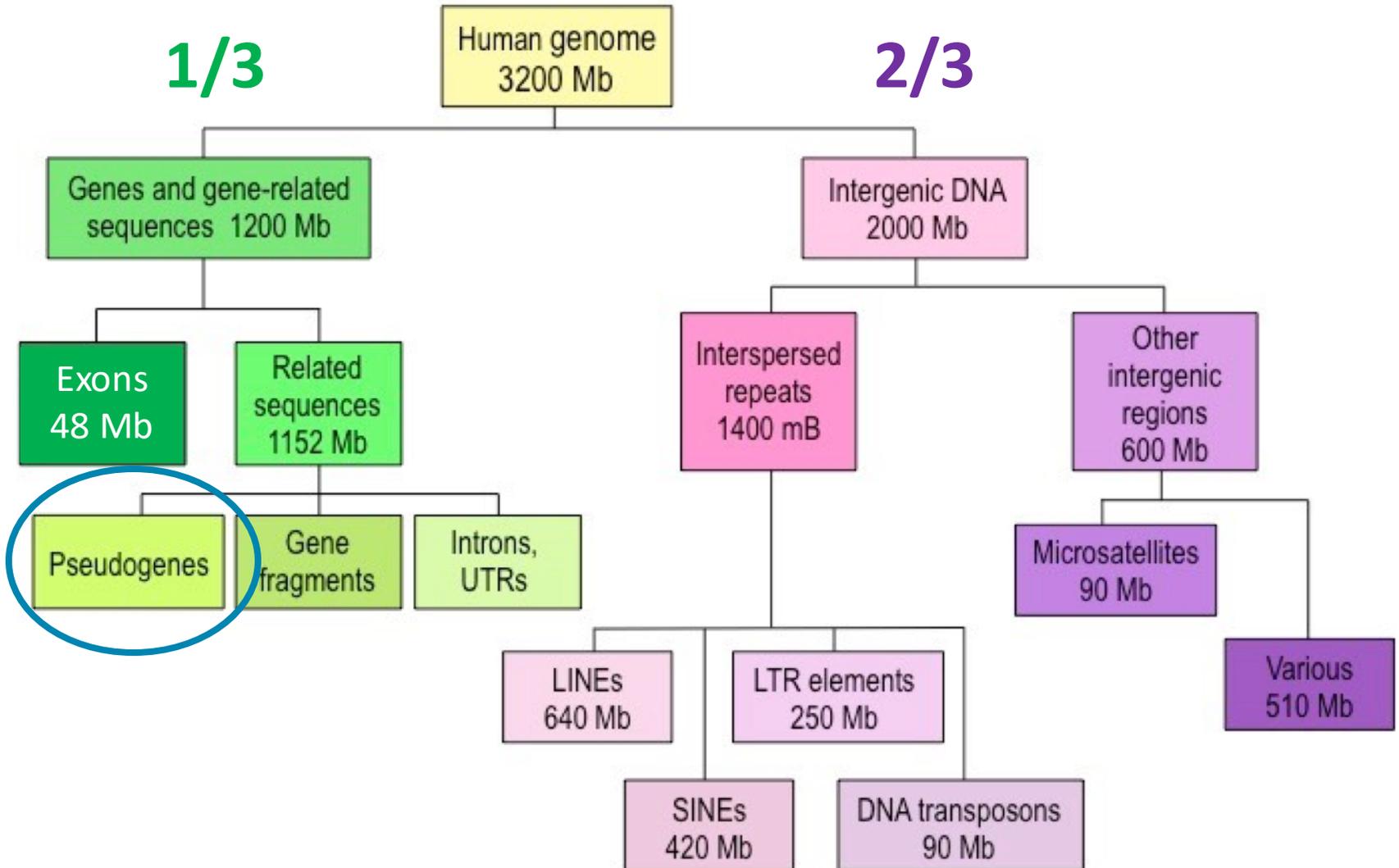


1 gène ⇨ plusieurs transcrits : Transcription Start Sites alternatifs

0 Plan

- Organisation du génome humain
- **Gènes**
 - Structure d'un gène eucaryote codant
 - **Pseudogènes**
- ADN intergénique

