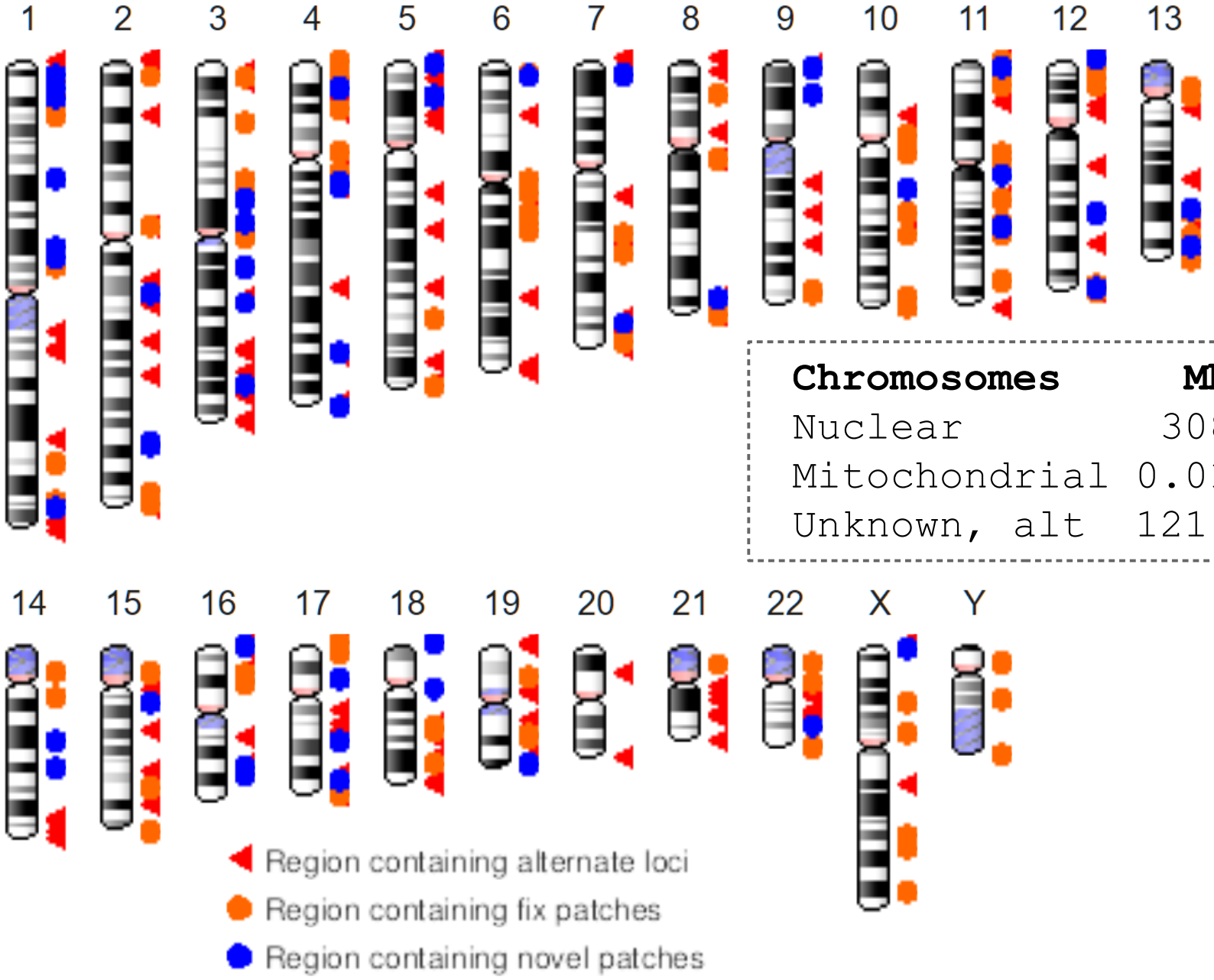
MUTATIONS IN SPACE:

GENES AND CONSEQUENCES

Lecture plan

- Human genome contents. Mitochondrial genome.
- Overview of human genes structure and processing
- Alternative splicing
- Epigenetics. Chromosomal imprinting.
- Variant annotation. ENSEMBL Variant Effect Predictor: impact and consequences
- Protein-truncating and loss-of-function variants
- Missense variants, inframe indels
- Synonymous and regulatory variants
- Variant effect, dominant and recessive variants, gain- and loss-of-function

Genome Reference Consortium GRCh38.p13



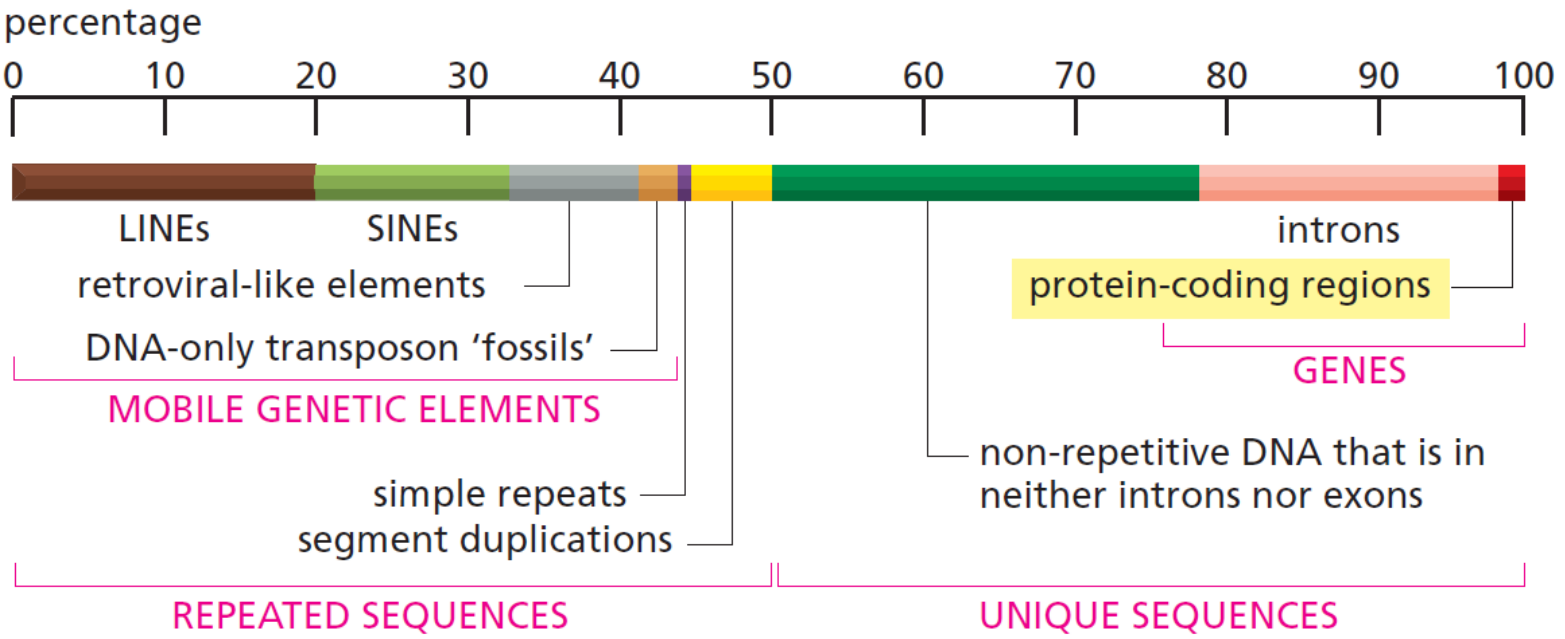
Chromosomes	Mbp
Nuclear	3088
Mitochondrial	0.017
Unknown, alt	121.0

https://www.ncbi.nlm.nih.gov/grc/human

Human genome contents

Regions	Length, Mbp	%	Description
Genes	1,200	37.5	Genomic locus where transcription occurs
Exons	48	1.5	Transcribed genomic region that remain in the RNA after splicing
Other (introns, UTRs)	1,152	36.0	Regions of a coding cDNA which are not translated
Unique and regulatory sequences	510	15.9	
Interspersed repeats	1,400	43.8	
LINEs	640	20.0	~850,000 Long Interspersed Elements (~7,000 bp). Retrotransposed elements containing open reading frames encoding (often inactive) reverse transcription machinery
SINEs, Alu repeats	420	13.1	~1,500,000 Short Interspersed Elements. Retrotransposed elements <500 bp that contain tRNA, snRNA and rRNA, which require other mobile elements to be transposed.
LTR retrotransposons	250	7.8	Transposable elements characterized by the presence of Long Terminal Repeats (LTRs) directly flanking an internal coding region
DNA transposons	90	2.8	Class II transposable elements that move through a DNA intermediate
Microsatellites	90	2.8	A region in the genomic sequence containing short tandem repeats of 2-10bp
Total	3,200	100.0	

Human genome contents



Alberts - *Essential Cell Biology*

Element	Transposition	Structure	Length	Copy number	Fraction of genome
LINEs	Autonomous	ORF1 ORF2 *(pol)* AAA	1–5 kb	20,000–40,000	21%
SINEs	Nonautonomous	AAA	100–300 bp	1,500,000	13%
DNA transposons	Autonomous	transposase	2–3 kb	300,000	3%
	Nonautonomous		80–3000 bp		

Griffiths -- *Introduction to Genetic Analysis*

ENSEMBL gene annotation GRCh38 v.99

Gene biotype	Genes (Transcripts)	%	Description
Proten coding	19,968 (153,197)	32.9	Genes that contain an open reading frame (ORF)
Pseudogenes	15,263	25.2	Genes that have homology to known protein-coding genes but contain a frameshift and/or stop codon(s) which disrupts the ORF
To be confirmed	1,060	1.7	Require experimental validation
T-cell receptors, immunoglobulins	408	0.7	Undergo somatic recombination before transcription
RNA genes	23,977	39.5	
lncRNA	16,880		A non-coding gene >200bp in length
snRNA	1,910		Processing of pre-messenger RNA
miRNA	1,879		A small RNA (~22bp) that silences the expression of target mRNA
snoRNA	942		Post-transcriptional modification of other RNAs
Other	2,366		rRNA, sRNA, scRNA, scaRNA, miscRNA
Total	60,676 (227,818)	100	

Q: which class is *Xist* RNA?

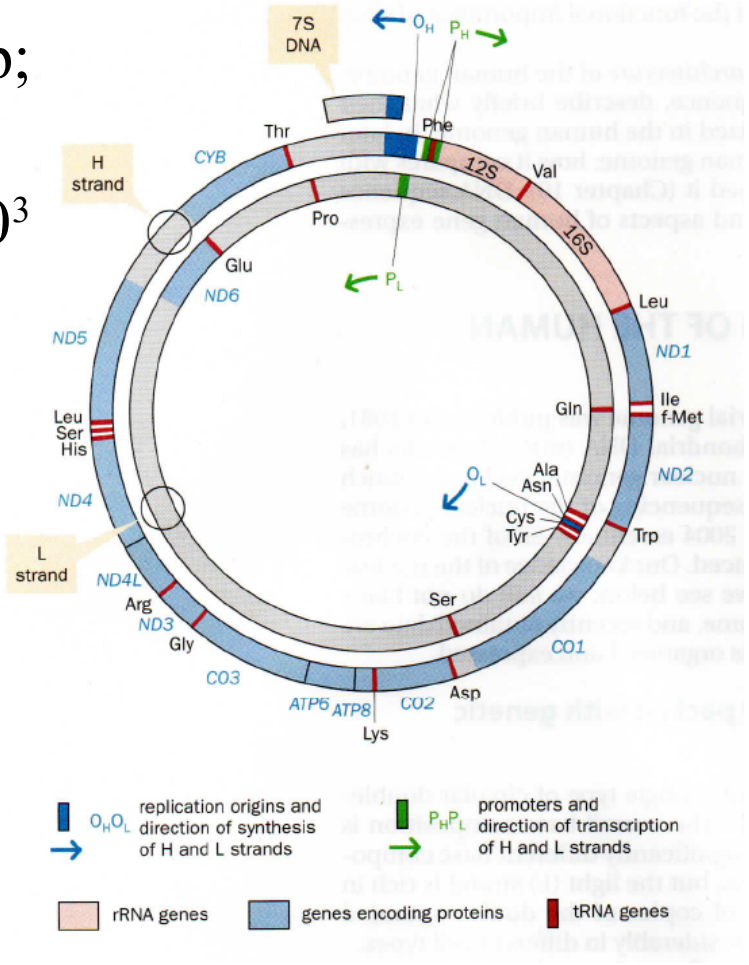
ENSEMBL gene annotation GRCh38 v.99

Chromosome	Approximate length (bp)	Protein-coding genes	Non-protein coding genes	Pseudogenes
1	248956422	2047	1964	1233
2	242193529	1303	1605	1033
3	198295559	1075	1160	768
4	190214555	753	984	732
5	181538259	881	1200	710
6	170805979	1041	989	803
7	159345973	989	977	893
8	145138636	670	1041	629
9	138394717	778	786	678
10	133797422	728	880	568
11	135086622	1312	1053	815
12	133275309	1036	1197	627
13	114364328	321	586	378
14	107043718	820	857	519
15	101991189	613	986	513
16	90338345	867	1033	467
17	83257441	1185	1198	531
18	80373285	269	608	246
19	58617616	1474	895	514
20	64444167	543	594	250
21	46709983	231	403	183
22	50818468	492	513	332
X	156040895	843	640	872
Y	57227415	63	108	392
Mitochondrial	16569	13	24	

Jackson (2018) *Essays Biochem*

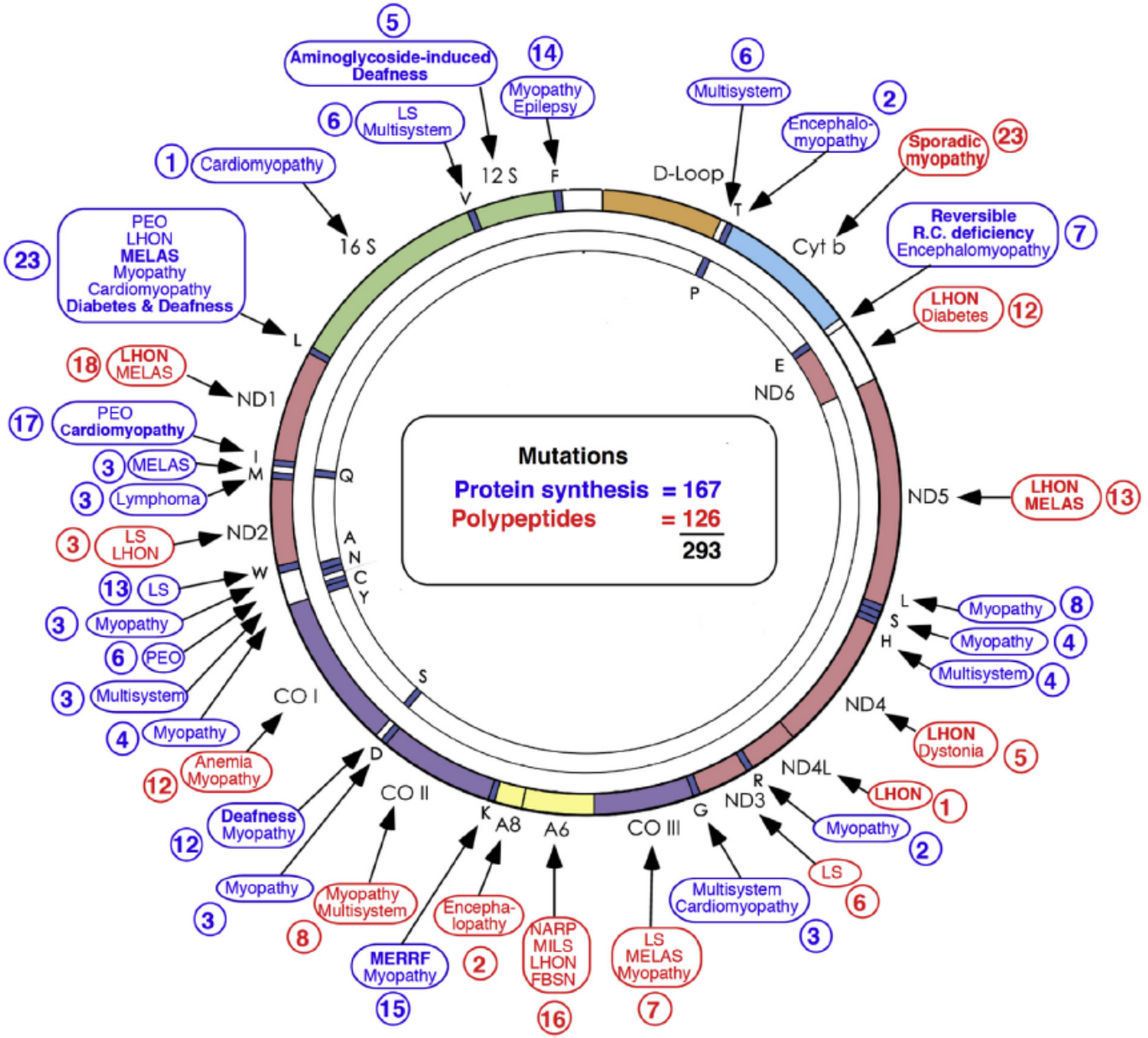
Mitochondrial genome

- **mtDNA**: circular, double-stranded, 16,569 bp; H and L chains; similar to bacteria
- Egg only, maternally inherited; each cell: ~10^3 copies; highly heterogeneous
- 37 genes: 22 tRNA + 2 rRNA + 13 coding
- 13 polypeptides are part of **mitochondrial respiratory complex** (Sugars → ATP), together with multiple nuclear genes
- mtDNA is to some extent autonomous, with its own genetic code
- Stop codons: TAA, TAG, AGA, AGG



