# Introduction to Galaxy tools : CHEAT SHEET

FROM PEAKS TO GENES Tutorial

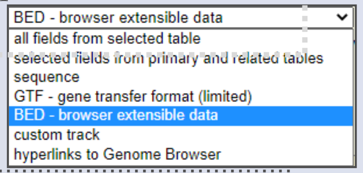
## Tools | Input parameters | Output | E.g.

### UCSC Main

**Goal:** retrieve and export data from the Genome Browser annotation track database

In a table browser select dataset, define region of interest (genome | position + identifiers) then retrieve or display the data

**File type as a plain text -** data is in ASCII format, or as ***.gzip** compressed archive format for Linux|Unix


BED - browser extensible data
all fields from selected table
selected fields from primary and related tables
sequence
GTF - gene transfer format (limited)
BED - browser extensible data
custom track
hyperlinks to Genome Browser

File formats

Retrieve list of genes of an animals, viruses, insects for i.e. mice etc.

### Select last

Keep the last X lines in a dataset.

Keep last lines| Keep from this line on

Input table

| | | | |
|---|---|---|---|
| chr1 | 10 | 20 | geneA |
| chr1 | 50 | 80 | geneB |
| chr5 | 10 | 40 | geneL |

File format is not restricted, dataset table

→

| | | | |
|---|---|---|---|
| chr1 | 50 | 80 | geneB |
| chr5 | 10 | 40 | geneL |

to compare the two files, to make sure that the chromosome names follow the same format

### Replace Text in a specific column

Perform find & replace operation on a specified column in a given file.

use **awk** tool for complex patterns

Choose column number to look at, use **"Find pattern"** and type expression to be replaced. **Replace with:** text, or & (ampersand) and \\1 (\\ is a term to find a digit) **Insert Replacement –** adds a new replacement for a new column

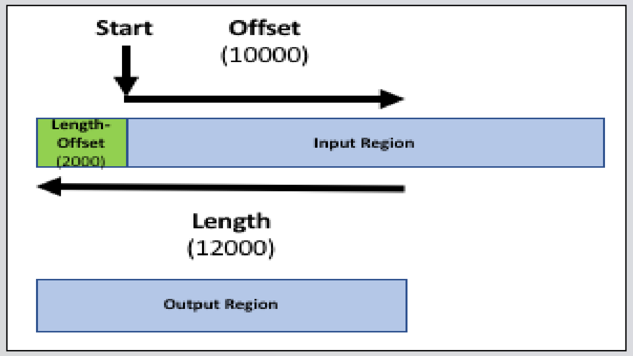
The same text file in a *.gzip with replacement or replacements

For i.e. to convert the chromosome names, to change 20 and 21 to X and Y

### Get flanks

Find the upstream and/or downstream flanking region(s) of all the selected regions

Every line should **contain at least 3 columns**: Chromosome number, Start and Stop co-ordinates



**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)

### Convert

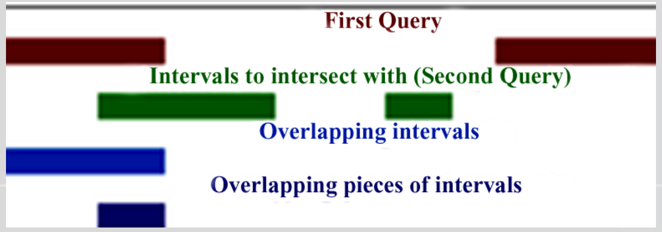
Converts Genomic Intervals To BED

Genomic intervals file

BED file

### Intersect

Choose to return Overlapping intervals OR Overlapping pieces of intervals for **TWO datasets in BED format. The order of datasets important!**


First Query
Intervals to intersect with (Second Query)
Overlapping intervals
Overlapping pieces of intervals

The intersection of two queries that are found in interval format (BED)

Extracting the genes which overlap/intersect with peaks regions in a dataset

Finding all exons containing repeats OR all regions that are both exonic and repetitive

# Introduction to Galaxy tools : CHEAT SHEET

## FROM PEAKS TO GENES Tutorial

| Tools | Input parameters | Output | E.g. |
|---|---|---|---|

### Group

**Goal:** group the input dataset by a particular column and perform aggregate functions

Dataset is in **TAB format.** **Group by column** number Chose **aggregate functions**: Mean, Median, …Randomly pick on any column(s). etc Possible to **ignore lines** with character or by case.

**If not? Text manipulation ->Convert**

Result in **TAB format**

**To group the table by chromosome and count the number of genes with peaks on each chromosome**

### Compute

Compute an expression for every row of a dataset and append the result as a new column (field)

**TAB data**

Columns are referenced with c and a number. For example, **c1** refers to **the first column** of a tab-delimited file. **Add an expression as a new column too a selected file.**

File format is not restricted, dataset table

**To generate a new BED file from the peak file that contains the positions of the peak summits.**

### Cut

Select (cuts out) specified columns from the dataset.

**This tool breaks column assignments.** **Dataset is in csv.** format . **Cut columns** for e.g. c1, c2 **Delimited by** Tab | Whitespace | Dot | Comma etc.

**To re-establish** column assignments
click on pencil icon ✎
in the latest history item.

Output is **always in tabular format** (e.g., if your original delimiters are commas, they will be **replaced with tabs**).

**Input dataset c1,c2,c3,c4,c5,c6 .**

**E.g. cut on** columns "**c6,c5,c4,c1**"

| | | | | | |
|---|---|---|---|---|---|
| a | 1 | # | % | 0 | + |
| b | 2 | $ | % | 0 | + |

→

| | | |
|---|---|---|
| a | % | + |
| b | % | + |

### Join two Datasets

Join lines of two datasets on a common field.

To change metadata assignments click on the "edit attributes" link of the history item generated by this tool.

**This tool will force the output data type to tabular.** **Two Input files** are in **TAB format, to join set** two column numbers that should be joined e.g. **using column 4 and 1**

| | | | |
|---|---|---|---|
| chr1 | 10 | 20 | geneA |
| chr1 | 50 | 80 | geneB |
| chr5 | 10 | 40 | geneL |

✖

| | |
|---|---|
| geneA | tumor suppressor |
| geneB | Foxp2 |
| geneC | Gnas1 |
| geneE | INK4a |

→

**BED** format file with flanking regions for every gene

**Result** (Keep the header lines –No)

| | | | | | |
|---|---|---|---|---|---|
| chr1 | 10 | 20 | geneA | geneA | Tumor suppressor |
| chr1 | 50 | 80 | geneB | geneB | Foxp2 |

**Result** (Keep the header lines –Yes)

| | | | | | |
|---|---|---|---|---|---|
| chr1 | 10 | 20 | geneA | geneA | Tumor suppressor |
| chr1 | 50 | 80 | geneB | geneB | Foxp2 |
| chr5 | 10 | 40 | geneL | | |

**To add in the end of BED file list of Gene names to RefSeq Gene identifiers in the table**

Joining 4 column in dataset with 1st of Dataset 2

### Sort

Sort the input file in a specific order

Select a **column** and **way it should be sorted. Initial format** can differ. Ways to sort columns: in Ascending | Descending order Flavor Fast numeric sort (-n), General numeric sort ( scientific notation -g),Natural/Version sort (-V) etc.

**Output file format equals input**

**For listing unique genes that was unsorted.**

E.g. Sort for „alphabetic order"

| |
|---|
| chr13 |
| chr2 |
| chr20 |
| chr4 |