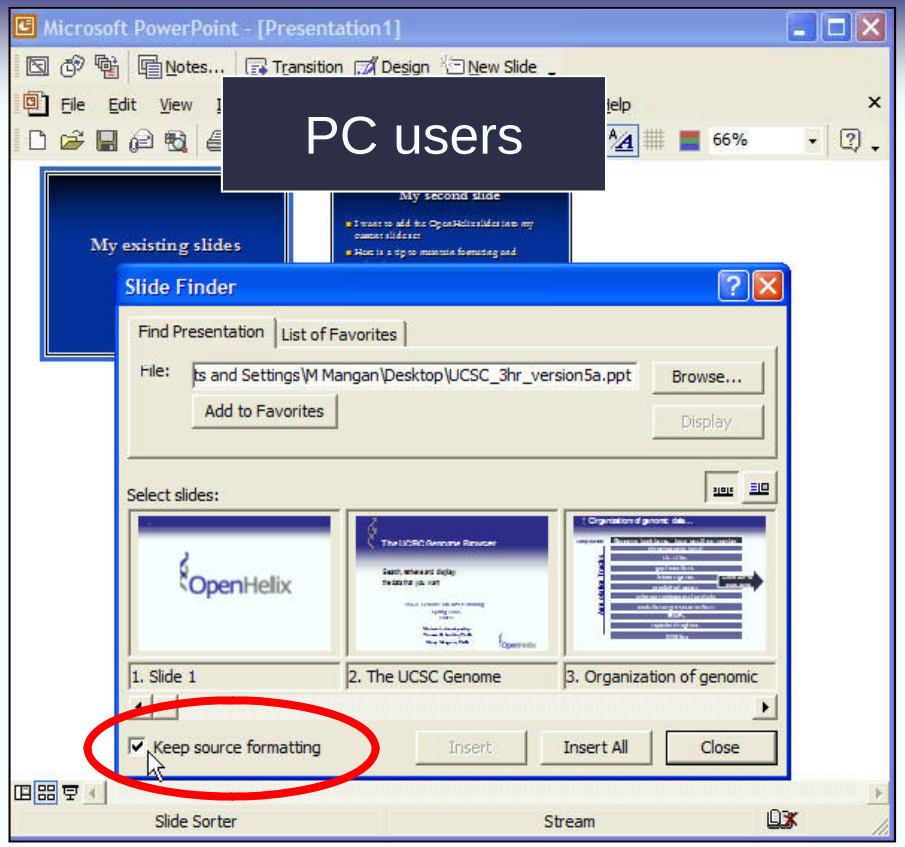


Copyright OpenHelix. No use or reproduction without express written consent

Important note to slide users:



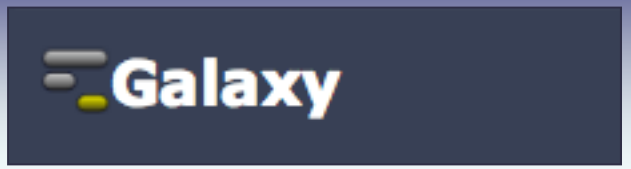


- **To maintain the color schemes/cues and the animations, if you import these slides into other slide sets please click the checkbox in the PowerPoint Insert window that maintains slide format. Otherwise important information may be lost.**

Copyright OpenHelix. No use or reproduction without express written consent



Galaxy



Materials prepared by:
Warren C. Lathe III, Ph.D.
www.openhelix.com

Updated: Q4 2010



Galaxy Agenda

- Introduction and Credits
- Interface
- Getting Data
- Preparing Data
- Analyzing Data
- Histories & Workflows
- Summary
- Exercise

Galaxy: http://galaxy.psu.edu/

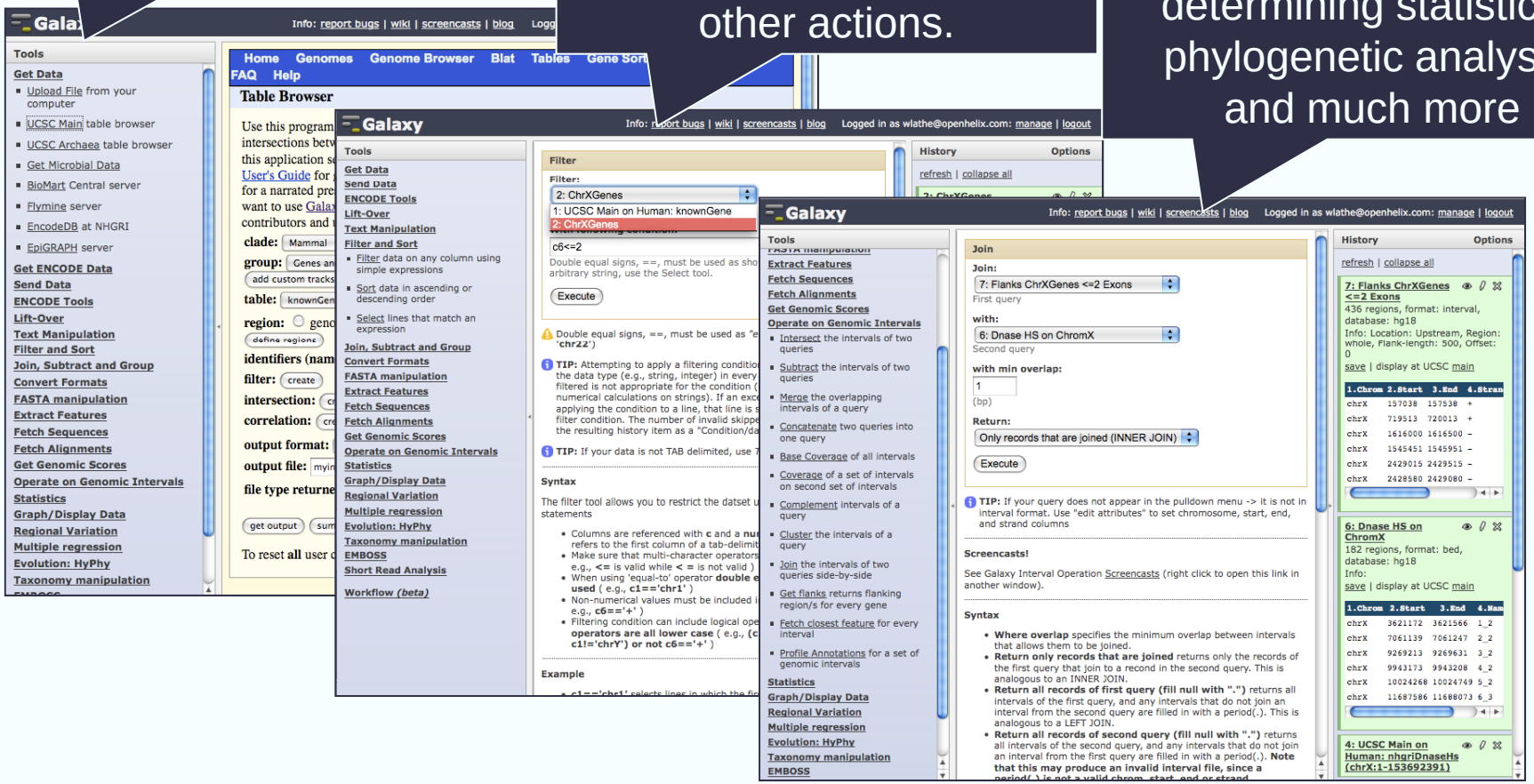
Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Conceptual Framework

Obtain data from many data sources including the UCSC Table Browser, BioMart, WormBase, or your own data.

Prepare data for further analysis by rearranging or cutting data columns, filtering data and many other actions.

Analyze data by finding overlapping regions, determining statistics, phylogenetic analysis and much more



Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Start



http://galaxy.psu.edu/

Use it now!

Galaxy = collaboration and reproducibility

Designed for biologists and developers

Why did we do it?

You are an experimental biologist. You keep watching databases fill with more and more data. You keep thinking: *even if I knew how to use Excel as a pro, it would probably not load 12,435,654 SNPs*. So how do you perform analyses without calling somebody on the Computer Science side of campus? Suppose you want to find human promoters with the highest SNP density. There is no straightforward way of doing it without learning programming first. And this is why...

- **Databases are not analyses tools** | Databases are where you get the data. Browsers are where you visualize the results. For a bench biologist there is not much in between besides spreadsheets or Perl scripting.
- **No tools for new datatypes** | Some datatypes generated by high throughput genomics are so new that there are no tools to analyze them. For example, how do you extract sequences of coding exons from the latest 28-way alignments of vertebrate genomes or analyze quality scores from 454/Solexa/SOLiD? With Galaxy.
- **Genomics is not really reproducible** | The Methods section of too many papers sound like *the data were analyzed using a collection of in-house scripts*. How do you repeat such a study? Galaxy saves every step of your analysis and allows you to share these workflows with others.
- **Too many tools** | *Bioinformatics* publishes hundreds of application notes per year. How does one know which tool to use? Galaxy integrates a multitude of different tools by giving them the same "look and feel" and linking them to data warehouses.

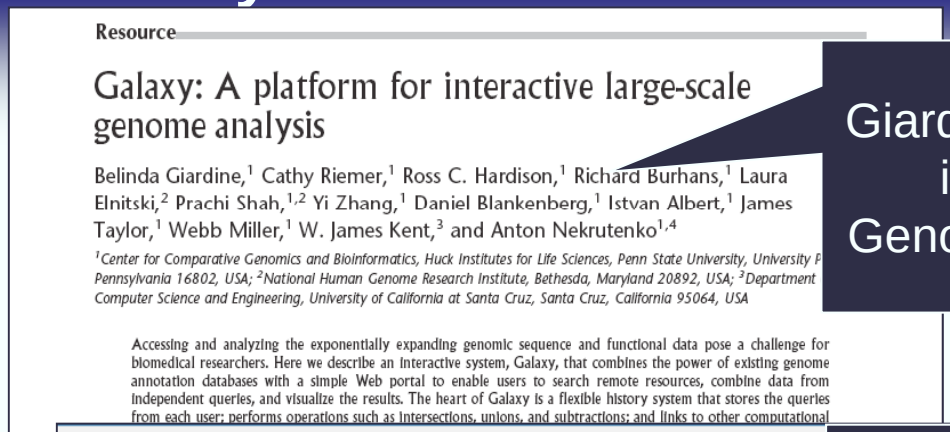
Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Team



Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Publications



Resource

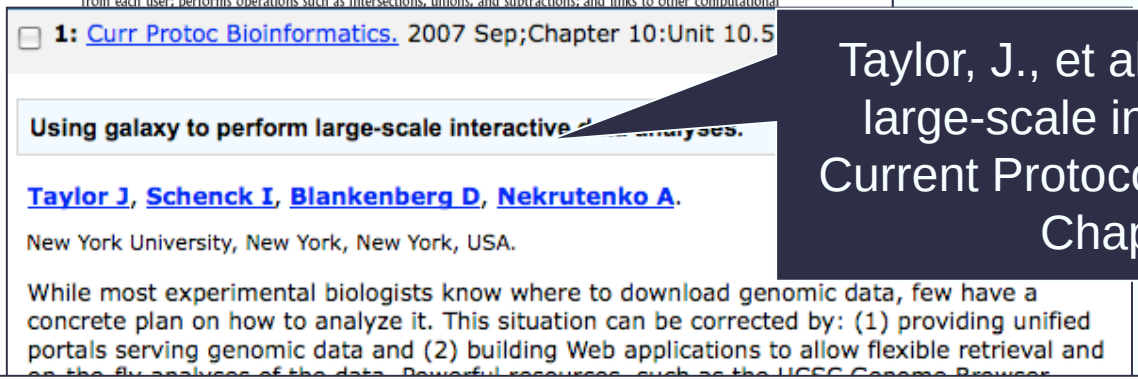
Galaxy: A platform for interactive large-scale genome analysis

Belinda Giardine,[1] Cathy Riemer,[1] Ross C. Hardison,[1] Richard Burhans,[1] Laura Elnitski,[2] Prachi Shah,[1,2] Yi Zhang,[1] Daniel Blankenberg,[1] Istvan Albert,[1] James Taylor,[1] Webb Miller,[1] W. James Kent,[3] and Anton Nekrutenko[1,4]

[1]Center for Comparative Genomics and Bioinformatics, Huck Institutes for Life Sciences, Penn State University, University P Pennsylvania 16802, USA; [2]National Human Genome Research Institute, Bethesda, Maryland 20892, USA; [3]Department Computer Science and Engineering, University of California at Santa Cruz, Santa Cruz, California 95064, USA

Accessing and analyzing the exponentially expanding genomic sequence and functional data pose a challenge for biomedical researchers. Here we describe an interactive system, Galaxy, that combines the power of existing genome annotation databases with a simple Web portal to enable users to search remote resources, combine data from independent queries, and visualize the results. The heart of Galaxy is a flexible history system that stores the queries from each user; performs operations such as intersections, unions, and subtractions; and links to other computational

Giardine, B., et al. Galaxy: a platform for interactive large-scale analysis. Genome Research (2005) 15:1451-1455



1: Curr Protoc Bioinformatics. 2007 Sep;Chapter 10:Unit 10.5

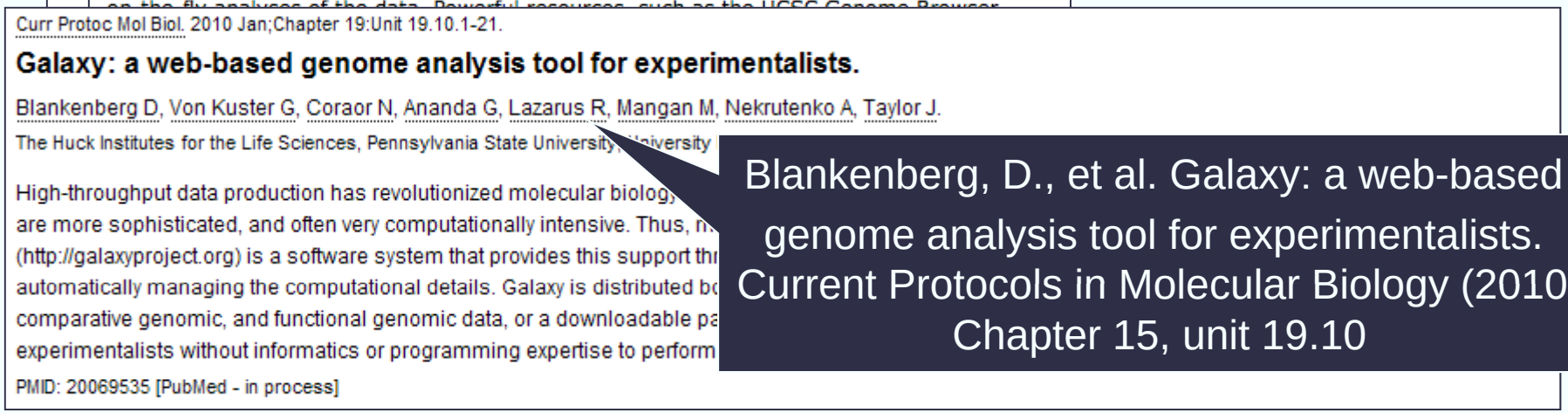
Using galaxy to perform large-scale interactive

Taylor J, Schenck I, Blankenberg D, Nekrutenko A.

New York University, New York, New York, USA.

While most experimental biologists know where to download genomic data, few have a concrete plan on how to analyze it. This situation can be corrected by: (1) providing unified portals serving genomic data and (2) building Web applications to allow flexible retrieval and

Taylor, J., et al. Using Galaxy to perform large-scale interactive data analyses. Current Protocols in Bioinformatics (2007) Chapter 10, unit 10.5



Curr Protoc Mol Biol. 2010 Jan;Chapter 19:Unit 19.10.1-21.

Galaxy: a web-based genome analysis tool for experimentalists.

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J.

The Huck Institutes for the Life Sciences, Pennsylvania State University,

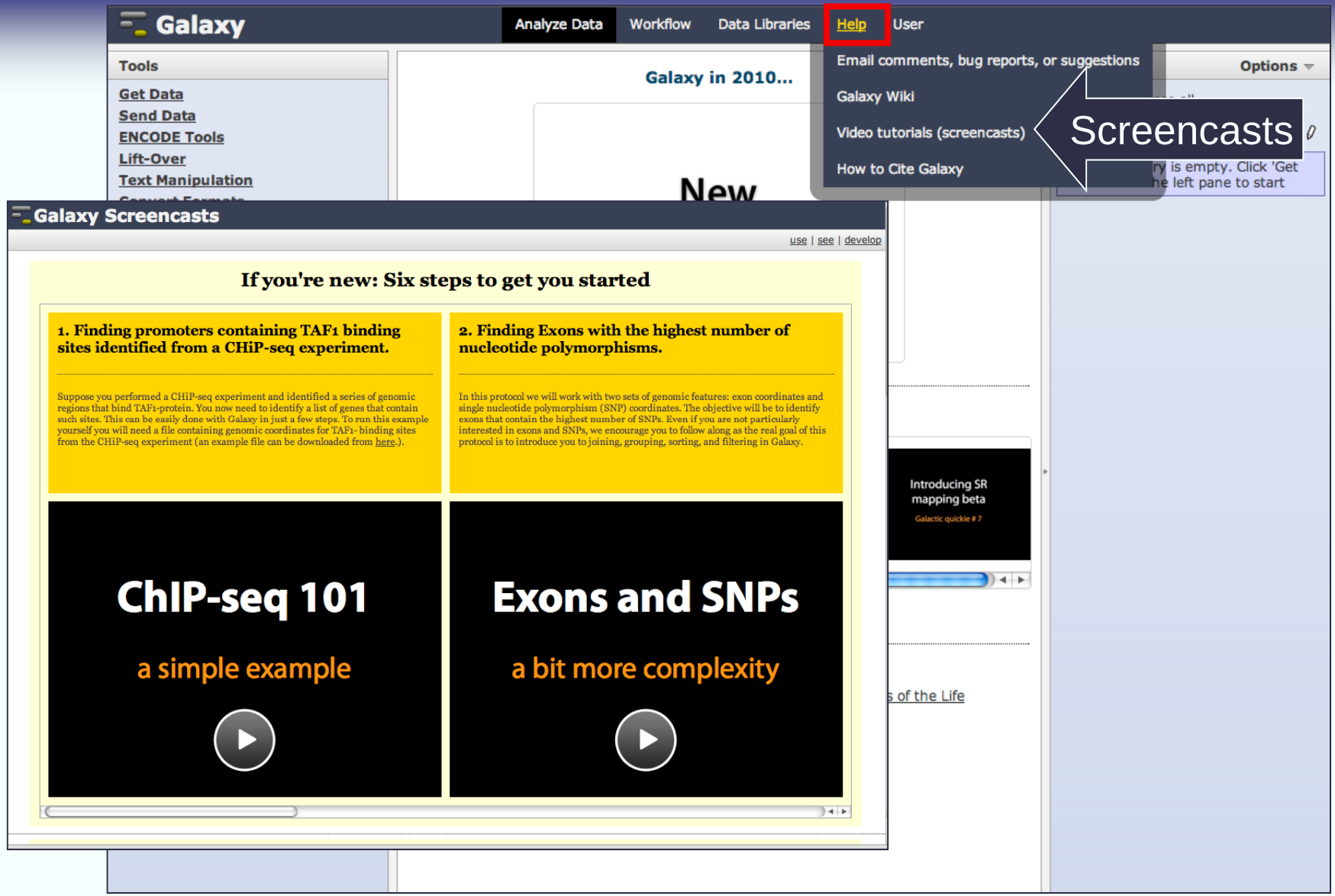
High-throughput data production has revolutionized molecular biology are more sophisticated, and often very computationally intensive. Thus, (http://galaxyproject.org) is a software system that provides this support th automatically managing the computational details. Galaxy is distributed b comparative genomic, and functional genomic data, or a downloadable pa experimentalists without informatics or programming expertise to perform

PMID: 20069535 [PubMed - in process]

Blankenberg, D., et al. Galaxy: a web-based genome analysis tool for experimentalists. Current Protocols in Molecular Biology (2010) Chapter 15, unit 19.10

Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Screencasts



Copyright OpenHelix. No use or reproduction without express written consent

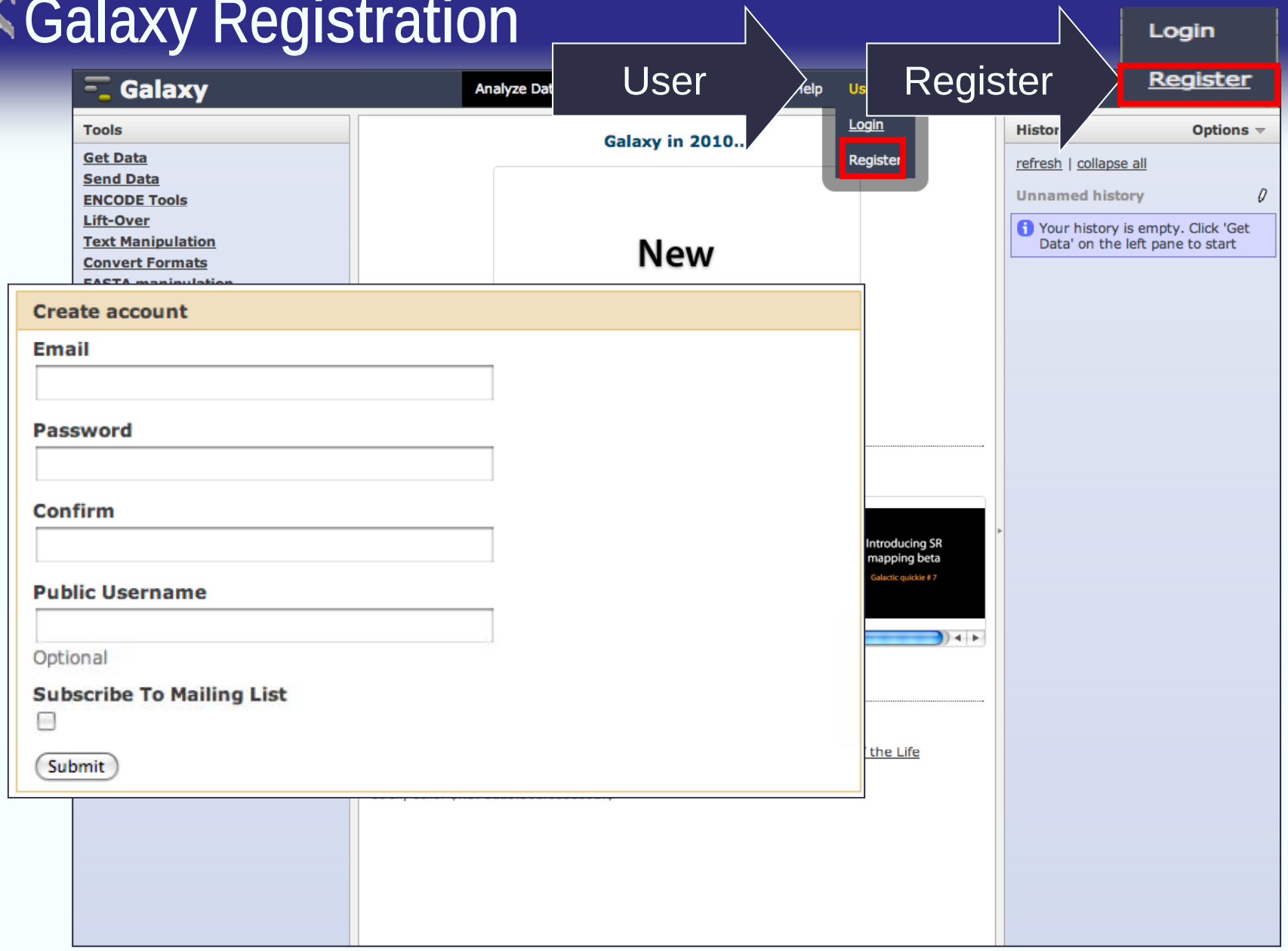
Galaxy Agenda

- Introduction and Credits
- Interface
- Getting Data
- Preparing Data
- Analyzing Data
- Histories & Workflows
- Summary
- Exercise

Galaxy: http://galaxy.psu.edu/

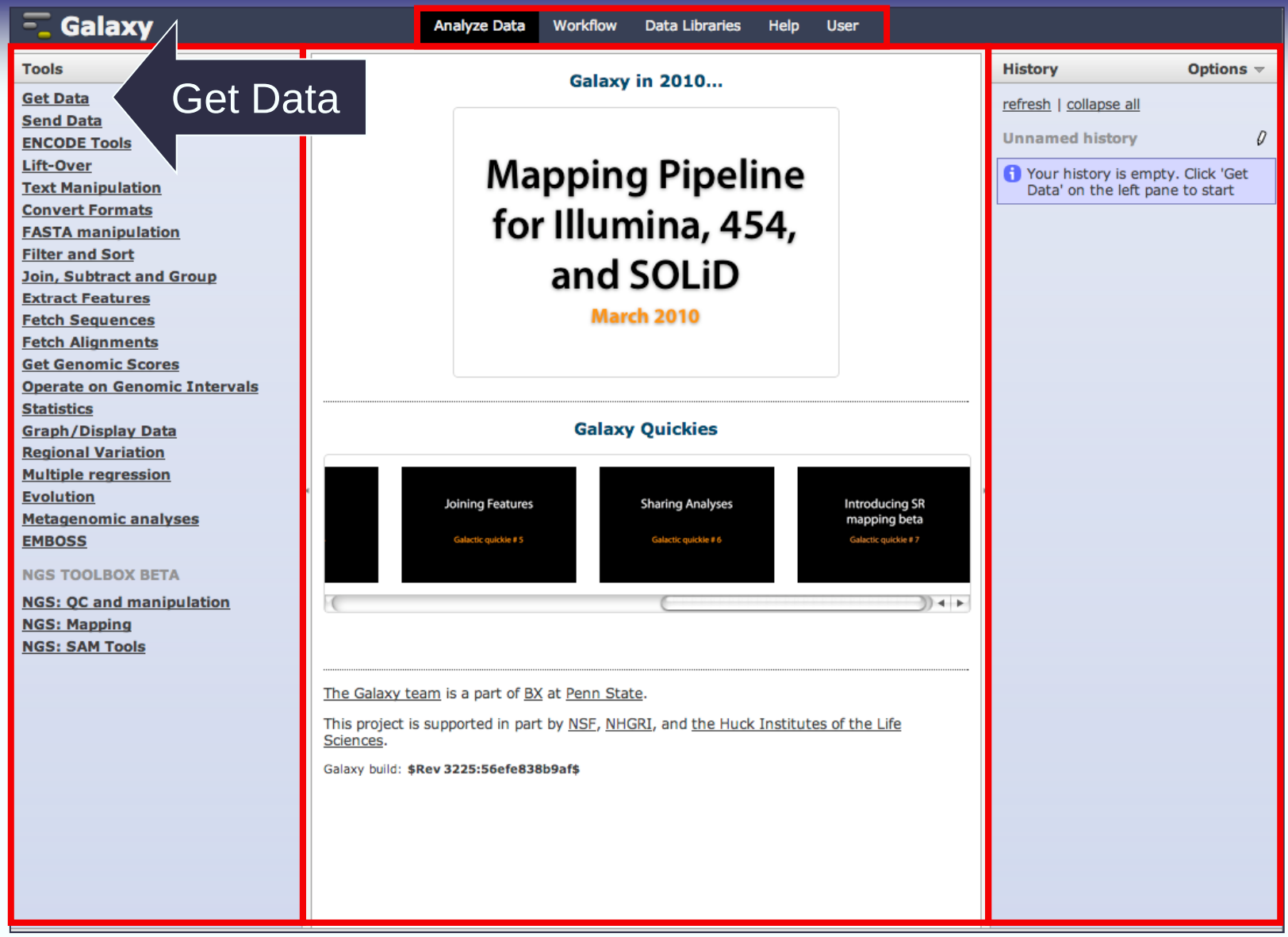
Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Registration



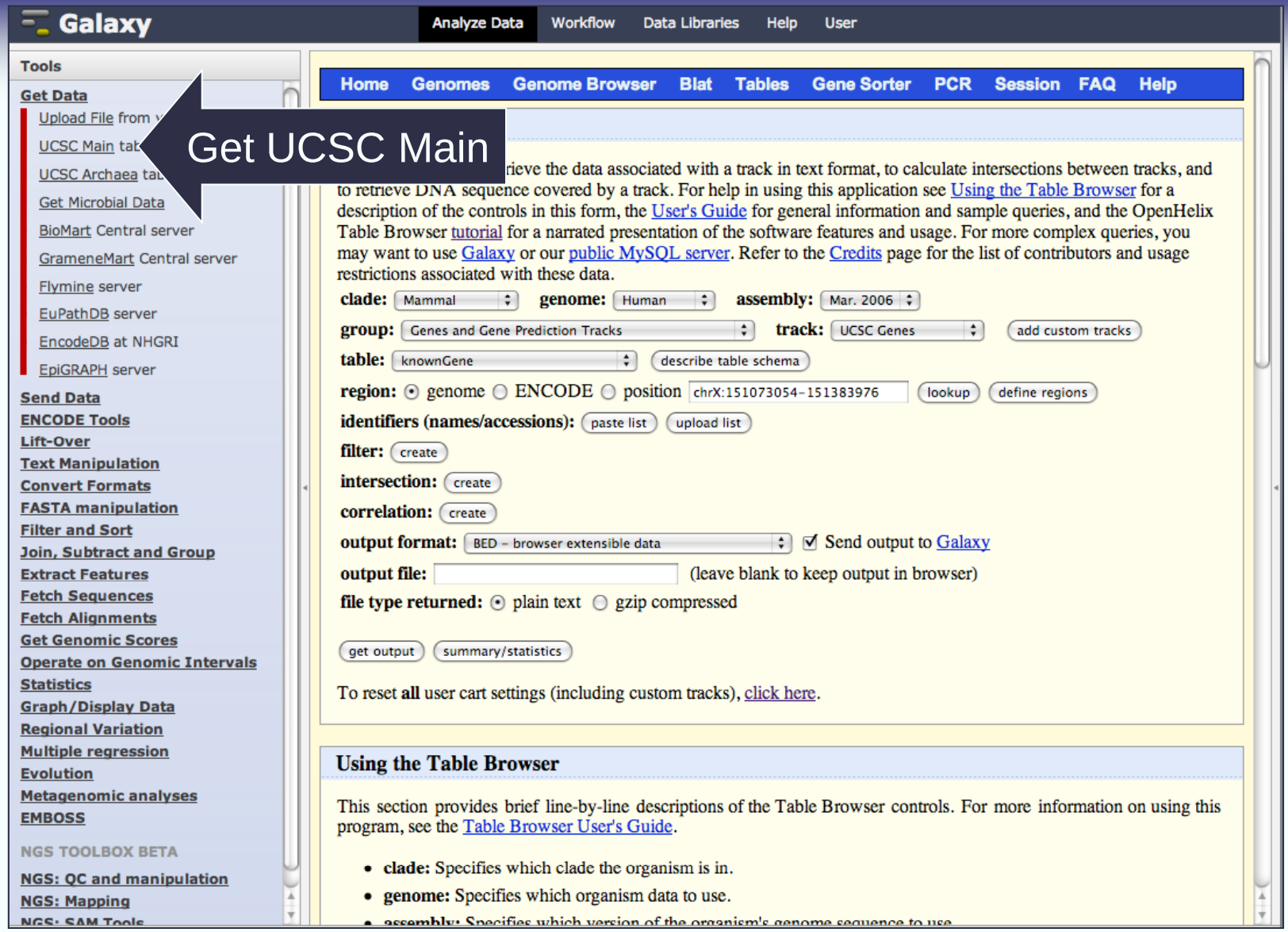
Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Interface Sections



