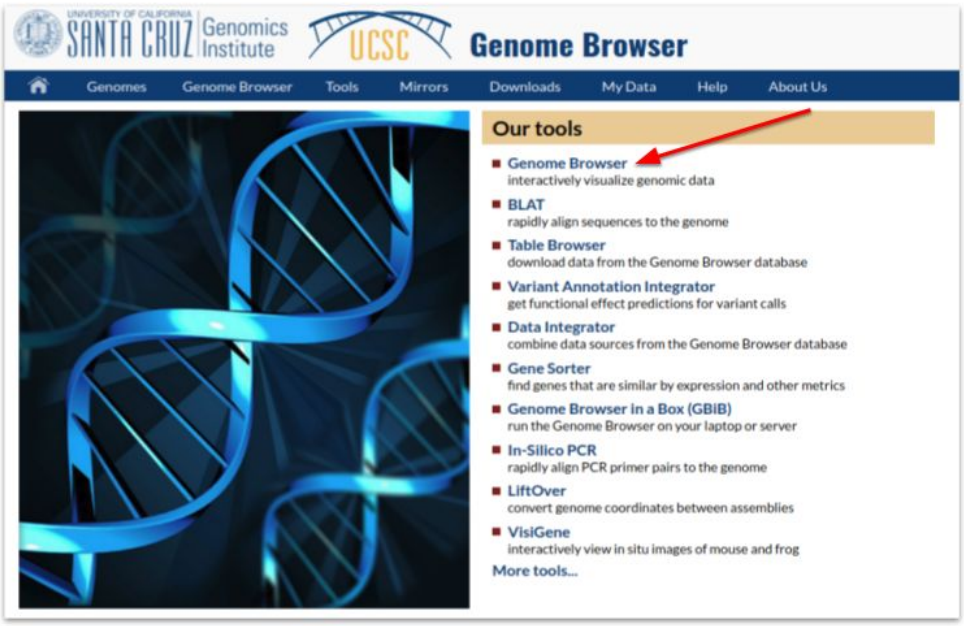
URPP Tutorial: Genomic Visualization I, UZH
Session I. Genome Browsers

