

URPP Tutorial: Genomic Visualization I, UZH

Session I. Genome Browsers

Part 1. UCSC Genome Browser

1. Navigating the Browser:

- Open your internet browser of choice (Firefox, Chrome, Opera, Safari)
- Go to <https://genome-euro.ucsc.edu/>
- Click on “**Genome Browser**”



- Select “**Human Assembly**” (usually it’s selected by default)
- Type **BRCA1** on the “**Position/Search Term**” field in the browser gateway and click “**GO**”. Look at the default tracks and try to answer the following questions:

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute UCSC Genome Browser Gateway

Genomes Genome Browser Tools Mirrors Downloads My Data Help About Us

Browse/Select Species

POPULAR SPECIES

Human Mouse Rat Fruitfly Worm Yeast

Enter species or common name

REPRESENTED SPECIES

Human Chimp Bonobo Gorilla Orangutan Gibbon Green monkey Crab-eating macaque Rhesus Baboon (anubis) Baboon (hamadryas)

Find Position

Human Assembly
Feb. 2009 (GRCh37/hg19)

Position/Search Term
BRCA1

Current position: chr17:41,067,594-41,432,945

GO

Human Genome Browser - hg19 assembly

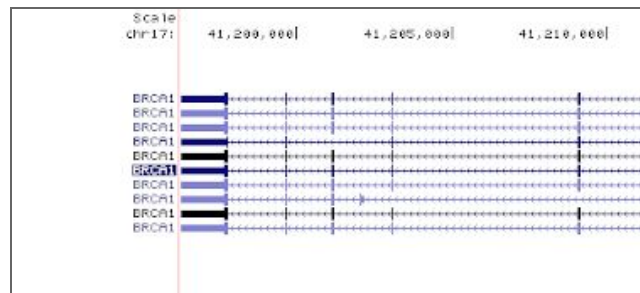
The February 2009 human reference sequence (GRCh37) was produced in the NCBI Assembly database.

Sample position queries

A genome position can be specified by the accession number of a sequence, keywords from the GenBank description of an mRNA. The following more information.

Request: Genome Browser Response:

- Which is the transcript with more exons? (hint: hover over the last exon of the isoforms)



- Which are the **two closest** genes to BRCA1? (hint: zoom-out and look immediate left and right of the gene)

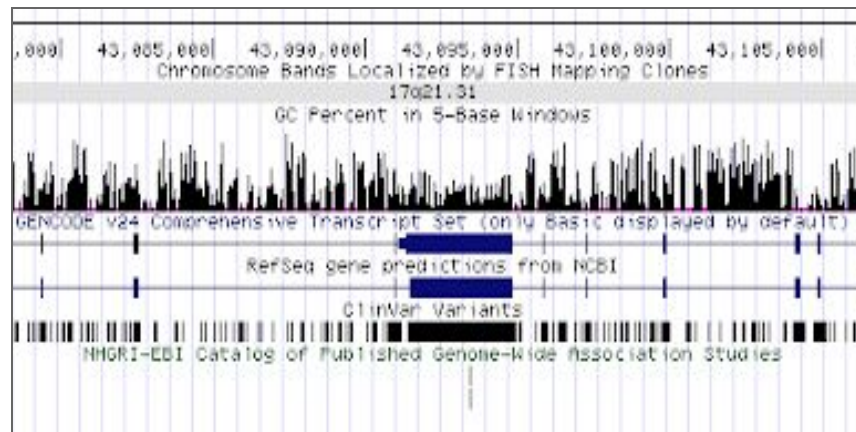
UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr17:41,196,312-41,277,500 81,189 bp. enter position, gene symbol, HGVS or search terms go

- In which tissue it has the **highest** expression? And **lowest**? (hint: click on the expression plot from GTEx)
- Hide all tracks **except** the **Gencode** track (hint: right click with the mouse of the left panels and hide the tracks)

- Can you identify the SNP ID? (Hint: Hover over it)



- What's the latest study about? (Hint: click directly on the SNP)
- What conclusions can you draw about this gene regarding its function?

