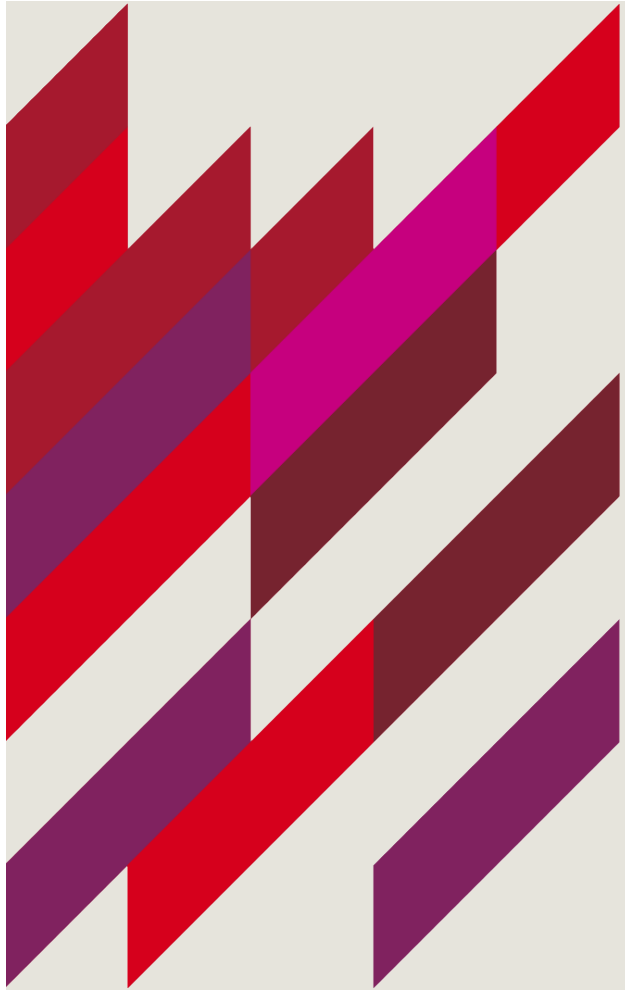
MACQUARIE University

The Human Genome

BIOL3120 Human Genetics and Evolutionary Medicine

The Human Genome

LEARNING OBJECTIVES

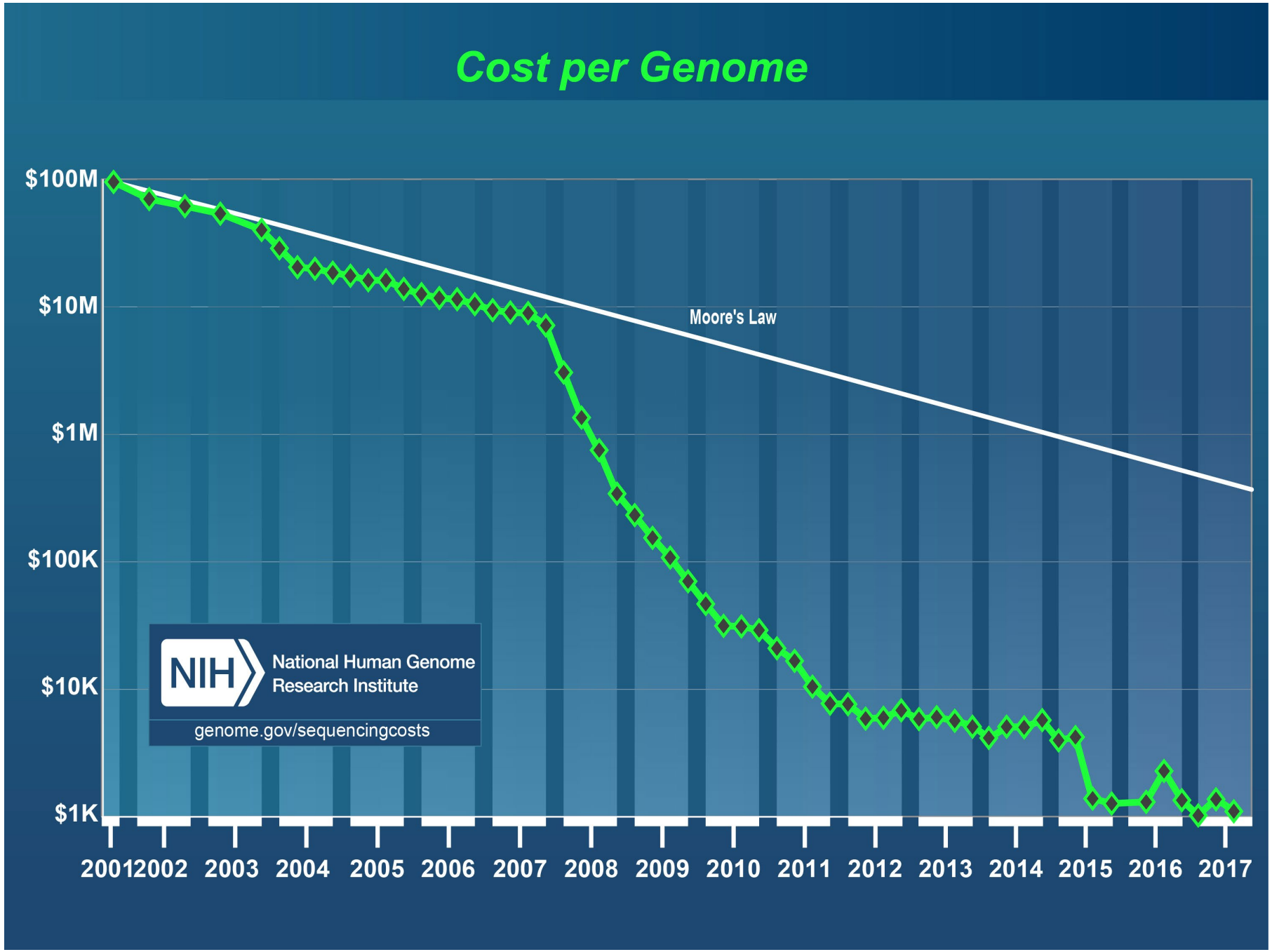


At the end of this lecture you should be able to:

- Define the terms used to describe the human genome
- Understand the structure of chromosomes
- Understand the coding regions of the human genome

The human genome

- First draft sequence February 2001
- Complete draft 2003
- Cost ~ 2.7 billion USD


Cost per Genome
$100M
$10M
$1M
$100K
$10K
$1K
Moore's Law
NIH
National Human Genome Research Institute
genome.gov/sequencingcosts
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017

