Organizace lidského genomu, mutace a instabilita lidské DNA

RNDr. Iveta Valášková , RNDr. Jitka Kadlecová
OLG, FN Brno *PřF 2005*

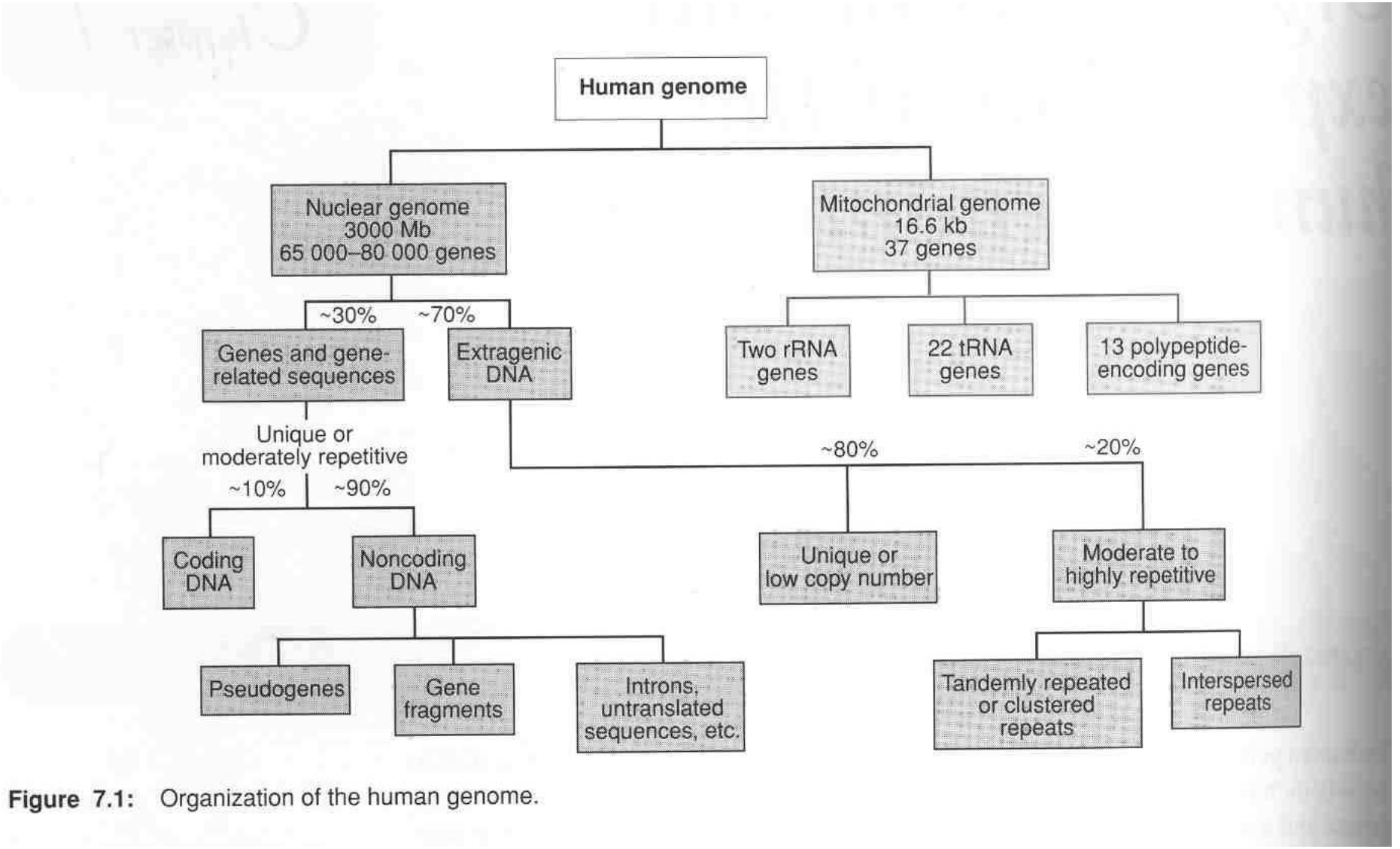


Figure 7.1: Organization of the human genome.

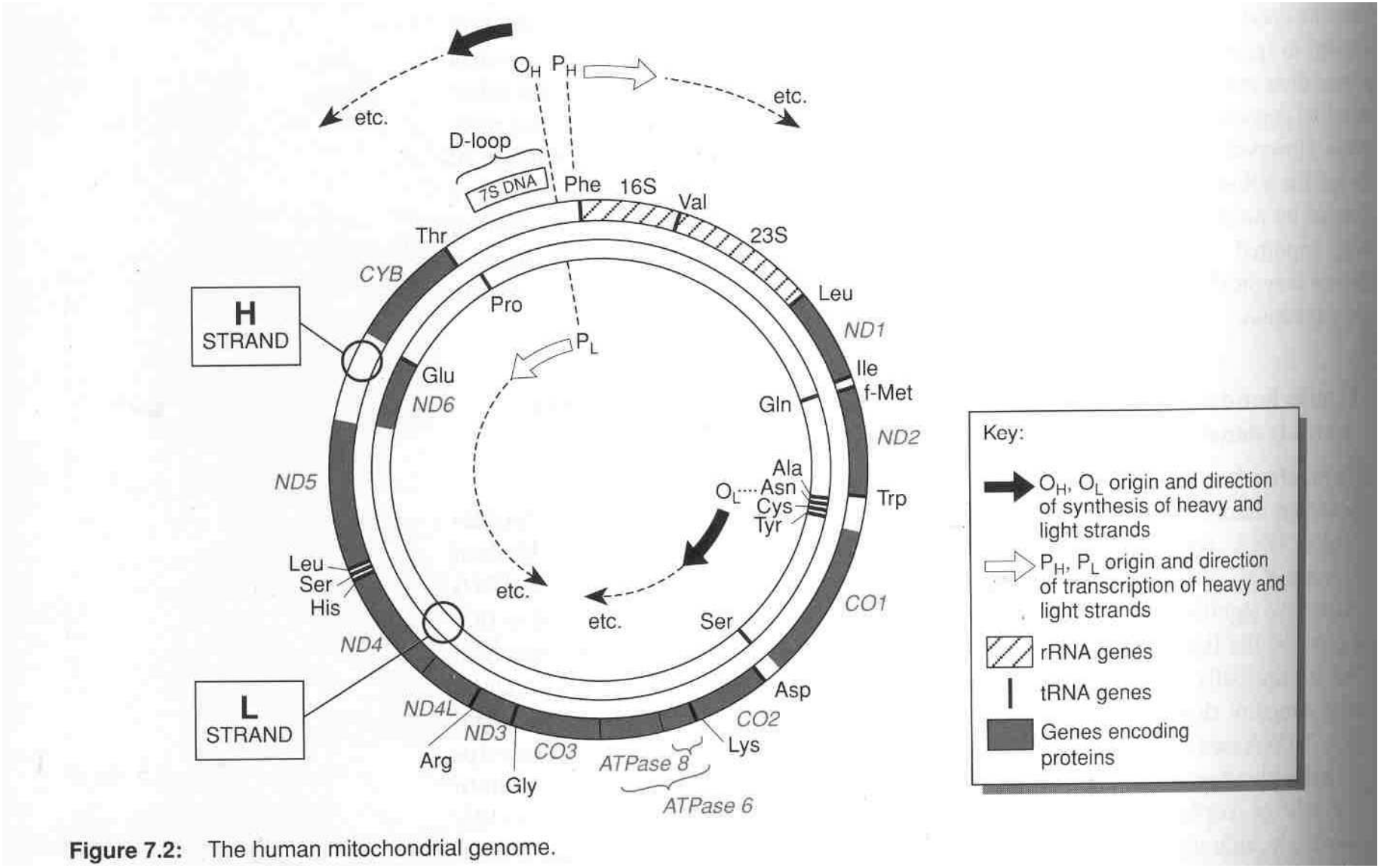


Figure 7.2: The human mitochondrial genome.

Jaderný a mitochondriální genom

- liší se v mnoha aspektech organizace a exprese

	Nuclear genome	Mitochondrial genome
Size	3000 Mb	16.6 kb
No. of different DNA molecules	23 (in XX) or 24 (in XY) cells, all linear	One circular DNA molecule
Total no. of DNA molecules per cell	23 in haploid cells; 46 in diploid cells	Several $\times 10^3$
Associated protein	Several classes of histone and nonhistone protein	Largely free of protein
No. of genes	~65 000–80 000	37
Gene density	~1/40 kb	1/0.45 kb
Repetitive DNA	Large fraction, see *Figure 7.1.*	Very little
Transcription	The great bulk of genes are transcribed individually	Continuous transcription of multiple genes
Introns	Found in most genes	Absent
% of coding DNA	~3%	~93%
Codon usage	See *Figure 1.22*	See *Figure 1.22*
Recombination	At least once for each pair of homologs at meiosis	None
Inheritance	Mendelian for sequences on X and autosomes; paternal for sequences on Y	Exclusively maternal

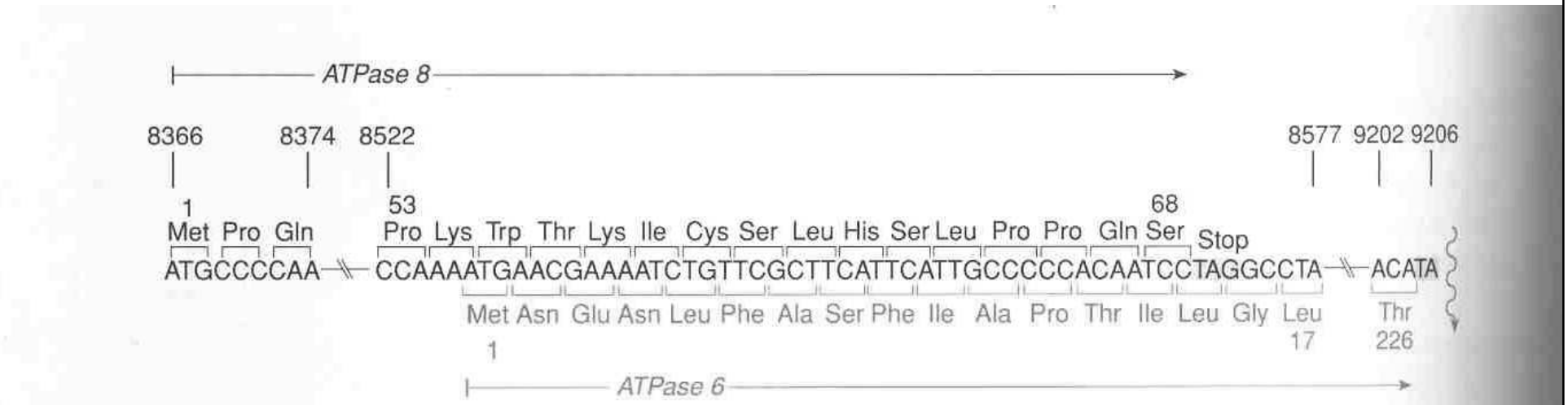


Figure 7.3: The genes for mitochondrial ATPase subunits 6 and 8 are partially overlapping and translated in different reading frames.