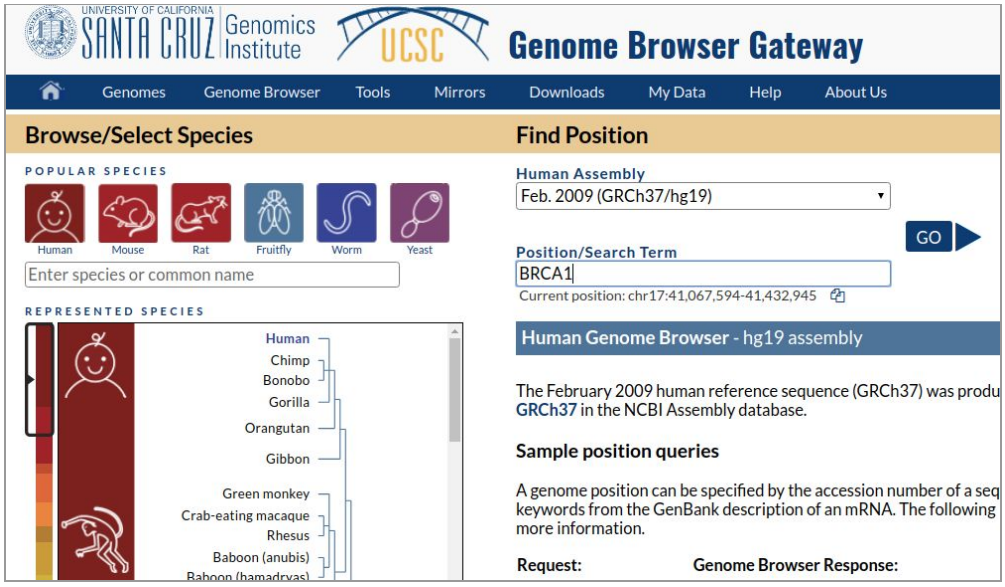
Part 1. UCSC Genome Browser

1. **Navigating the Browser:**

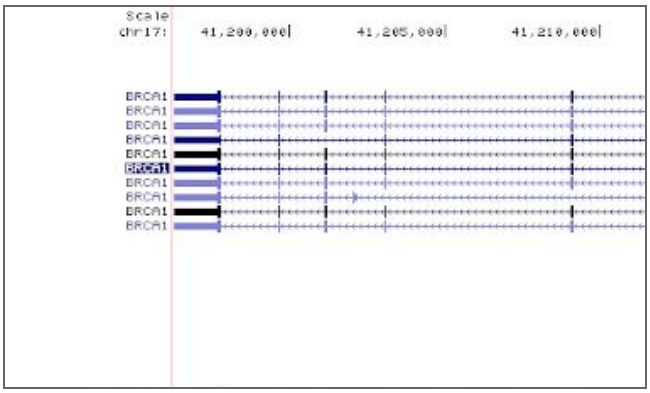
- Open your internet browser of choice (Firefox, Chrome, Opera, Safari)
- Go to https://genome-euro.ucsc.edu/
- Click on "**Genome Browser"**



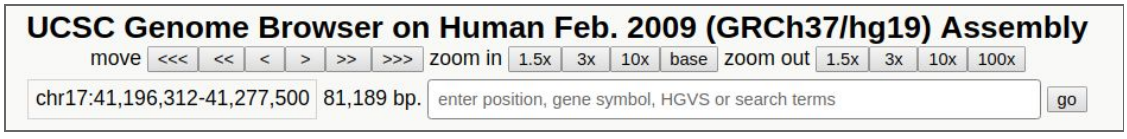
- Select "**Human Assembly**" (usually it's selected by default)
- Type **BRCA1** on the "**Position/Search Term**" field in the browser gateway and click "**GO**". Look at the default tracks and try to answer the following questions:



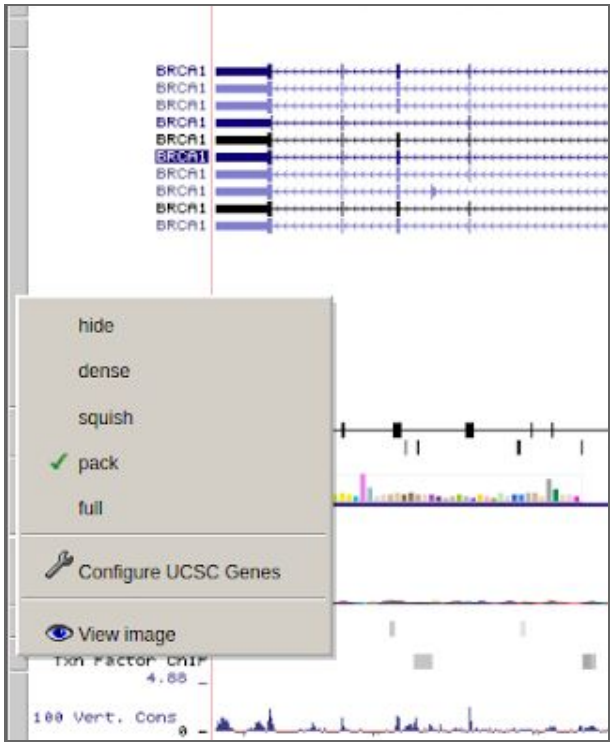
- Which is the transcript with more exons? (hint: hover over the last exon of the isoforms)