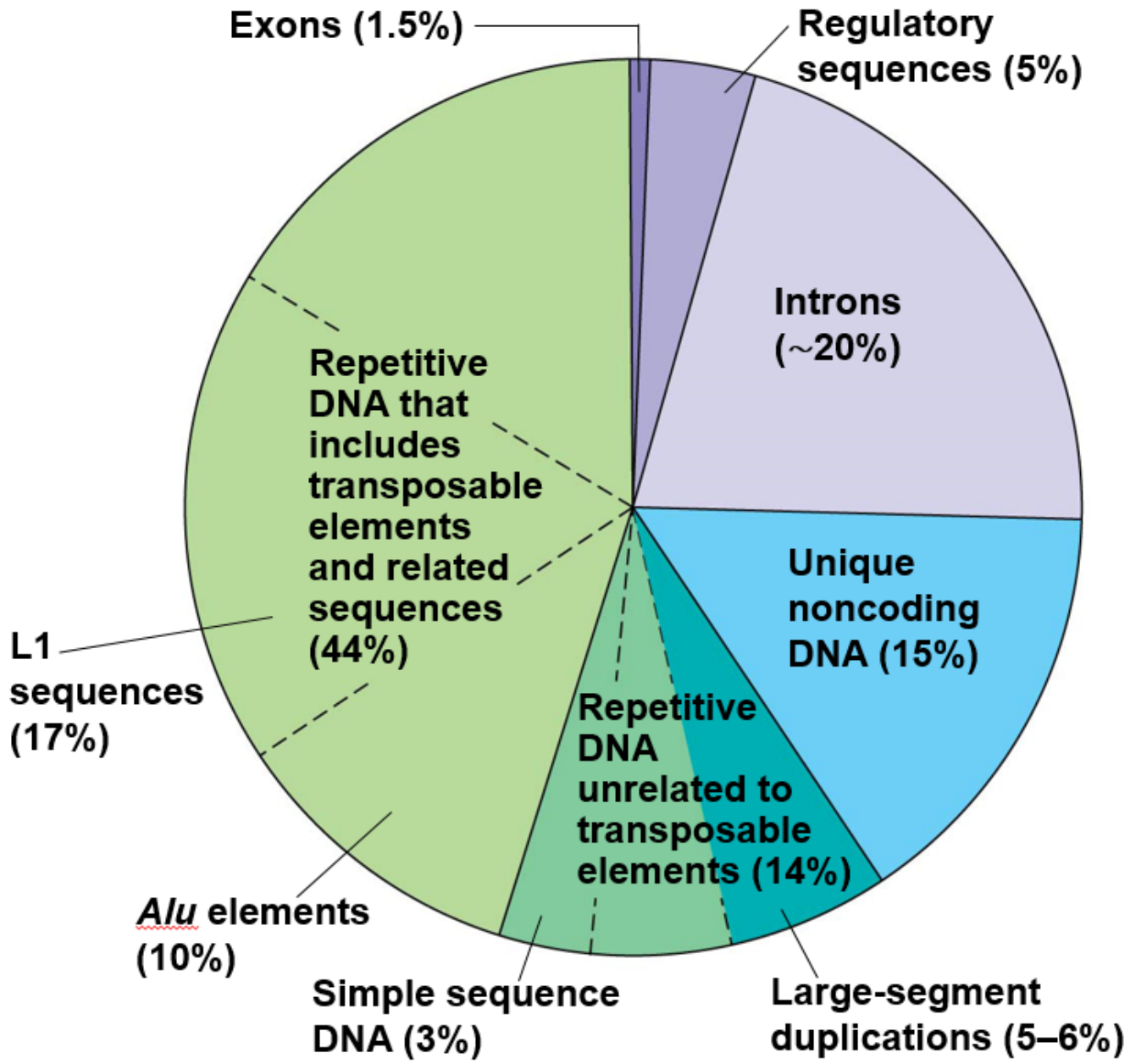
The human genome

- First draft sequence February 2001
- Complete draft 2003
- Cost ~ 2.7 billion USD

- Sequence determined but little understanding of its meaning

Quick overview

- 3.2 billion basepairs x 2
- Over 23 chromosome pairs
- ~20,000 protein-coding genes

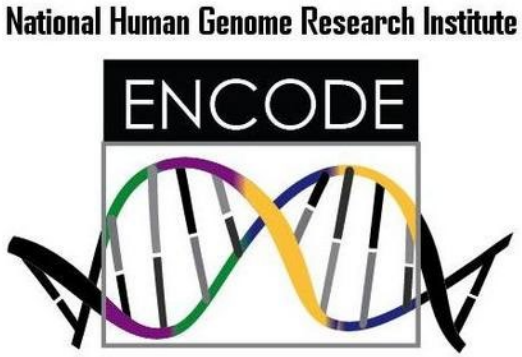


Junk DNA?

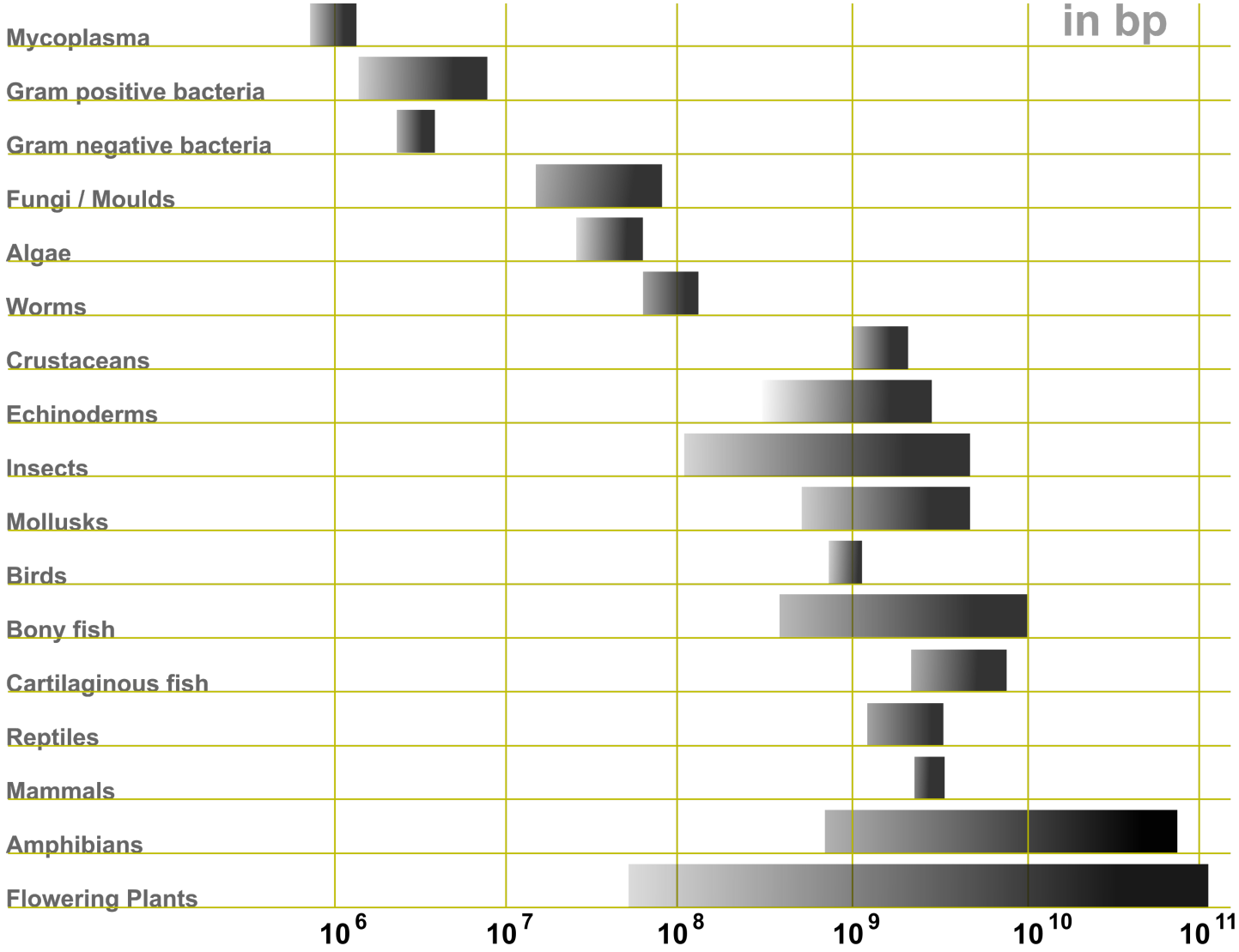
- Historically, non-coding DNA referred to as junk DNA

- ENCODE project (2012):
  - “The vast majority (80.4%) of the human genome participates in at least one biochemical RNA- and/or chromatin-associated event in at least one cell type”
  - (think about gene definition)