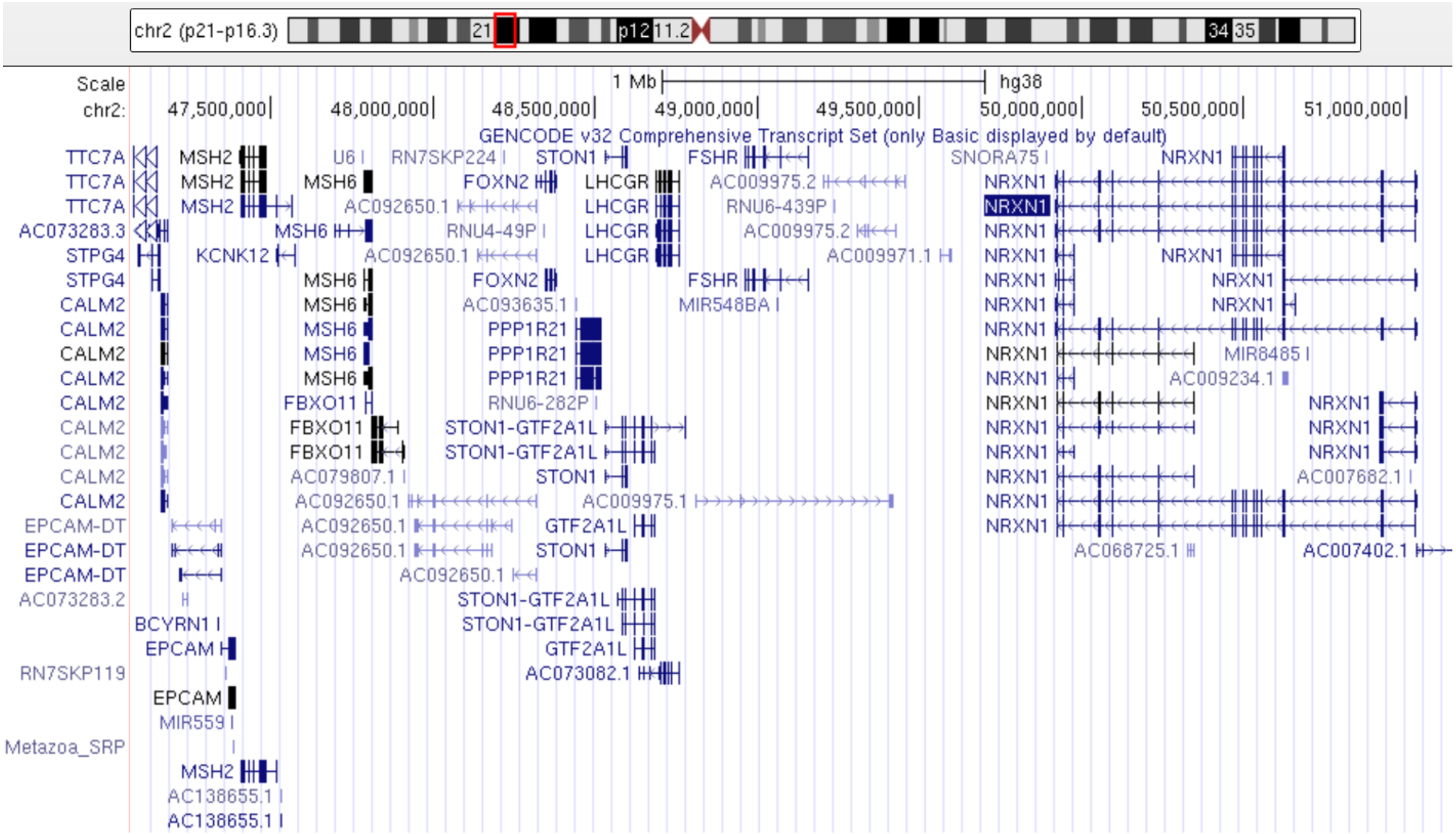
- **Mitochondrial diseases**: a heterogeneous group of inherited anomalies in oxidative phosphorylation due to mutations in the mitochondrial (70%) or nuclear DNA (30%)
- ~300 disease-causing point mutations known in mtDNA

Pathogenic mutations in mtDNA

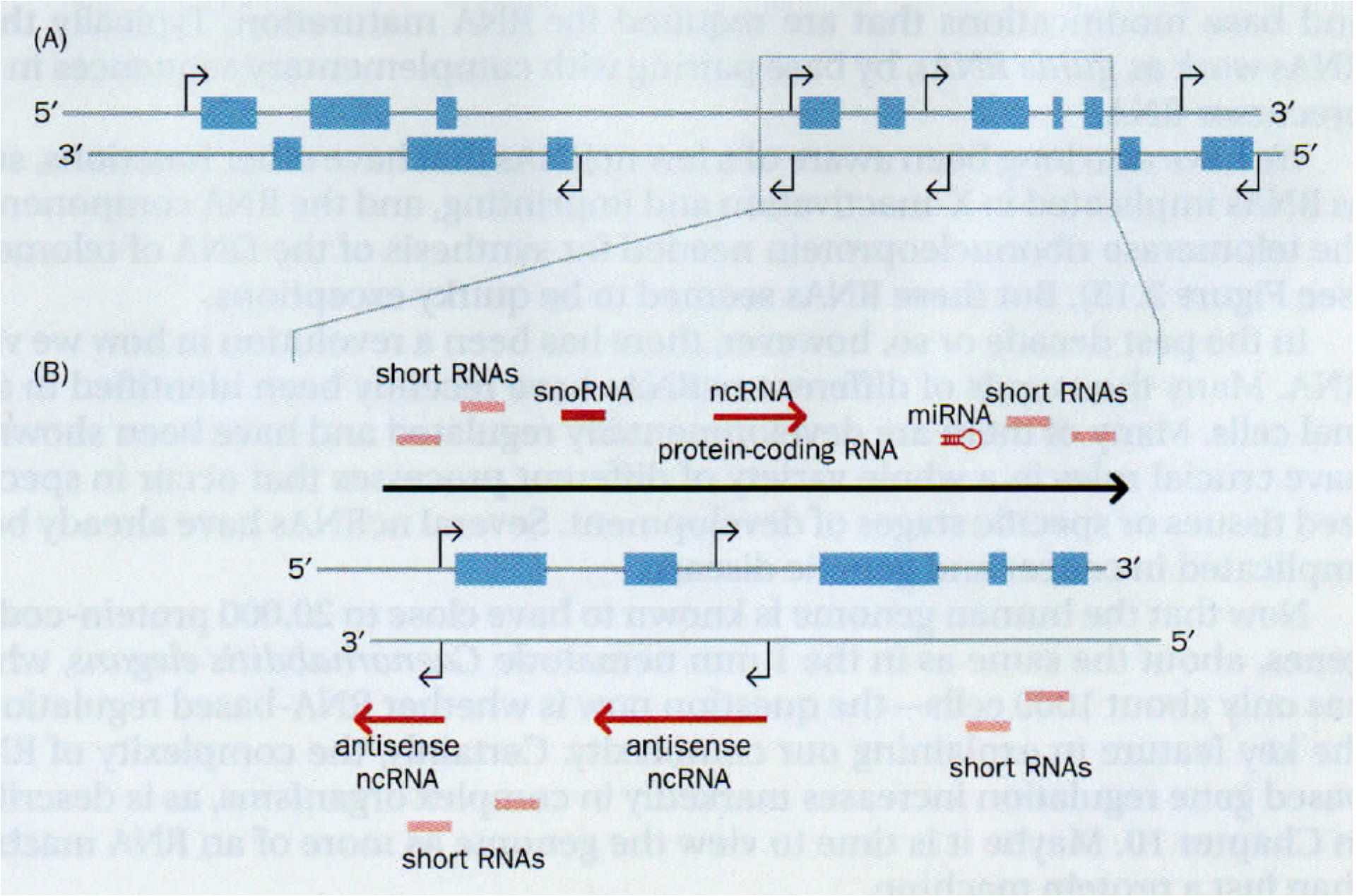


DiMauro (2017) Mitochondrial Encephalomyopathies

UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly

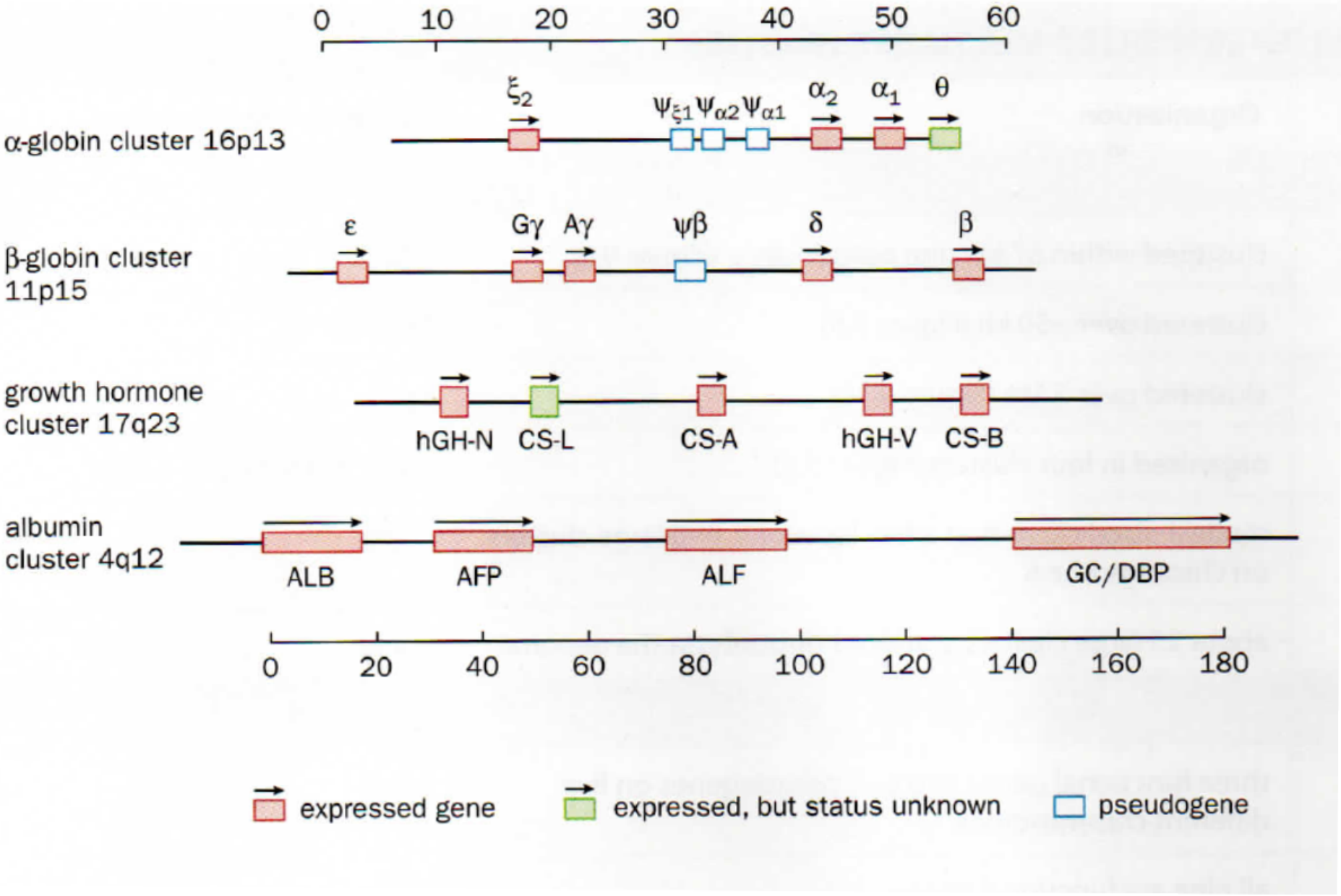


Blurring of gene boundaries



Strachan, Read – *Human Molecular Genetics*

Multigene families



Strachan, Read – *Human Molecular Genetics*

Multigene families

TABLE 9.6 EXAMPLES OF CLUSTERED AND INTERSPERSED MULTIGENE FAMILIES

Family	Copy no.	Organization	Chromosome location(s)
CLUSTERED GENE FAMILIES			
Growth hormone gene cluster	5	clustered within 67 kb; one pseudogene (Figure 9.8)	17q24
α-Globin gene cluster	7	clustered over ~50 kb (Figure 9.8)	16p13
Class I HLA heavy chain genes	~20	clustered over 2 Mb (Figure 9.10)	6p21
HOX genes	38	organized in four clusters (Figure 5.5)	2q31, 7p15, 12q13, 17q21
Histone gene family	61	modest-sized clusters at a few locations; two large clusters on chromosome 6	many
Olfactory receptor gene family	> 900	about 25 large clusters scattered throughout the genome	many
INTERSPERSED GENE FAMILIES			
Aldolase	5	three functional genes and two pseudogenes on five different chromosomes	many
PAX	9	all nine are functional genes	many
NF1 (neurofibromatosis type I)	> 12	one functional gene at 22q11; others are nonprocessed pseudogenes or gene fragments (Figure 9.11)	many, mostly pericentromeric
Ferritin heavy chain	20	one functional gene on chromosome 11; most are processed pseudogenes	many

Human protein classes

PANTHER Protein Class

Total # Genes: 20996 Total # protein class hits: 11214

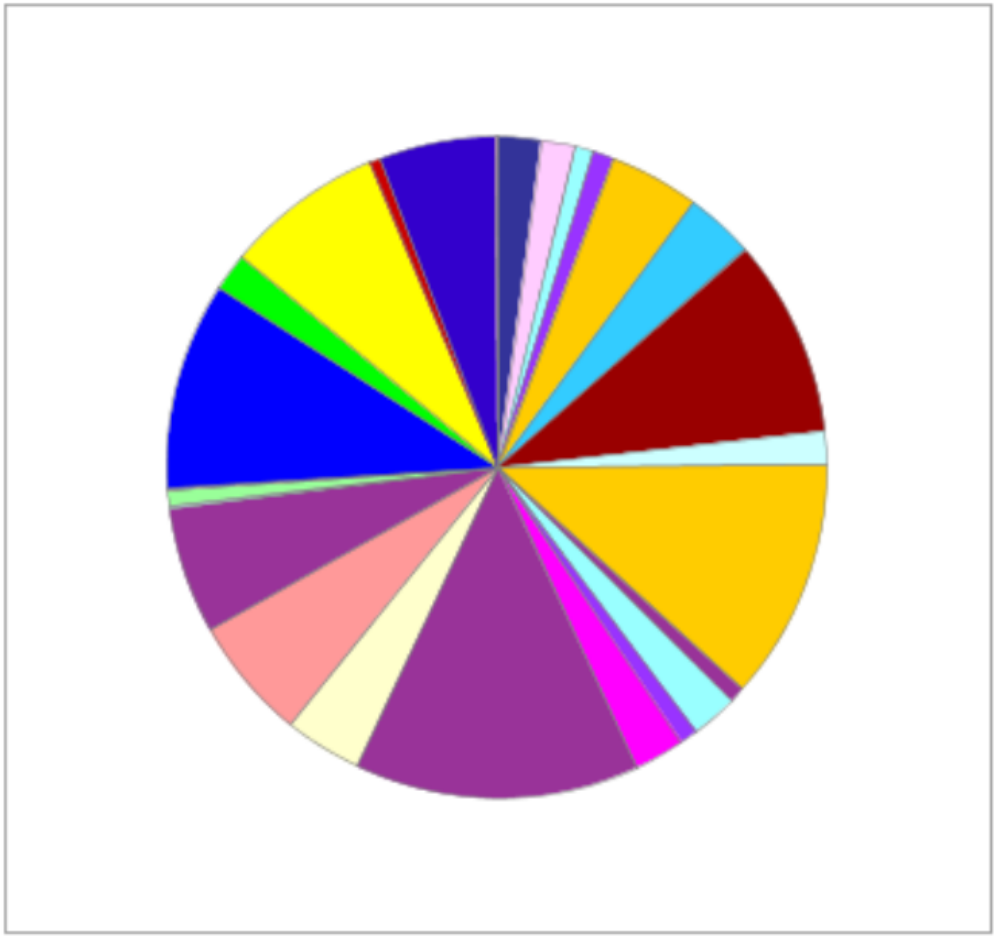


**Chart tooltips are read as: Category name (Accession): # genes; Percent of gene hit against total # genes; Percent of gene hit against total # Protein Class hits

Click to get gene list for a category:

- calcium-binding protein (PC00060)
- cell adhesion molecule (PC00069)
- cell junction protein (PC00070)
- chaperone (PC00072)
- cytoskeletal protein (PC00085)
- defense/immunity protein (PC00090)
- enzyme modulator (PC00095)
- extracellular matrix protein (PC00102)
- hydrolase (PC00121)
- isomerase (PC00135)
- ligase (PC00142)
- lyase (PC00144)
- membrane traffic protein (PC00150)
- nucleic acid binding (PC00171)
- oxidoreductase (PC00176)
- receptor (PC00197)
- signaling molecule (PC00207)
- storage protein (PC00210)
- structural protein (PC00211)
- surfactant (PC00212)
- transcription factor (PC00218)
- transfer/carrier protein (PC00219)
- transferase (PC00220)
- transmembrane receptor regulatory/adaptor protein (PC00226)
- transporter (PC00227)
- viral protein (PC00237)

Human protein classes

1	Nucleic acid binding (PC00171)	1567
2	Hydrolase (PC00121)	1322
3	Transcription factor (PC00218)	1138
4	Enzyme modulator (PC00095)	1079
5	Transferase (PC00220)	867
6	Signaling molecule (PC00207)	693
7	Receptor (PC00197)	675
8	Transporter (PC00227)	638
9	Cytoskeletal protein (PC00085)	497
10	Oxidoreductase (PC00176)	424
11	Defense/immunity protein (PC00090)	386
12	Membrane traffic protein (PC00150)	280
13	Ligase (PC00142)	250
14	Calcium-binding protein (PC00060)	237
15	Transfer/carrier protein (PC00219)	203
16	Cell adhesion molecule (PC00069)	195
17	Extracellular matrix protein (PC00102)	190
18	Chaperone (PC00072)	111
19	Cell junction protein (PC00070)	98
20	Lyase (PC00144)	97
21	Isomerase (PC00135)	85
22	Structural protein (PC00211)	84
23	Transmembrane receptor regulatory/adaptor protein (PC00226	64
24	Storage protein (PC00210)	18
25	Viral protein (PC00237)	8
26	Surfactant (PC00212)	8
27	Unknown	9782
	Total	20996

Exercise: think of appropriate questions

HGNC

HUGO Gene Nomenclature Committee

The resource for approved human gene nomenclature

UniProt

UniProtKB

BLAST Align Retrieve/ID mapping Peptide search

GeneCards®: The Human Gene Database

GeneCards is a searchable, integrative database that provides comprehensive, user-friendly information on all annotated and predicted human genes. The knowledgebase automatically integrates gene-centric data from ~150 web sources, including genomic, transcriptomic, proteomic, genetic, clinical and functional information.