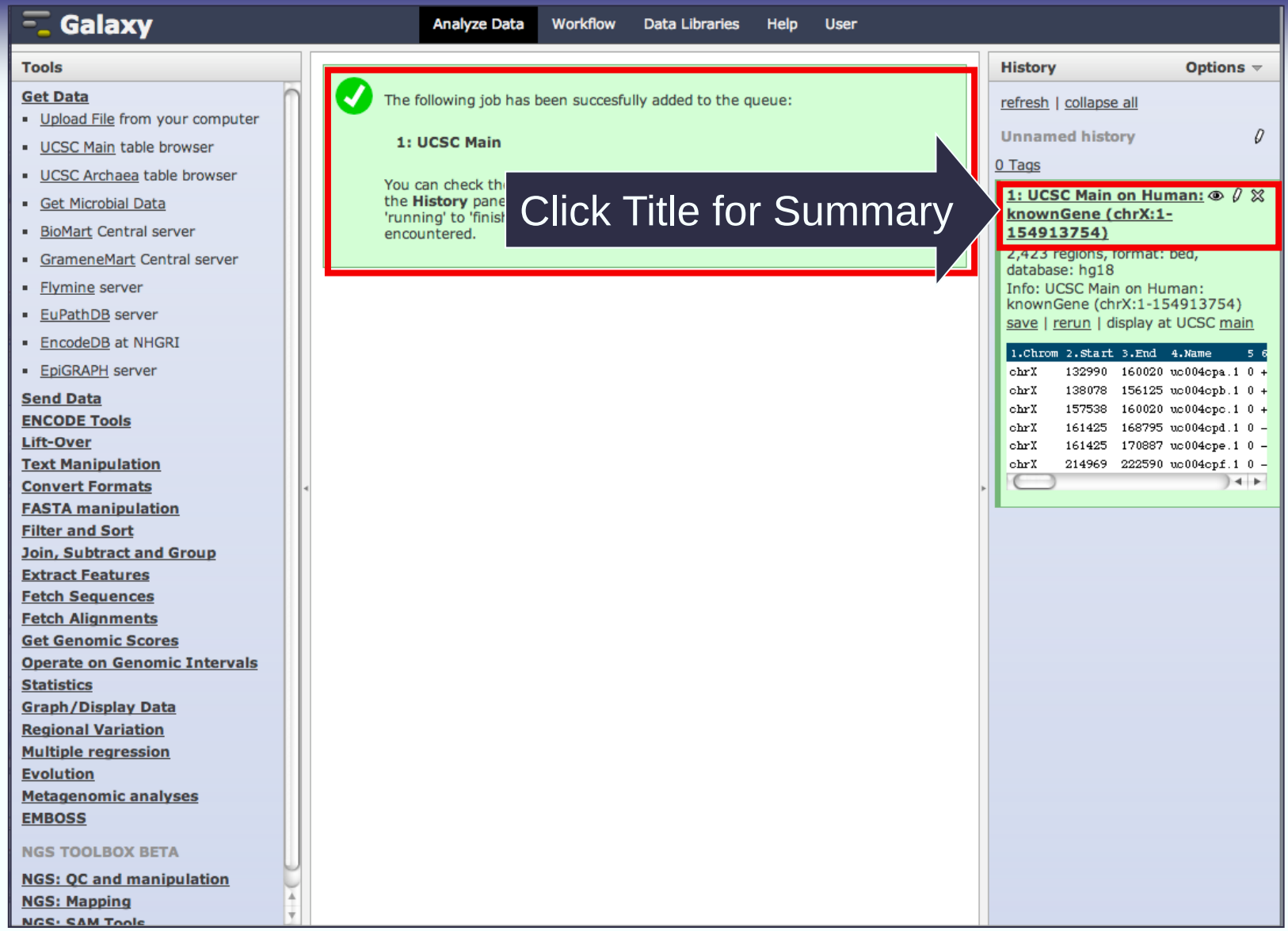
Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Interface Behavior



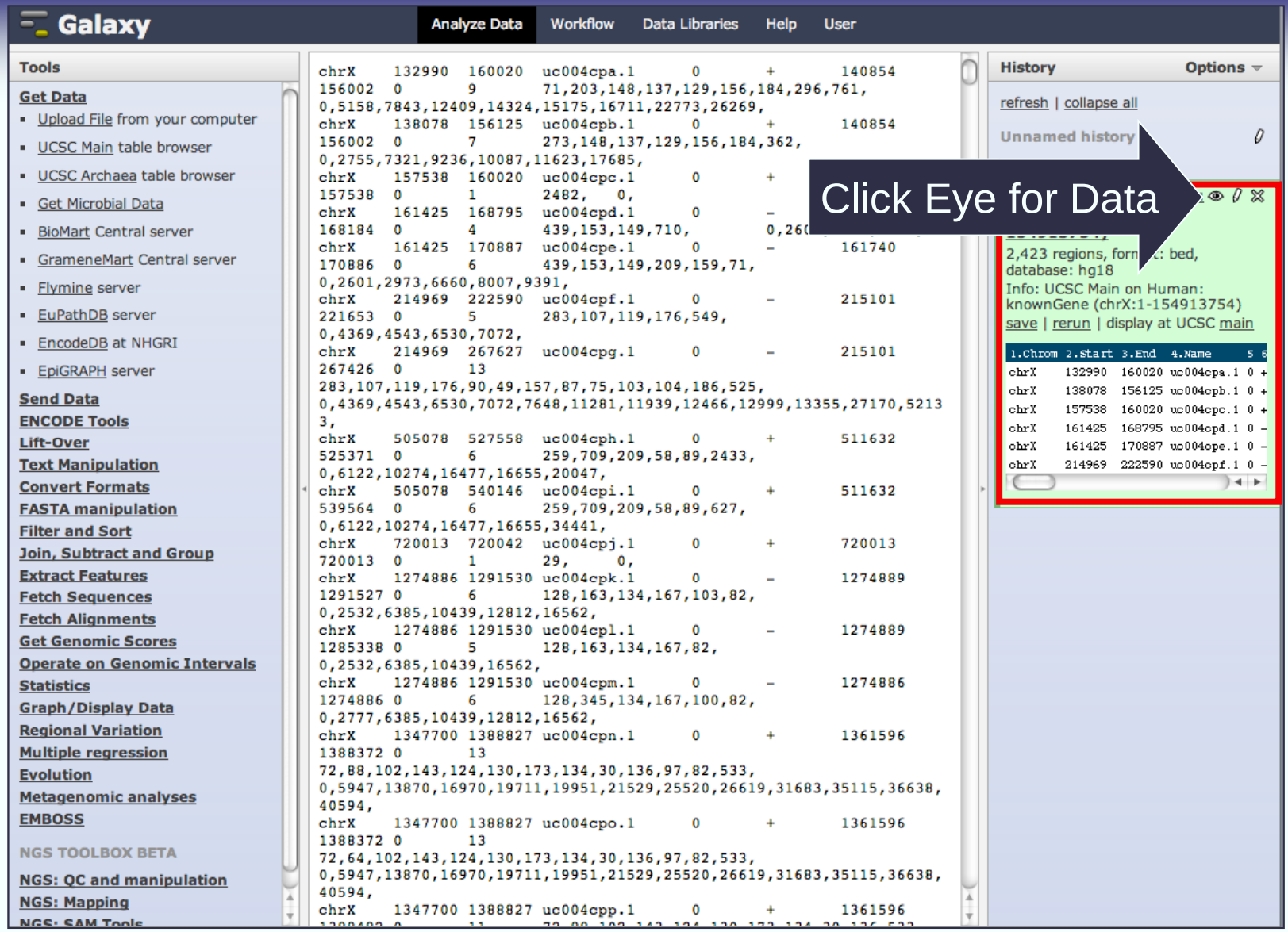
Copyright OpenHelix. No use or reproduction without express written consent

Data Uploaded



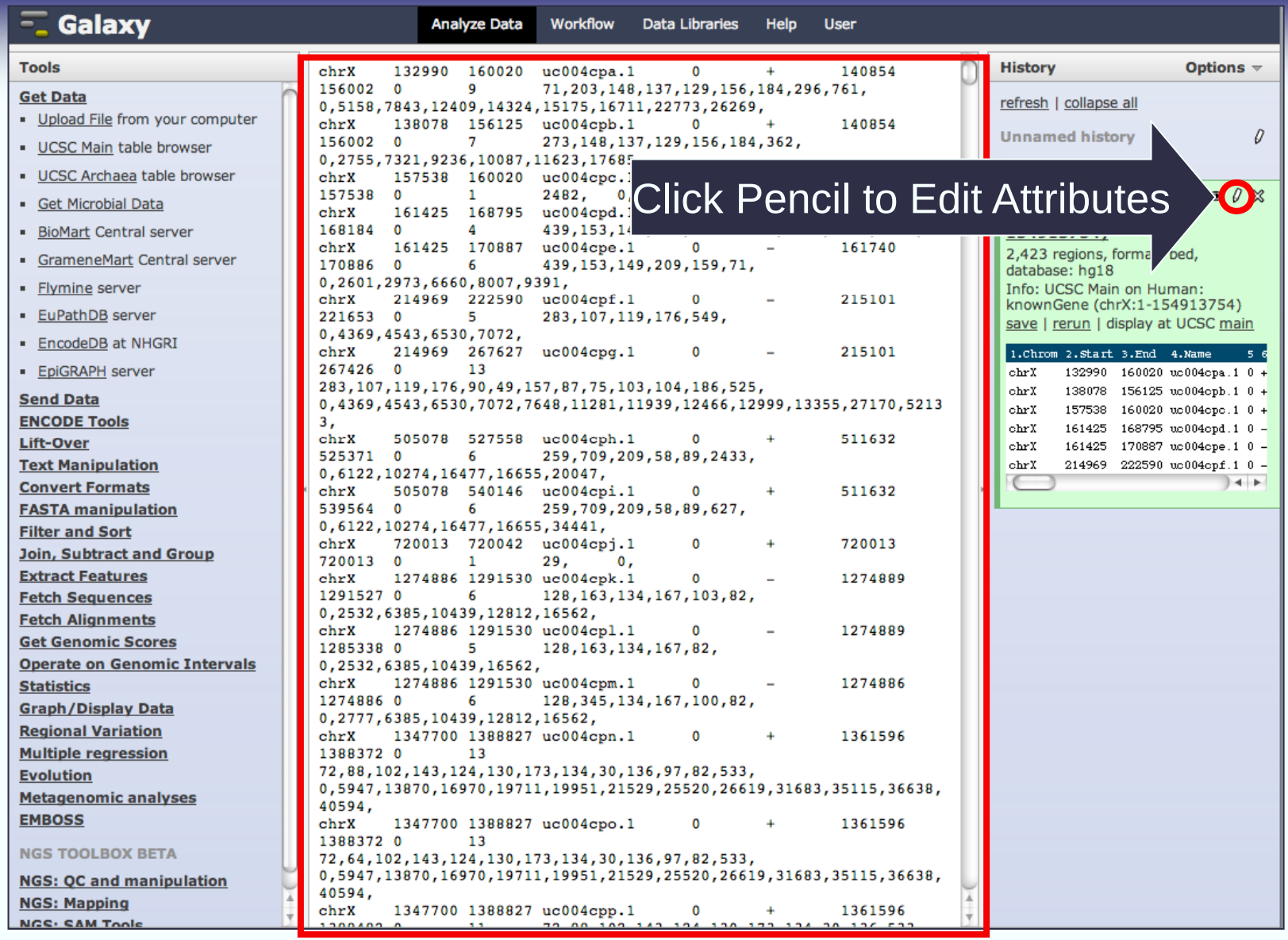
Copyright OpenHelix. No use or reproduction without express written consent

View Summary



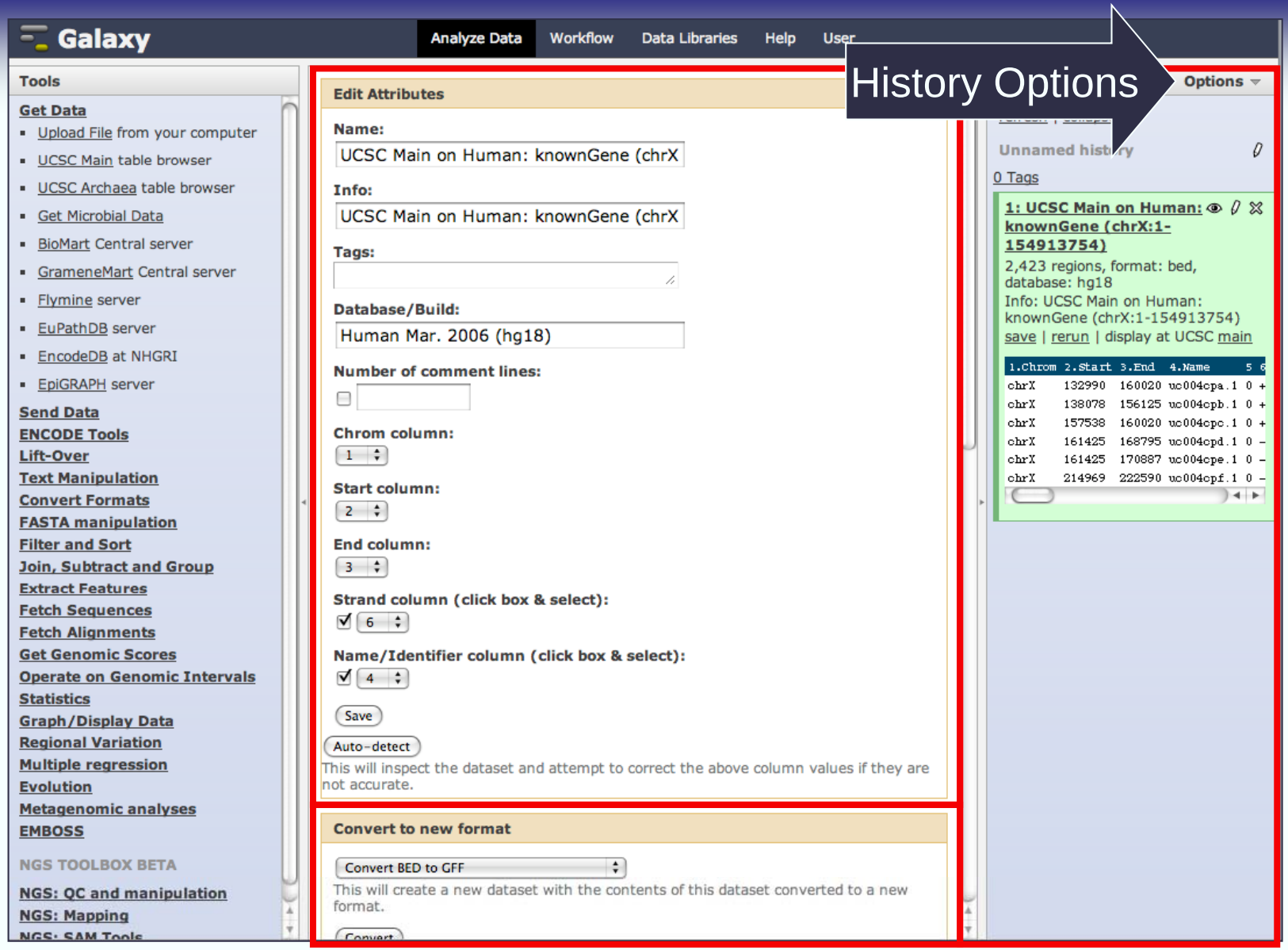
Copyright OpenHelix. No use or reproduction without express written consent

View Data



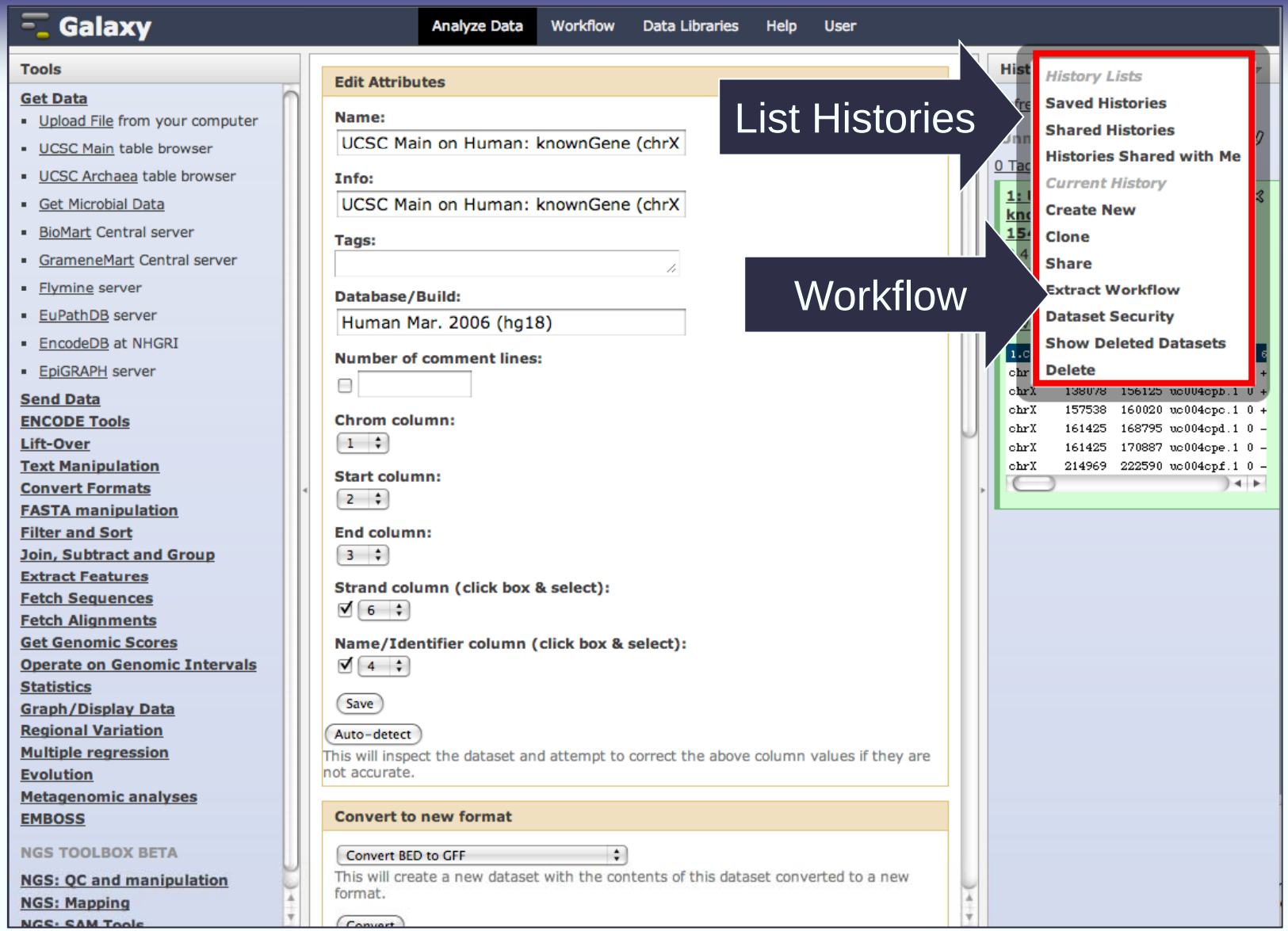
Copyright OpenHelix. No use or reproduction without express written consent

Edit Attributes



Copyright OpenHelix. No use or reproduction without express written consent

History and Workflow



Copyright OpenHelix. No use or reproduction without express written consent

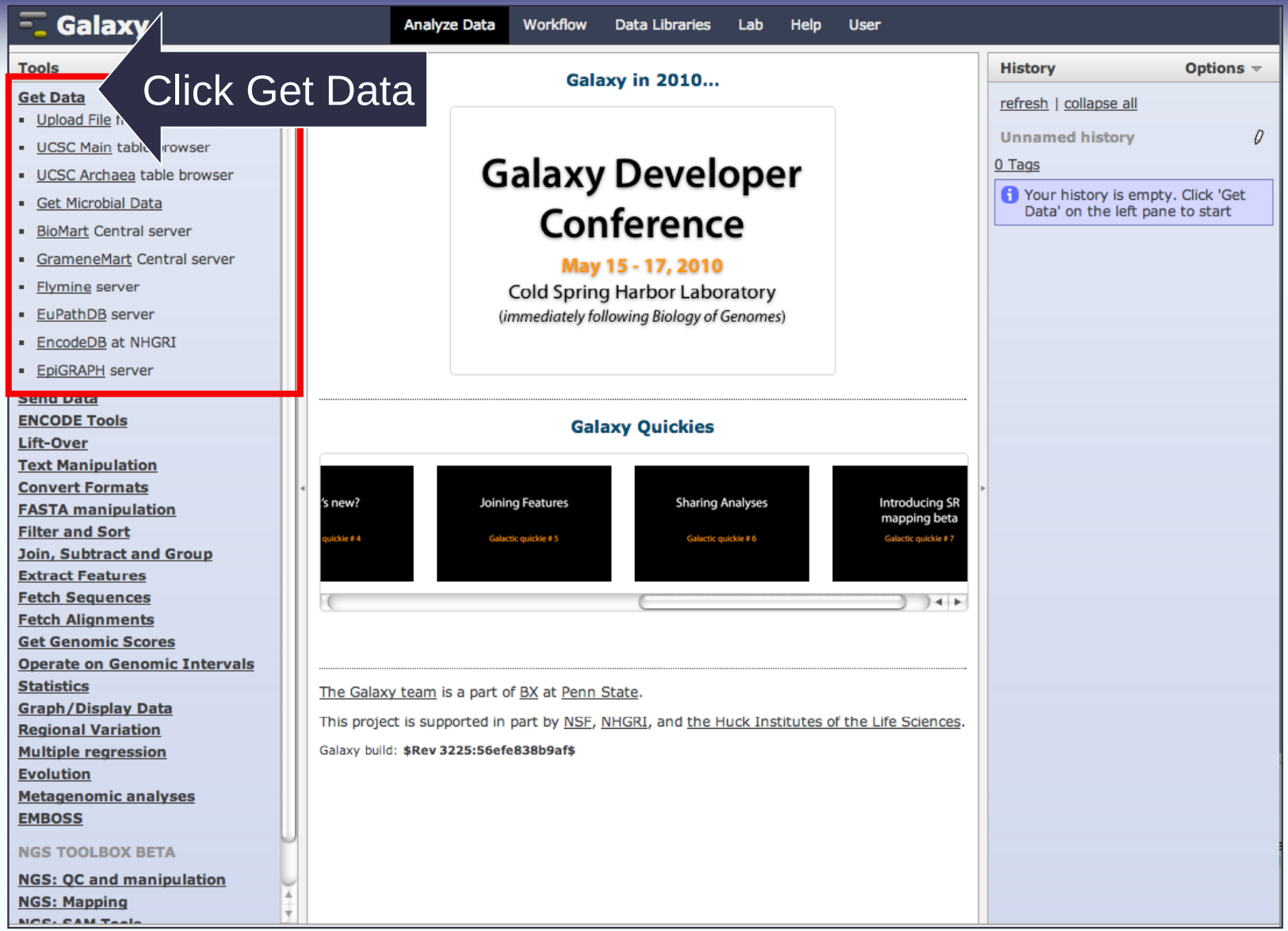
Galaxy Agenda

- Introduction and Credits
- Interface
- Getting Data
- Preparing Data
- Analyzing Data
- Histories & Workflows
- Summary
- Exercise

Galaxy: http://galaxy.psu.edu/

Copyright OpenHelix. No use or reproduction without express written consent

Getting Data



Copyright OpenHelix. No use or reproduction without express written consent

Getting Data: Table Browser



Copyright OpenHelix. No use or reproduction without express written consent

Getting Data: UCSC Table Browser

clade: Mammal
genome: Human
assembly: Mar. 2006

group: Genes and…
track: UCSC Genes
table: knownGene

region: position, chrX

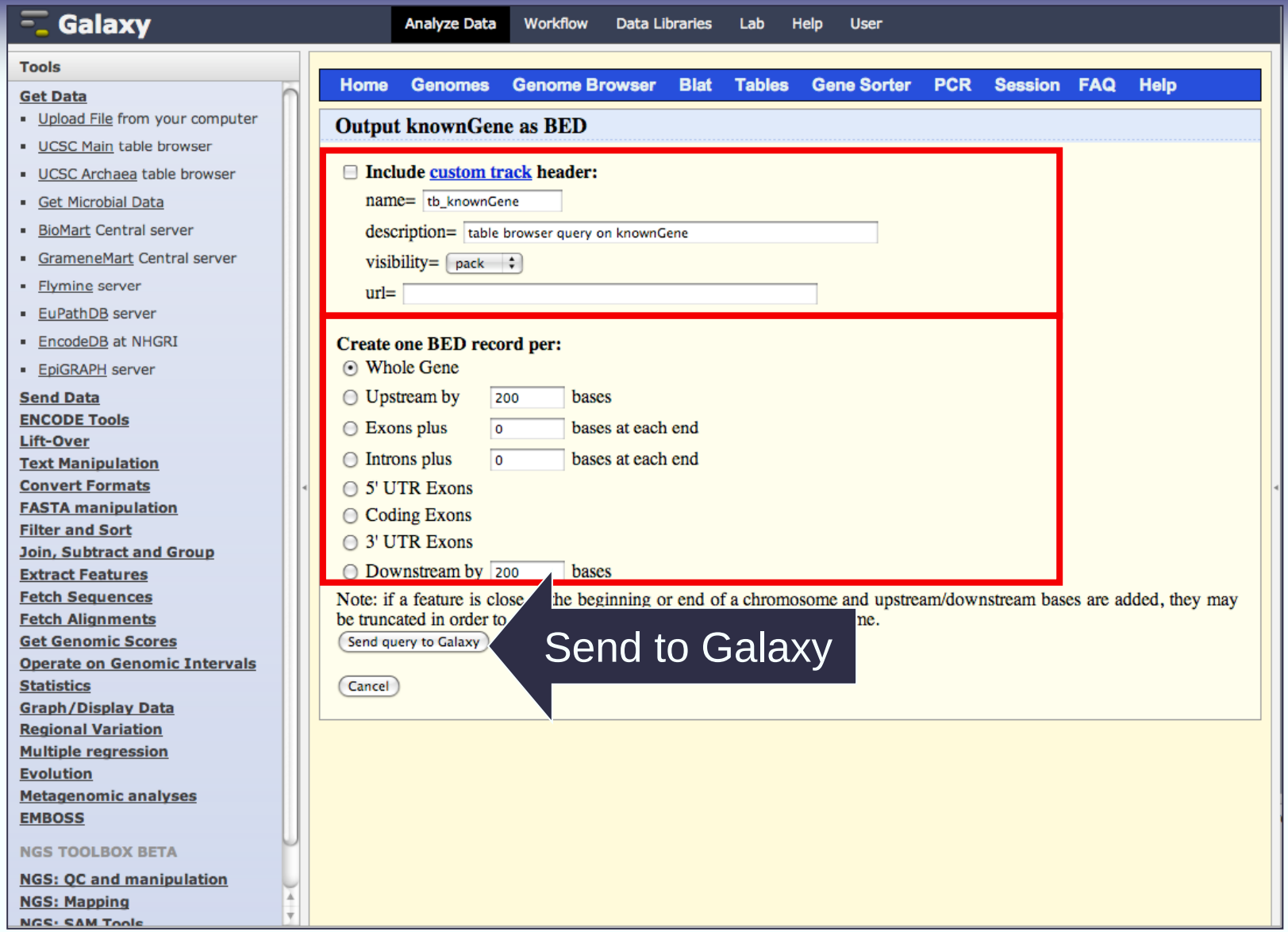
Output format: BED, and check Send output to Galaxy

