

Ch 1 – L 2.2

A first look at Genome Browsers

Introduction to *GENOME BROWSERS*

Genomic database have developed a way to «see» genes and sequences

ENSEMBL -

NCBI - Gene (<https://www.ncbi.nlm.nih.gov/gene/?term=>)

UCSC -

UWASH

and others

NCBI is National Center for Biotechnological Information

is based in the National Library of Medicine at NIH
(National Institutes of Health)

USA – It is a public domain (still.... Trump permitting)

The National Institutes of Health

<https://www.nih.gov/>

The National Library of Medicine

<https://www.nlm.nih.gov/>

The National Center for Biotechnological Information

<https://www.ncbi.nlm.nih.gov/>







EMBO – European Molecular Biology Organization
EBI – European Bioinformatics Institute



Europe's flagship laboratory for the life sciences

EMBL is at the forefront of innovation in life sciences research, technology development and transfer, and provides outstanding training and services to the scientific community in its member states. This publicly-funded non-profit institute is housed at six sites in Europe whose expertise covers the whole spectrum of molecular biology.

Locations

 <p>EMBL Heidelberg Germany MAIN LABORATORY / GENERAL INFORMATION</p>	 <p>EMBL Barcelona Spain TISSUE BIOLOGY AND DISEASE MODELLING</p>	 <p>EMBL Grenoble France STRUCTURAL BIOLOGY</p>
 <p>EMBL Hamburg Germany STRUCTURAL BIOLOGY</p>	 <p>EMBL-EBI Hinxton United Kingdom EUROPEAN BIOINFORMATICS INSTITUTE</p>	 <p>EMBL Rome Italy EPIGENETICS AND NEUROBIOLOGY</p>

Tools

[All tools](#)

BioMart >

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Search

All species for

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

All genomes

-- Select a species --

- [View full list of all Ensembl species](#)
- [Edit your favourites](#)

Favourite genomes



Human
GRCh38.p13

[Still using GRCh37?](#)



Mouse
GRCm38.p6



Zebrafish
GRCz11

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 99 (January 2020)

- Update to GENCODE 33 for human
- Update to dbSNP153 for human
- Import of updated VISTA enhancers for human and mouse
- New genomes: 10 mammals (including 2 dog breeds), 11 birds, 15 fish and 4 reptiles
- Updated genome assemblies: zebra finch, fugu, Nile tilapia and Asian bonytongue

[More release news](#) on our blog

Other news from our blog





- 28 Feb 2020: [Cool stuff the Ensembl VEP can do: install using Docker](#)
- 26 Feb 2020: [What's coming in Ensembl 100 / Ensembl Genomes 47](#)
- 25 Feb 2020: [Bug report: GRCh37 BLAT queries between 21st January – 20th February 2020](#)

Search Human (*Homo sapiens*)

Search all categories ▾ Search Human...

e.g. BRCA2 or 17:63992802-64038237 or rs699 or osteoarthritis

Genome assembly: GRCh38.p13 (GCA_000001405.28)

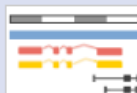
-  [More information and statistics](#)
-  [Download DNA sequence \(FASTA\)](#)
-  [Convert your data to GRCh38 coordinates](#)
-  [Display your data in Ensembl](#)

Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart



[View karyotype](#)



[Example region](#)

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

-  [More about this genebuild](#)
-  [Download FASTA files for genes, cDNAs, ncRNA, proteins](#)
-  [Download GTF or GFF3 files for genes, cDNAs, ncRNA, proteins](#)
-  [Update your old Ensembl IDs](#)





[Example gene](#)



[Example transcript](#)

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.




-  [More about comparative analysis](#)
-  [Download alignments \(EMF\)](#)



[Example gene tree](#)

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

-  [More about variation in Ensembl](#)
-  [Download all variants \(GVF\)](#)
-  [Variant Effect Predictor](#)



[Example variant](#)



[Example phenotype](#)

Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.



Gene counts (Primary assembly)

<u>Coding genes</u>	20,449 (incl 657 <u>readthrough</u>)
Non coding genes	23,992
Small non coding genes	4,871
Long non coding genes	16,900 (incl 308 <u>readthrough</u>)
Misc non coding genes	2,221
<u>Pseudogenes</u>	15,220 (incl 8 <u>readthrough</u>)
<u>Gene transcripts</u>	227,980

What is «readthrough» ?