Part 1.1. UCSC Table Browser

1. Go to **Tools > Table Browser**

Table Browser

Use this program to retrieve the data associated with a track in text format, to calculate intersections between tracks, and to retrieve DNA description of the controls in this form, and the [User's Guide](#) for general information and sample queries. For more complex queries, you can use the [Query Builder](#) through annotation enrichments, send the data to [GREAT](#). Send data to [GenomeSpace](#) for use with diverse computational tools. Refer to the [Sequence and Annotation Downloads](#) page.

clade: genome: assembly:

group: track:

table:

region: ☒ genome ☐ ENCODE Pilot regions ☐ position

identifiers (names/accessions):

filter:

intersection:

correlation:

output format: Send output to ☐ Galaxy ☐ GREAT ☐ GenomeSpace

output file: (leave blank to keep output in browser)

file type returned: ☒ plain text ☐ gzip compressed

To reset **all** user cart settings (including custom tracks), [click here](#).

2. Create a **custom** track:

- Use the **human GRCh37/h19 assembly**.

- Select → **Group: regulation** and **Track: CpG Islands**
- **Import** the custom track to the **UCSC Genome browser** by changing the output format to **custom track** and clicking on **get output** → **get custom track in genome browser**

Output cpGISlandExt as Custom Track

[Custom track](#) header:

name=

description=

visibility=

url=

Create one BED record per:

☒ Whole Gene

☐ Upstream by bases

☐ Downstream by bases

Note: if a feature is close to the beginning or end of a chromosome an

- Is there any **CpG island in or near BRCA1**? (hint: remember to zoom-out)

Part 1.3 UCSC Genome Graphs

1. Go to **Tools > Genome Graphs > import**

[Home](#)
[Genomes](#)
[Genome Browser](#)
[Tools](#)
[Mirrors](#)
[Downloads](#)
[My Data](#)
[View](#)
[Help](#)
[About Us](#)

Human Genome Graphs

clade:
genome:
assembly:

graph
in
,

in

significance threshold:

No graph data is available for this assembly. Upload your own data or import from a table or custom track.

Using Genome Graphs

Genome Graphs is a tool for displaying genome-wide data sets such as the results of genome-wide SNP association studies, linkage studies and h instructions, see the [Genome Graphs User's Guide](#).

2. Import the **knownGene** track by clicking on import (remember to name it!)

Import Table to Genome Graphs

group: Genes and Gene Predictions track: GENCODE v24

table: knownGene

name of data set: Genes

description: gencode 24

display min value: max value:

label values:

draw connecting lines between markers separated by up to 25000000 bases.

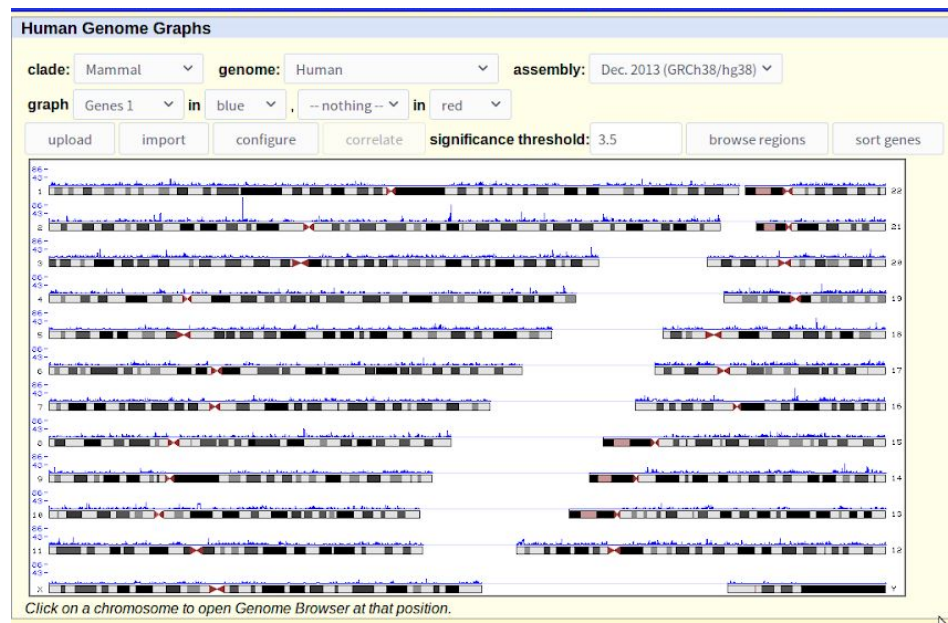
conversion type: ☒ depth ☐ coverage

Note: Loading some tables can take up to a minute. If you are importing more than one data set please give them different names. Only the most recent you may have to import them again.

submit

Import table types

3. Load the **Genes** track. Click on **configure**.



4. Change the **chromosome layout** to **one per line**.

5. Change the **width** to 2000 and change the **number of graphs** and **number of lines** to two.
6. Now **import the track of repetitive elements** (Repeats) by doing the same that we previously did to upload the track of genes.
7. Remember to load the tracks, you can customize the **colours** and **overlap** graphs. You can upload custom tracks as well.
8. Can you conclude anything about the repetitive elements around the centromeres?