Get Output



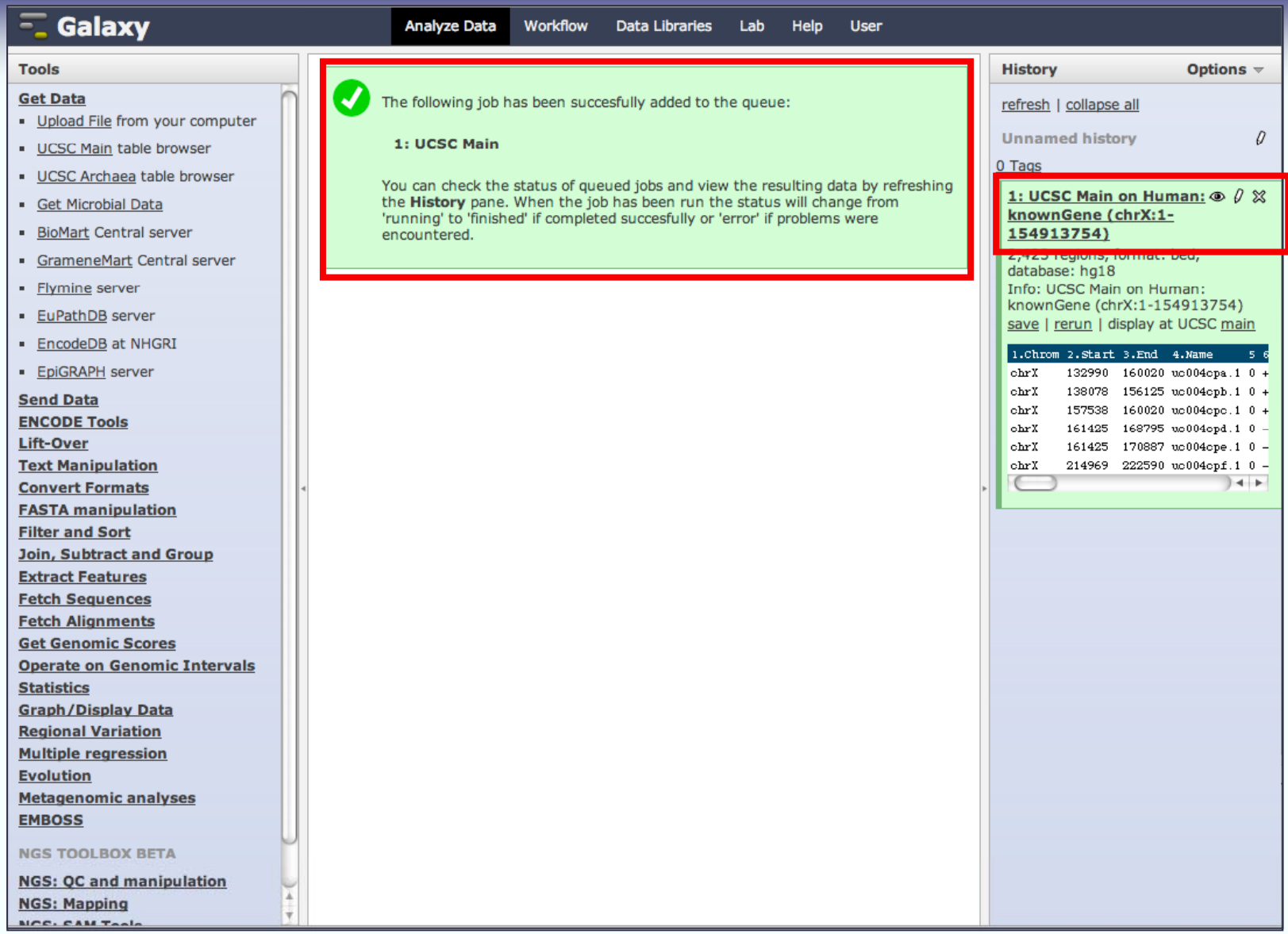
Copyright OpenHelix. No use or reproduction without express written consent

Getting Data: Send Table Results to Galaxy



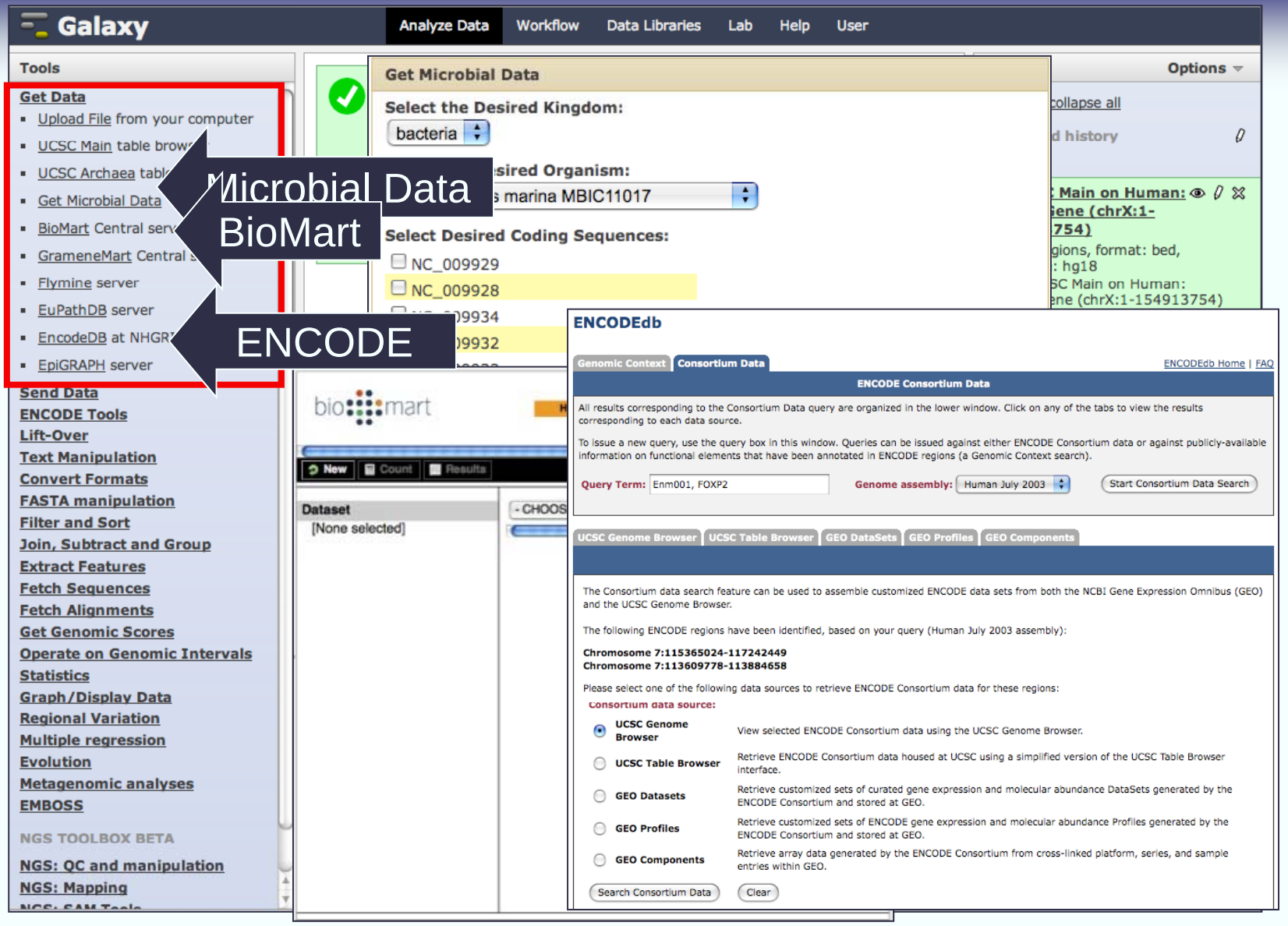
Copyright OpenHelix. No use or reproduction without express written consent

Getting Data: UCSC Table Browser Results



Copyright OpenHelix. No use or reproduction without express written consent

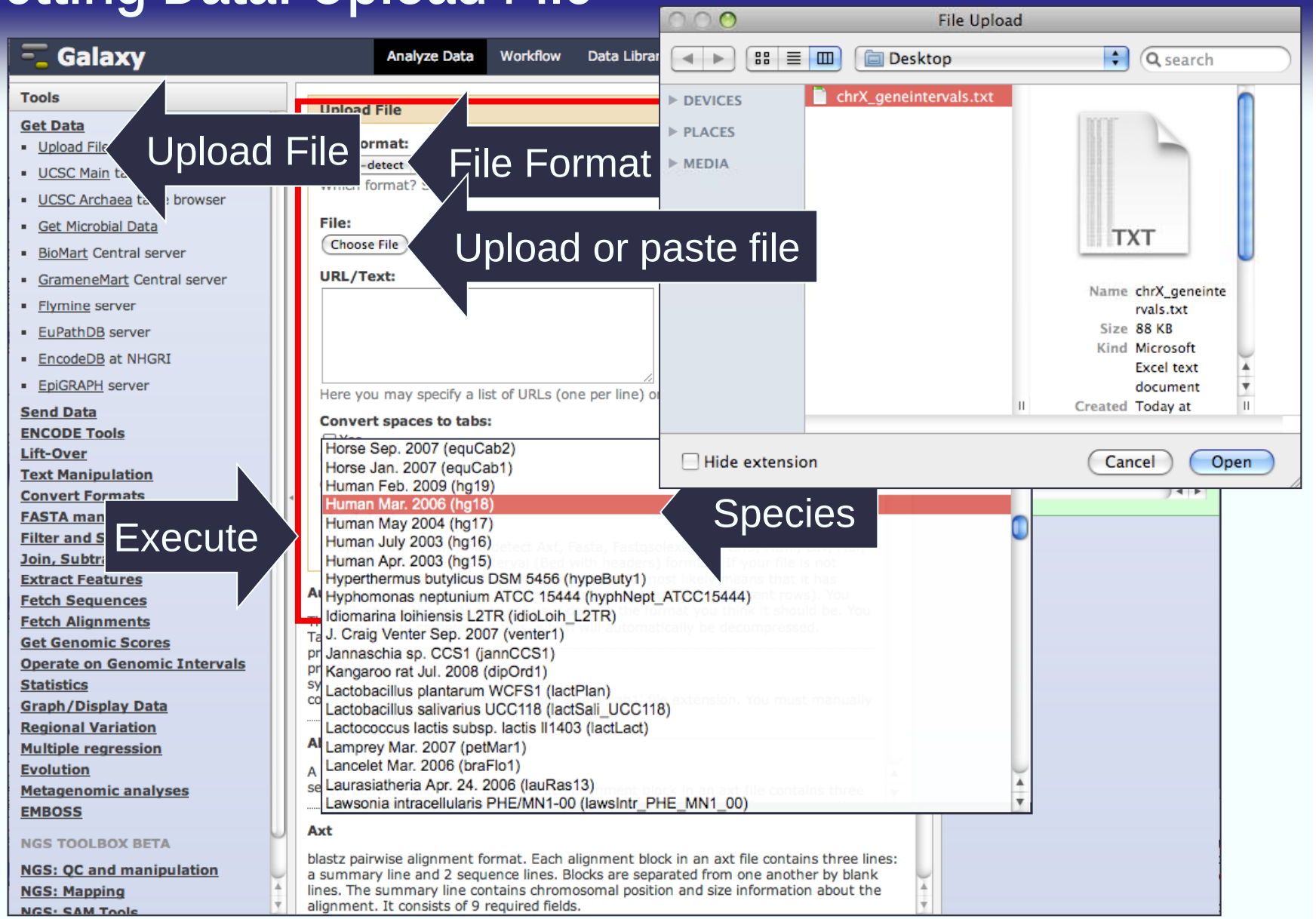
Getting Data: Additional Data Sources



Copyright OpenHelix. No use or reproduction without express written consent

Getting Data: Upload File

http://www.openhelix.com/galaxy



Copyright OpenHelix. No use or reproduction without express written consent

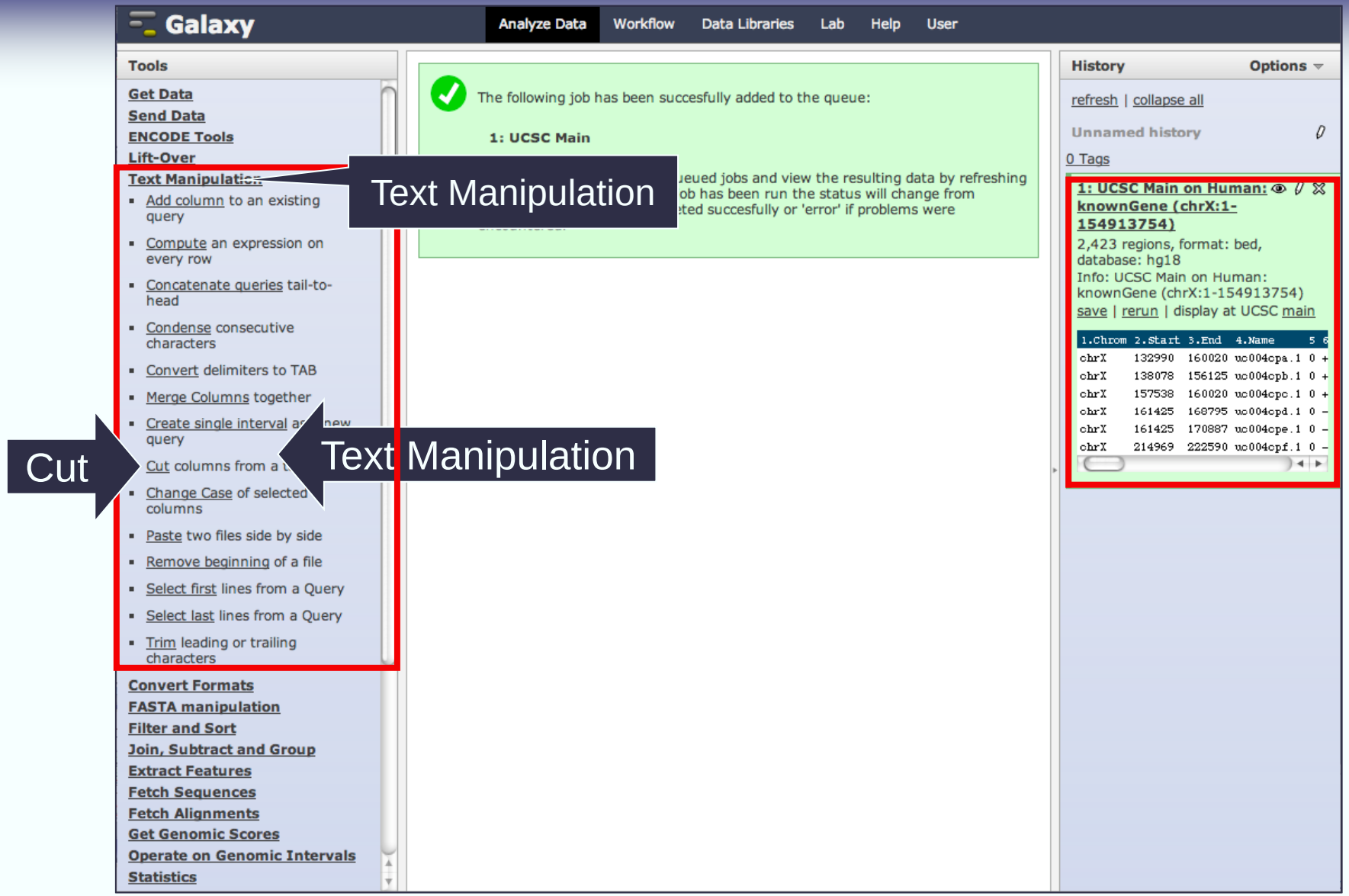
Galaxy Agenda

- Introduction and Credits
- Interface
- Getting Data
- Preparing Data
- Analyzing Data
- Histories & Workflows
- Summary
- Exercise

Galaxy: http://galaxy.psu.edu/

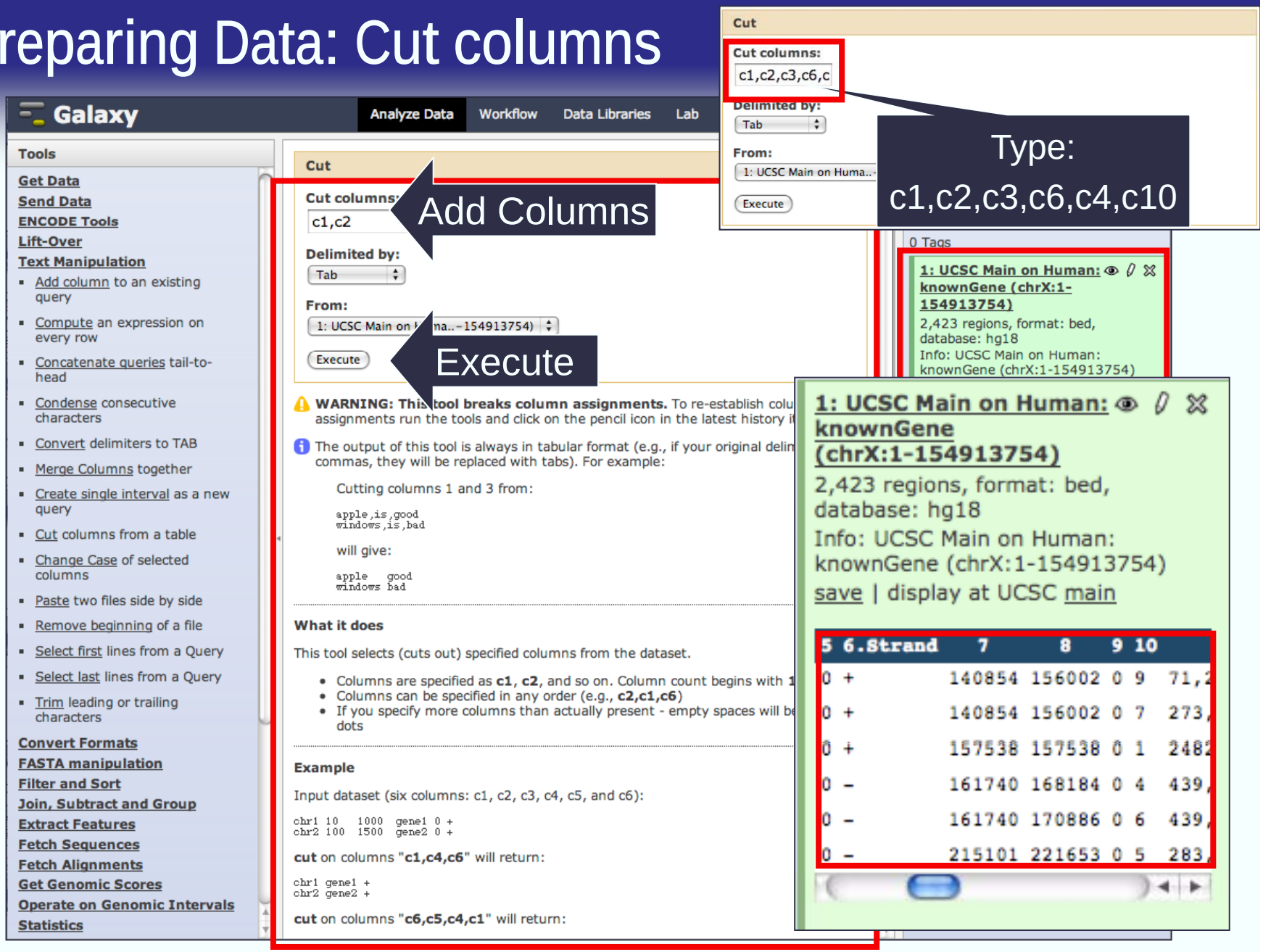
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Text Manipulation



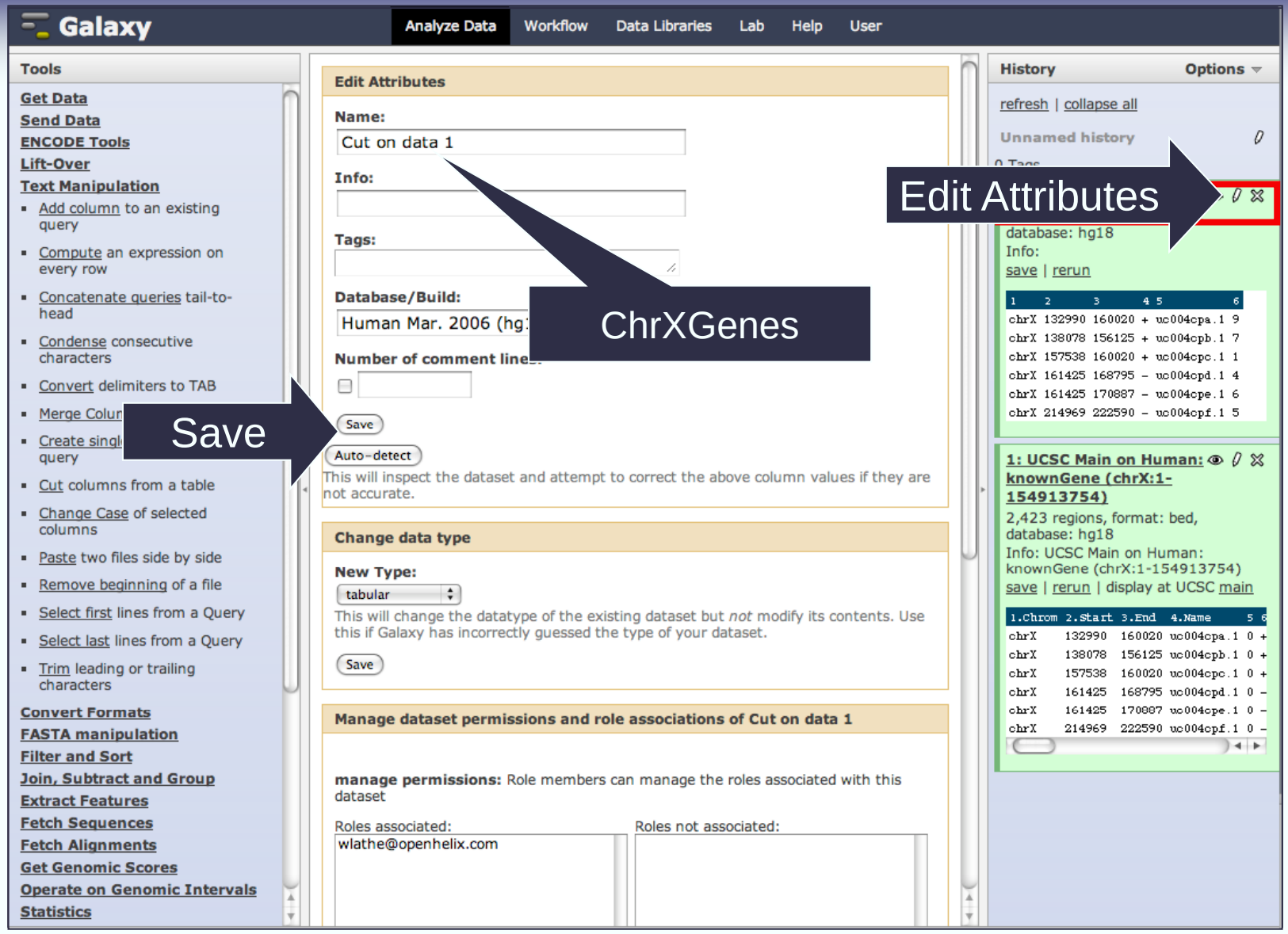
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Cut columns



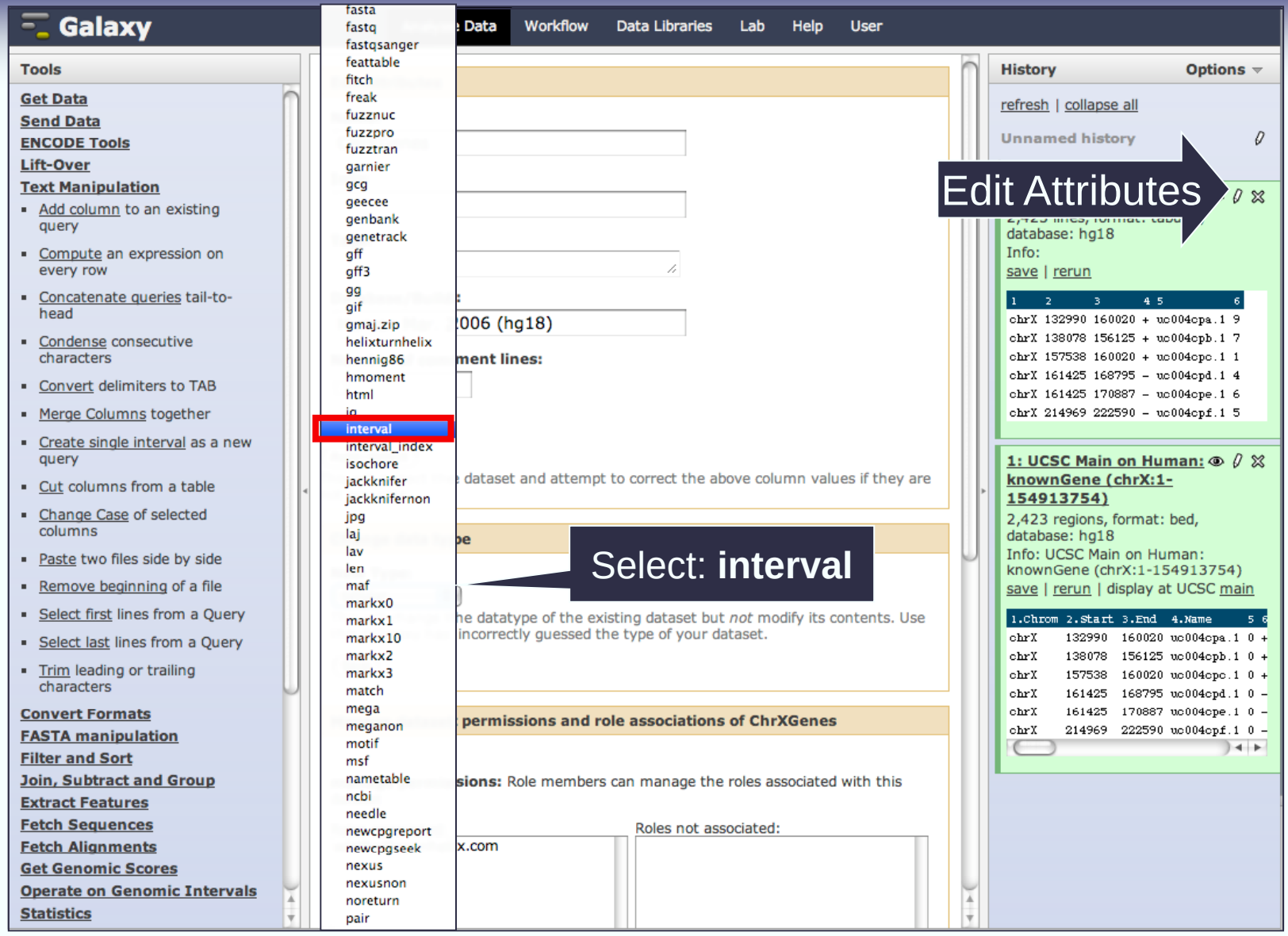
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Change Attributes



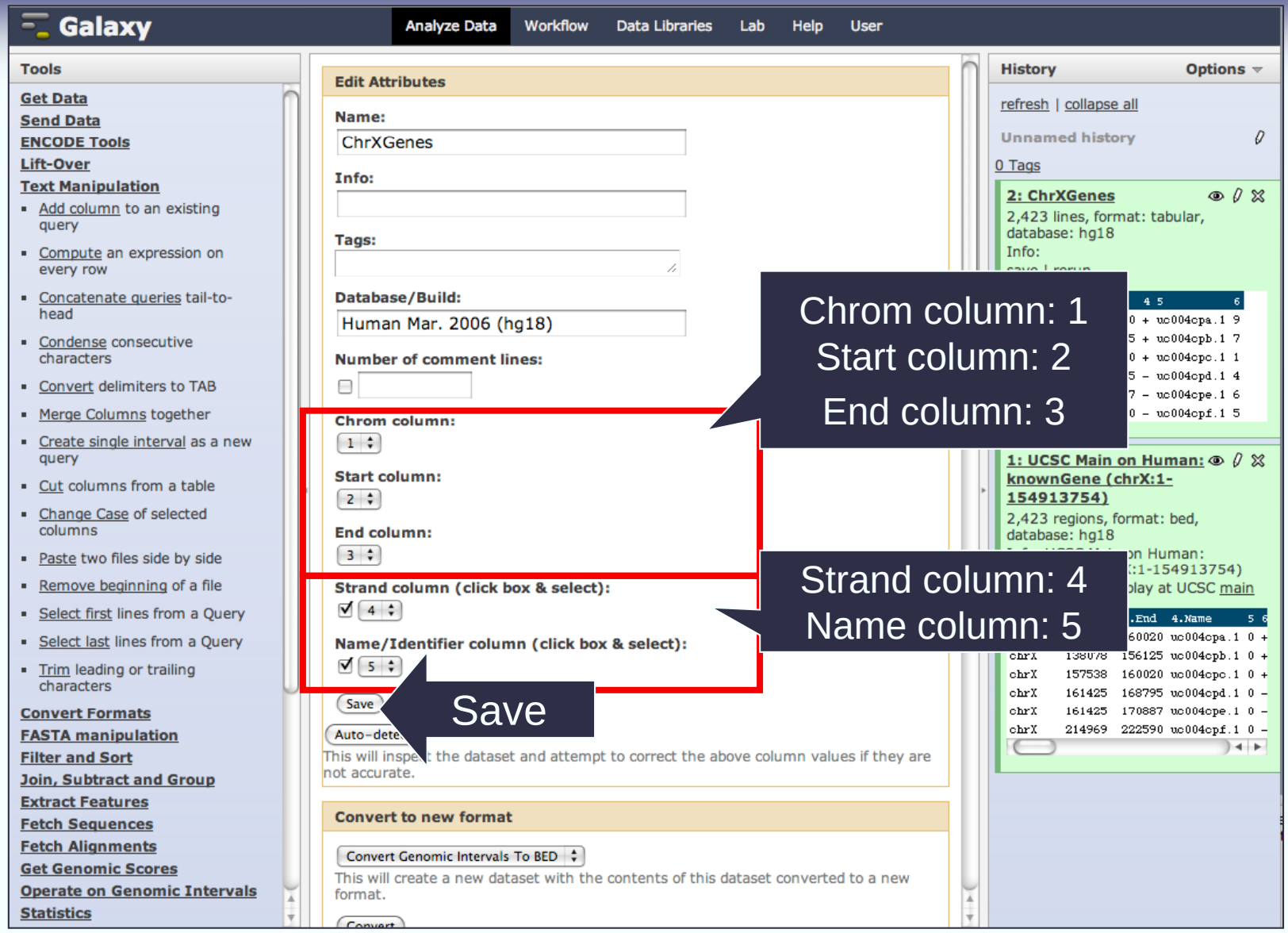
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Change Format



