Galaxy

Galaxy 101: Genomic Intervals

www.galaxyproject.org

@jxtx / #usegalaxy

Basic Analysis

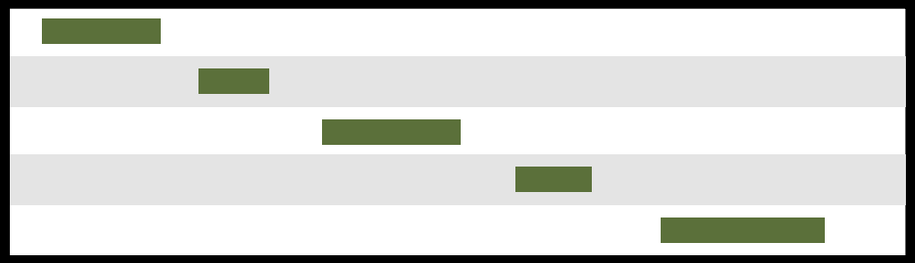
On human chromosome 22,
which coding exons have the most
repeats in them?

(~ http://usegalaxy.org/galaxy101)

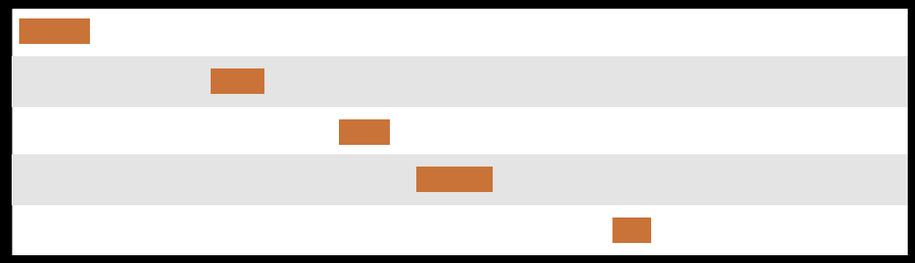
Exons & Repeats: A General Plan

- Get some data
  - Get Data → UCSC Table Browser
- Identify which exons have Repeats
- Count Repeats per exon
- Save, download, ... exons with most Repeats

(~ http://usegalaxy.org/galaxy101)