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

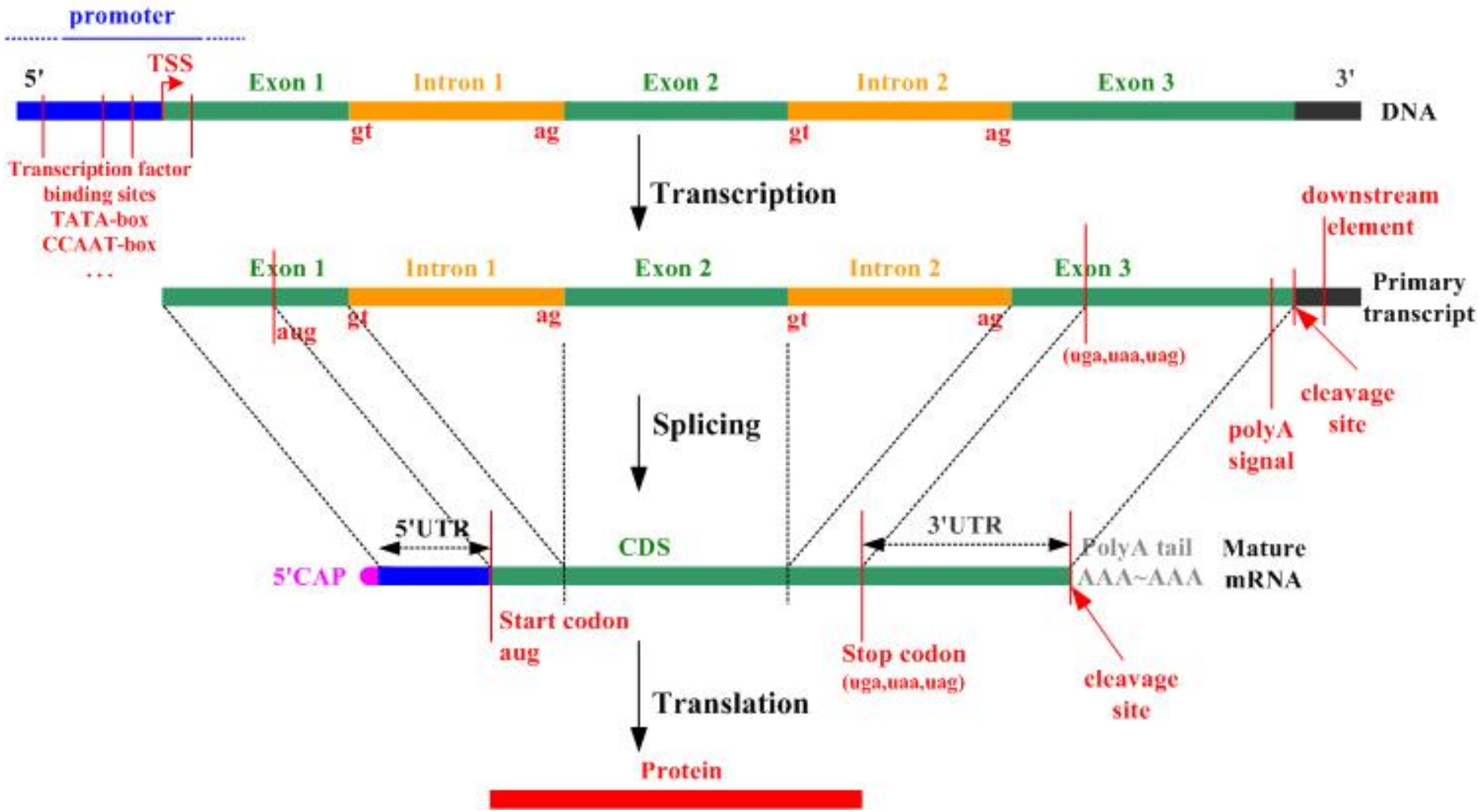
Human (GRCh38.p13) ▾

Search Human (*Homo sapiens*)

Search all categories ▾ | *Search Human...* | Go

e.g. **BRCA2** or **17:63992802-64038237** or **rs699** or **osteoarthritis**

Human gene structure and processing



Note: CDS (coding sequence) vs. mRNA, splicing sites, stop and start codons

Exercise: draw a typical human gene

Human gene structure and processing

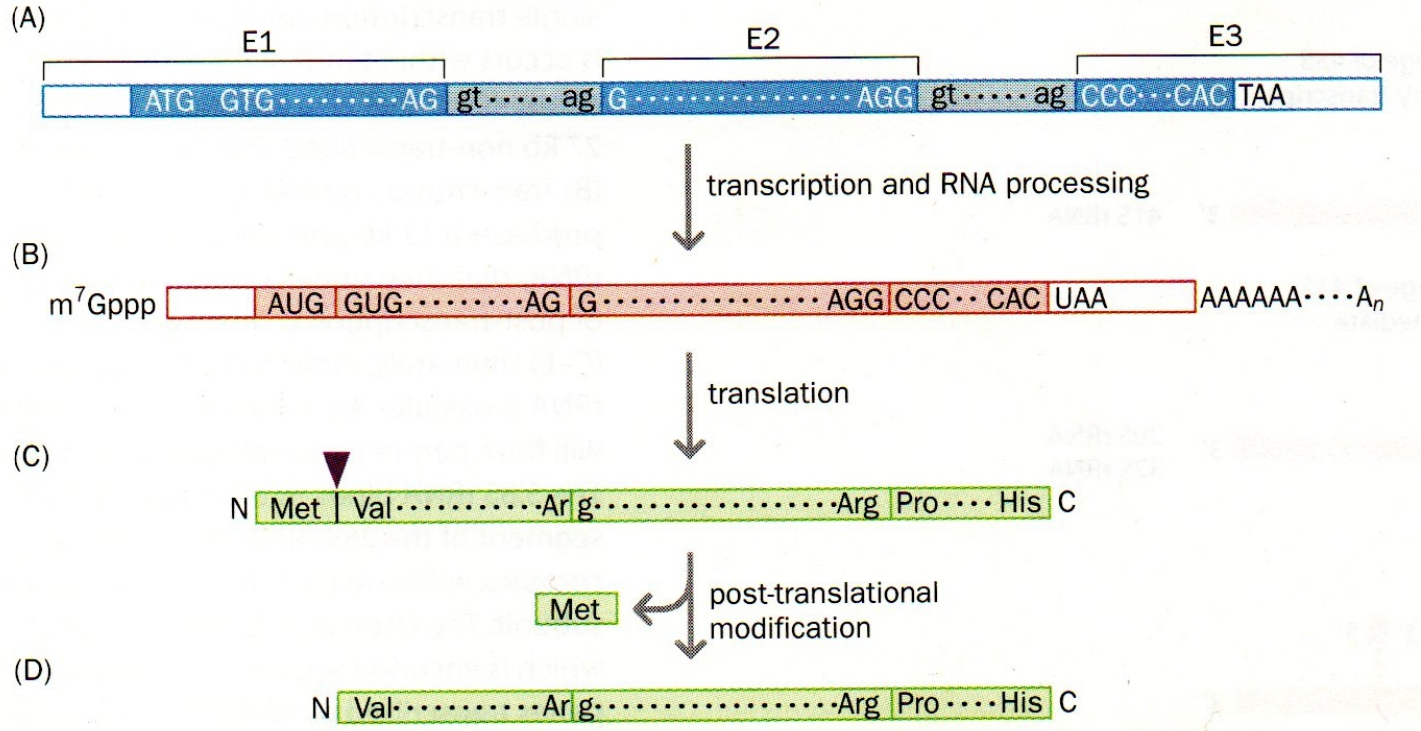


Figure 1.23 Transcription and translation of the human β-globin . (A) The β-globin gene comprises three exons (El-E3) and two introns. The 5'-end sequence of El and the 3' end sequence of E3 are noncoding sequences (unshaded sections). (B) These sequences are transcribed and so occur at the 5' and 3' ends (unshaded sections) of the β-globin mRNA that emerges from RNA processing. (C) Some codons can be specified by bases that are separated by an intron. The Arg104 is encoded by the last three nucleotides (AGG) of exon 2 but the Arg30 is encoded by an AGG codon whose first two bases are encoded by the last two nucleotides of exon 1 and whose third base is encoded by the first nucleotide of exon 2. (D) During post-translational modification the 147·amino acid precursor polypeptide undergoes cleavage to remove ils *N*-terminal methionine residue, to generate the mature 146-residue β-globin protein. The flanking *N* and *C* symbols to the left and right, respectively, in (C) and (D) depict the *N*-terminus and *C*-terminus.

Human gene structure and processing

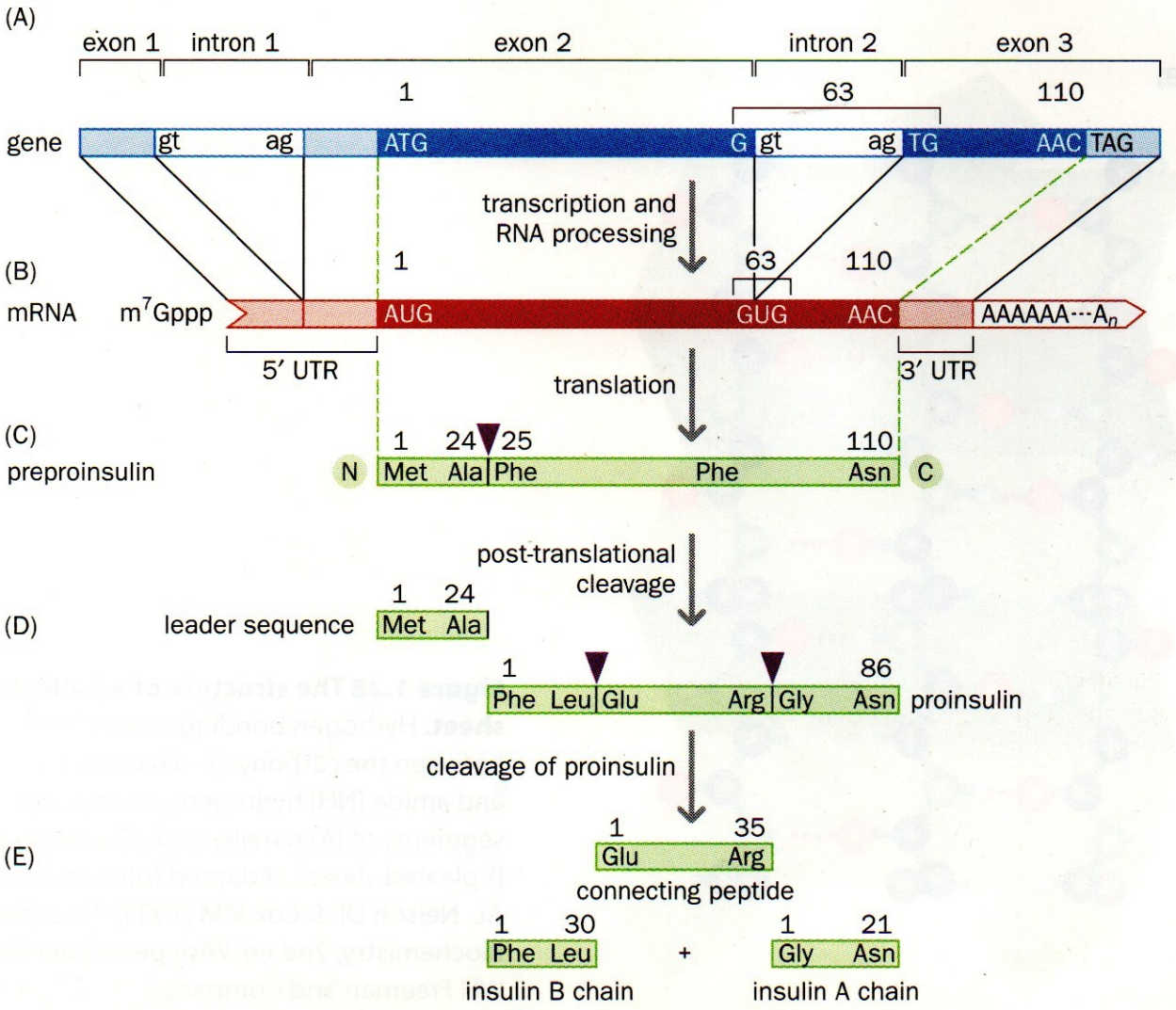


Figure 1.26 Insulin synthesis involves multiple post-translational cleavages of polypeptide precursors. (A) The human insulin gene comprises three exons and two introns. The coding sequence (the part that will be used to make polypeptide) is shown in deep blue. It is confined to the 3′ sequence of exon 2 and the 5′ sequence of exon 3. (B) Exon 1 and the 5′ part of exon 2 specify the 5′ untranslated region (5′ UTR), and the 3′ end of exon 3 specifies the 3′ UTR. The UTRs are transcribed and so are present at the ends of the mRNA. (C) A primary translation product, preproinsulin, has 110 residues and is cleaved to give (D) a 24-residue N-terminal *leader sequence* (that is required for the protein to cross the cell membrane but is thereafter discarded) plus an 86-residue proinsulin precursor. (E) Proinsulin is cleaved to give a central segment (the connecting peptide) that may maintain the conformation of the A and B chains of insulin before the formation of their interconnecting covalent disulfide bridges (see Figure 1.29).

Examples of post-translational processing

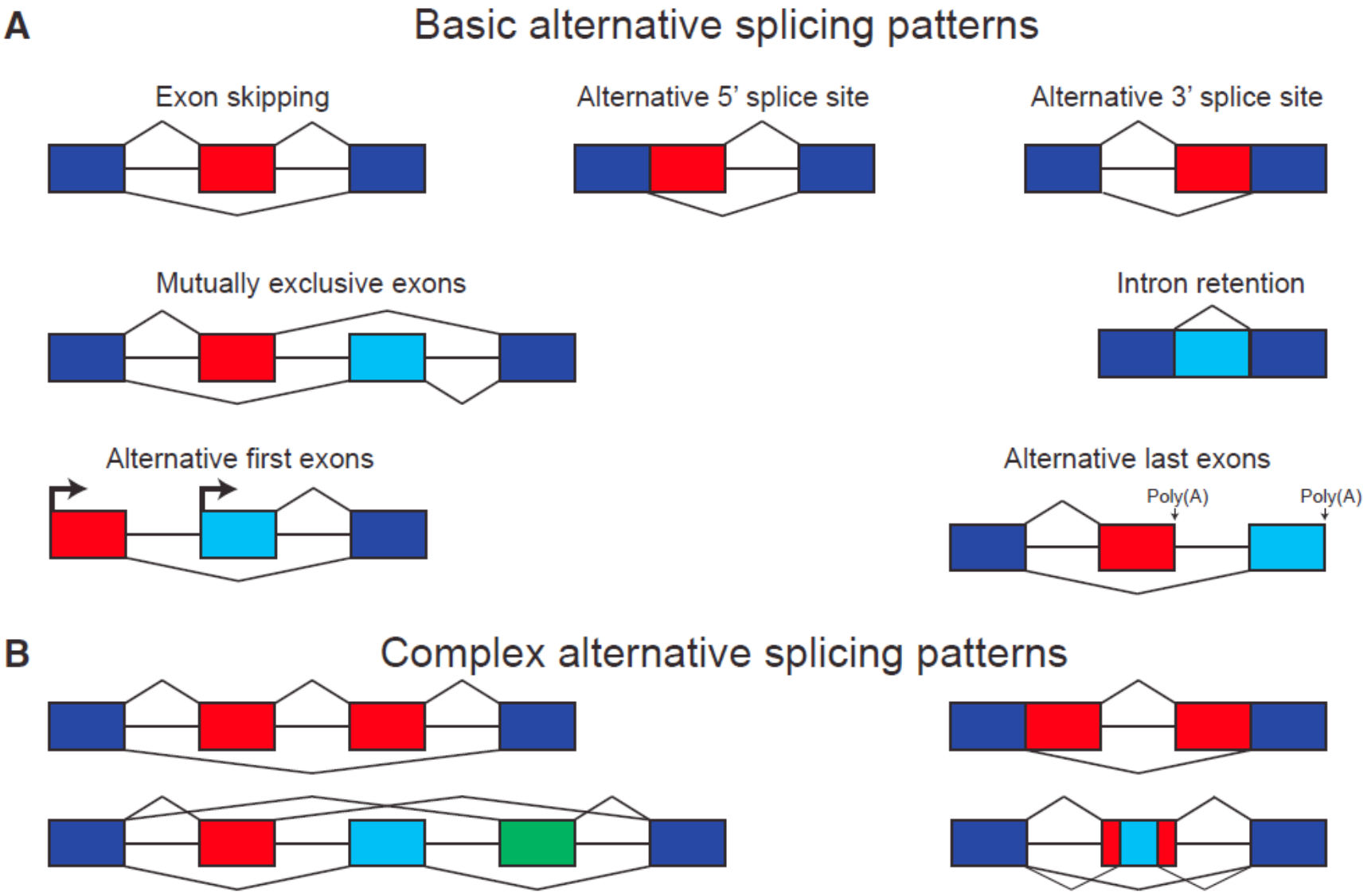
Human gene structure and processing

TABLE 9–1 SOME VITAL STATISTICS FOR THE HUMAN GENOME	
DNA length	3.2×10^9 nucleotide pairs*
Number of genes	approximately 25,000
Largest gene	2.4×10^6 nucleotide pairs
Mean gene size	27,000 nucleotide pairs
Smallest number of exons per gene	1
Largest number of exons per gene	178
Mean number of exons per gene	10.4
Largest exon size	17,106 nucleotide pairs
Mean exon size	145 nucleotide pairs
Number of pseudogenes**	more than 20,000
Percentage of DNA sequence in exons (protein coding sequences)	1.5%
Percentage of DNA in other highly conserved sequences***	3.5%
Percentage of DNA in high-copy repetitive elements	approximately 50%

Q: what gene (exon) is the largest?

