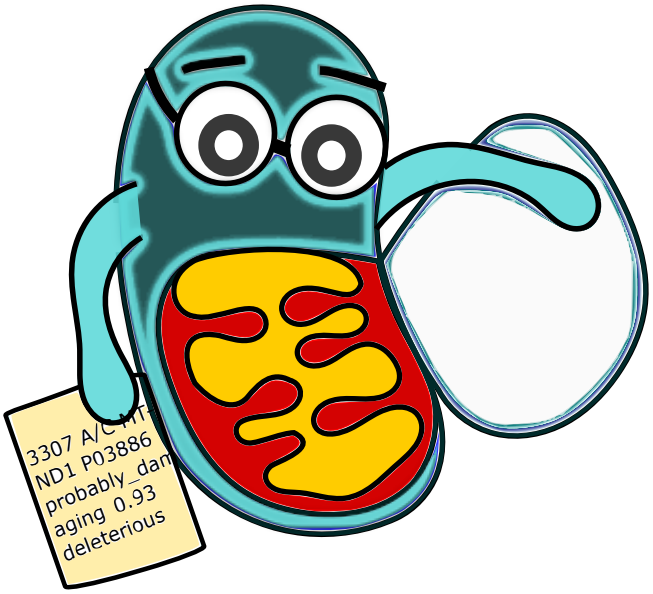

MitoLink Framework
User Manual 1.0

Introduction

MitoLink is an integrated workflow system to facilitate understanding of genotype-phenotype correlations in cases of mitochondrial dysfunction. The workflow system is implemented using the open source workflow architecture, Galaxy. MitoLink may be accessed at http://ab-openlab.csir.res.in/mito and is freely available.

There are six modules in MitoLink, described below. All the modules are accessible without registering to the system. However, for maintaining user-sessions it is recommended that anyone who is interested in creating data or task intensive workflow should register. The benefit of registration includes user sessions, saved histories, visualization, generation and execution of workflow and many others.

All the modules of MitoLink and some default modules by Galaxy are accessible through a web-based interface which has following components:

1. Navigation Panel: It provides the links to major components of the server like Tools Page (Analyze Data), Workflow System, Shared Libraries, Visualization, Help Section and User Login/Registration.

2. Tool Panel: This panel lists all the tools available in MitoLink along with default utilities in Galaxy.

3. Detail Panel (Canvas): This panel displays the interface of all the tools along with Input Parameters required to run a tool. It also provides help and examples to run a tool. This panel also displays the Output of a tool after its execution when user clicks on the eye  icon show in **History Panel**.

4. History Panel: This panel shows the information about the tools which are executed by a user. The information can include result after completion of a tool execution or error generated while running the tool. The workflow(s) are generated by extracting tasks from history panel.

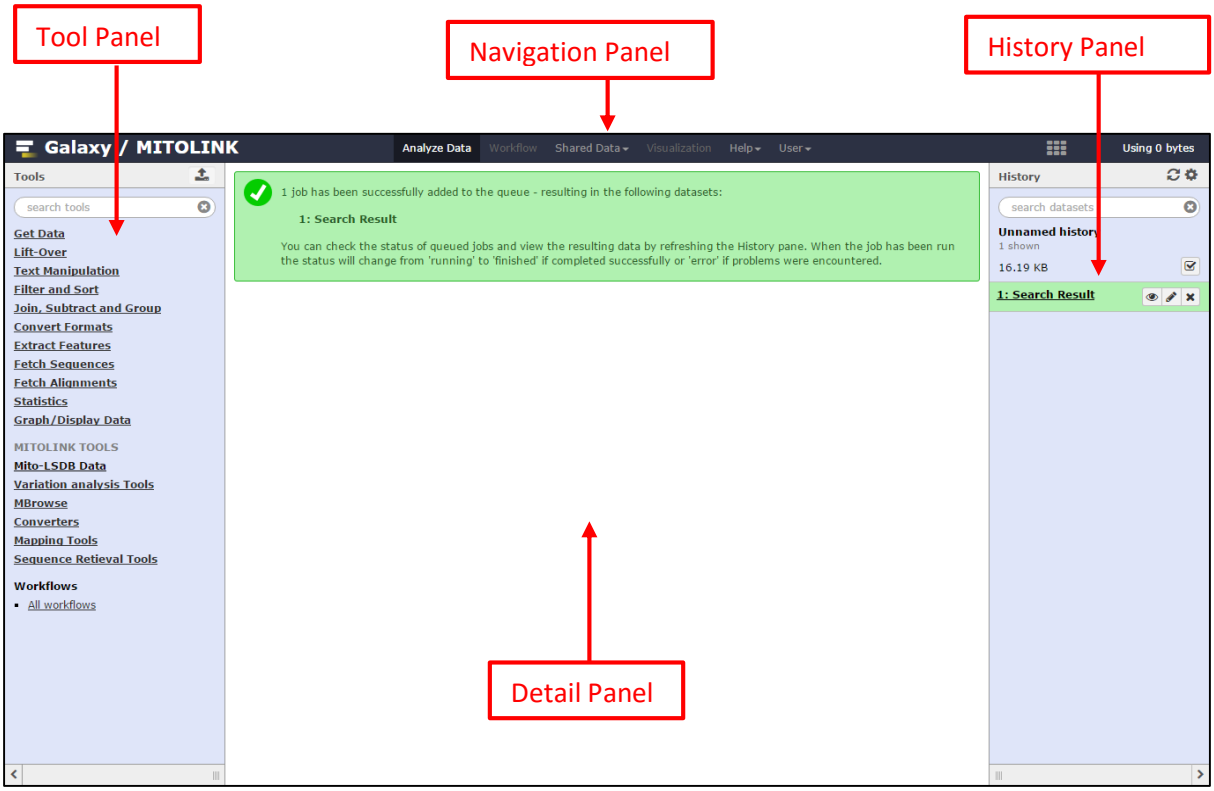

Figure 1: MitoLink Homepage

I. Navigation Panel

1. Analyze Data

The data analysis page is where everything happens. There, you can run any available tools on the data, run complete workflows, browse or download a results, and share files with other users. It is the default page when you open Galaxy in your browser, but you can also access it any time by clicking on "Analyze Data" in the Navigation Panel.

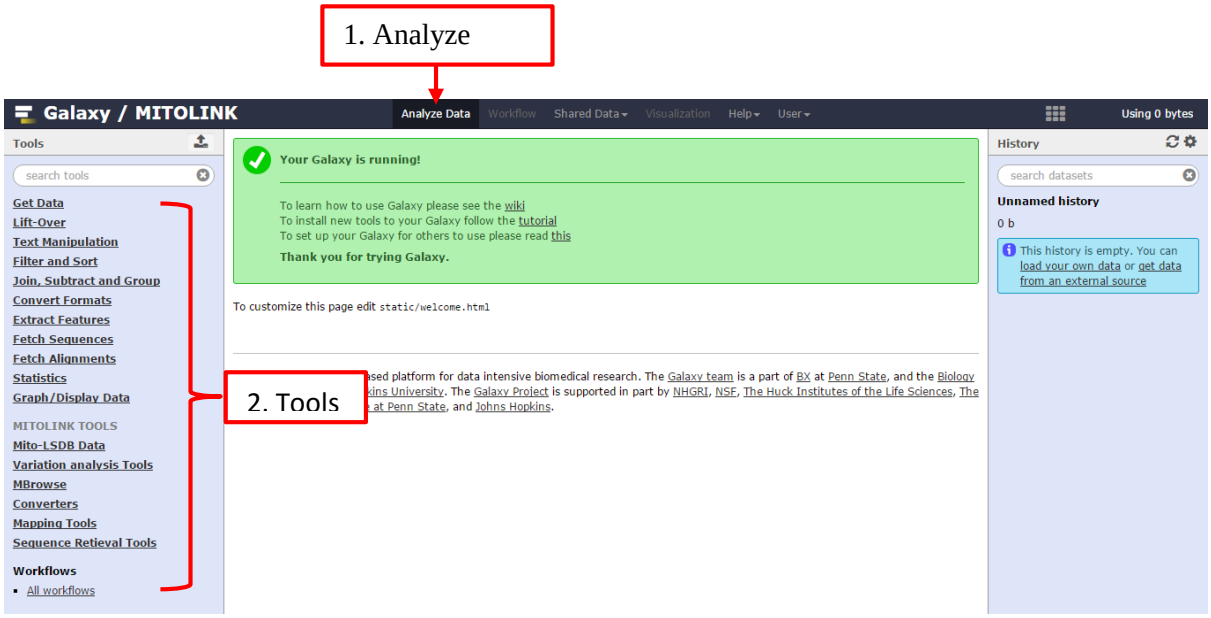

Figure 2: Analyze Data view

2. Work flow

Workflows are analyses that are intended to be executed (one or more times) with different user-provided input Datasets. Workflow can be reused over and over, not only reducing tedious work, but enhancing reproducibility by applying the same exact methods to all of your data. Workflow is nothing but creating pipeline, user can use it again and again or user can published it.

Workflow can be created through navigation panel or from tool panel. In workflow section user can create workflow or can upload or import the workflow. The canvas is where inputs, tools, and noodles are added and connected as you build and modify your workflow (Figure 3). Selecting Edit opens the workflow editor view (Figure 2). The navigator provides a full view of your workflow in a condensed format (Figure 3). Accessed by clicking on the gear icon on the right side of the center Workflow Canvas upper bar, the workflow editor menu (Figure 3) is for global editor actions. It consists of Save, Run, Edit Attributes, Auto re-layout, Close

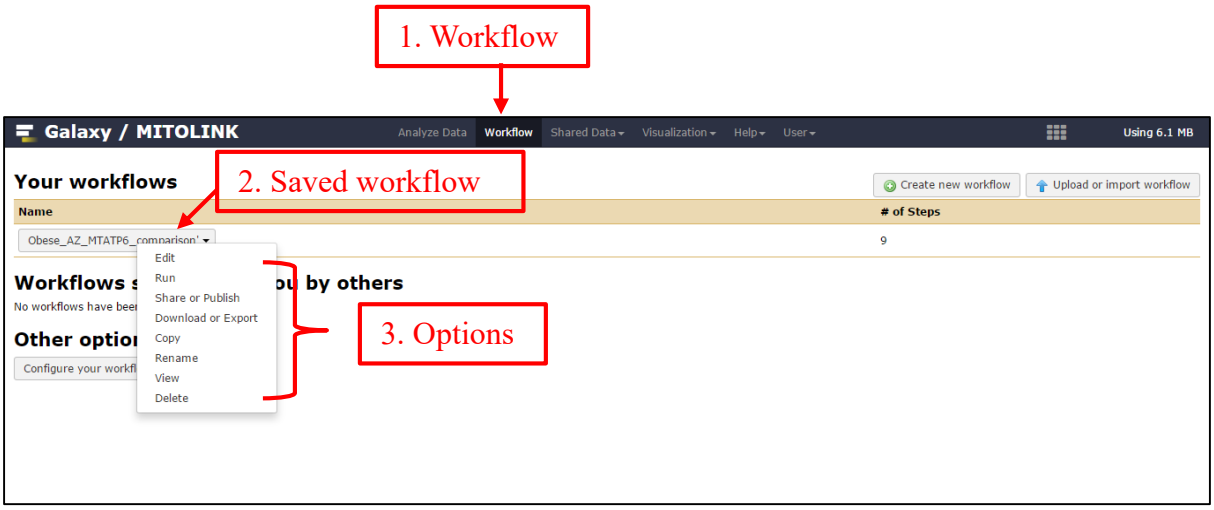

Figure 3: Options for workflow

The following example of workflow shows the comparison between Obese and Alzheimer disease with MTATP6 gene. The generated result of Obese and Alzheimer disease from Query Builder passed to the MtSNPscore. MtSNPscore generate result in five different form. Then we can do various analysis on generated result. Here we did filteration and comaparison between Obese and Alzheimer disease.