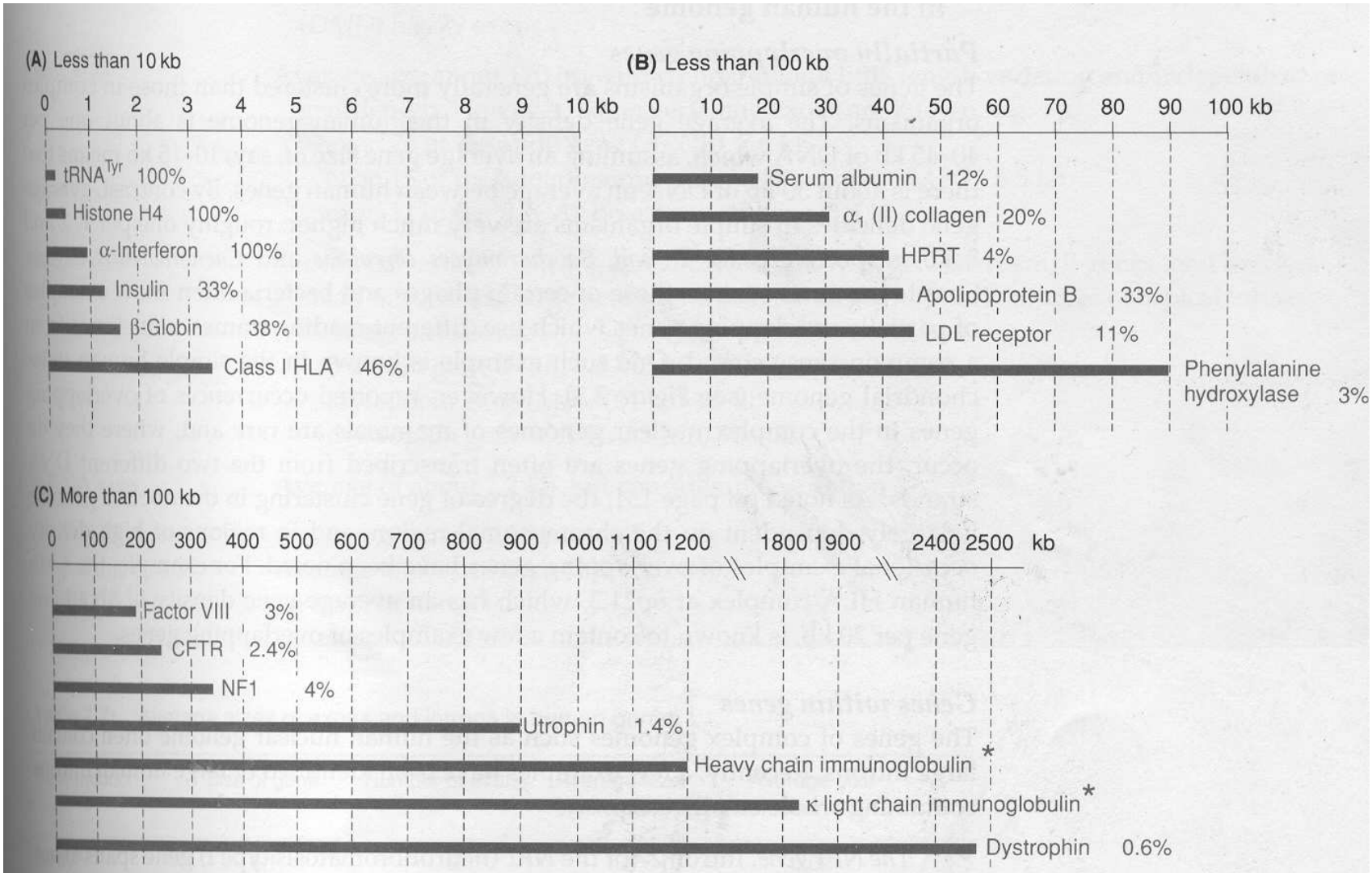


Figure 7.6: Human genes vary enormously in size and exon content.

Exon content is shown as a percentage of the lengths of indicated genes. *Note* the generally inverse relationship between gene length and percentage of exon content. Asterisks emphasize that the lengths given for the indicated Ig heavy chain and light chain loci correspond to the germline organizations. Immunoglobulin and T-cell receptor genes have unique organizations, requiring cell-specific somatic rearrangements in order to be expressed in B or T lymphocytes respectively (see page 177). CFTR, cystic fibrosis transmembrane regulator; HPRT, hypoxanthine phosphoribosyl transferase; NF1, neurofibromatous type 1.

Table 7.7: Average sizes of exons and introns in human genes

Gene product	Size of gene	Number of exons	Average size of exon (bp)	Average size of intron (kb)
$tRNA^{tyr}$	0.1	2	50	0.02
Insulin	1.4	3	155	0.48
β-Globin	1.6	3	150	0.49
Class I HLA	3.5	8	187	0.26
Serum albumin	18	14	137	1.1
Type VII collagen	31	118	77	0.19
Complement C3	41	29	122	0.9
Phenylalanine hydroxylase	90	26	96	3.5
Factor VIII	186	26	375	7.1
CFTR (cystic fibrosis)	250	27	227	9.1
Dystrophin	2400	79	180	30.0

Table 7.2: Human chromosome groups

Group	Chromosomes	Description
A	1–3	Largest; 1 and 3 are metacentric but 2 is submetacentric
B	4,5	Large; submetacentric with two arms very different in size
C	6–12, X	Medium size; submetacentric
D	13–15	Medium size; acrocentric with satellites
E	16–18	Small; 16 is metacentric but 17 and 18 are submetacentric
F	19, 20	Small; metacentric
G	21, 22, Y	Small; acrocentric, with satellites on 21 and 22 but not on the Y

Note that numbering of autosomes is in decreasing order in size, except that chromosome 21 is now known to be smaller than chromosome 22

Table 7.3: DNA content of human chromosomes[a]

Chromosomes	Percentage of total length	Amount of DNA (Mb)	Chromosome	Percentage of total length	Amount of DNA (Mb)
1	8.3	250	13	3.6	110
2	7.9	240	14	3.5	105
3	6.4	190	15	3.3	100
4	6.1	180	16	2.8	85
5	5.8	175	17	2.7	80
6	5.5	165	18	2.5	75
7	5.1	155	19	2.3	70
8	4.5	135	20	2.1	65
9	4.4	130	21	1.8	55
10	4.4	130	22	1.9	60
11	4.4	130	X	4.7	140
12	4.1	120	Y	2.0	60

[a]The DNA content is given for chromosomes prior to entering the S (DNA replication) phase of cell division (see *Figure 1.3*).
Data abstracted from Stephens *et al.* (1990).

Table 7.8: Examples of intragenic repetitive coding DNA (see also page 266)

Gene product	Size of encoded repeat in amino acids	No. of copies	Nucleotide sequence homology between copies
Ubiquitin (*UbB* and *UbC* genes)	76	3 (*UbB*) 9 (*UbC*)	High homology
Involucrin	10	59	High homology for central 39 repeats
Apolipoprotein (a)	114 = kringle 4-like repeat[a]	37	High homology; 24 of the repeats are identical in sequence
Plasminogen	~75–80	5	Low homology but conserved protein domains (kringles[a])
Collagen	18	57	Low homology but conserved amino acid motifs based on $(Gly\text{-}X\text{-}Y)_6$
Serum albumin	195	3	Low homology
Proline-rich protein genes	16–21	5	Low homology
Tropomyosin α-chain	42	7	Low homology
Immunoglobulin ε-chain, C region	108	4	Low homology
Dystrophin	109	24	Low homology

[a]A kringle is a cysteine-rich sequence that contains three internal disulfide bridges and forms a pretzel-shaped structure.

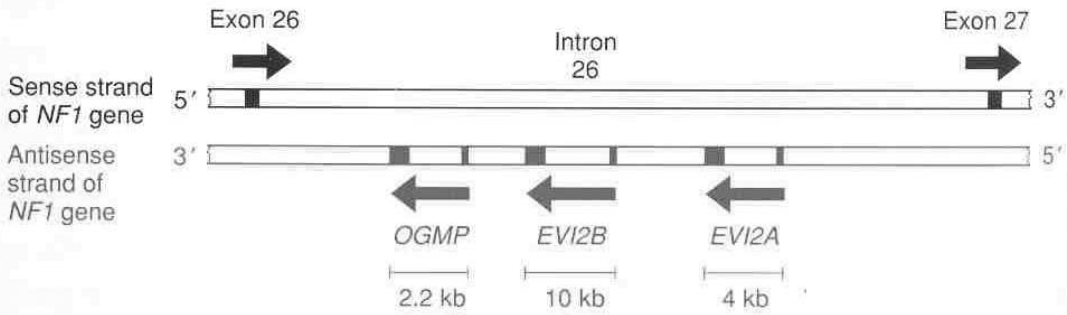


Figure 7.7: Genes within genes: intron 26 of the gene for neurofibromatosis type I (NF1) contains three internal genes each with two exons.

Note that the three internal genes are transcribed from the opposing strand to that used for transcription of the *NF1* gene. Genes are: *OGMP*, oligodendrocyte myelin glycoprotein; *EVI2A* and *EVI2B*, human homologs of murine/genes thought to be involved in leukemogenesis, and located at ecotropic viral integration sites.

Projekt HUGO

Table 11. Genome overview.

Size of the genome (including gaps)	2.91 Gbp
Size of the genome (excluding gaps)	2.66 Gbp
Longest contig	1.99 Mbp
Longest scaffold	14.4 Mbp
Percent of A+T in the genome	54
Percent of G+C in the genome	38
Percent of undetermined bases in the genome	9
Most GC-rich 50 kb	Chr. 2 (66%)
Least GC-rich 50 kb	Chr. X (25%)
Percent of genome classified as repeats	35
Number of annotated genes	26,383
Percent of annotated genes with unknown function	42
Number of genes (hypothetical and annotated)	39,114
Percent of hypothetical and annotated genes with unknown function	59
Gene with the most exons	Titin (234 exons)
Average gene size	27 kbp
Most gene-rich chromosome	Chr. 19 (23 genes/Mb)
Least gene-rich chromosomes	Chr. 13 (5 genes/Mb), Chr. Y (5 genes/Mb)
Total size of gene deserts (>500 kb with no annotated genes)	605 Mbp
Percent of base pairs spanned by genes	25.5 to 37.8*
Percent of base pairs spanned by exons	1.1 to 1.4*
Percent of base pairs spanned by introns	24.4 to 36.4*
Percent of base pairs in intergenic DNA	74.5 to 63.6*
Chromosome with highest proportion of DNA in annotated exons	Chr. 19 (9.33)
Chromosome with lowest proportion of DNA in annotated exons	Chr. Y (0.36)
Longest intergenic region (between annotated + hypothetical genes)	Chr. 13 (3,038,416 bp)
Rate of SNP variation	1/1250 bp

*In these ranges, the percentages correspond to the annotated gene set (26, 383 genes) and the hypothetical + annotated gene set (39,114 genes), respectively.