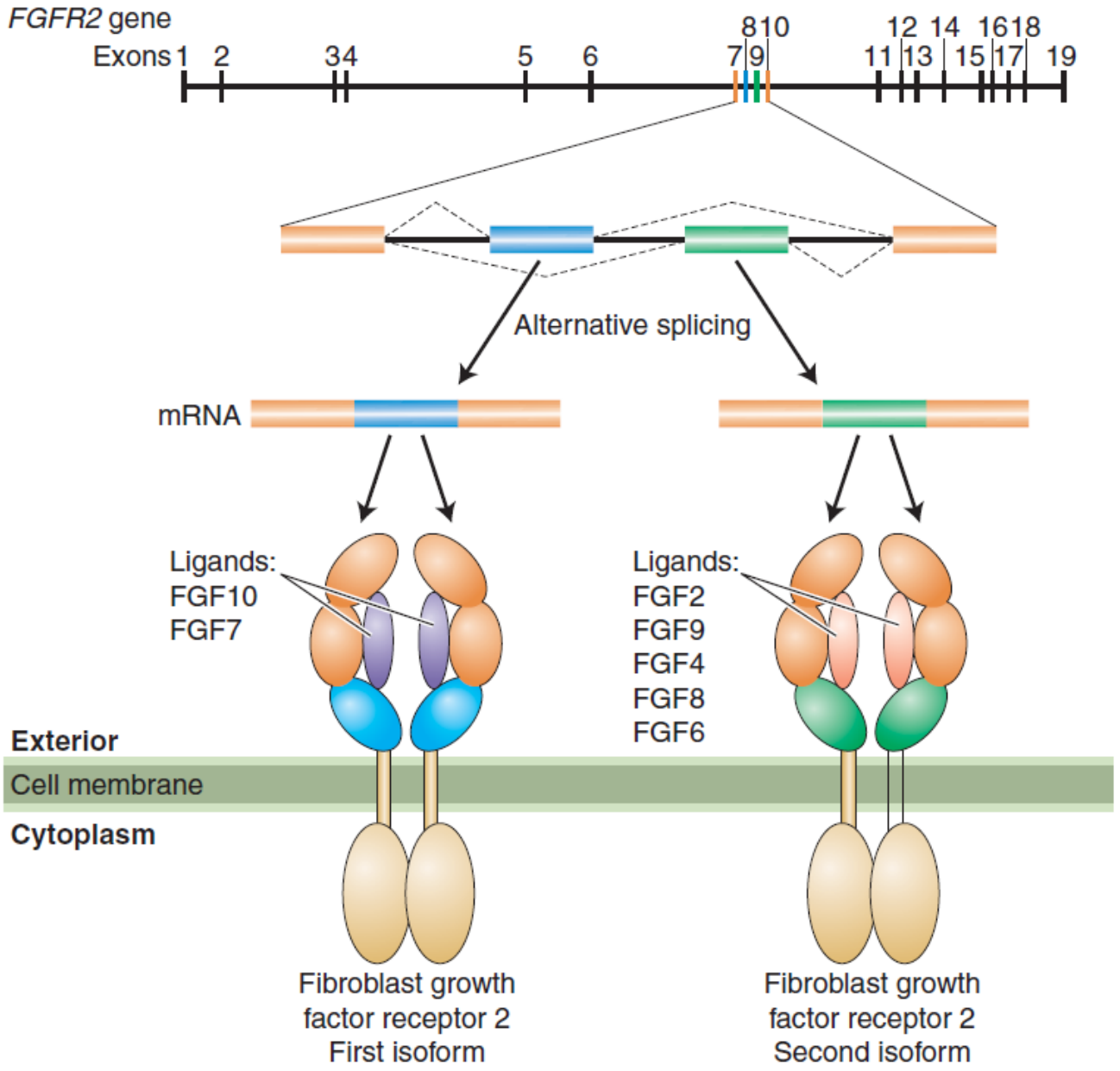
Alberts – *Essential Cell Biology*

Alternative splicing of human genes



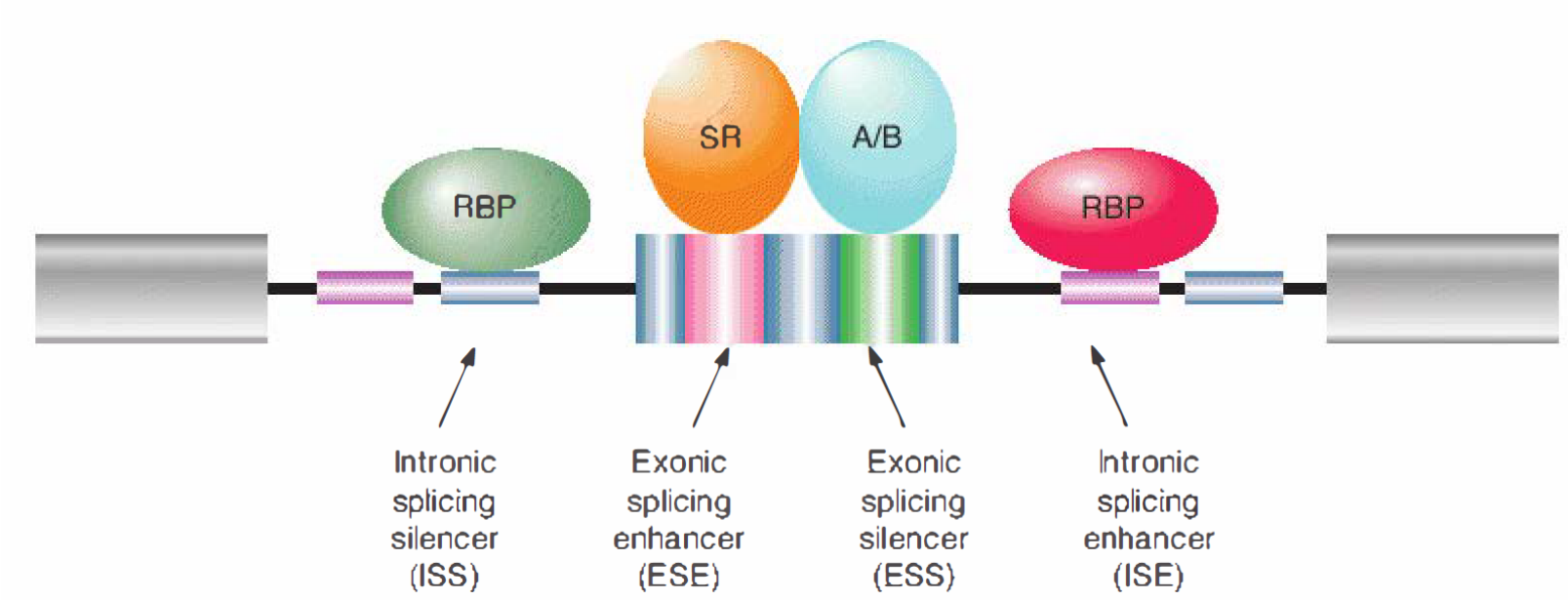
Park (2018) *Am J Hum Genet*

Alternative splicing of human genes

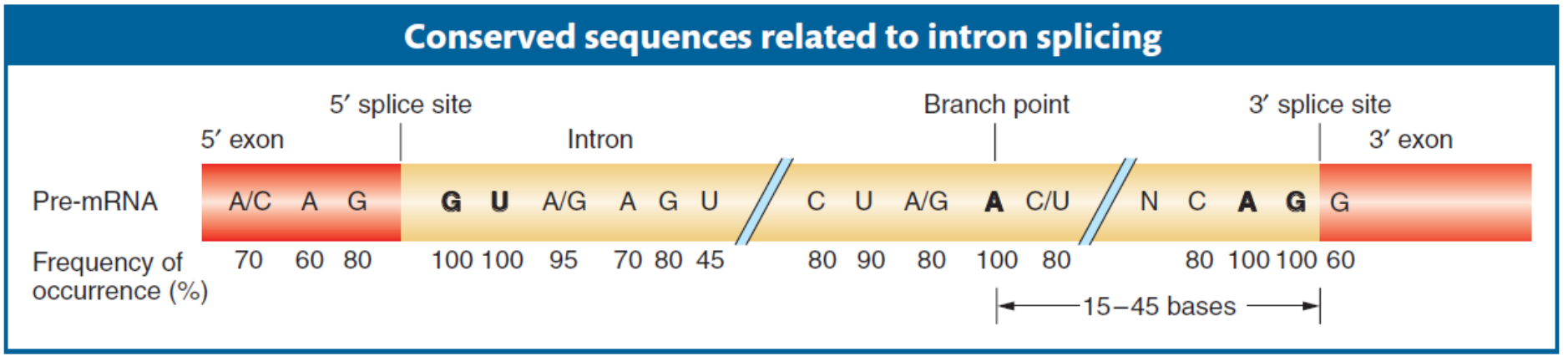


Griffiths -- *Introduction to Genetic Analysis*

Alternative splicing of human genes



Lewin – *Genes XI*



Griffiths -- *Introduction to Genetic Analysis*

Alternative splicing of human genes

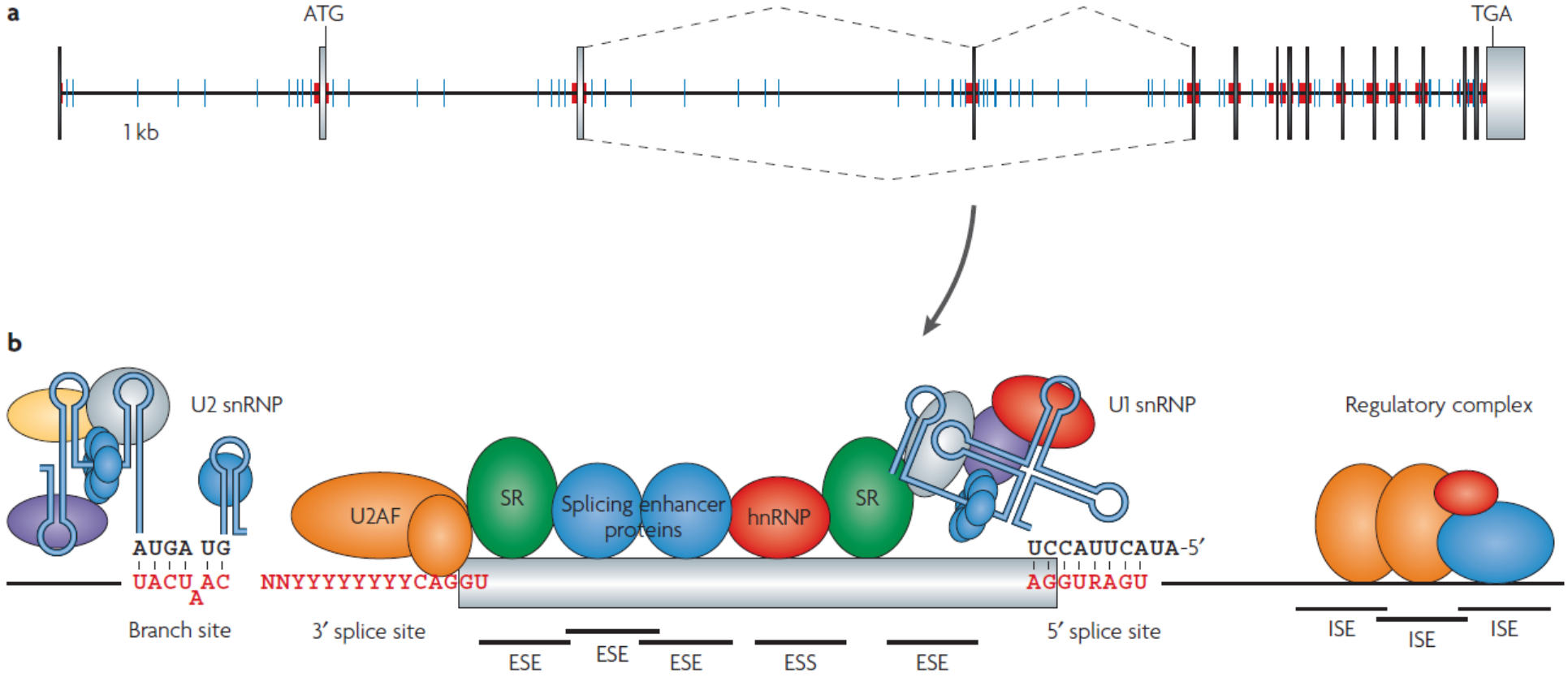


Figure 1 | **The splicing code. a** | A pre-mRNA as it might appear to the spliceosome. Red indicates consensus splice site sequences at the intron–exon boundaries. Blue indicates additional intronic cis-acting elements that make up the splicing code. **b** | cis-elements within and around an alternative exon are required for its recognition and regulation. The 5′ splice site and branch site serve as binding sites for the RNA components of U1 and U2 small nuclear ribonucleoprotein (snRNPs), respectively. This RNA:RNA base pairing determines the precise joining of exons at the correct nucleotides. Mutations in the pre-mRNA that disrupt this base pairing decrease the efficiency of exon recognition. Exons and introns contain diverse sets of enhancer and suppressor elements that refine bone fide exon recognition. Some exon splicing enhancers (ESEs) bind SR proteins and recruit and stabilize binding of spliceosome components such as U2AF. Exon splicing suppressors (ESSs) bind protein components of heterogeneous nuclear ribonucleoproteins (hnRNP) to repress exon usage. Some intronic splicing enhancers (ISEs) bind auxiliary splicing factors that are not normally associated with the spliceosome to regulate alternative splicing.

Wang (2007) *Nat Rev Genet*

Alternative splicing of human genes

- ENSEMBL GRCh38 v.99, protein-coding genes and transcripts:
  - 1 transcript: 22.2% (no alternative splicing)
  - 2-5 transcripts: 52.9%
  - >5 transcripts: 24.9%
  - More than 75 transcripts: *ADGRG1, ANK2, KCNMA1, MAPK10, NDRG2, PAX6, TCF4*

- Longest transcript designated as **canonical** (≠ most biologically relevant)
- AS contribution to proteome complexity and transcript functionality is still debated: transcripts ≠ protein isoforms
- AS transcripts that introduce premature stop codon are subject to NMD (**nonsense-mediated decay**)
- Microexons (3-30 nt): misregulated in autistic brain (Irimia (2014) *Cell*).

Aberrant splicing in disease

- **Cis-acting splicing mutations**: disruption of the splicing code, **15-60% of human disease mutations** (Wang 2007 *Nat Rev Genet*)

Examples: synonymous mutations in *CFTR* ⟹ cystic fibrosis;

Splice site mutations in *MITF* ⟹ Waardenburg syndrome type 2 (WS2), a dominantly inherited syndrome of hearing loss and pigmentary disturbances

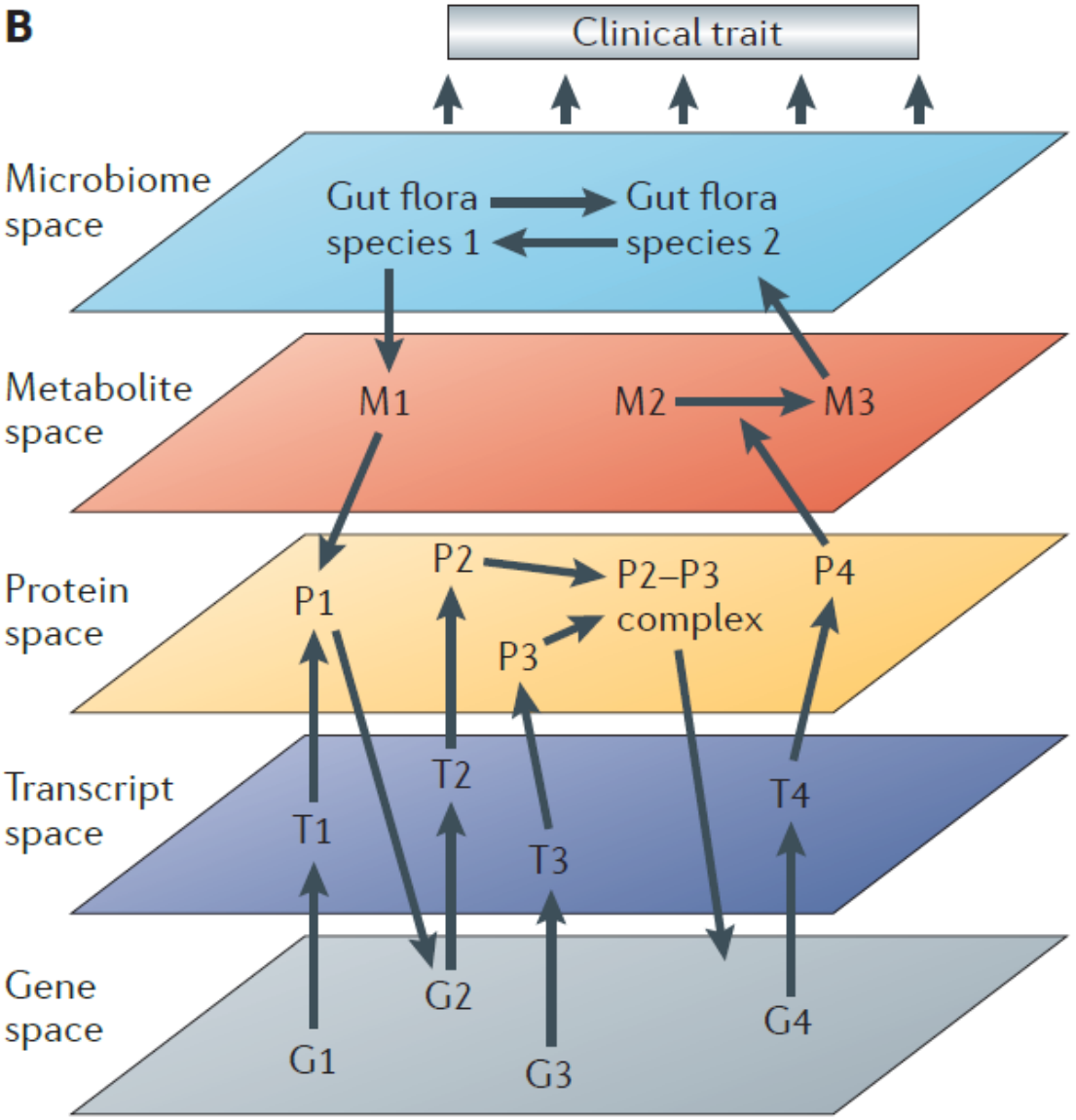
- **Trans-acting mutations**: disruption of the splicing RNA-protein machinery.

