

# The human reference genome

[www.ensembl.org/info/data/ftp/index.html](http://www.ensembl.org/info/data/ftp/index.html)

<https://www.gencodegenes.org/human/>

## Single species data

Popular species are listed first. You can customise this list via our [home page](#)

Show 10 entries							Show/hide columns
★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	
Y	<a href="#">Human</a> <i>Homo sapiens</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	
Y	<a href="#">Mouse</a> <i>Mus musculus</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	
Y	<a href="#">Zebrafish</a> <i>Danio rerio</i>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	<a href="#">FASTA</a>	

## Fasta files

Content	Regions	Description	Download
Transcript sequences	CHR	<ul style="list-style-type: none"><li>Nucleotide sequences of all transcripts on the reference chromosomes</li></ul>	<a href="#">Fasta</a>
Protein-coding transcript sequences	CHR	<ul style="list-style-type: none"><li>Nucleotide sequences of coding transcripts on the reference chromosomes</li><li>Transcript biotypes: protein_coding, nonsense_mediated_decay, non_stop_decay, IG_*_gene, TR_*_gene, polymorphic_pseudogene</li></ul>	<a href="#">Fasta</a>
Protein-coding transcript translation sequences	CHR	<ul style="list-style-type: none"><li>Amino acid sequences of coding transcript translations on the reference chromosomes</li><li>Transcript biotypes: protein_coding, nonsense_mediated_decay, non_stop_decay, IG_*_gene, TR_*_gene, polymorphic_pseudogene</li></ul>	<a href="#">Fasta</a>
Long non-coding RNA transcript sequences	CHR	<ul style="list-style-type: none"><li>Nucleotide sequences of long non-coding RNA transcripts on the reference chromosomes</li></ul>	<a href="#">Fasta</a>
Genome sequence (GRCh38.p12)	ALL	<ul style="list-style-type: none"><li>Nucleotide sequence of the GRCh38.p12 genome assembly version on all regions, including reference chromosomes, scaffolds, assembly patches and haplotypes</li><li>The sequence region names are the same as in the GTF/GFF3 files</li></ul>	<a href="#">Fasta</a>
Genome sequence, primary assembly (GRCh38)	PRI	<ul style="list-style-type: none"><li>Nucleotide sequence of the GRCh38 primary genome assembly (chromosomes and scaffolds)</li><li>The sequence region names are the same as in the GTF/GFF3 files</li></ul>	<a href="#">Fasta</a>

 [Homo sapiens.GRCh38.dna.primary\\_assembly.fa.gz](#)

840 MB