|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tool** | **What for** | **What in** | **What out** | ***Example operation*** |
| **UCSC Main** table browser | **Retrieve and export data from the Genome Browser** annotation track database | What in: **select**  dataset, **define** region of interest (genome | position + identifiers),  **retrieve and display data**  **Send output to Galaxy:** displays results of query in [Galaxy](https://usegalaxy.org/), a framework for interactive genome analysis.  **Send output to GREAT:** displays the functional enrichments of the query results in [GREAT](http://great.stanford.edu/), a tool for analysis of the biological function of cis-regulatory regions.  **get output:** Submits a data query based on the specified parameters and returns the output.  **Summary/statistics:** Displays statistics about the data specified by the parameters. | output formats    file type returned  **plain text** - data is in ASCII format  **\*.gzip** compressed archieve format for Linux|Unix | Retrieve list of genes of an animals, viruses, insects for i.e. mice etc.  [*https://genome.ucsc.edu/cgi-bin/hgTables?GALAXY\_URL=https%3A//usegalaxy.eu/tool\_runner&tool\_id=ucsc\_table\_direct1&sendToGalaxy=1&hgta\_compressType=none&hgta\_outputType=bed*](https://genome.ucsc.edu/cgi-bin/hgTables?GALAXY_URL=https%3A//usegalaxy.eu/tool_runner&tool_id=ucsc_table_direct1&sendToGalaxy=1&hgta_compressType=none&hgta_outputType=bed) |
| **Select last** lines from a dataset (tail) | **Keep the last X lines in a file**. It is a tool from text processing tools | **Keep last** **lines**| Keep from this line on  C:\Users\Ludmilla\Desktop\Galaxy\cheetsheets\peaks and genes\select_last_input_file.png | **C:\Users\Ludmilla\Desktop\Galaxy\cheetsheets\peaks and genes\select_last_output.pngOutput** | to compare the two files, to make sure that the chromosome names follow the same format    https://usegalaxy.eu/root?tool\_id=toolshed.g2.bx.psu.edu/repos/bgruening/text\_processing/tp\_tail\_tool/1.1.0 |
| **Replace Text in a specific column** | **Performs find & replace operation** on a specified column in a given file.   For more complex patterns, use the ***awk***tool. | **Parameters**  **In column:** i.e. 5  **Find pattern:** i.e. any symbol or letter expression „hello“  **Replace with:** text, or & (ampersand) and \\1 \\2 \\3  **Insert Replacement** (add new replacement for a new column) | The same text file in a \*.gzip with replacement or replacements | For i.e. to convert the chromosome names, to change 20 and 21 to X and Y  [*https://usegalaxy.eu/root?tool\_id=toolshed.g2.bx.psu.edu/repos/bgruening/text\_processing/tp\_replace\_in\_column/1.1.3*](https://usegalaxy.eu/root?tool_id=toolshed.g2.bx.psu.edu/repos/bgruening/text_processing/tp_replace_in_column/1.1.3) |
| **Get flanks** returns flanking region/s for every gene | **This tool finds the** **upstream and/or downstream flanking region(s)** of all the selected regions in the input file. | Every line should contain at least 3 columns: Chromosome number, Start and Stop co-ordinates.  **Parameters**  **Region:** Whole feature| Around start | Around end  **Location of the flanking region/s:**  **Offset:**  Use positive values to offset co-ordinates in the direction of transcription and negative values to offset in the opposite direction**.**  **Length of the flanking region(s)**  Use non-negative value | **Result**  BED format file with flanking regions for every gene | Adding promoter regions, i.e. to get regions 2kb bases upstream of the start of the gene to 10kb bases downstream of the start (12kb in length)    *https://usegalaxy.eu/root?tool\_id=toolshed.g2.bx.psu.edu/repos/devteam/get\_flanks/get\_flanks1/1.0.0* |

Formats from Peaks and Genes

GNU zip (gzip) –compressed archive format for Linux and Unix systems

Txt. – plain text file

**Interval** format is a Galaxy format for representing genomic intervals. It is tab-separated, but has the added requirement that three of the columns must be:

* chromosome ID
* start position (0-based)
* end position (end-exclusive)

An optional strand column can also be specified, and an initial header row can be used to label the columns, which do not have to be in any special order. Unlike BED format (see below) arbitrary additional columns can also be present.

The **BED - Browser Extensible Data** format provides a flexible way to encode gene regions. BED lines have three required fields:

* chromosome ID
* start position (0-based)
* end position (end-exclusive)

There can be up to and nine additional optional fields, but the number of fields per line must be consistent throughout any single set of data.