Example: mutations in *SMN1* ⟹ loss of snRNP production ⟹ spinal muscular atrophy (SMA). Nusinersen, an antisense oligonucleotide drug for correcting splicing in spinal muscular atrophy.

Park, E., Pan, Z., Zhang, Z., Lin, L., and Xing, Y. (2018). The Expanding Landscape of Alternative Splicing Variation in Human Populations. *Am. J. Hum. Genet*. 102, 11–26.

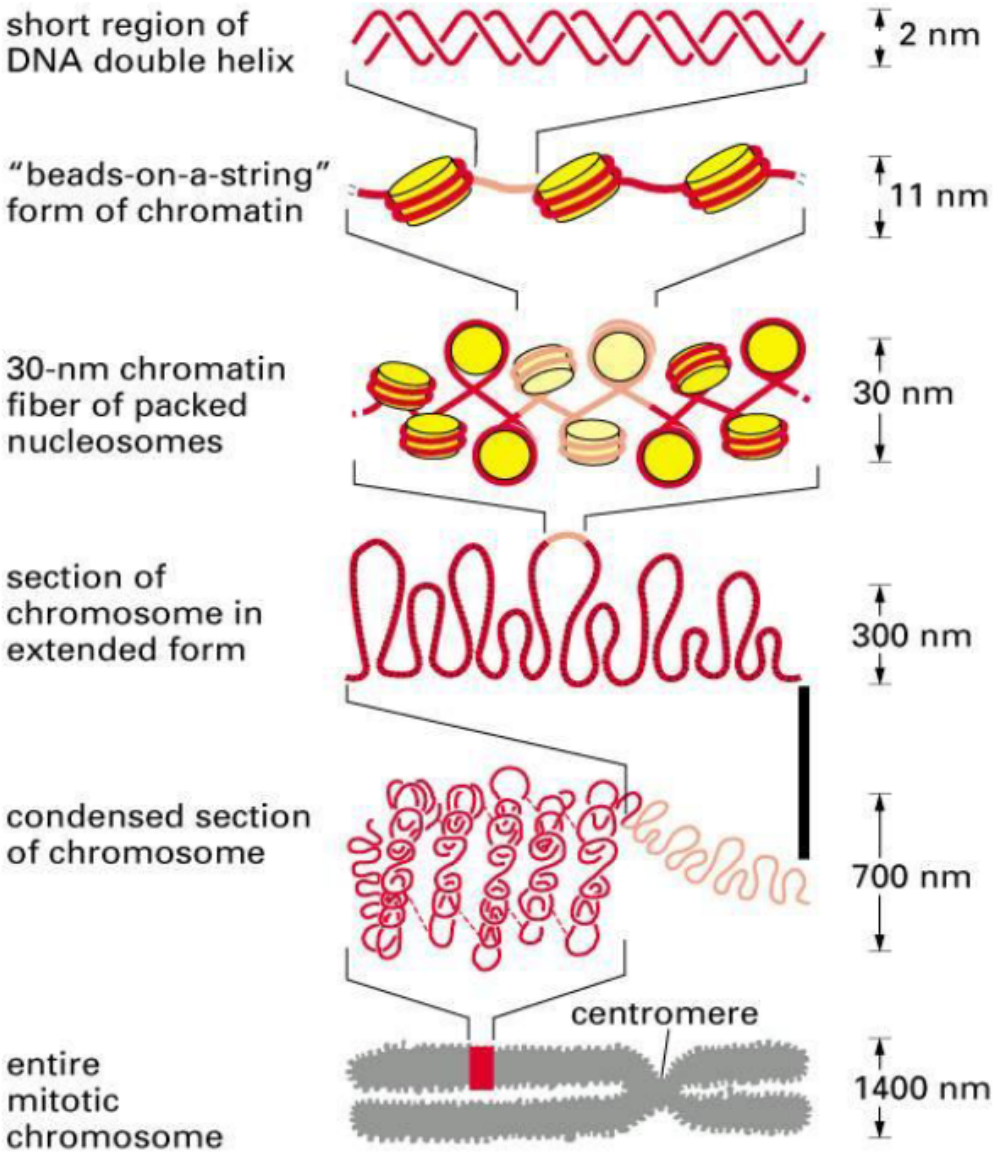
Wang, G.-S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet*. 8, 749–761.

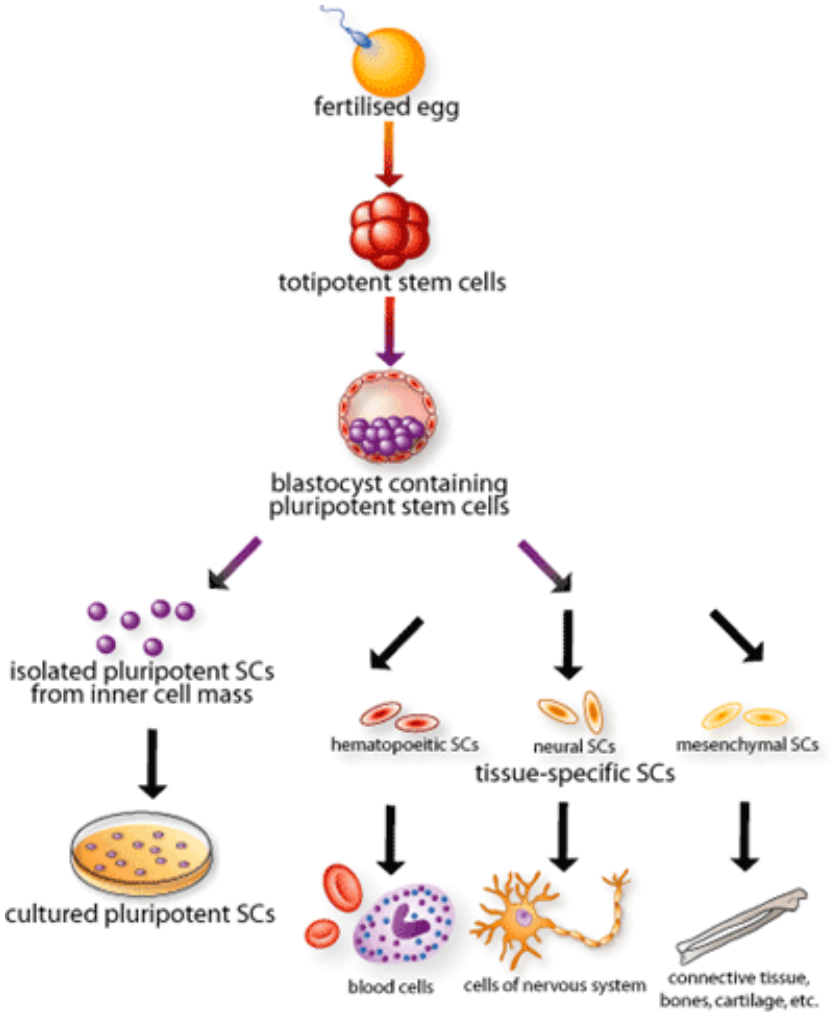
Human genome in action



Civelek (2014) *Nat Rev Genet*

More realistic picture





Molecular Biology of the Cell, 4th ed.

Chaudrey (2004) *Stem Cell Bioeng*

Epigenetics

Epigenetics: heritable phenotype changes that do not involve alterations in the DNA sequence

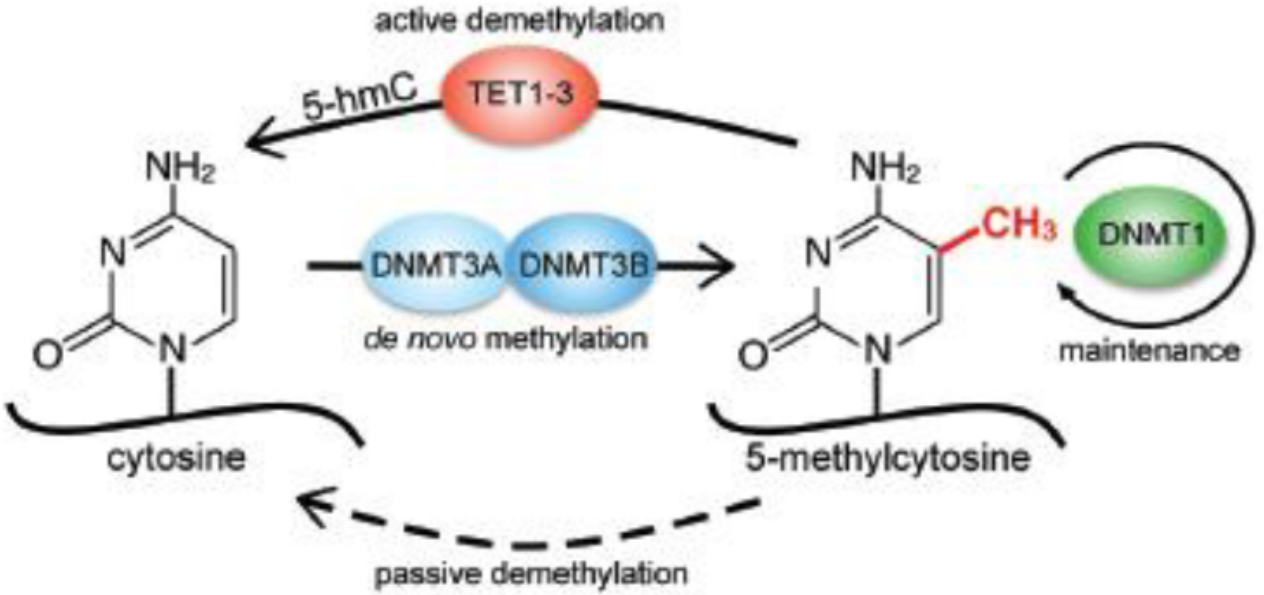
Epigenetic regulation:

1. DNA methylation at CpG dinucleotides
2. Covalent modification of histone proteins
3. Noncoding RNAs

- *Above the genetis*: instructions on using instructions, or gene expression control mechanisms
- Methylation and histone modifications are reversible
- Maintained at cell division and erased during early embriogenesis
- Affected by internal (development, aging) and environmental (chemicals, drugs, diet, lifestyle) factors

DNA methylation

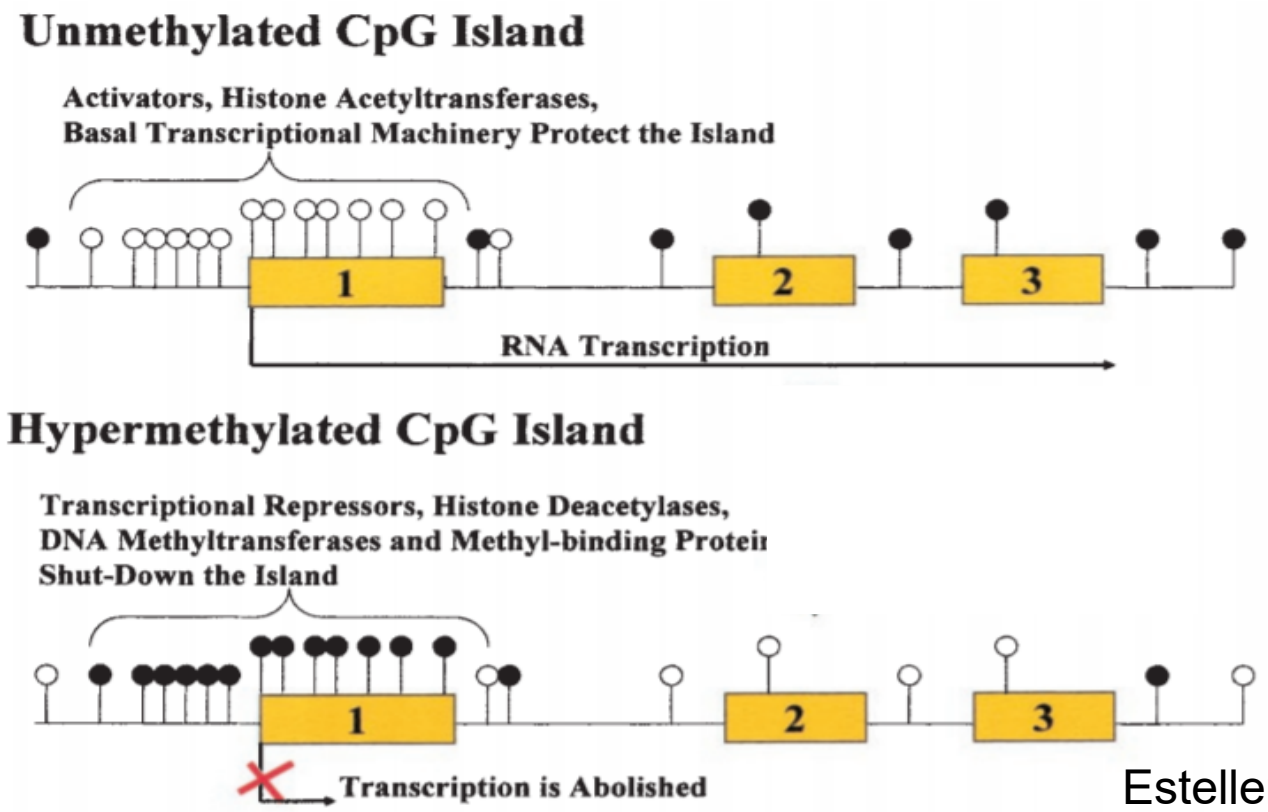
- The only known epigenetic modification of DNA in mammals is methylation of cytosine at position C_5 in CpG dinucleotides
- DNA methyltransferases (DNMTs) establish and maintain DNA methylation patterns
- Methyl-CpG binding proteins (MBDs) read them
- Patterns of CpG methylation may be person-specific, tissue-specific, or locus-specific



Ambrosi (2017) *J Mol Biol*

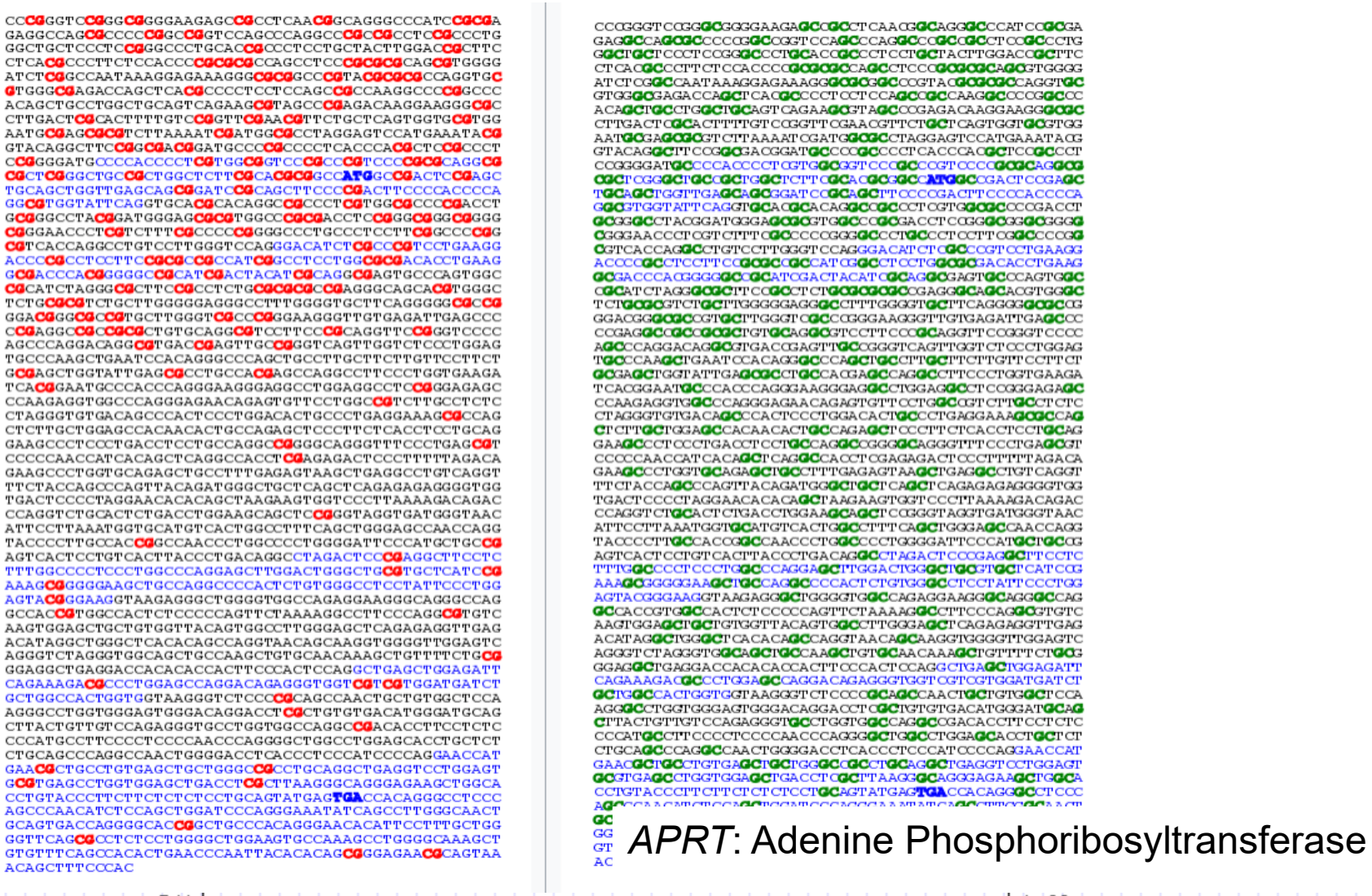
CpG dinucleotides and islands

- Few CpG dinucleotides interrupted by so-called CpG islands
- **CpG island** *ad hoc* definition: length >200 bp, CG >50%, observed-to-expected CpG ratio >60%
- ~30,000 CpG islands in the human genome
- ~70% of human promoters have high CpG content (Saxonov 2006 *PNAS*)
- Methylation of CpG islands silences gene expression