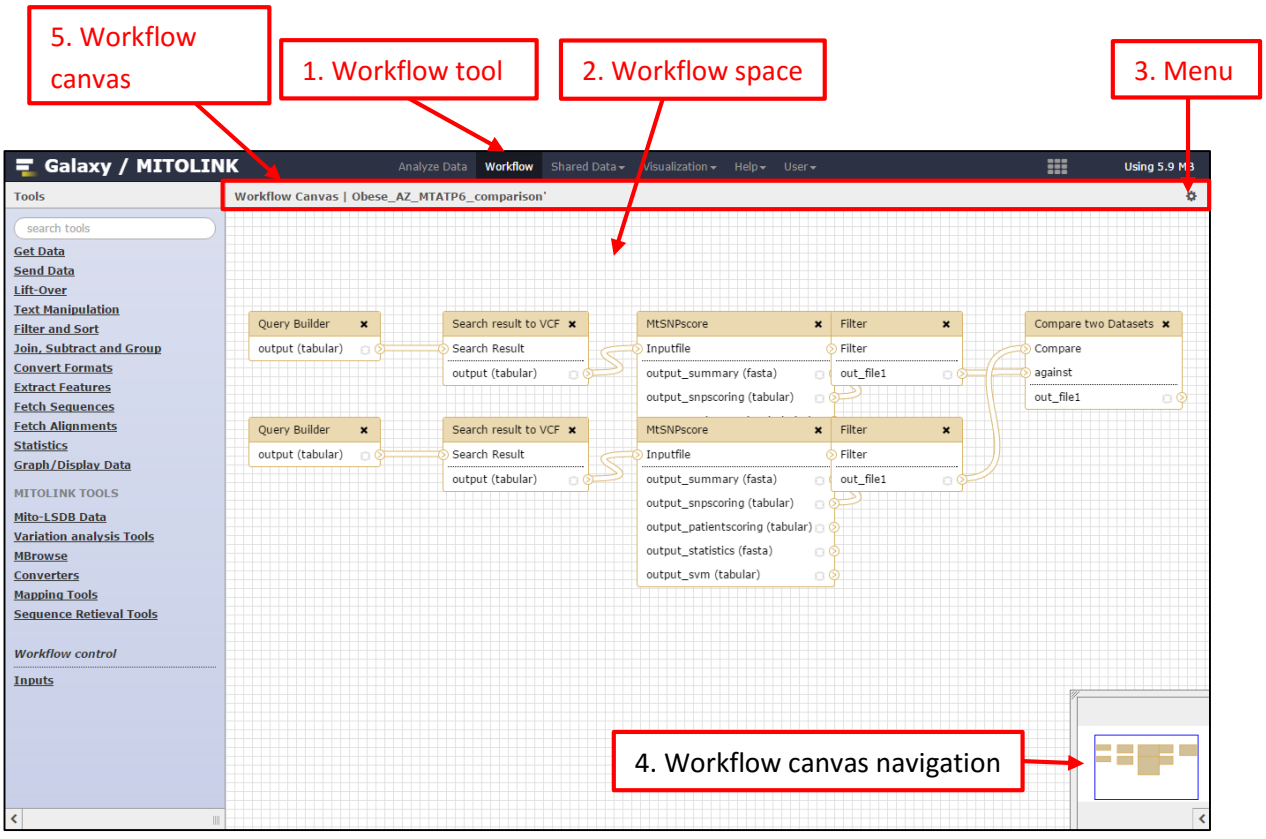

Figure 4: workflow overview

3. Shared data library

Data libraries are collections of Datasets that are accessible from within a Galaxy instance. Libraries are designed for sharing datasets in between users or groups. The data library of

MitoLink consists of Mitochondrial protein list. Some of the actions that can be performed on data libraries are accessed by clicking the pop-up menu icon just right of the data library name.

- View Information – Shows the information about dataset.
- Import this dataset into your current history - this creates an item in your current history on which you can perform analysis. The item is a pointer to the library dataset disk file, so the file is not copied on disk.
- Download this dataset - this allows you to download a local copy of the dataset.

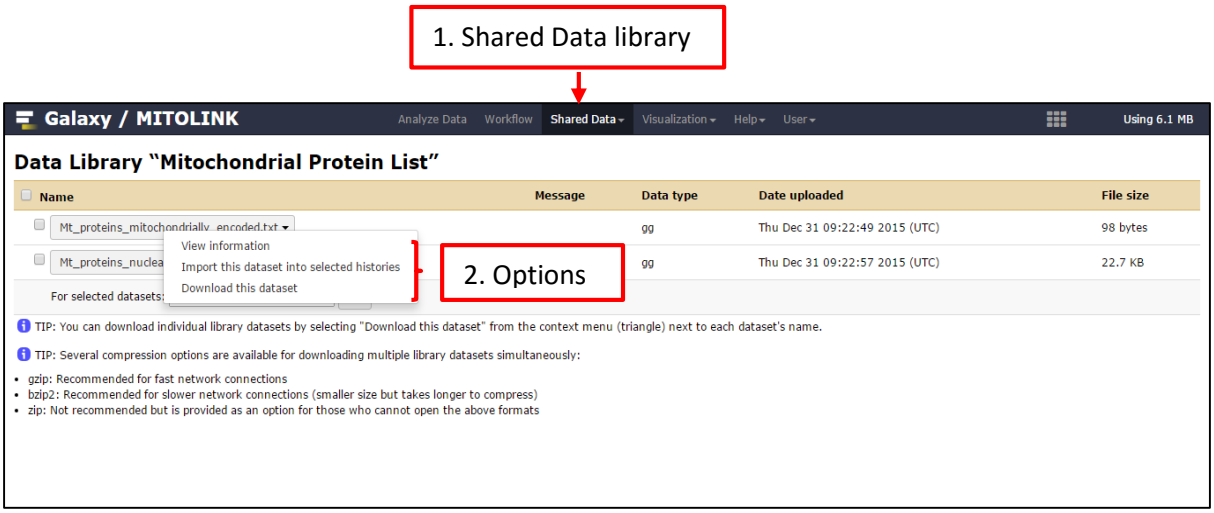

Figure 5: Data library view

4 Help

The help section of Galaxy consist of Support, Search, Mailing List, Videos, Wiki and How to cite Galaxy. User can fine user manual in help section.

5. User

Login option and register option can get in user section. It is recommended that user register their account before using framework. Although unregistered users have access to tools available but their history is stored temporarily. On the other hand, registered users can save and retrieve their results in history panel later too.

II. Tool Panel

The user can get tools in tool panel. There are two categories of tools galaxy inbuilt tool and MitoLink tools. Galaxy tools consists of Data importing, Manipulation, Filtering, Sorting, Format conversion etc. MitoLink tool are specific for finding Genotype-Phenotype correlation.

1. Importing data to the MitoLink

A user can upload the data using the Galaxy tool **Get Data.** The uploaded data can be used for MtSNPscore analysis. There are many other services like UCSC browser, EBI ENA services, BioMart server, Flymine server etc. available in Get Data section which can be used for importing the data. If the data is not available in appropriate form, converter can be used. User can perform a query on the MitoLSDB data and pass the result to the MtSNPscore for further analysis. The following figure shows the file upload method.

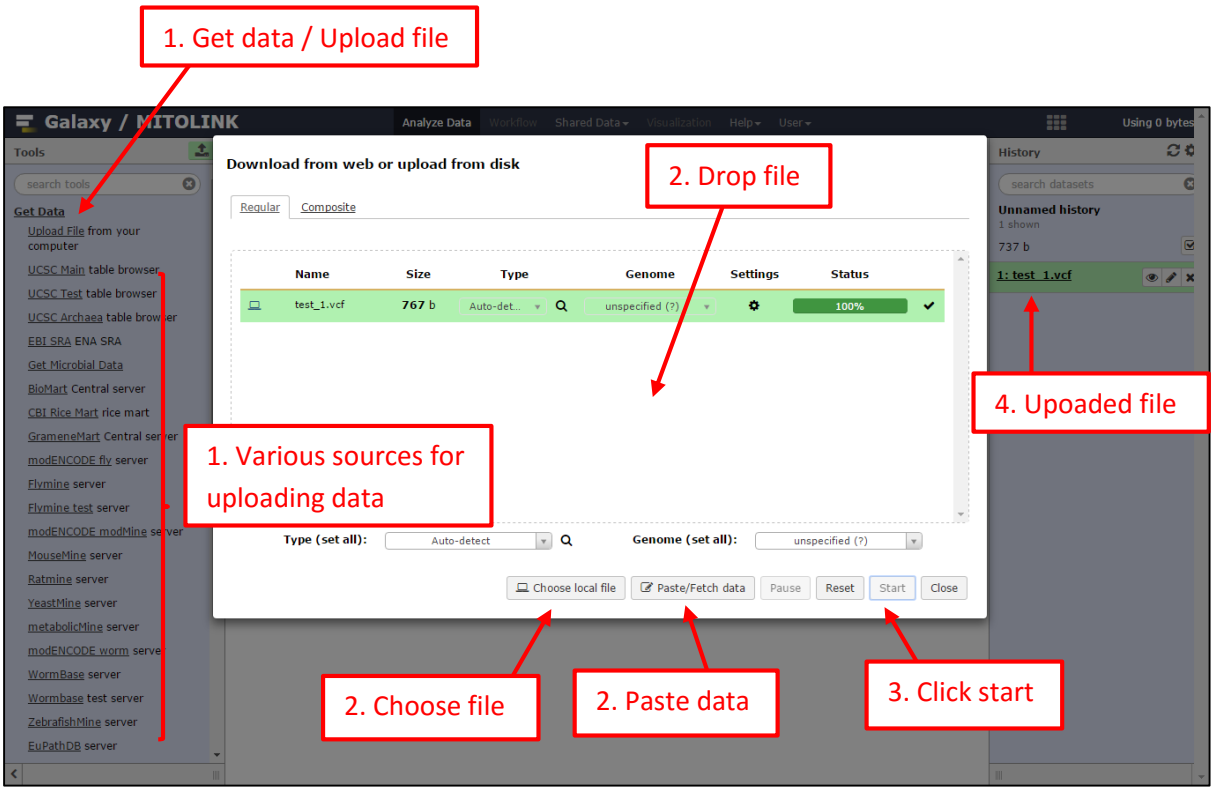

Figure 6: Get Data screen

2. Mito-LSDB Data – Query Builder

MitoLSDB is an integrated platform to catalogue disease association studies on mtDNA. MitoLSDB work according to the Gene, variant DNA position, variant DNA wild-type, variant DNA muted, variant protein position, variant protein wild-type, variant protein muted, variant codon change wild-type, variant codon change mutated, variant frequency, disease and population. The generated result can be used as an input for MtSNPscore for further analysis.

