Using Galaxy:

[What is Galaxy from main page to the use galaxy server page: 1](#_Toc60229530)

[On human Chromosome 22, which coding exons have the most repeats in them? 2](#_Toc60229531)

[Get Data: 2](#_Toc60229532)

[A) Get the Codons data for Chromosome 22: 2](#_Toc60229533)

[B) Get the repeats data for Chromosome 22: 3](#_Toc60229534)

[C) Look for places where these two datasets overlap: 4](#_Toc60229535)

[D) Count of number of repeats overlapping: 4](#_Toc60229536)

[E) extra: Incorporate the overlap count with the rest of the Exon information: 5](#_Toc60229537)

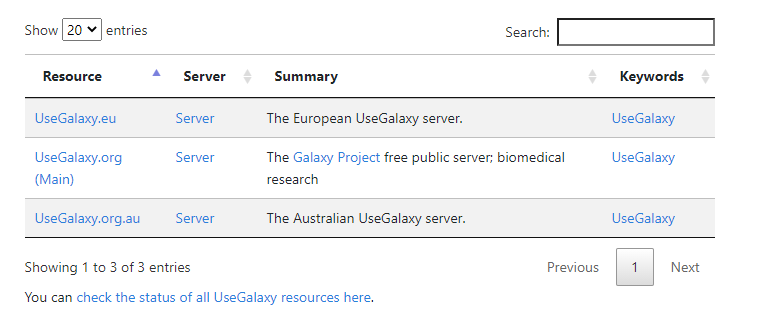
# What is Galaxy from main page to the use galaxy server page:

Steps to create an account:

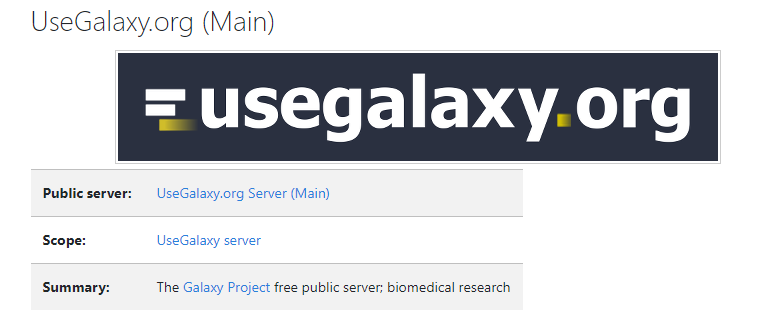
1 - <https://galaxyproject.org/>

2 – Menu – hit ‘Use’: <https://galaxyproject.org/use/>

3 – In the options below – choose “UseGalaxy.org” (Main Server)



4 – And then again use the Public Server “UseGalaxy.org Server (Main)



5 – This will ultimately take you here <https://usegalaxy.org/> - main page to se the Open source. Here go to the Login/Register Menu and register for a free open source account.

* Open Source
* Web-based platform
* For Data intensive biomedical research

Few things to keep in mind about Galaxy:

* Interactive analysis is performed using the tools to operate on dataset, set within Galaxy.
* Datasets are immutable and running tools always creates one or more new datasets.
* Datasets are available through history, which provides complete provenance for each dataset.

# On human Chromosome 22, which coding exons have the most repeats in them?

(also, on: <http://usegalaxy.org/galaxy101>)