Table 10.1: Incidence of mutation classes in the human genome

Mutation class	Type of mutation	Incidence
Base substitutions	All types	Comparatively common type of mutation in coding DNA but also common in noncoding DNA
	Transitions and transversions	Unexpectedly, transitions are commoner than transversions, especially in mitochondrial DNA
	Synonymous and nonsynonymous substitutions	Synonymous substitutions are considerably more common than nonsynonymous substitutions in coding DNA; conservative substitutions are more common than nonconservative
	Gene conversion-like events (multiple base substitution)	Rare except at certain tandemly repeated loci or clustered repeats
Insertions	Of one or a few nucleotides	Very common in noncoding DNA but rare in coding DNA where they produce frameshifts
	Triplet repeat expansions	Rare but can contribute to several disorders, especially neurological disorders (see page 266)
	Other large insertions	Rare; can occasionally get large-scale tandem duplications and also insertions of transposable elements (page 271)
Deletions	Of one or a few nucleotides	Very common in noncoding DNA but rare in coding DNA where they produce frameshifts
	Larger deletions	Rare, but often occur at regions containing tandem repeats (page 268) or between interspersed repeats (see page 254 and *Figure 10.9*)
Chromosomal abnormalities	Numerical and structural	Rare as constitutional mutations, but can often be pathogenic (see page 51ff.) Much more common as somatic mutations and often in tumor cells

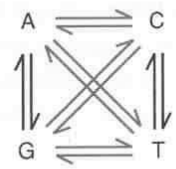


Figure 10.1: Transversions are theoretically expected to be twice as frequent as transitions. Red arrows, transversions; black arrows, transitions.

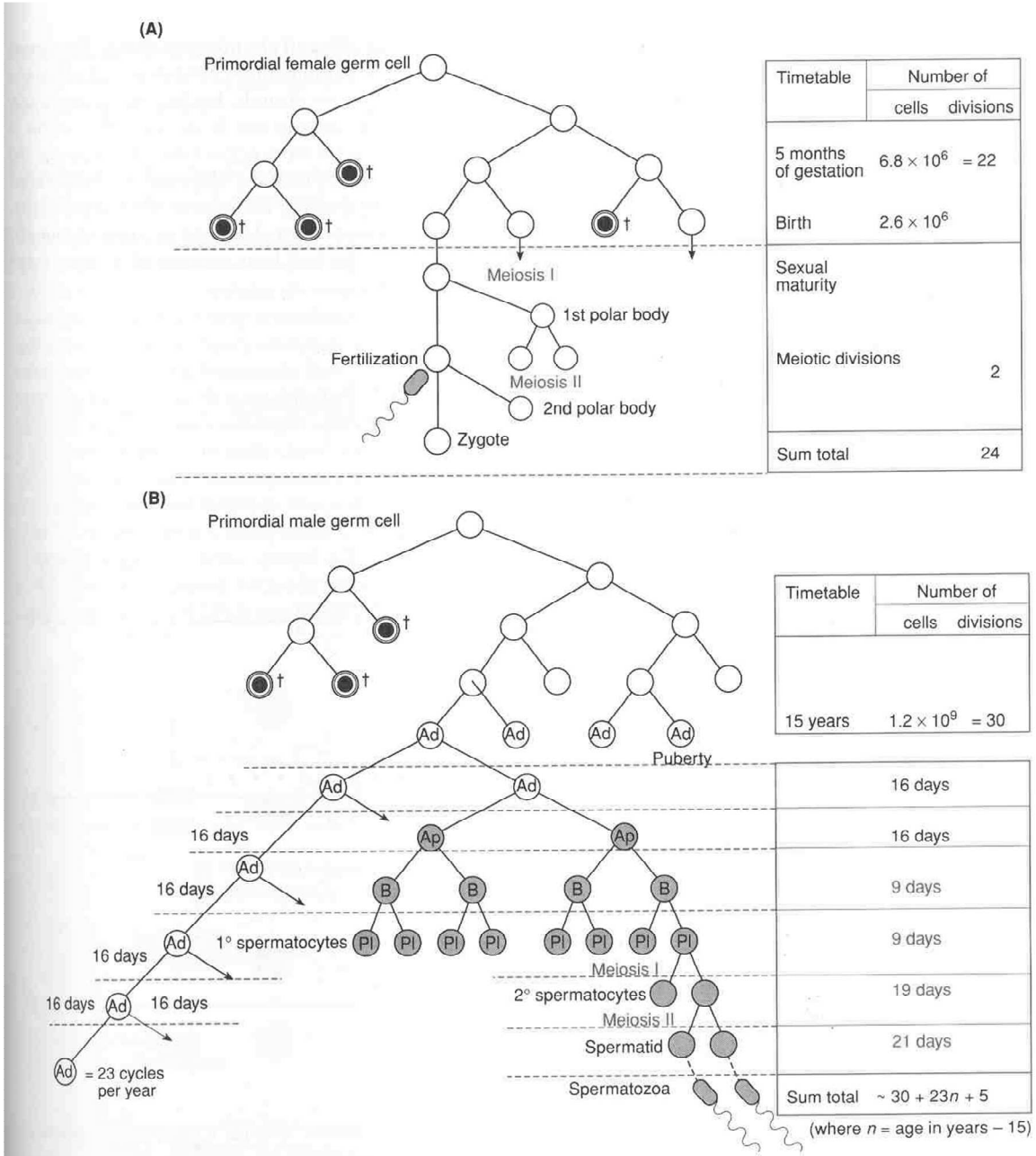


Figure 10.4: The number of cell divisions that are required to produce a human sperm cell is much greater than the number required to produce an egg cell.

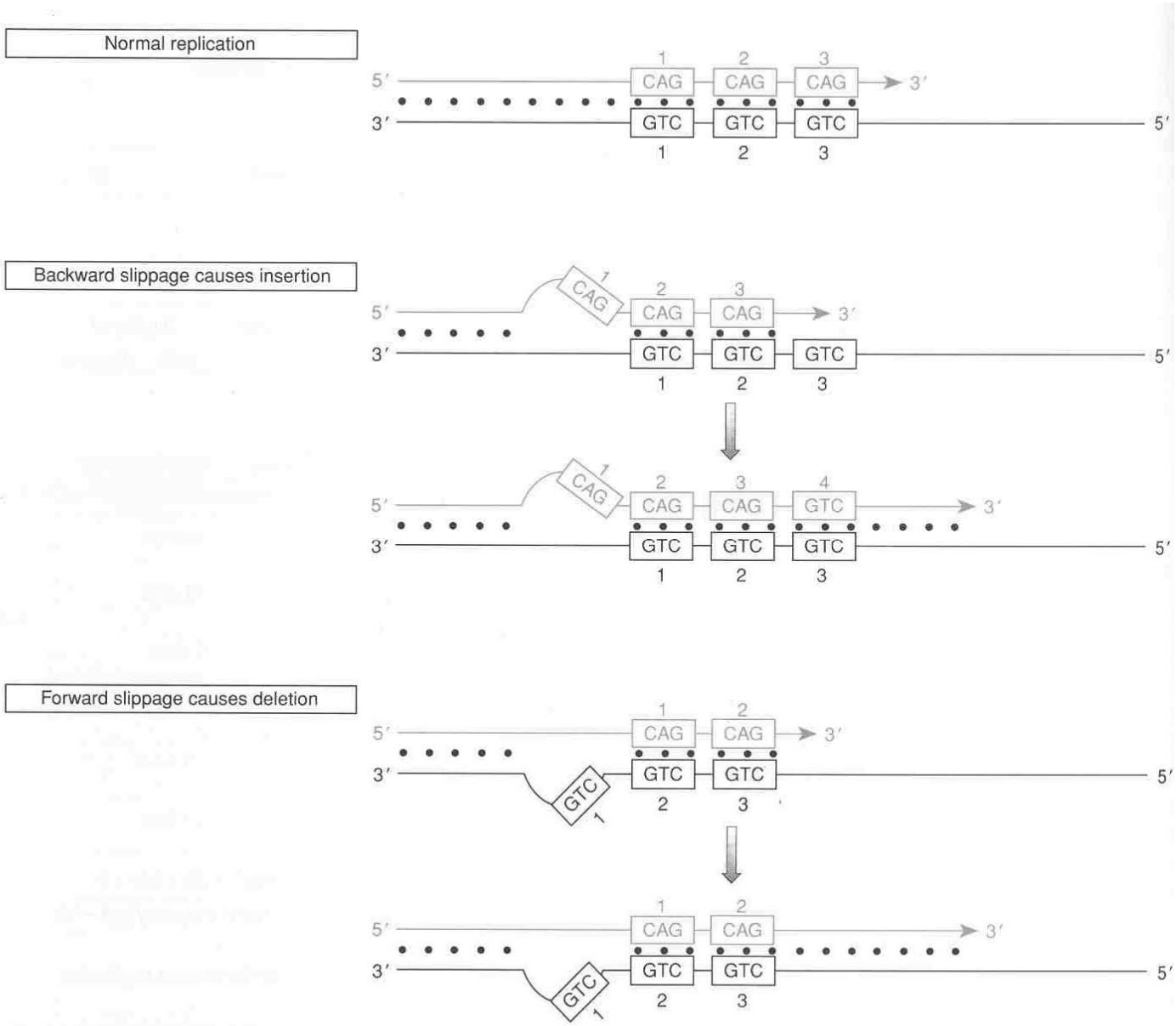


Figure 10.5: Slipped strand mispairing during DNA replication can cause insertions or deletions.