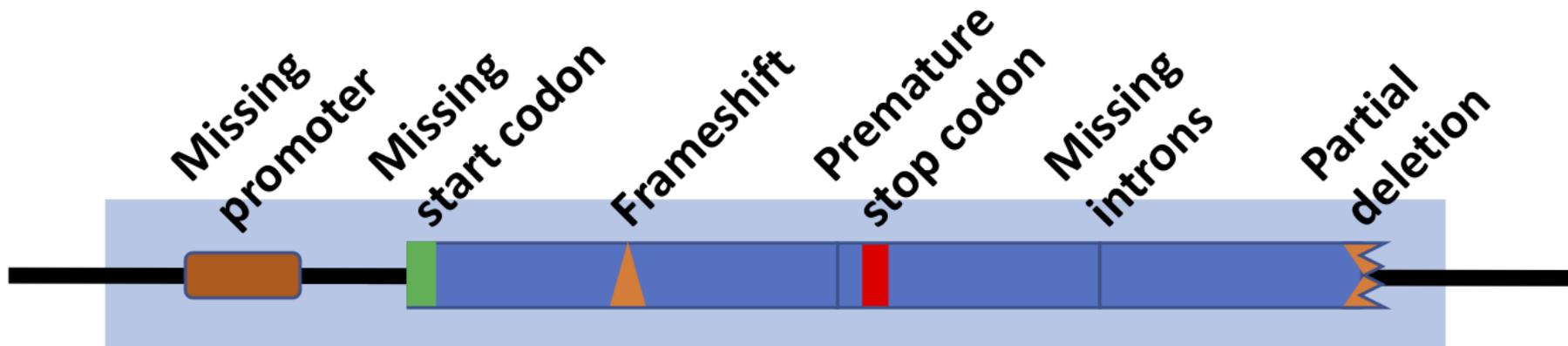
2 Génome : Composition du génome



2 Génome : Pseudogènes

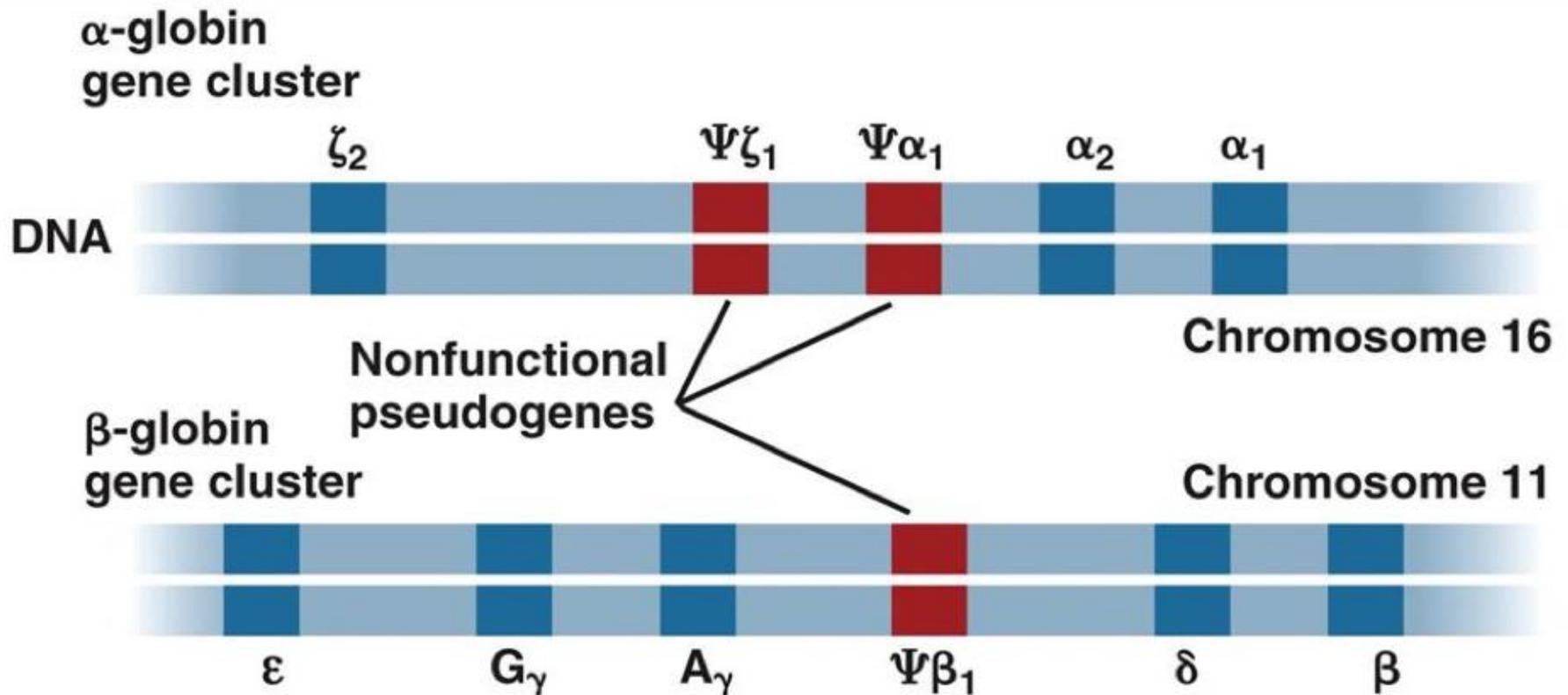
- Segments d'ADN présentant des **homologies de séquence avec des gènes actifs**
- **Perte de fonction par rapport au gène parental** ⇨ pas de protéine ou protéine non fonctionnelle

Common defects of pseudogenes:



2 Génome : Pseudogènes

- Ex: clusters des gènes codant pour les **globines**



PRINCIPLES OF LIFE, Figure 12.11
© 2012 Sinauer Associates, Inc.

Problème pour le séquençage et l'interprétation des variants...

Posez vos questions pendant le cours



1

Allez sur
wooclap.com

2

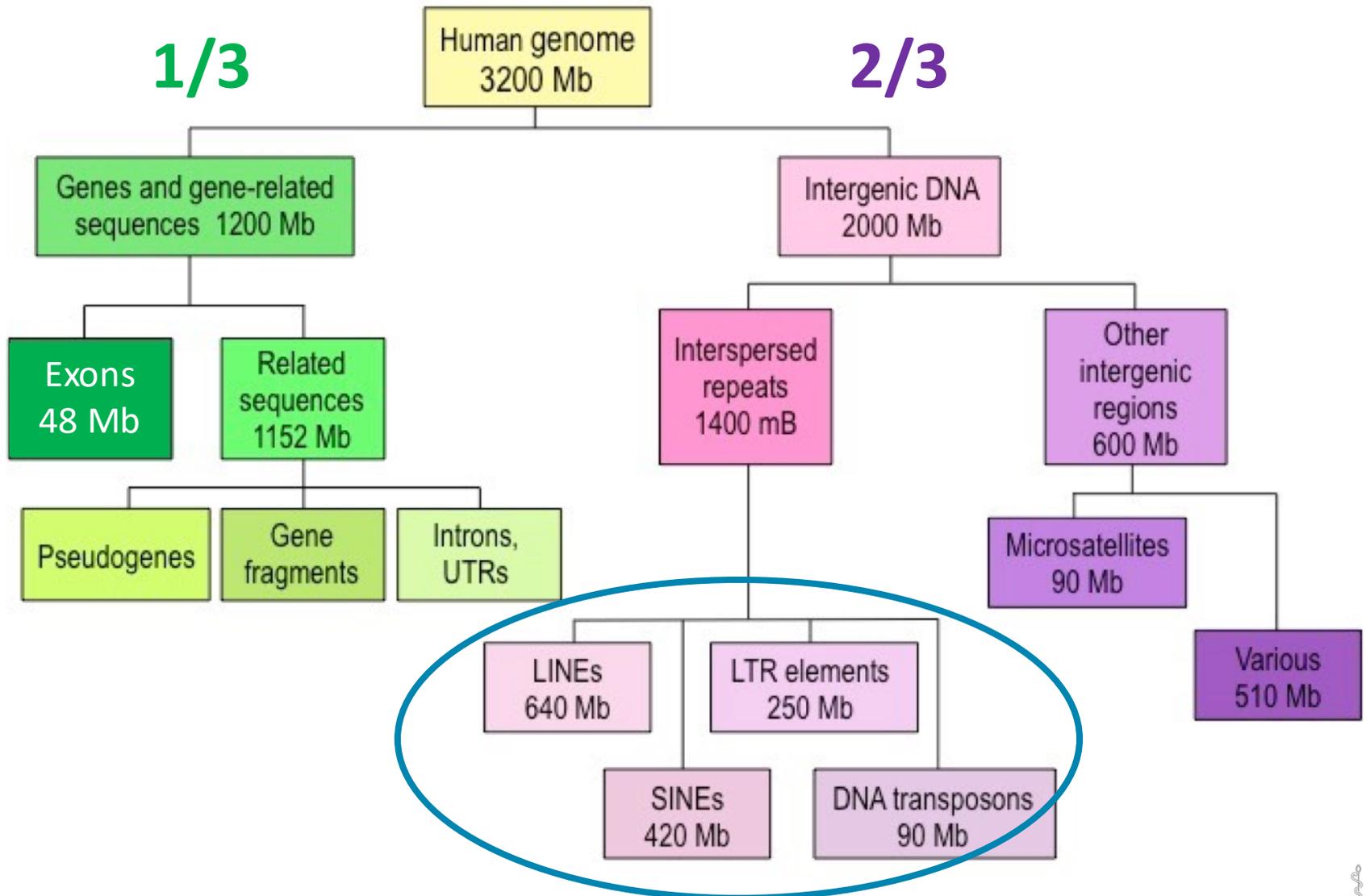
Entrez le code
d'événement dans
le bandeau
supérieur