  - https://www.nature.com/articles/nature11247

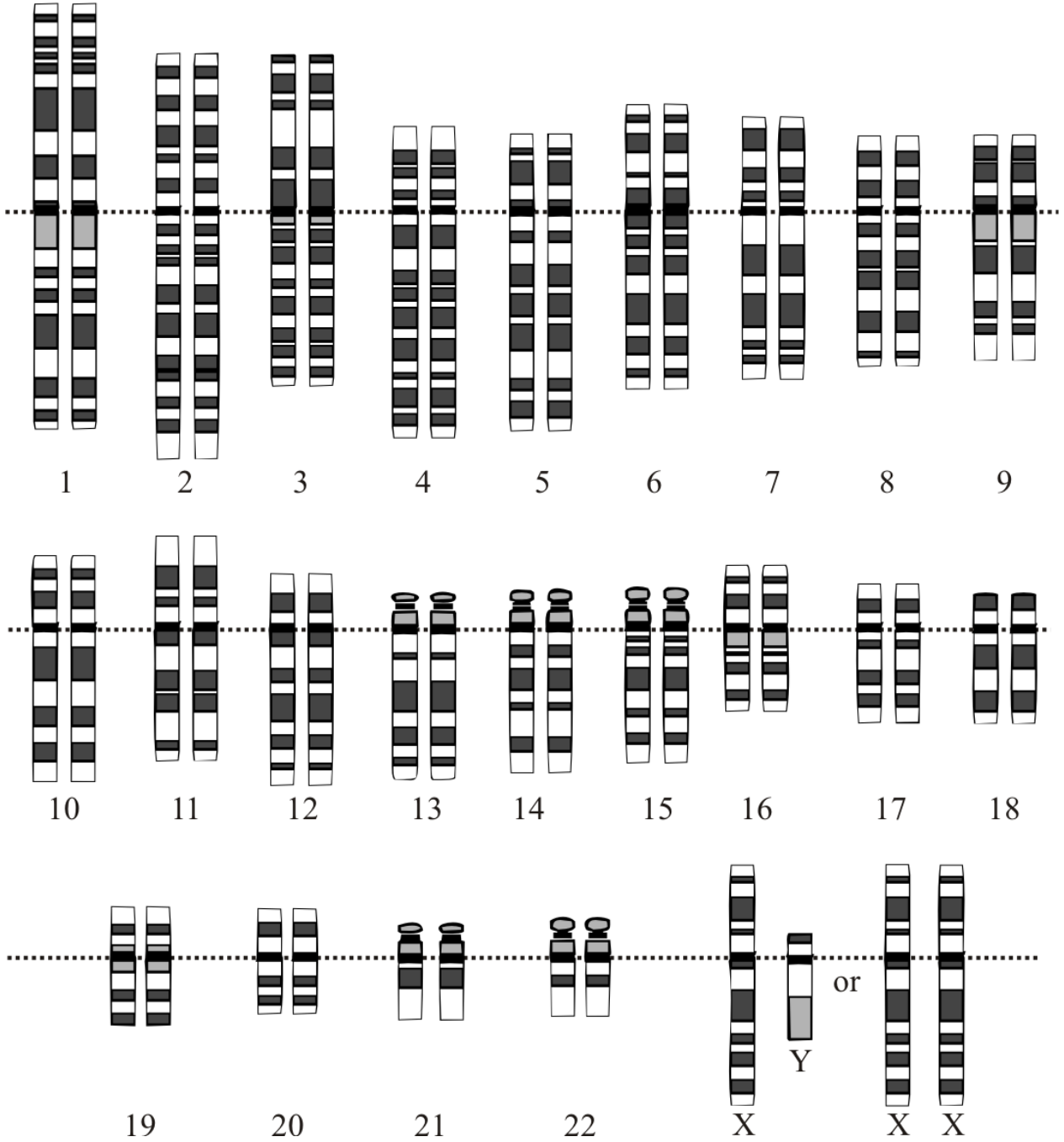


Example genome sizes



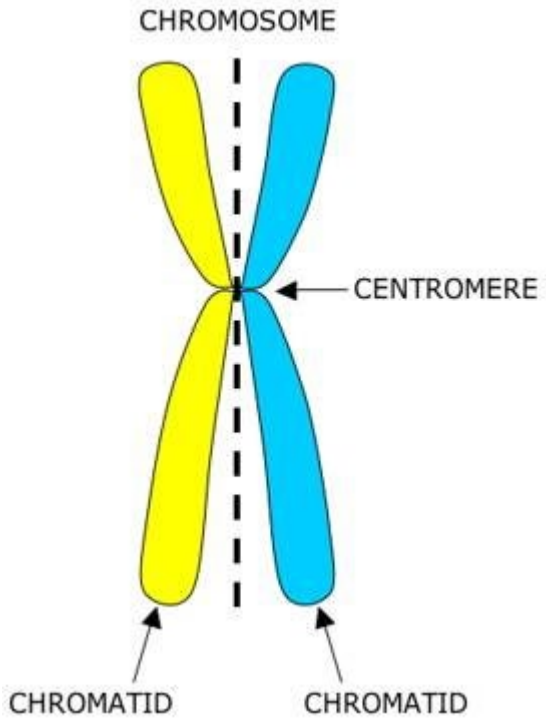
Terms used to describe the human genome

Chromosomes



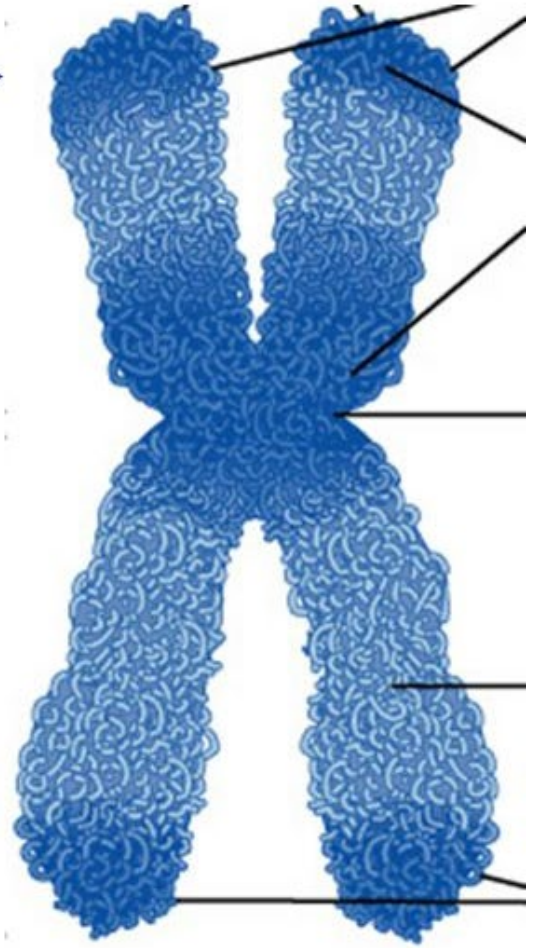
Chromosome related nomenclature

- Chromatin: the material of which chromosomes are made.
  - complex of DNA, and proteins (histones and non-histones)
- Chromatid: one of the two copies of a chromosome
  - joined at the centromere
  - two homologous chromatids are called sister chromatids.
- Centromere is DNA normally in a heterochromatin state
  - Site of spindle attachment to chromosomes (important in meiosis)