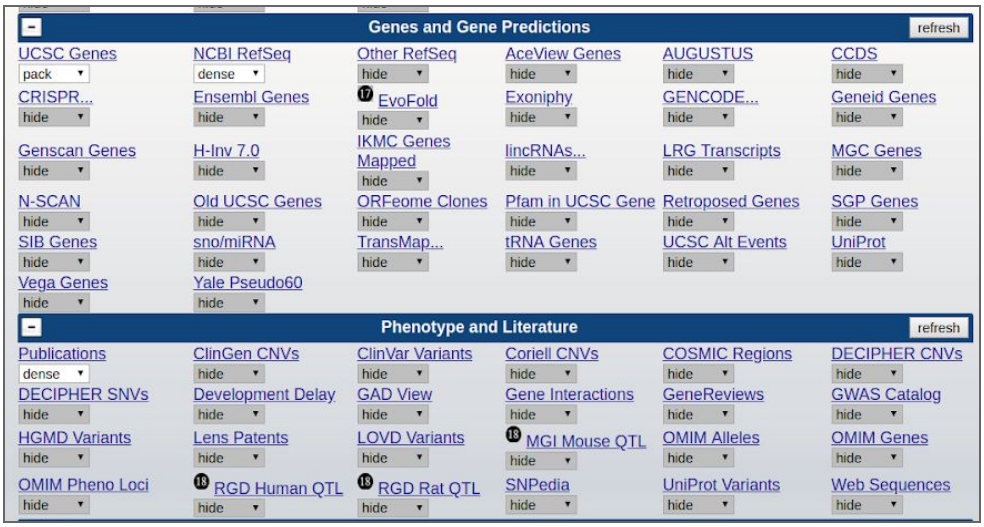
- Which are the **two closest** genes to BRCA1? (hint: zoom-out and look immediate left and right of the gene)



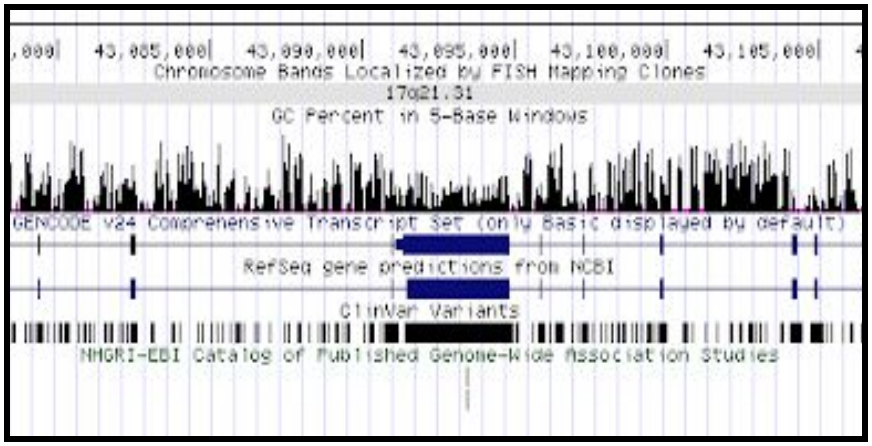
- In which tissue it has the **highest** expression? And **lowest**? (hint: click on the expression plot from GTEx)
- Hide all tracks **except** the **Gencode** track (hint: right click with the mouse of the left panels and hide the tracks)



- Activate the track “**GWAS Catalog**” (Hint: Check the “Phenotype and Literature” track)