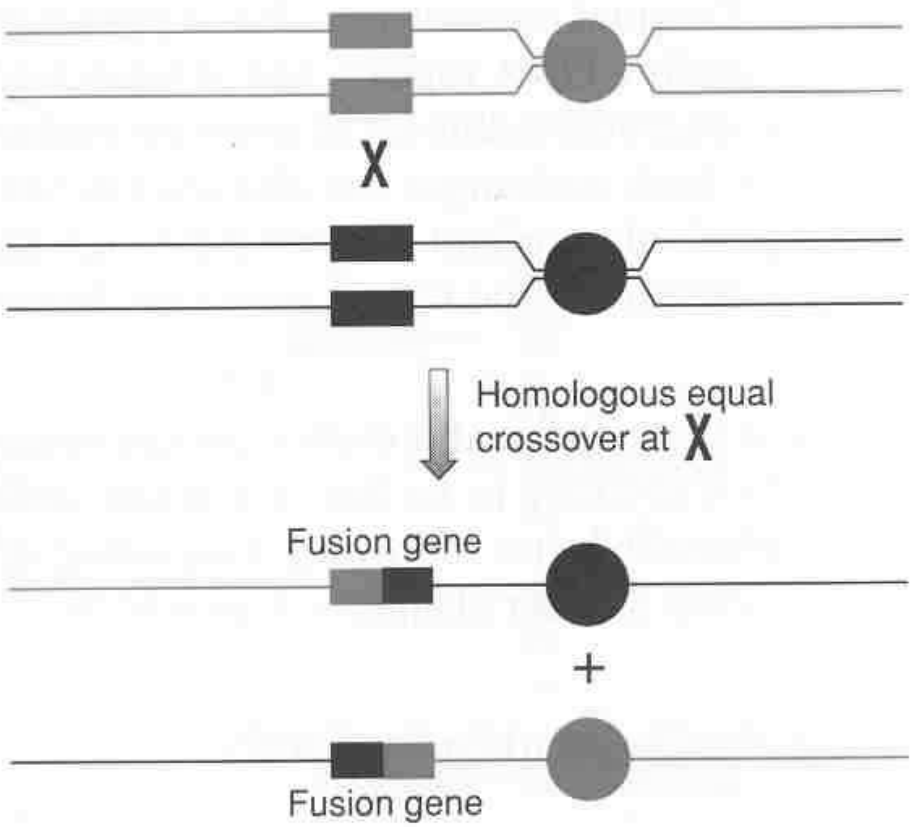
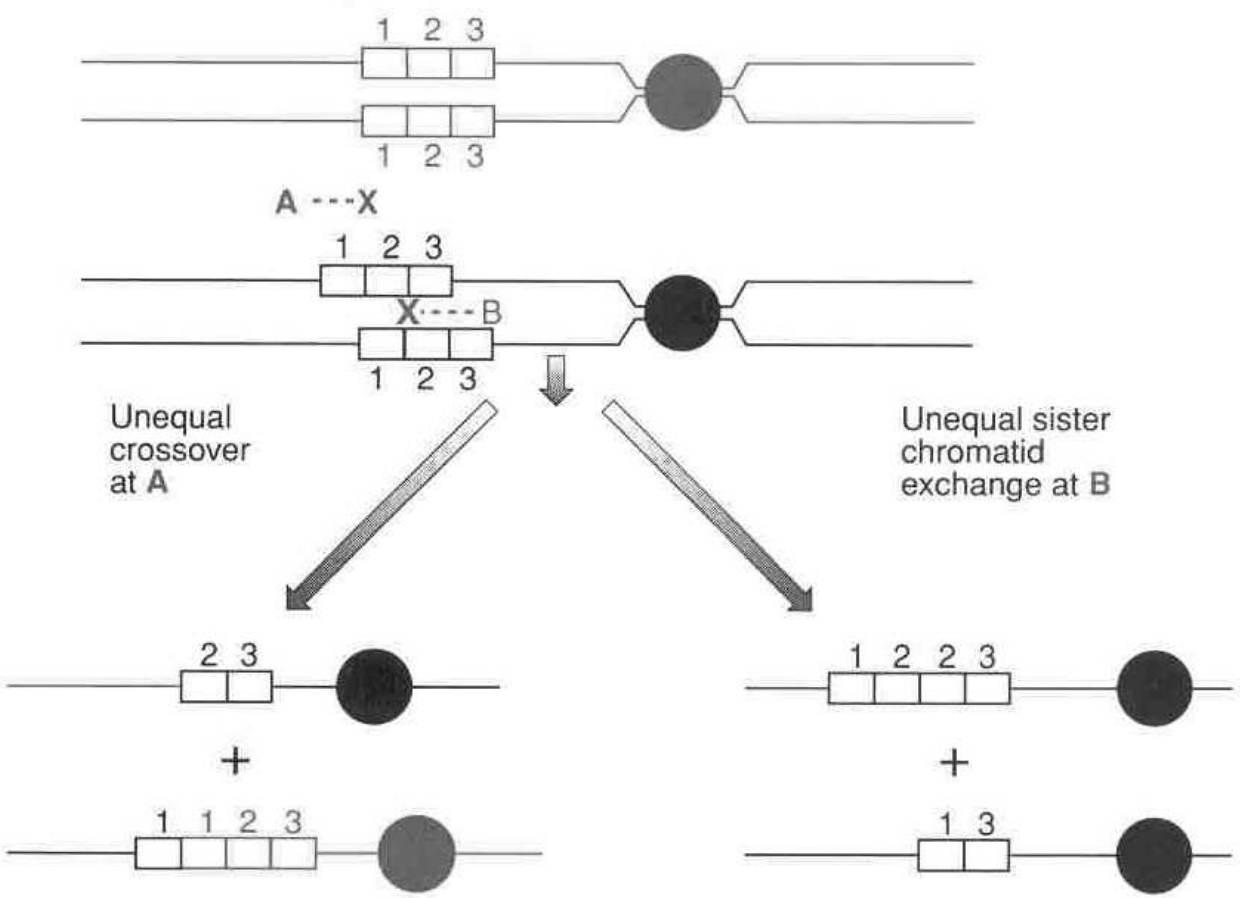


fig.: Vznik fúzního genu homologní rekombinací



Inzerce a delece způsobená nerovnoměrným crossing-overem a nerovnoměrnou sesterskou chromatidovou výměnou

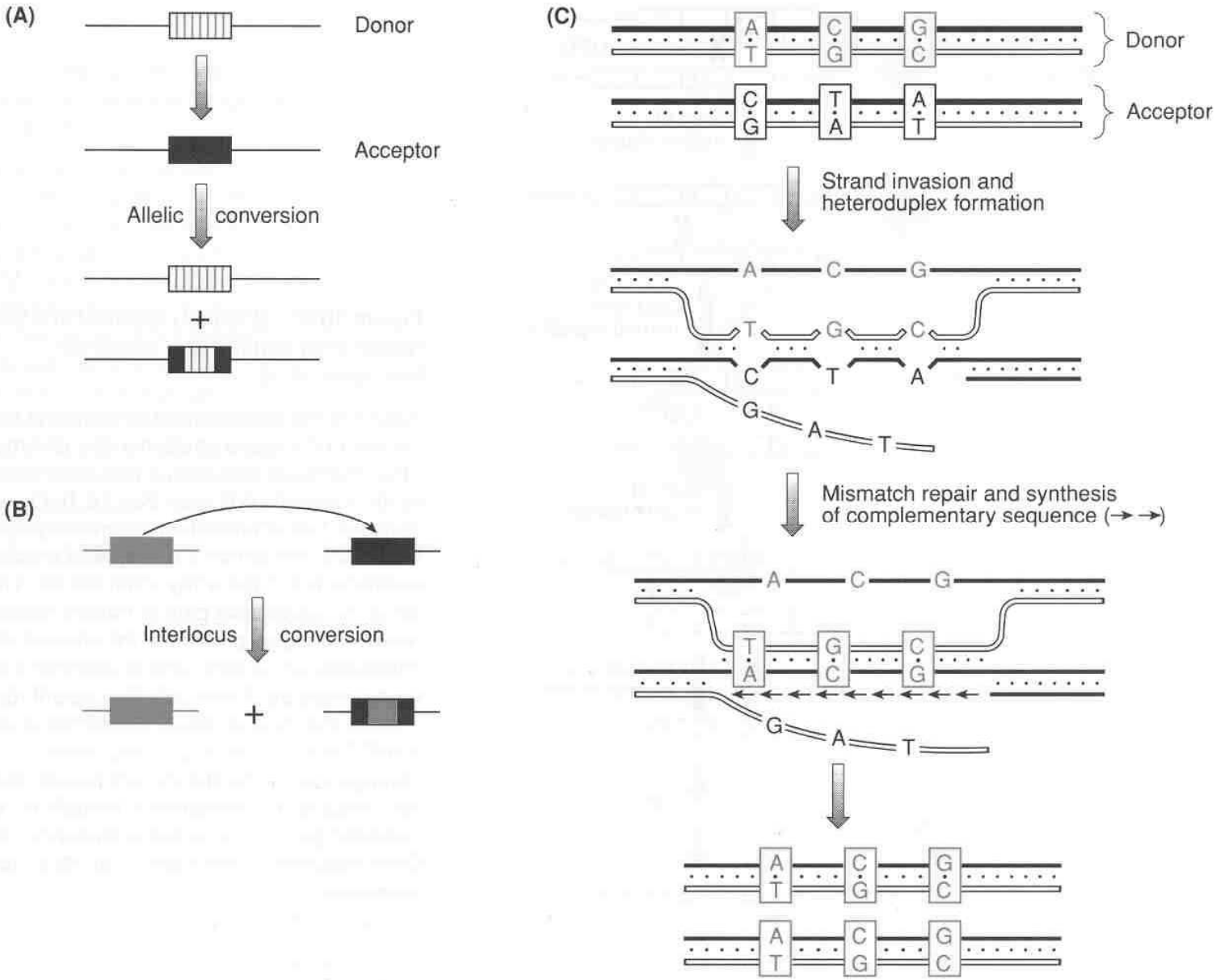
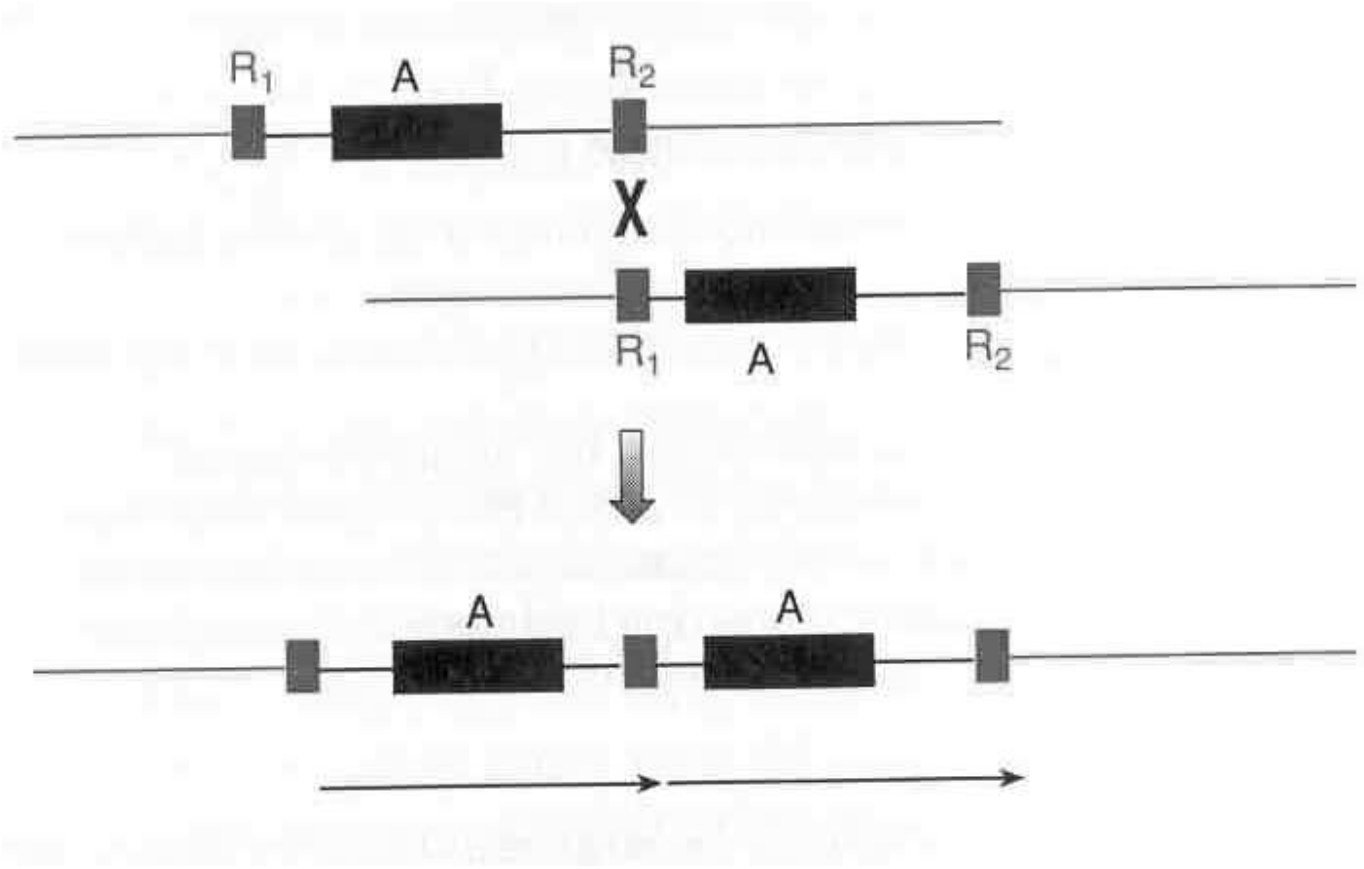


Figure 10.10: Gene conversion involves a nonreciprocal sequence exchange between allelic or nonallelic genes.