Code
d'événement
PJINGO

0 Plan

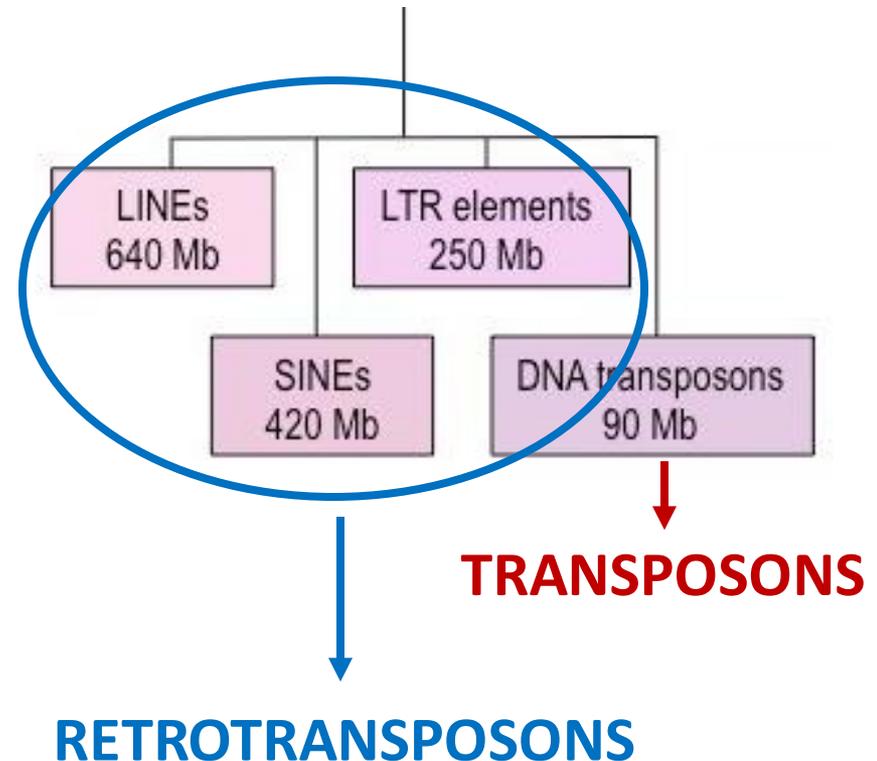
- Organisation du génome humain
- Gènes
- **ADN intergénique**
 - Séquences répétées
 - **Microsatellites**

3 Génome : Composition du génome



3 Repeats : Séquences répétées dispersées

- 45% du génome humain
- Dispersées dans tout le génome
- 0,1 → 6kb
- Moyennement répétitif
- Dérivent d'éléments "mobiles" ou "transposables" capables de se déplacer
- Autonomes ou non



« gènes sauteurs »

3 Repeats : Séquences répétées dispersées

SCIENCES • GÉNÉTIQUE

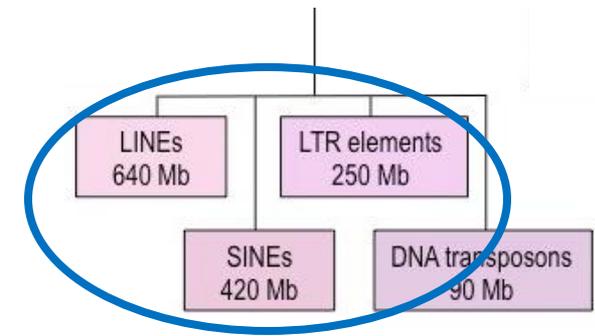
Ces virus qui ont colonisé notre génome : amis ou ennemis ?

Plus de 60 % de notre génome héberge de l'ADN issu d'entités virales qui ont, jadis, infecté de lointains ancêtres. Sans ce puissant moteur de l'évolution, nous ne serions pas là aujourd'hui à remonter le fil de cette improbable saga. Mais cet ADN trahit parfois ses hôtes, provoquant des maladies comme des cancers.