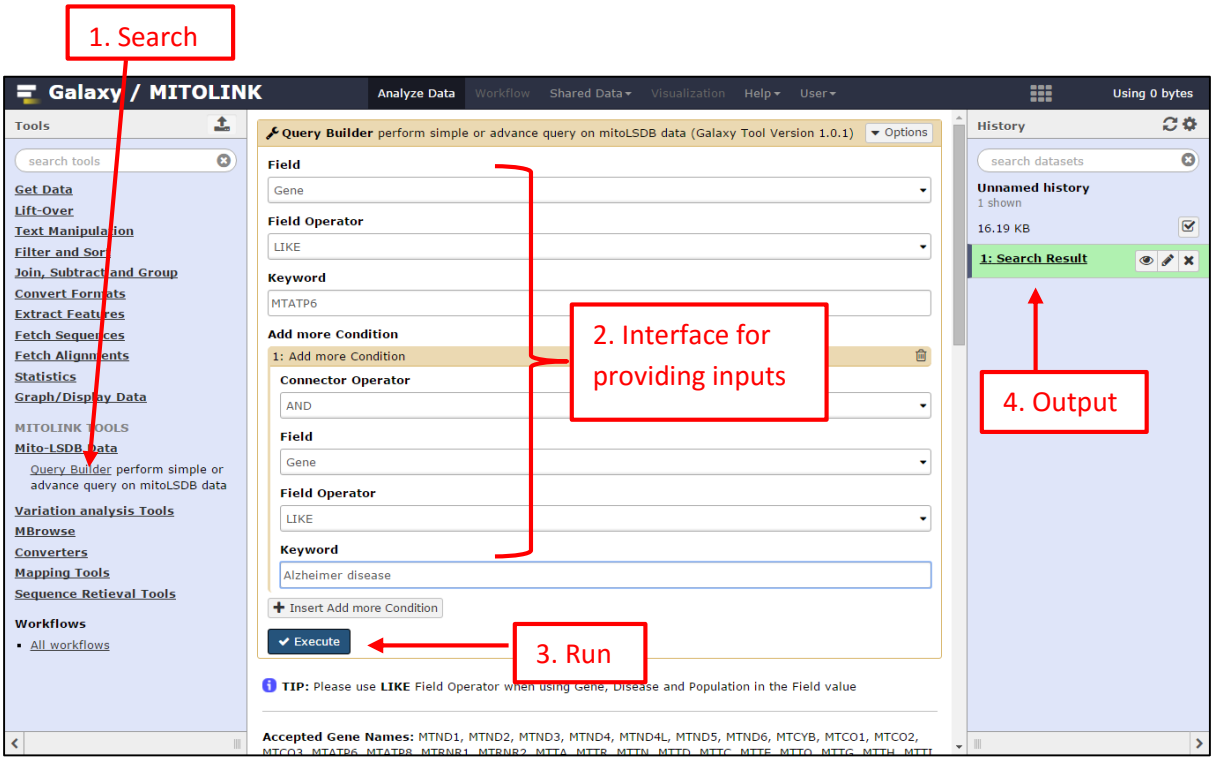

Figure 7: Query Builder tool view

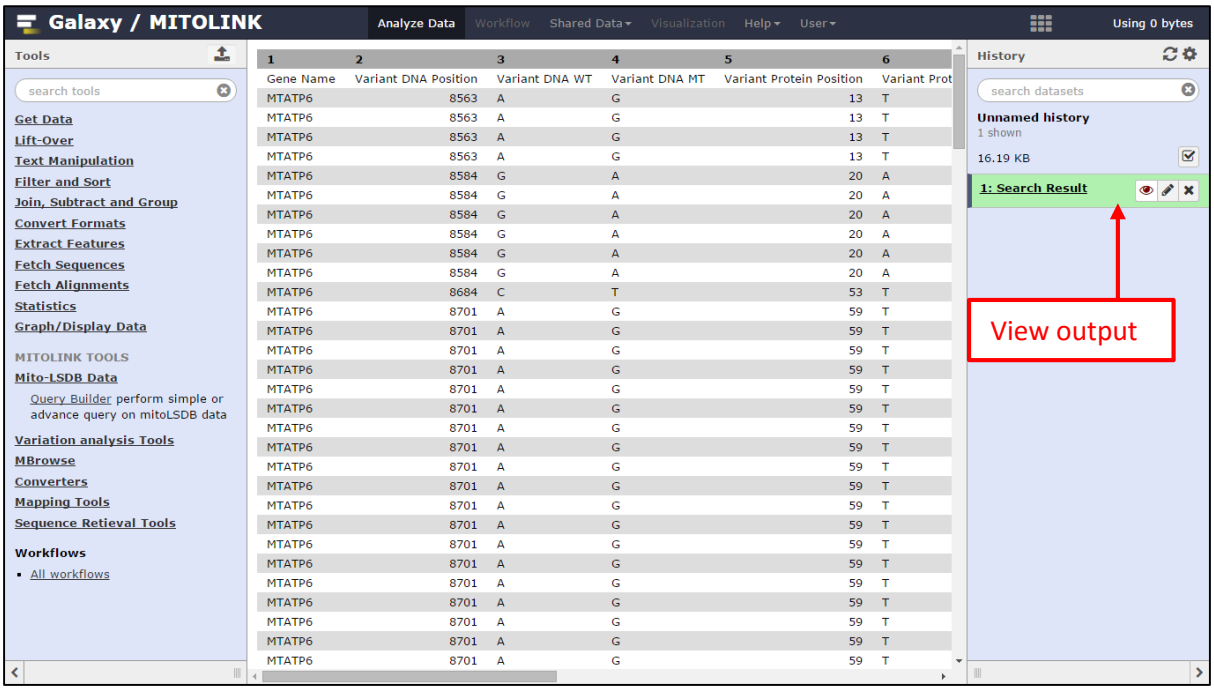

Figure 8: Query builder output

3. Variation analysis Tools

Variation analysis tools can be used for Nuclear Genome and Mitochondrial Genome Variation Analysis.

A. PolyPhen

PolyPhen is a nuclear genome variant analysis tool. PolyPhen is a tool which predict possible impact of amino acid substitution using structural and evolutionary consideration for all mitochondrial proteins. User can get list mitochondrial protein which are mitochondrial encoded and mitochondrial protein which are nuclear encoded in the shared data library.

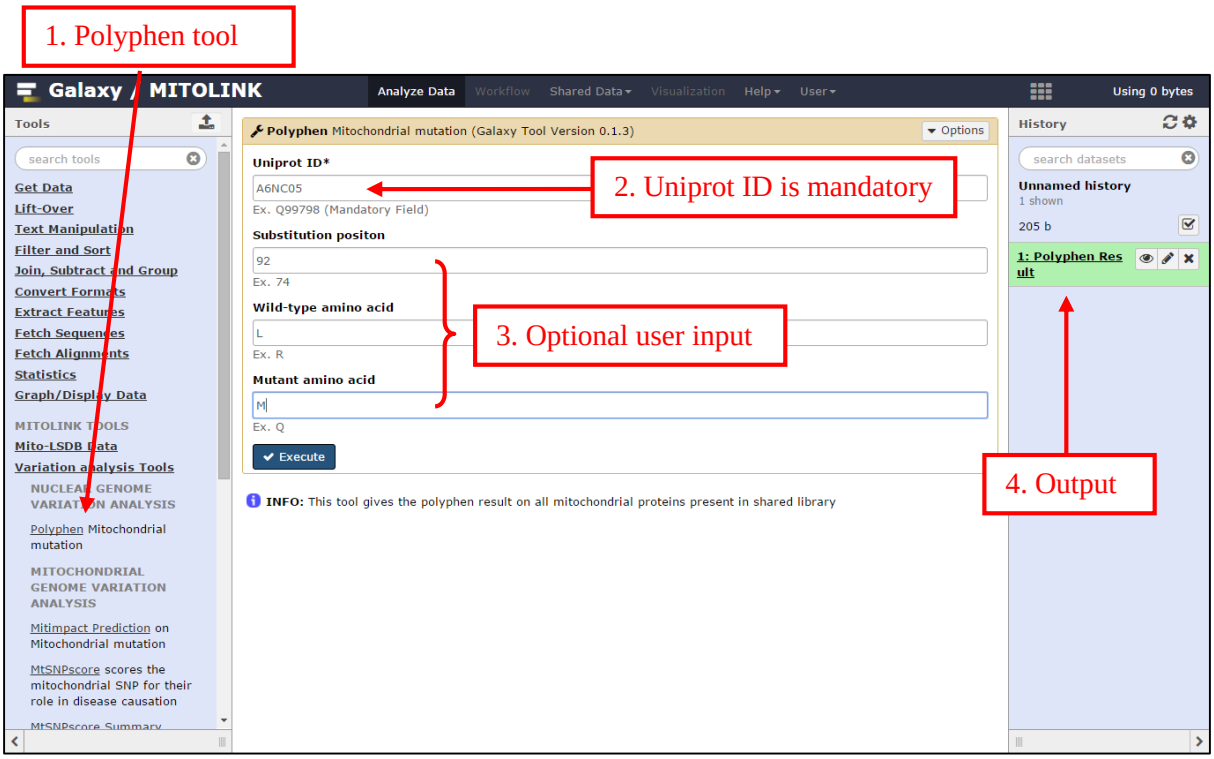

Figure 9: PolyPhen tool view

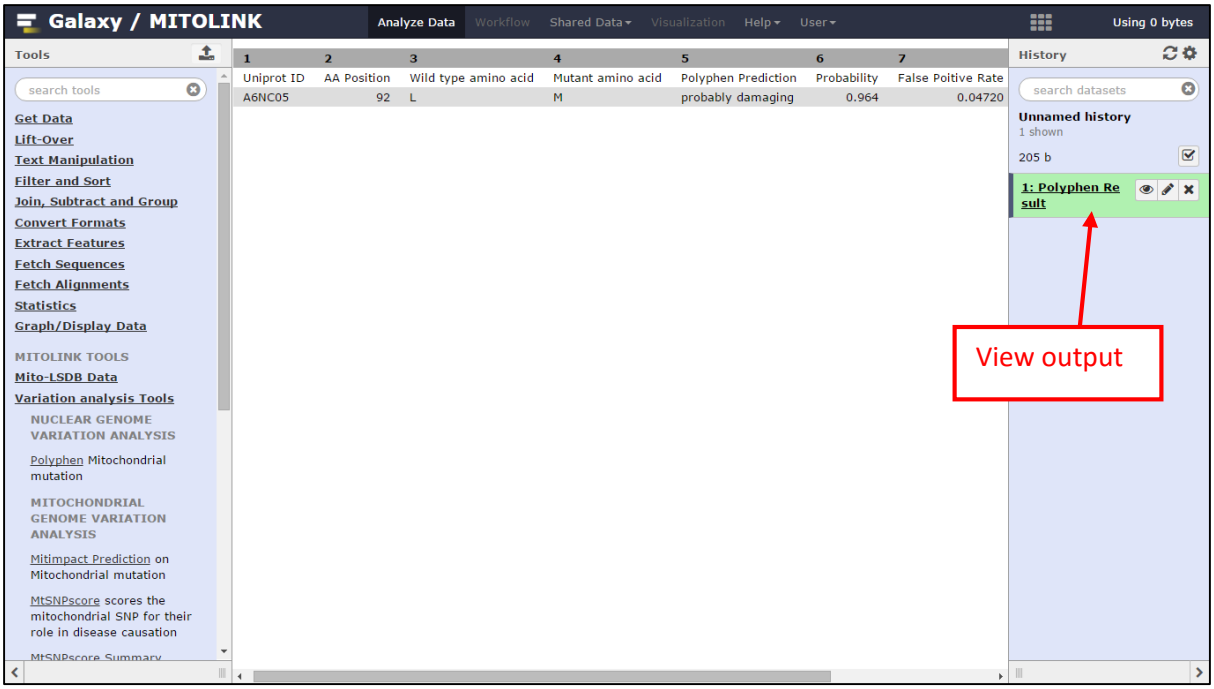

Figure 10: Polyphen output

B. MitImpact Prediction

MitImpact is a mitochondrial variant analysis tool. MitImpact is a collection of pre-computed pathogenicity predictions for all nucleotide changes that cause non-synonymous substitutions in human mitochondrial protein coding genes. It counts 24,115 amino acid variations in all the 13 coding genes of the Homo sapiens mitochondrion (cf. NCBI Reference Sequence: NC_012920.1). The input option for this tool is genomic position. According to the genomic position it show result for nucleotide substitution, and gives score like PolyPhen2, SIFT, FatHmm score, CADD score etc.