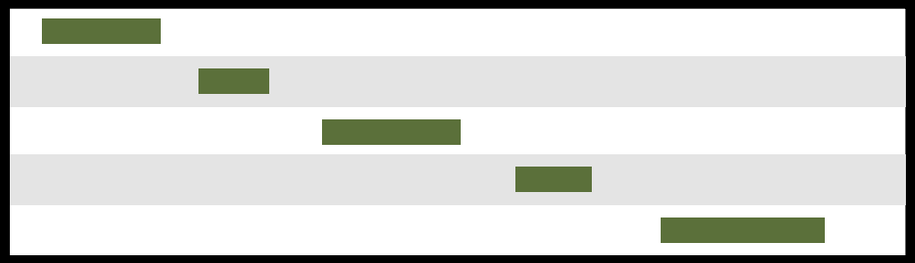


Exons

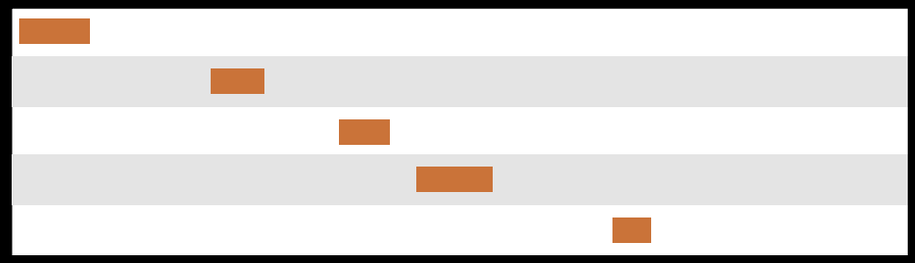


Repeats

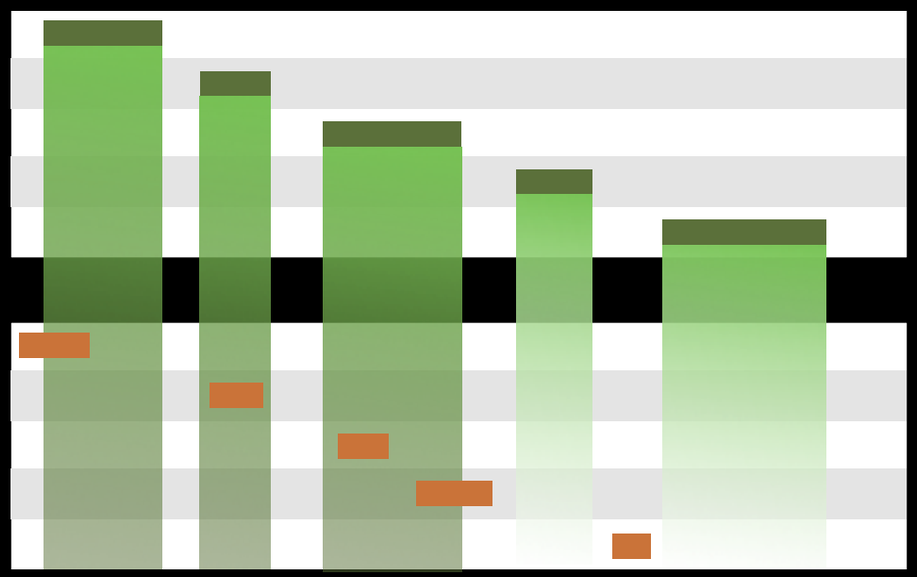
(Identify which exons have Repeats)