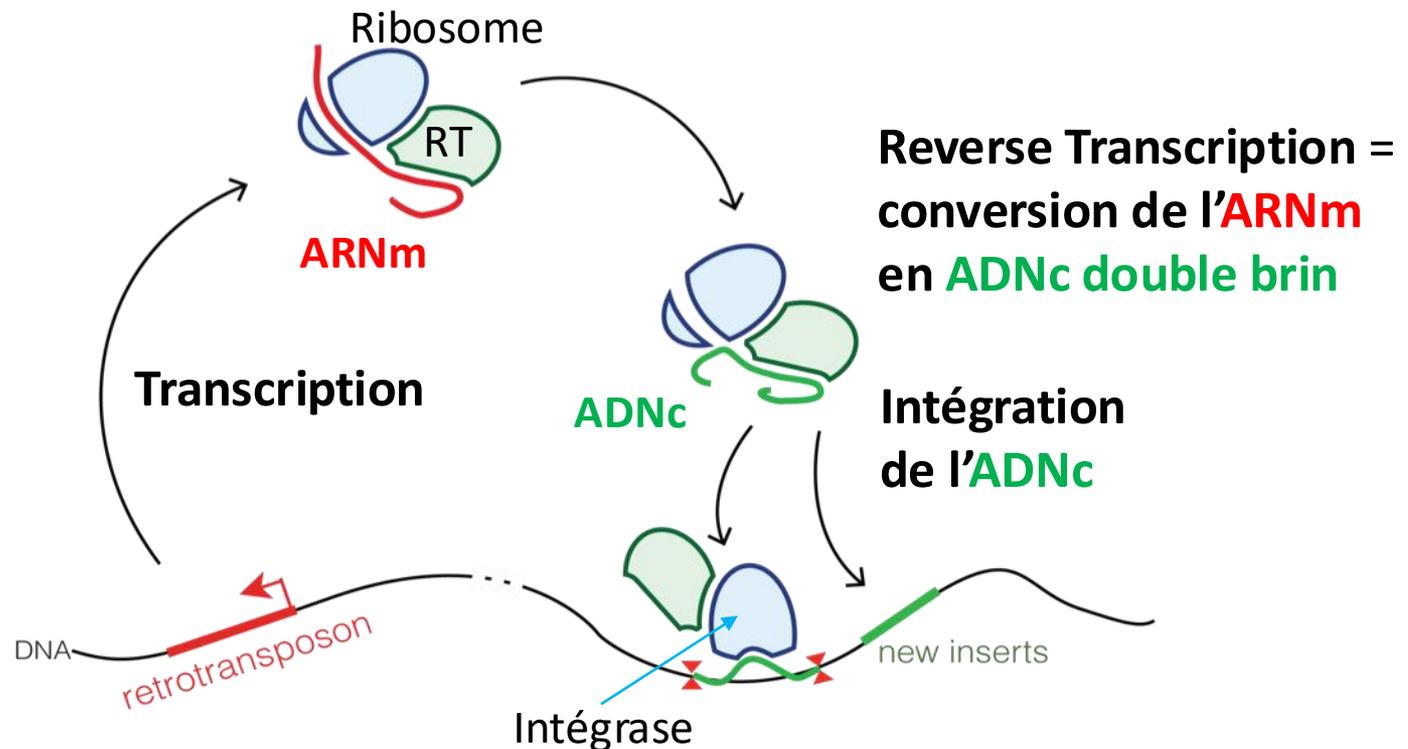
Par Florence Rosier

Publié le 26 février 2024 à 18h00, modifié le 22 avril 2024 à 12h41 • ⌚ Lecture 10 min.

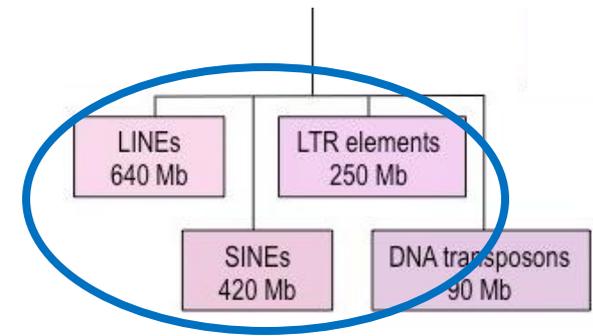
3 Repeats : **Rétrotransposons**



- **Intermédiaire ARNm, rétrotranscription en ADNc puis intégration**



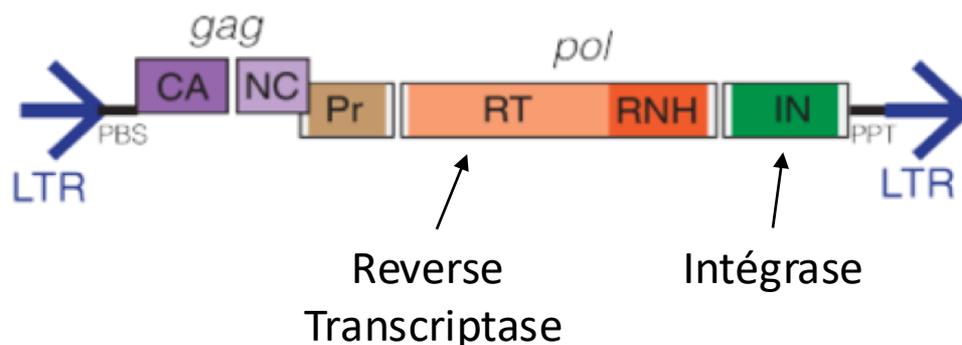
3 Repeats : **Rétrotransposons**



■ **LTR-rétrotransposons**

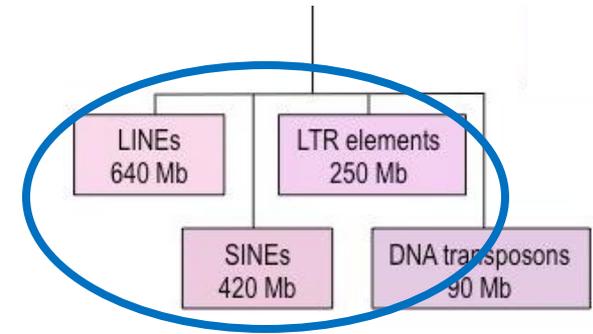
- contiennent des **Long Terminal Repeats**
- codent pour une **Reverse Transcriptase** et une **Intégrase** similaires à celles des rétrovirus.
- **Rétrovirus endogènes (HERV) : pas d'enveloppe (ENV) ⇒ non infectieux**

LTR-retrotransposon (5-20 kb)



LTR: Long terminal repeat. **gag** and **pol:** polyprotein-coding genes **CA** and **NC:** capsid and nucleocapsid protein. **Pr** and **IN:** Protease and Integrase protein **RT** and **RNH:** Reverse transcriptase and RNase H