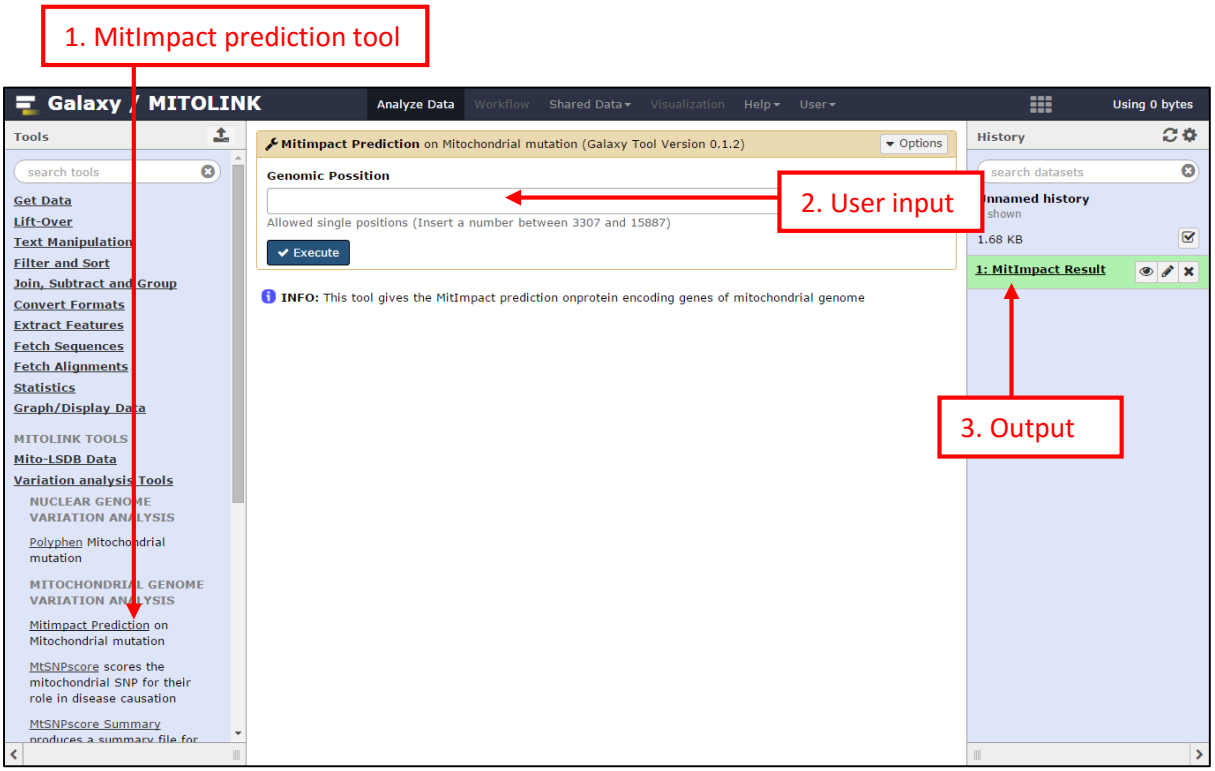

Figure 11: MitImpact prediction tool view

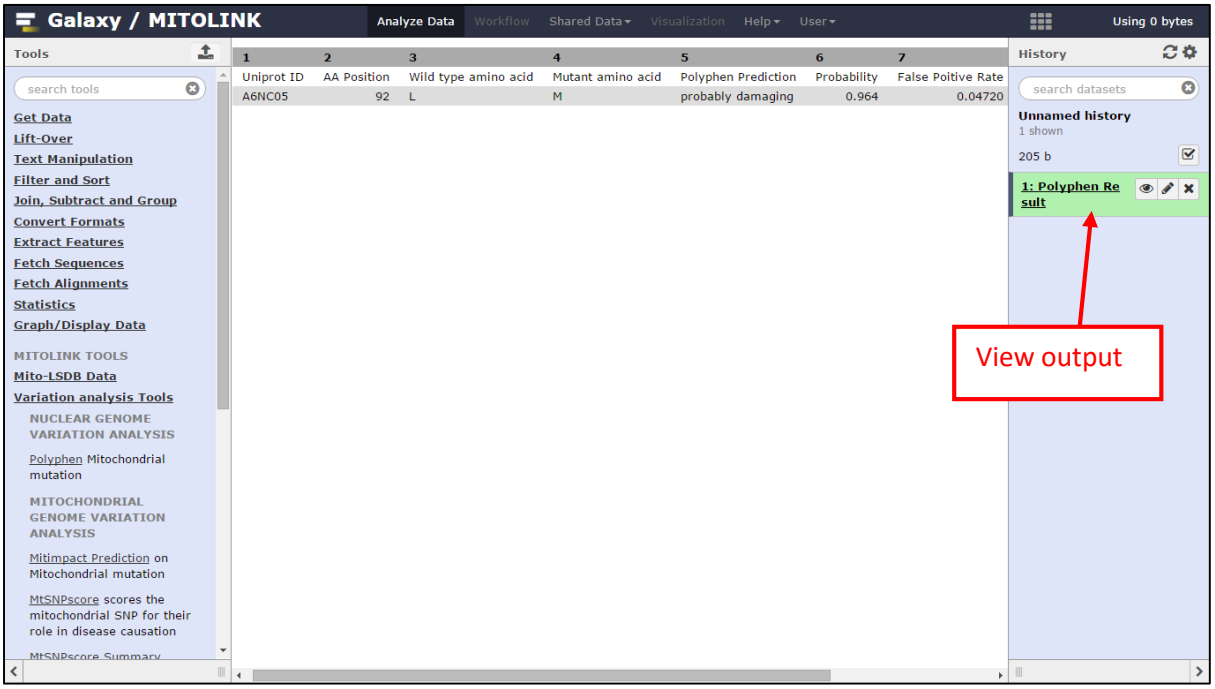

Figure 12: MitImpact prediction output

C. MtSNPscore tool

MtSNPscore is comprehensive weighted scoring system used for identification of mtDNA variations that can impact pathogenicity and would likely be associated with disease. The tool generates five outputs as follows: MtSNPscore summary which summarize the input data. MtSNPscore SNP scores provides detailed analysis result. It consists of Sample ID, SNPsites, Locus, local position, protein position, SIFT tolerated or not etc. MtSNPscore Patients score provide score for patient, MtSNPscore statistics provide general statistics of the output, statistics only reported when rCRS is not selected as Normal Population. MtSNPscore SVM file provide output in SVM readable format, SVM file also only generated when rCRS is not selected as Normal Population

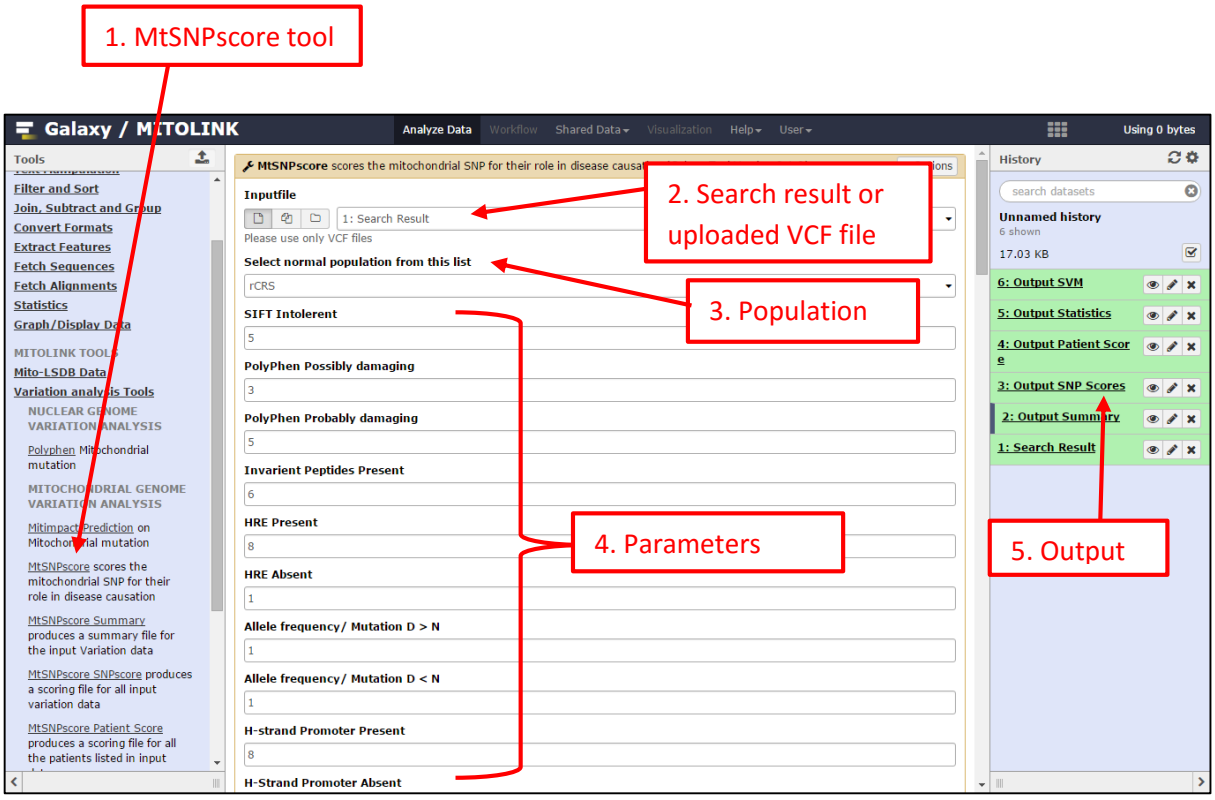

Figure 13: MitSNPscore tool view

4. MBrowse

MBrowse is a combination of database and interactive web pages for displaying annotations on genomes. It is used to display detailed view of genome. The analysis result generated by MitoLSDB, variation analysis tools can be display in MBrowse by importing it in GFF format.

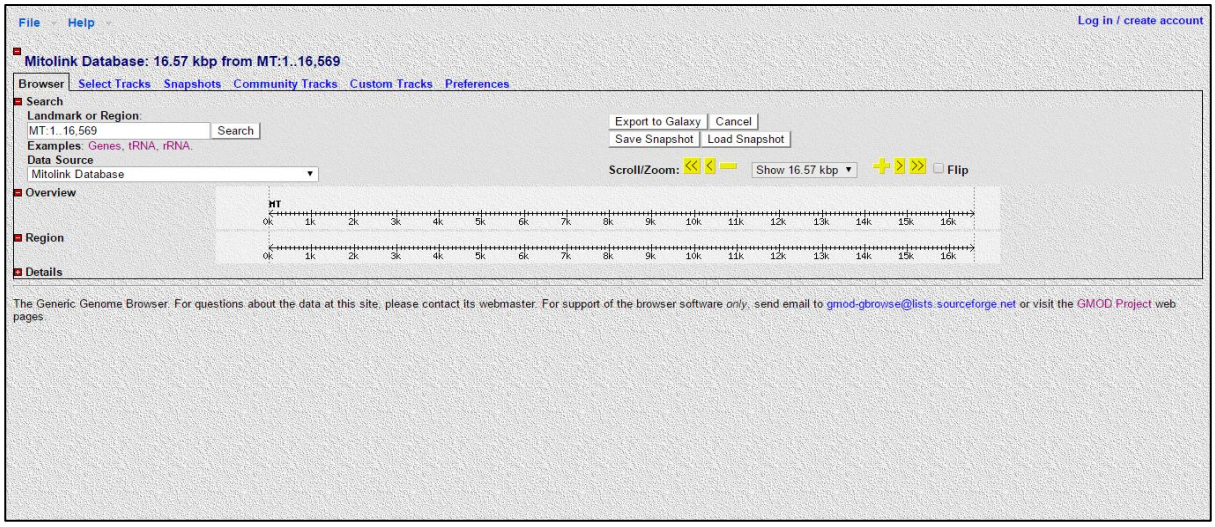

Figure 14: Mbrowse view

For Annotation user have to upload data from Custom track. Data can be upload by entering the data in text box or from URL of by direct file upload option. GBrowse accepts custom track data in a variety of formats including BED, GFF, and GFF3. For dense quantitative data use Wiggle (WIG) format.