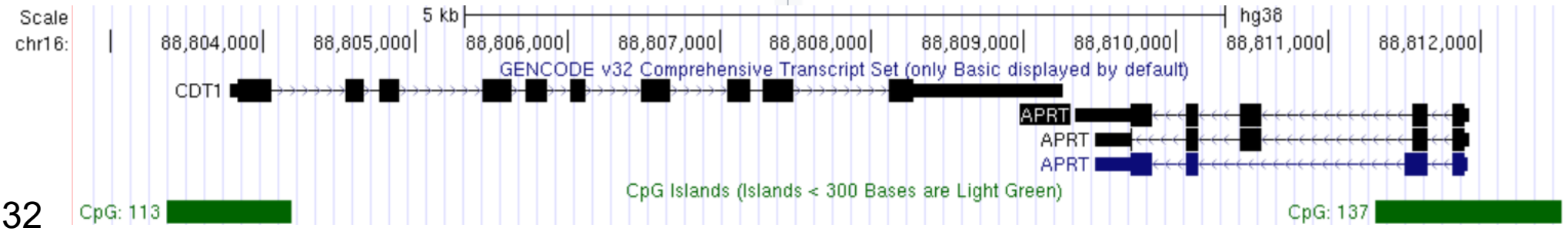


Esteller (2002) *Oncogene*

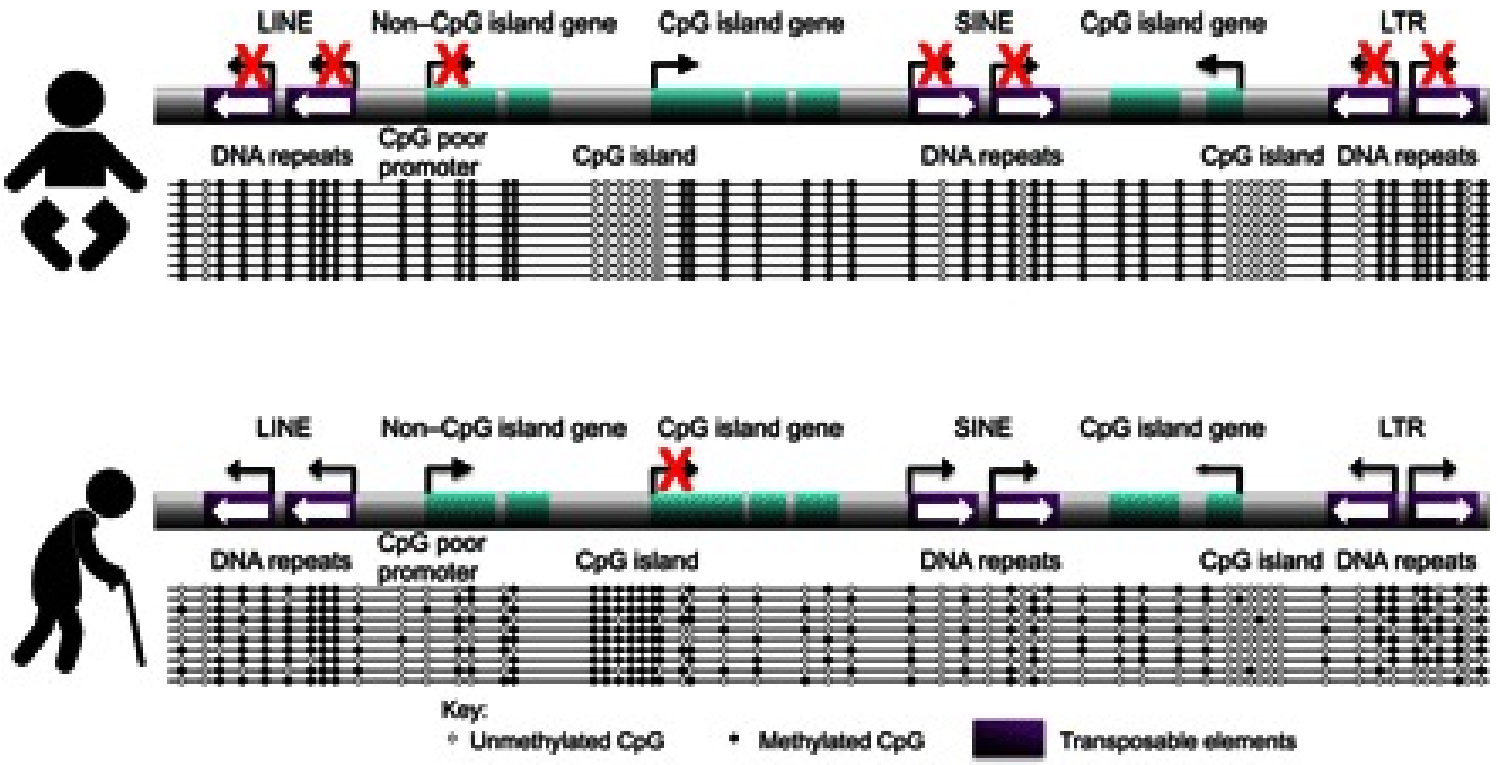
CpG dinucleotides and islands



APRT: Adenine Phosphoribosyltransferase



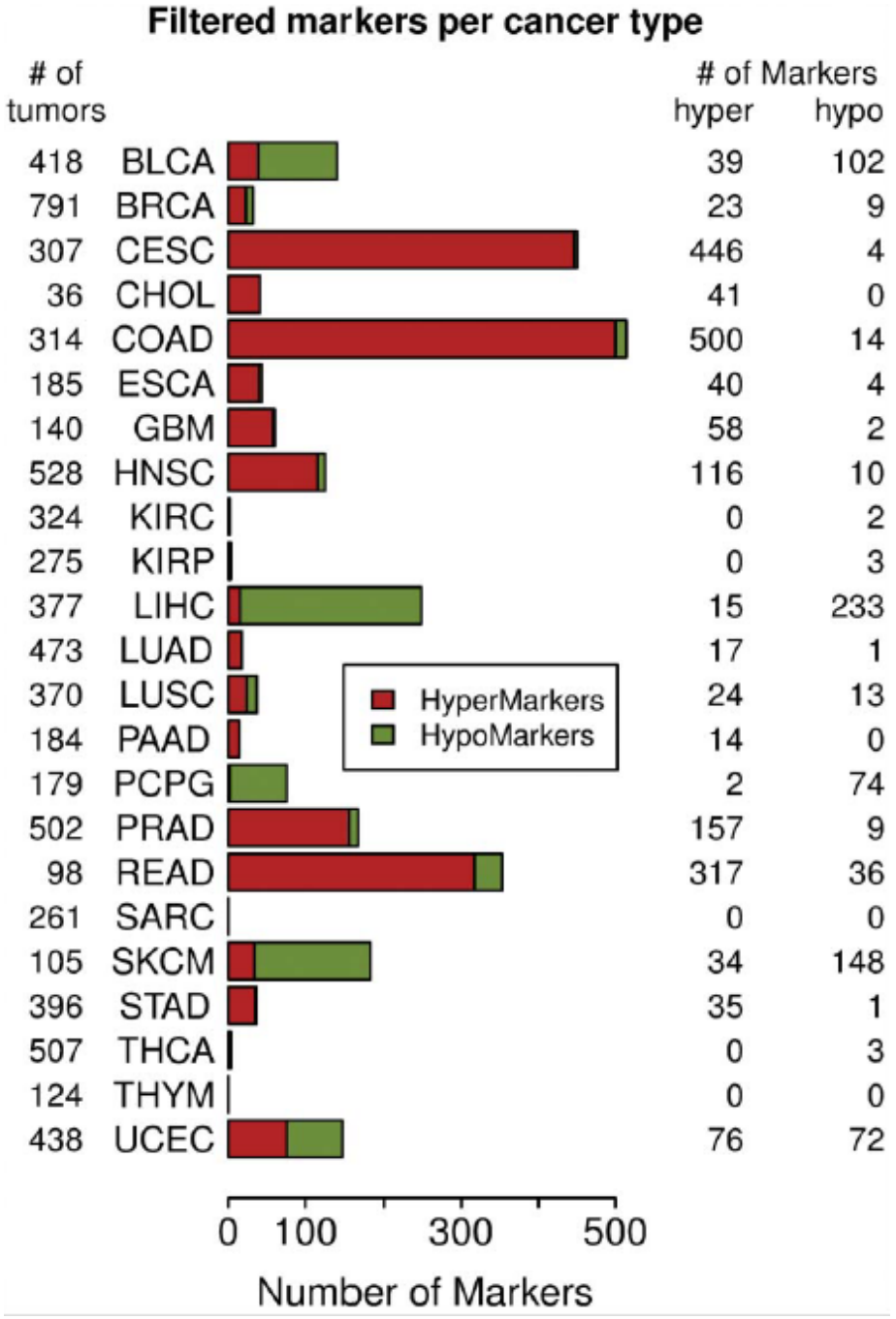
DNA methylation and aging



Young mammalian cells are characterized by DNA hypermethylation over the genome, with the exception of CpG islands within the promoters of expressed genes. In particular, DNA repeats, such as LINE, SINE, and long terminal repeat (LTR) transposable elements, are heavily DNA-methylated, helping to maintain them in a constitutive heterochromatin state. During aging, there is general DNA hypomethylation over the genome, which mostly occurs in a stochastic manner within the cell population. Loss of DNA methylation leads to activation of normally silenced DNA sequences like the transposable elements. However, DNA methylation also increases in a nonstochastic manner over the CpG islands of certain genes, correlating with their heterochromatinization and silencing.

Pal & Tyler (2016) *Sci Adv*

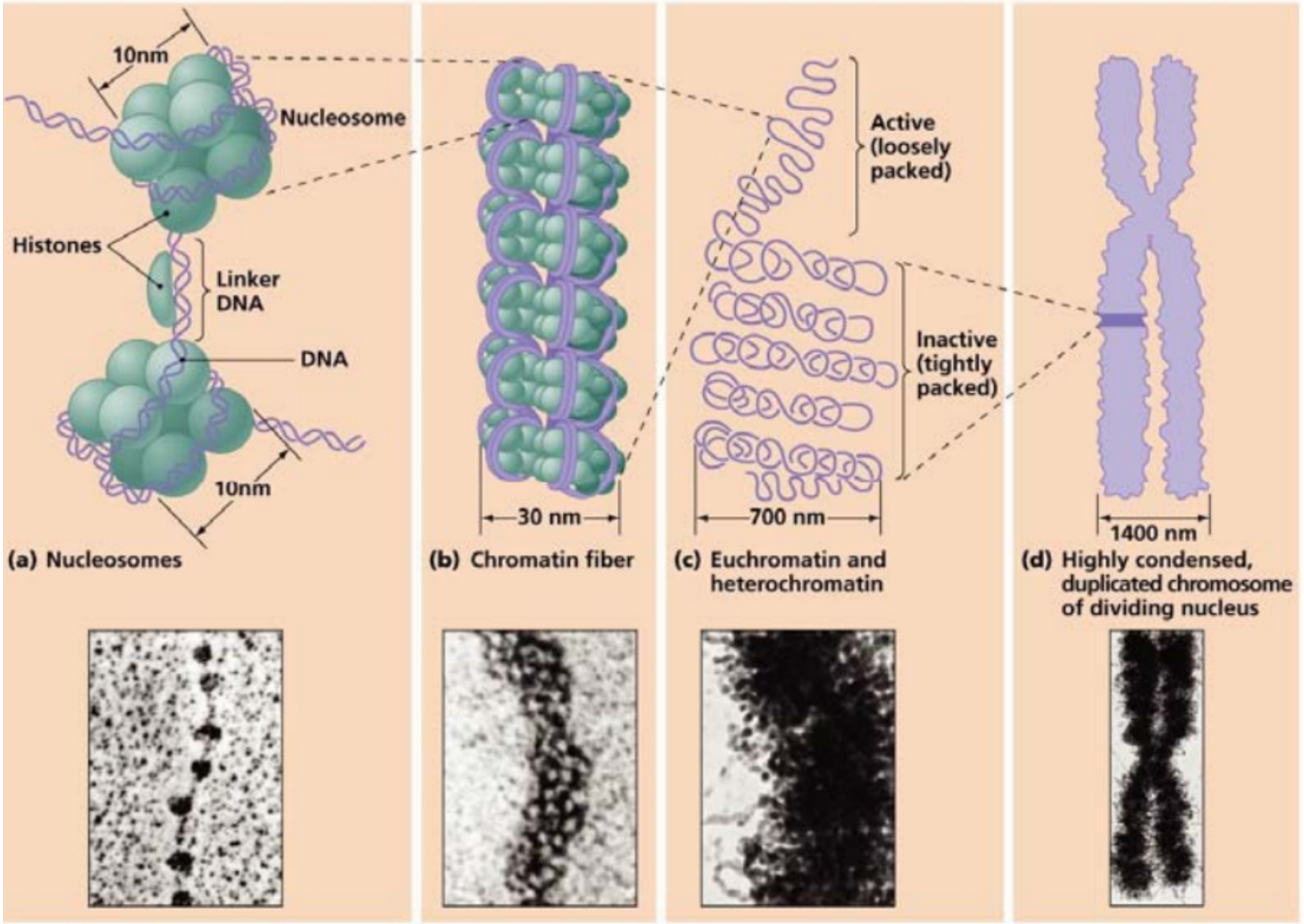
DNA methylation and cancer



We identified **differentially methylated regions for individual cancer types** and those were further filtered against data from normal tissues to obtain marker regions with cancer-specific methylation, resulting in a total of 1,250 hypermethylated and 584 hypomethylated marker CpGs. From hypermethylated markers, optimal sets of six markers for each TCGA cancer type were chosen that could identify most tumors with high specificity and sensitivity [area under the curve (AUC): 0.969-1.000] and a universal 12 marker set that can detect tumors of all 33 TCGA cancer types (AUC >0.84).

Vrba & Futscher (2018) *Epigenetics*

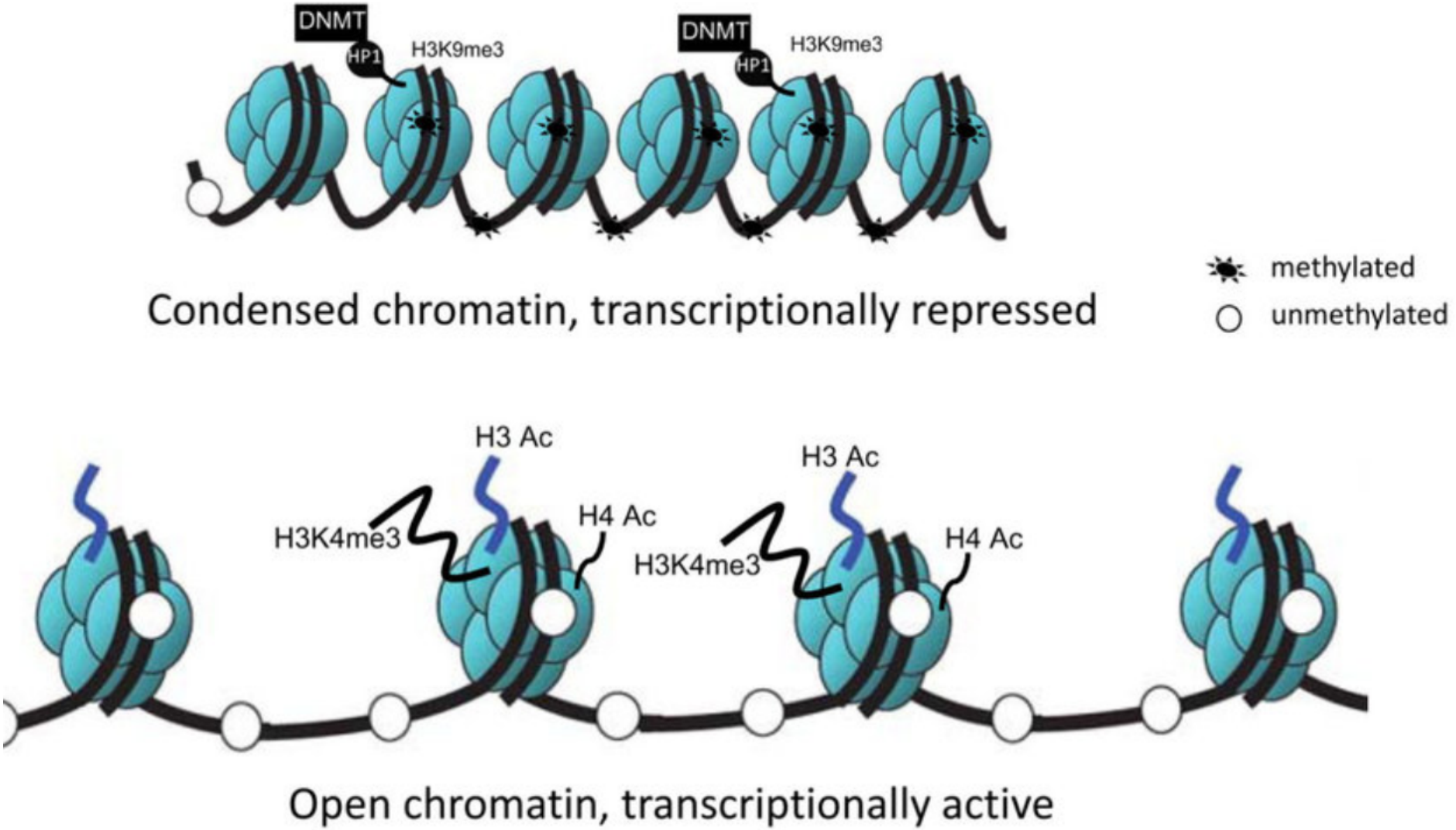
Histone modifications



Copyright © 2006 Pearson Education, Inc., publishing as Benjamin Cummings.