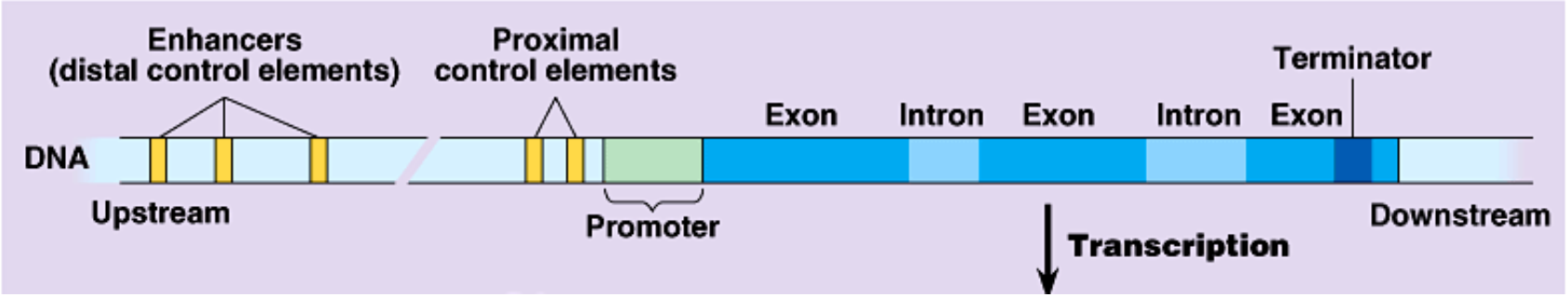


Chromosome parts

- Heterochromatin:
  - More condensed
  - Silenced genes (methylated)
  - Gene poor (high AT content)
  - Stains darker
- Euchromatin:
  - Less condensed
  - Gene expressing
  - Gene rich (higher GC content)
  - Stains lighter



DNA & Gene parts

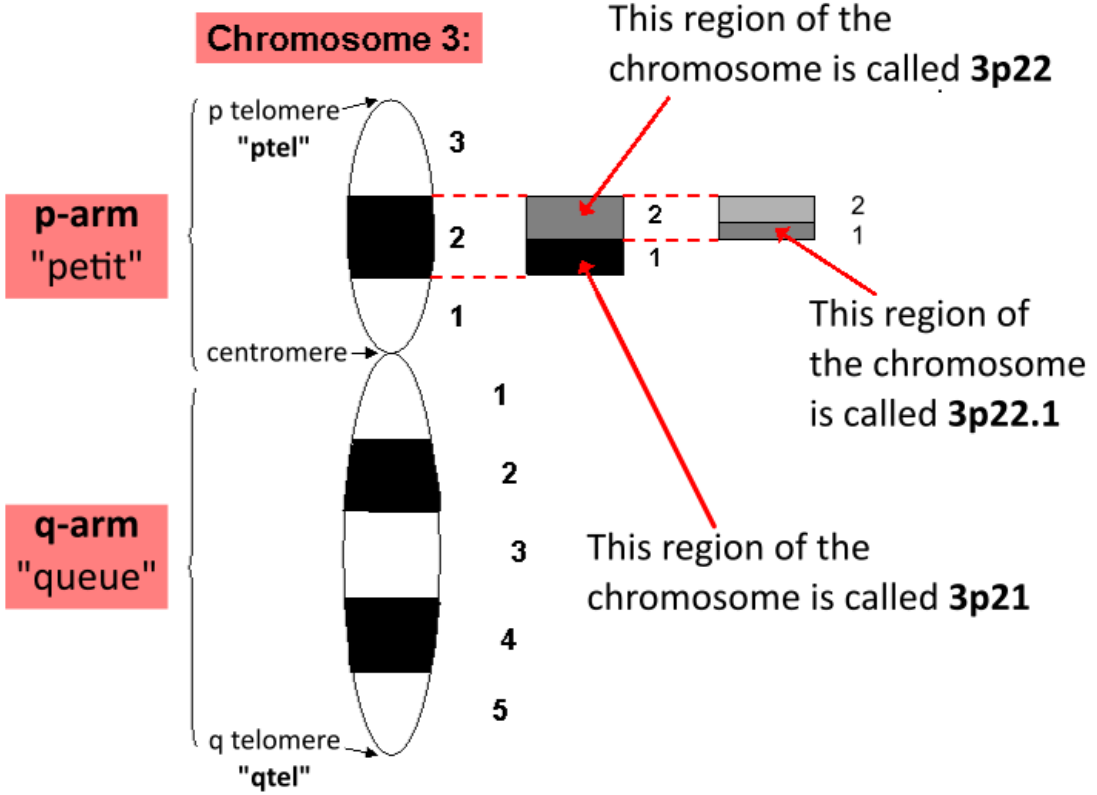


- Enhancers: genomic DNA (gDNA) which can alter the expression of a gene
- Promoters: gDNA region where transcription factors bind to initiate transcription of a gene
- Exons: gDNA that is the coding regions of a gene
- Introns: a portion of a gene that does not code for an amino acid
- Gene: a sequence of nucleotides in DNA that codes for a molecule that has a function

