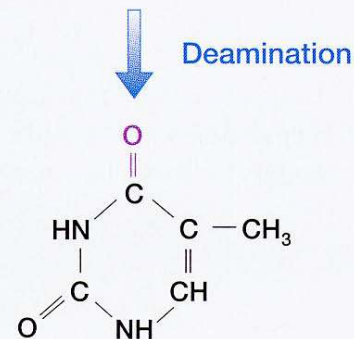
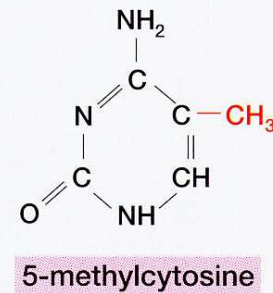
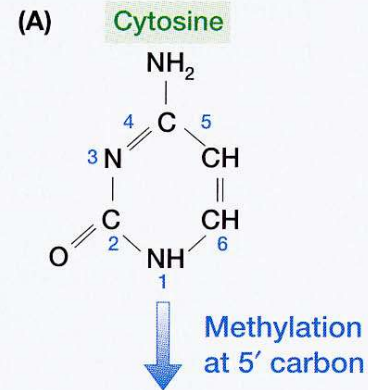
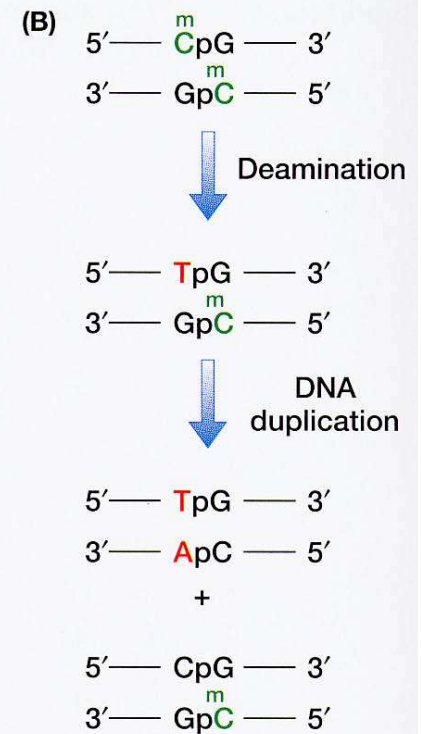
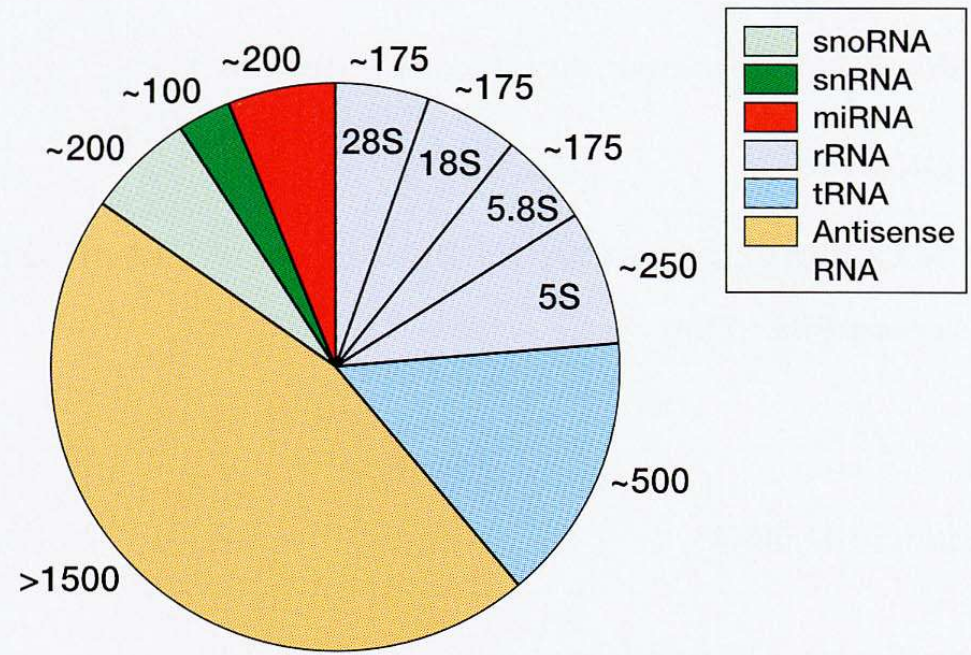


CpG Islands



Thymine
(forms mismatch with G;
inefficiently recognized by
DNA repair system)