Exons

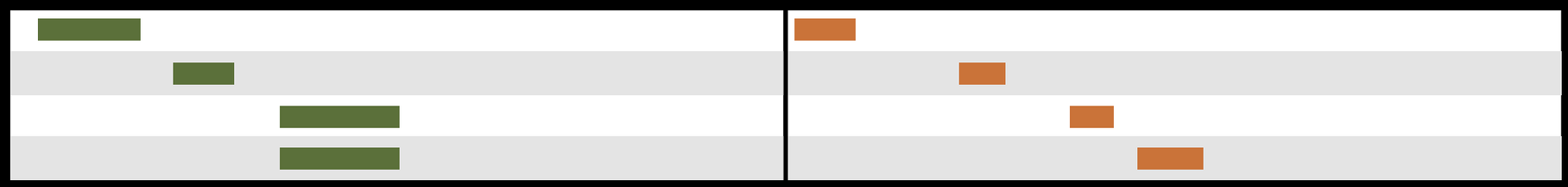


Repeats



Exons

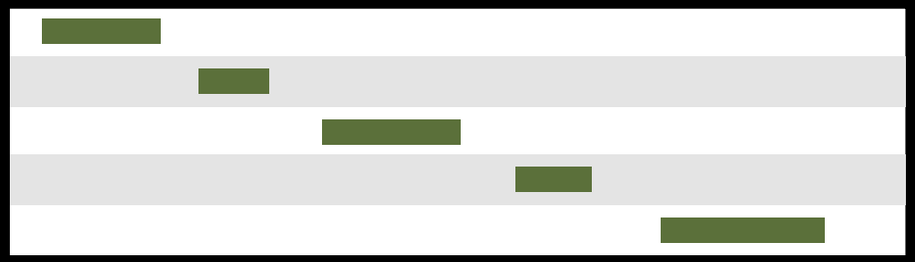
Repeats



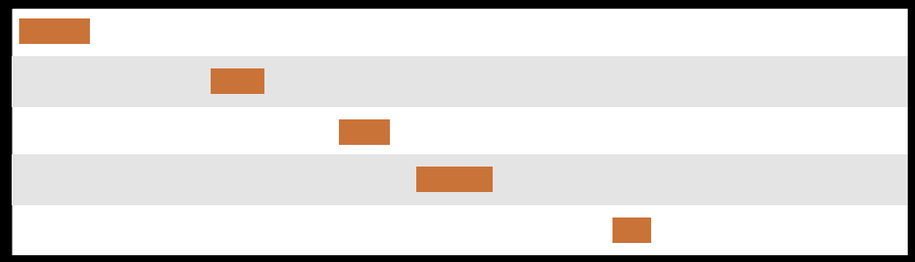
Overlap pairings

Operate on Genomic Intervals → **Join**

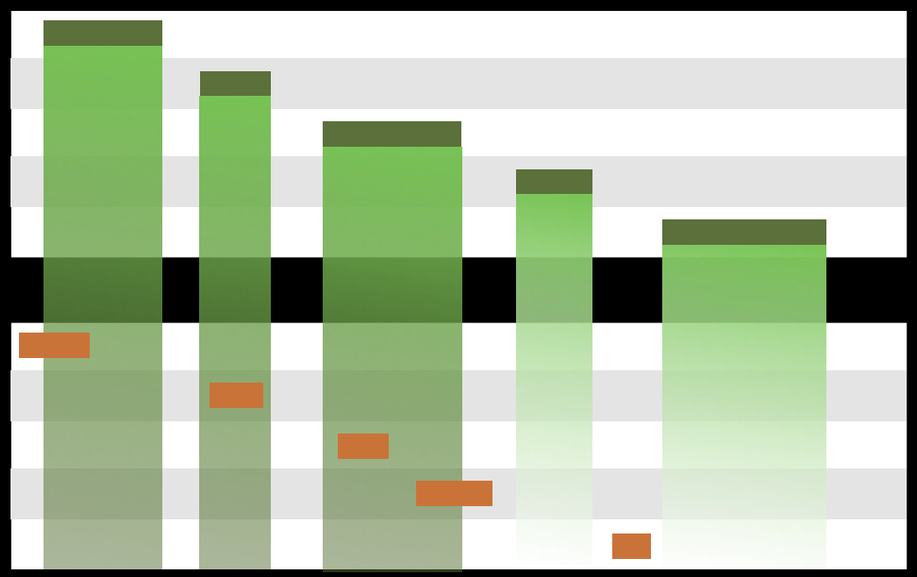
(Identify which exons have Repeats)



Exons



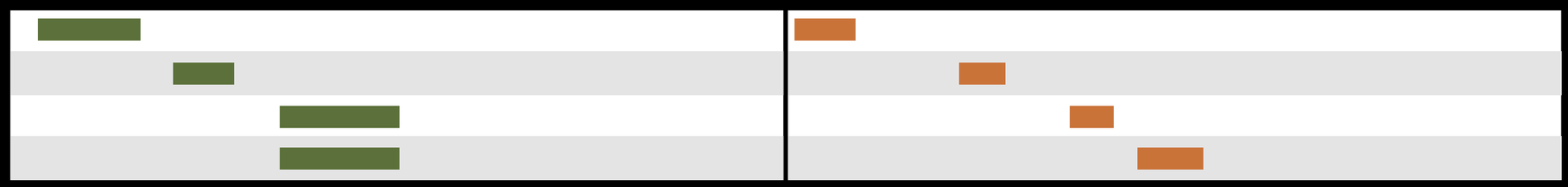
Repeats



Exons

Repeats

Overlap pairings

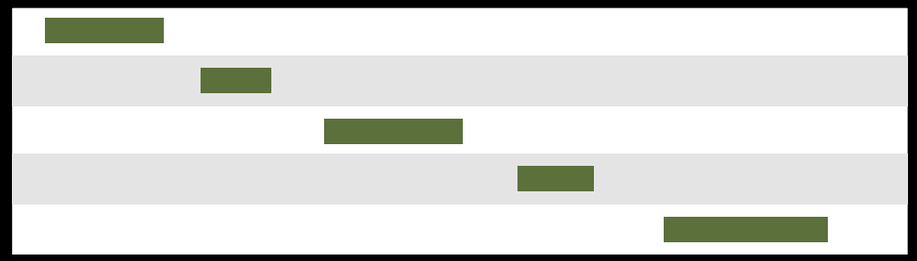


Exon overlap counts

Join, Subtract, and Group → Group

(Count Repeats per exon)