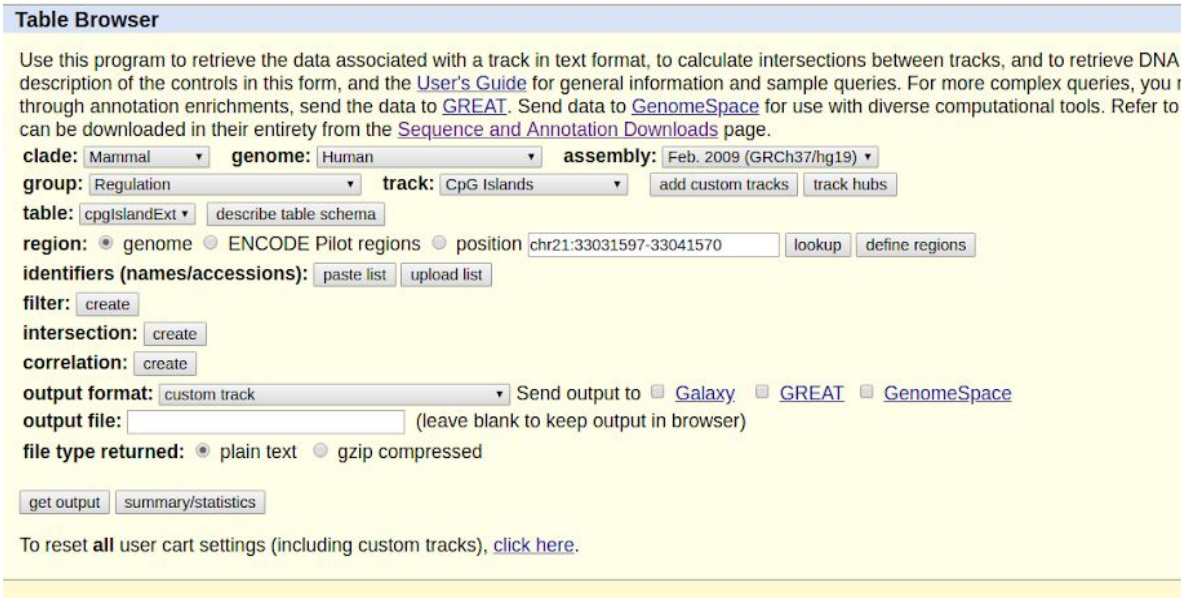
- Can you identify the SNP ID? (Hint: Hover over it)



- 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

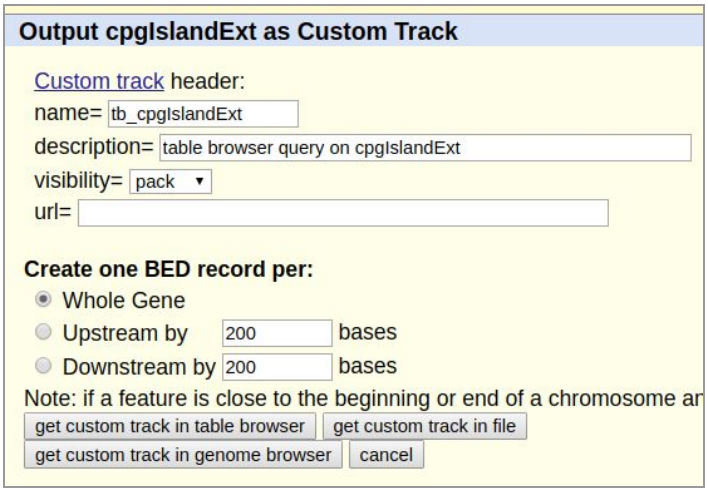
1. Go to **Tools** >**Table Browser**



2. Create a **custom** track:

- Use the **human CRCh37/h19 assembly**.

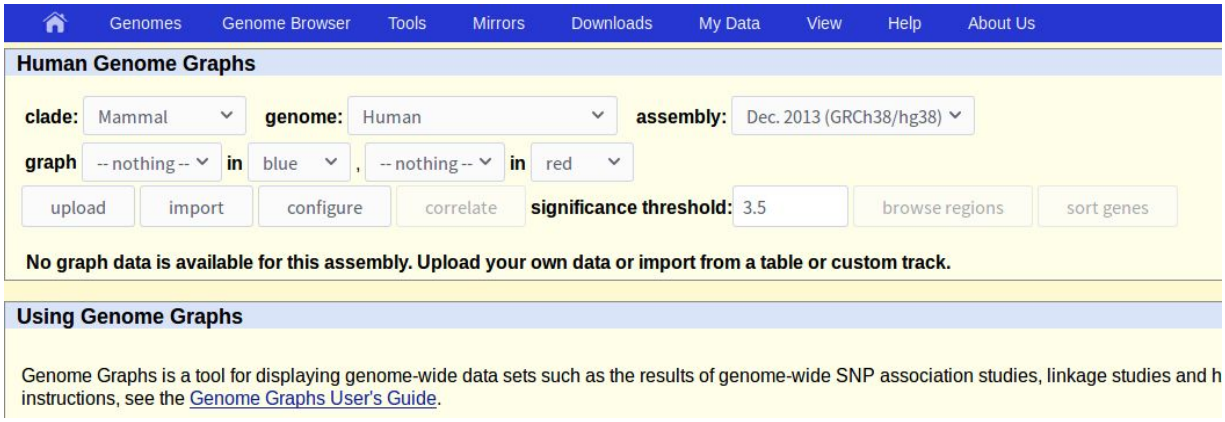
- Select → **Group**: **regulation** and **Track: CpG Islands**
- **Import** the custom track to the **UCSC Genome browser** by changing the output format to **custom track** and clicking on **get output → get custom track in genome browser**