3 Repeats : **Rétrotransposons**

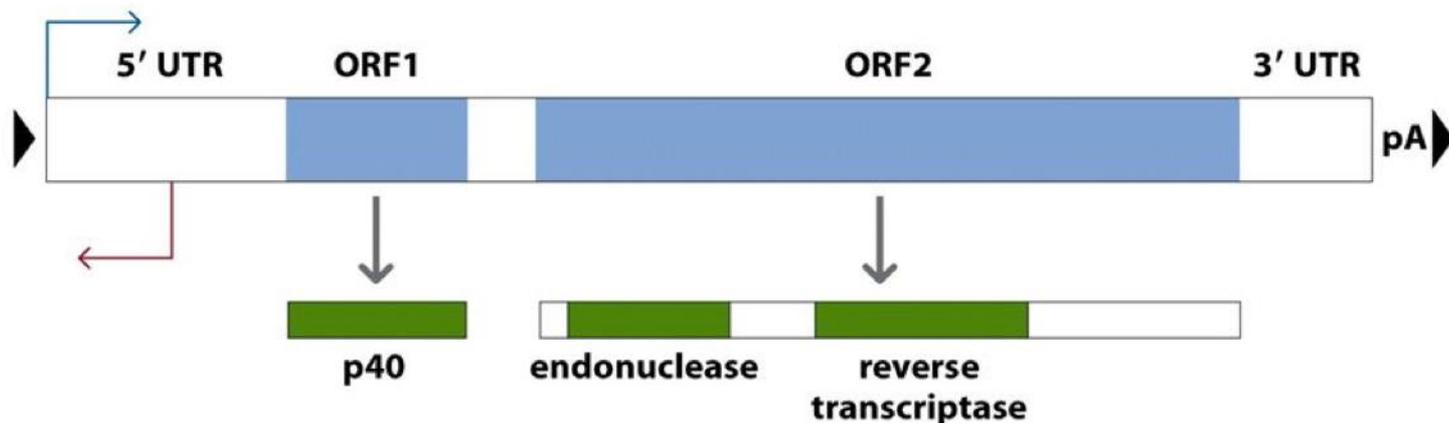


■ Rétroposons : pas de LTRs

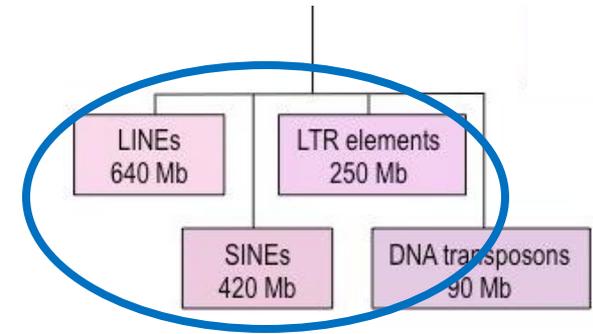
■ LINEs = Long Interspread Nuclear Elements

- $\approx 6\text{kb}$
- codent pour une RT (\rightarrow ADNc) et une endonuclease (\rightarrow intégration)
- transcrits par RNAPol II, 2 cadres de lectures non chevauchants

(A) LINE-1 repeat element



3 Repeats : **Rétrotransposons**

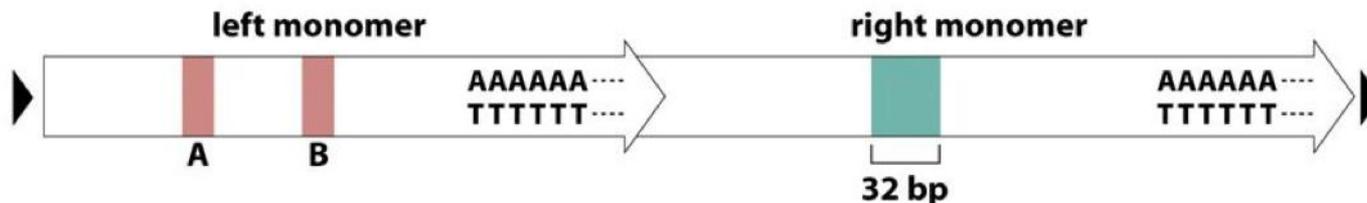


■ Rétroposons : pas de LTRs

■ SINEs = Short Interspersed Nuclear Elements

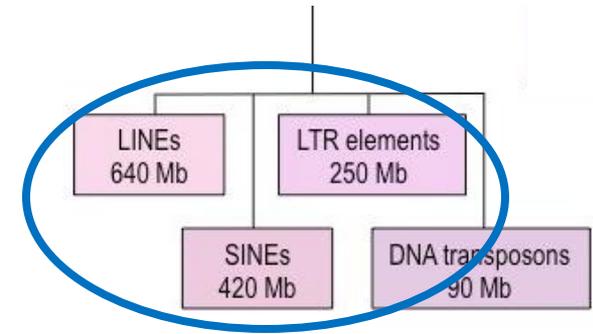
- 50-500bp
- ne codent PAS pour une RT ⇔ NON autonomes
- transcrits par RNAPol III
- ARNt, ARNr 5S, ARN 7SL (**Alu**, B1)
- Séquences Alu : 350bp, site *AluI*, 1.1 millions de copies