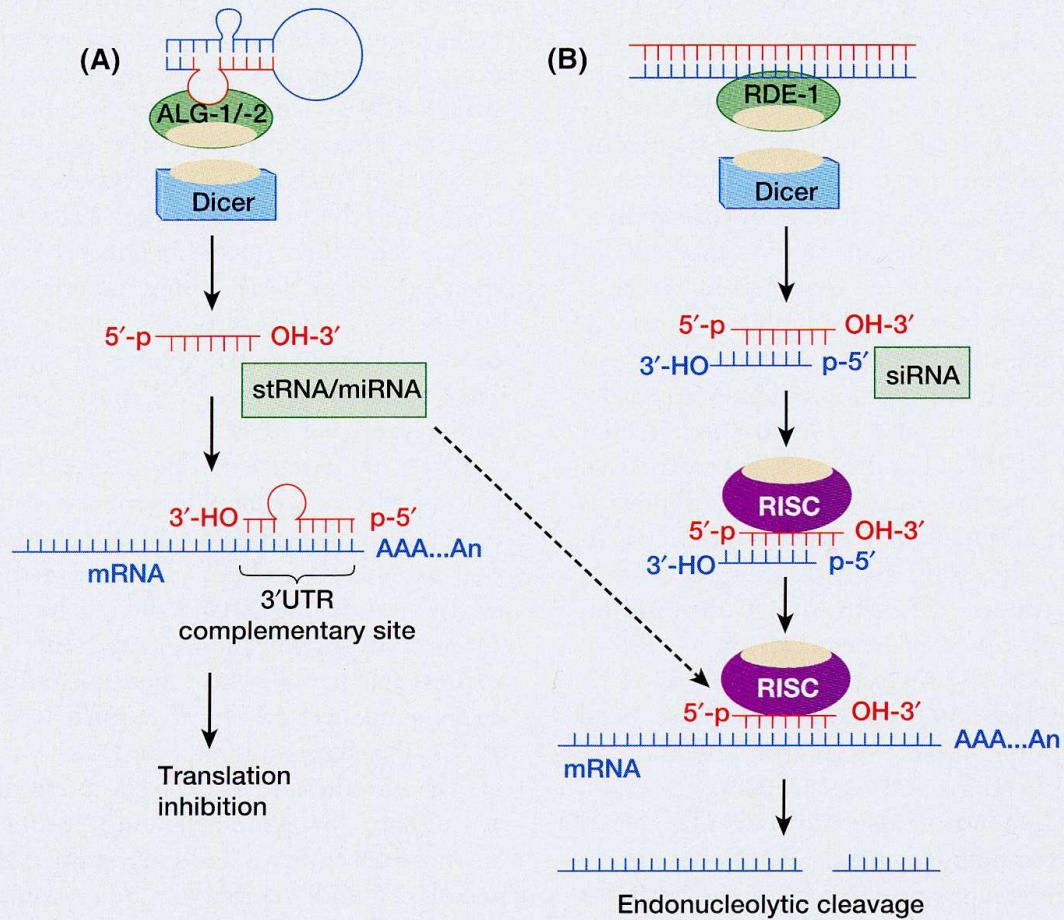


RNA Klassen: Hauptklassen

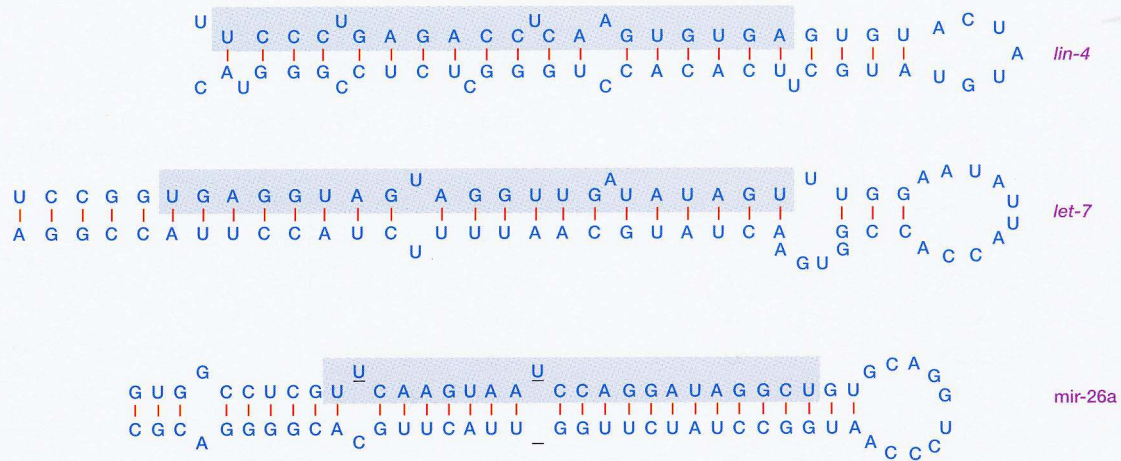
- Ribosomale RNA (rRNA)
- Transfer RNA (tRNA)
- Small nuclear RNA (snRNA)
- Small nucleolar RNA (snoRNA)

Andere RNA Klassen

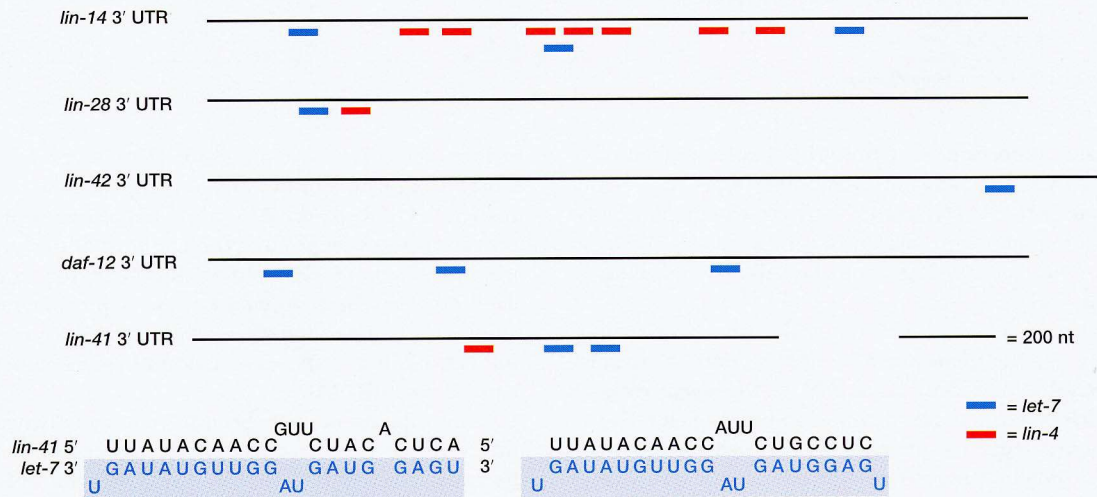
- Micro RNA
- X-chromosome inactivation associated
- Imprinting associated
- Nervous system specific
- Antisense RNA
- Andere



(A)

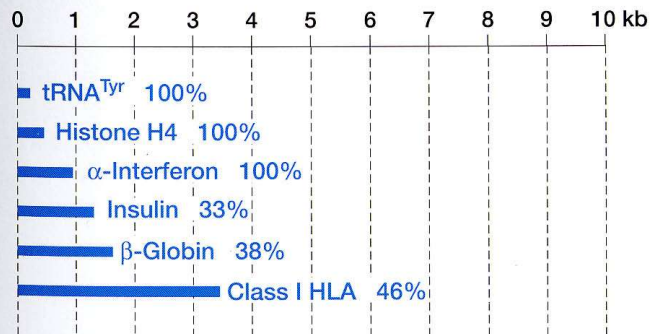


(B)

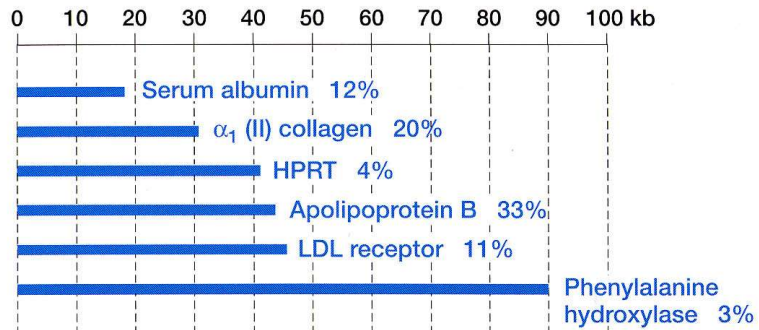


Proteinkodierende Gene

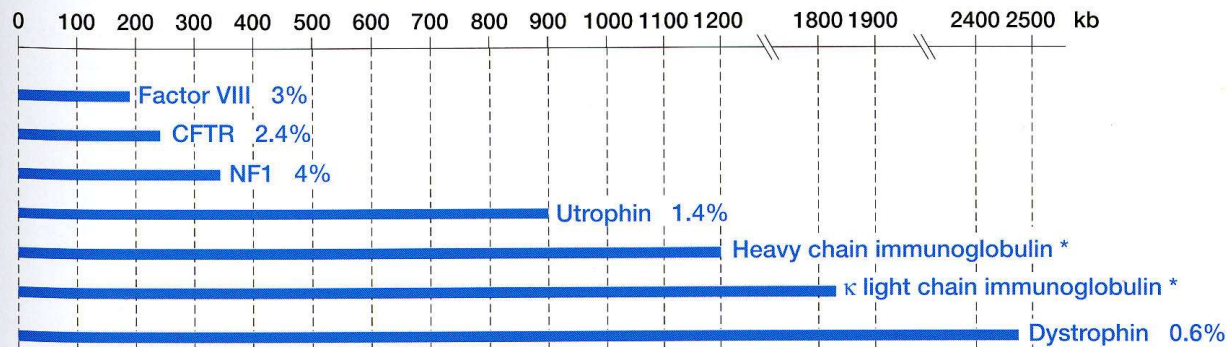
(A) Less than 10 kb



(B) Less than 100 kb



(C) More than 100 kb



Beispiele für humane Gene mit ununterbrochener Coding sequence

- 37 mitochondriale Gene
- Viele RNA Gene
- Retrogene
- Interferone
- Histone
- Viele Ribonuclease Gene
- Heat shock Proteine.....

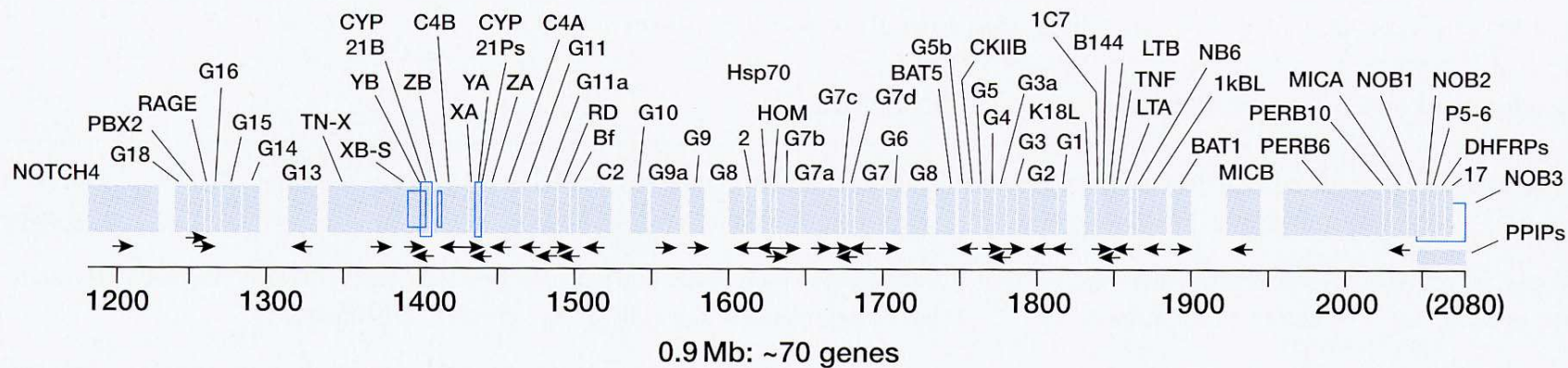
Verteilung der Gene im Genom

- Funktionell identische Gene (Histone, Ubiquitin)
- Funktionell ähnliche Gene (z.B. Hox Gene)
- Funktionell verwandte Gene