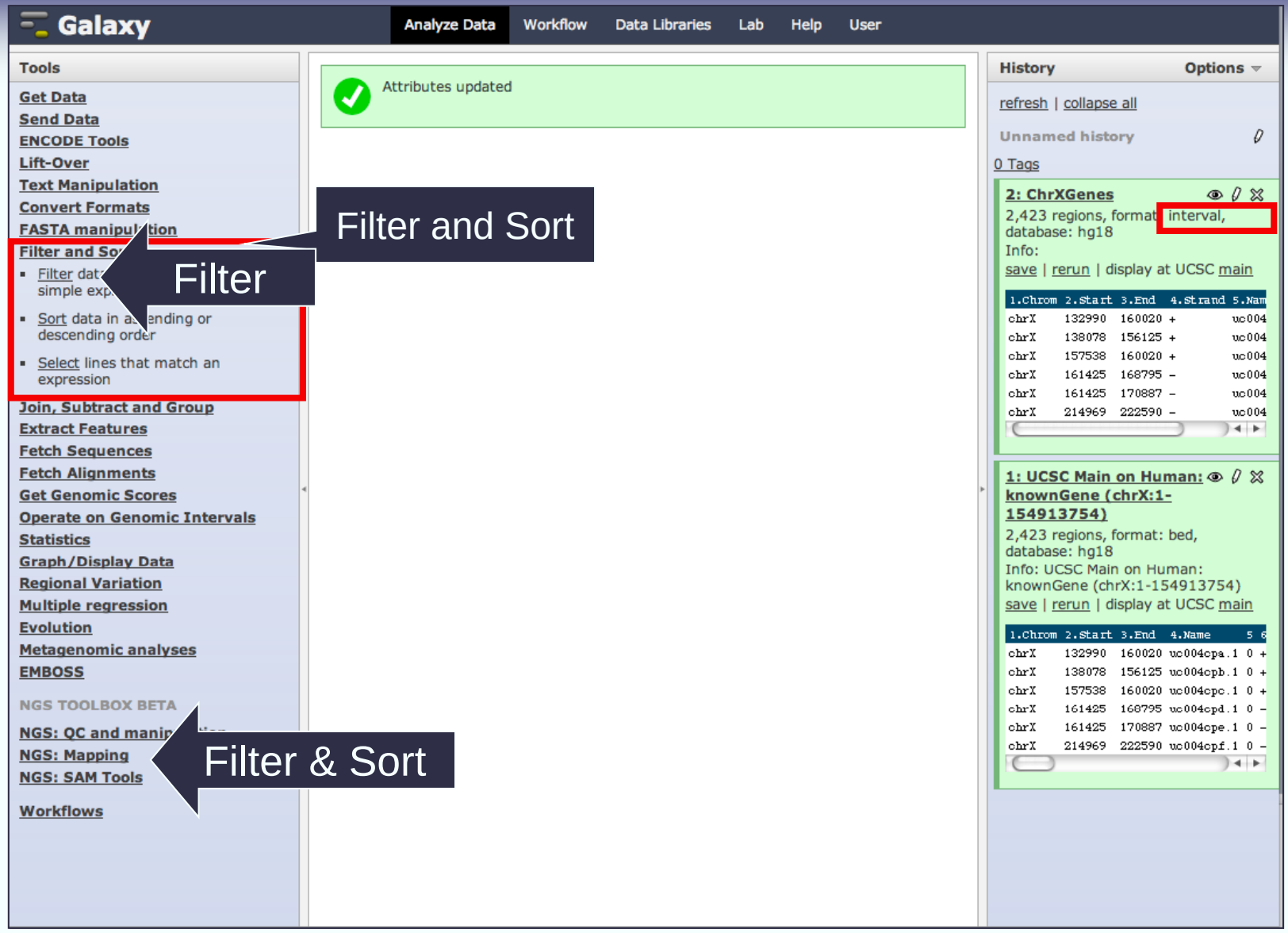
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Select Columns



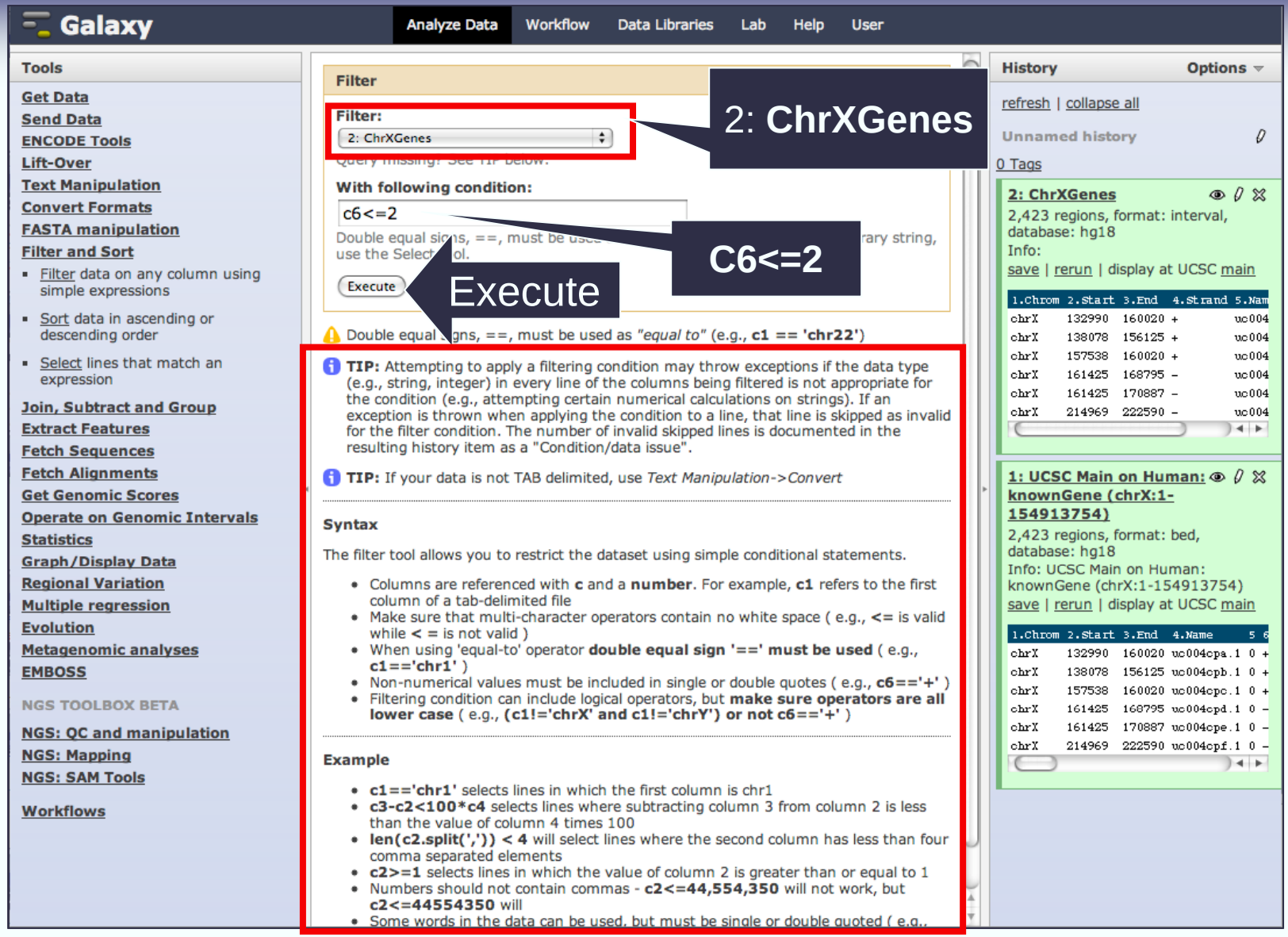
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Filtering and Sorting



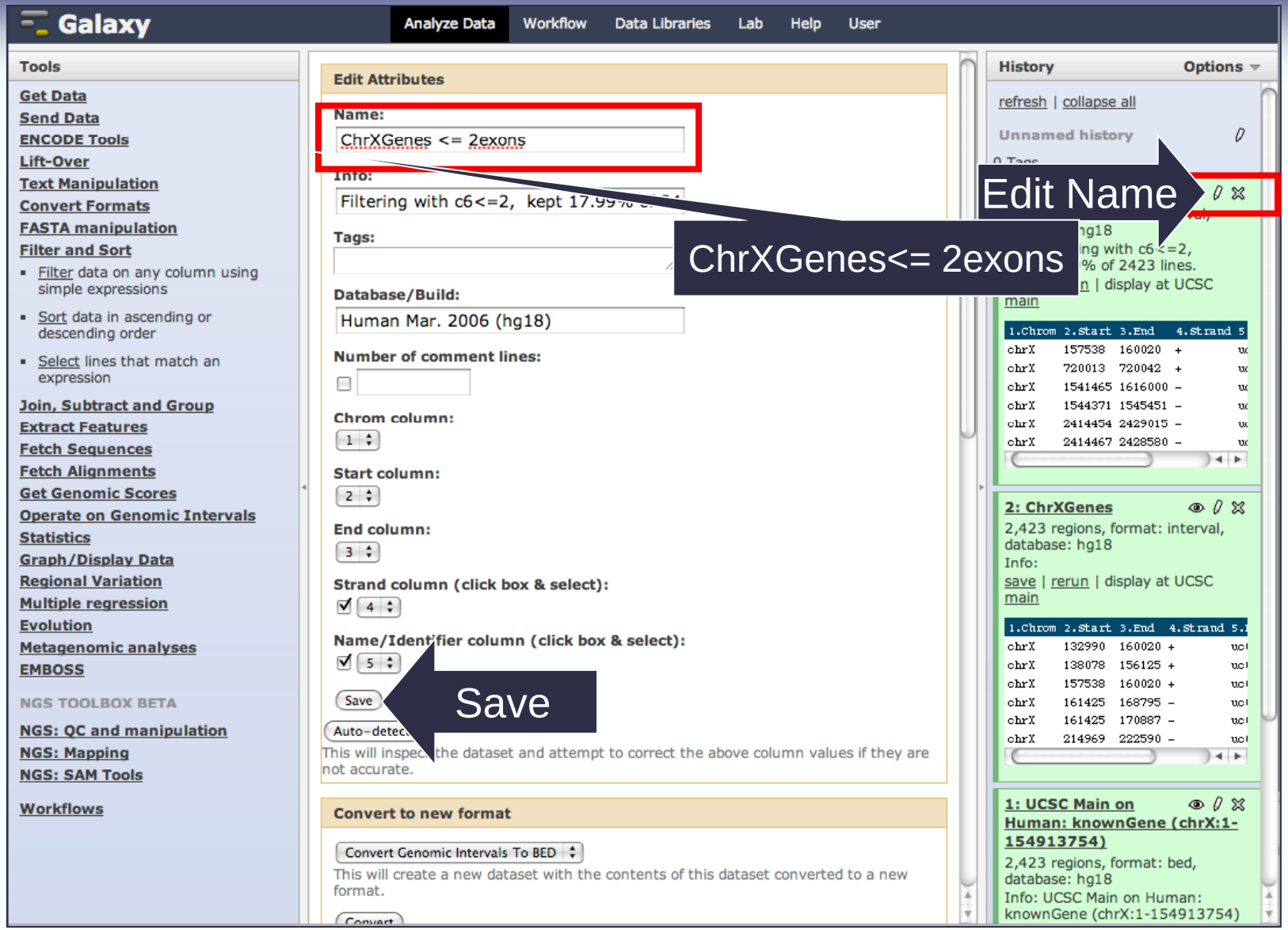
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Filter on Column



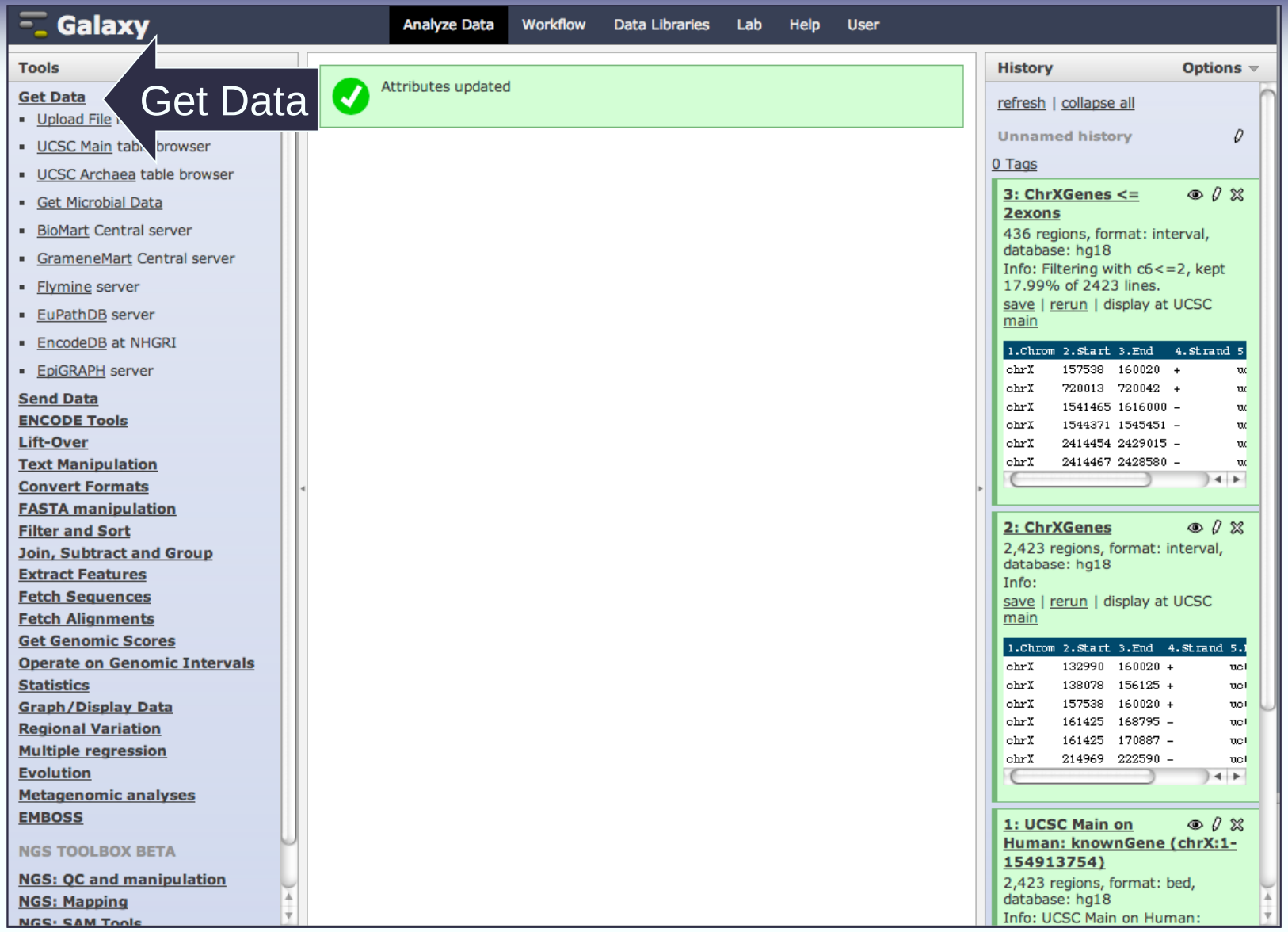
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Change Name



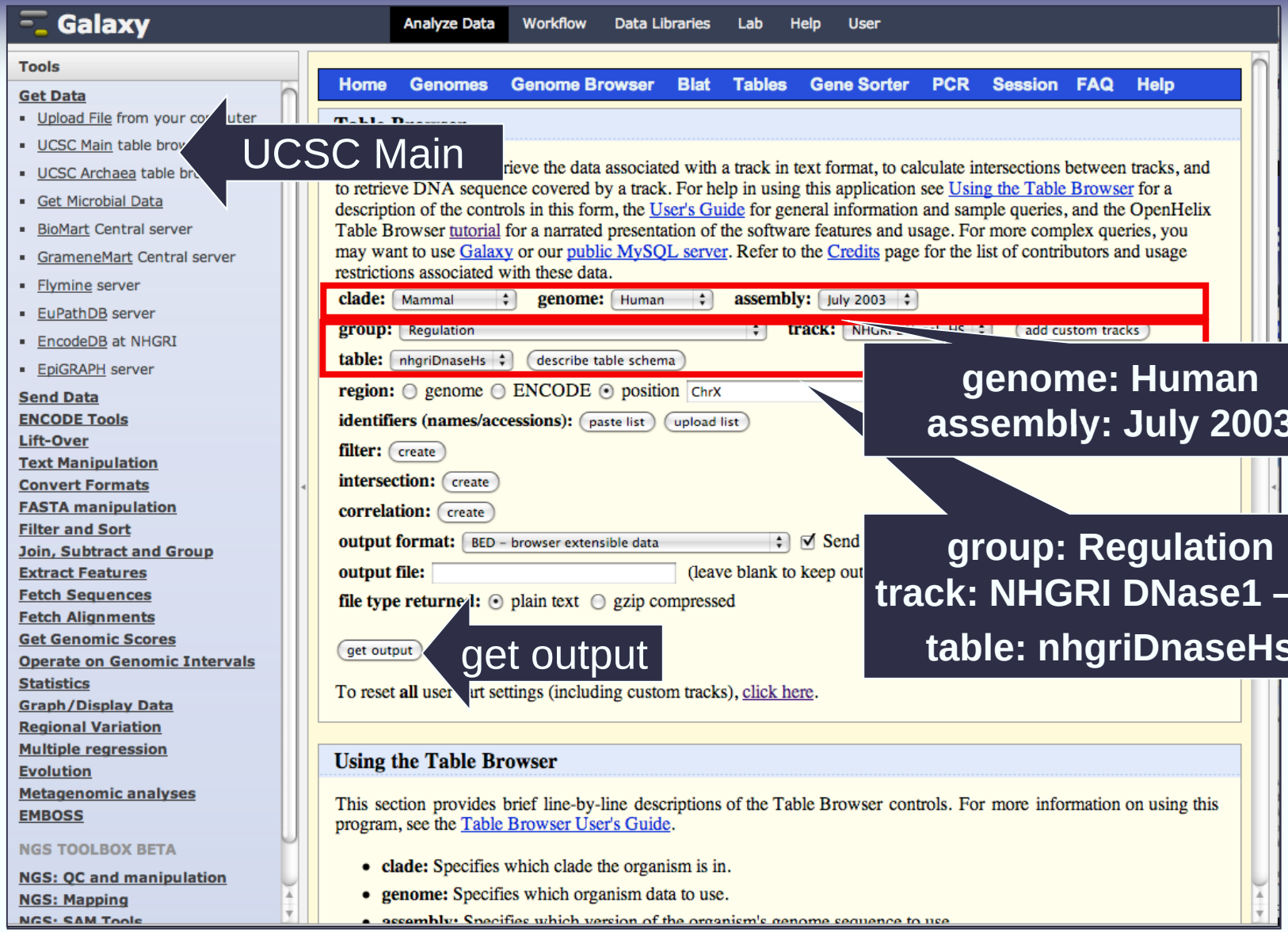
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Download second dataset



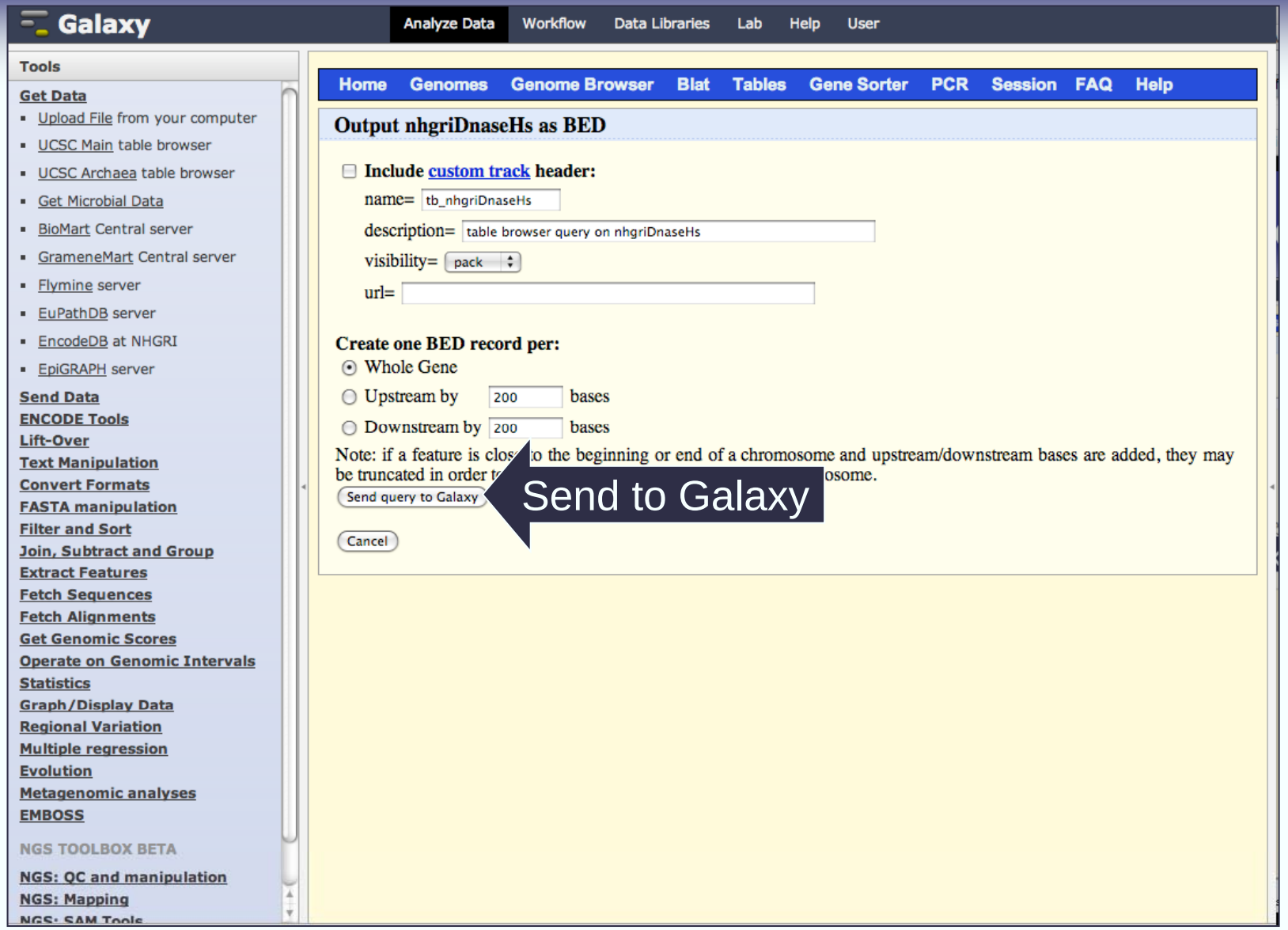
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Download second dataset



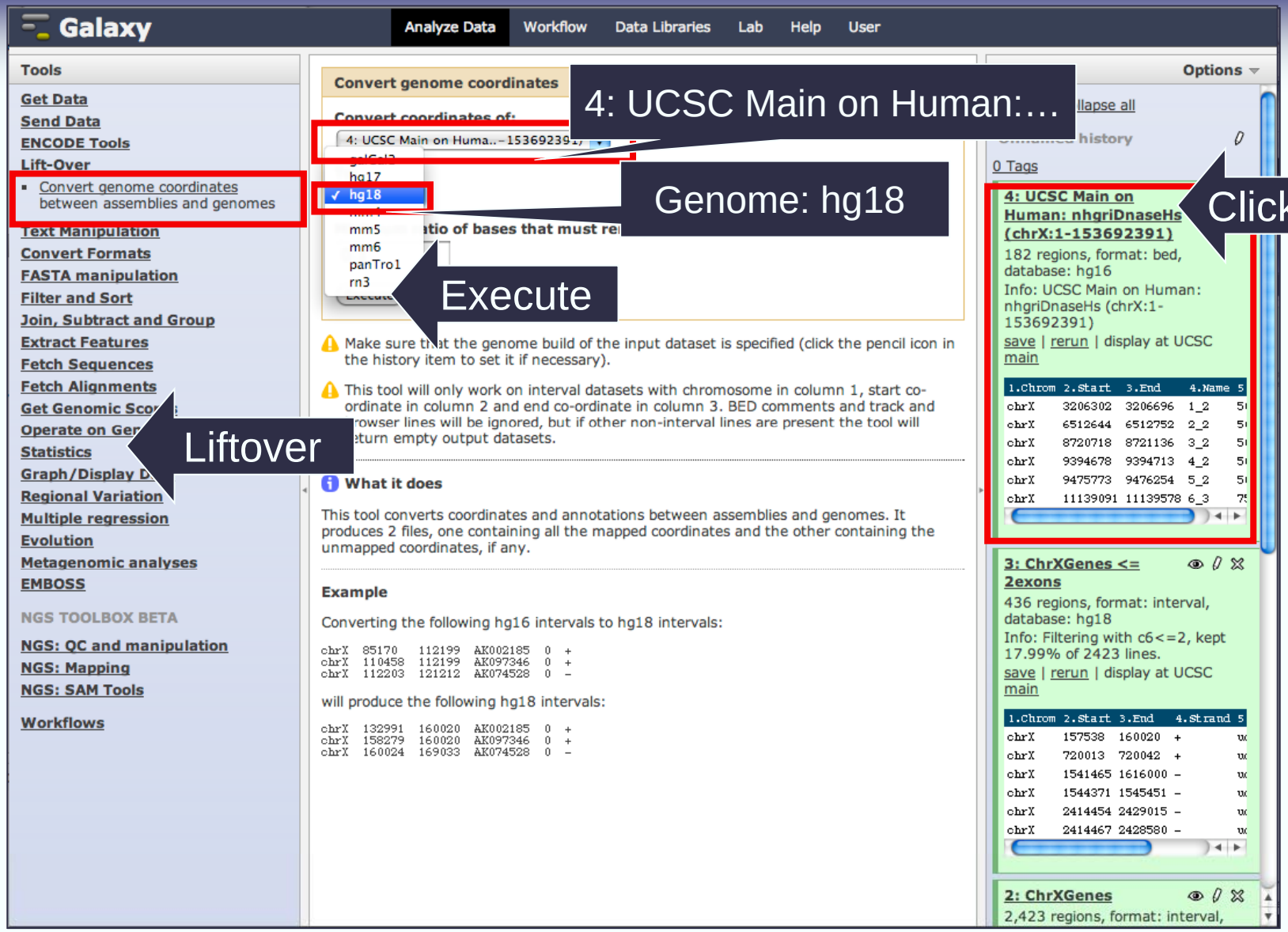
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Download second dataset



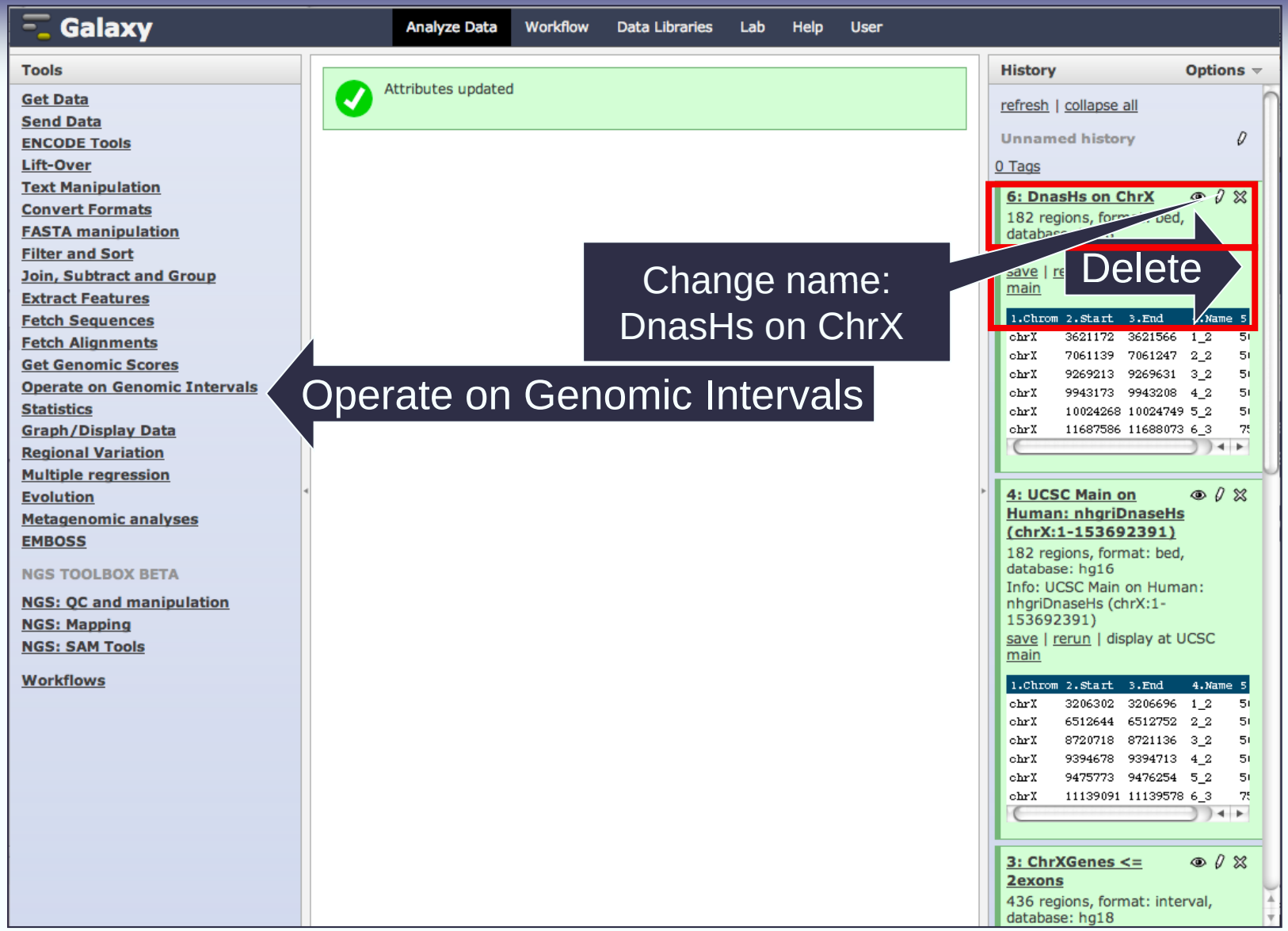
Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Liftover