(B) Alu repeat element

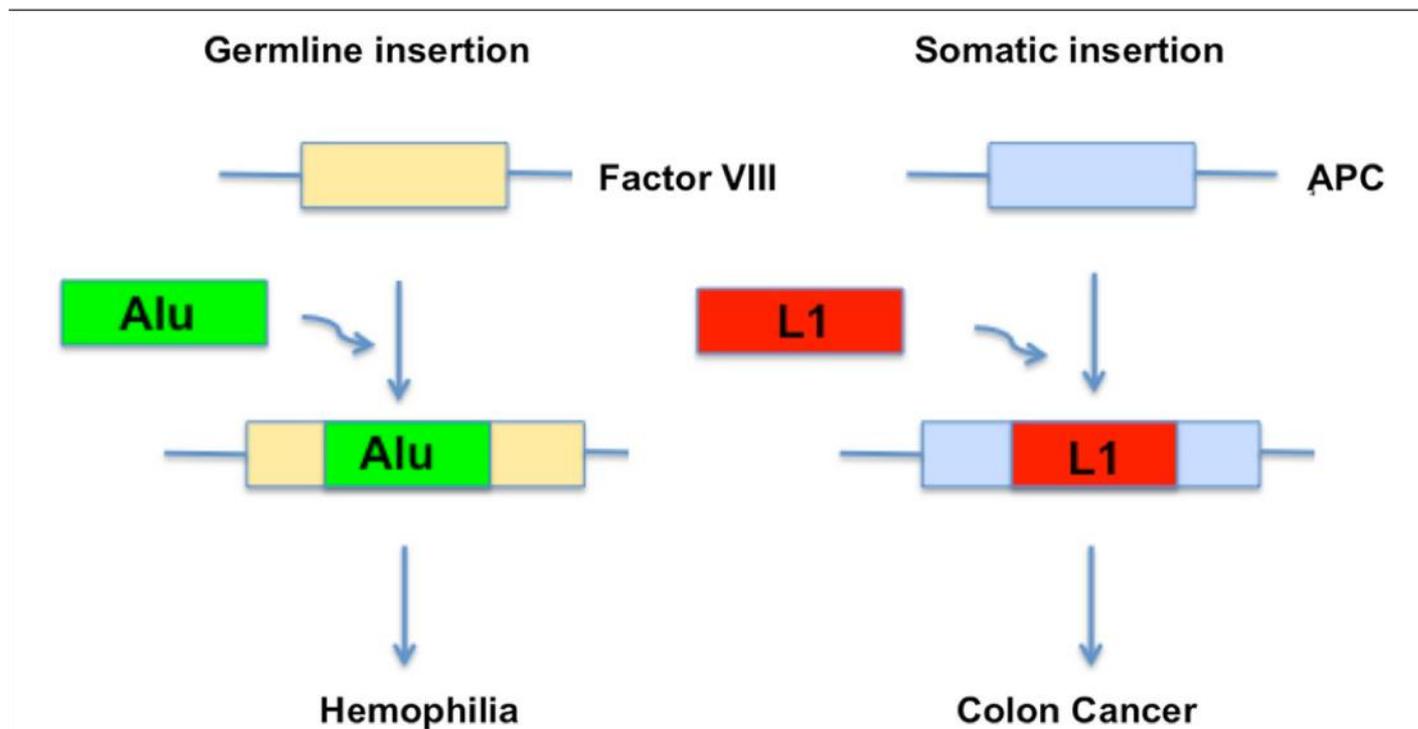


Promoteur RNAPol III

3 Repeats : **Rétrotransposons**

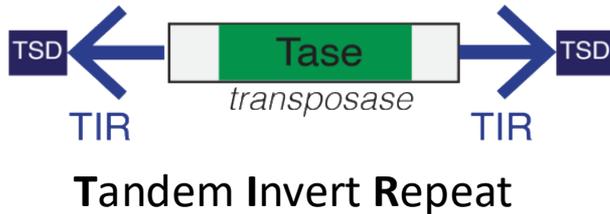


- L'insertion d'éléments mobiles dans les gènes à l'origine de maladies



3 Repeats : **Transposons**

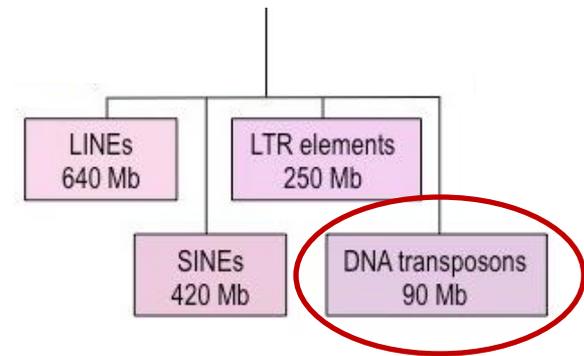
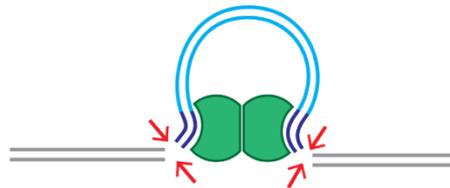
Transposon ADN code pour une **TRANSPOSASE**



1/Liaison des transposases



2/Clivage

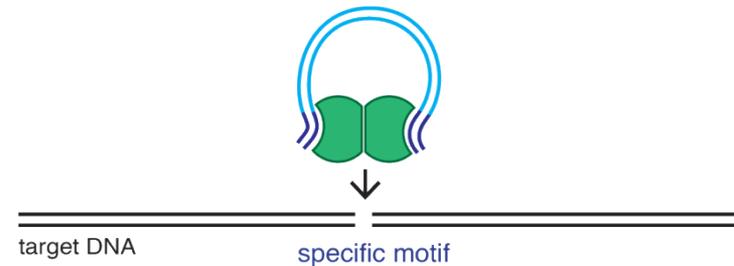


Aucun transposon encore actif chez l'Homme

(bactéries ou plantes +++)

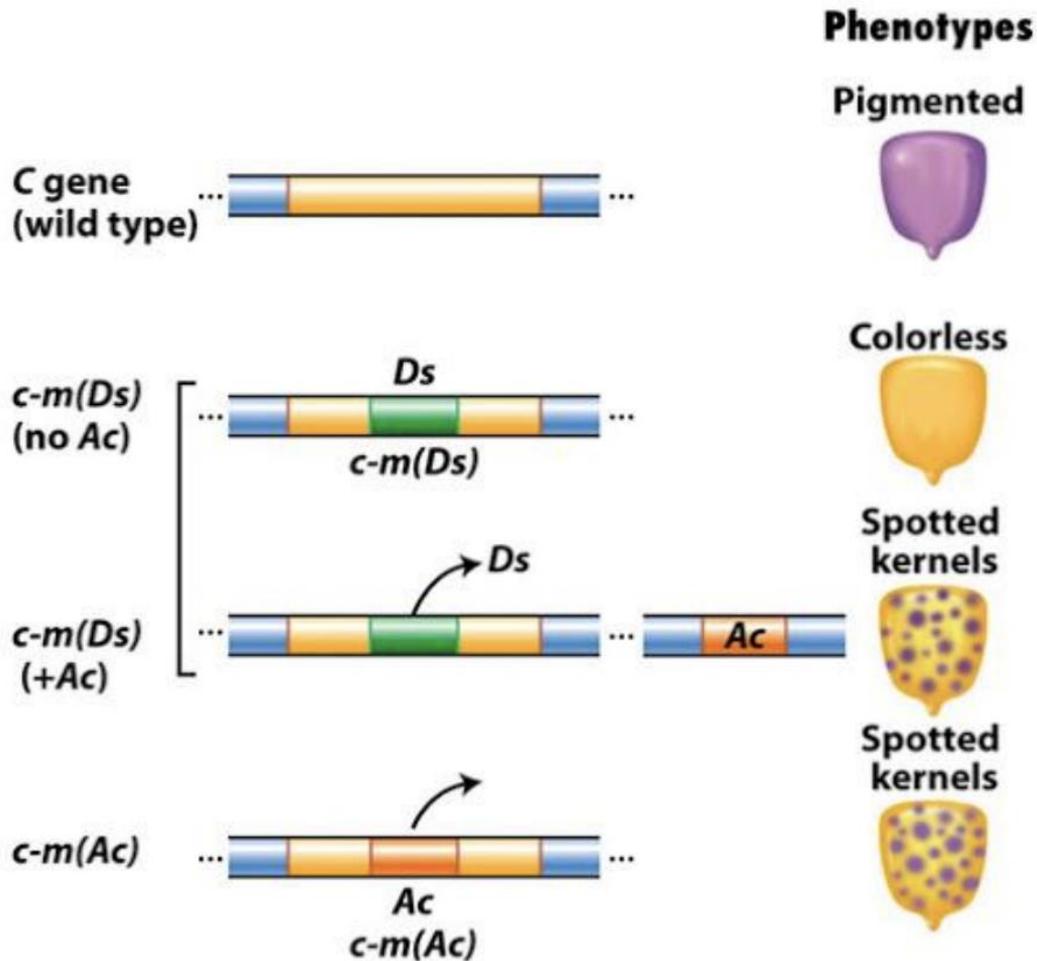
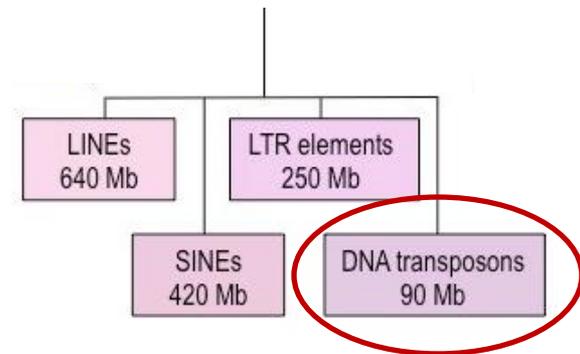
⇒ vestiges évolutifs

⇒ rôle ?