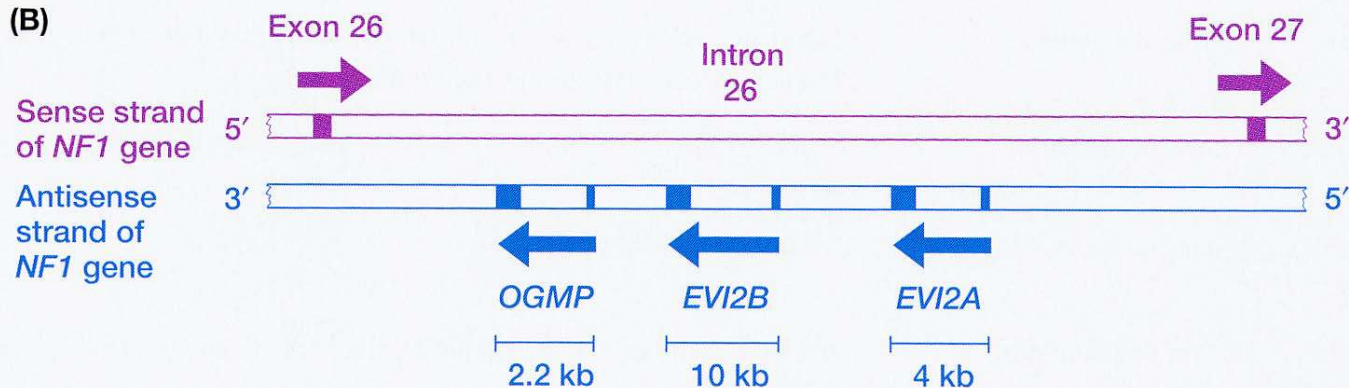
Organisation von Genen

- Überlappende Gene in Säugern selten
- Genes within genes: z.B. snoRNA, NF1, Rb1
- Polycistronische Transkriptionseinheit (z.B. Mitochondrium)
- Bicistronische Transkriptionseinheit (z.B. InsulinA und B, Ubiquitin und ribosomale Proteine)

(A)



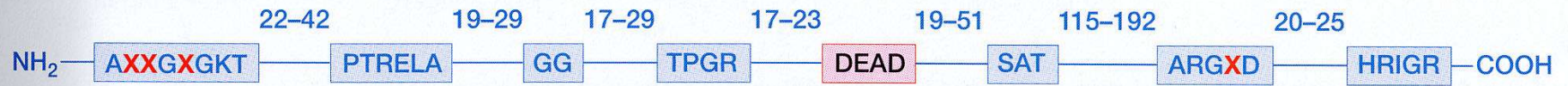
(B)



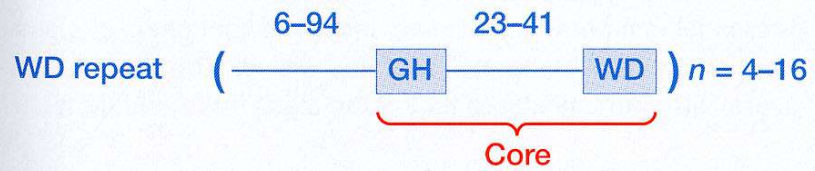
Genfamilien

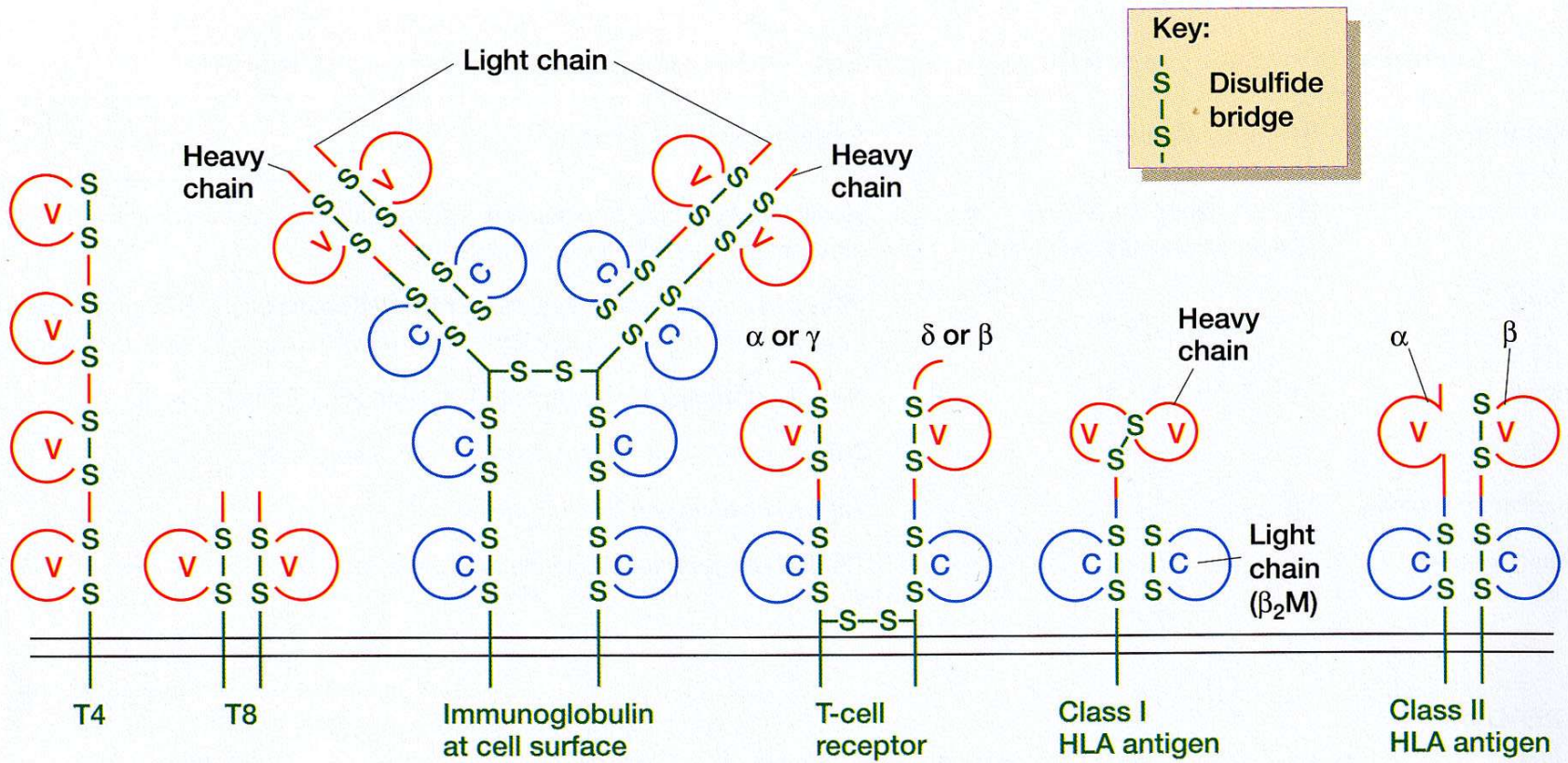
- Klassische Genfamilien
- Genfamilien mit großen, hochkonservierten Domänen
- Genfamilien mit kurzen konservierten Aminosäure Motiven
- Gensuperfamilien

(A) DEAD box



(B)





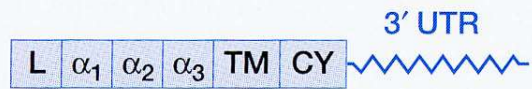
Organisation der Genfamilien

- Single cluster
 - Tandem organization
 - Close clustering
 - Compound cluster
- Multiple gene cluster
 - Olfactory receptor gene family
- Dispersed genfamilies
 - Different genomes
 - Ancient duplication events
 - retrotransposition