Tandemová genová duplikace jako výsledek nerovnoměrného crossing-overu a nerovnoměrné sesterské chromatidové výměny způsobená krátkými rozptýlenými repeticemi

Table 10.4: Effect of location and class of mutation on gene function

Location and nature of mutation	Effect on gene function	Comments
Extragenic mutation	Normally none	Rare mutations may result in inactivation of distant regulatory elements required for normal gene expression (see *Figure 8.6*)
Multigene deletion	Abolition	Associated with contiguous gene syndromes (see *Figure 15.4*)
Whole gene deletion	Abolition	
Whole gene duplication	Can have effect due to altered gene dosage	Large duplications including the peripheral myelin protein 22 gene can cause Charcot–Marie–Tooth syndrome (see *Figure 15.6*)
Whole exon deletion	Abolition or modification	May cause shift in reading frame; protein often unstable
Within exon	Abolition	If loss/change of key amino acids, shift of the reading frame or introduction of premature stop codon
	Modification	If nonconservative substitutions, small in-frame insertions or other mutations at some locations
	None	If conservative/silent substitutions or mutation at nonessential sites
Whole intron deletion	None	
Splice site mutation	Abolition or modulation of expression	Conserved GT and AG signals are critically important for normal gene expression. Mutations may induce exon skipping
Promoter mutation	Abolition or modulation of expression	Deletion, insertion or substitution of nucleotides within promoter may alter expression. Complete deletion abolishes function
Mutation of termination codon	Modification	Additional amino acids are included at the end of the protein until another stop codon is reached
Mutation of poly(A) signal	Abolition or modulation of expression	Deletion, insertion or substitution of nucleotides within poly(A) site may alter expression. Complete deletion abolishes function
Elsewhere in introns/UTS	Usually none	

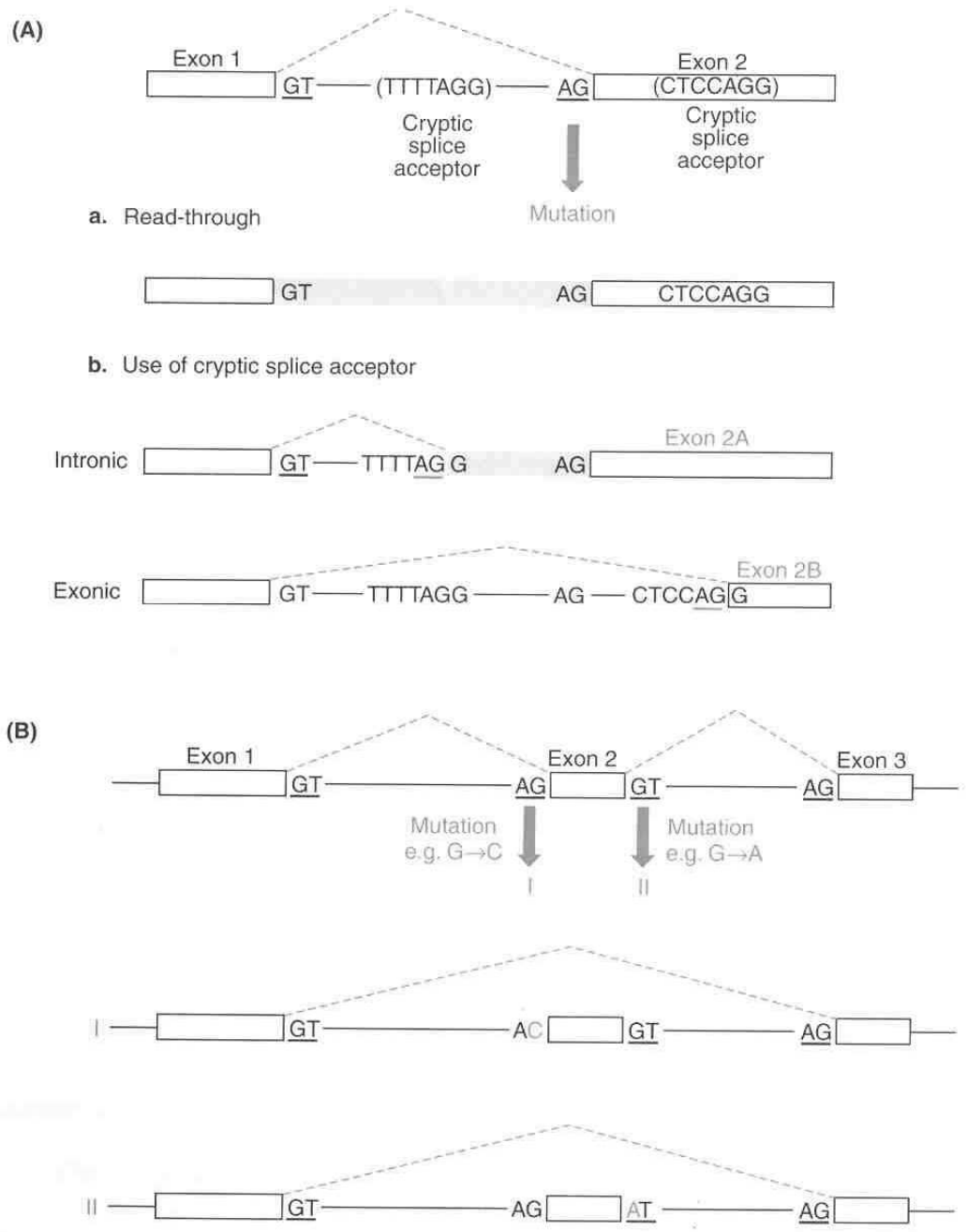


Figure 10.11: Mutations in conserved splice sites can cause altered exons or exon skipping.

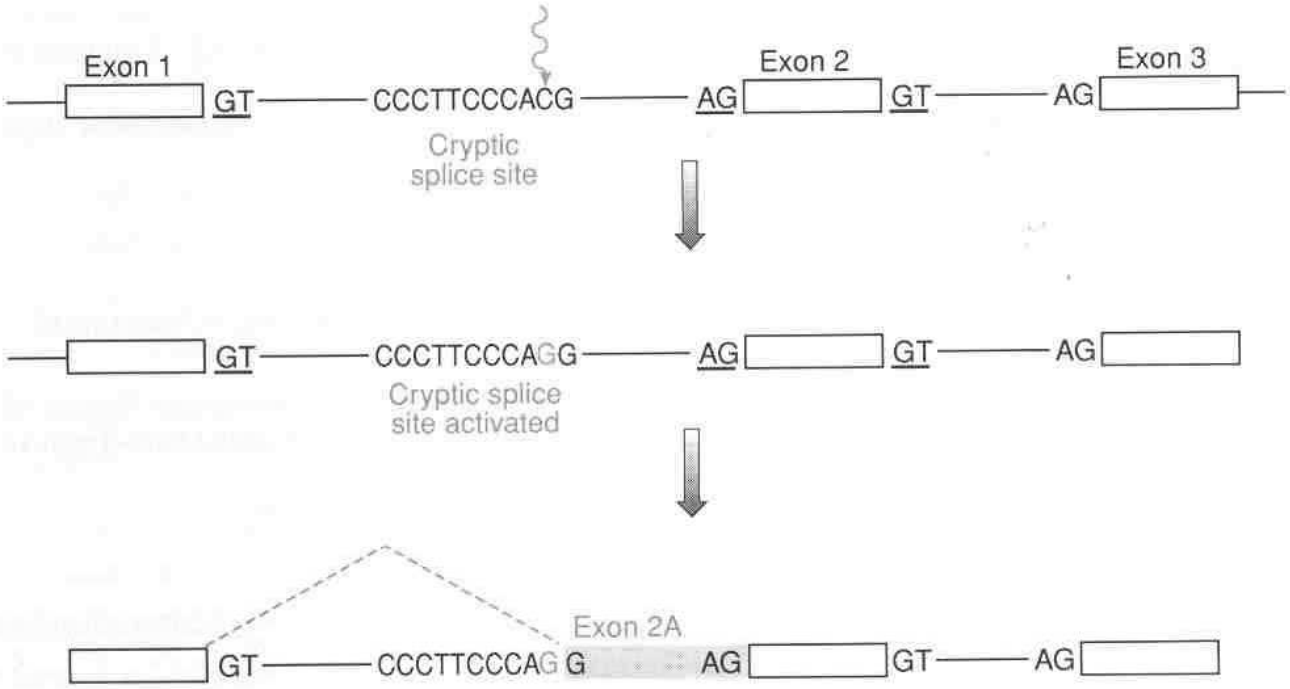


Figure 10.12: Mutations can cause abnormal RNA splicing by activation of cryptic splice sites.

A mutation can result in the alteration of a sequence which is not important for RNA splicing so as to create a new, alternative splice site. In the example illustrated, the mutation is envisaged to change a single nucleotide in intron 1. The nucleotide happens to occur within a *cryptic splice site* sequence that is closely related to the splice acceptor consensus sequence but, unlike the cryptic splice acceptor sites in *Figure 10.11,* shows a difference with respect to the conserved AG dinucleotide (see *Figure 1.15*). The mutation overcomes this difference and so can activate the cryptic splice site so that it competes with the natural splice acceptor site. If it is used by the splicing apparatus, a novel exon, exon 2A, results, which contains additional sequence which may or may not result in a frameshift.

Patologický potenciál repetitivních sekvencí

Type of repeated DNA	Type of mutation	Mechanism and examples
Tandem repeats		
Very short repeats within genes	Deletion	Slipped strand mispairing (see *Figure 10.5*). Examples in *Figure 10.13*
	Frameshifting insertion	Slipped strand mispairing
	Triplet repeat expansion	Initially by slipped strand mispairing?; subsequently large-scale expansion by unknown mechanism
Moderate sized intragenic repeats	Intragenic deletion	UEC/UESCE[a] (see *Figure 10.7*)
Large tandem repeats containing whole genes	Partial or total gene deletion	UEC/UESCE[a] (*Figure 10.7*). Examples in *Figure 10.15*
	Alteration of gene sequence	Gene conversion (*Figure 10.10*). Examples in *Figures 10.15 and 10.16*
	Duplication causing gene dosage-related aberrant expression	UESCE[a] – 1.5 Mb duplication in Charcot–Marie–Tooth 1A (see *Figure 15.6*)
Interspersed repeats		
Short direct repeats	Deletion	Slipped strand mispairing or intrachromatid recombination?
Interspersed repeat elements (e.g. *Alu* repeats)	Deletion	UEC/UESCE[a]
	Duplication	UEC/UESCE[a]
Inverted repeats	Inversion	Intrachromatid exchange, e.g. Factor VIII (see *Figure 10.18*)
Active transposable elements	Intragenic insertion by retrotransposons	Retrotransposition (*Figures 8.7* and *8.11*). Examples, see page 271

[a]UEC, unequal crossover; UESCE, unequal sister chromatid exchange.