3/Transfert de la séquence capturée

3 Repeats : Transposons



Chez le Maïs

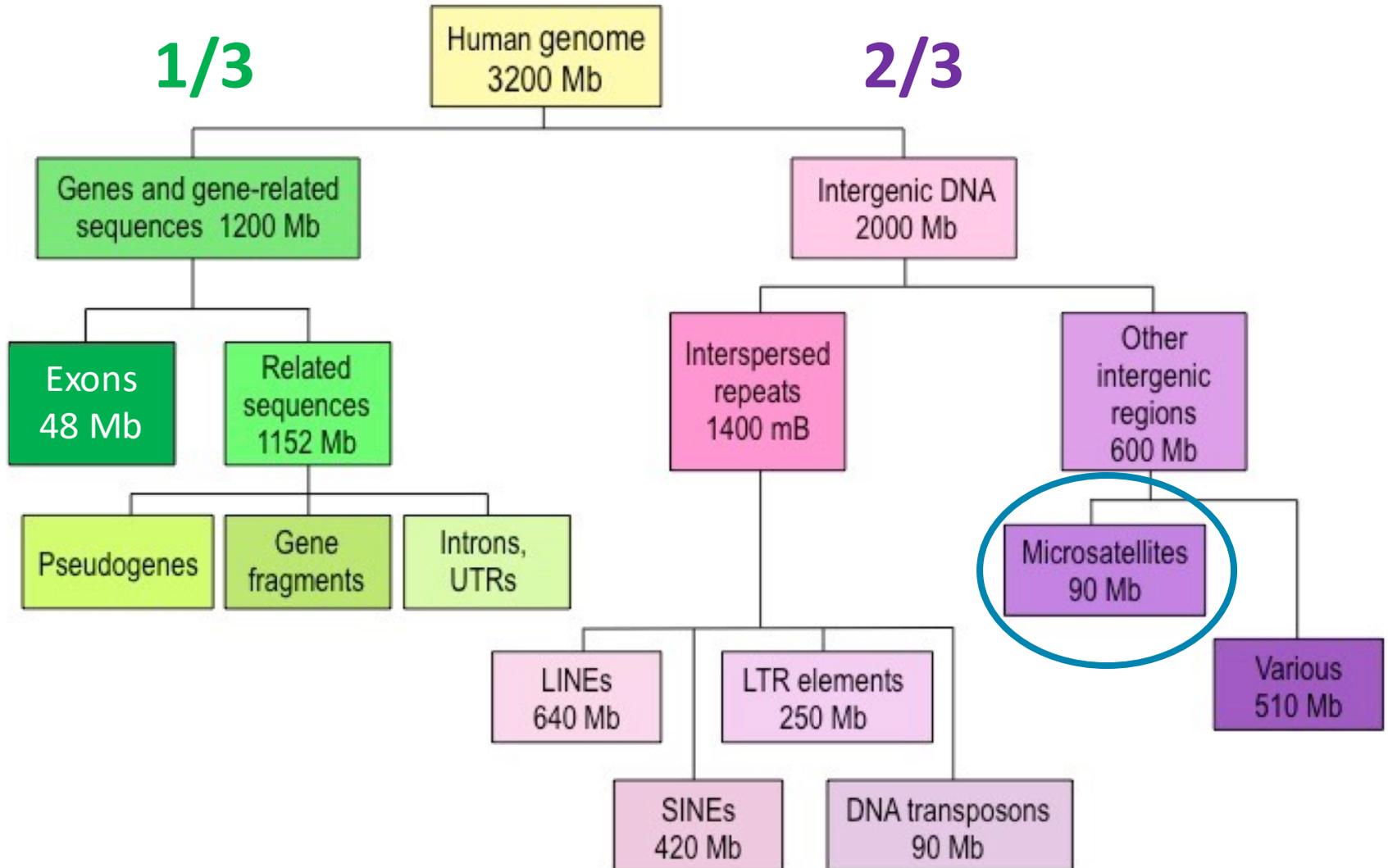
Figure 14-4
 Introduction to Genetic Analysis, Ninth Edition
 © 2008 W. H. Freeman and Company



0 Plan

- Organisation du génome humain
- Gènes
- **ADN intergénique**
 - Séquences répétées
 - **Microsatellites**

2 Génome : Composition du génome



3 Repeats : **Séquences répétées en tandem**

- **10% du génome**
- **Dispersées dans le génome mais localisées** (tjs aux mêmes endroits)
- **Qqs pb à qqs centaines de pb**
- **Hautement répétitif**