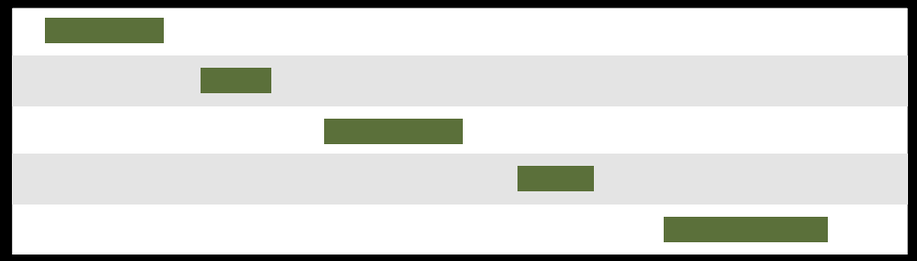
Exon overlap counts



Exons

We’ve answered our question, but we can do better.
Incorporate the overlap count with rest of Exon information

Exon overlap counts



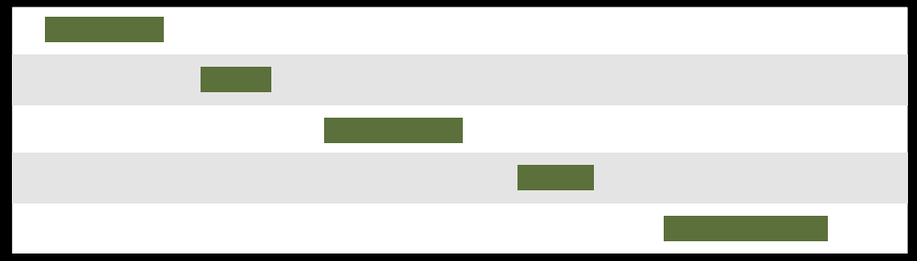
Exons

Join on exon name

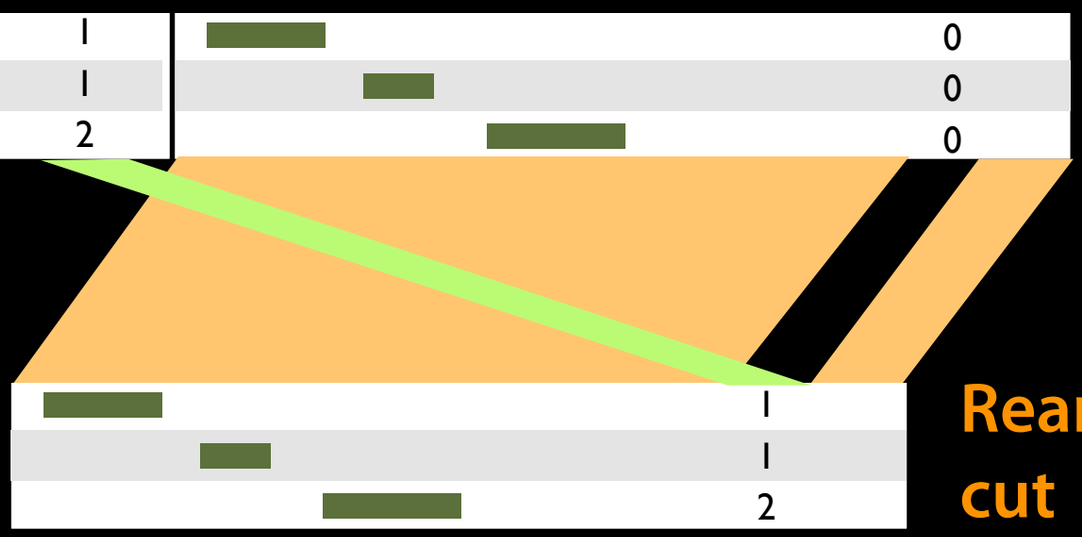
Join, Subtract, and Group → Join

(Incorporate the overlap count with rest of Exon information)

Exon overlap counts



Exons



Join on exon name

Rearrange columns w/ cut

Text Manipulation → Cut

(Incorporate the overlap count with rest of Exon information)

Summary

Interactive analysis in Galaxy is performed by using *tools* to operate on *datasets*

Datasets are immutable, and running tools always creates one or more new datasets

Datasets are available through the *history*, which provides complete provenance for each dataset