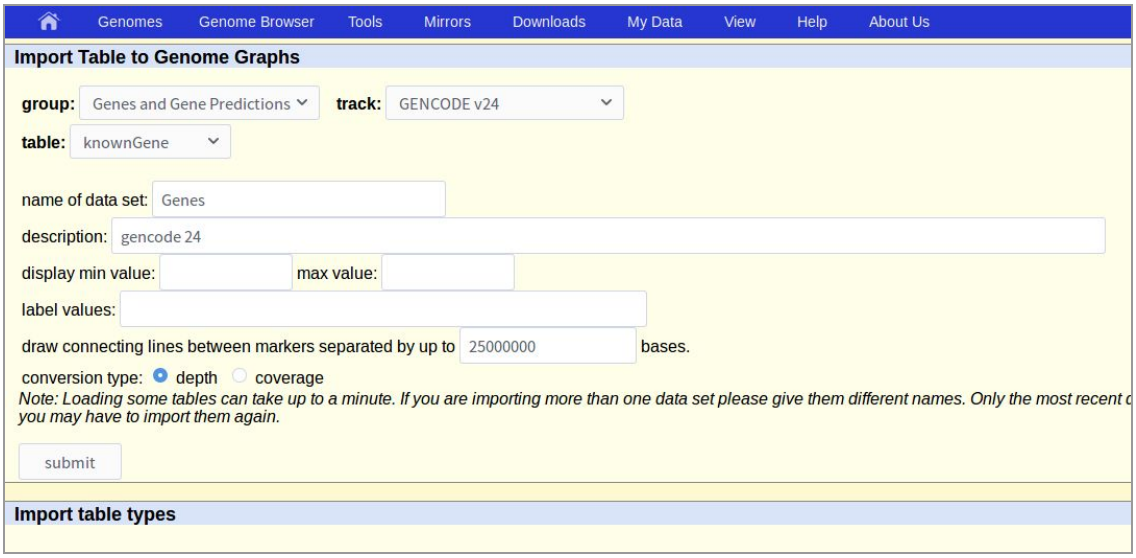
- Is there any **CpG island in or near BRCA1**? (hint: remember to zoom-out)