Copyright OpenHelix. No use or reproduction without express written consent

Preparing Data: Change Name



Copyright OpenHelix. No use or reproduction without express written consent

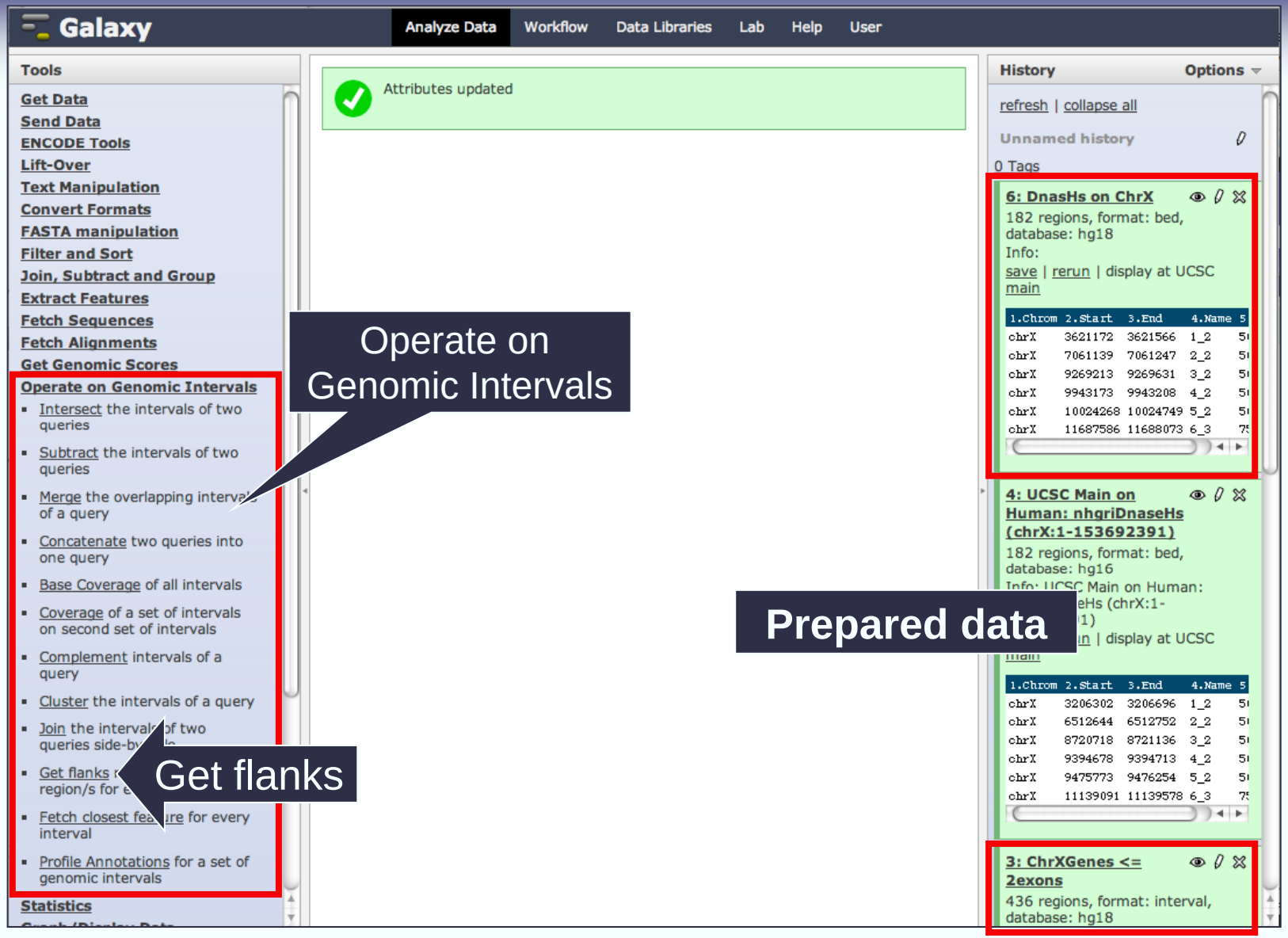
Galaxy Agenda

- Introduction and Credits
- Interface
- Getting Data
- Preparing Data
- Analyzing Data
- Histories & Workflows
- Summary
- Exercise

Galaxy: http://galaxy.psu.edu/

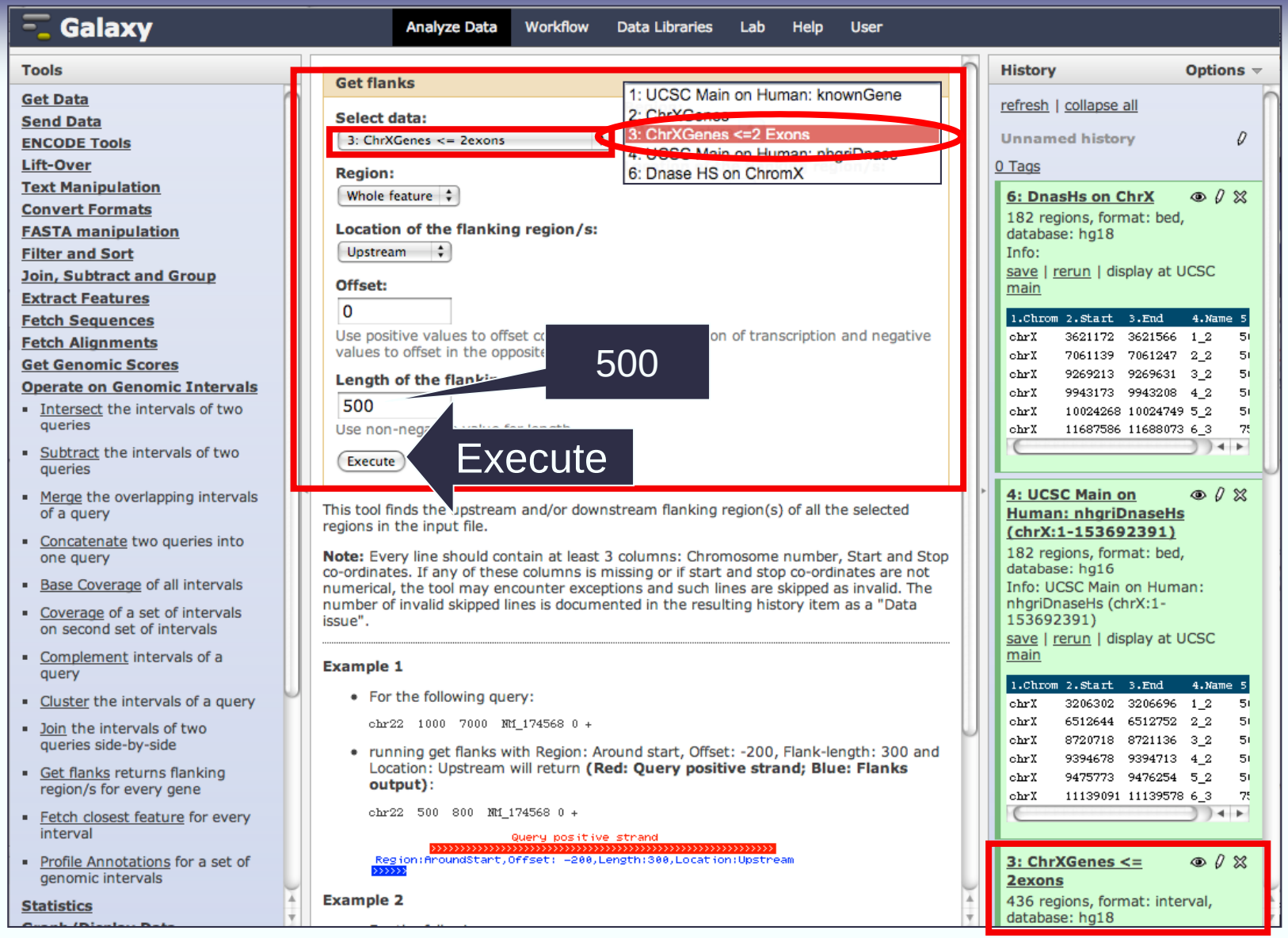
Copyright OpenHelix. No use or reproduction without express written consent

Analyzing Data: Getting Flanking regions



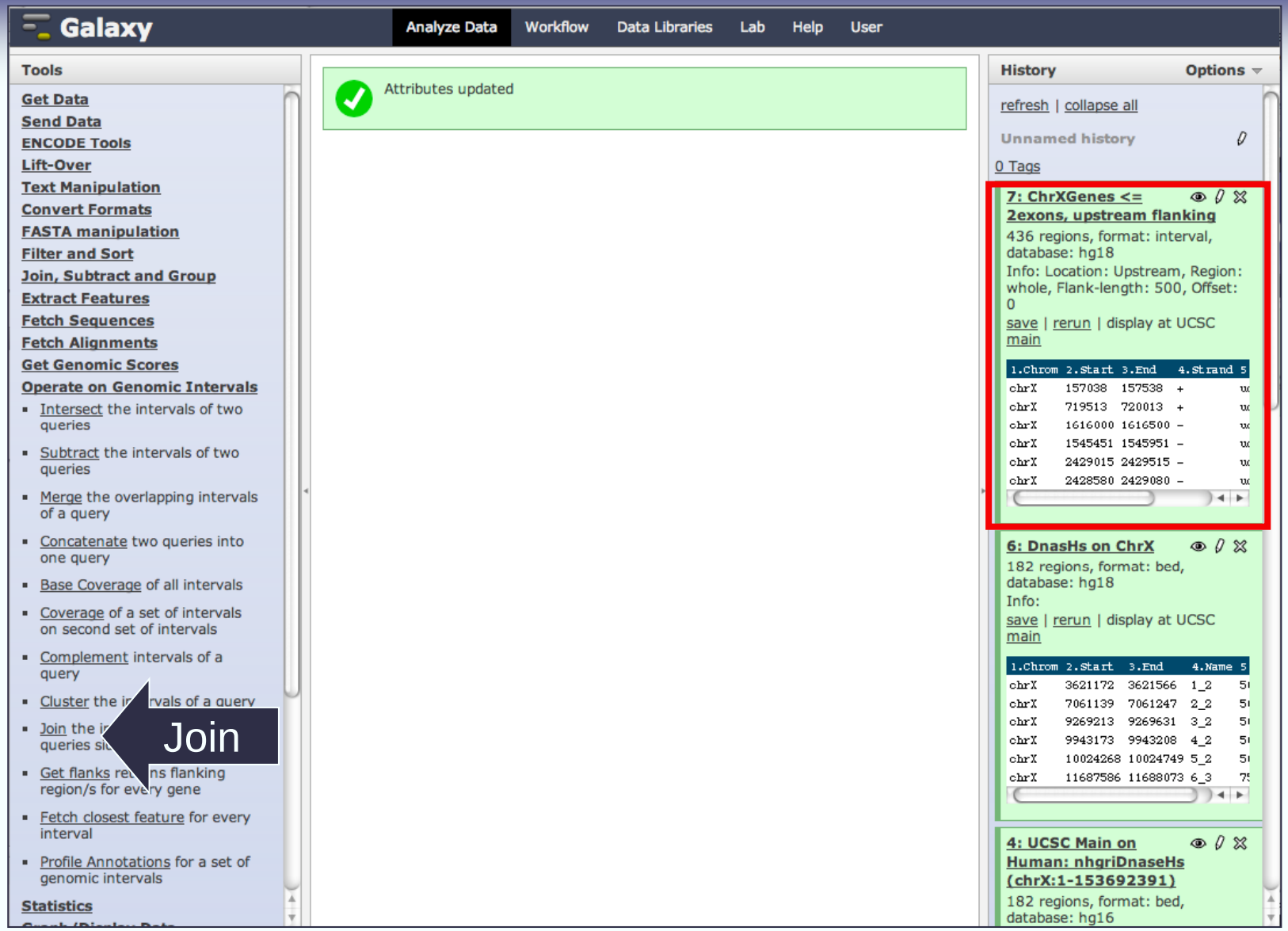
Copyright OpenHelix. No use or reproduction without express written consent

Analyzing Data: Flanking regions



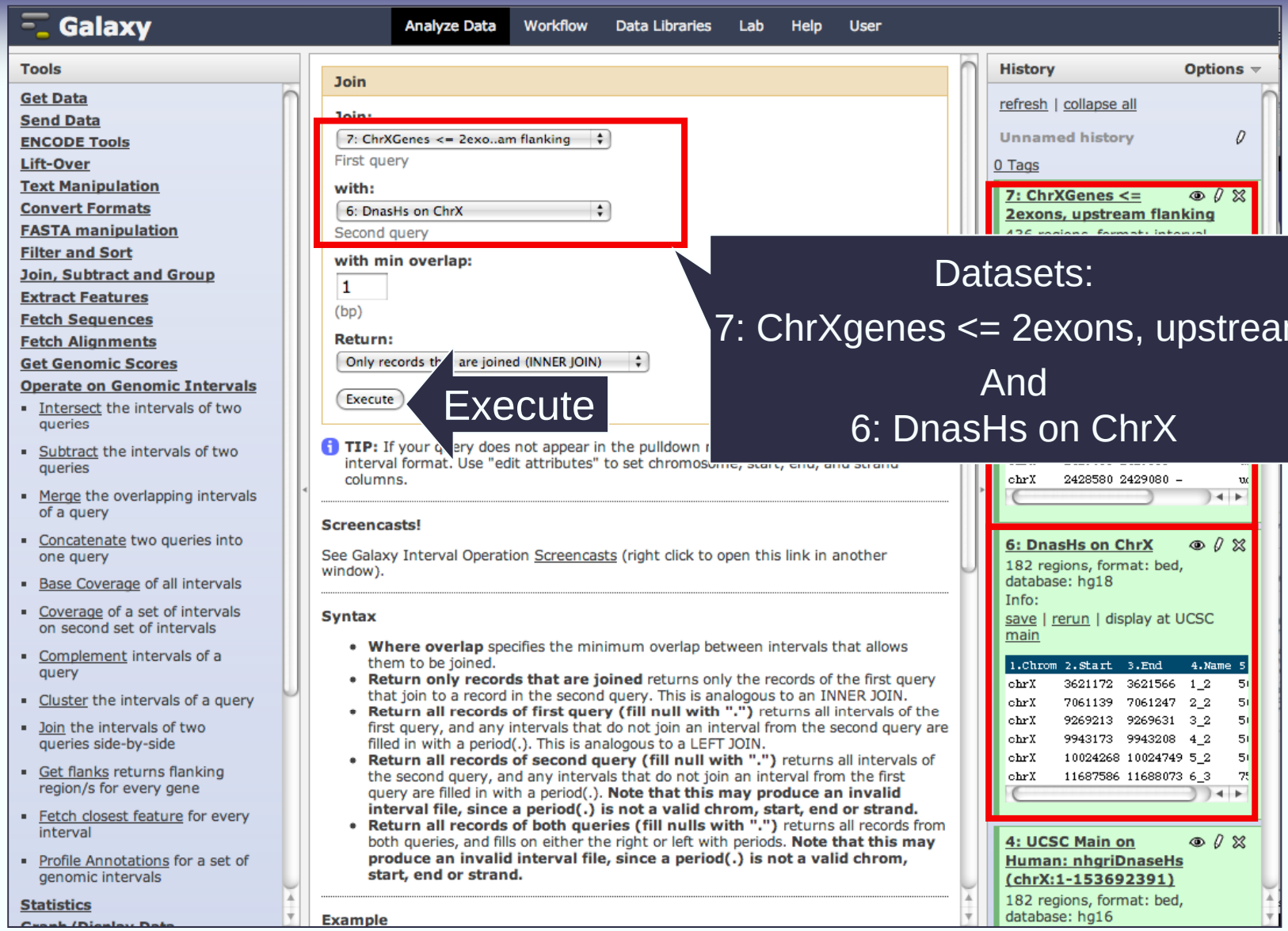
Copyright OpenHelix. No use or reproduction without express written consent

Analyzing Data: Joining Two Datasets



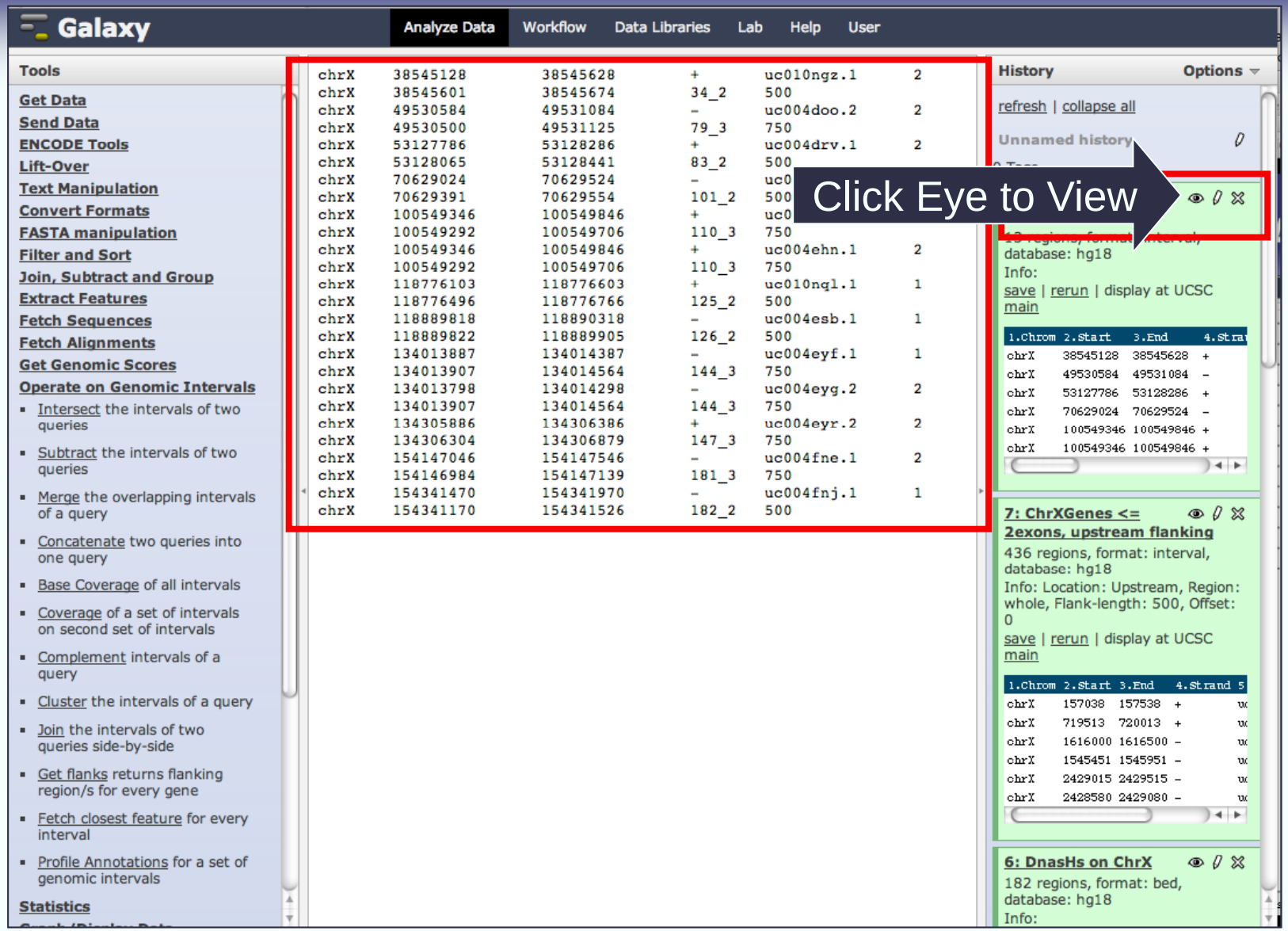
Copyright OpenHelix. No use or reproduction without express written consent

Analyzing Data: Joining Two Datasets



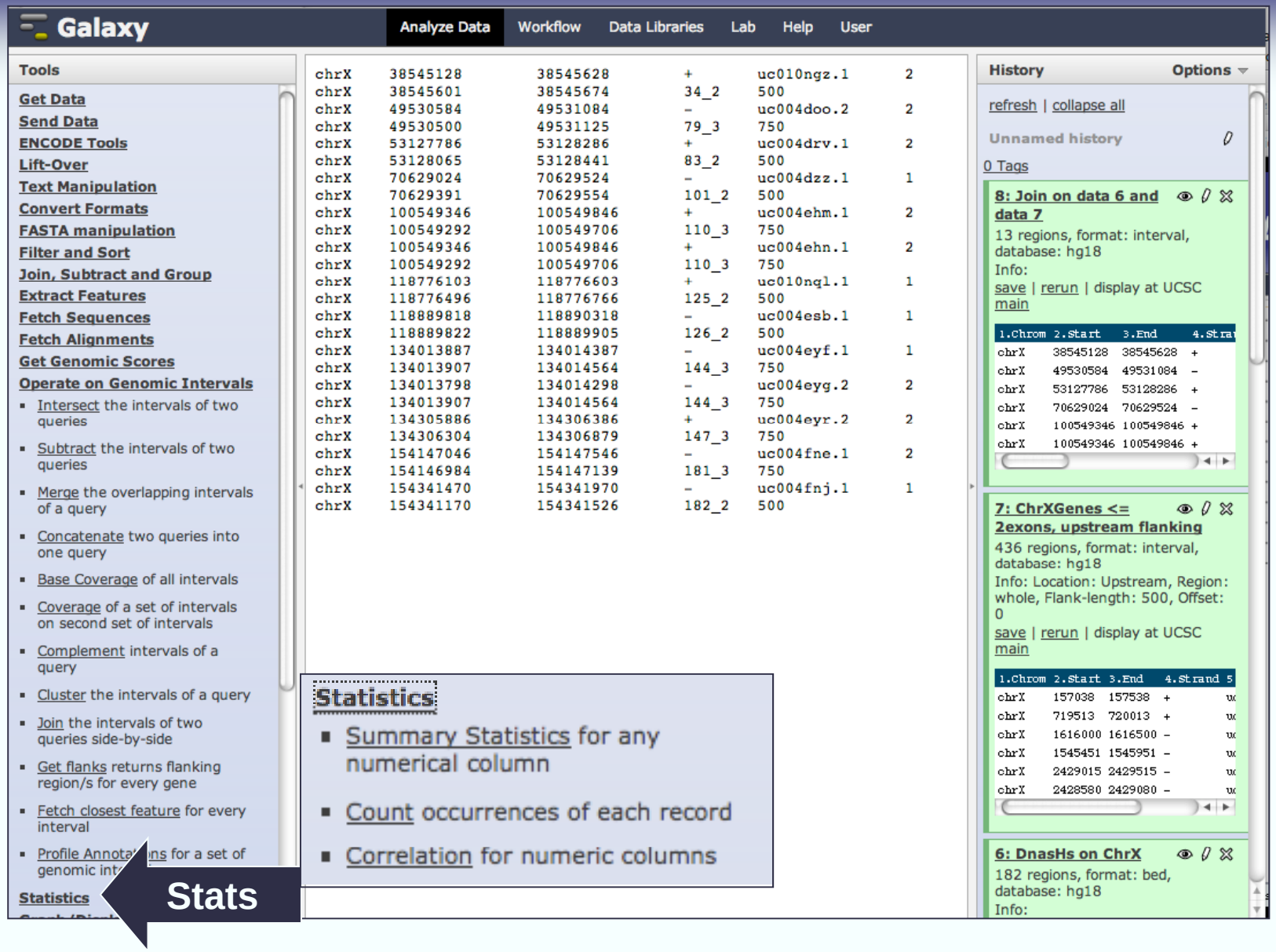
Copyright OpenHelix. No use or reproduction without express written consent

Analyzing Data: Viewing Data Analysis



Copyright OpenHelix. No use or reproduction without express written consent

Analyzing Data: Many Analysis Tools



Copyright OpenHelix. No use or reproduction without express written consent

Analyzing Data: Next Generation Sequencing

Galaxy | Analyze Data | Workflow | Data Libraries | Lab | Help | User

Tools

- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Regional Variation
- Multiple regression
- Evolution
- Metagenomic analyses
- EMBOSS

NGS TOOLBOX BETA

- NGS: QC and manipulation
  - GENERIC FASTQ DATA
    - FASTQ Groomer converts any FASTQ to Sanger
    - Quality format converter (ASCII-Numeric)
    - Compute quality statistics
    - Draw quality score boxplot
    - Draw nucleotides distribution chart
    - Split paired end reads
  - ROCHE-454 DATA
    - Build base quality distribution
    - Select high quality segments
  - AB-SOLID DATA
    - Convert SOLiD output to fastq
    - Compute quality statistics for SOLiD data
    - Draw quality score boxplot for SOLiD data
- NGS: Mapping
- NGS: SAM Tools

Workflows

Draw

Draw nucleotides distribution chart

Statistics Text File:

output of 'FASTX Statistics' tool

Execute

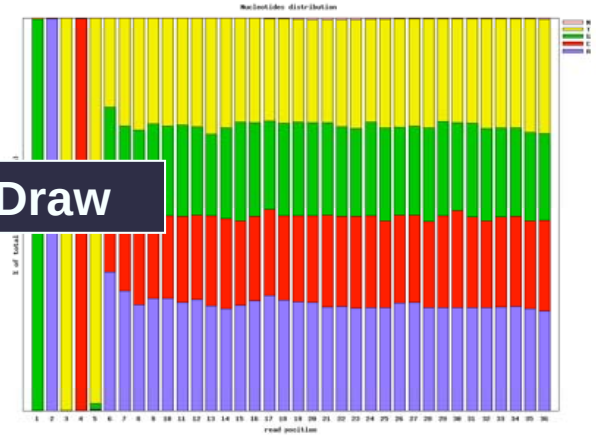
What it does

Creates a stacked-histogram graph for the nucleotide distribution in the Solexa library.

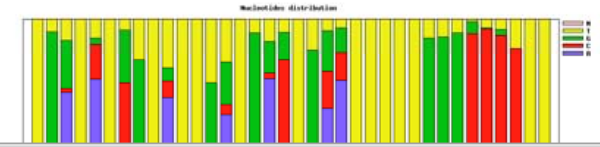
TIP: Use the **FASTQ Statistics** tool to generate the report file needed for this tool.

Output Examples

The following chart clearly shows the barcode used at the 5'-end of the library: **GATCT**



In the following chart, one can almost 'read' the most abundant sequence by looking at the dominant values: **TGATA TCGTA TTGAT GACTG AA...**



History Options

refresh | collapse all

Unnamed history

0 Tags

Your history is empty. Click 'Get Data' on the left pane to start

Copyright OpenHelix. No use or reproduction without express written consent

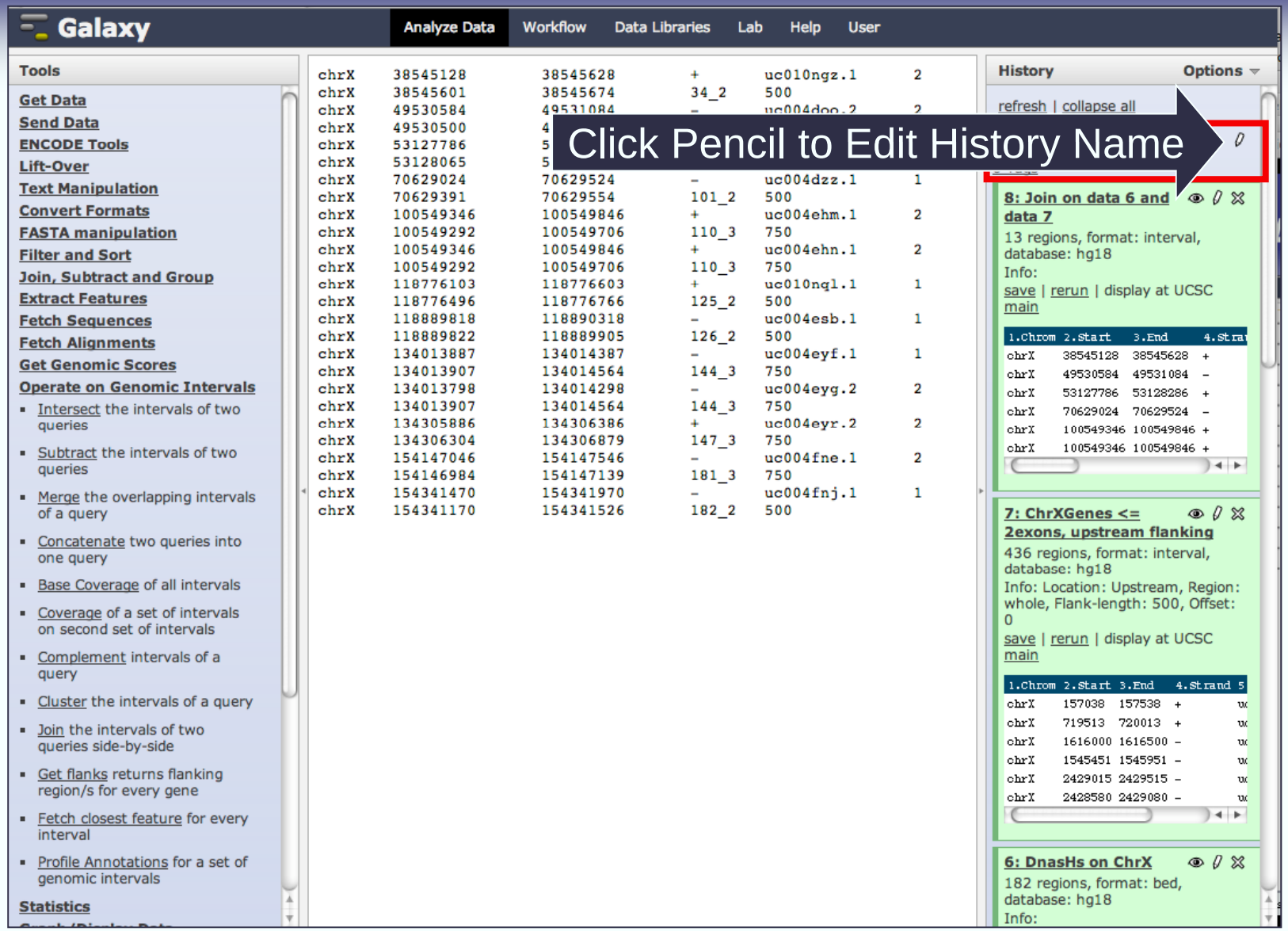
Galaxy Agenda

- Introduction and Credits
- Interface
- Getting Data
- Preparing Data
- Analyzing Data
- Histories & Workflows
- Summary
- Exercise