Figure 15: Custom track options

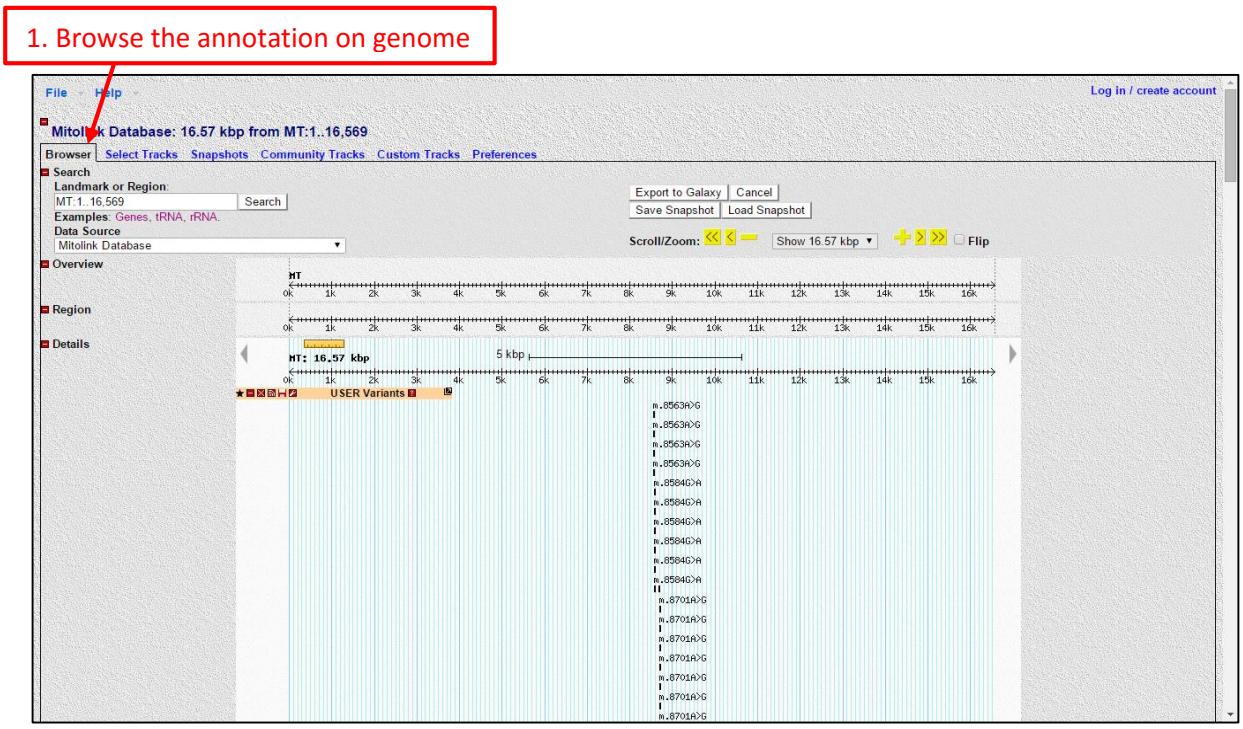

Figure 16 Browse panel view

To select the tracks you would like to have displayed in your MBrowse genome viewer, go to the 'Select Tracks' tab at the top of the MBrowse window and choose your tracks of interest. Then hit the "Back to Browser" links at the top and bottom of the page take you back to the MBrowse genome viewer, which will now have all of the selected tracks displayed.

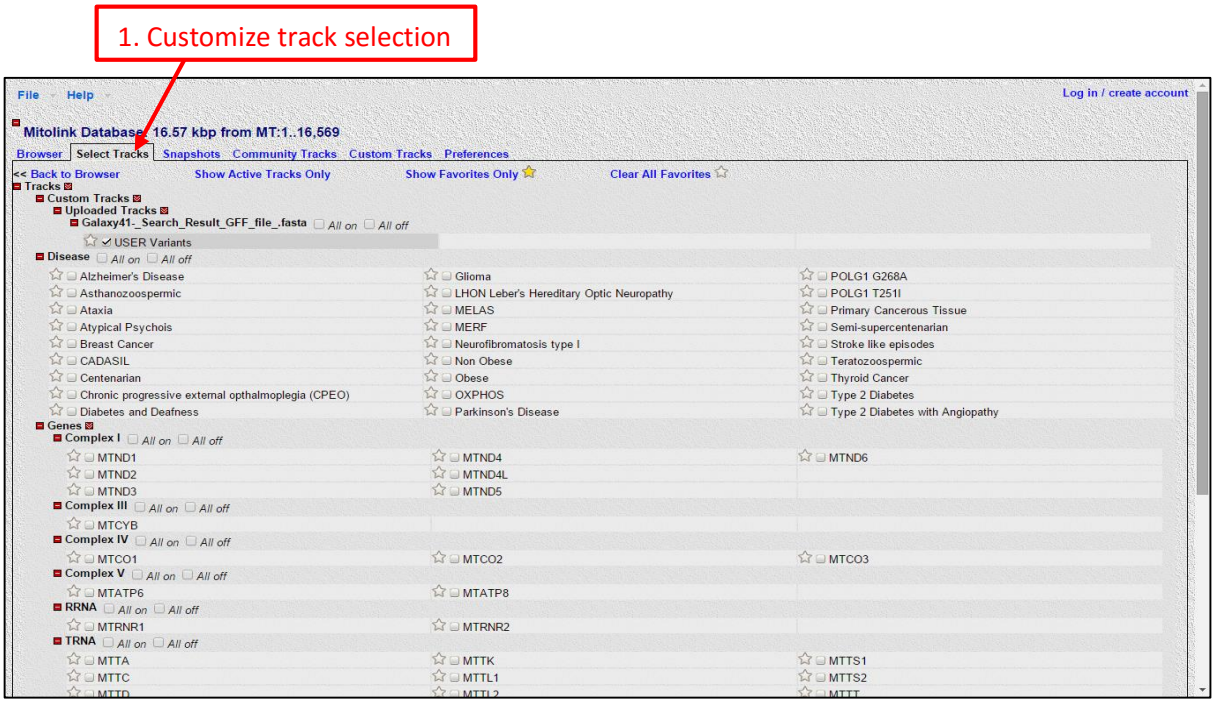

Figure 17: Select track view

5. Converters

A. VCF to GFF

Converts a VCF file to GFF file format. Converted GFF file can be downloaded and used as a custom track on the MBrowse for visualization.

Figure 18: VCF to GFF converter view

B. Search result to GFF

Converts the output of MitoLSDB search into GFF file format. It takes, output of a search on MitoLSDB data as the input. Converted GFF file can be downloaded and used as a custom track on the MBrowse for visualization.

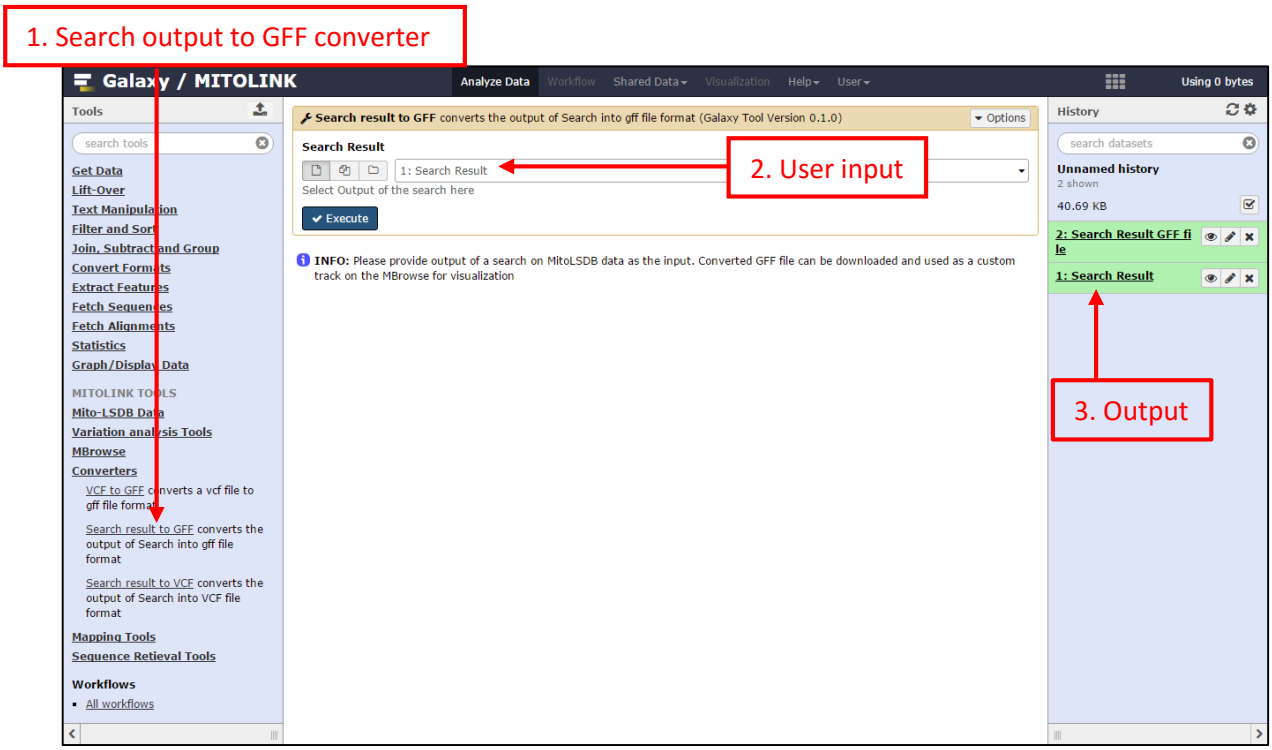

Figure 19: Search result to GFF converter view

C. Search result to VCF

Converts the output of Search into VCF file format. It takes, output of a search on MitoLSDB data as the input. Converted VCF file can be used as input to the MtSNPscore tool for analysis.

