

Supplementary File 1

This document provides a picture based tutorial on how to use uniqueome BED files to perform genomic analyses, such as calculating the unique proportion of RefSeq genes. Galaxy can be accessed from <http://main.g2.bx.psu.edu/>.

1. Select "Get Data", then "UCSC Main table browser"

2. Select gene model. eg. RefSeq genes

3. Select "BED" format, and Ensure that "Send output to Galaxy" is checked.

4. Click "get output"

The screenshot shows the Galaxy web interface with the 'Table Browser' tool selected. The left sidebar lists various tools under 'Get Data'. The main panel shows the 'Table Browser' configuration page. The 'group' is set to 'Genes and Gene Prediction Tracks', 'table' is 'refGene', 'region' is 'chr9:92376712-92376734', and 'output format' is 'BED - browser extendable data'. The 'Send output to Galaxy' checkbox is checked. The 'get output' button is visible at the bottom of the configuration section.

5. Select appropriate parameters for your experiment.

6. Click "send query to Galaxy"

The screenshot shows the Galaxy web interface with the 'Output refGene as BED' tool selected. The left sidebar lists various tools under 'Get Data'. The main panel shows the 'Output refGene as BED' configuration page. The 'name' is 'refGene', 'description' is 'Table browser query on refGene', and 'visibility' is 'public'. Under 'Create one BED record per:', the 'Whole Gene' radio button is selected. The 'Send query to Galaxy' button is visible at the bottom of the configuration section.

14. Select "BED" and the URL to upload.
eg. "http://grimmond.imb.uq.edu.au/uniqueome/downloads/hg19_uniqueome.unique_starts.color-space.35.3.negative.BED.gz"

15. Select the Genome
eg. hg19

16. Press "Execute"

13. Upload your "starts" data. Click "Get Data" then "Upload File".

18. Select the RefSeq exons, then the uniqueome exons

19. Press "Execute". Repeat for the other strand.

17. Determine the number of unique bases in the RefSeq exons. Click "Operate on Genomic Intervals", then click "Coverage".

21. Select Join results for both strands.

22. Press "Execute"

20. Concatenate the results of both strands. Click "Concatenate".

24. Type c3-c2 (end coordinate minus start coordinate).

25. Select concatenated results.

26. Select "YES".

27. Press "Execute"

23. Calculate the total length of each exon. Select "Text Manipulation" and then "Compute".

29. Select the concatenated results, and group by ID (column 4). Select the “Add new Operation” button.

30. Select “Sum” by column 7 to add the unique lengths of exons per gene.

31. Select “Sum” by column 9 to add the total lengths of exons per gene.

32. Press “Execute”

28. Sum all the exon based values. Click “Join, Subtract, and Group” then click “Group”.

33. Download the results to your computer as a tab delimited text file. Select the results. eg. “11:Group on data 10”.

34. Press “Save”