The Structure of Chromosomes

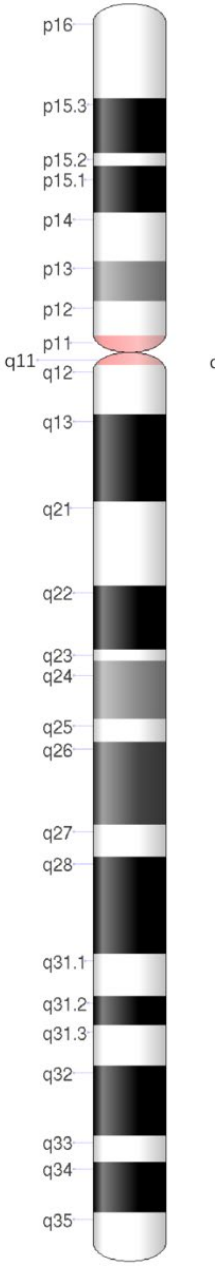
Chromosome numbering

- 1-22 Arranged in order of decreasing size
- q = long arm; p = short arm



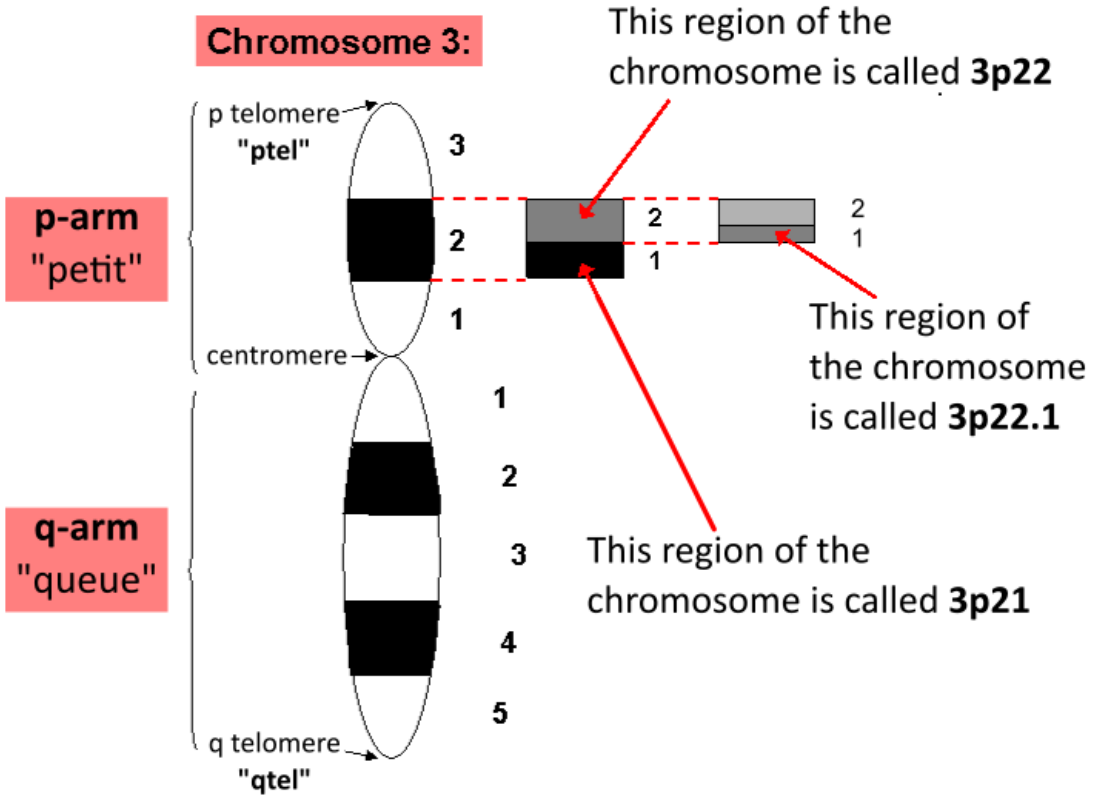
G-banding example: Chromosome 4

- Giemsa stain
  - darker on heterochromatin (AT rich) regions, less accessible
  - Lighter on euchromatin (GC rich) regions, more accessible and active
  - 92% of genes in euchromatin in humans



Chromosome numbering

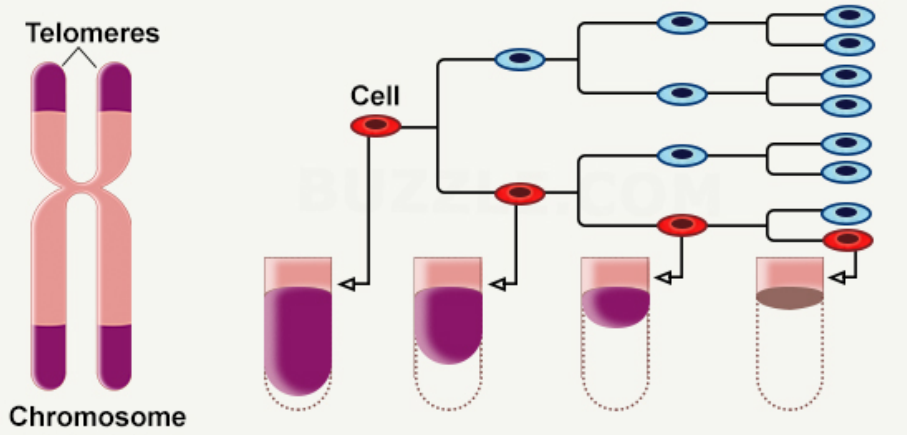
- 1-22 Arranged in order of decreasing size
- q = long arm; p = short arm
- Black and white stripes are because of staining: G-banding
- ‘Bands’ have the same appearance on homologous chromosomes so identification became easier
- Banding allows:
  - Unequivocal identification of each human chromosome
  - Detection of rearrangements (e.g. translocations)

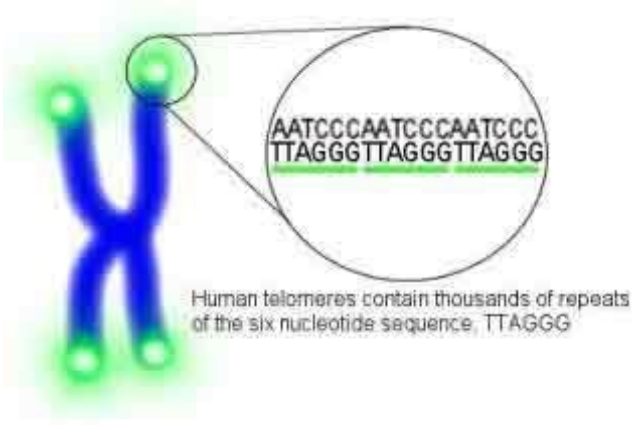


Chromosomes

Telomeres

- protect functional regions (containing genes) by capping the ends with a ‘buffer’ sequence
- Telomeres consist of variable numbers of a repetitive sequence (TTAGGG)
- Telomere length is reduced every time a cell divides
- average telomere length declines from ~ 11 kilobases at birth to less than 4 kilobases in old age, decline greater in men
- Marker of cell age – signal for apoptosis (cell self-destruction)
- **“telomeric** DNA is **lost** at an **average rate** of '25.7–27.7 base pairs' per year



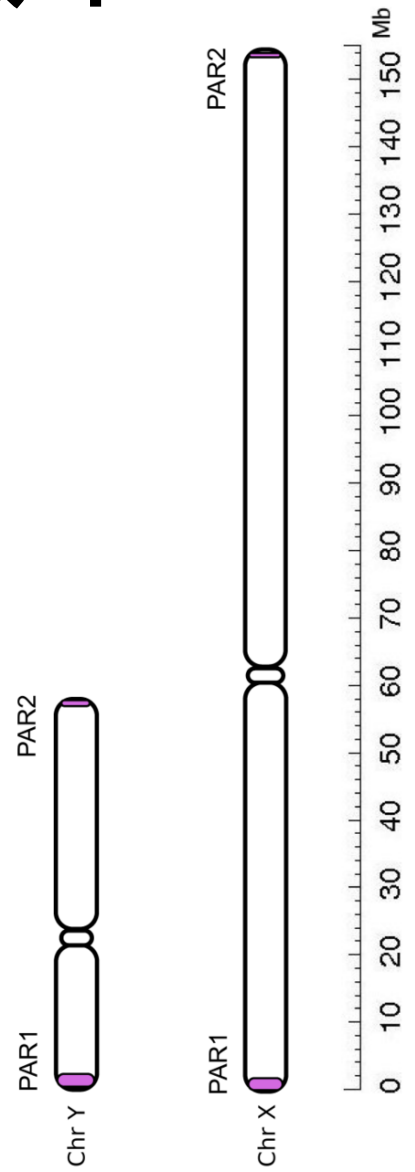


Telomeres

- Various premature ageing disease characterised by critically short telomeres
- Cancer cells often do not show telomere length shortening - telomerase protein not normally active has become active in many cancer cell types
- Telomerase and telomere-binding proteins are potential targets for anti-cancer and anti-ageing therapies

Pseudoautosomal regions on X & Y chromosomes

- Homologous regions at ends of X and Y chromosomes
  - Allows pairing during meiosis
  - Regions tend not to be silenced in X-inactivation
  - Inherited in an autosomal manner
  - These regions are called PAR1 and PAR2