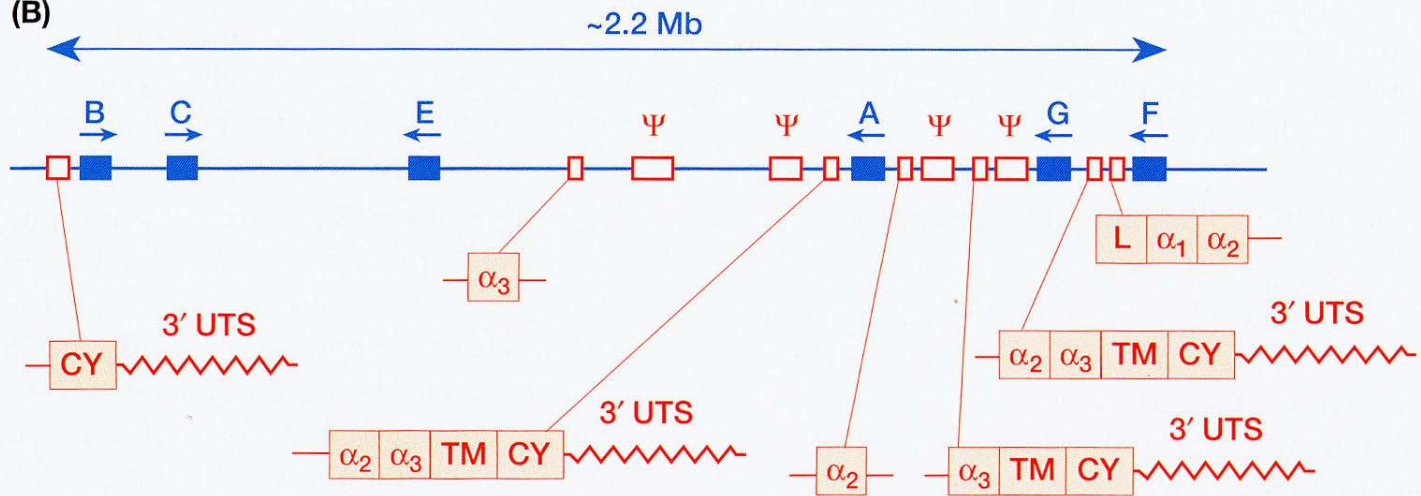
Pseudogene

- Defekte Kopie
- Processed und non processed
- Retrogene
 - Polypeptid kodierend
 - RNA Gen

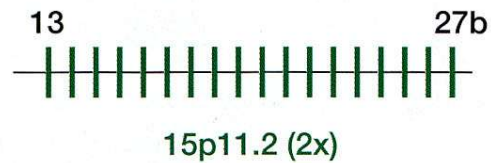
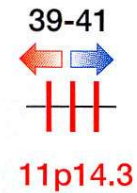
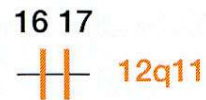
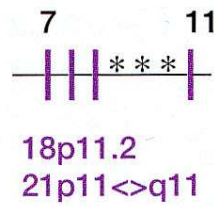
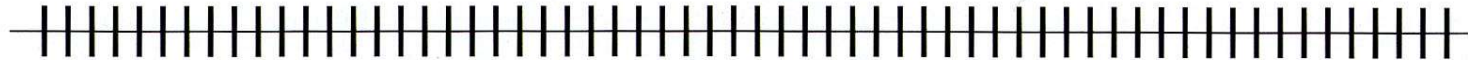
(A)

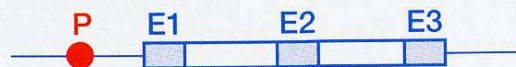


(B)



NF1 gene 17q11.2; 60 exons

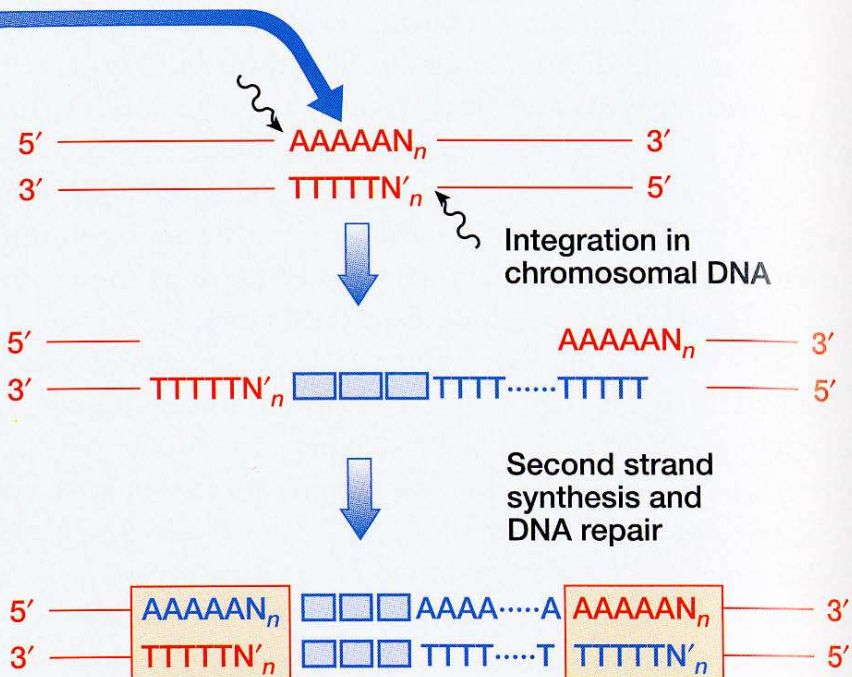
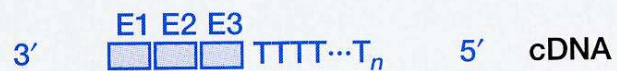




Transcription and
RNA processing

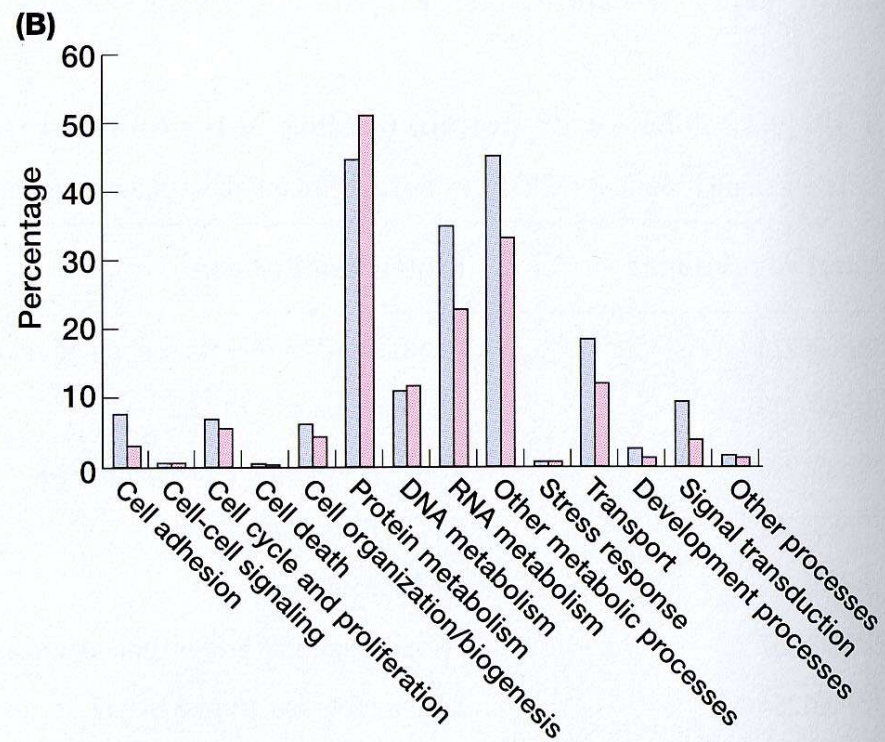
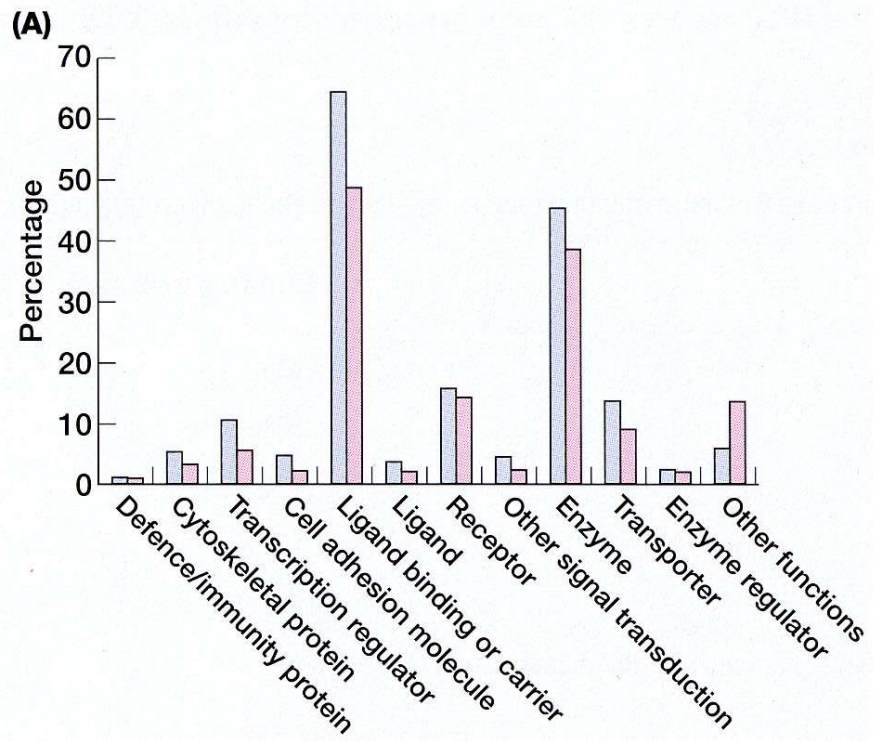