Galaxy: http://galaxy.psu.edu/

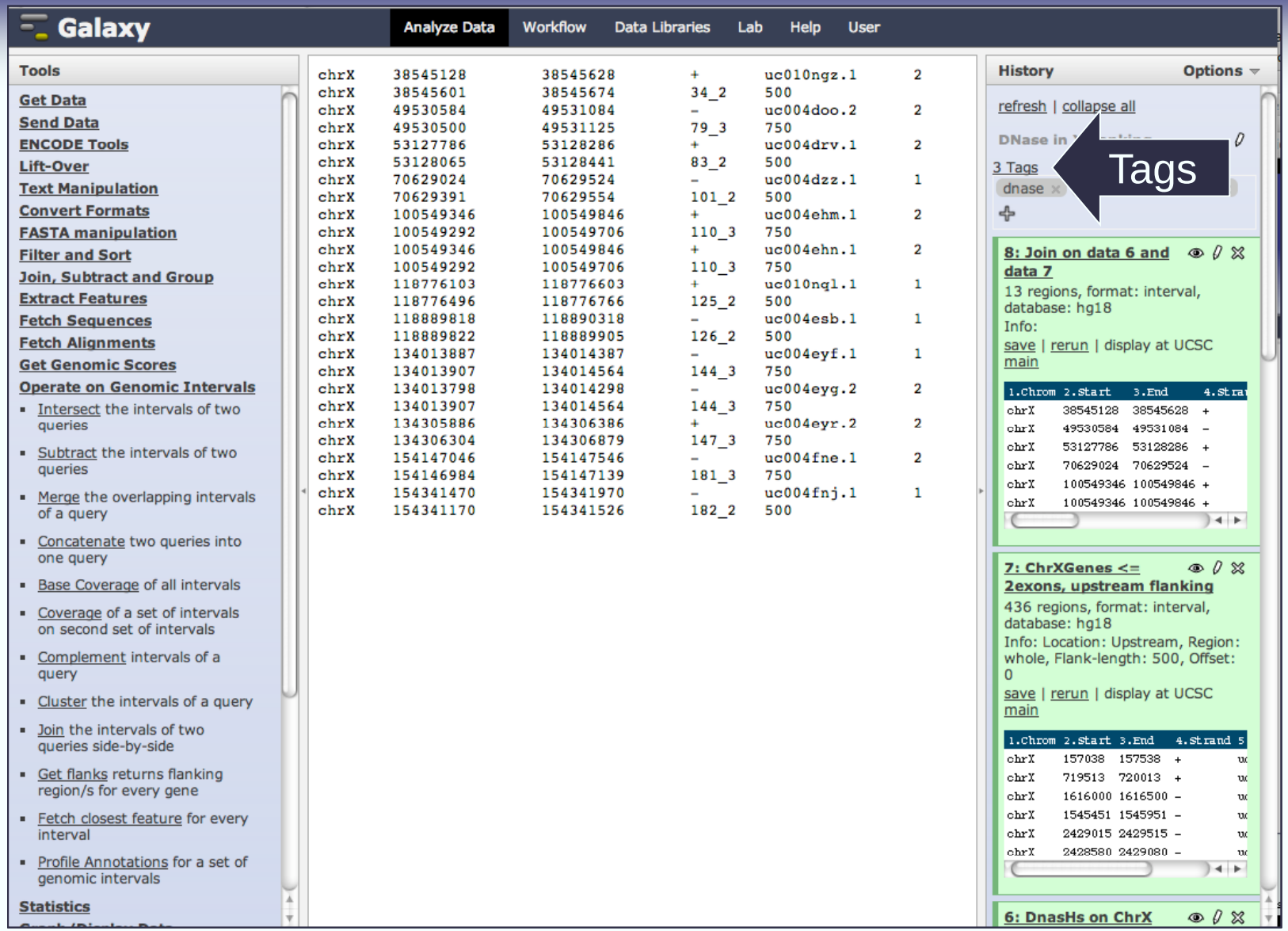
Copyright OpenHelix. No use or reproduction without express written consent

History: Name History



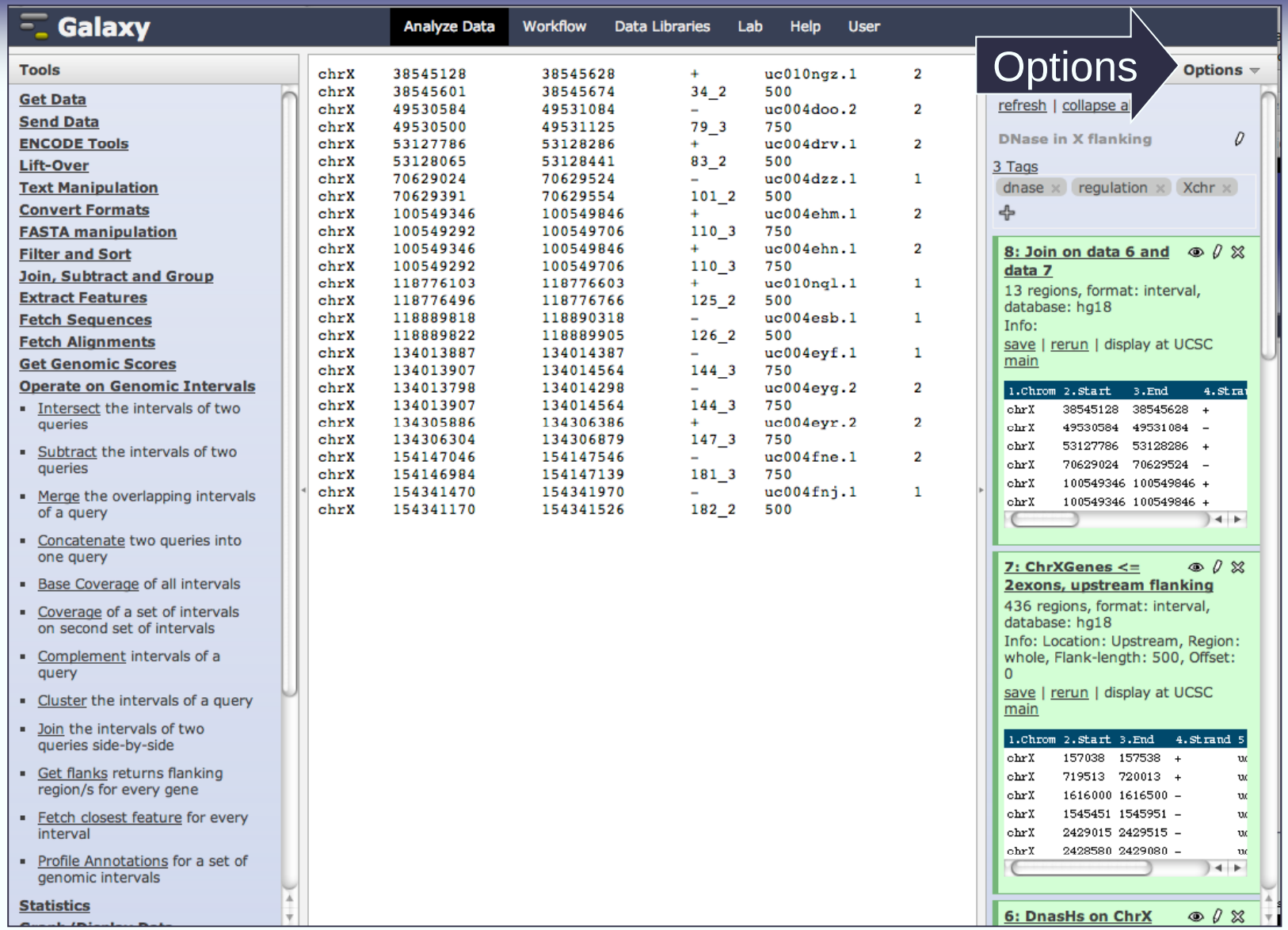
Copyright OpenHelix. No use or reproduction without express written consent

History: Name History & Add Tags



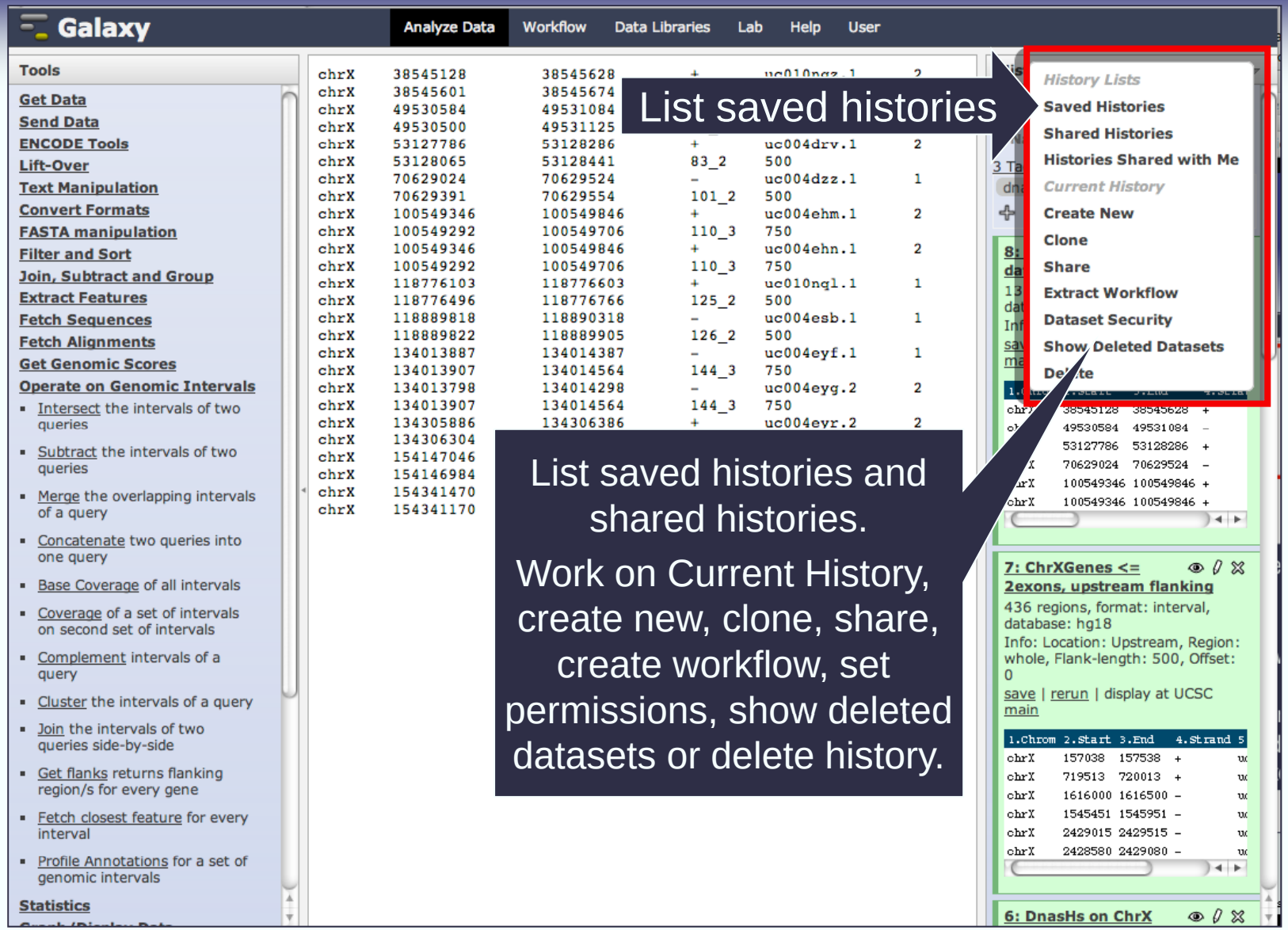
Copyright OpenHelix. No use or reproduction without express written consent

History: Name History



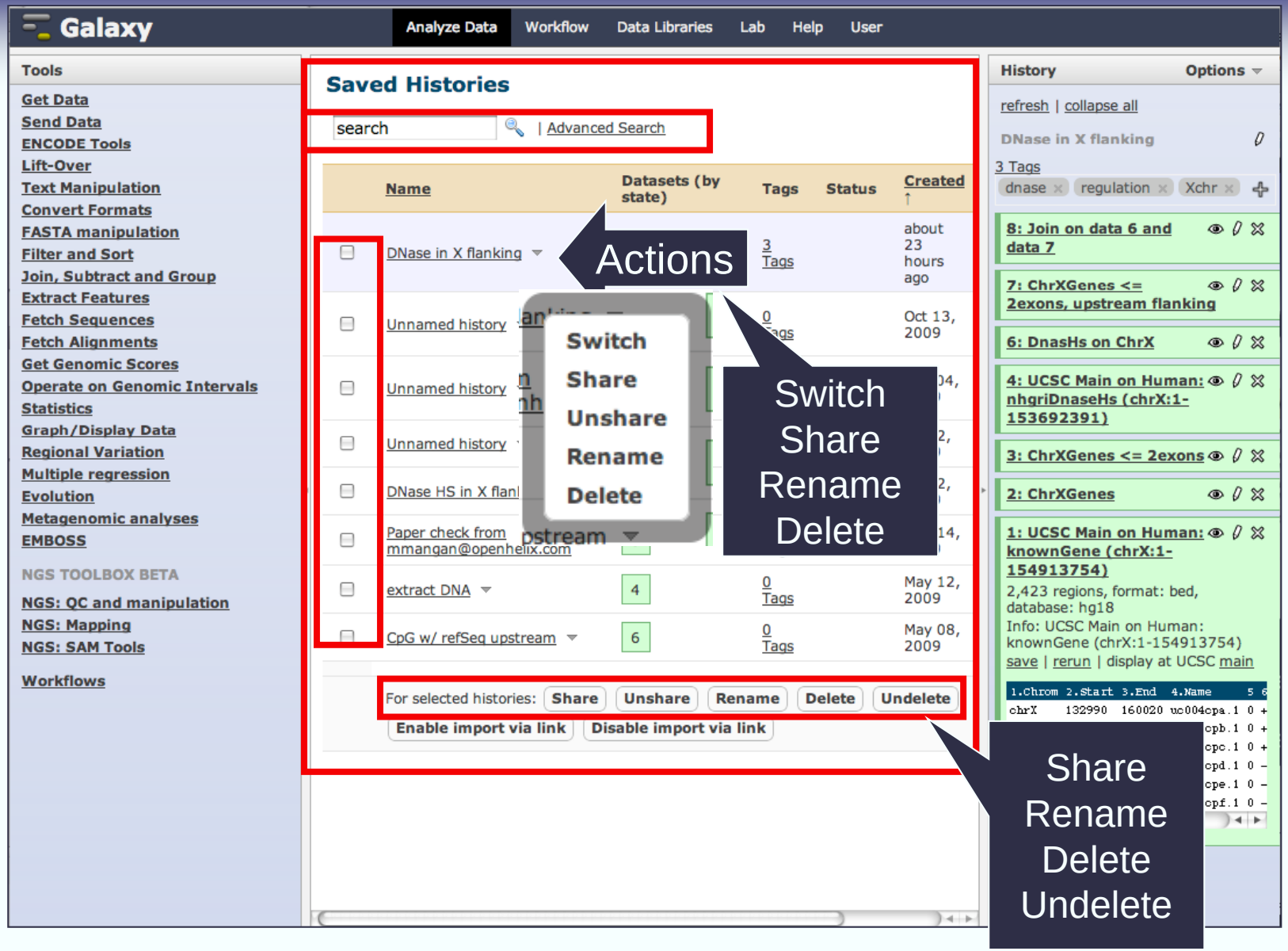
Copyright OpenHelix. No use or reproduction without express written consent

History: History Options



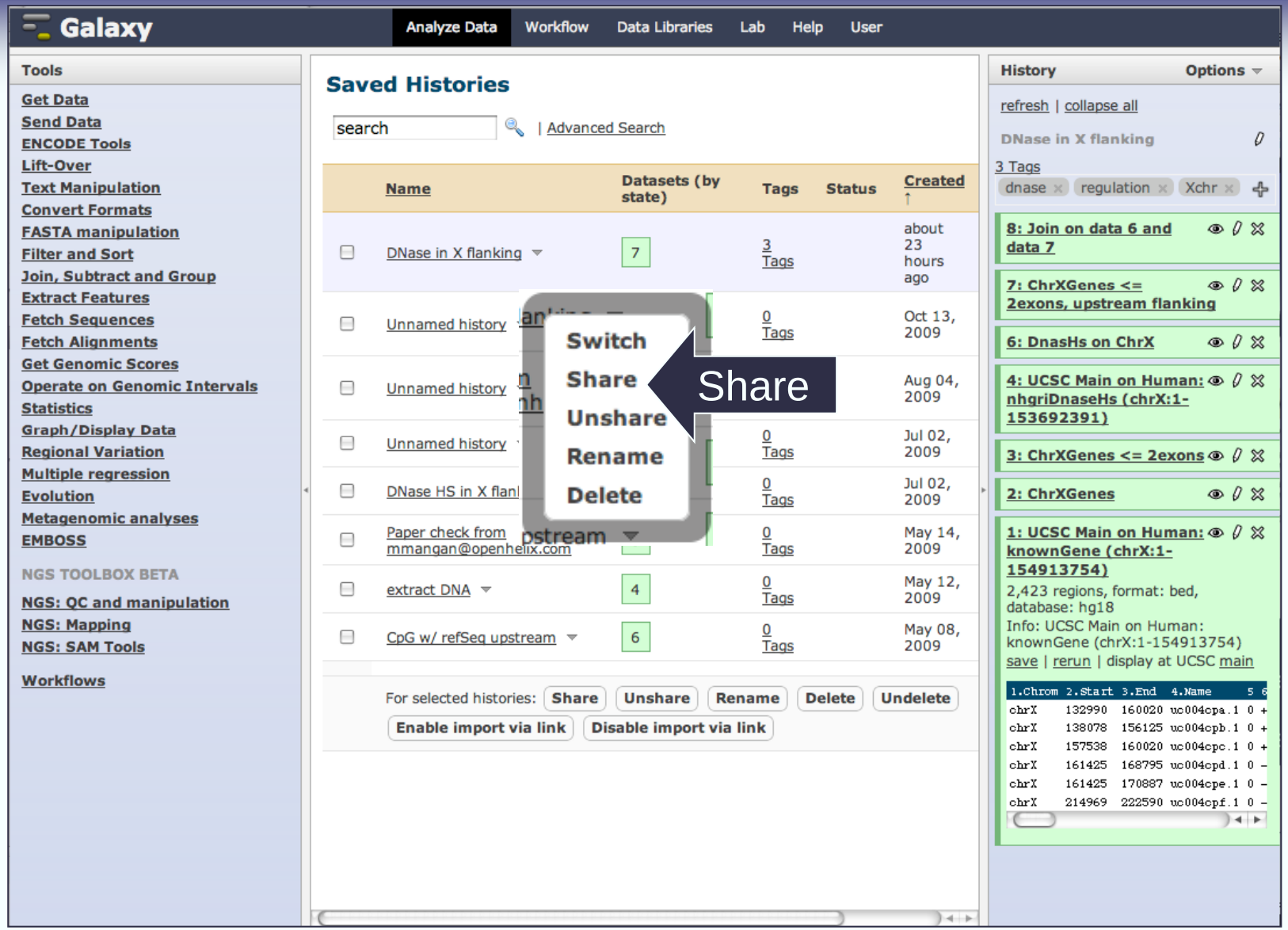
Copyright OpenHelix. No use or reproduction without express written consent

History: Saved Histories



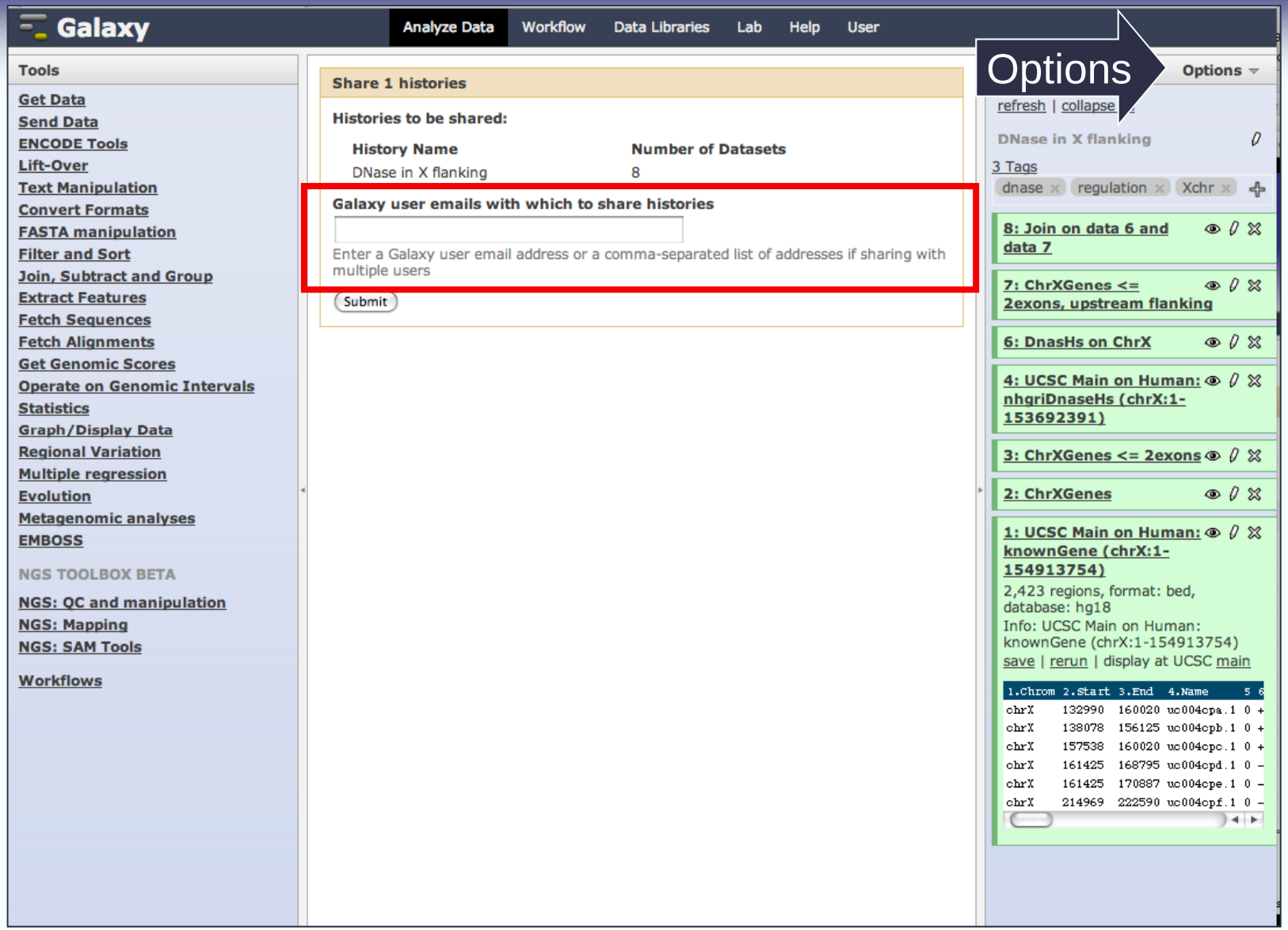
Copyright OpenHelix. No use or reproduction without express written consent

History: Share, Rename, Delete History



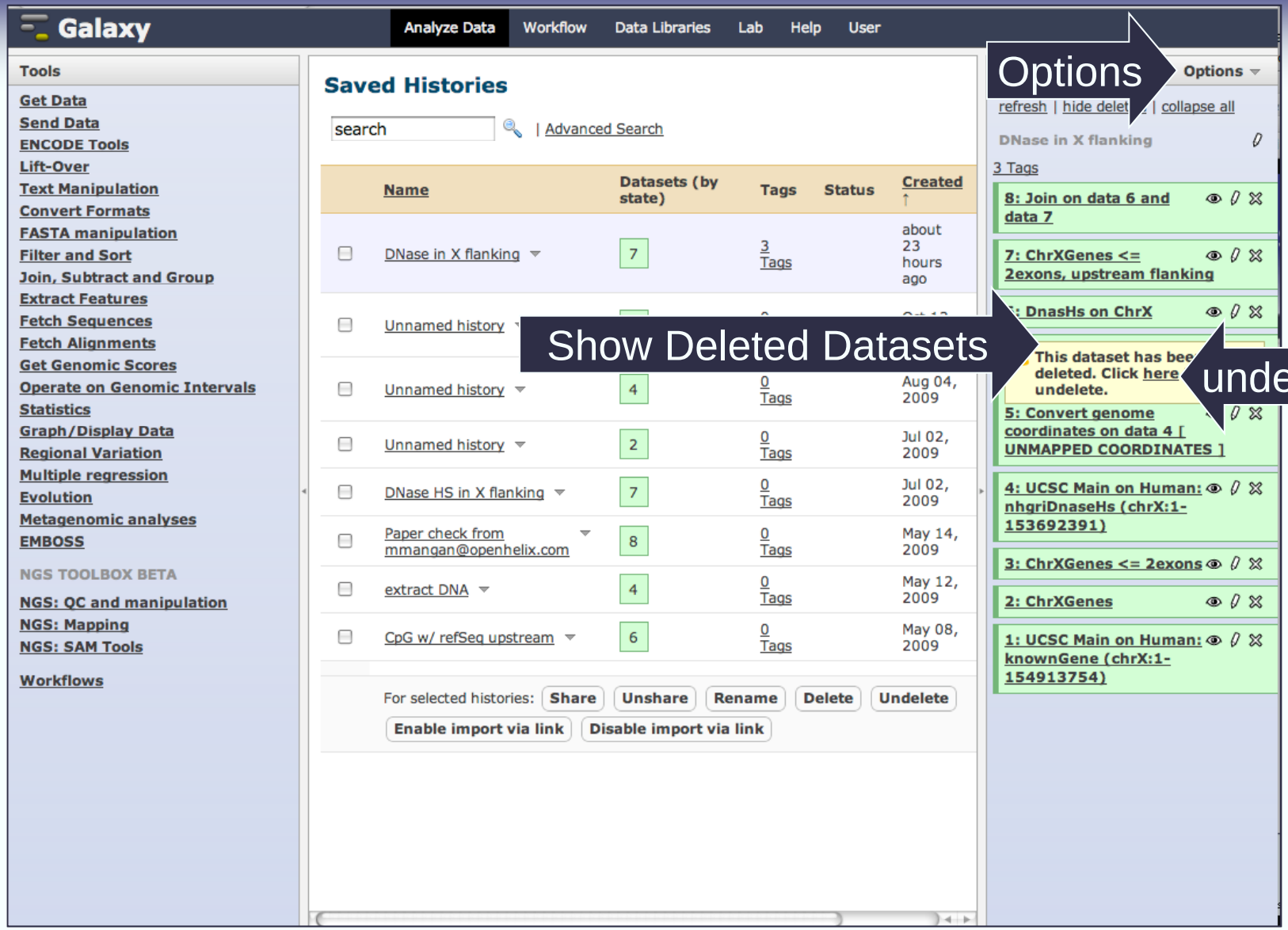
Copyright OpenHelix. No use or reproduction without express written consent

History: Share History



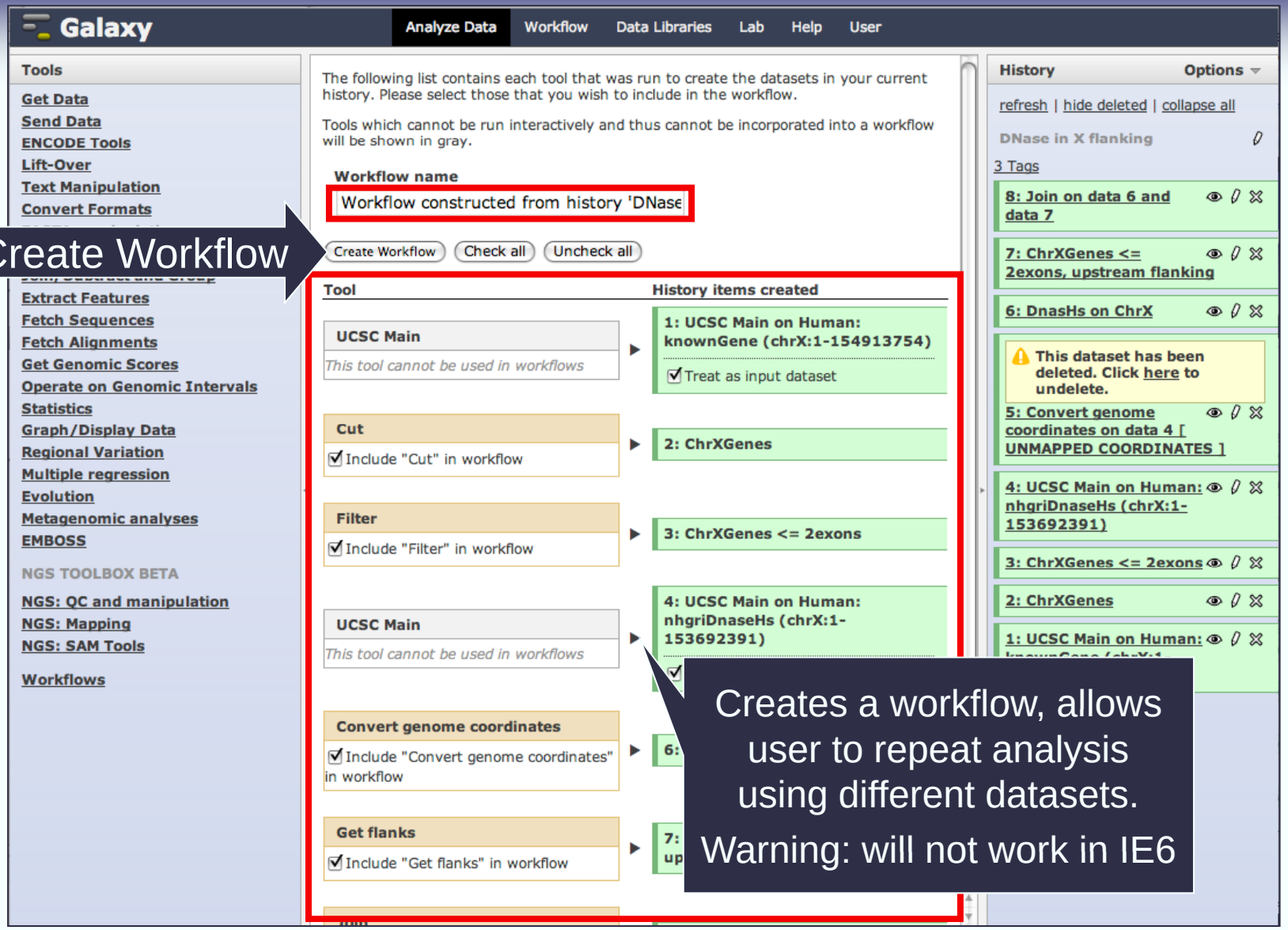
Copyright OpenHelix. No use or reproduction without express written consent