TR —mutace→ **TREs** ⟹ **TRED**

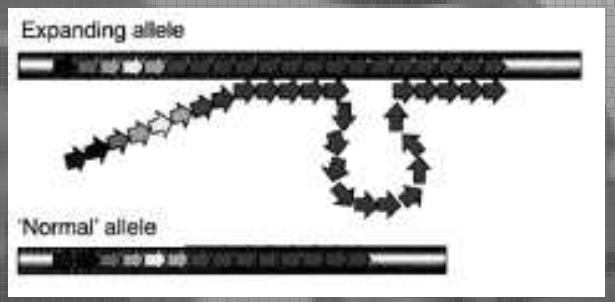
trinucleotide repeat

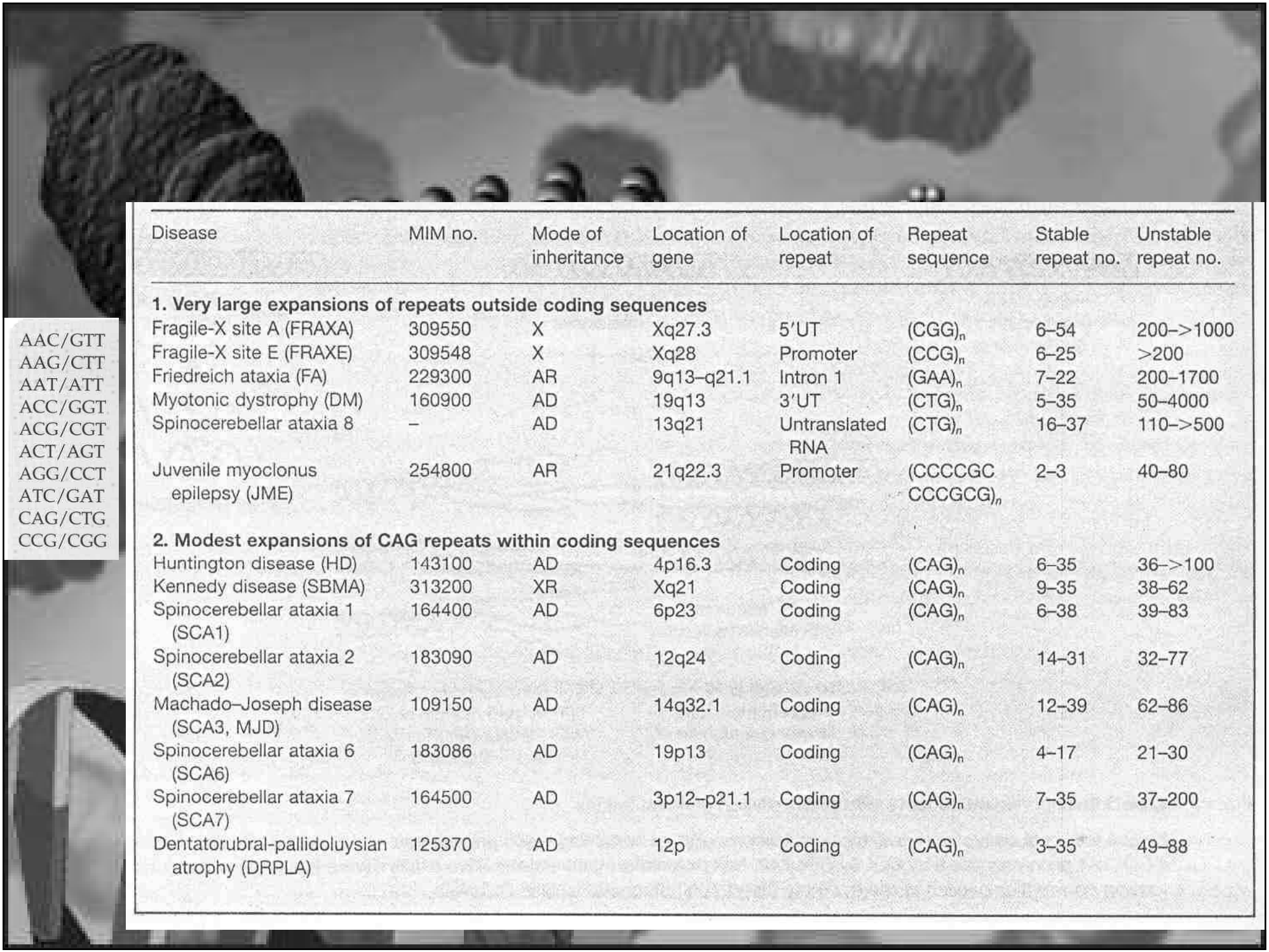
trinucleotide repaet expansion

trinucleotide repeat expansion diseases

expanze

nový typ mutace, popsán 1991





AAC/GTT
AAG/CTT
AAT/ATT
ACC/GGT
ACG/CGT
ACT/AGT
AGG/CCT
ATC/GAT
CAG/CTG
CCG/CGG

Disease	MIM no.	Mode of inheritance	Location of gene	Location of repeat	Repeat sequence	Stable repeat no.	Unstable repeat no.
1. Very large expansions of repeats outside coding sequences							
Fragile-X site A (FRAXA)	309550	X	Xq27.3	5′UT	$(CGG)_n$	6–54	200–>1000
Fragile-X site E (FRAXE)	309548	X	Xq28	Promoter	$(CCG)_n$	6–25	>200
Friedreich ataxia (FA)	229300	AR	9q13–q21.1	Intron 1	$(GAA)_n$	7–22	200–1700
Myotonic dystrophy (DM)	160900	AD	19q13	3′UT	$(CTG)_n$	5–35	50–4000
Spinocerebellar ataxia 8	–	AD	13q21	Untranslated RNA	$(CTG)_n$	16–37	110–>500
Juvenile myoclonus epilepsy (JME)	254800	AR	21q22.3	Promoter	$(CCCCGC CCCGCG)_n$	2–3	40–80
2. Modest expansions of CAG repeats within coding sequences							
Huntington disease (HD)	143100	AD	4p16.3	Coding	$(CAG)_n$	6–35	36–>100
Kennedy disease (SBMA)	313200	XR	Xq21	Coding	$(CAG)_n$	9–35	38–62
Spinocerebellar ataxia 1 (SCA1)	164400	AD	6p23	Coding	$(CAG)_n$	6–38	39–83
Spinocerebellar ataxia 2 (SCA2)	183090	AD	12q24	Coding	$(CAG)_n$	14–31	32–77
Machado–Joseph disease (SCA3, MJD)	109150	AD	14q32.1	Coding	$(CAG)_n$	12–39	62–86
Spinocerebellar ataxia 6 (SCA6)	183086	AD	19p13	Coding	$(CAG)_n$	4–17	21–30
Spinocerebellar ataxia 7 (SCA7)	164500	AD	3p12–p21.1	Coding	$(CAG)_n$	7–35	37–200
Dentatorubral-pallidoluysian atrophy (DRPLA)	125370	AD	12p	Coding	$(CAG)_n$	3–35	49–88

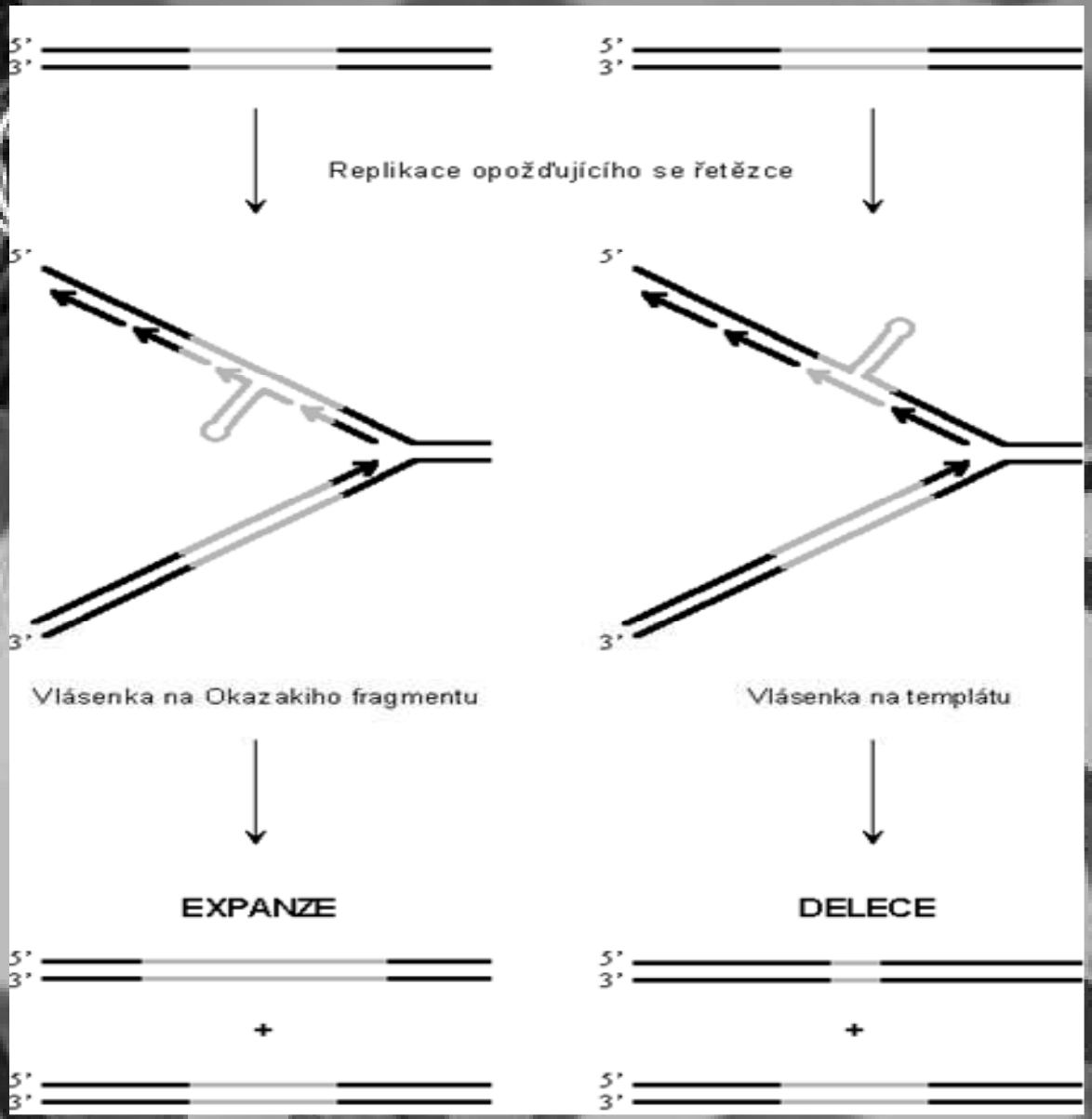
TR - trinukleotidové repetice

- široce rozšířeny v lidském genomu
  * v intronech
  * uvnitř čtecích rámců (exonech)
    - v překládaných
    - nepřekládaných