Reverse
transcriptase



Proteome Klassifikation

- Datenbanken
 - Interpro (EBI)
 - Pfam Wellcome Trust Sanger Institute
- Kategorien
 - Proteinfamilien
 - Proteindomänen
 - Proteinrepeats
- GO



Repetitive DNA

- Satellite DNA
- Minisatellite DNA
- Microsatellite DNA

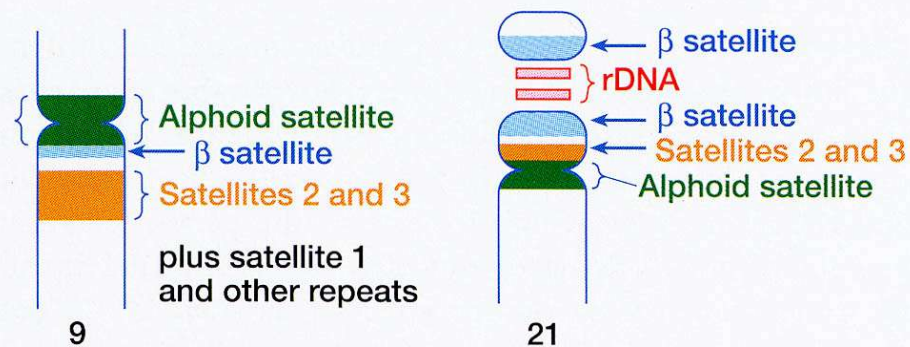


Table 9.14: Major classes of tandemly repeated human DNA

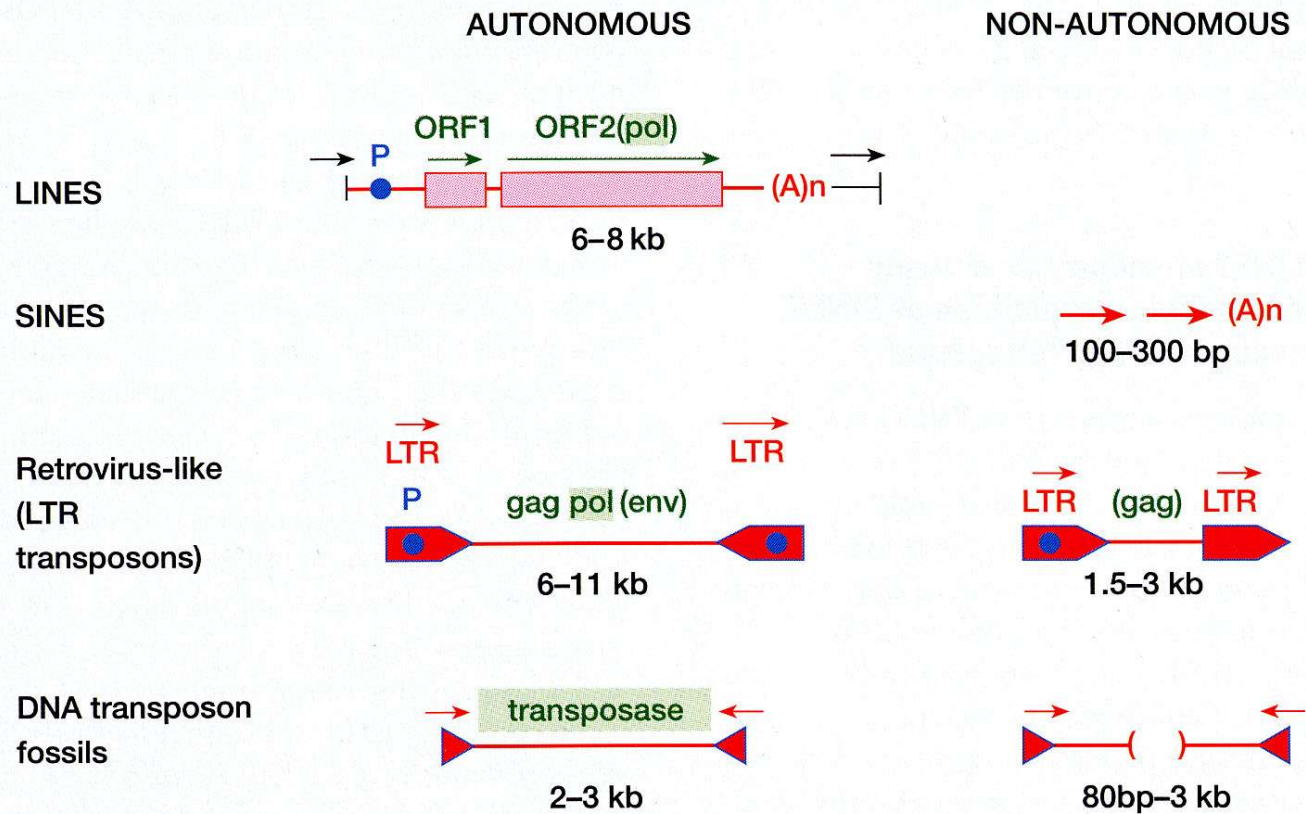
Class	Size of repeat unit (bp)	Major chromosomal location(s); transcriptional status
Satellite DNA (arrays often within 100 kb to several Mb size range)	5–171	Especially at centromeres; not transcribed
α (alphoid DNA)	171	Centromeric heterochromatin of all chromosomes
β (<i>Sau3A</i> family)	68	Notably the centromeric heterochromatin of 1, 9, 13, 14, 15, 21, 22 and Y
Satellite 1 (AT-rich)	25–48	Centromeric heterochromatin of most chromosomes and other heterochromatic regions
Satellites 2 and 3	5	Most, possibly all, chromosomes
Minisatellite DNA (arrays often within the 0.1–20 kb range)	9–64	At or close to telomeres of all chromosomes; vast majority not transcribed
Telomeric family	6	All telomeres
Hypervariable family	9–64	All chromosomes, often near telomeres
Microsatellite DNA (= simple sequence repeats, SSR) (arrays typically < 100 bp)	12	Dispersed throughout all chromosomes; some small arrays of very simple sequence