themysteryofepigenetics.blogspot.com

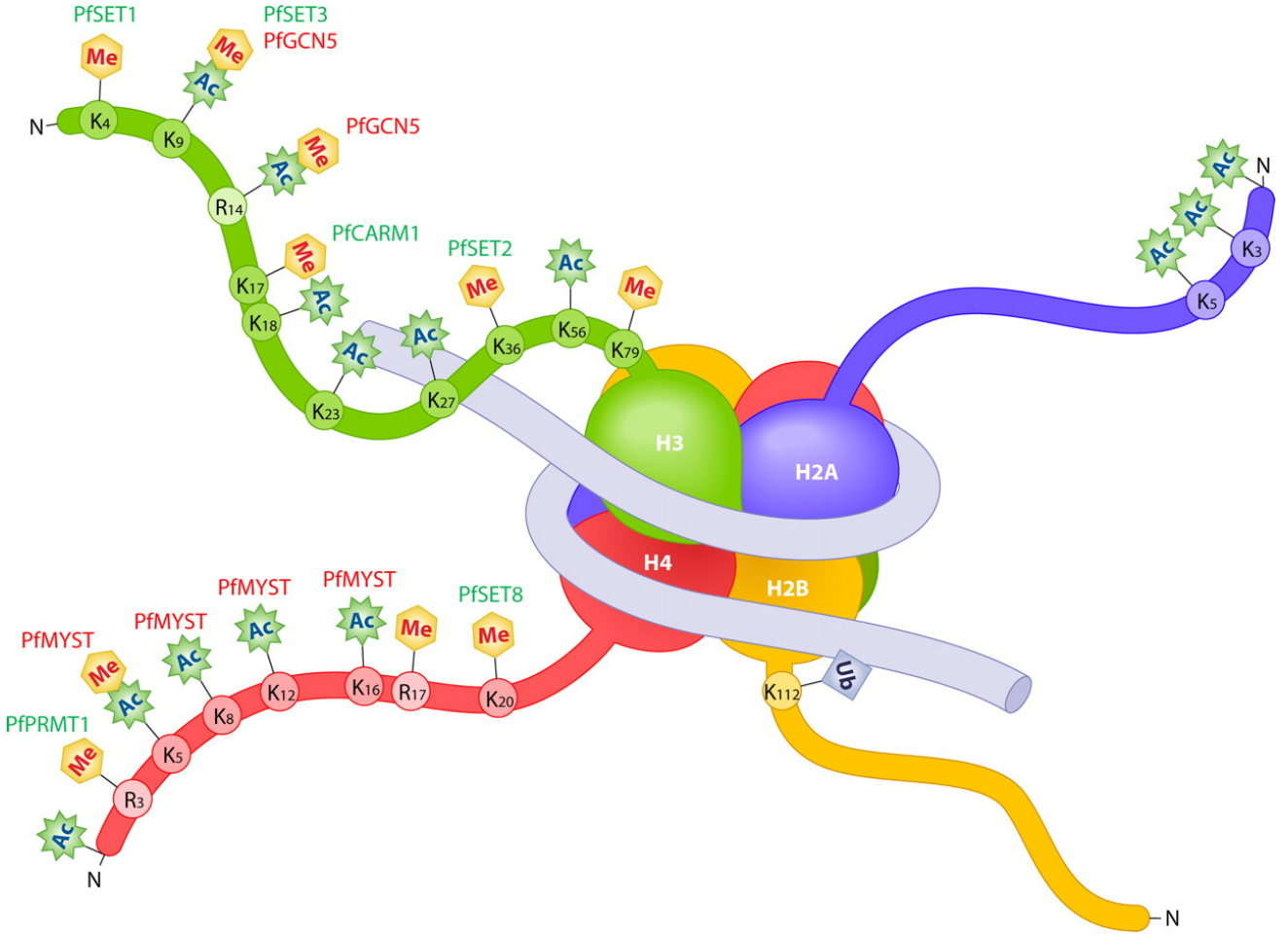
Histone modifications



Bansal (2017) *Pediatric Diabetes*

Histone modifications

- **Histone code**: post-translational modifications of histone N-ends (Lys, Arg, Cys) by phosphorylation, acetylation, methylation and ubiquitylation.
- These changes regulate gene expression by modulating the access of regulatory factors to the DNA



themysteryofepigenetics.blogspot.com

Histone modifications

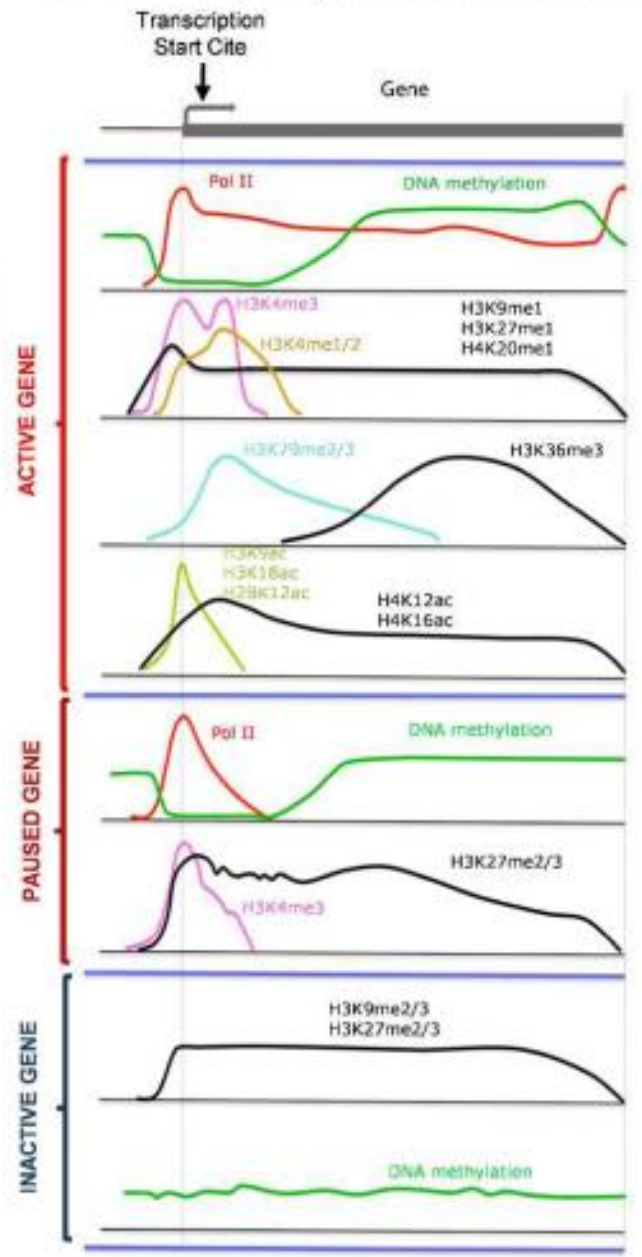
Table 1. The histone code.

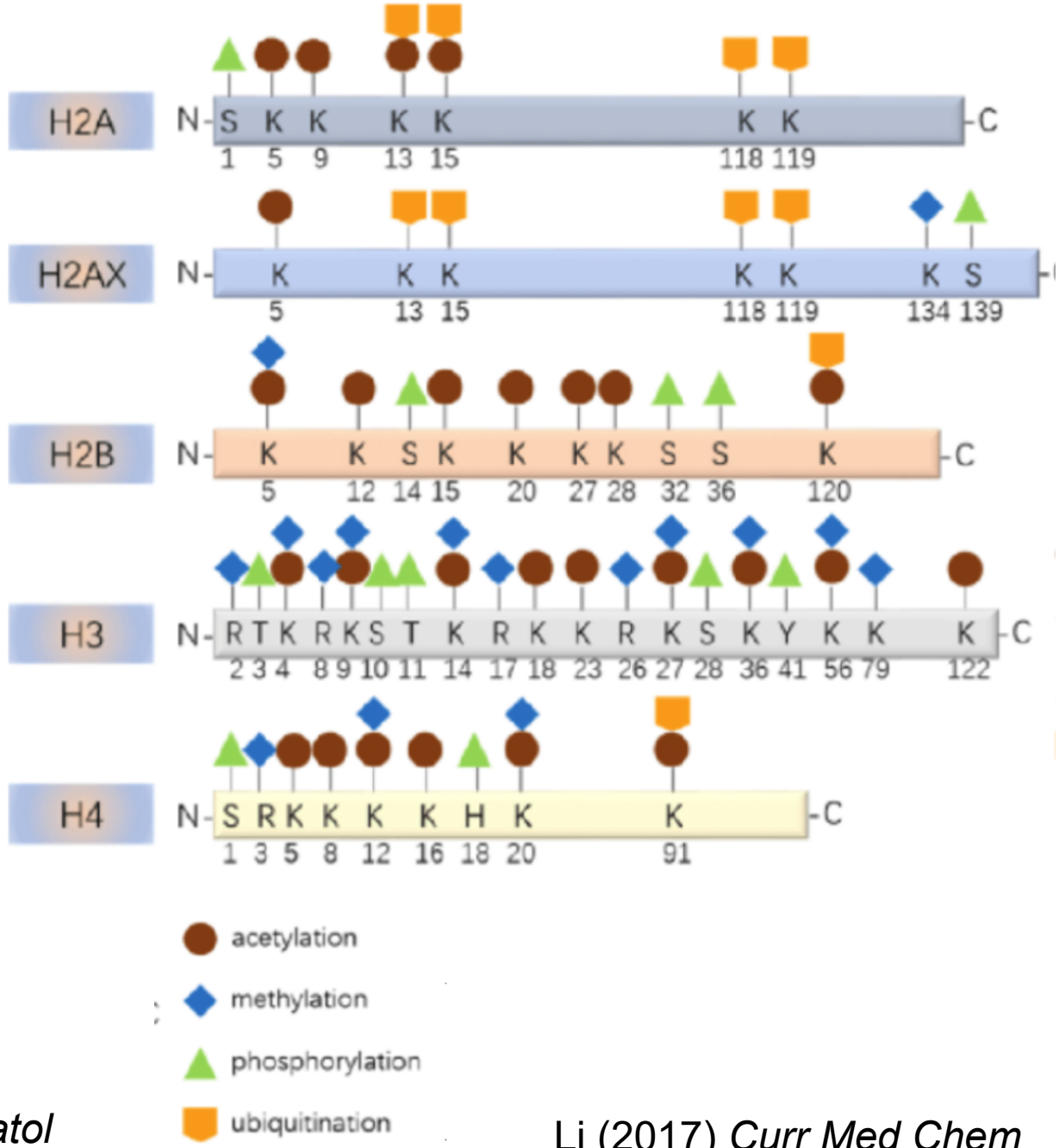
Histone code	Methylation			Acetylation	Ubiquitination
	Monomethylation	Dimethylation	Trimethylation		
H2AK119	–	–	–	–	Repression
H2BK5	Activation	–	Repression	–	–
H3K4	Activation	Activation	Activation	–	–
H3K9	Activation	Repression	Repression	Activation	–
H3K14	–	–	–	Activation	–
H3K18	–	–	–	Activation	–
H3K27	Activation	Repression	Repression	Activation	–
H3K36	Repression	Activation	Activation	–	–
H3K56	–	–	–	Activation	–
H3K79	Activation	Activation	Activation, repression	–	–
H4K12	–	–	–	Activation	–
H4K20	Activation		Repression	–	–

For each post-translational modification, the known functional association on gene transcription is shown. By reading the combinatorial and/or sequential histone modifications that constitute the histone code, it may be possible to predict which gene products will be transcribed. However, this code is controversial, since some gene loci present marks both associated with transcriptional activation and linked with repression. These bivalent domains are posited to be poised for either up- or down-regulation and to provide an epigenetic blueprint for lineage determination, and are usually found in stem cells.

Bauge (2014) *Future Med Chem*

Histone modifications



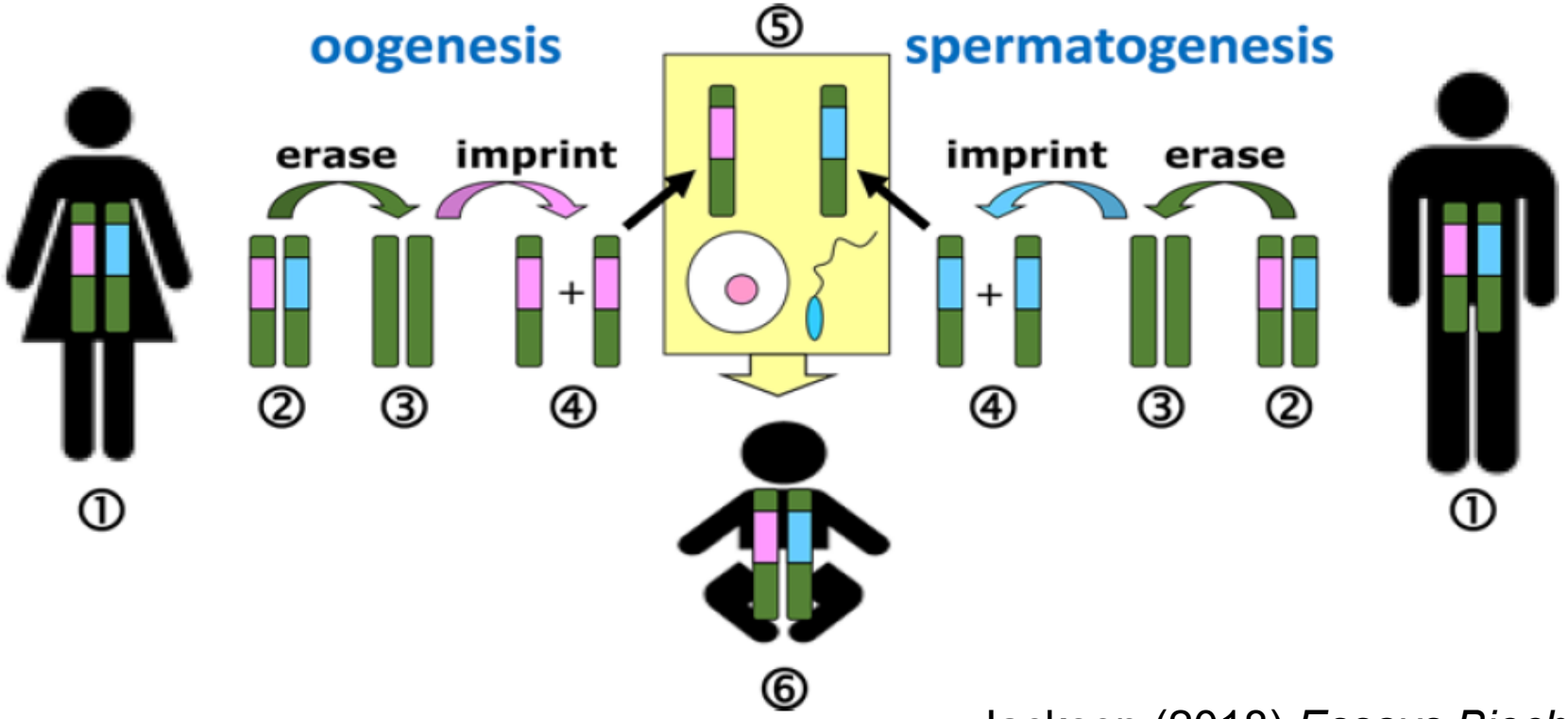


Botchkarev (2012) *J Invest Dermatol*

Li (2017) *Curr Med Chem*

Chromosomal imprinting

- **Chromosomal imprinting**: ~100 genes on various chromosomes, one copy is inactive by epigenetic mechanisms depending upon parent of origin
- For some genes (~70) only the paternal allele is active, while the maternal copy is epigenetically silenced throughout the life of the individual, and vice versa (~30 genes)
- Mutations in an active copy of a gene result in **imprinting disorders**

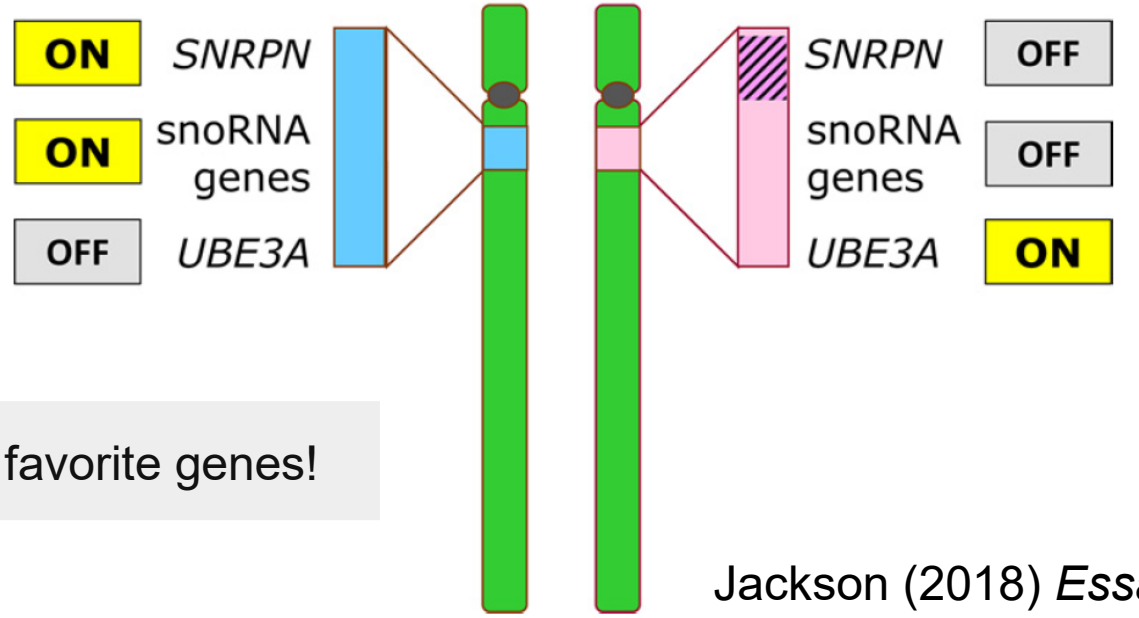


Jackson (2018) *Essays Biochem*

Chromosomal imprinting

Gene	Aliases	Location	Status	Expressed Allele
MAGEL2	nM15, NDNL1	15q11-q12 *AS*	**Imprinted**	Paternal
MKRN3	D15S9, RNF63, ZFP127, ZNF127, MGC88288	15q11-q13	**Imprinted**	Paternal
UBE3A	AS, ANCR, E6-AP, HPVE6A, EPVE6AP, FLJ26981	15q11-q13 *AS*	**Imprinted**	Maternal
NPAP1	C15orf2	15q11-q13	**Imprinted**	Unknown
ZNF127AS	MKRN3AS, Znp127as	15q11-q13	**Unknown**	Unknown
SNORD109A	HBII-438A	15q11.2	**Imprinted**	Paternal
SNORD108	HBII-437, HBII-437 C/D box snoRNA	15q11.2	**Imprinted**	Paternal
SNORD107	HBII-436, HBII-436 C/D box snoRNA	15q11.2	**Imprinted**	Paternal
SNORD109B	HBII-438B, HBII-438B C/D box snoRNA	15q11.2	**Imprinted**	Paternal
ATP10A	ATPVA, ATPVC, ATP10C, KIAA0566	15q11.2 *AS*	**Imprinted**	Maternal
SNRPN	SMN, PWCR, SM-D, RT-LI, HCERN3, SNRNP-N, FLJ33569, FLJ36996, FLJ39265, MGC29886, SNURF-SNRPN, DKFZp762N022, DKFZp686C0927, DKFZp761I1912, DKFZp686M12165	15q11.2	**Imprinted**	Paternal

http://www.geneimprint.com/site/genes-by-species



Exercise: check your favorite genes!