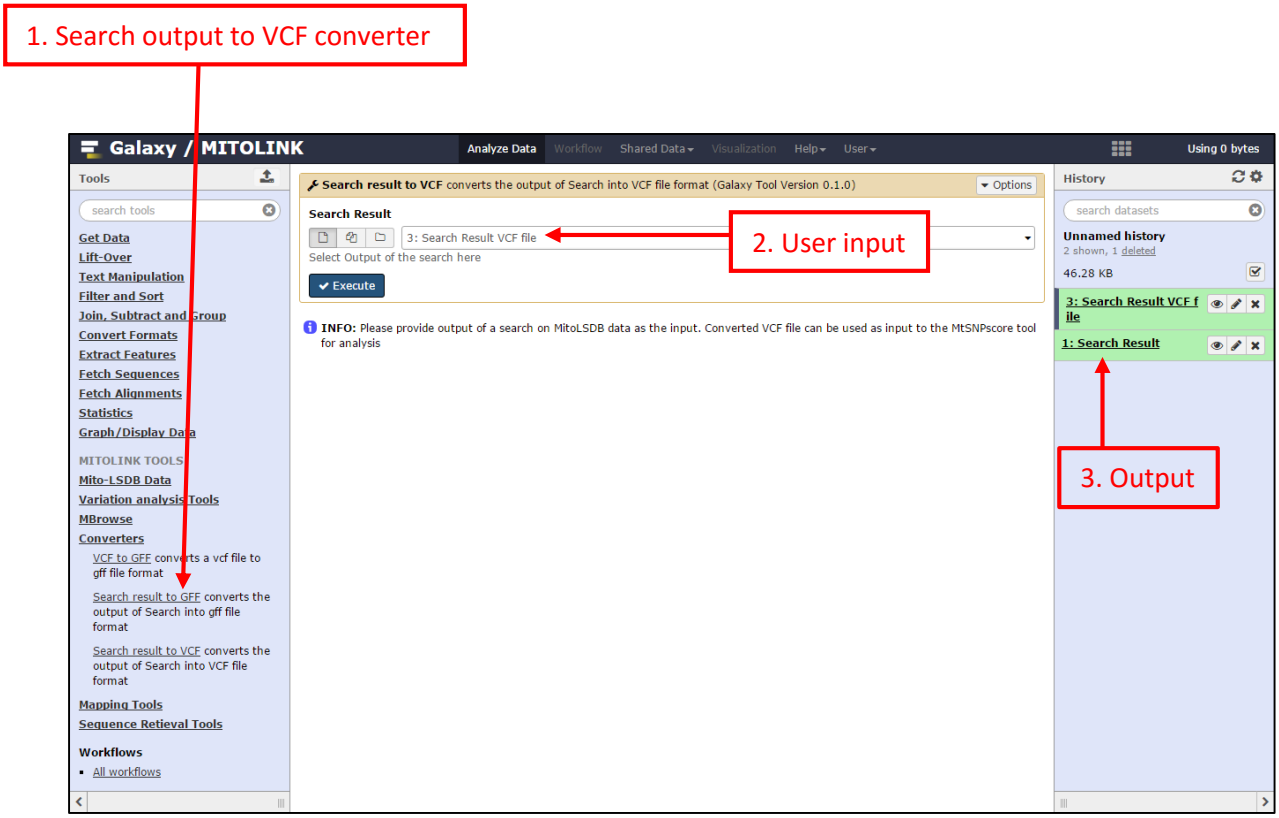

Figure 20: Search result to VCF converter view

6. Mapping tools

Mapping tools is useful for gene search, genomic position from gene location, reference nucleotide search, and mitochondrial protein search.

A. Gene Mapping

Gene mapping tool searches a genomic position from a gene location for mitochondrial protein.

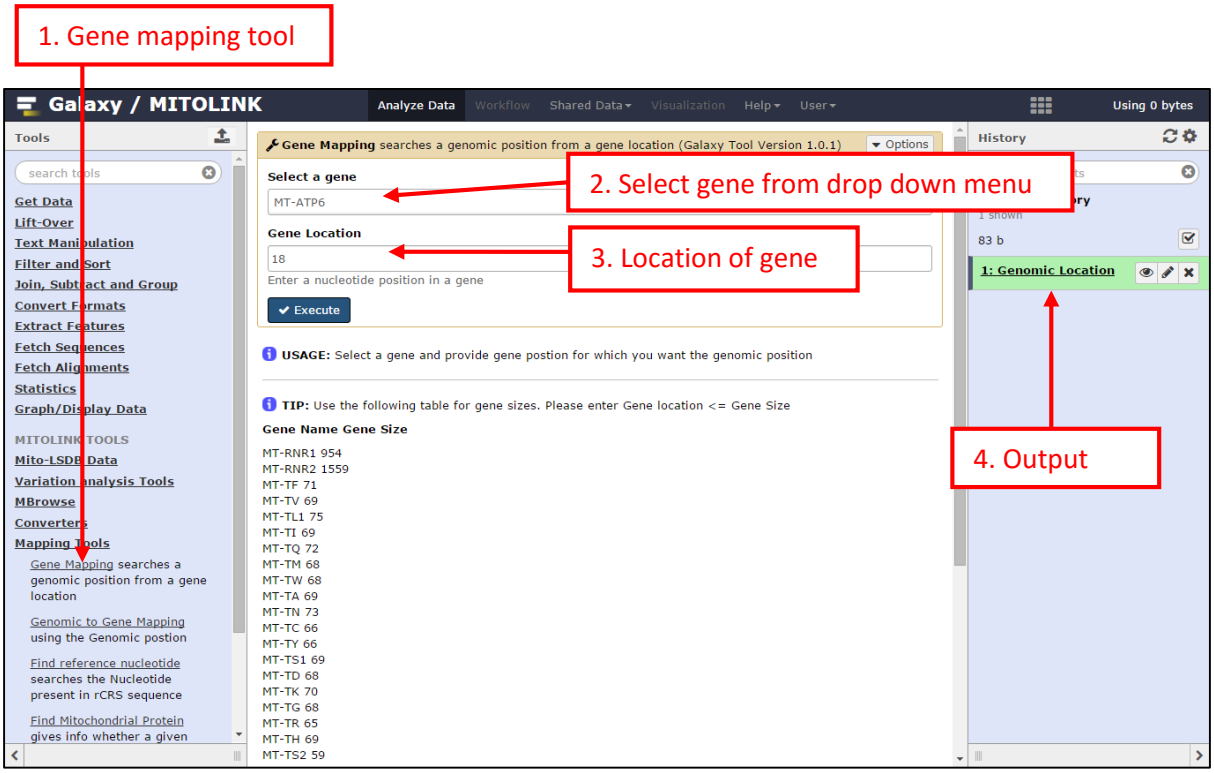

Figure 21: Gene mapping tool view

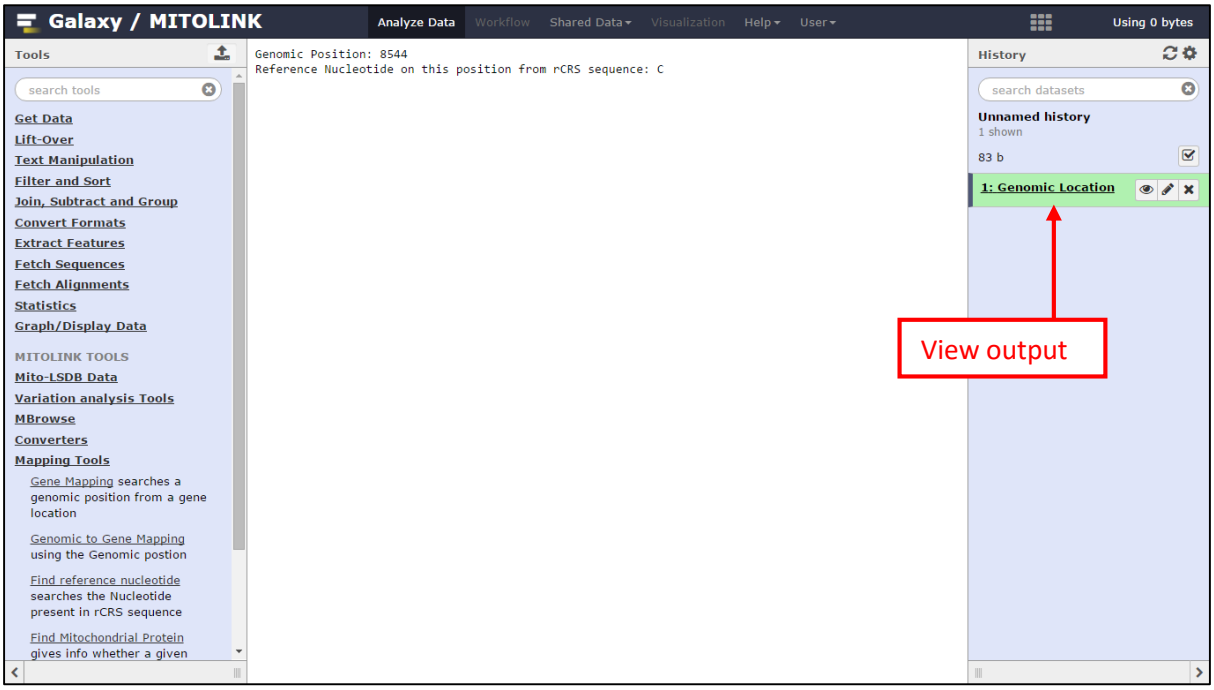

Figure 22: Gene mapping tool output

B. Genomic to Gene Mapping

Genomic to Gene mapping tool uses nucleotide genomic position along with Wild type and Mutated nucleotide in HGVS format and tool will display its gene location, gene name and other information.