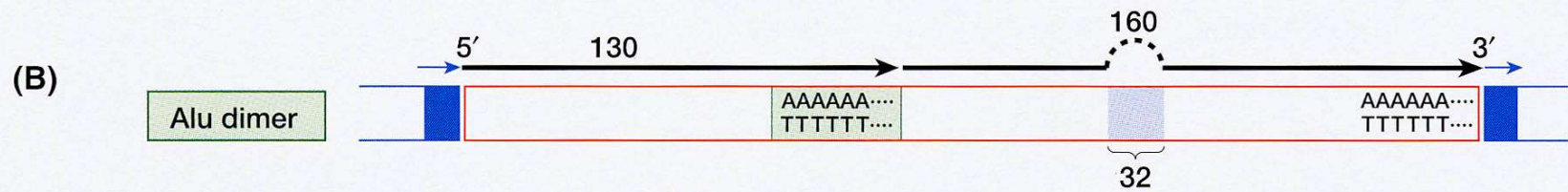
Interspersed Repetitive noncoding DNA

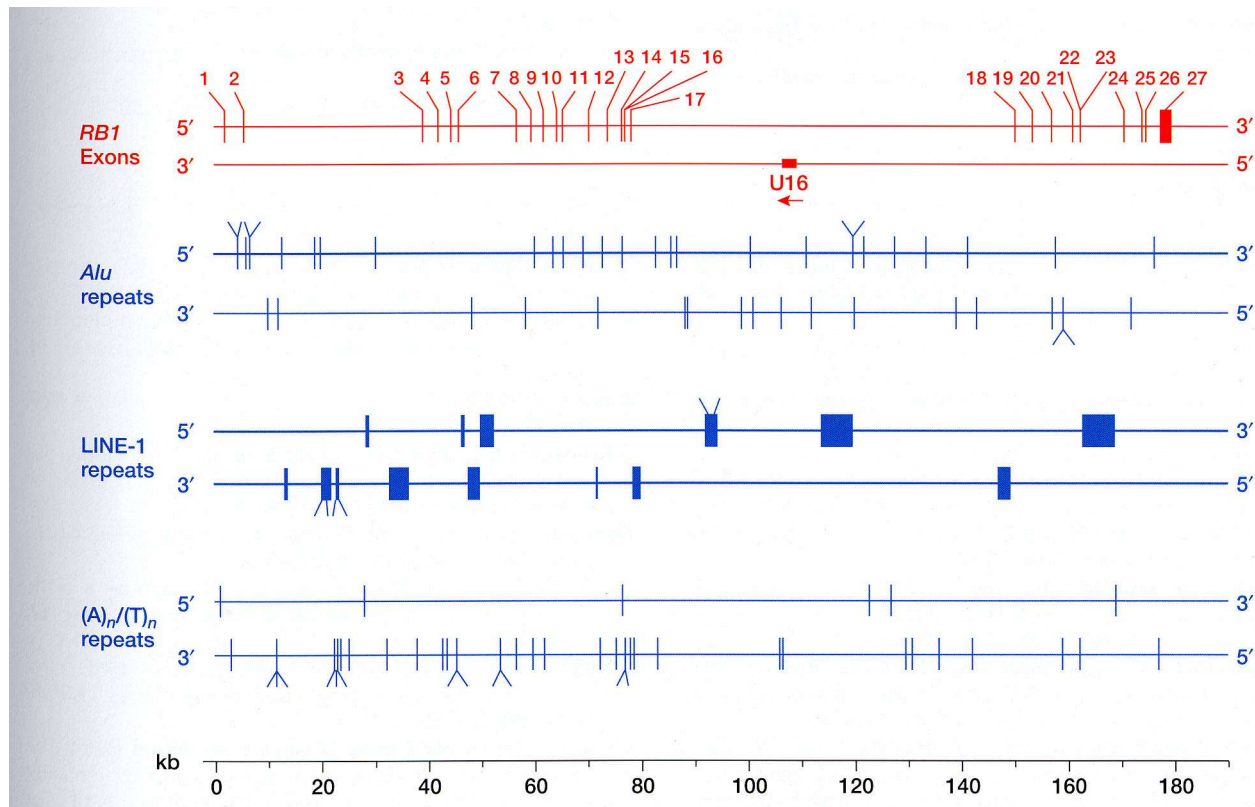
- Retrotransposons
- DNA Transposons

Vier Hauptklassen

- Human LTR transposons
- Human DNA transposons
- Line
- Sine



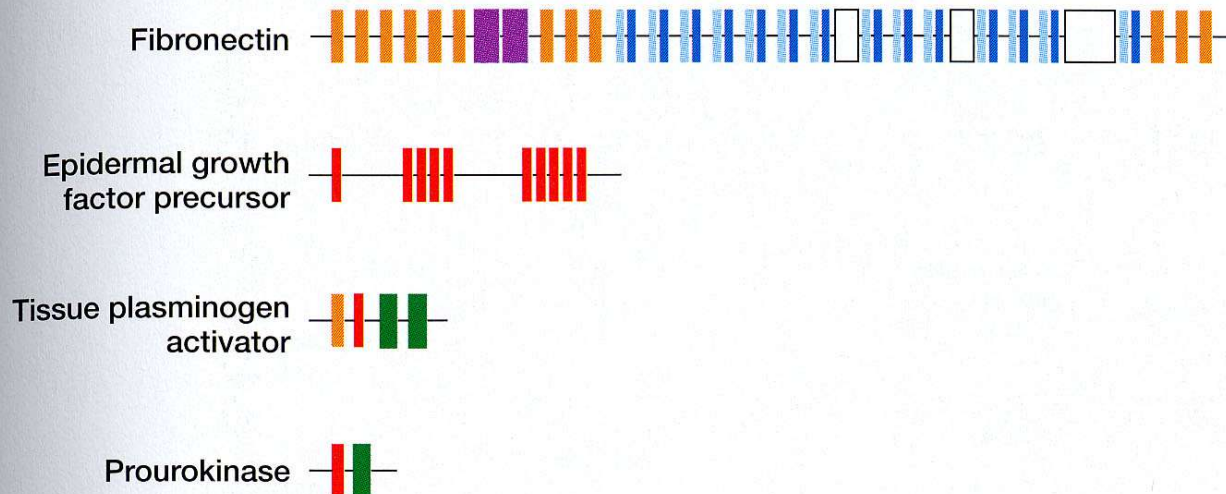




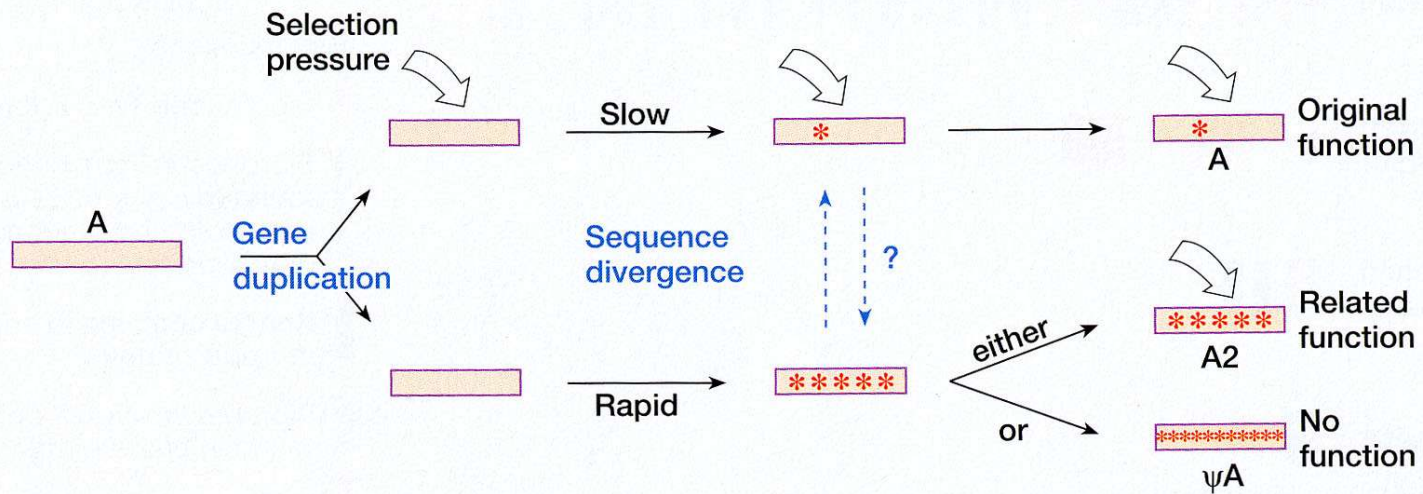
Genome size	~ 3200 Mb
Nuclear genome	~ 3200 Mb
Mitochondrial genome	37 kb
Euchromatic component	~ 2900–3000 Mb
Constitutive heterochromatin	> 200 Mb (<i>Table 9.2; Figure 2.15</i>)
Highly conserved fraction	> 100 Mb (>3%)
Coding DNA	~ 50 Mb (~1.5%)
Other (regulatory etc.)	~ 100 Mb (3%)
Segmentally duplicated DNA	>150 Mb (>5%)
Noncoding repetitive DNA	> 50% of genome
Transposon-based repeats	~ 1400 Mb (~ 43%; see <i>Table 9.15</i>)
Gene number	~ 30 000–35 000
Nuclear genome	~ 30 000–35 000 (<i>Section 9.1.3</i>).
Mitochondrial genome	37 (<i>Section 9.1.2</i>).
Per chromosome	Average of ~ 1400; but depends on chromosome length and type (see <i>Figure 8.4</i>); ~ 60 per band in a 550-band chromosome preparation
Polypeptide-encoding genes	~ 30 000 but considerable uncertainty
RNA genes	~ 3000, but some uncertainty (see <i>Figure 9.4</i>)
Pseudogenes	~20 000
Gene density	~ 1/100 kb in nuclear genome; 1/0.45 kb in mitochondrial genome
Gene size (genomic extent)	Average = 27 kb, but enormous variation (see <i>Figure 9.7</i>).
Intergenic distance	Average = ca. 75 kb in nuclear genome.
CpG island number	~ 30 000 (in genome sequence filtered to remove noncoding repeats)
Exon number	Average = 9. Generally correlated with gene length, but wide variation.
Largest number	363 (in the titin gene)
Smallest number	1 (that is, no introns – <i>Table 9.5</i>)
Exon size	Average = 122 bp for internal exons with comparatively little length variation, but 3' exons can be considerably longer (Zhang, 1998).
Largest exons	Many kb long, e.g. exon 26 of the apoB gene (<i>APOB</i>) is 7.6 kb
Smallest exons	< 10 bp
Intron size	Enormous variation; strong direct correlation with gene size (see <i>Table 9.6</i>): Hundreds of kb e.g. intron 8 of the human <i>WWOX</i> gene is ~ 800 kb.
Largest introns	Tens of bp
Smallest introns	
mRNA size	Average of about 2.6 kb, but considerable variation (titin mRNA is > 115 kb long!)
5' UTR	Average of about 0.2–0.3 kb
3' UTR	Average of about 0.77 kb but likely to be an underestimate because of under- reporting of long 3' UTRs
Noncoding RNA size	Highly variable; from ~ 21–22 nucleotides (microRNA) to many kb e.g. <i>XIST</i> (17 kb)
Polypeptide size	Average of about 500–550 amino acids
Largest polypeptide	Titin: 38 138 codons in titin gene (but significant length variation)
Smallest polypeptides	Tens of amino acids e.g. various small hormones etc.