History: View Deleted Datasets



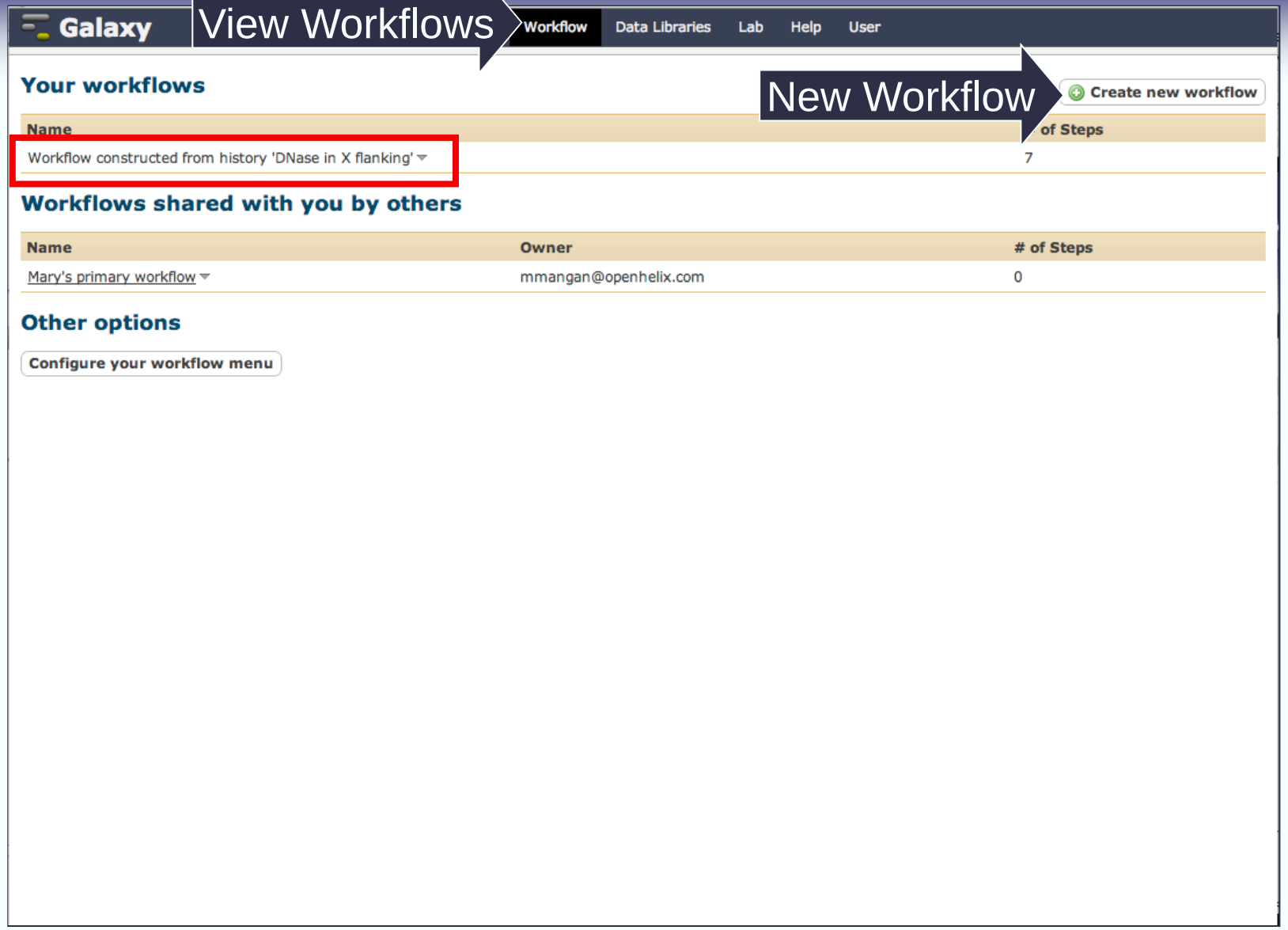
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Create Workflow from History



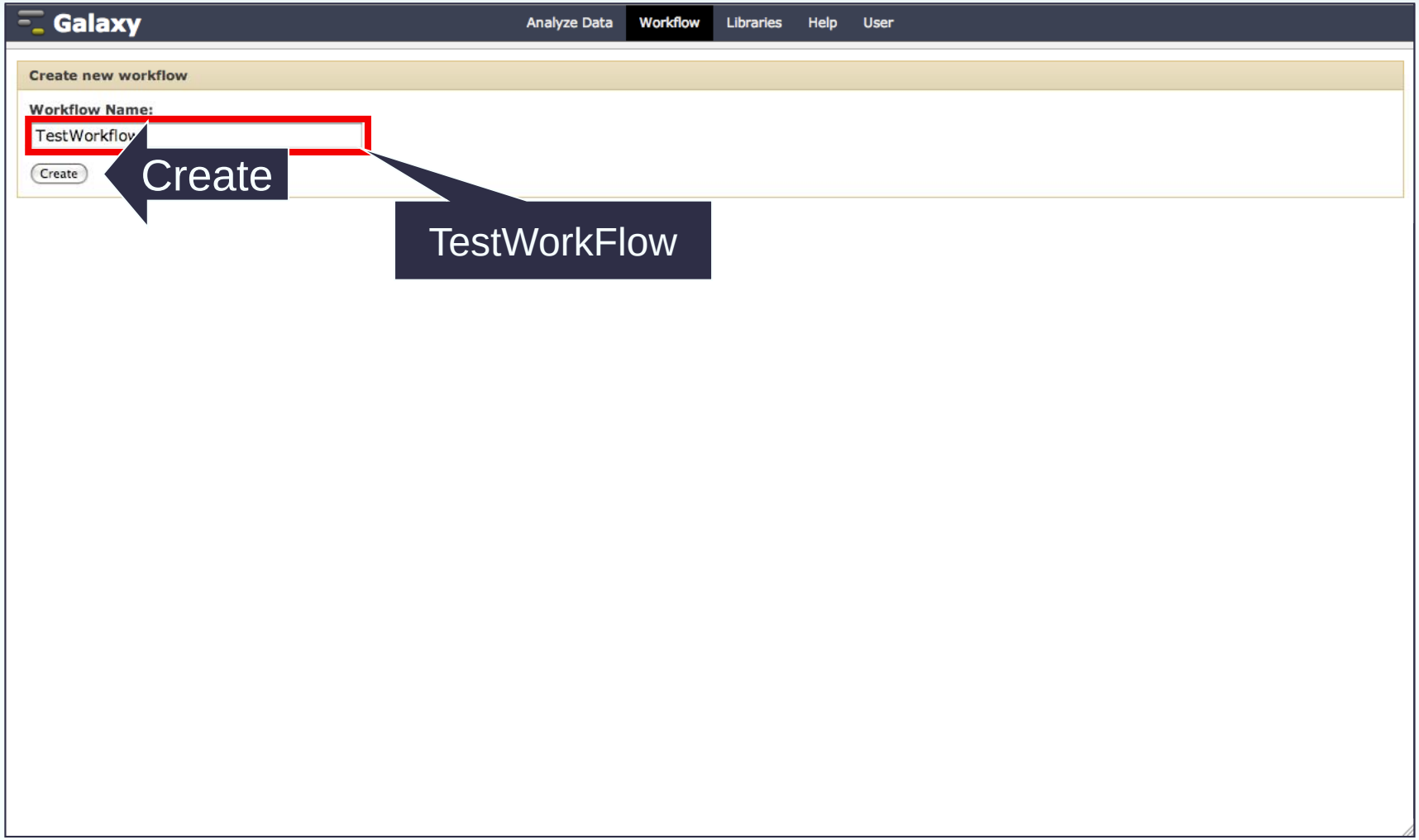
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: View and Run Workflows



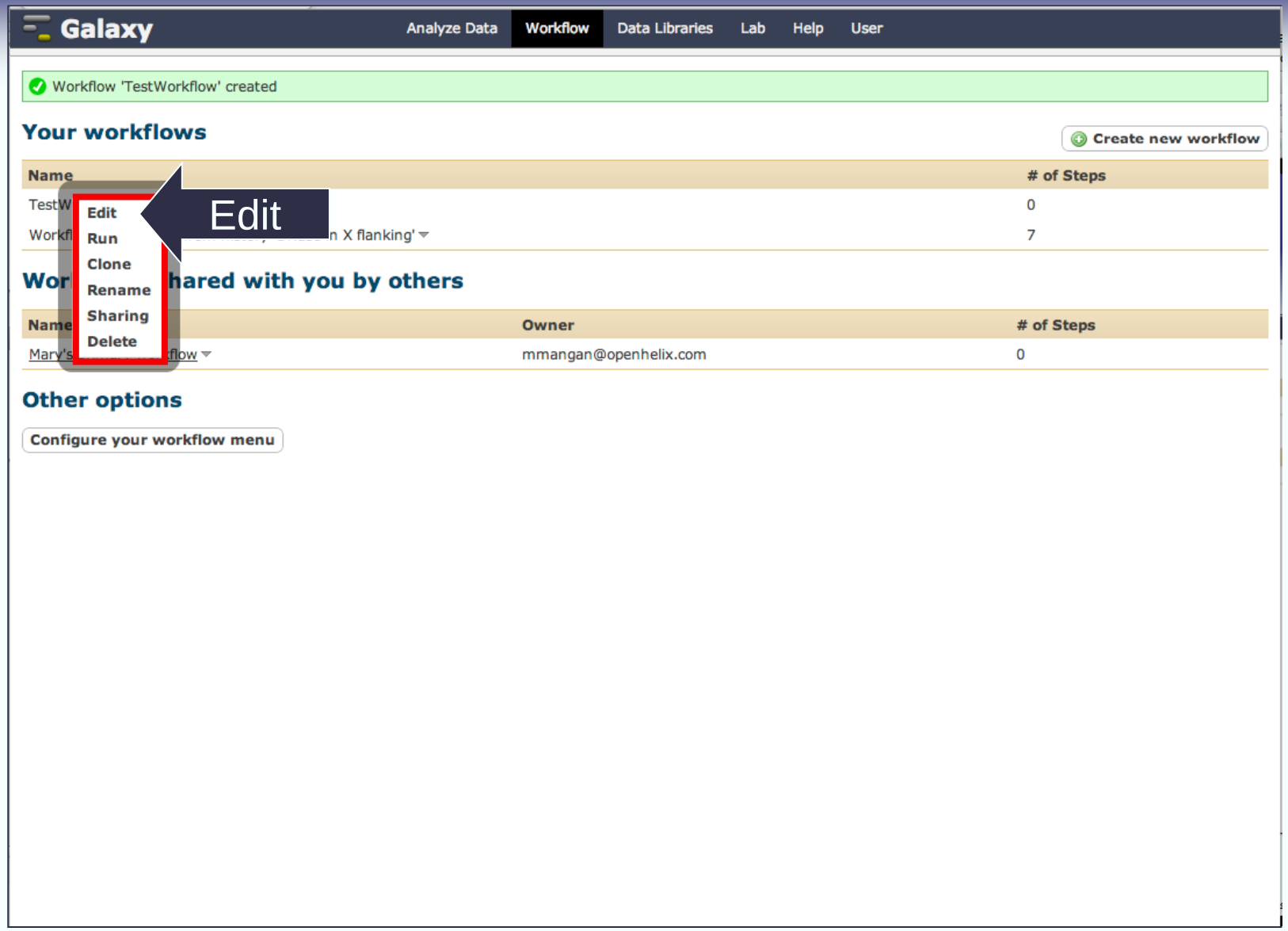
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Create New Workflow



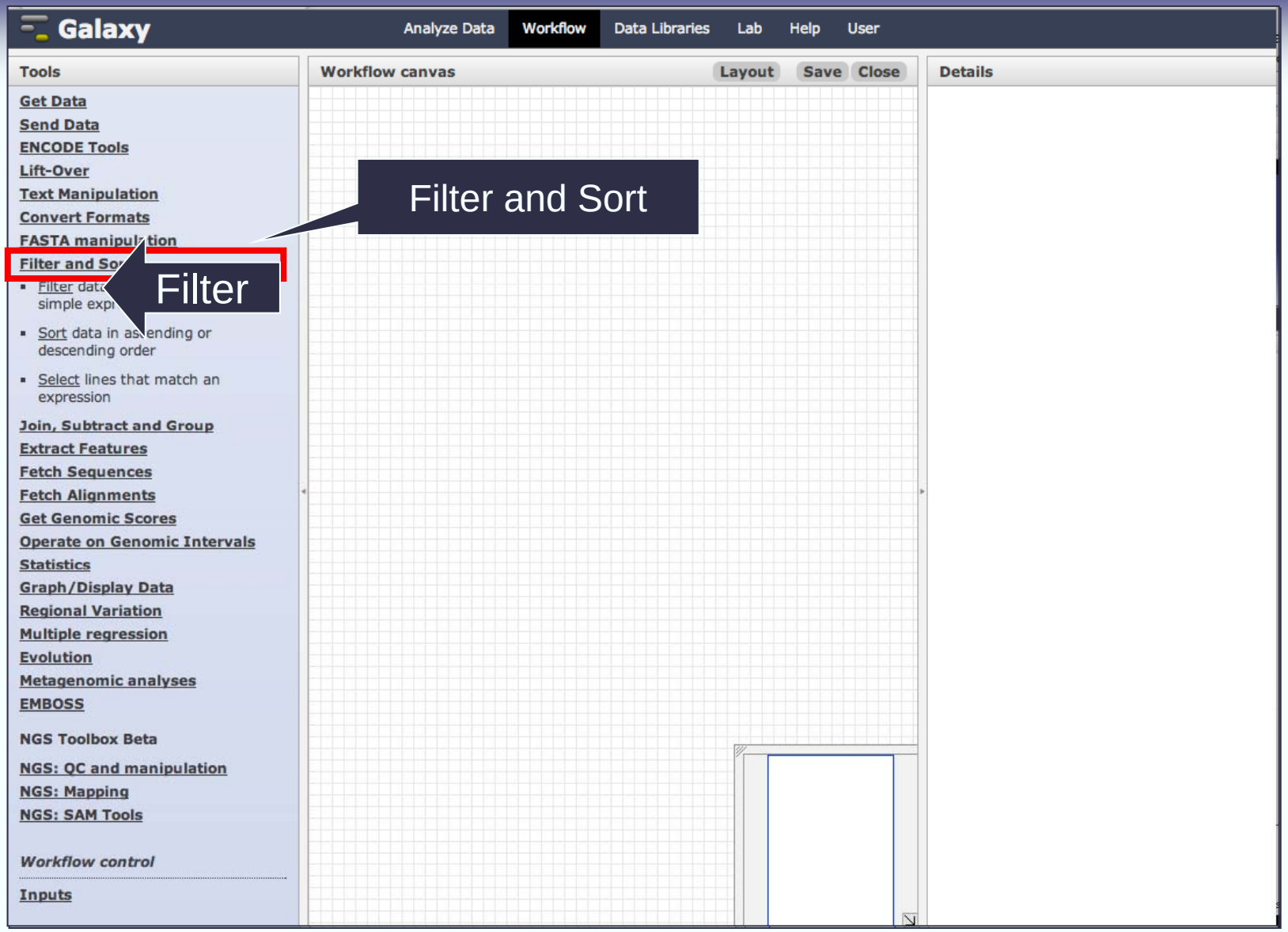
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Edit Workflow



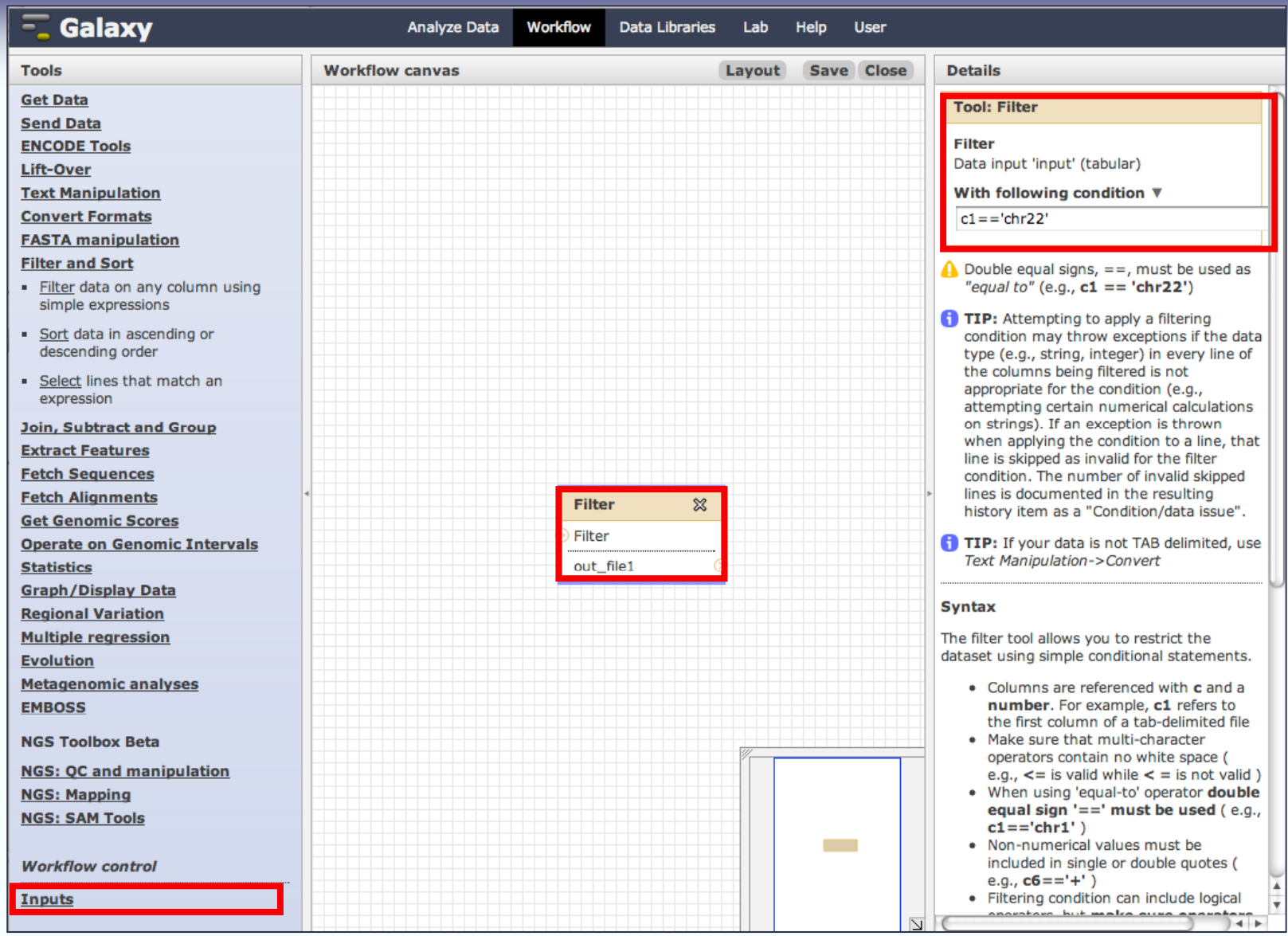
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Add Filter



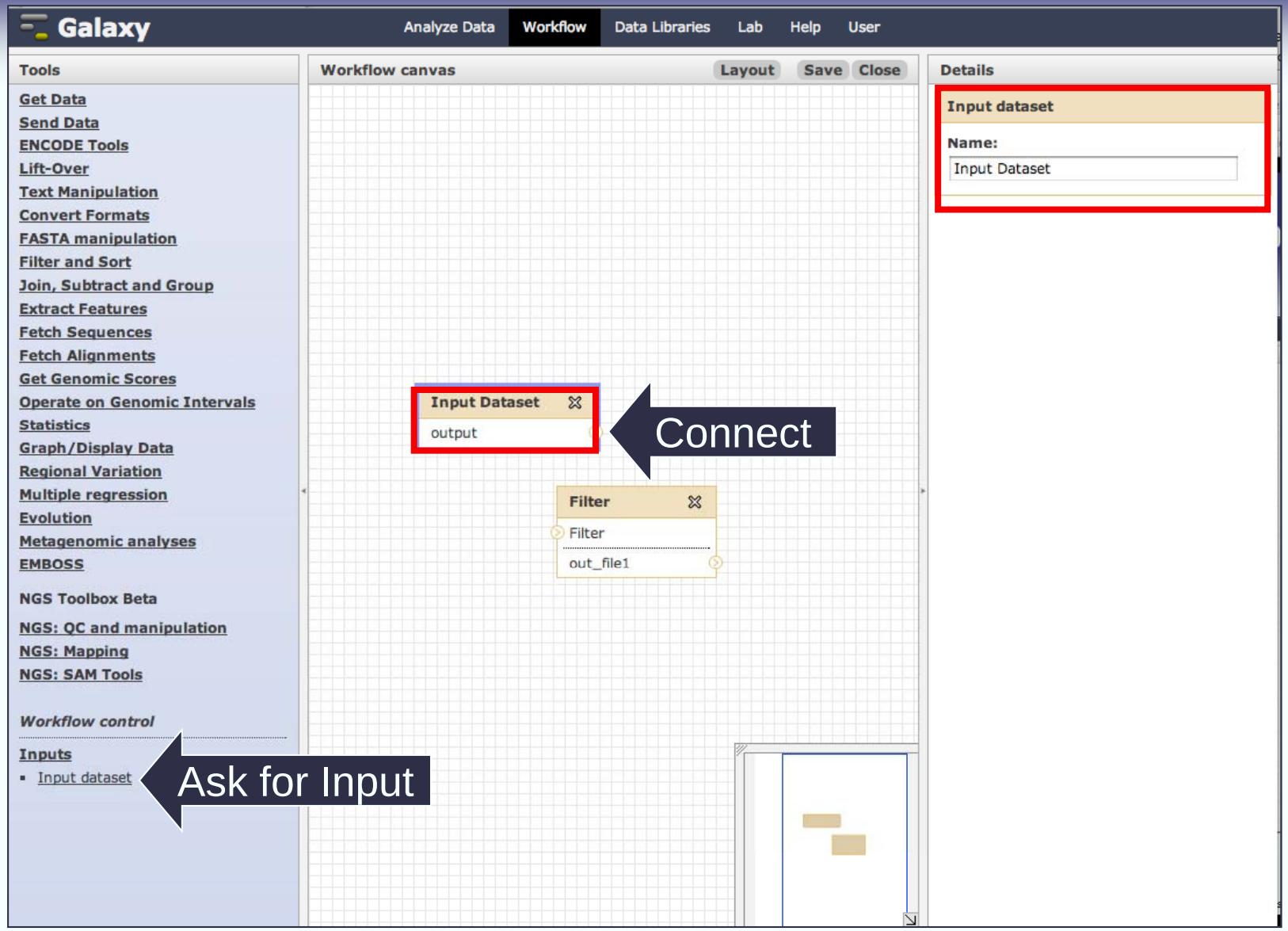
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Workflow Layout



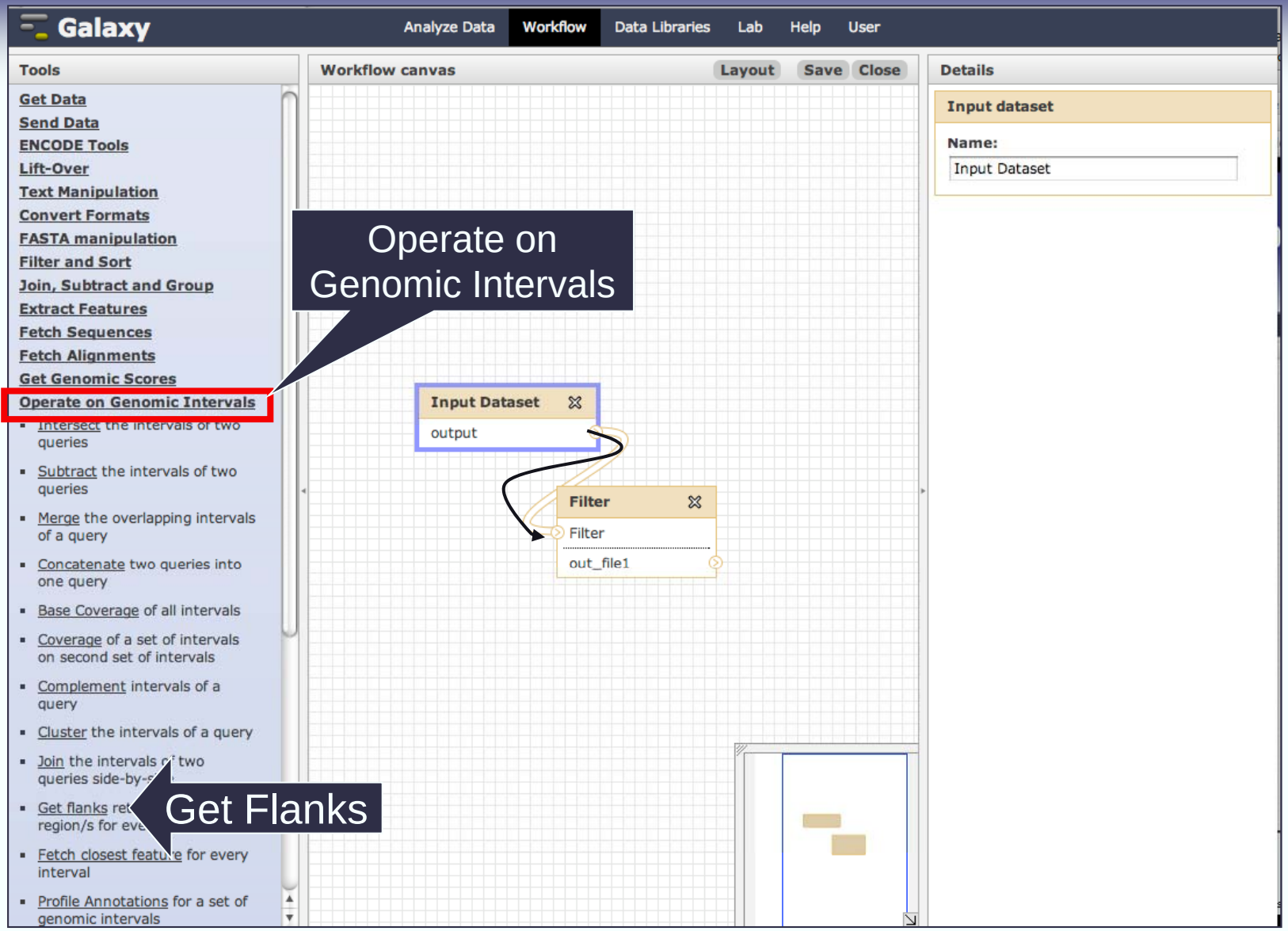
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Input Data and Connect Steps



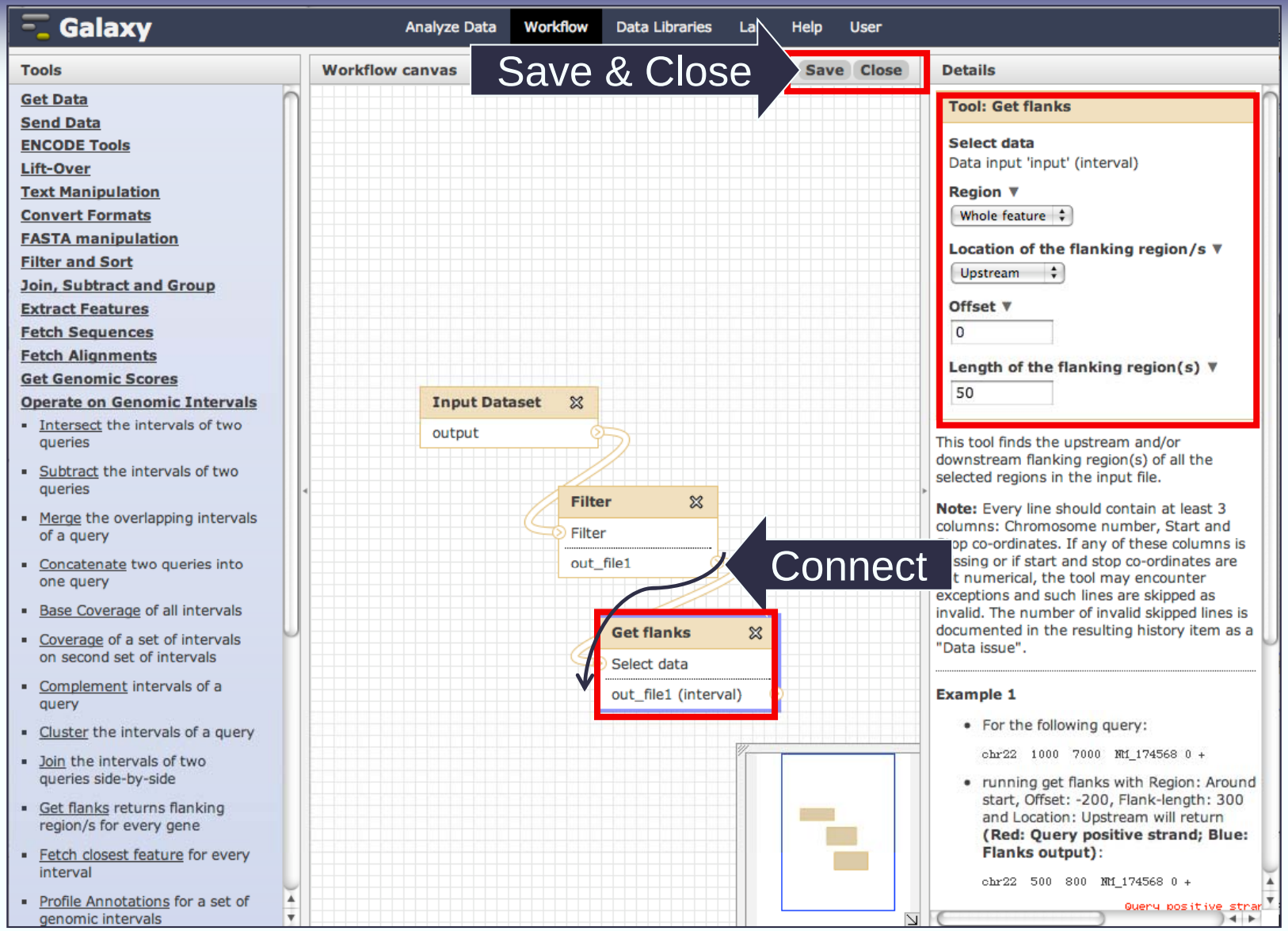
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Add Additional Steps, Flanks