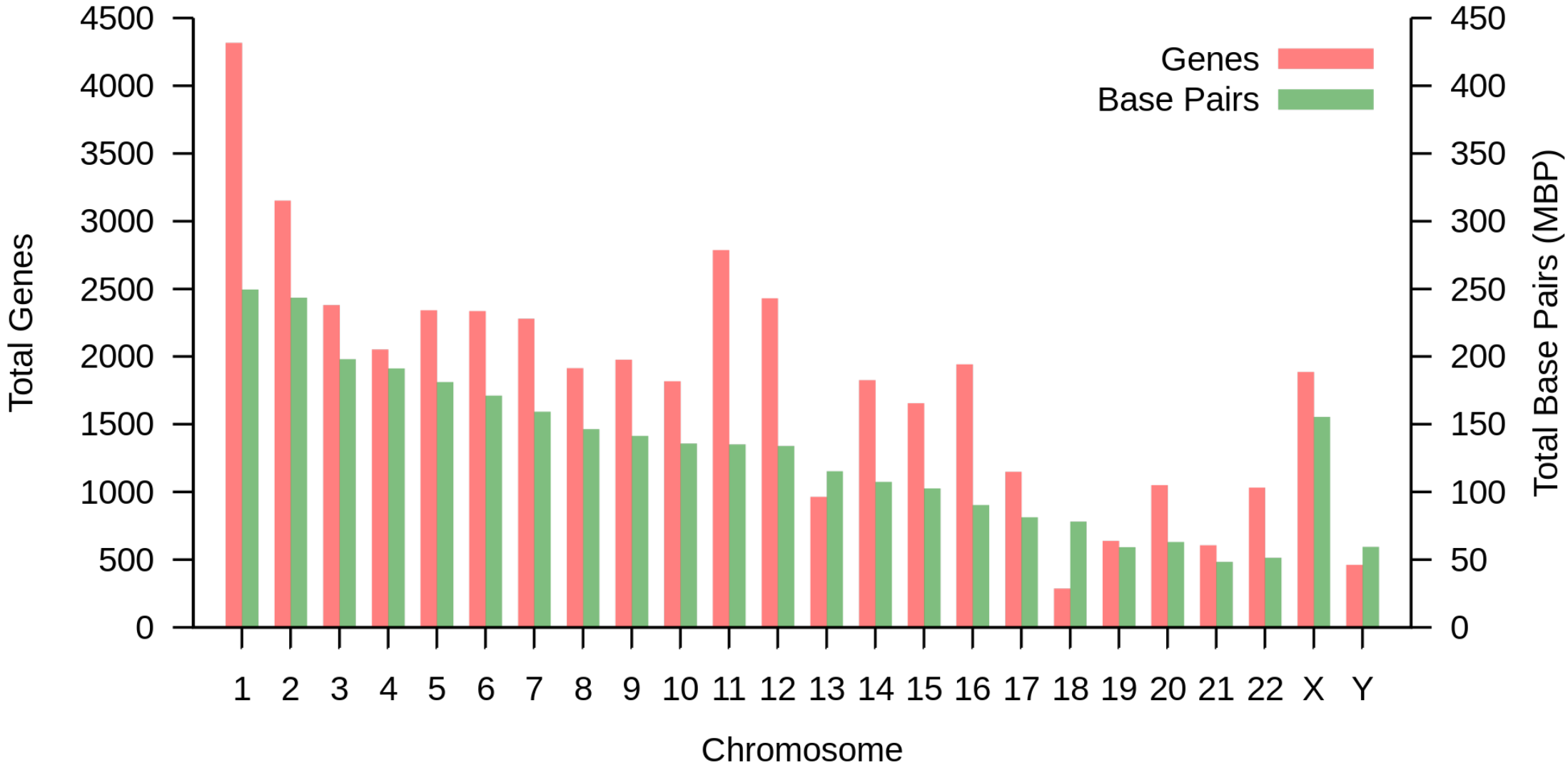


Number of genes on each chromosome

Chromosome	Length (mm)	Base pairs	Variations	Protein-coding genes	Pseudo-genes	Total long ncRNA	Total small ncRNA	miRNA	rRNA	snRNA	snoRNA	Misc ncRNA	Links	Centromere position (Mbp)	Cumulative (%)
1	85	248,956,422	12,151,146	2058	1220	1200	496	134	66	221	145	192	EBI	125	7.9
2	83	242,193,529	12,945,965	1309	1023	1037	375	115	40	161	117	176	EBI	93.3	16.2
3	67	198,295,559	10,638,715	1078	763	711	298	99	29	138	87	134	EBI	91	23
4	65	190,214,555	10,165,685	752	727	657	228	92	24	120	56	104	EBI	50.4	29.6
5	62	181,538,259	9,519,995	876	721	844	235	83	25	106	61	119	EBI	48.4	35.8
6	58	170,805,979	9,130,476	1048	801	639	234	81	26	111	73	105	EBI	61	41.6
7	54	159,345,973	8,613,298	989	885	605	208	90	24	90	76	143	EBI	59.9	47.1
8	50	145,138,636	8,221,520	677	613	735	214	80	28	86	52	82	EBI	45.6	52
9	48	138,394,717	6,590,811	786	661	491	190	69	19	66	51	96	EBI	49	56.3
10	46	133,797,422	7,223,944	733	568	579	204	64	32	87	56	89	EBI	40.2	60.9
11	46	135,086,622	7,535,370	1298	821	710	233	63	24	74	76	97	EBI	53.7	65.4
12	45	133,275,309	7,228,129	1034	617	848	227	72	27	106	62	115	EBI	35.8	70
13	39	114,364,328	5,082,574	327	372	397	104	42	16	45	34	75	EBI	17.9	73.4
14	36	107,043,718	4,865,950	830	523	533	239	92	10	65	97	79	EBI	17.6	76.4
15	35	101,991,189	4,515,076	613	510	639	250	78	13	63	136	93	EBI	19	79.3
16	31	90,338,345	5,101,702	873	465	799	187	52	32	53	58	51	EBI	36.6	82
17	28	83,257,441	4,614,972	1197	531	834	235	61	15	80	71	99	EBI	24	84.8
18	27	80,373,285	4,035,966	270	247	453	109	32	13	51	36	41	EBI	17.2	87.4
19	20	58,617,616	3,858,269	1472	512	628	179	110	13	29	31	61	EBI	26.5	89.3
20	21	64,444,167	3,439,621	544	249	384	131	57	15	46	37	68	EBI	27.5	91.4
21	16	46,709,983	2,049,697	234	185	305	71	16	5	21	19	24	EBI	13.2	92.6
22	17	50,818,468	2,135,311	488	324	357	78	31	5	23	23	62	EBI	14.7	93.8
X	53	156,040,895	5,753,881	842	874	271	258	128	22	85	64	100	EBI	60.6	99.1
Y	20	57,227,415	211,643	71	388	71	30	15	7	17	3	8	EBI	12.5	100
mtDNA	0.0054	16,569	929	13	0	0	24	0	2	0	0	0	EBI	N/A	100
total		3,088,286,401	155,630,645	20412	14600	14727	5037	1756	532	1944	1521	2213			

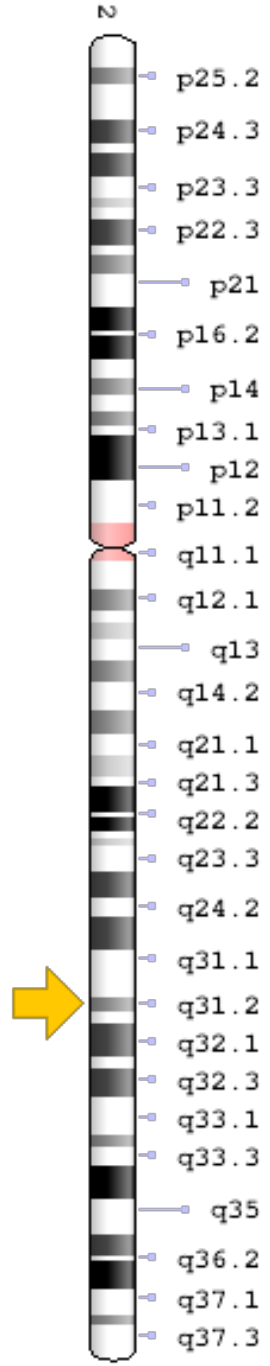
https://en.wikipedia.org/wiki/Human_genome

Chromosome size vs number of genes



Gene size in humans

- Titin gene (TTN)
  - 365,719bp total gene size (including introns)
  - Exons: 80,781bp across 365 exons
- DMD gene (Duchenne muscular dystrophy)
  - 2,220,390 bp including introns
  - ~14kb mRNA
- Average gene size ~8kb