Part 1.3 UCSC Genome Graphs

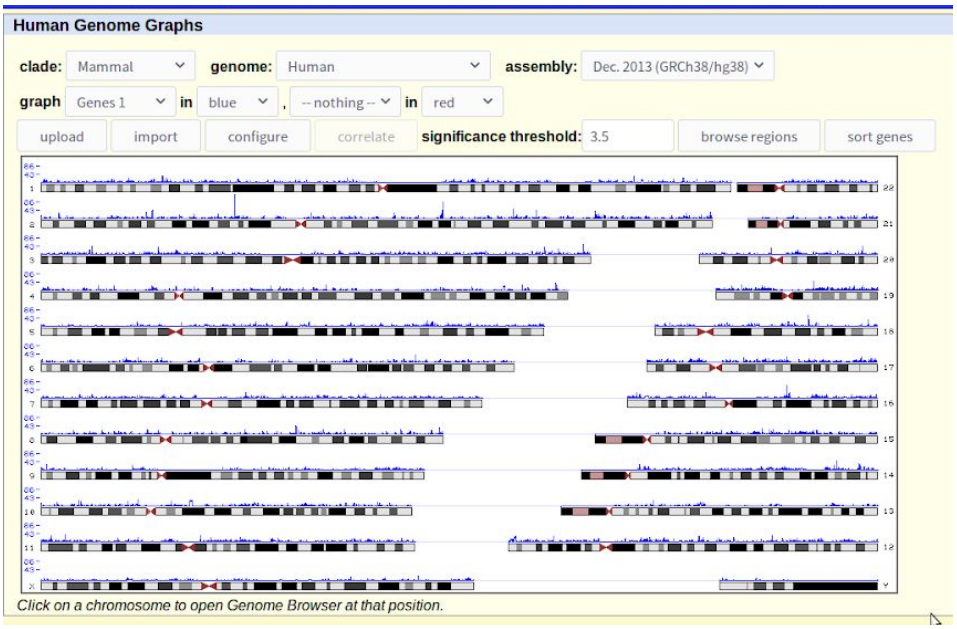
1. Go to **Tools** > **Genome Graphs > import**



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



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

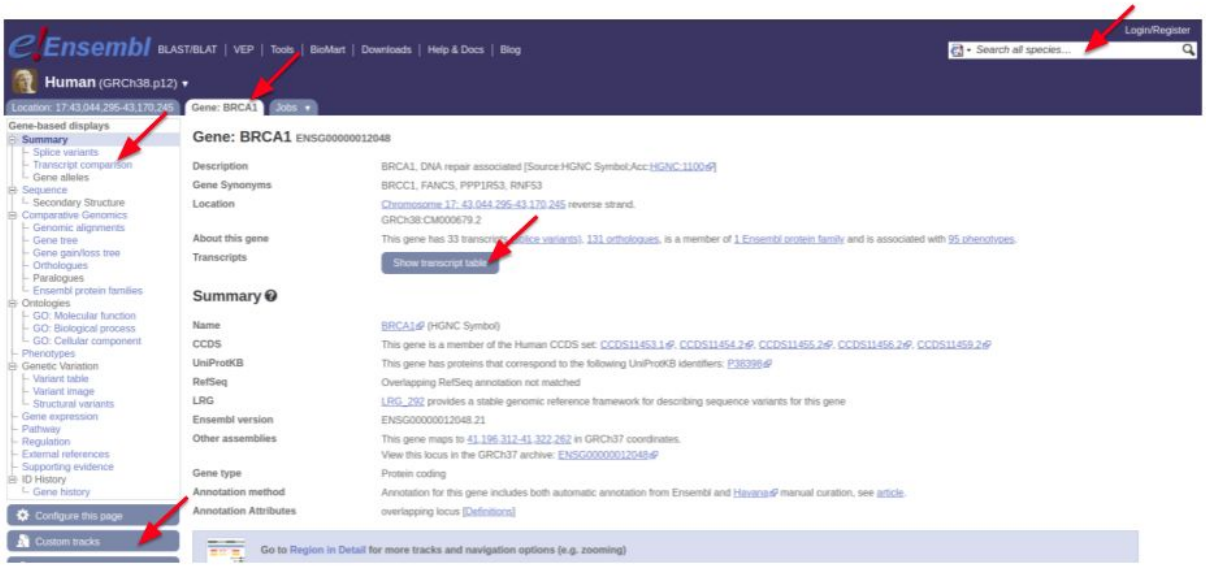


4. Change the **chromosome layout** to **one per line**.

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?