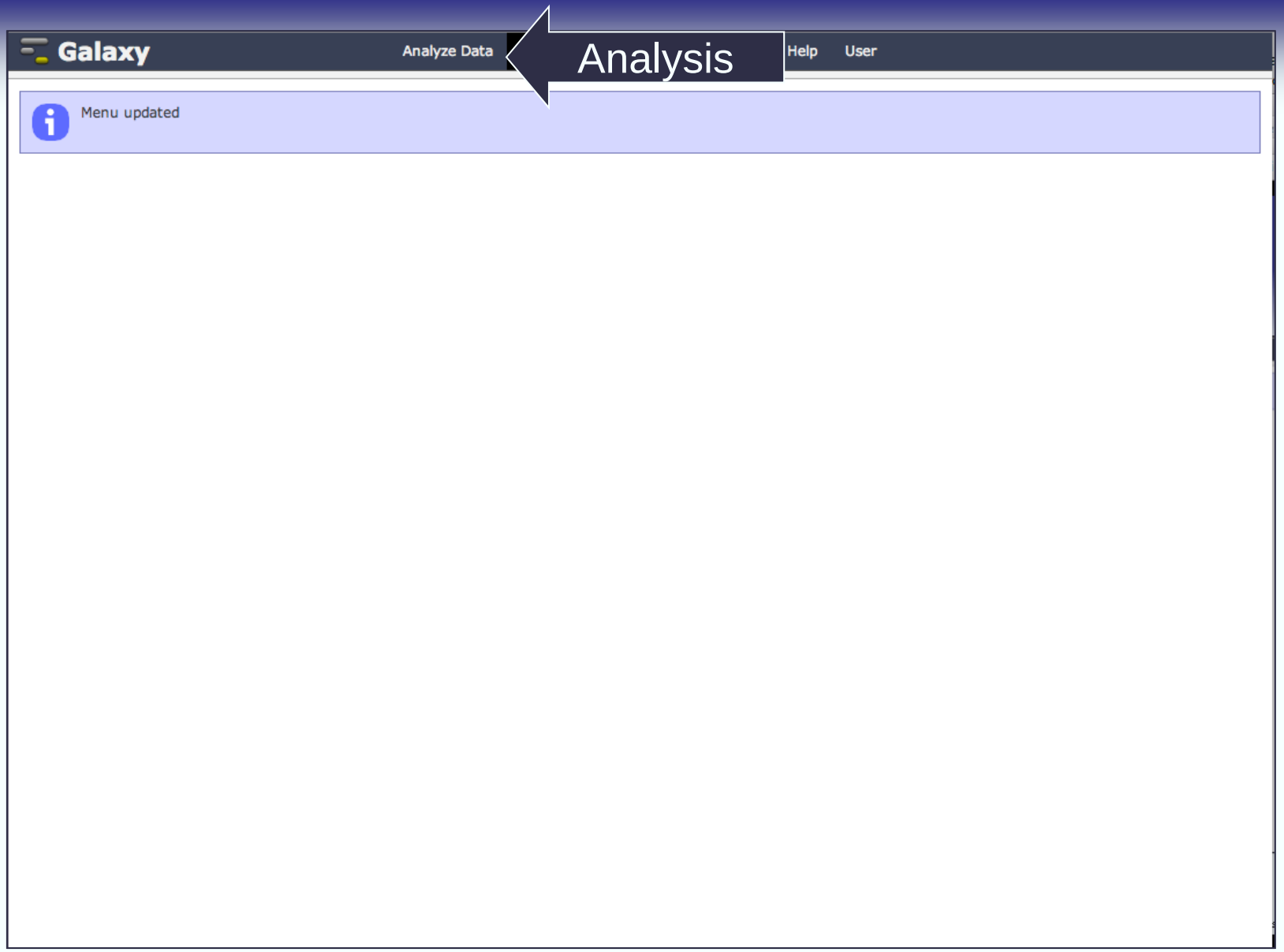
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Save and Close



Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Add to Menu



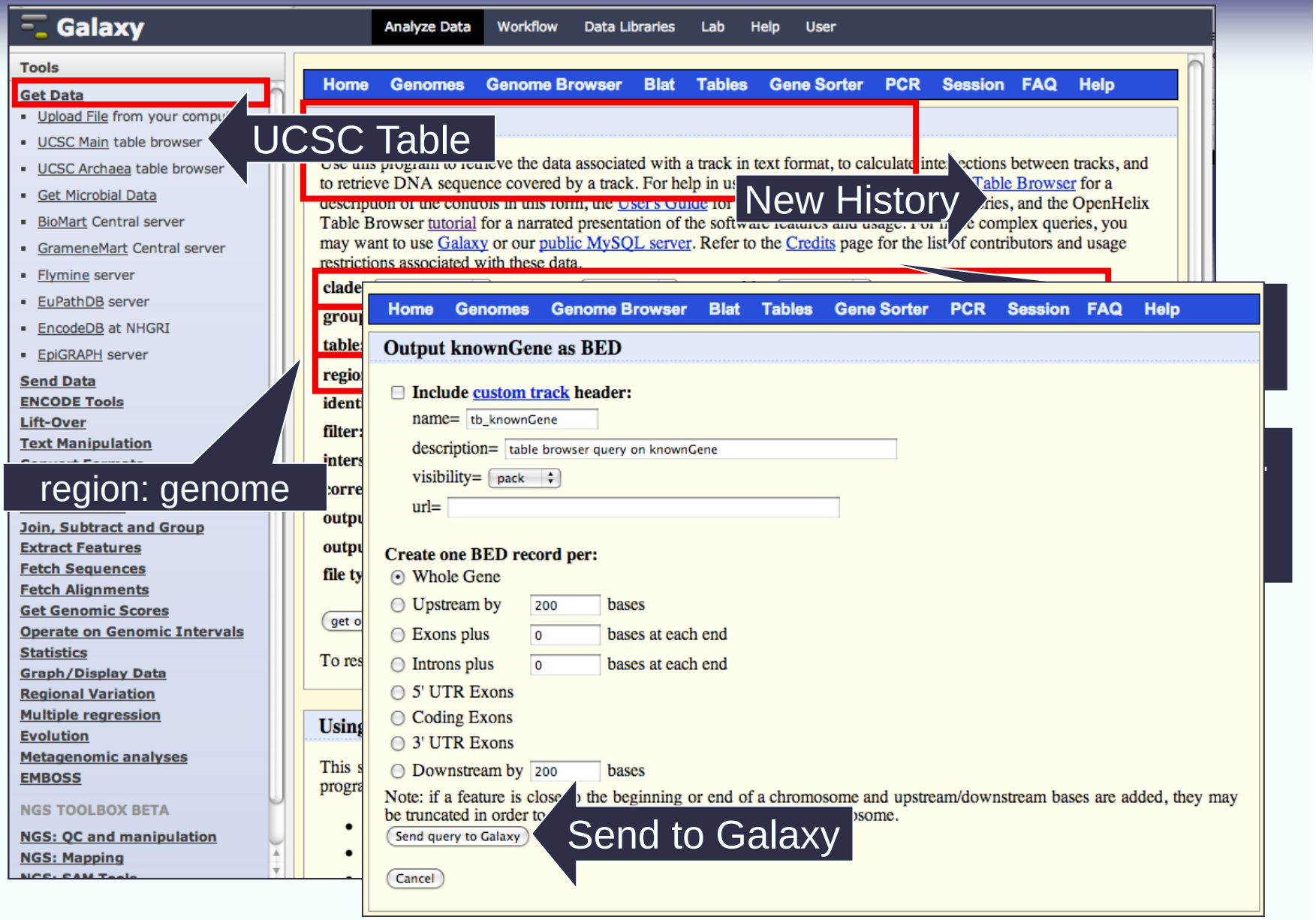
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: View from Menu



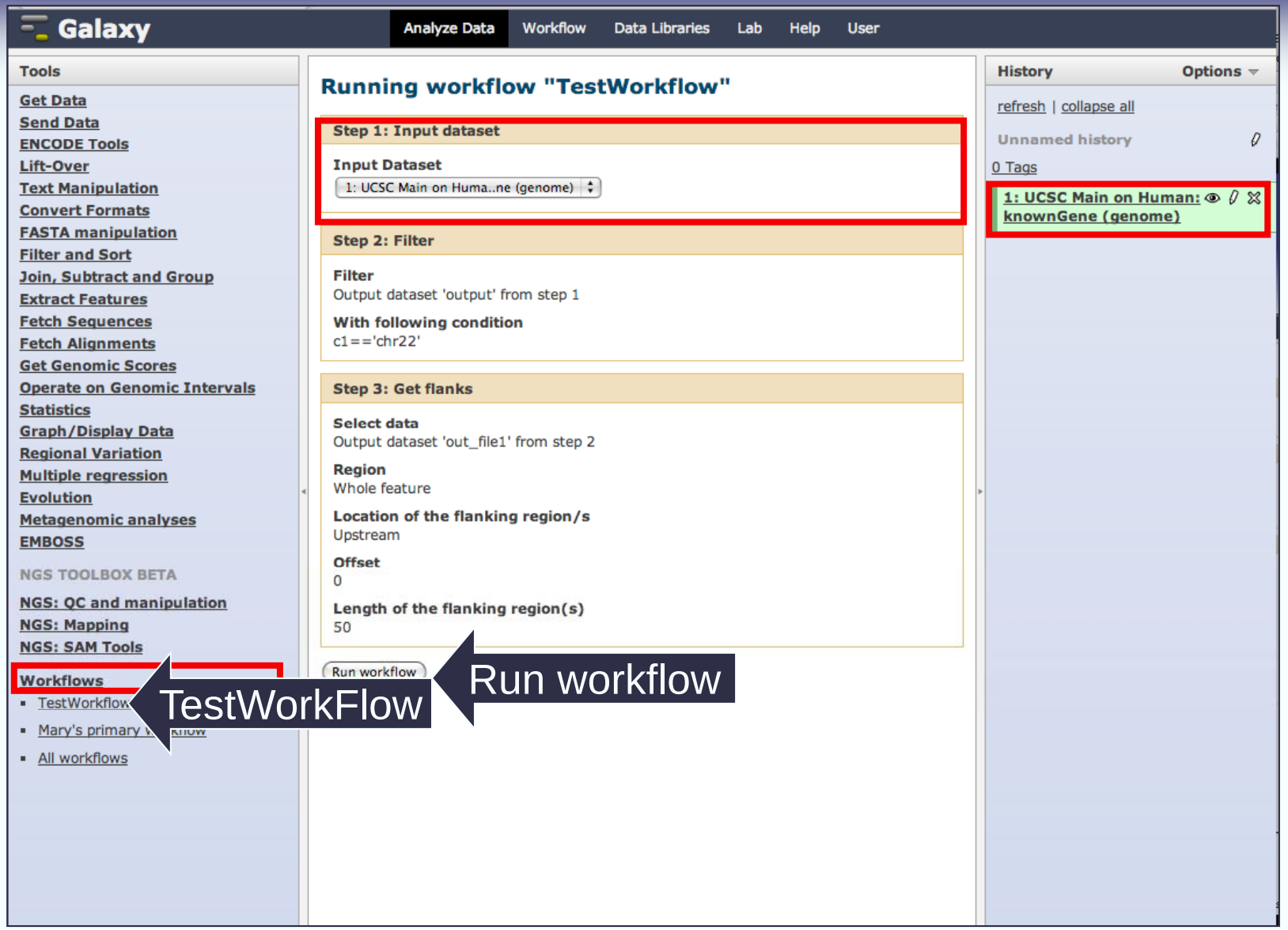
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Add Dataset Input



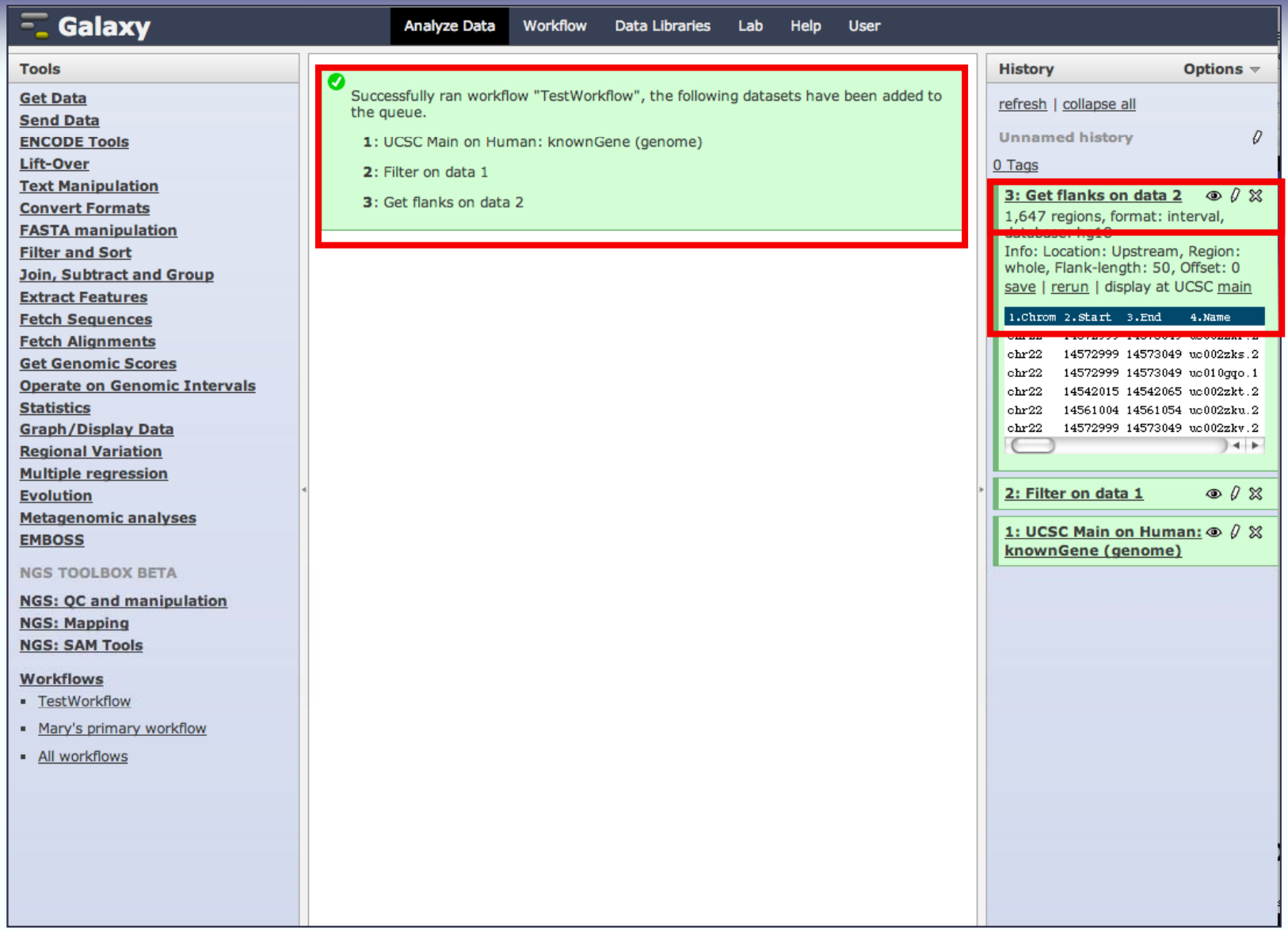
Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Run Workflow



Copyright OpenHelix. No use or reproduction without express written consent

Workflow: Get Resulting Data!



Copyright OpenHelix. No use or reproduction without express written consent



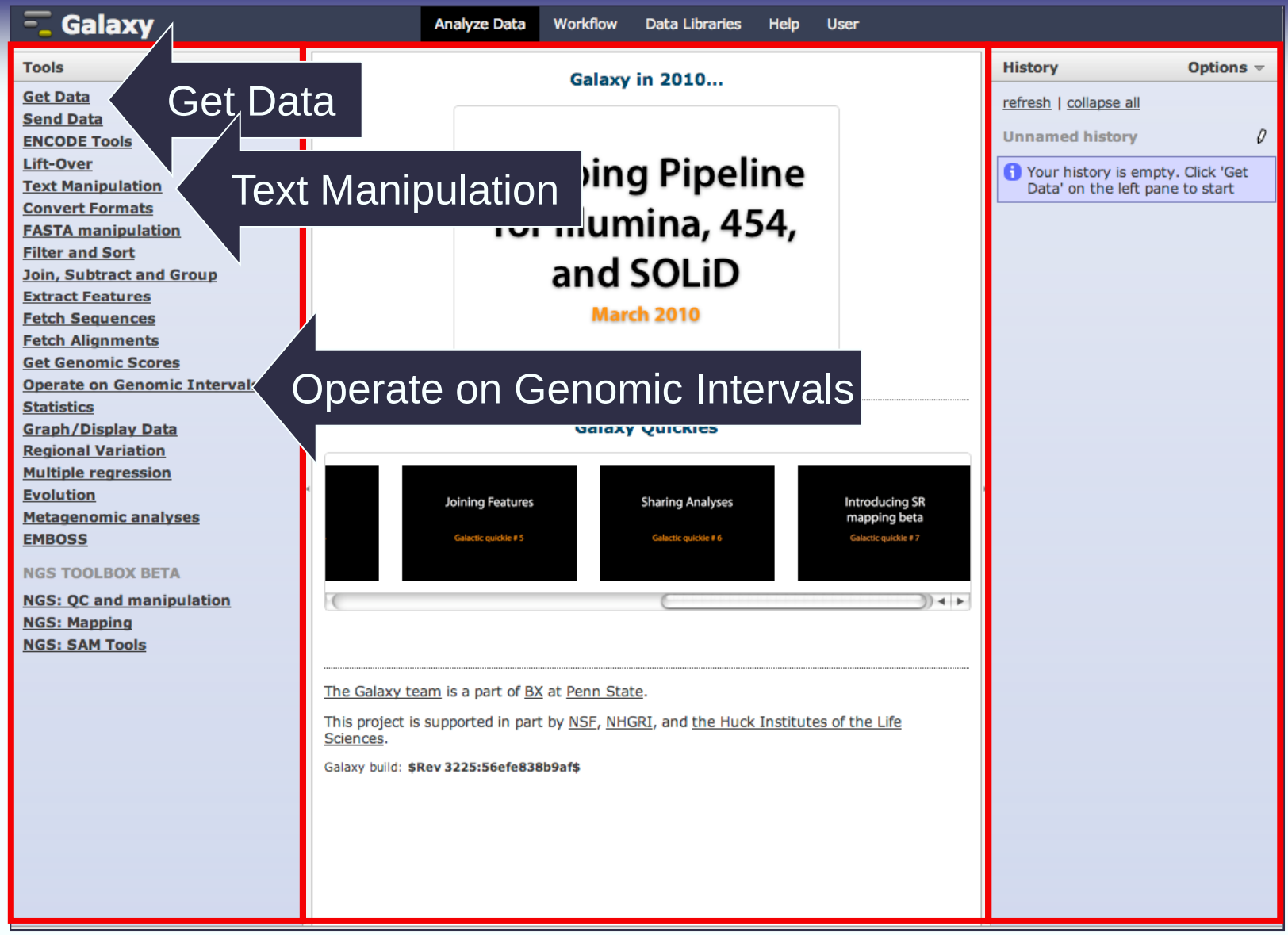
Galaxy Agenda

- Introduction and Credits
- Interface
- Getting Data
- Preparing Data
- Analyzing Data
- Histories & Workflows
- Summary
- Exercise

Galaxy: http://galaxy.psu.edu/

Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Interface



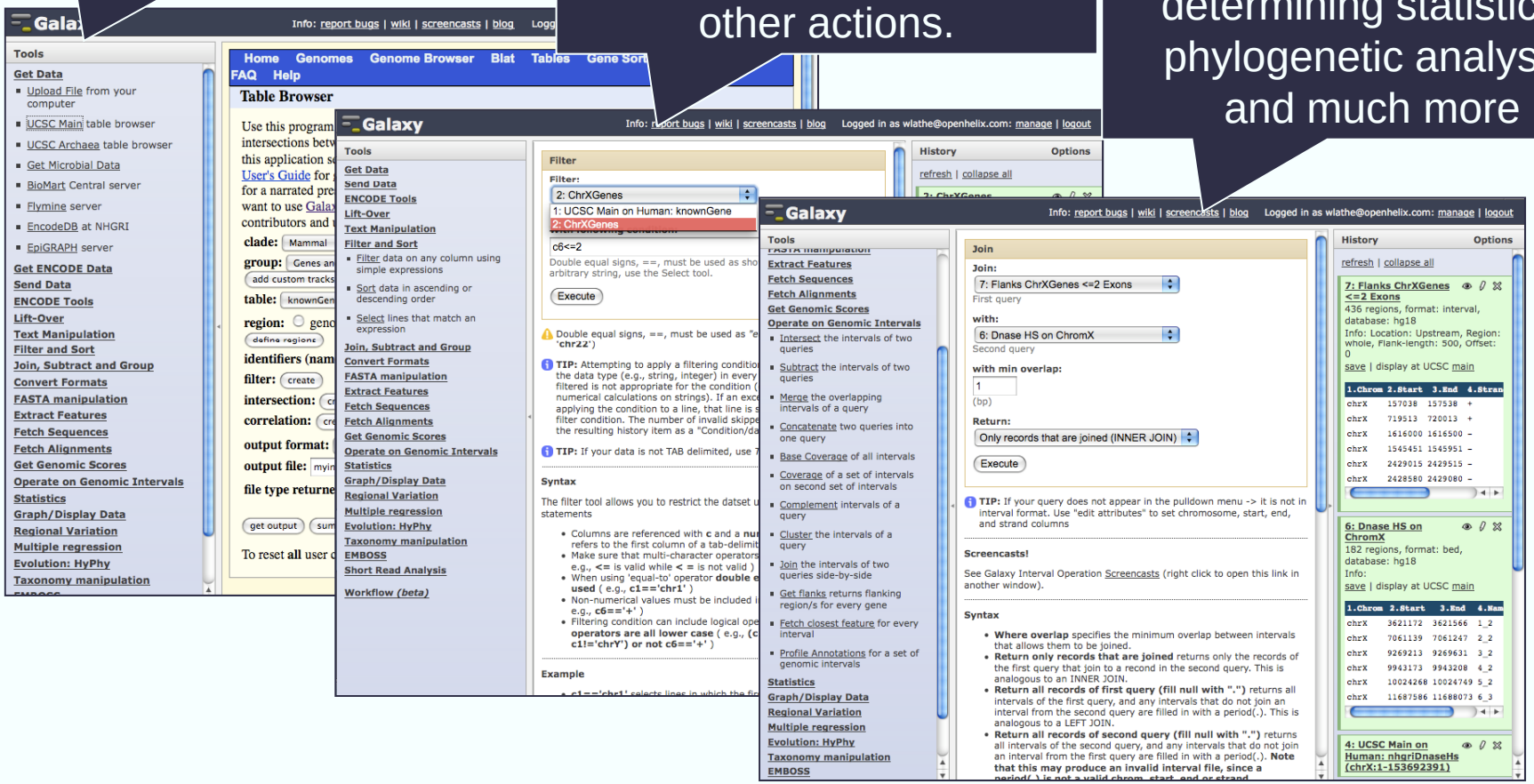
Copyright OpenHelix. No use or reproduction without express written consent

Galaxy

Obtain data from many data sources including the UCSC Table Browser, BioMart, WormBase, or your own data.

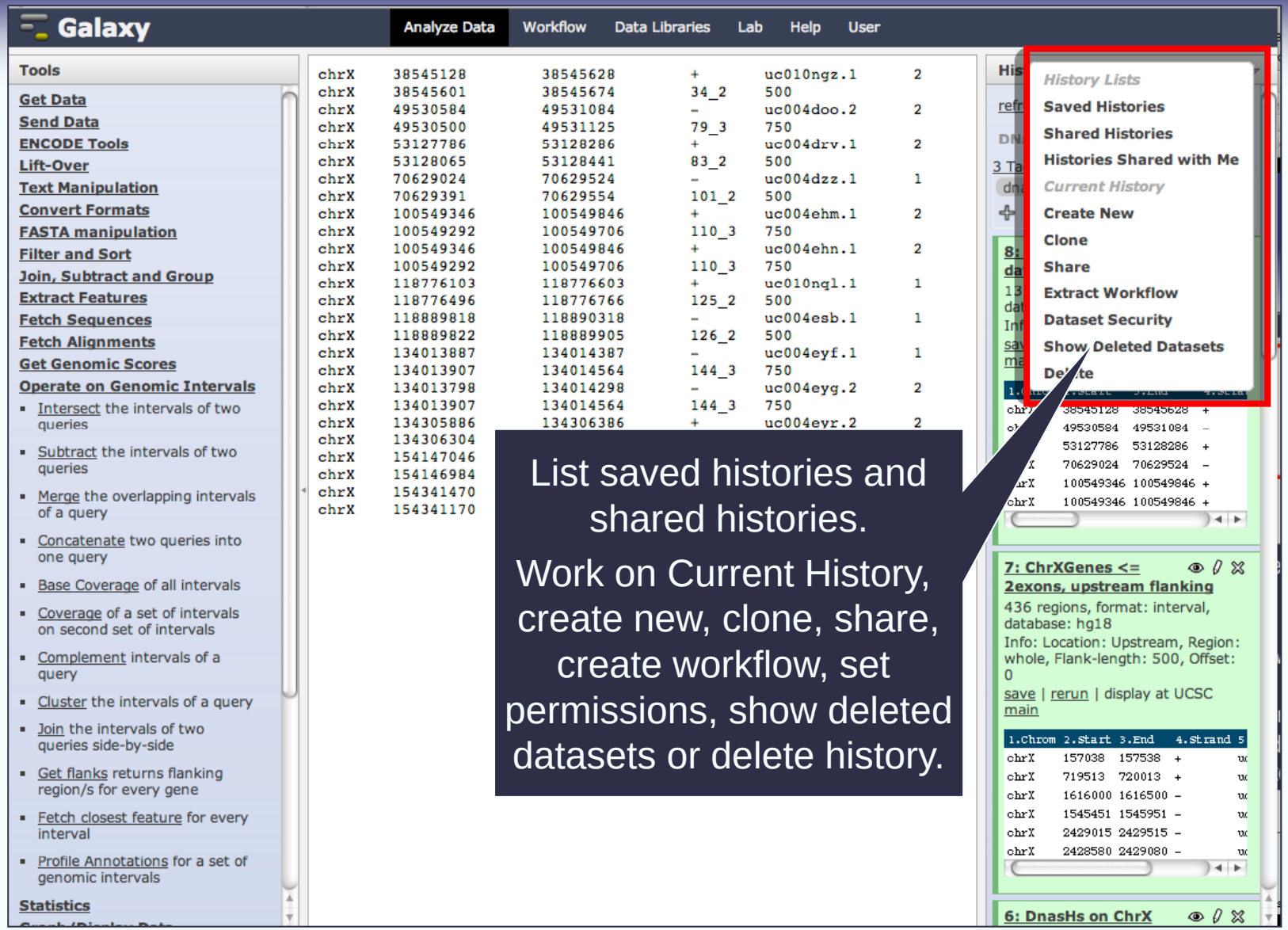
Prepare data for further analysis by rearranging or cutting data columns, filtering data and many other actions.

Analyze data by finding overlapping regions, determining statistics, phylogenetic analysis and much more



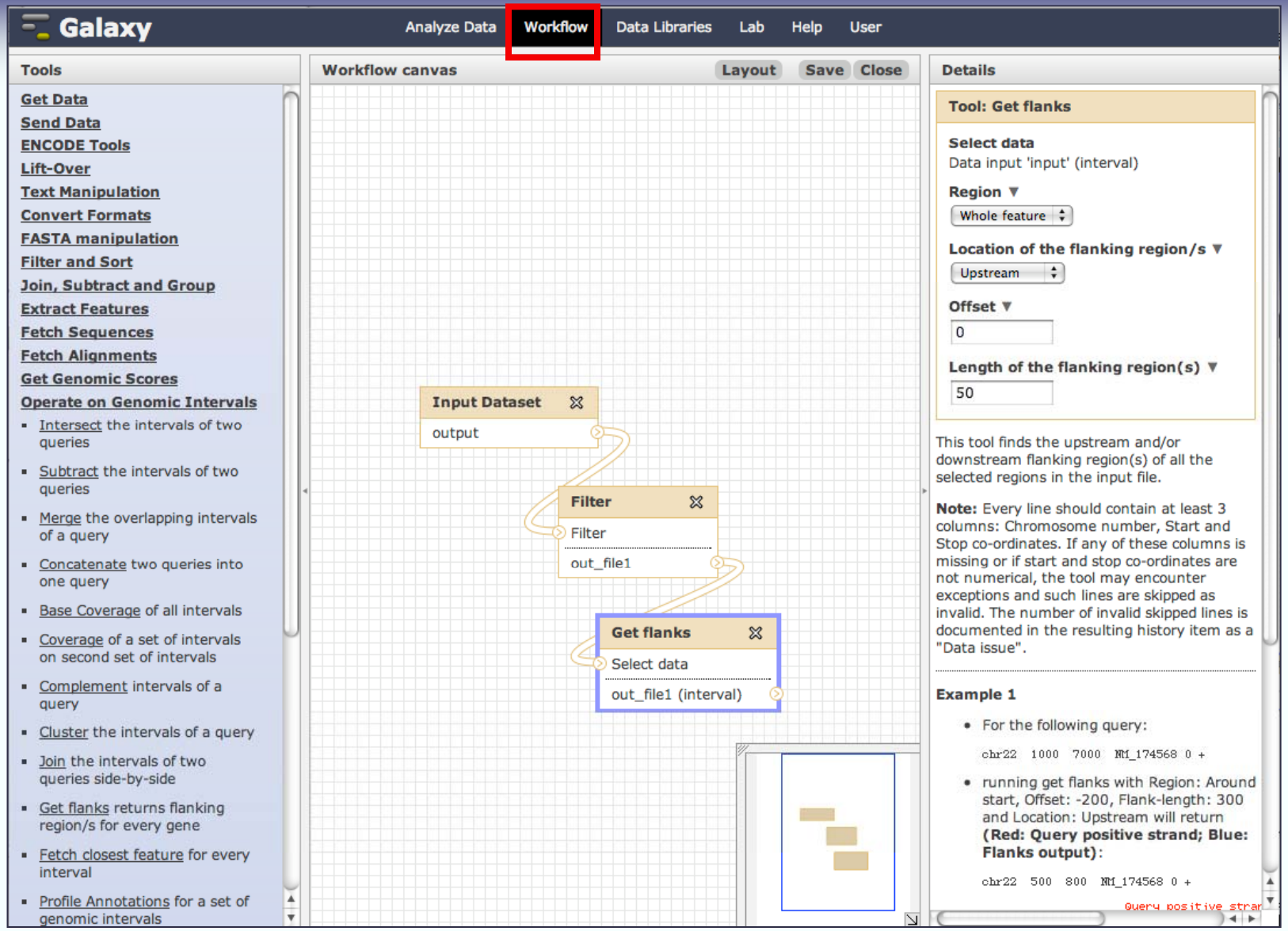
Copyright OpenHelix. No use or reproduction without express written consent

Histories



Copyright OpenHelix. No use or reproduction without express written consent

Workflows



Copyright OpenHelix. No use or reproduction without express written consent

Galaxy Agenda

- Introduction and Credits
- Interface
- Getting Data
- Preparing Data
- Analyzing Data
- Histories & Workflows
- Summary
- Exercise

Galaxy: http://galaxy.psu.edu/

Copyright OpenHelix. No use or reproduction without express written consent



Copyright OpenHelix. No use or reproduction without express written consent