  > oblastech

  - nestabilita, závisící na:
    * typu sekvence
    * délce repetitivní sekvence

Expanze a delece tinukleotidů při replikaci



Rozdělení chorob zapříčiněných expanzí trinukleotidových repetic podle lokalizace TREs

1 TR lokalizovány uvnitř ORF

↳ expanzí narušená struktura proteinů
(Huntingtonova chorea)

2 TR lokalizovány vně ORF - v 3´UTR nebo 5´UTR
- v intronech

↳ patrně inaktivují nebo ovlivňují expresi genu
(Myotonická dystrofie)

Klinická anticipace

spojená s expanzivním prodlužováním trinukleotidových repetic u

asymptomatický prarodič

(MD1 : 7 - 50 CTG repetic)

rodič s mírným klinickým projevem

(MD1: 100 CTG repetic)

potomek s těžkým průběhem choroby

(MD: 1500 CTG repetic)

Myotonická dystrofie 1
MD1

- autozomálně dominantní choroba
- nejčastější forma svalové dystrofie
  - celosvětová frekvence výskytu 1 : 8000

patří mezi onemocnění TREDs

příčina

expandované trinukleotidové repetice (TREs)

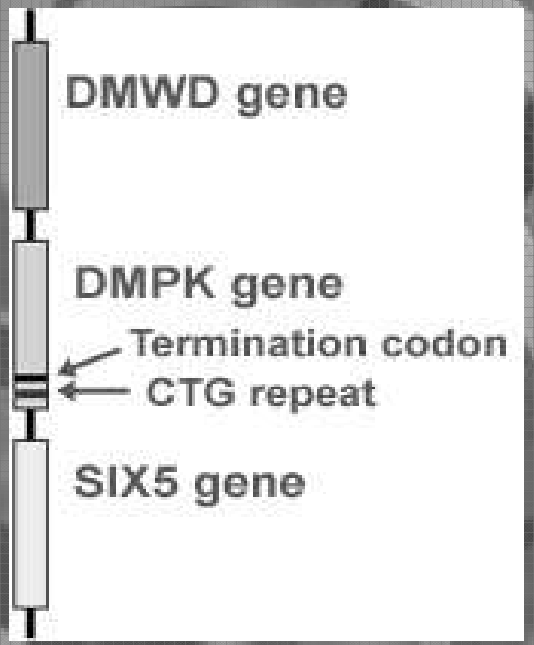
Myotonická dystrofie 1

příčina

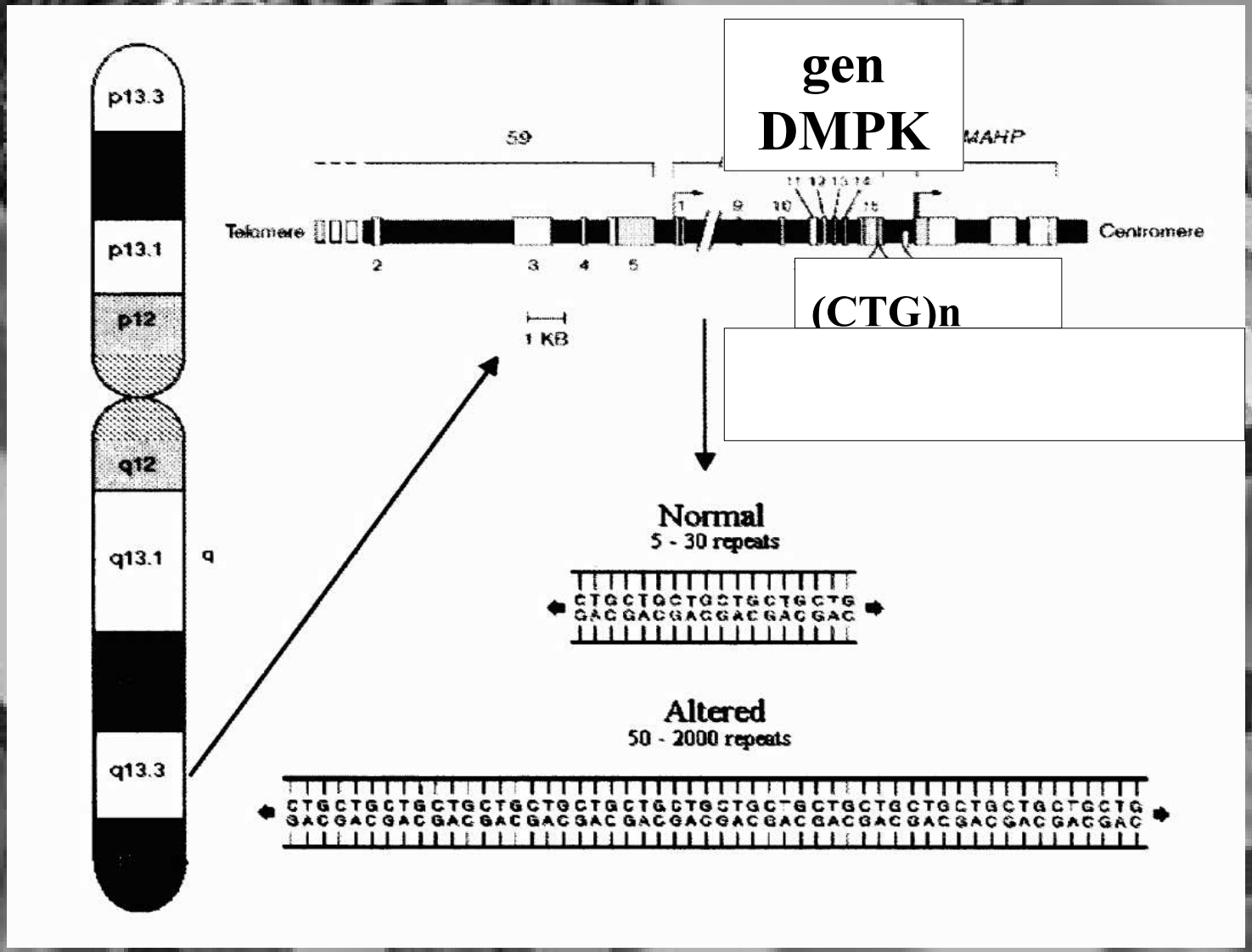
expanze trinukleotidu CTG

ve 3´UTR (untranslated region) genu DMPK

lokus 19q13.3

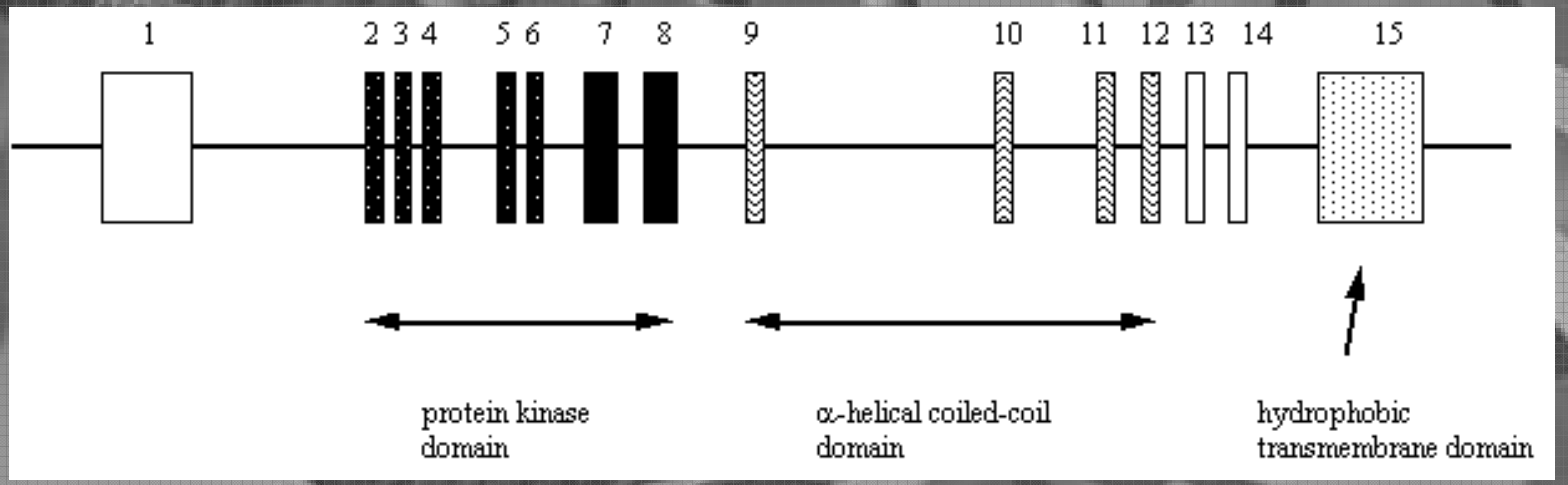


Expanze CTG trinukleotidu a její lokalizace



DMPK

Dystrophia Myotonica Protein Kinase



Exprese: především v srdci, kosterním svalstvu, mozku
Funkce: zapojen do signálních drah

hraje centrální roli v buněčné regulaci


DMPK gen
normální produkt
DMPK protein
expanzivní
mutace
neovlivněná funkce DMPK proteinu
vznik patologického fenotypu neobjasněn
pravděpodobné příčiny vzniku MD:
• narušený transport a úprava mRNA
• narušení struktury chromatinu expandovaným repetitivním traktem ⟶ porucha exprese genů lokalizovaných v okolí genu DMPK
• vysycení DNA vazebných proteinů ⟶ narušení funkce genů v okolí genu DMPK

Korelace genotypu a fenotypu u MD1

	premutace	klinická forma myotonické dystrofie		
		mírná	klasická	neonatální
počet CTG trinukleotidů	35 - cca. 49	50 - cca. 150	100 - 1000 až 1500	1000 až cca. 3000
propuknutí choroby ve věku	-	21 - 40 let	11 - 20 let	od narození do cca. 10 let
průměrná délka života	normální	64 let	48 - 55 let	přežije-li neonatální období až 45 let
klinické projevy	žádné postižení vyjímečně katarata katarakta	katarakta mírná myotonie	svalová slabost myotonie katarakta předčasné plešatění srdeční arytmie postižení endokrinního systému další	těžká hypotonie respirační problémy postižení srdce mentální retardace další typický výraz obličeje "maska"