Part 2. Ensembl Browser

1. Go to <https://www.ensembl.org/>
2. In the **Search** select human and type your favorite gene (i.e. BRCA1) and click on “Go”

The screenshot shows the Ensembl browser interface for the BRCA1 gene. The top navigation bar includes links for BLAST/BLAT, VEP, Tools, BioMart, Downloads, Help & Docs, and Blog. A search bar is located in the top right corner. The main content area is divided into a left sidebar with a 'Gene-based displays' menu and a main panel for the 'Gene: BRCA1 ENSG0000012048'. The sidebar menu includes options like Summary, Splice variants, Transcript comparison, Gene alleles, Sequence, Secondary Structure, Comparative Genomics, Genomic alignments, Gene tree, Gene gain/loss tree, Orthologues, Paralogs, Ensembl protein families, Ontologies, GO: Molecular function, GO: Biological process, GO: Cellular component, Phenotypes, Genetic Variation, Variant table, Variant image, Structural variants, Gene expression, Pathway, Regulation, External references, Supporting evidence, ID History, and Gene history. The main panel displays the gene's description, synonyms, location, and a summary of its features. A 'Show transcript table' button is visible in the 'Transcripts' section. At the bottom left, there are buttons for 'Configure this page' and 'Custom tracks'.

- How many transcripts does BRCA1 has?
 - How many transcripts are annotated as non-coding?
 - How many **one-to-one** orthologs? (hint: click on orthologues)
3. Go to menu on the left and click on **Genomic alignments** (under **Comparative Genomics**) and select: **Multiple > 12 Primates EPO > View an image of this alignment**

Genomic alignments ?

Alignment: [Select another alignment](#)

[Download alignment](#)

Hidden

The following 1 species in the alignment are not shown. Use "Configure this page -> 12-way EPO alignment" on the left to show them.

- Ancestral sequence

[View an image of this alignment](#)

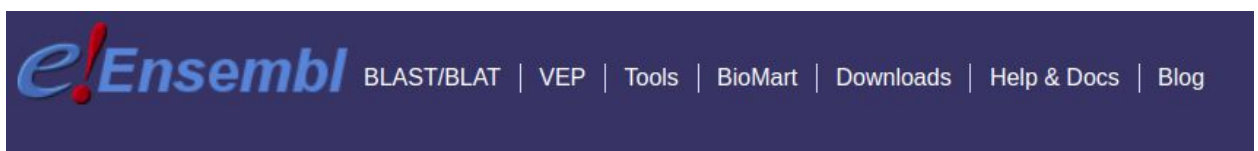
A total of 3 alignment blocks have been found. Please select an alignment to view by selecting a Block from the Alignment column.

Alignment (click to view)	Length (bp)	Location on Human	Additional species
Block 1	101548	17:43069298-43170845	11
Block 2	23619	17:43045679-43069297	8
Block 3	667	17:43043695-43044361	7

- You can either **download the data** for further use or **quickly** assess if your gene of interest is **conserved** or/and **syntenic** among species of interest.

Part 2. BioMart

- Go to BioMart** by clicking **BioMart** on the ensembl webpage



- Create a BED file using BioMart.
- Select **Ensembl Genes 94**
- Choose your favorite species
- Click on **filter** and select **protein_coding**

e!Ensembl BLAST/BLAT | VEP | Tools | BioMart | Downloads | Help & Docs | Blog

New Count Results URL XML Perl Help

Dataset 22686 / 64914 Genes
Human genes (GRCh38.p12)

Filters

Gene type: protein_coding

Attributes

Chromosome/scaffold name
Gene start (bp)
Gene end (bp)
Gene stable ID
Gene type
Strand
Karyotype band

Dataset
[None Selected]

☐ Transcript count <=

☒ Gene type

☐ Transcript type

☐ Source (gene)

☐ Source (transcript)

protein_coding
pseudogene
antisense
3prime_overlapping_ncRNA
antisense
bidirectional_promoter_lncRNA
IG_C_gene
IG_C_pseudogene
IG_D_gene
IG_J_gene
IG_J_pseudogene
IG_pseudogene
IG_V_gene

ensembl

ensembl

- Click on “**count**” and see how many protein-coding genes are (hint: look next to Dataset).
- Now click on **Attributes** and unselect everything
- Select with this specific order: **Chromosome/scaffold name** > **Gene start (bp)** > **Gene end (bp)** > **Gene stable ID** > **Gene type** > **Strand**