1 – Get data into Galaxy

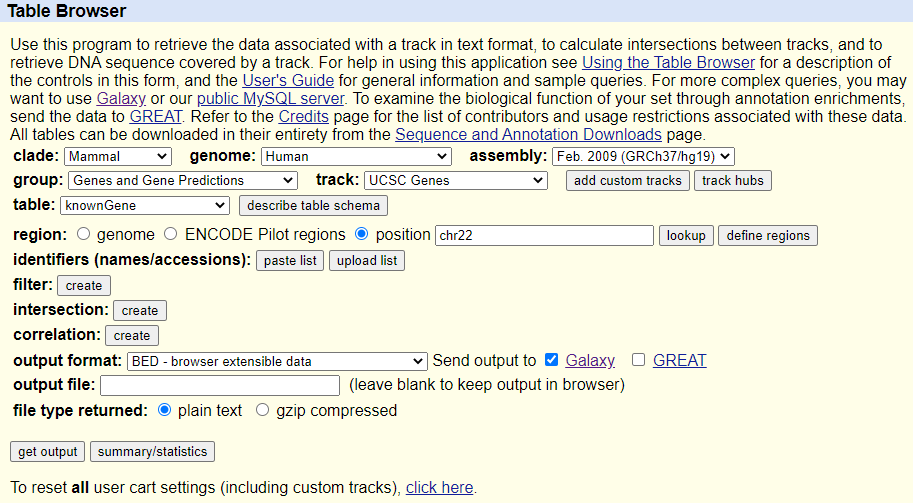
2 – Identify which exons have repeats

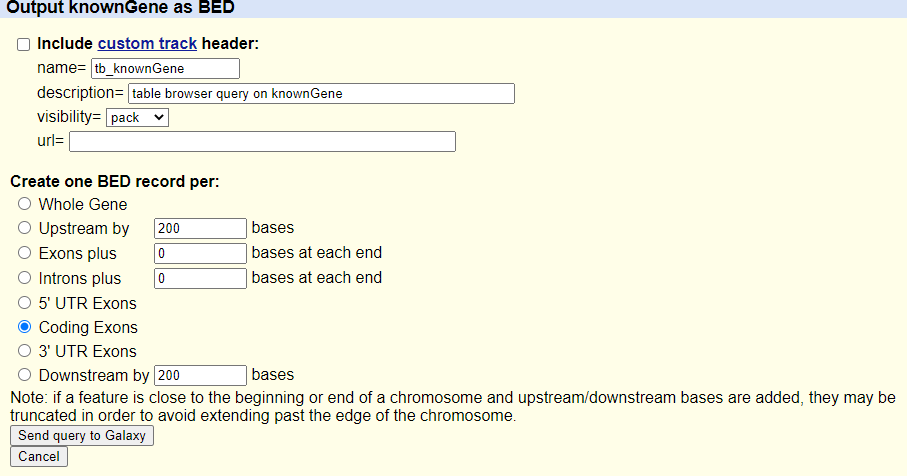
3 – Count the number of repeats

## Get Data:

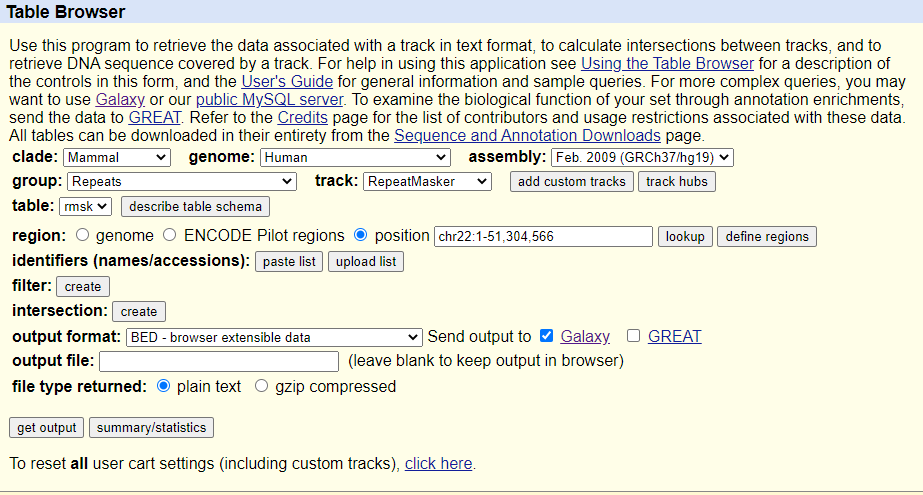
### A) Get the Codons data for Chromosome 22:

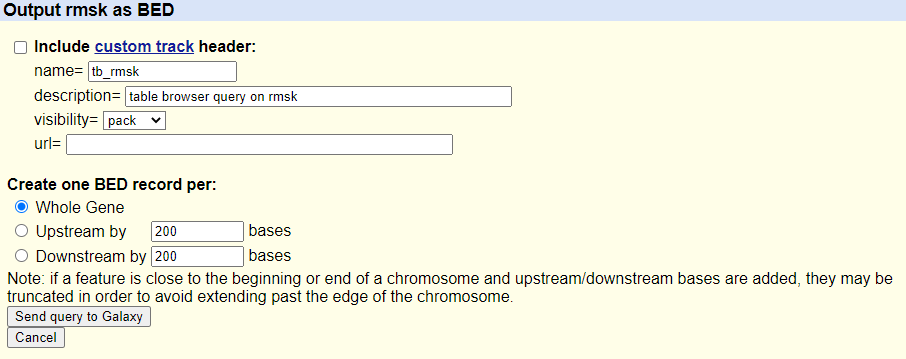
The following settings in the UCSC Main has been used:





### B) Get the repeats data for Chromosome 22:





### C) Look for places where these two datasets overlap:

a JOIN on both data sets. So matching the repeats to the Chromosome coding regions

>> Found in: “Operate on Genomic Intervals” >> Join

### D) Count of number of repeats overlapping:

Grouping tool >> Found in: “Join, Subtract, and Group” >> Group:

1. Insert Operation
   1. Count
   2. Column 4
   3. Execute
2. Results will show:

>> The exon repeat and how often it was matched in the join.

### E) extra: Incorporate the overlap count with the rest of the Exon information:

Joining, but this time on the Exon Name:

Incorporate the overlap count with rest of the Exon information

>> Found in: “Join, Subtract, and Group” >> Join two Datasets:

Join:

A – Chromosome Gene information (the first dataset): Column 4

B - Dataset of the repeats match between the two datasets: Column 1

Result: Full Exon information along with the count

### F) Furthermore, you can extract just the columns that you are looking to have in the final output.

If you want to get the data out of galaxy, you can use the download icon (disk icon) to your computer in the raw format.