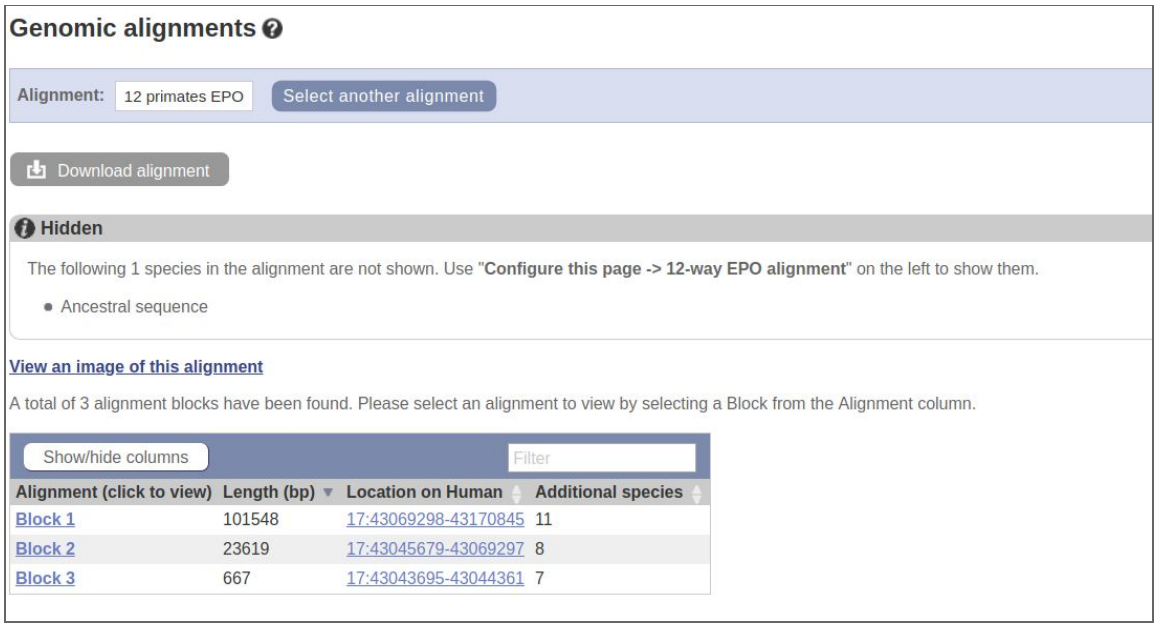
Part 2. Ensembl Browser

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**"



- How many transcripts does BRCA1 has?
- How many transcripts are annotated as non-coding?
- How many **one-to-one** orthologs? (hint: click on orthologues)

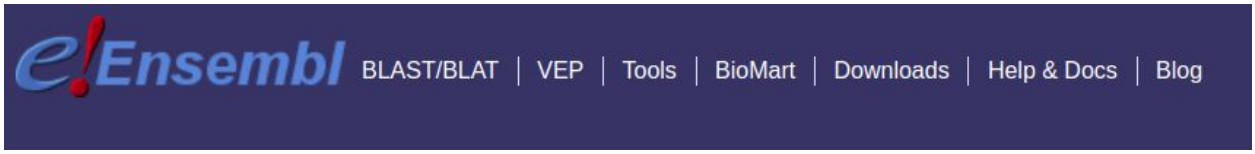
3. Go to menu on the left and click on **Genomic alignments** (under **Comparative Genomics**) and select: **Multiple** > **12 Primates EPO** > **View an image of this alignment**