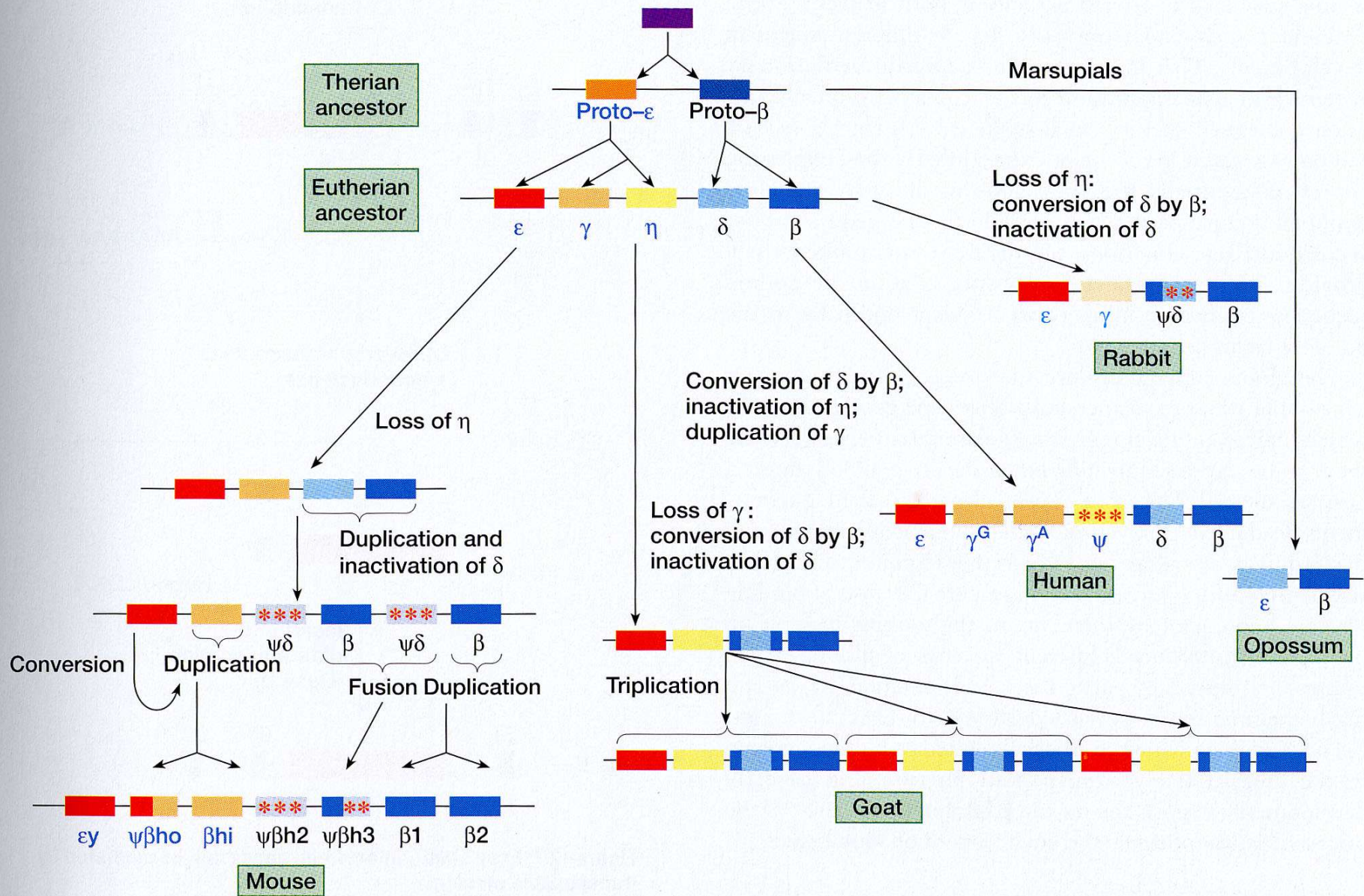
Introns

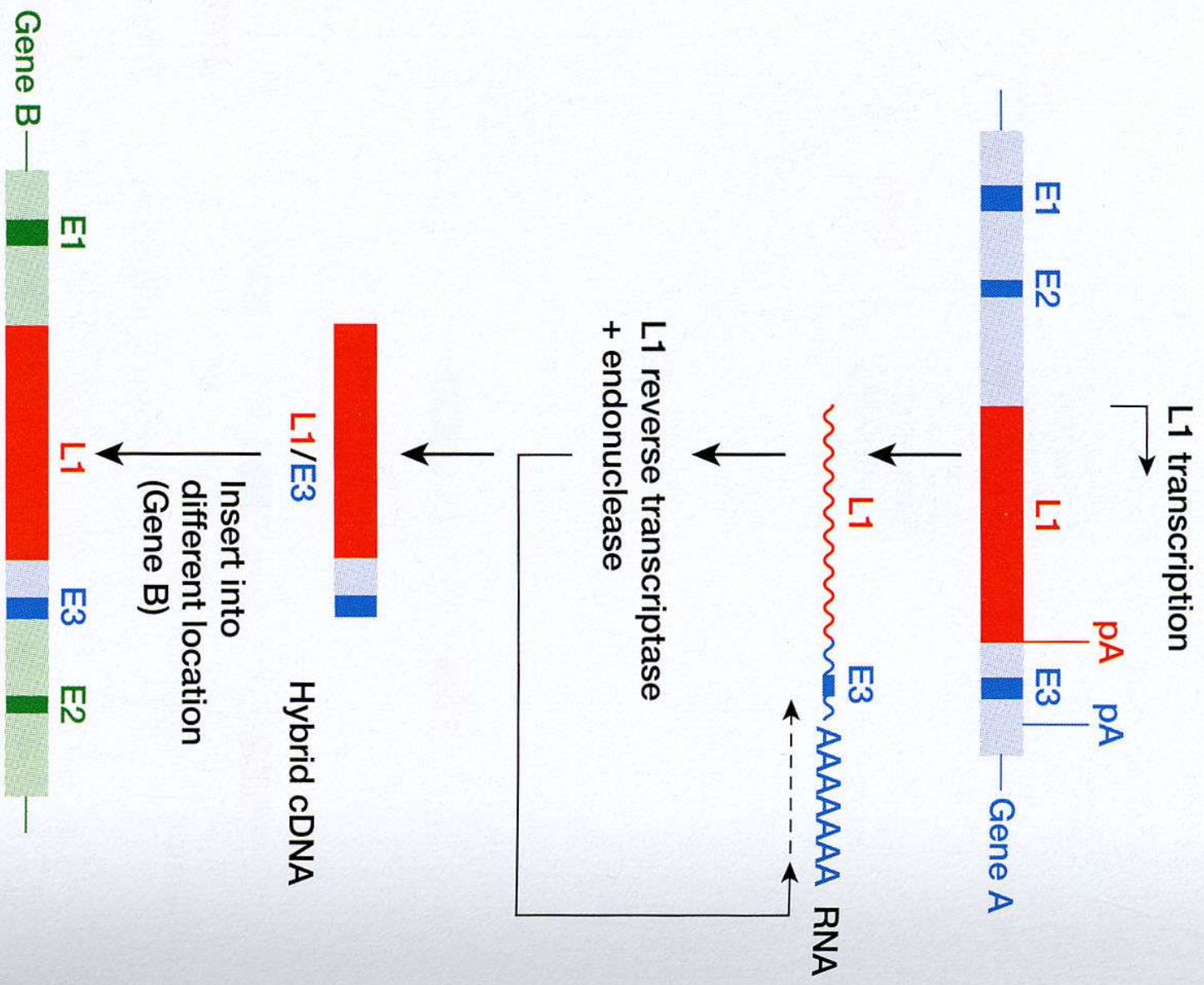
- Spliceosomal introns
- Group I und group II introns
- Archaeal introns

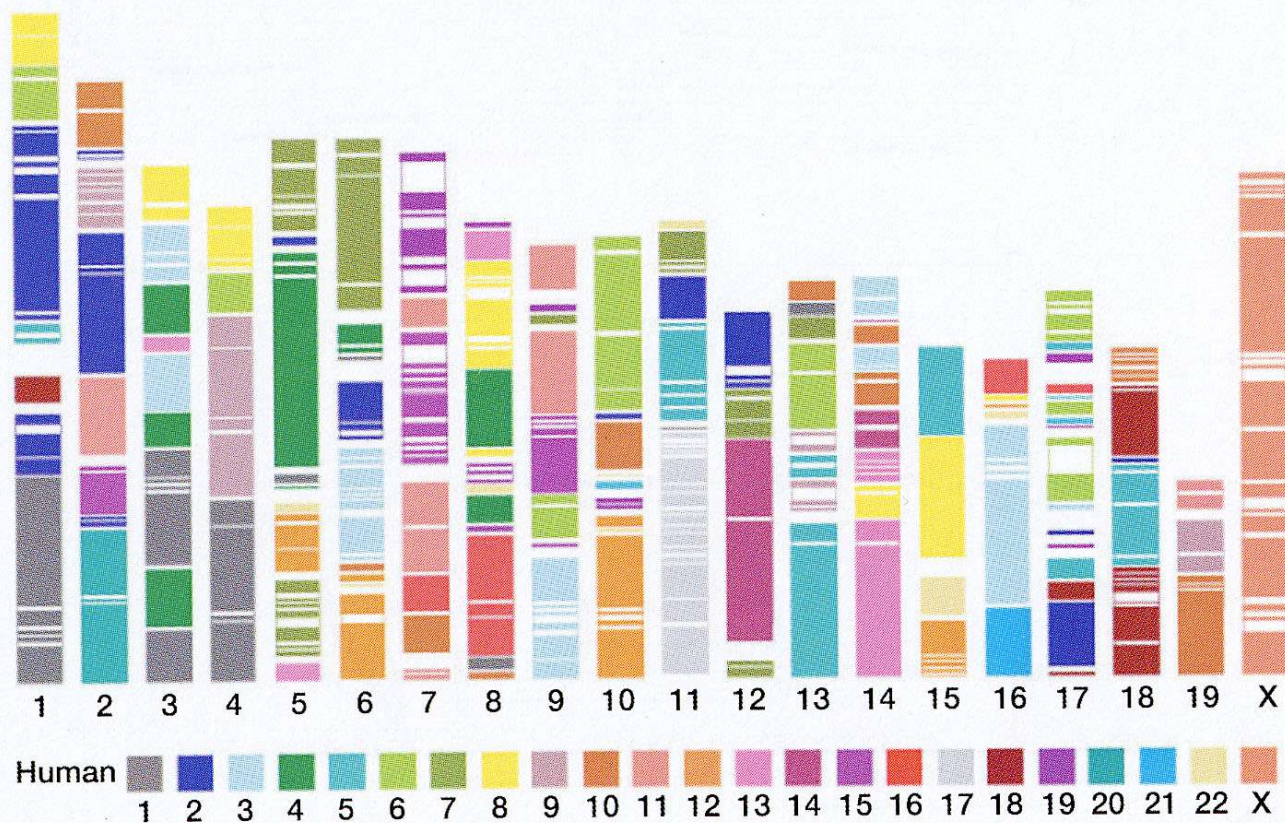


- Coding potential**
- Key for exons:**
- █ Fibronectin Type I domain