Gene counts (Alternative sequence)

<u>Coding genes</u>	3,054 (incl 44 <u>readthrough</u>)
Non coding genes	1,555
Small non coding genes	297
Long non coding genes	1,071 (incl 45 <u>readthrough</u>)
Misc non coding genes	187
<u>Pseudogenes</u>	1,798
<u>Gene transcripts</u>	21,636

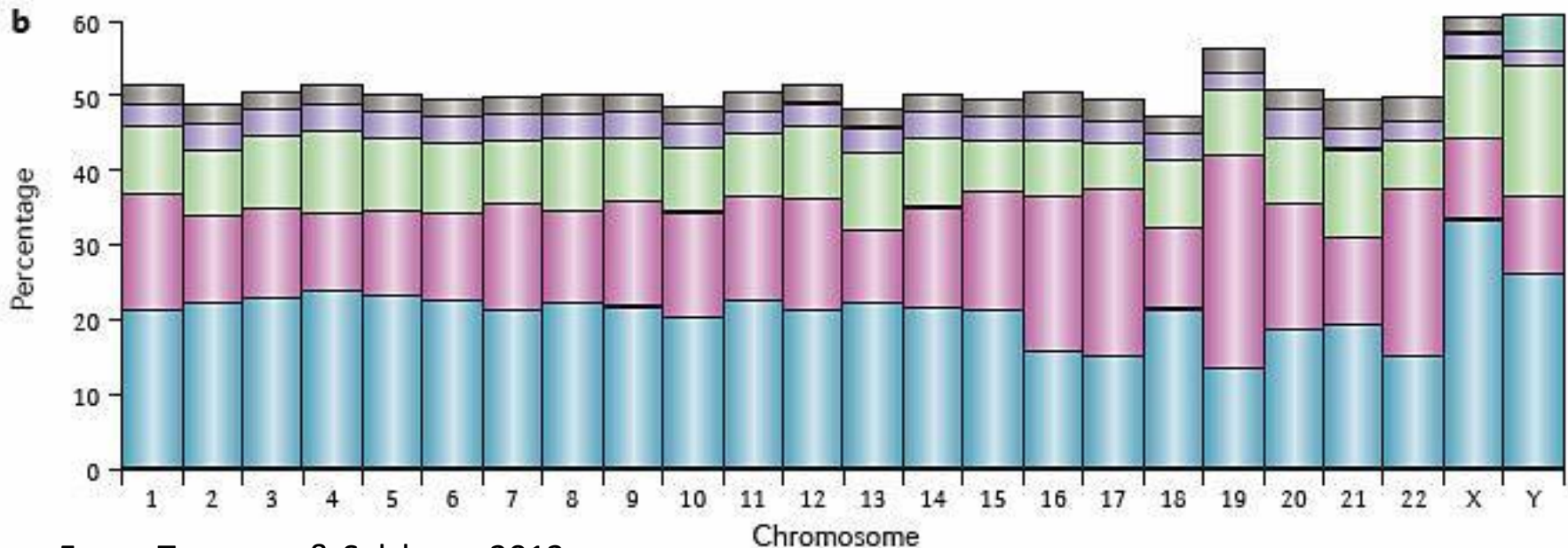
Other

Genscan gene predictions	51,756
Short Variants	678,017,608
Structural variants	6,073,012

Repetitive sequences cover nearly half of the Human Genome

a

Repeat class	Repeat type	Number (hg19)	Cvg	Length (bp)
Minisatellite, microsatellite or satellite	Tandem	426,918	3%	2–100
SINE	Interspersed	1,797,575	15%	100–300
DNA transposon	Interspersed	463,776	3%	200–2,000
LTR retrotransposon	Interspersed	718,125	9%	200–5,000
LINE	Interspersed	1,506,845	21%	500–8,000
rDNA (16S, 18S, 5.8S and 28S)	Tandem	698	0.01%	2,000–43,000
Segmental duplications and other classes	Tandem or interspersed	2,270	0.20%	1,000–100,000



From: Treangen & Salzberg, 2012

For those of you interested in Transposable Elements biology:

Current Opinion in Genetics and Development, April 2018 issue:

<https://www.sciencedirect.com/journal/current-opinion-in-genetics-and-development/vol/49/suppl/C>

How to study Adv Mol Biol

1. Follow lessons and activities as much as you can
2. Study every group of lessons immediately using «TextBook» reviews

How to prepare your exam

1. Study the 4 (four) Research papers thoroughly
2. Study the subjects using exclusively the TextBook reviews.
3. Do not study on lesson slides
4. Do not study on the notes you've taken during my lectures

How the exam is organized:

1. one moodle quiz with 15 questions primarily on the second module
2. one interview composed of:
 1. 10 minutes (max) free talking on one subject from the list given in advance
 2. few quick questions on your talk
 3. analysis of one Research Paper
3. 1 and 2 will compose the average grade with ***internal activities*** (1:1:1)

Internal activities are;

Participation to Exercises, quizzes, lessons in the moodle format,
Wikis & books on different subjects

Students' report: this activity consists traditionally in the choice, study and analysis of a recently published research paper on course themes, and preparation of a short report to be tell to the class in 20-25 minutes. This activity is made by students in couple. However, due to the difficulty linked to the coronavirus emergency, we may discuss and set out different kinds of reporting.

Research Paper – Reddy et al, 2008

Ch. 1 Textbooks (Review articles):

1. Bannister 2011 on PTMs
2. VanSteensel 2017 on LADs