Diagnostika MD

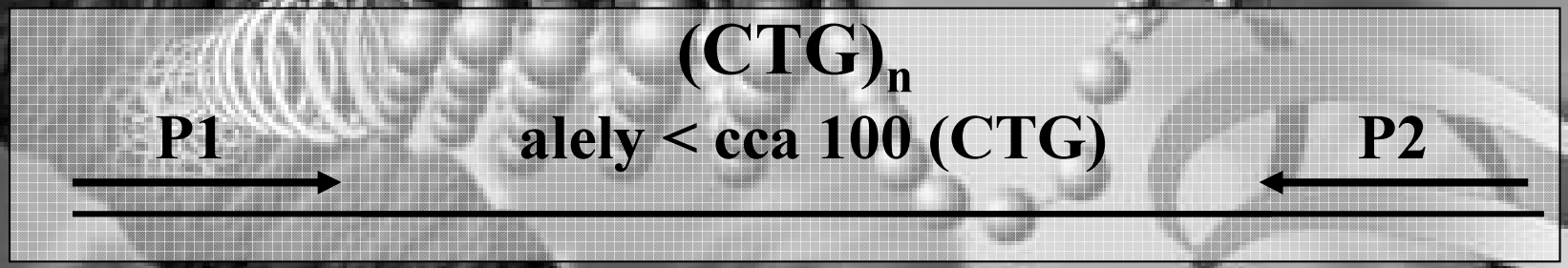
Klinické vyšetření
neurologické vyšetření
EMG, EEG
kardiologické vyšetření

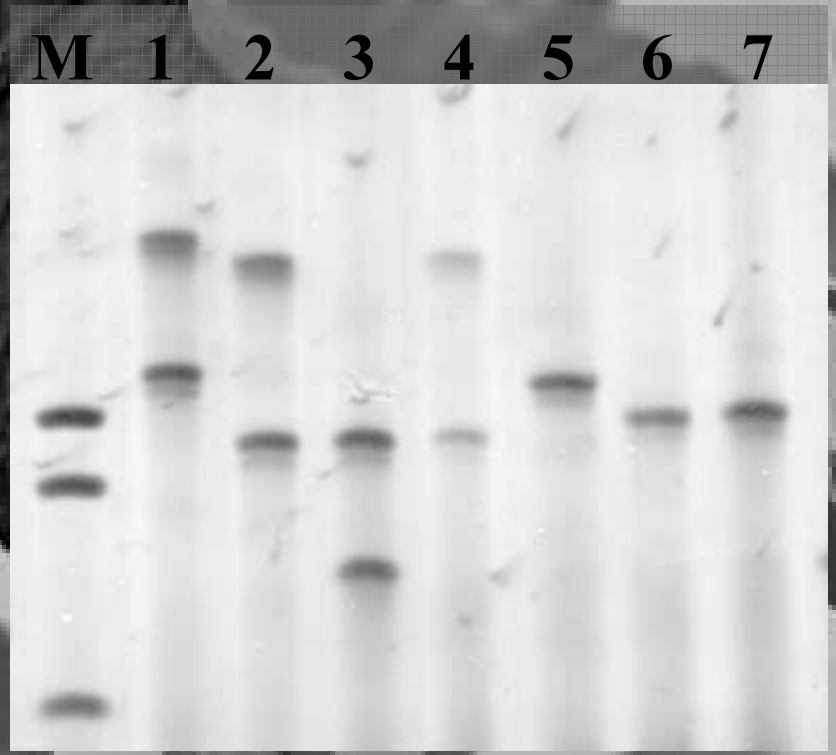
Histologické vyšetření postižených svalů
histochemie
elektron. mikroskopie

Molekulárně genetické vyšetření
PCR, TP PCR

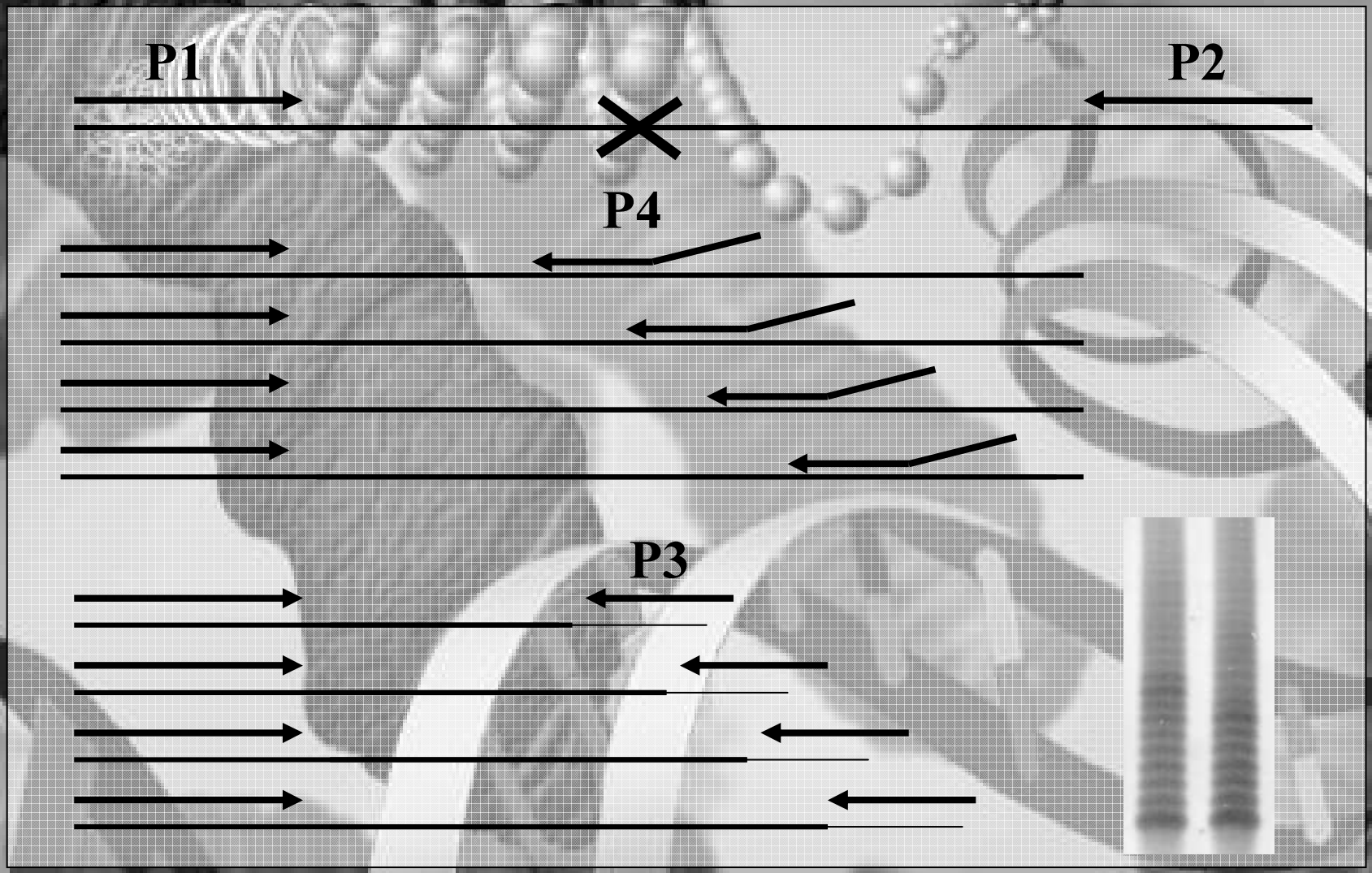
jednoznačně vyvrátí nebo potvrdí diagnózu

PCR P1/P2

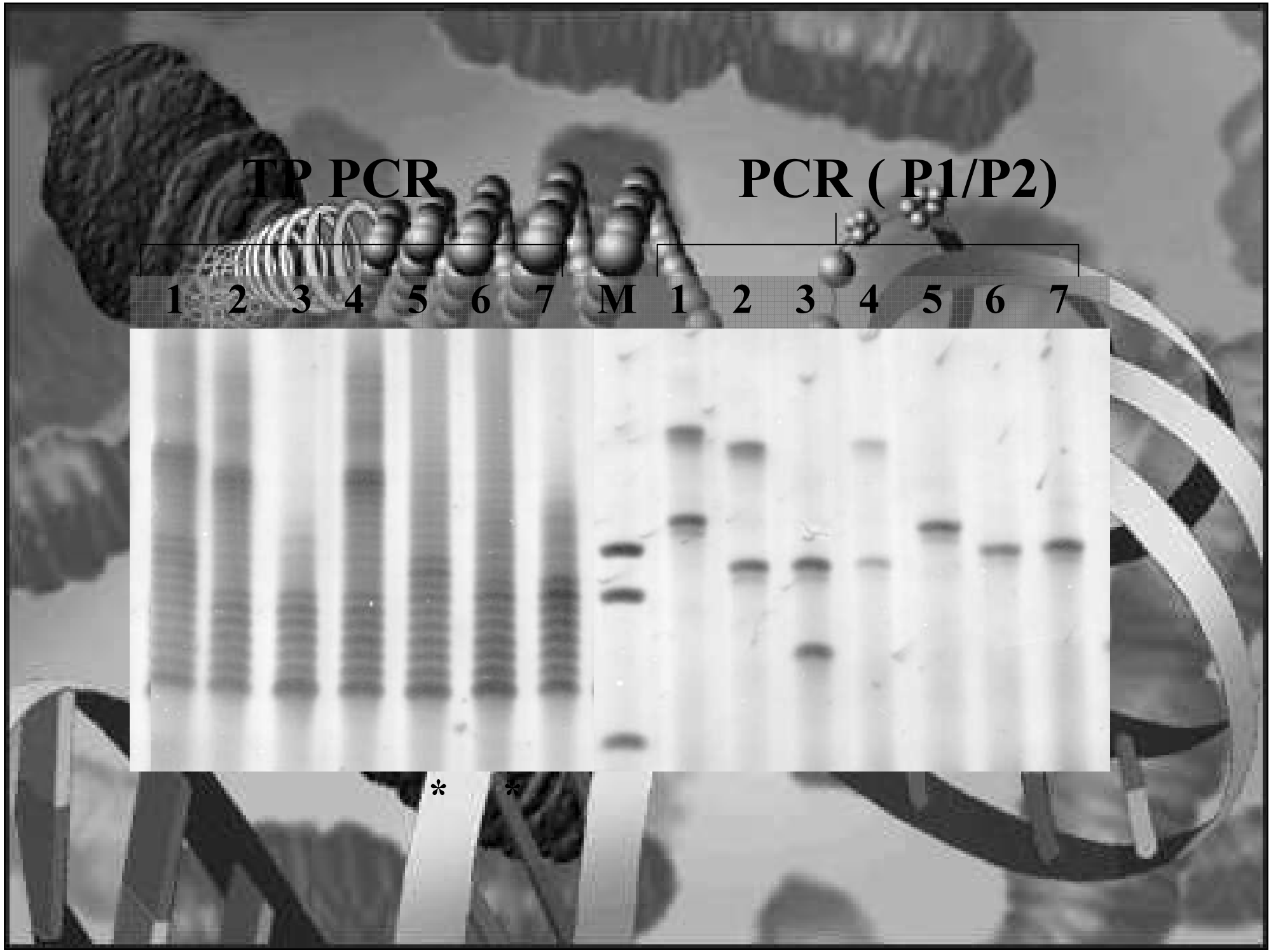




Triplet Primer PCR



Strategie molekulárně genetického vyšetření MD1


TP PCR
PCR (P1/P2)
1 2 3 4 5 6 7 M 1 2 3 4 5 6 7
* *