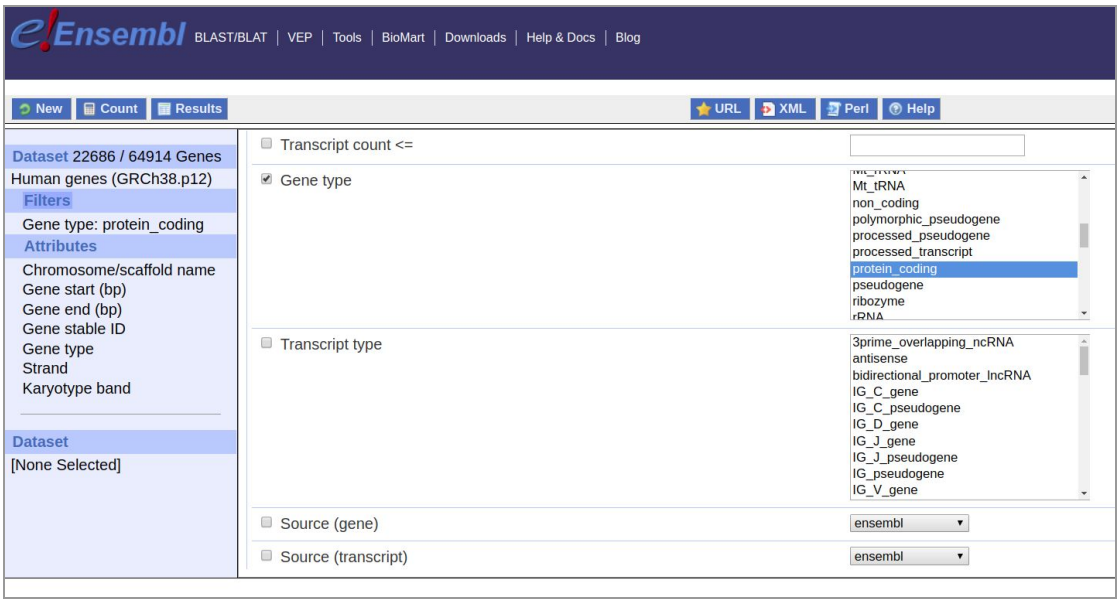
4. You can either **download the data** for further use or **quickly** assess if your gene of interest is **conserved** or/and **syntenic** among species of interest.

Part 2. BioMart

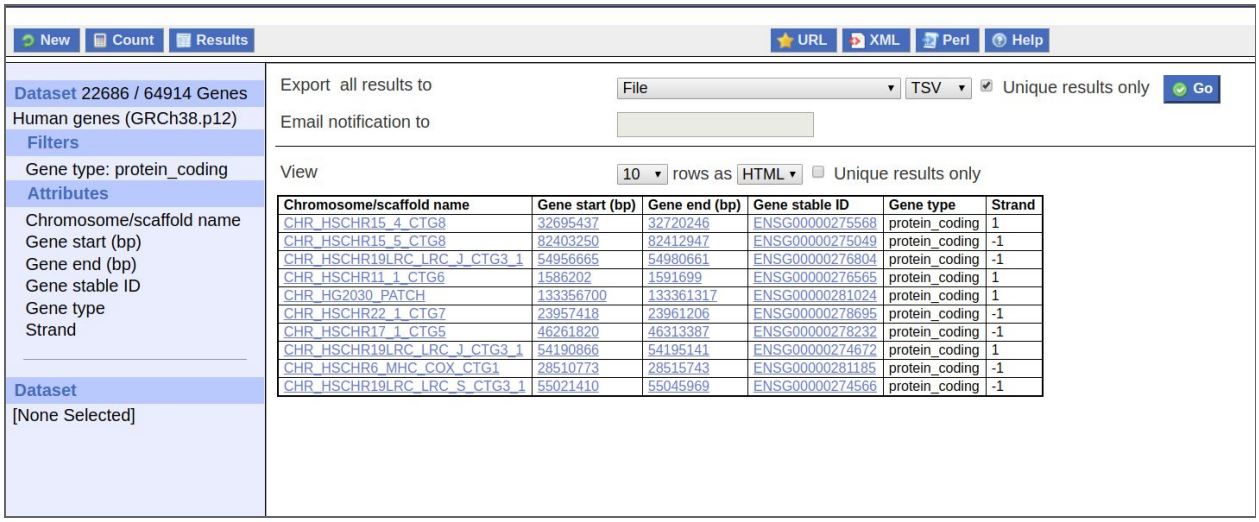
1. **Go to BioMart** by clicking **BioMart** on the ensembl webpage



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



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



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

9. Click on **Results**
10. Mark "**unique results only**" > **Go**
11. The download will start and you will have a bed file.
12. 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 folde**r > **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
```