New Count Results URL XML Perl Help

Dataset 22686 / 64914 Genes
Human genes (GRCh38.p12)

Filters

Gene type: protein_coding

Attributes

Chromosome/scaffold name
Gene start (bp)
Gene end (bp)
Gene stable ID
Gene type
Strand

Dataset
[None Selected]

Export all results to File TSV Unique results only Go

Email notification to

View 10 rows as HTML Unique results only

Chromosome/scaffold name	Gene start (bp)	Gene end (bp)	Gene stable ID	Gene type	Strand
CHR_HSCHR15_4_CTG8	32695437	32720246	ENSG00000275568	protein_coding	1
CHR_HSCHR15_5_CTG8	82403250	82412947	ENSG00000275049	protein_coding	-1
CHR_HSCHR19LRC_LRC_J_CTG3_1	54956665	54980661	ENSG00000276804	protein_coding	-1
CHR_HSCHR11_1_CTG6	1586202	1591699	ENSG00000276565	protein_coding	1
CHR_HG2030_PATCH	133356700	133361317	ENSG00000281024	protein_coding	1
CHR_HSCHR22_1_CTG7	23957418	23961206	ENSG00000278695	protein_coding	-1
CHR_HSCHR17_1_CTG5	46261820	46313387	ENSG00000278232	protein_coding	-1
CHR_HSCHR19LRC_LRC_J_CTG3_1	54190866	54195141	ENSG00000274672	protein_coding	1
CHR_HSCHR6_MHC_COX_CTG1	28510773	28515743	ENSG00000281185	protein_coding	-1
CHR_HSCHR19LRC_LRC_S_CTG3_1	55021410	55045969	ENSG00000274566	protein_coding	-1

- Click on **Results**
- Mark “**unique results only**” > **Go**
- The download will start and you will have a bed file.
- You can create multiple files with lots of information.

- If you selected human as your favorite species --can you find genes only involved in Schizophrenia? (hint: go to **filters** > **phenotype**). How many genes did you find?

13. You can explore other phenotypes, protein features and additional external information.

NOTE: There is **Ensembl Bacteria**: <https://bacteria.ensembl.org/index.html> for bacterial species.

Part 3. Integrative Genome Viewer

1. If you have the **VM** from other courses IGV is under **home/student/software**
2. Else **download** and **install IGV**:

- **For Windows/Mac/Linux:**

<https://software.broadinstitute.org/software/igv/download>

- **From the command:**

```
wget \
http://data.broadinstitute.org/igv/projects/downloads/2.4/IGV\_2.4.14.zip
unzip IGV_2.4.14.zip
```

- **Run IGV from within the installed distribution:**

```
./igv.sh
```

- **Or if you are in Linux:**

```
sudo apt-get install igv #install
Igv                      #run
```

3. You will work with data from the 1000 Genomes Project:

- Download the **.bam file** and its **index .bai** (the \ is to be careful of the space, there cannot be any newlines when you copy it in the command line):

```
wget \
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data/HG00096/exome\_alignment/HG00096.chrom11.ILLUMINA.bwa.GBR.exome.20120522.bam
```

```
wget \
ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/data/HG00096/exome\_alignment/HG00096.chrom11.ILLUMINA.bwa.GBR.exome.20120522.bam.bai
```

Note: Both files have to be in the same folder.

- Load the files on IGV: **File > Load from folder > exome_alignment/HG00096.chrom11.ILLUMINA.bwa.GBR.exome.20120522.bam**
- Navigate to the gene **SSRP1** in chr11. Remember what **the colours meant?**
- Look at the neighboring gene **P2RX3**. Is there anything curious about the way the reads are distributed?

4. Load the data from the server:

Load from server > 1000 Genomes > Phase 3 sites > Phase 3 Genotypes

- Navigate to **BRCA1**. What can you tell about the allele frequencies?

Part 4. **Circos** (*we will not use it*)

1. Download Circos:

```
wget http://circos.ca/distribution/circos-0.69-6.tgz
tar zxvf circos-0.69-6.tgz      #unzip
```

- Check for missing Perl modules:

```
sudo perl -MCPAN -e shell
install Config::General
...
exit
```

- If the installation of modules fails:

```
sudo apt-get install libgd-perl
```

- Run Circos within your distribution:

```
bin/circos
```