 - █ Fibronectin Type III domain found in many cell surface receptors and other extracellular matrix proteins
 - █ Kringle domain; 38 copies in ApoA protein
 - █ Fibronectin Type II domain also found in blood clotting factors
 - █ Epidermal growth factor domain
 - No known domain

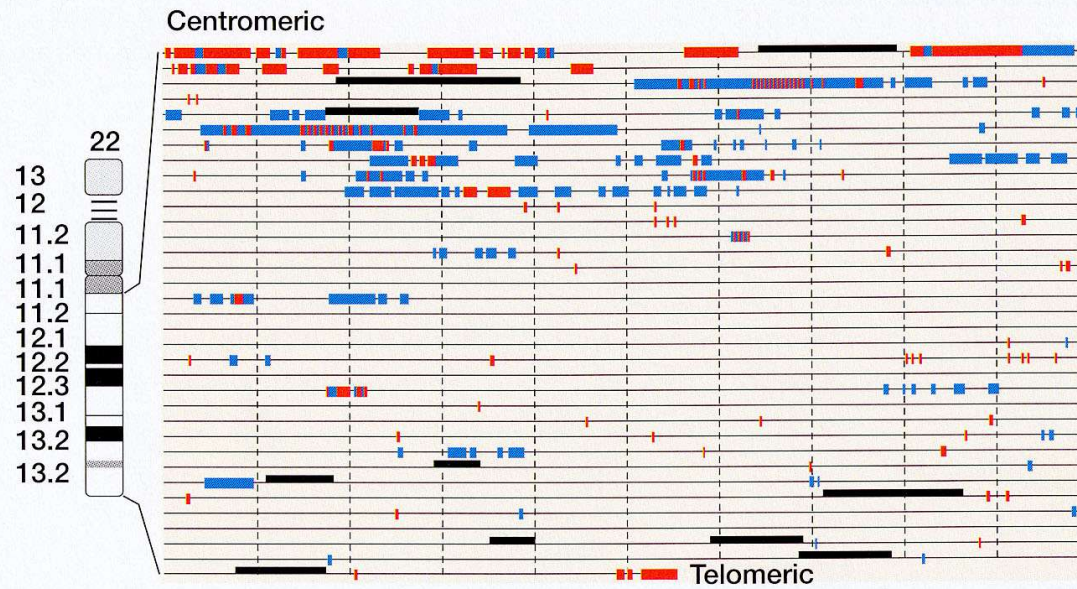








(A)



(B)