3 Repeats : Séquences répétées en tandem

■ Plusieurs types

- **Satellites** ⇨ au niveau des **centromères**

- **Minisatellites (Variable Number of Tandem**

Repeats ou VNTR) ⇨ 10-100pb répétées qqs

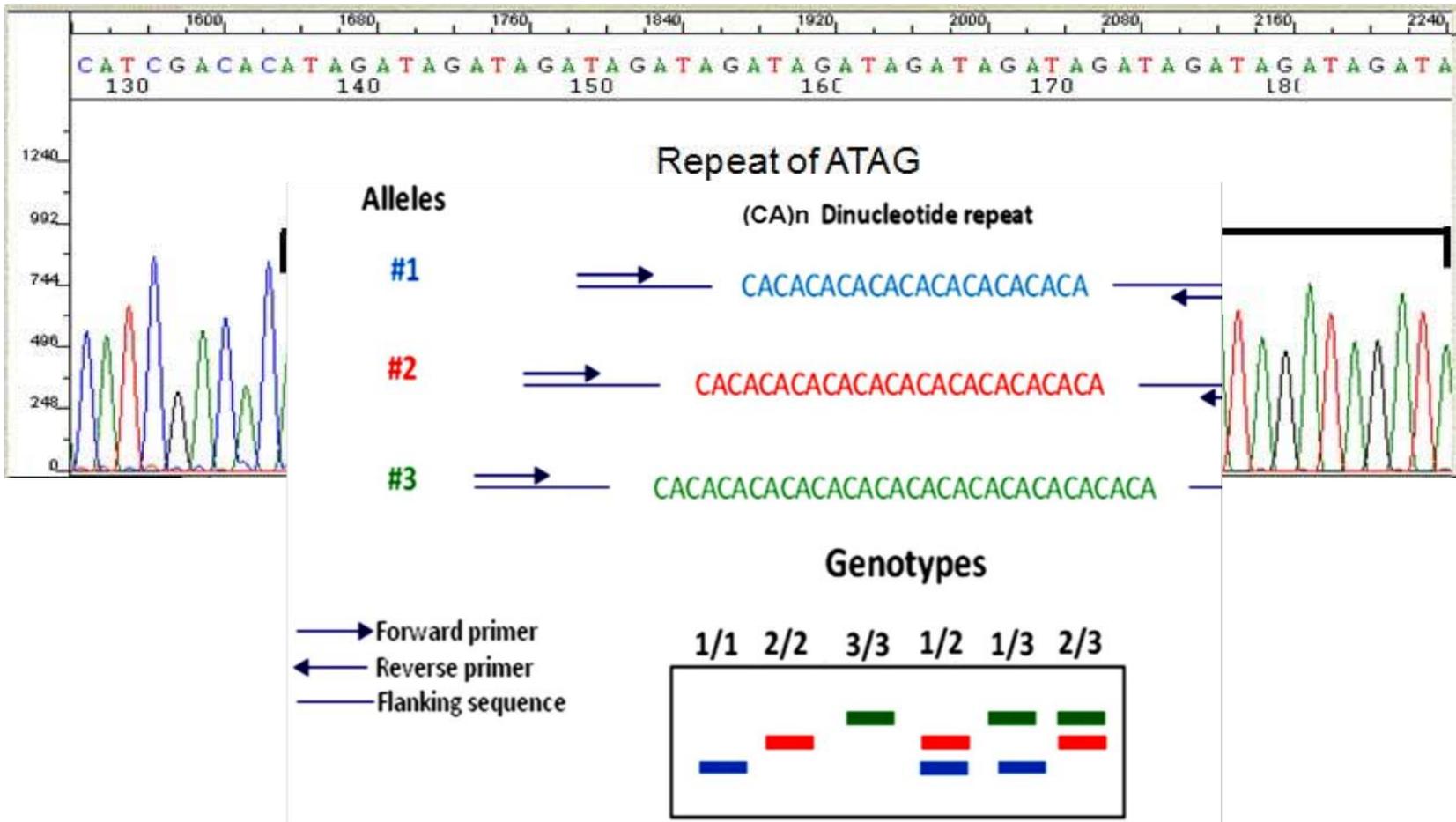
milliers de fois

Ex: motif **TTAGGG** répété dans les **téломères**

3 Repeats : Séquences répétées en tandem

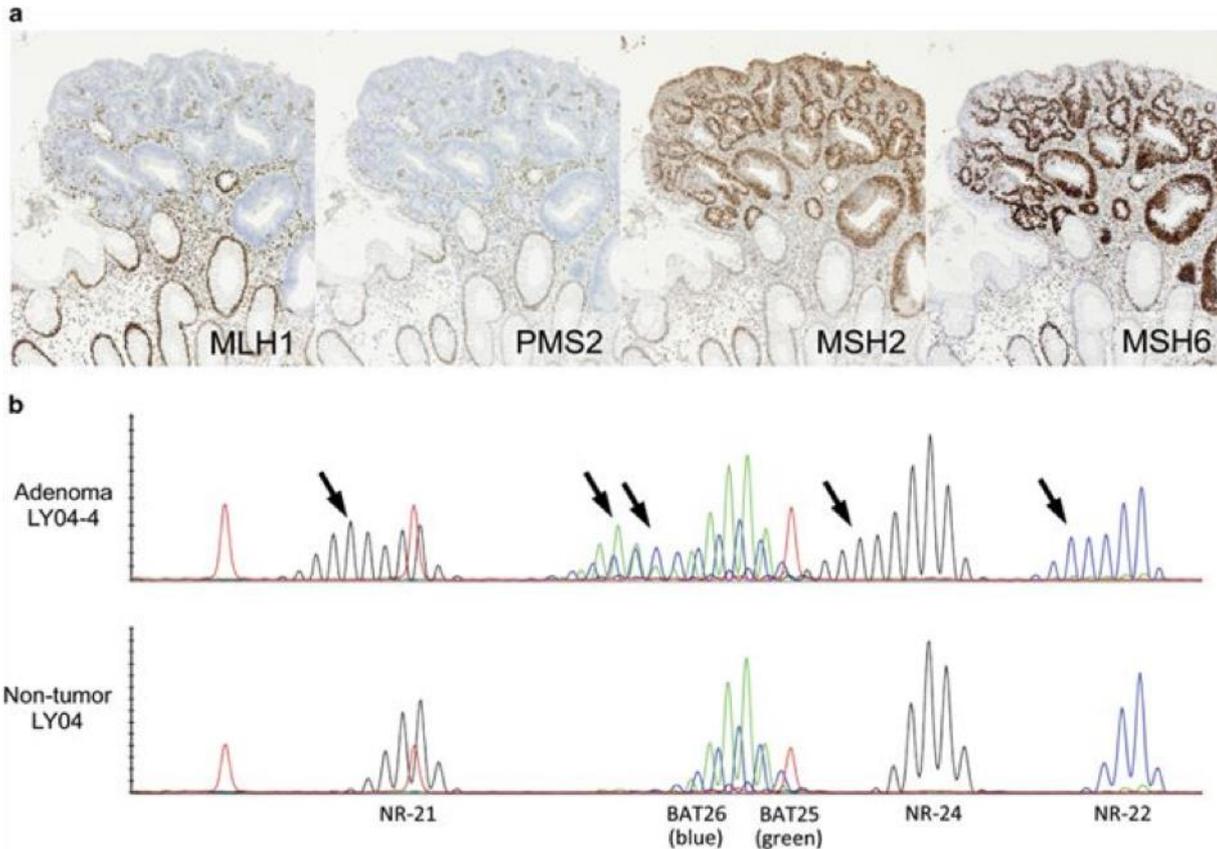
■ Microsatellites (Short Tandem Repeats ou STR)

⇒ 1 à 8pb répétées en tandem de 5 à 50 fois



3 Repeats : Séquences répétées en tandem

■ Microsatellites (Short Tandem Repeats ou STR)



Instabilité des microsatellites (MSI) dans certains cancers colorectaux par perte de protéines de réparation de l'ADN
⇒ **réponse à l'immunothérapie**

Posez vos questions pendant le cours



1

Allez sur
wooclap.com

2

Entrez le code
d'événement dans
le bandeau
supérieur

Code
d'événement
PJINGO

Génome : évaluez-vous !



1

Allez sur
wooclap.com

2

Entrez le code
d'événement dans
le bandeau
supérieur

Code
d'événement
PJINGO