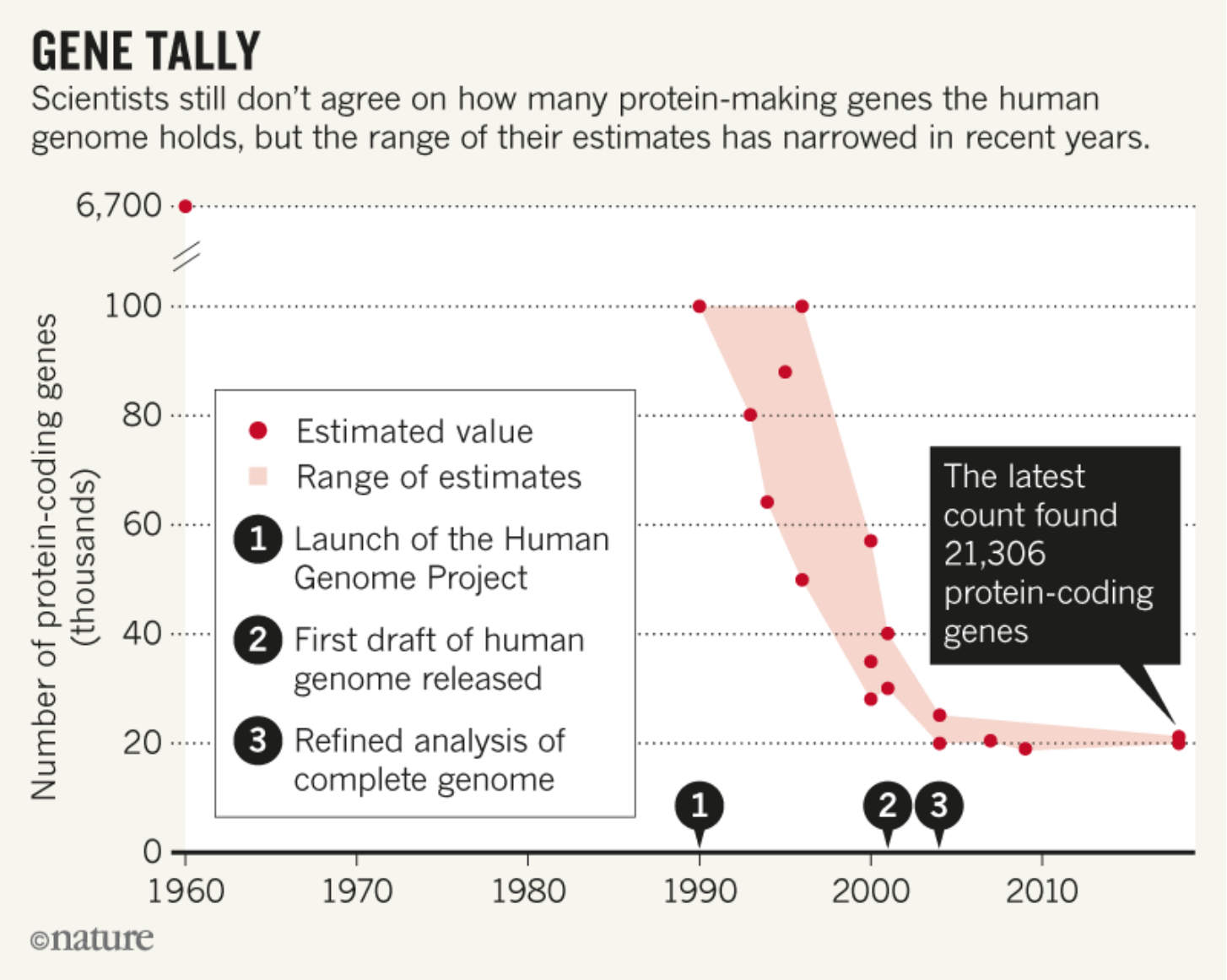


Coding Regions Human Genome

Human genome summary

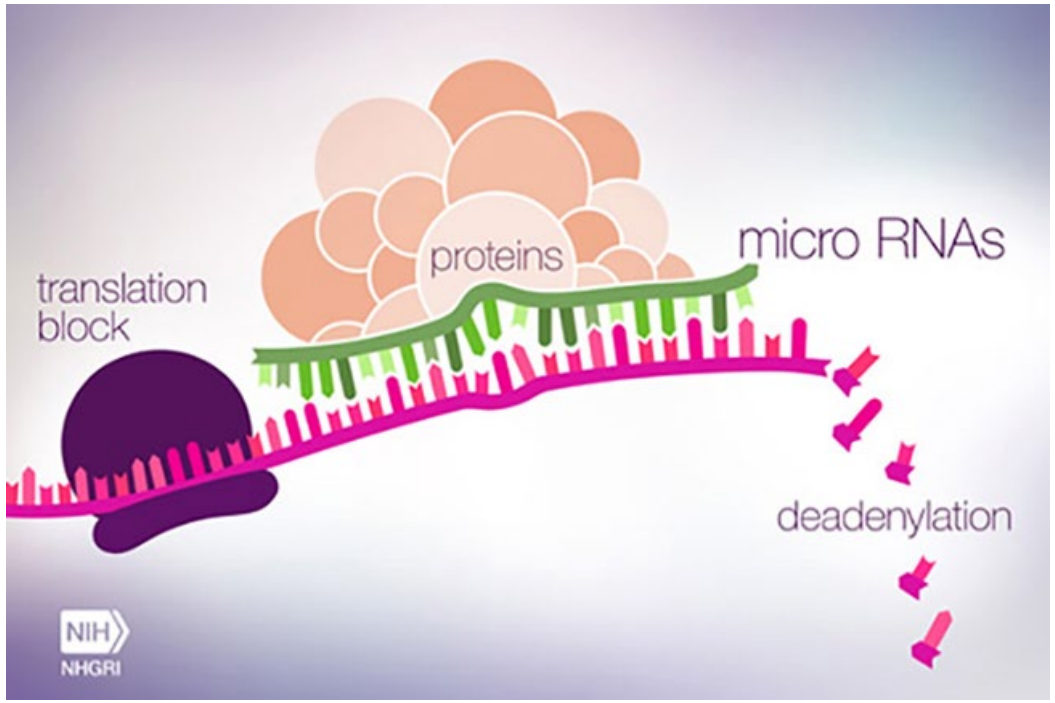
- 3.2 billion base pairs x 2 across 23 chromosome pairs
- ~20,000 protein coding genes, but other functional molecules
- Only 1-2% codes for protein, but up to 80% of genome regulatory function

How many genes?

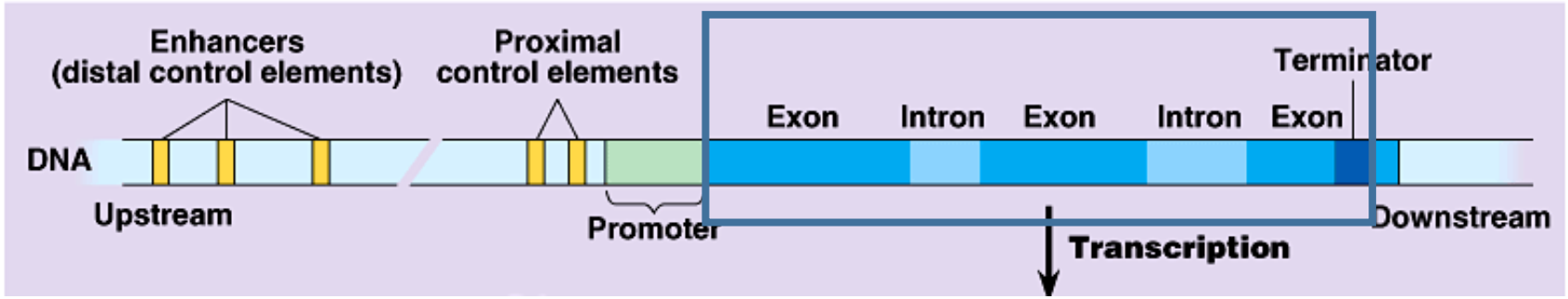


What is a gene?