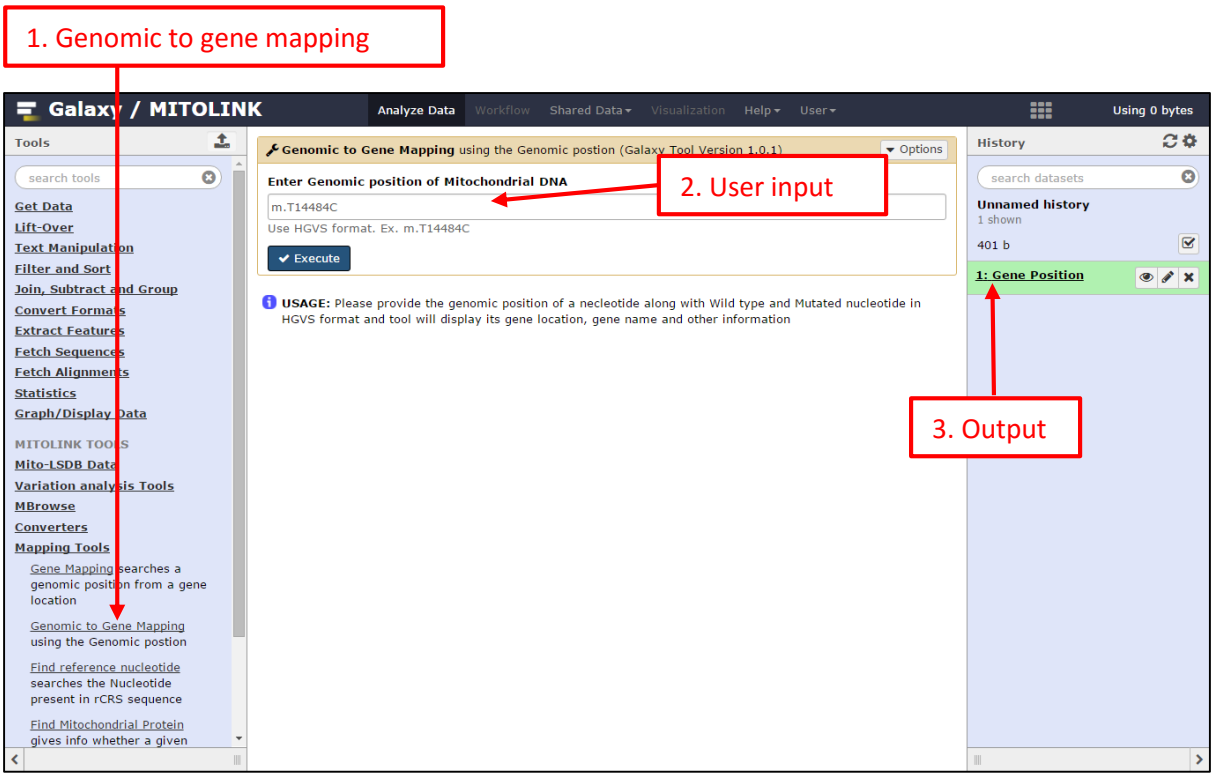

Figure 23: Genomic to gene mapping tool view

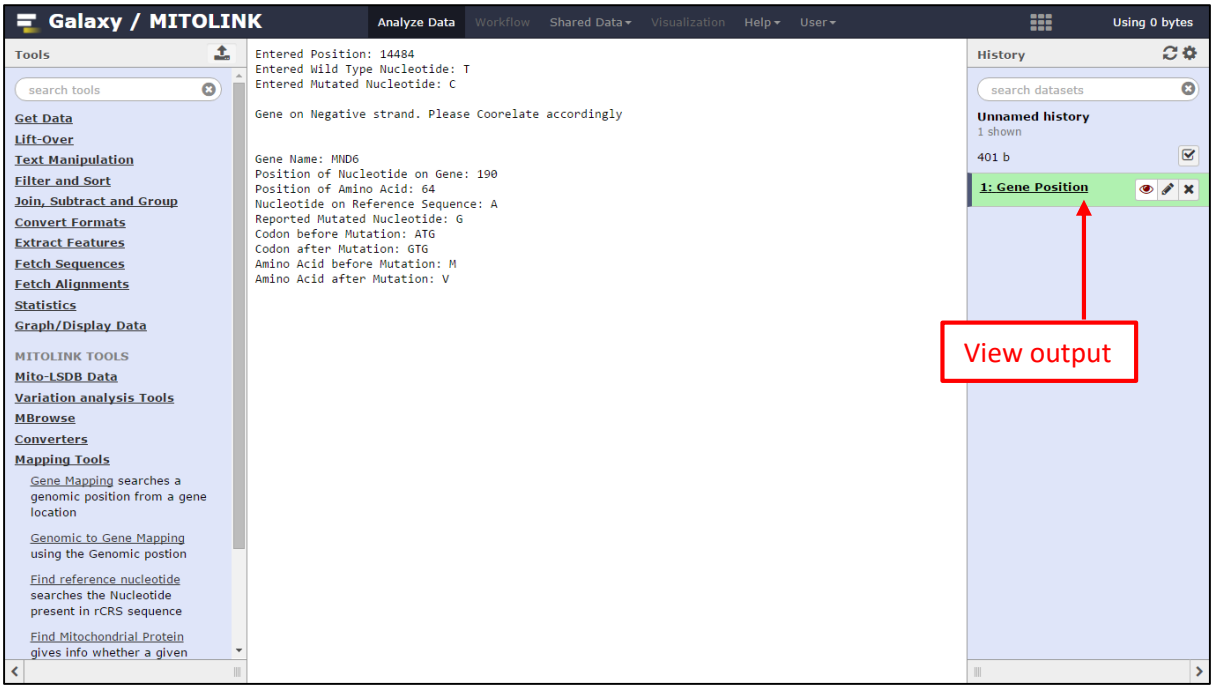

Figure 24: Genomic to gene mapping tool output

C. Reference Nucleotide search

Find reference nucleotide tool searches the Nucleotide present in rCRS sequence. It takes input as genomic position and returns the Nucleotide present at that position in Reference (rCRS) genome

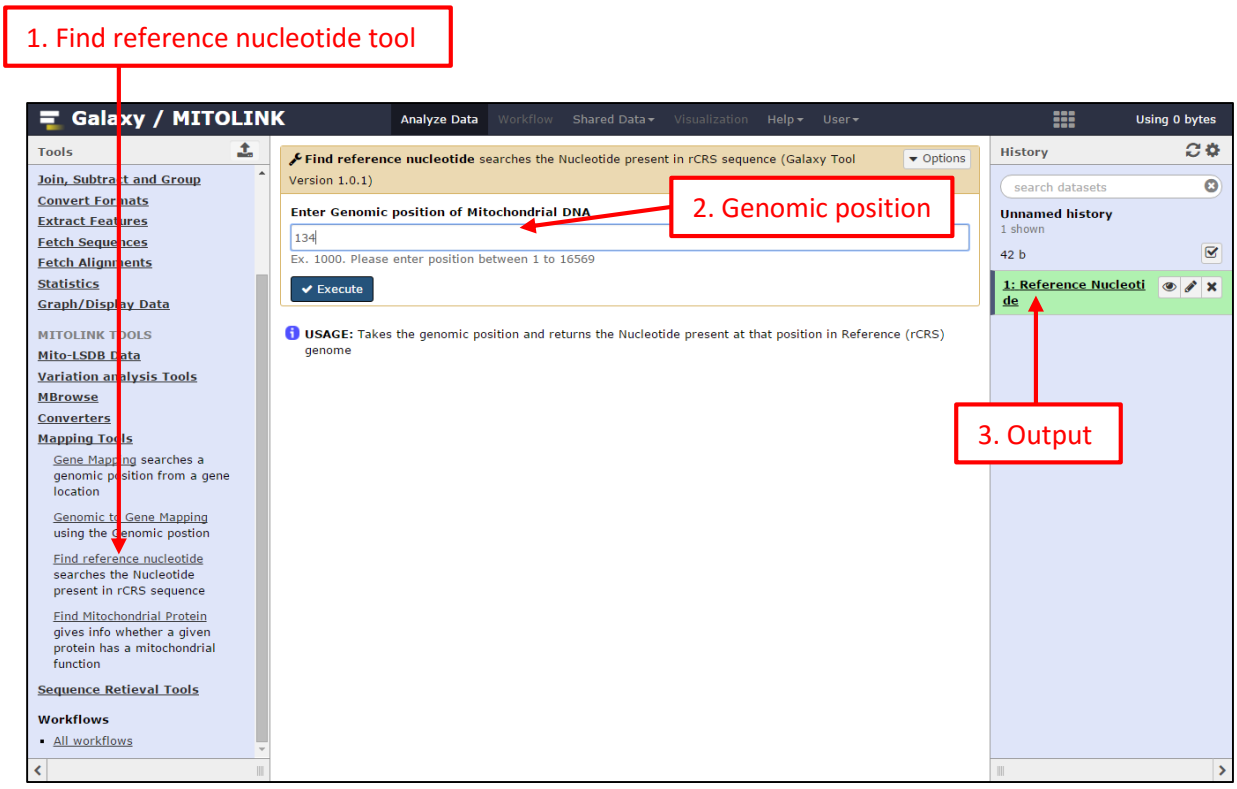

Figure 25: Reference nucleotide searching tool

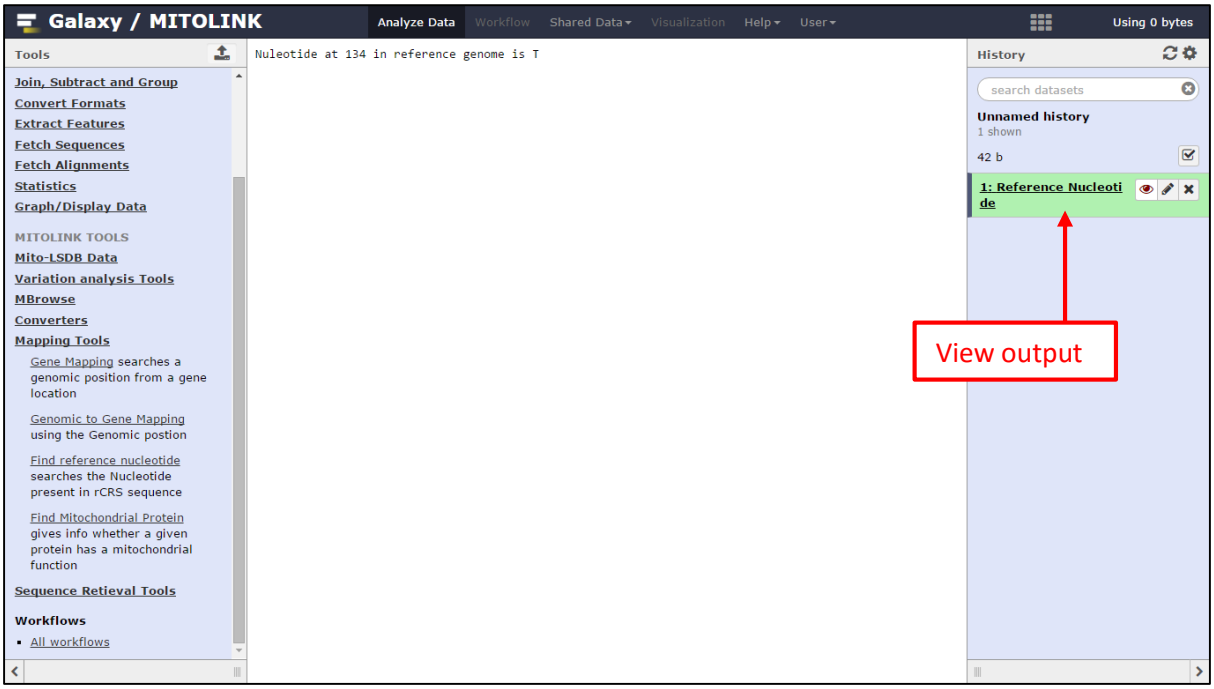

Figure 26: Reference nucleotide search output

D. Find Mitochondrial Protein

The tool searches the entered protein with in-house mitochondrial protein library and gives the various evidence for mitochondrial function along with various other information.