Jackson (2018) *Essays Biochem*

Imprinting disorders

	Angelman syndrome	*Prader-Willi syndrome*
Key features	* Moderate to severe ID (IQ ~25–54) * Jerky, puppet-like movements * Happy and sociable disposition * Seizures	* Mild to moderate ID (IQ ~60–70) * Insatiable appetite leading to morbid obesity * Behaviour problems
Frequency in the population	~1/20,000	~1/15,000
Underlying genetic abnormality (in some cases, the underlying cause has not been determined)	– Maternal 15q11.2 deletion (~70%) – Paternal UPD (~4%) – Imprinting defect (~8%) – Pathogenic variant in UBE3A (~6%)	– Paternal 15q11.2 deletion (~70%) – Maternal UPD (~20%) – Imprinting defect (~5%)
Key genes	*UBE3A* encoding a ubiquitin ligase	*SNORD116* gene cluster encoding snoRNAs (other genes in the imprinted region may also influence the phenotype)

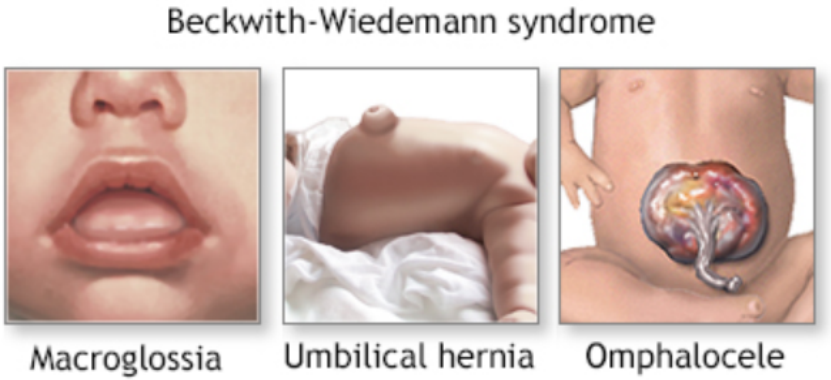
Jackson (2018) *Essays Biochem*

Imprinting disorders

- IGF2 is a hormone that stimulates growth during embryonic and fetal development // not the IGF2 receptor gene!
- Normally maternally silenced in humans
- **Epimutation** (missing methyl tags) can result in two active copies

Activation of the maternal *IGF2* gene during egg formation or very early in development causes **Beckwith-Wiedemann Syndrome (BWS):**

– overgrowth
– an increased risk of cancer, especially during childhood
– variety of other symptoms



Frequency: ~15,000 births. However, in babies that were conceived in the laboratory with the help of artificial reproductive technology, the rate of BWS may be as high as 1/4,000.

https://learn.genetics.utah.edu/content/epigenetics/imprinting

Non-coding RNAs in the genome

Name	Size	Location	Number in humans	Functions	Illustrative examples
Short ncRNAs					
miRNAs	19–24 bp	Encoded at widespread locations	>1,424	Targeting of mRNAs and many others	miR-15/16, miR-124a, miR-34b/c, miR-200
piRNAs	26–31bp	Clusters, intragenic	23,439	Transposon repression, DNA methylation	piRNAs targeting *RASGRF1* and LINE1 and IAP elements
tiRNAs	17–18bp	Downstream of TSSs	>5,000	Regulation of transcription?	Associated with the *CAP1* gene
Mid-size ncRNAs					
snoRNAs	60–300 bp	Intronic	>300	rRNA modifications	U50, SNORD
PASRs	22–200 bp	5′ regions of protein-coding genes	>10,000	Unknown	Half of protein-coding genes
TSSa-RNAs	20–90 bp	–250 and +50 bp of TSSs	>10,000	Maintenance of transcription?	Associated with *RNF12* and *CCDC52* genes
PROMPTs	<200 bp	–205 bp and –5 kb of TSSs	Unknown	Activation of transcription?	Associated with *EXT1* and *RBM39* genes
Long ncRNAs					
lincRNAs	>200 bp	Widespread loci	>1,000	Examples include scaffold DNA–chromatin complexes	*HOTAIR*, *HOTTIP*, *lincRNA-p21*
T-UCRs	>200 bp	Widespread loci	>350	Regulation of miRNA and mRNA levels?	uc.283+, uc.338, uc160+
Other lncRNAs	>200 bp	Widespread loci	>3,000	Examples include X-chromosome inactivation, telomere regulation, imprinting	*XIST*, *TSIX*, TERRAs, *p15AS*, *H19*, *HYMAI*

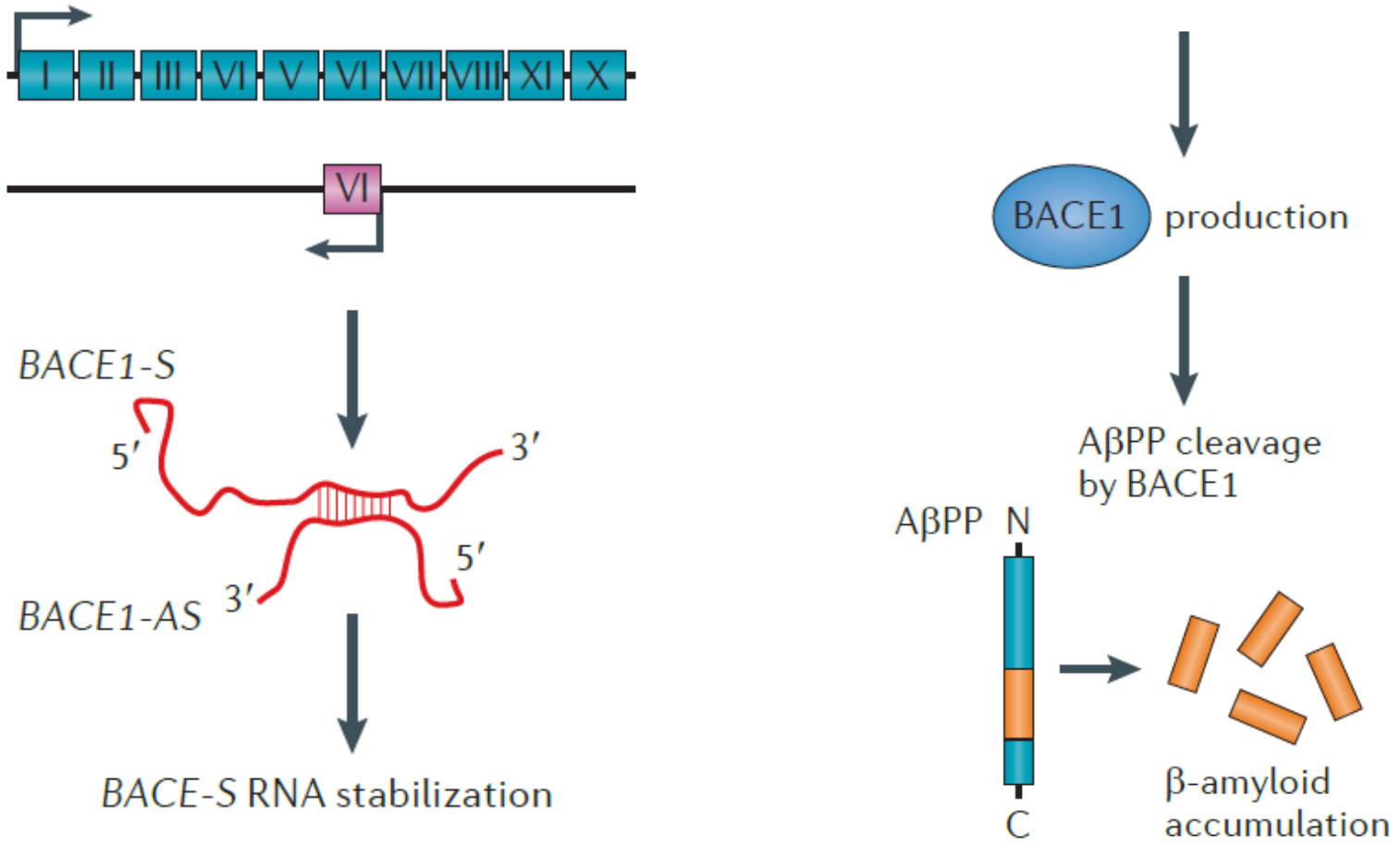
Esteller (2011) *Nat Rev Genet*

Non-coding RNAs in non-cancer disease

Disease	Involved ncRNAs	ncRNA type
Spinal motor neuron disease	miR-9	miRNA
Spinocerebellar ataxia type 1	miR-19, miR-101, miR-100	miRNA
Amyotropic lateral sclerosis	miR-206	miRNA
Arrhytmia and hypertension	miR-1	miRNA
Atheromatosis and atherosclerosis	miR-10a, miR-145, mR-143 and miR-126	miRNA
Atheromatosis and atherosclerosis	Circular ncRNA linked to the CDKN2A locus	lncRNA
Cardiac hypertrophy	miR-21	miRNA
Rett's syndrome	miR-146a, miR-146b, miR-29 and miR-382	miRNA
5q syndrome	miR-145 and miR-146a	miRNA
ICF syndrome	miR-34b, miR-34c, miR-99b, let-7e and miR-125a	miRNA
Crohn's disease	miR-196	miRNA
Prader–Willi and Angelman syndromes	snoRNA cluster at 15q11–q13 imprinted locus	snoRNA
Beckwith–Wiedeman syndrome	lncRNAs *H19* and *KCNQ1OT1*	lncRNA
Uniparental disomy 14	snoRNA cluster at 14q32.2 imprinted locus	snoRNA
Silver–Russell syndrome	lncRNA *H19*	lncRNA
Silver–Russell syndrome	miR-675	miRNA
McCune–Albright syndrome	lncRNA *NESP-AS*	lncRNA
Deafness	miR-96	miRNA
Alzheimer's disease	miR-29, miR-146 and miR-107	miRNA
Alzheimer's disease	ncRNA antisense transcript for *BACE1*	lncRNA

Esteller (2011) *Nat Rev Genet*

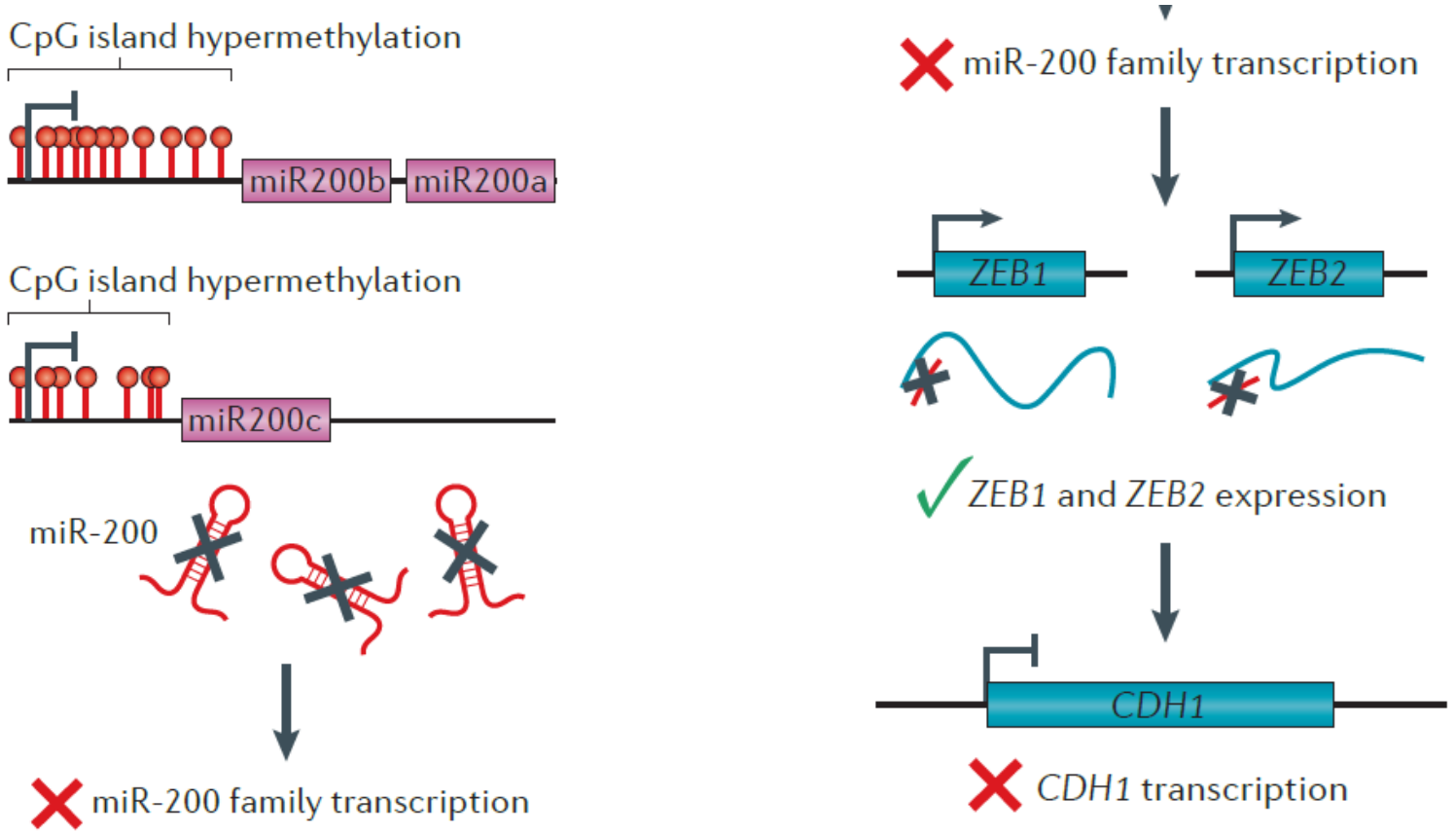
Non-coding RNAs in Alzheimer's disease



An antisense lncRNA, *BACE1-AS*, regulates the expression of the sense *BACE1* gene (labelled *BACE1-S* in the figure) through the stabilization of its mRNA. *BACE1-AS* is elevated in Alzheimer's disease, increasing the amount of BACE1 protein and, subsequently, the production of β-amyloid peptide.

Esteller (2011) *Nat Rev Genet*

Non-coding RNAs in cancer



Alterations in the epigenetic regulation of the miR-200 family are involved in epithelial-to-mesenchymal transition in cancer. Specifically, CpG island hypermethylation-associated silencing of these miRNAs in human tumours causes an upregulation of the zinc finger E-box-binding homeobox (HOX) 1 (*ZEB1*) and *ZEB2* transcriptional repressors, which, in turn, leads to a downregulation of E-cadherin *CDH1*

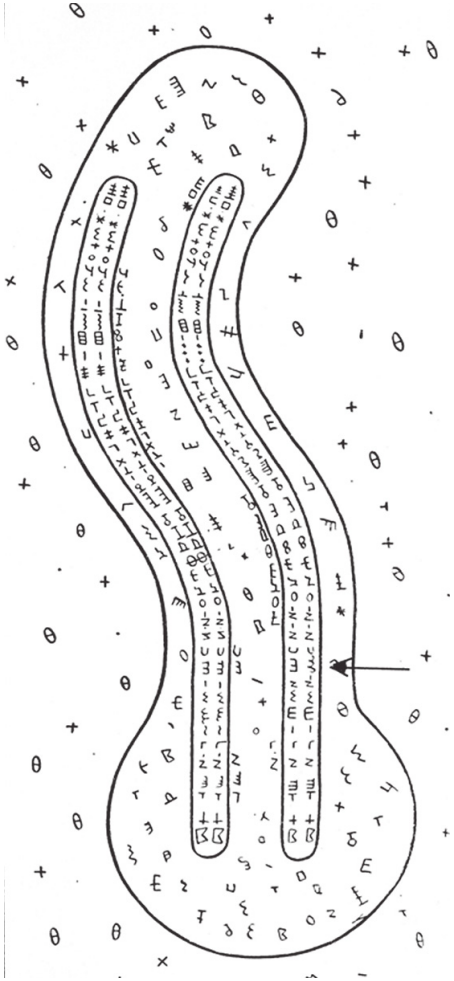
Esteller (2011) *Nat Rev Genet*

Epigenetic effects of smoking

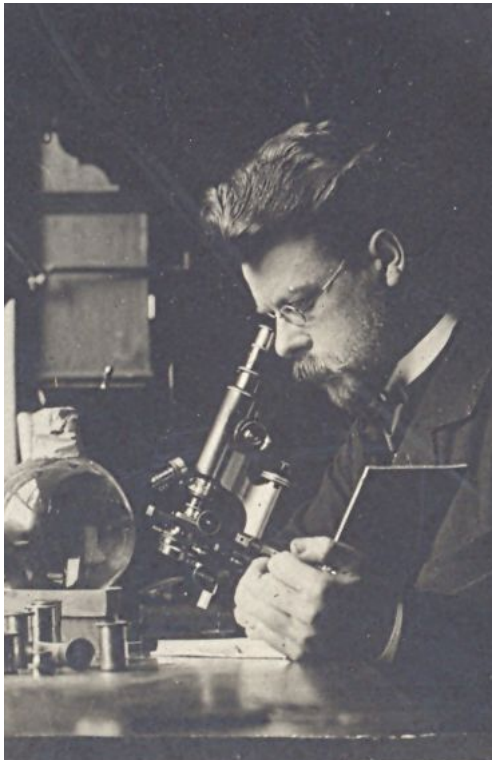
From Wikipedia, the free encyclopedia

Contents [hide]

Николай Конст. Кольцов (1872-1940)



- 1915: «Следует признать гены способными... к мутациям. Ведь во всяком органическом соединении атом водорода может быть скачкообразно заменен группой CH_3»
- 1927: *Omnis molecula ex molecula:* гипотеза о матричном воспроизведении молекул наследственности



Кольцов 1927 → Тимофеев-Ресовский, Циммер, Дельбрюк, Шредингер 1935-1945 → Уотсон, Крик 1953