- Classically, region of DNA which codes for a protein
- Nowadays:
  - a sequence of nucleotides in DNA that codes for a molecule that has a function
  - This might code for a protein, or just be transcribed into RNA e.g. t-RNAs, non-coding RNAs

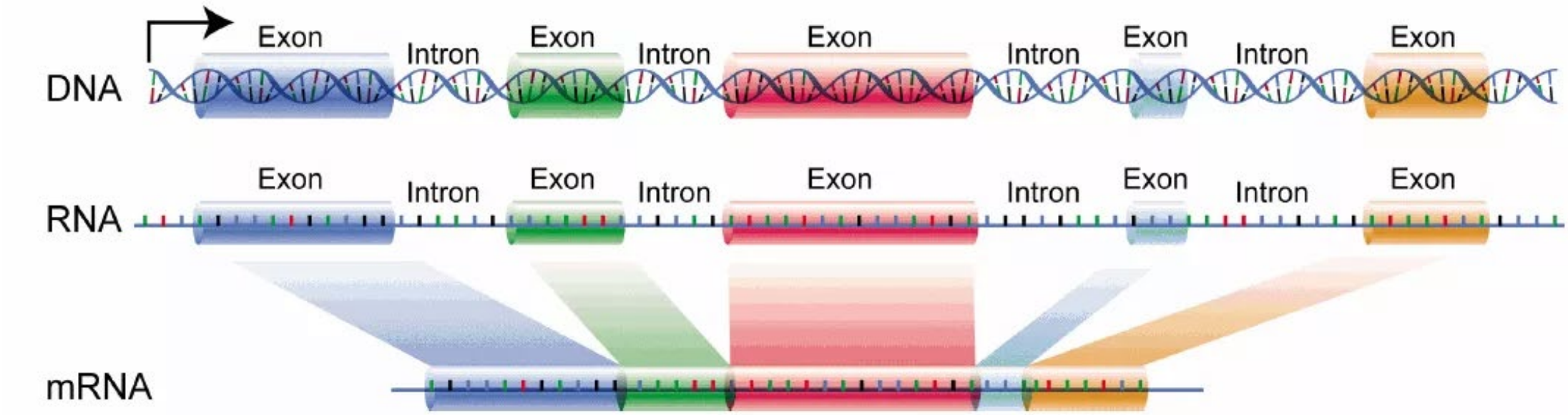


What is a gene?



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

Exons and introns

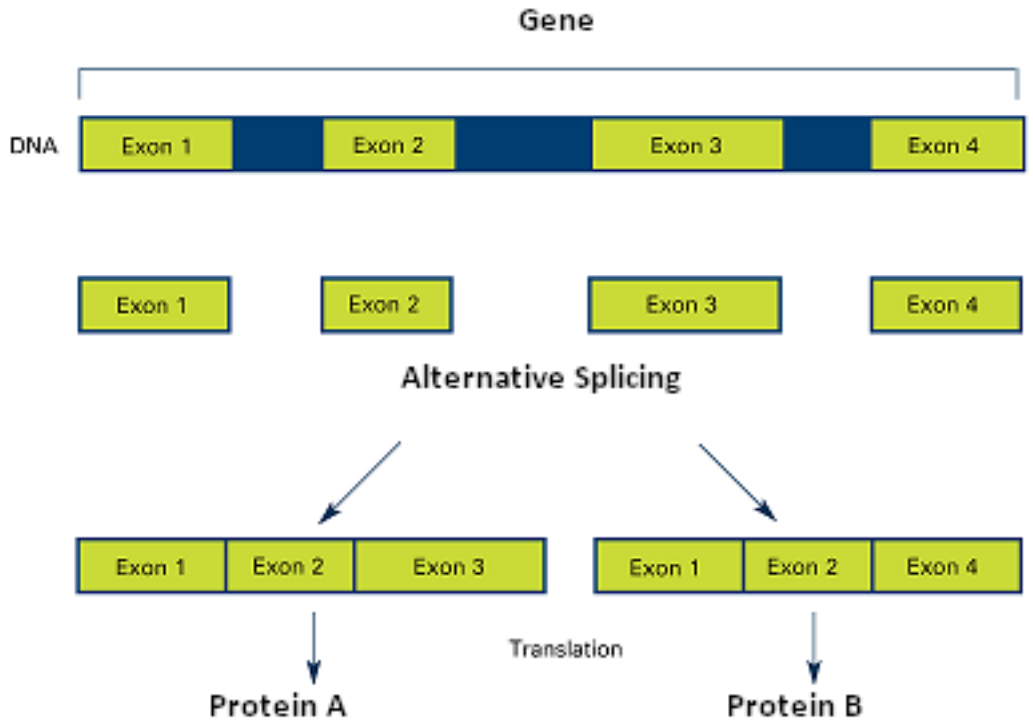


- Exon: part of a gene that will encode a part of the final mature RNA
- Intron: any nucleotide sequence within a gene that is removed by RNA splicing during maturation of the final RNA product
- 1.1% of the genome is spanned by exons, whereas 24% is in introns, with 75% of the genome being intergenic DNA

Multiple transcripts / alternate splicing

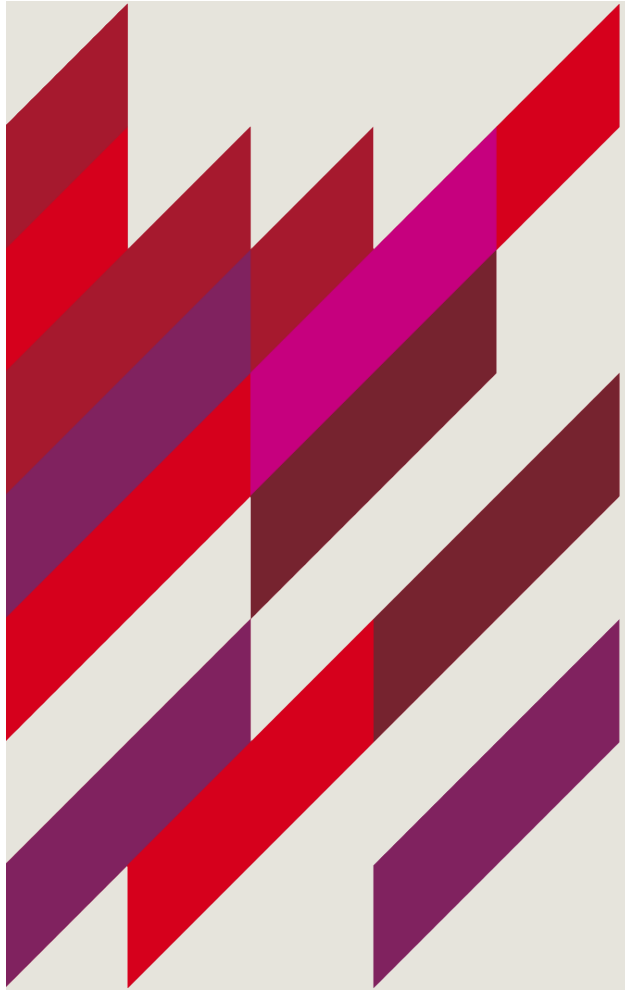
- Exons can be alternatively arranged to create different proteins from the same gene

- Different types of cells often use different transcripts more/less frequently



The Human Genome

LEARNING OBJECTIVES



At the end of this lecture you should be able to:

- Define the terms used to describe the human genome
- Understand the structure of chromosomes
- Understand the coding regions of the human genome