

TABLE 9.5 HUMAN GENOME AND HUMAN GENE STATISTICS

SIZE OF GENOME COMPONENTS	
Mitochondrial genome	16.6 kb
Nuclear genome	3.1 Gb ^a
Euchromatic component	2.9 Gb (~93%)
Highly conserved fraction	~150 Mb (~5%)
Protein-coding DNA sequences	~35 Mb (~1.1%)
Other highly conserved DNA	~115 Mb (~3.9%)
Segmentally duplicated DNA	~160 Mb (~5.5%)
Highly repetitive DNA	~1.6 Gb (~50%)
Constitutive heterochromatin	~200 Mb (~7%; Table 9.3)
Transposon-based repeats	~1.4 Gb (~45%; Table 9.12)
DNA per chromosome	48 Mb—249 Mb (Table 9.3)
GENE NUMBER	
Nuclear genome	> 26,000
Mitochondrial genome	37
Protein-coding genes	~ 20,000–21,000
RNA genes	> 6000 (exact figure not known)
Pseudogenes related to protein-coding genes	> 12,000
GENE DENSITY	
Nuclear genome	>1 per 120 kb (but considerable uncertainty)
Mitochondrial genome	1 per 0.45 kb
LENGTH OF PROTEIN-CODING GENES	
Average length	53.6 kb
Smallest	a few hundred base pairs long (several examples)
Largest	2.4 Mb (dystrophin)
EXON NUMBER IN PROTEIN-CODING GENES	
Average number of exons in one gene ^b	9.8
Largest number in one gene	363 (in the titin gene)
Smallest number in one gene	1 (no introns—see Tables 9.4 and 9.7 for example)
EXON SIZE IN PROTEIN-CODING GENES	
Average exon size	288 bp (but exons at 3' end of genes tend to be large)
Smallest	< 10 bp (various; e.g. exon 3 of the troponin I gene <i>TNNI1</i> is just 4 bp long)
Largest	18.2 kb (exon 6 in <i>MUC16</i> isoform-201)
INTRON SIZE IN PROTEIN-CODING GENES	
Smallest	< 30 bp (various)
Largest	1.1 Mb (intron 5 in <i>KCNIP4</i>)
RNA SIZE	
Smallest noncoding RNA	< 20 nucleotides (e.g. many transcriptional start site-associated RNAs are 18 nucleotides)
Largest noncoding RNA	Several hundred thousand nucleotides; e.g. <i>UBE3A</i> antisense RNA is likely to be close to 1 Mb
Largest mRNA	> 103 kb (titin mRNA, NF-2A isoform)
POLYPEPTIDE SIZE	
Smallest	tens of amino acids (various neuropeptides)
Largest	34,350 amino acids (titin, NF-2A isoform)

^aNote that the total size can vary between haplotypes because of copy-number variation. ^bFor the longest isoform. Data were obtained largely from ENSEMBL release 55 datasets.