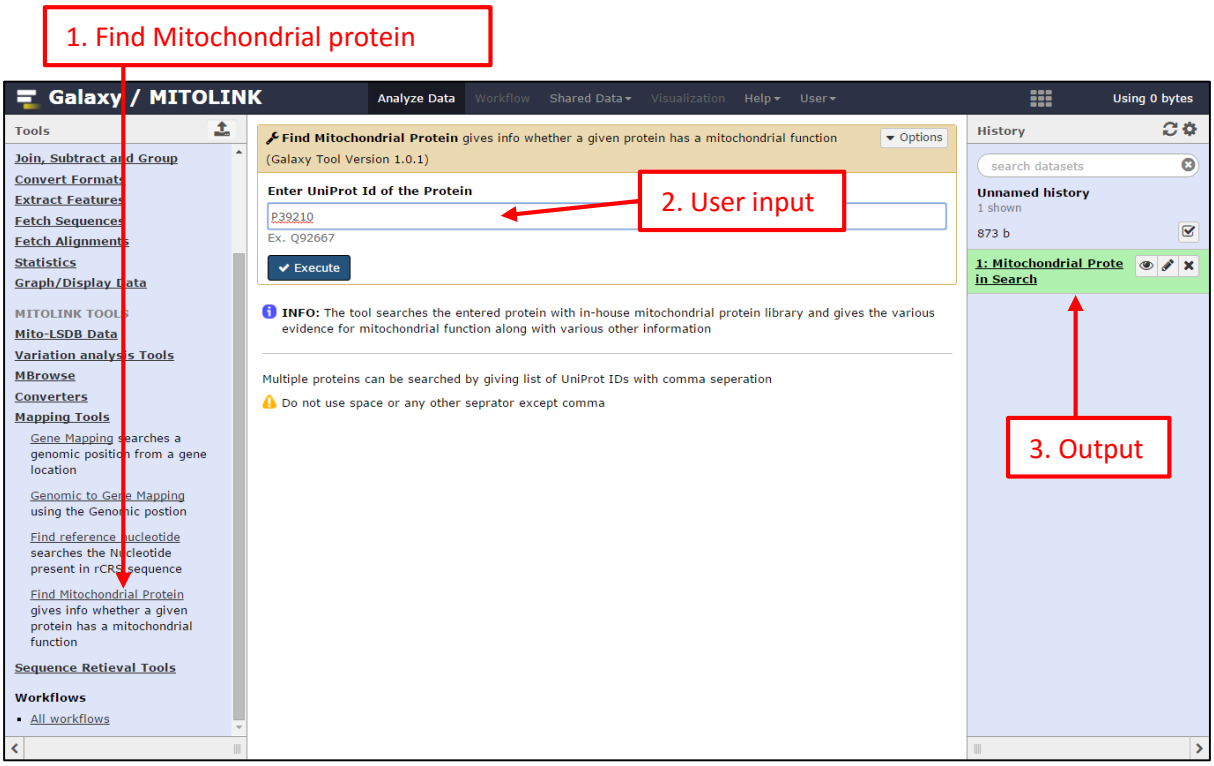

Figure 27: Mitochondrial protein search tool view

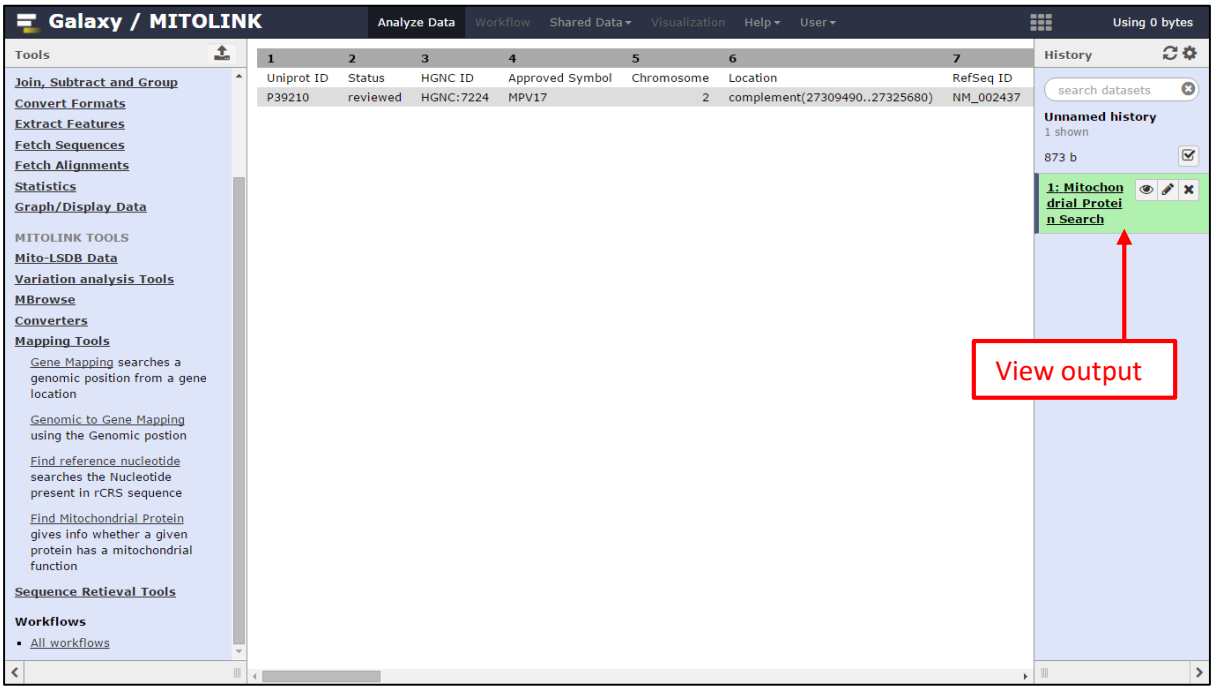

Figure 28: Mitochondrial protein search output

7. Sequence Retieval Tools

A. Gene sequence retreival

The tool will return the gene sequence from reference (rCRS) genome for selected gene.

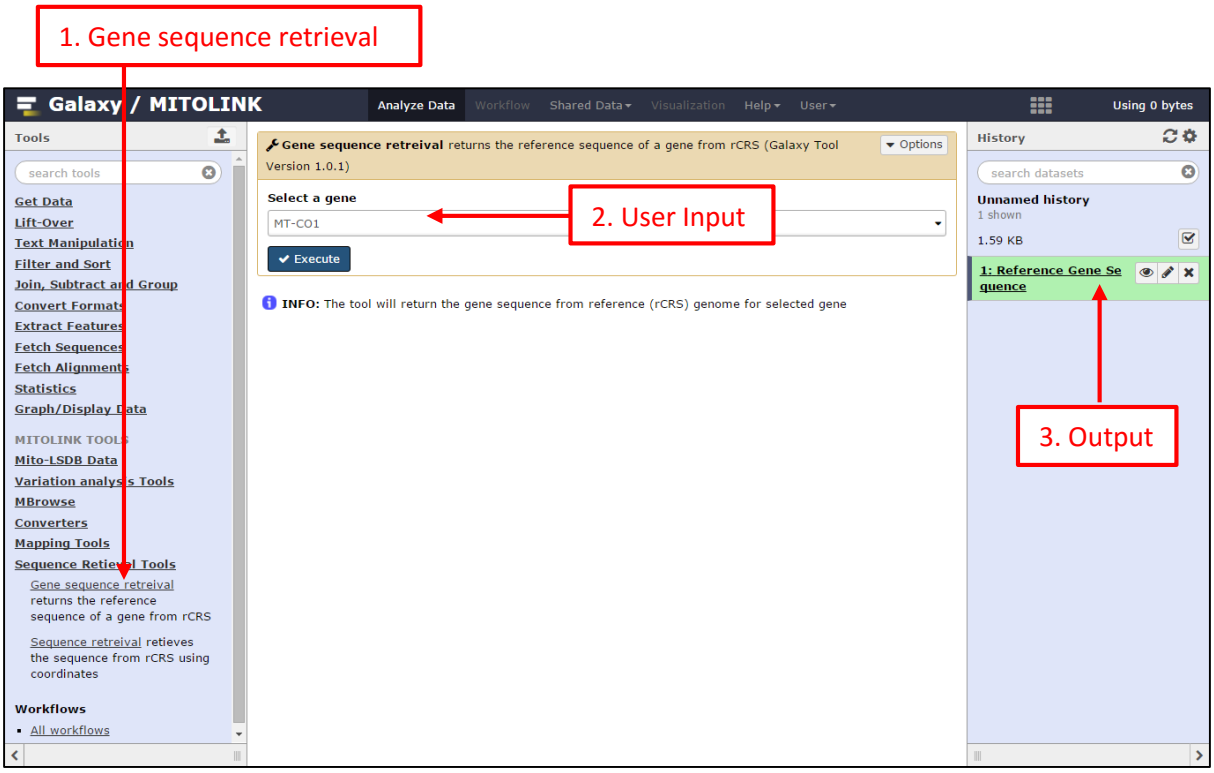

Figure 29 Gene sequence retrieval tool

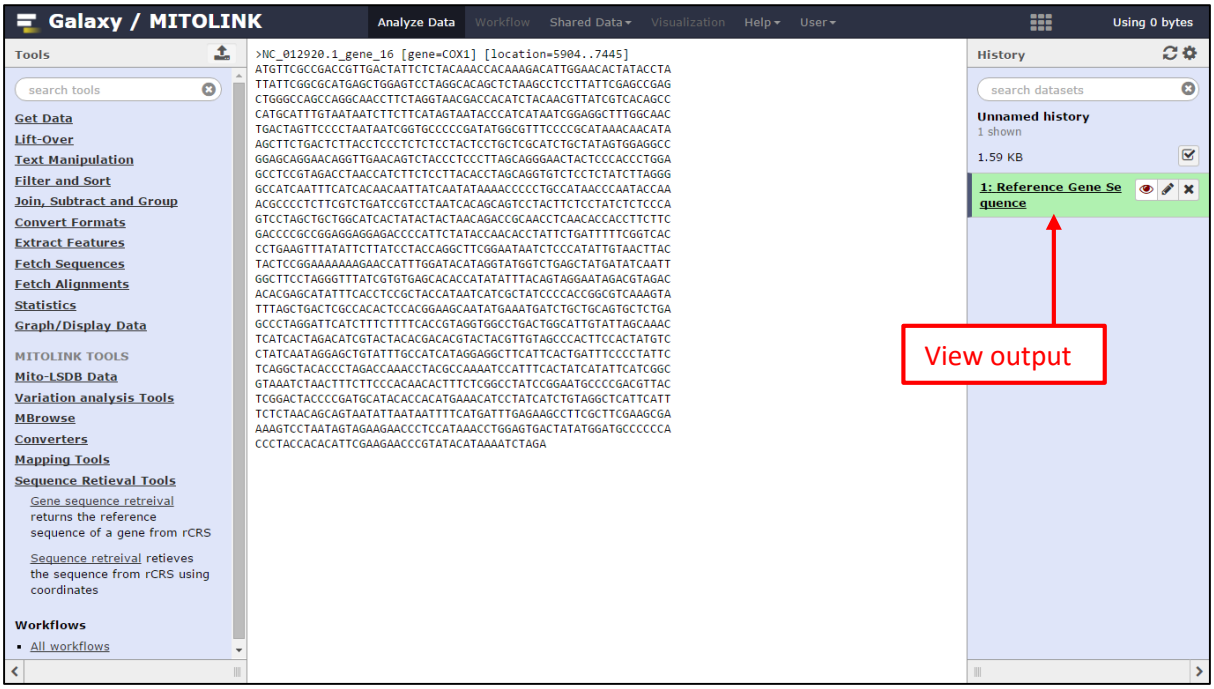

Figure 30: Gene sequence retrieval output

B. Sequence retrieval

The tool will retrieve the sequence starting and ending with the given coordinates from reference (rCRS) genome

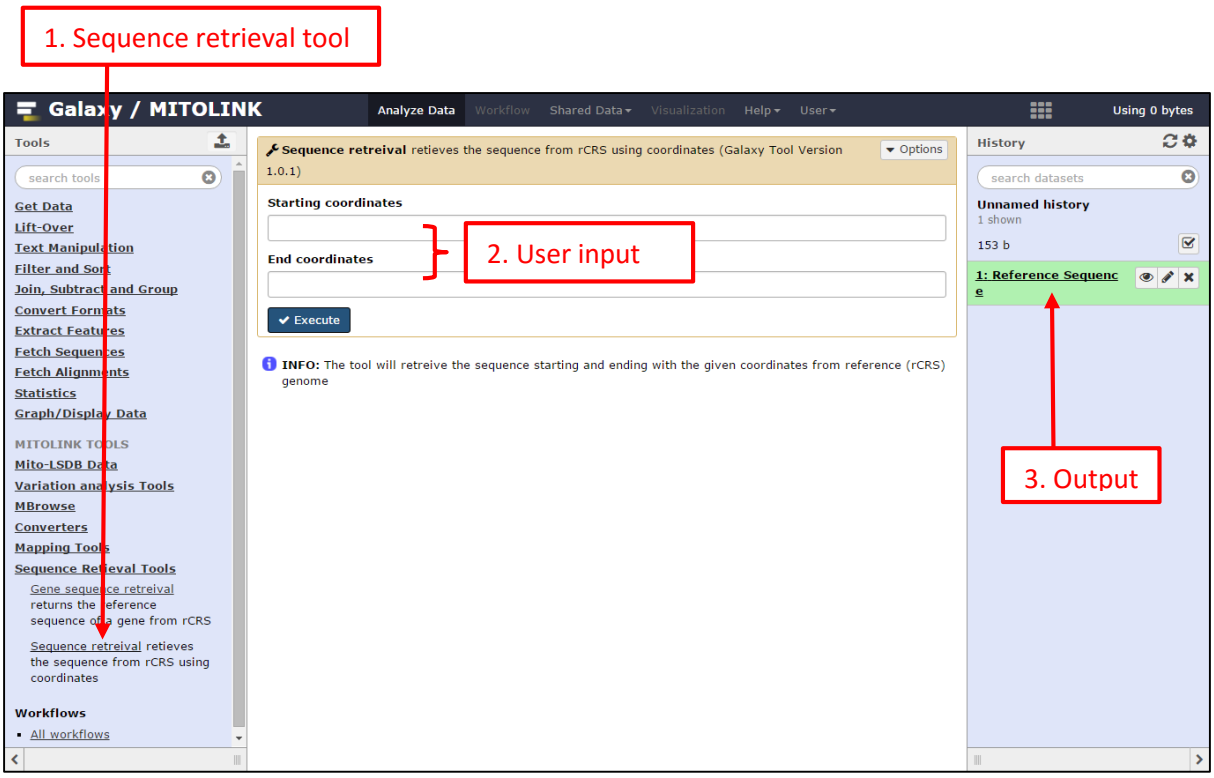

Figure 31: Sequence retrieval tool view

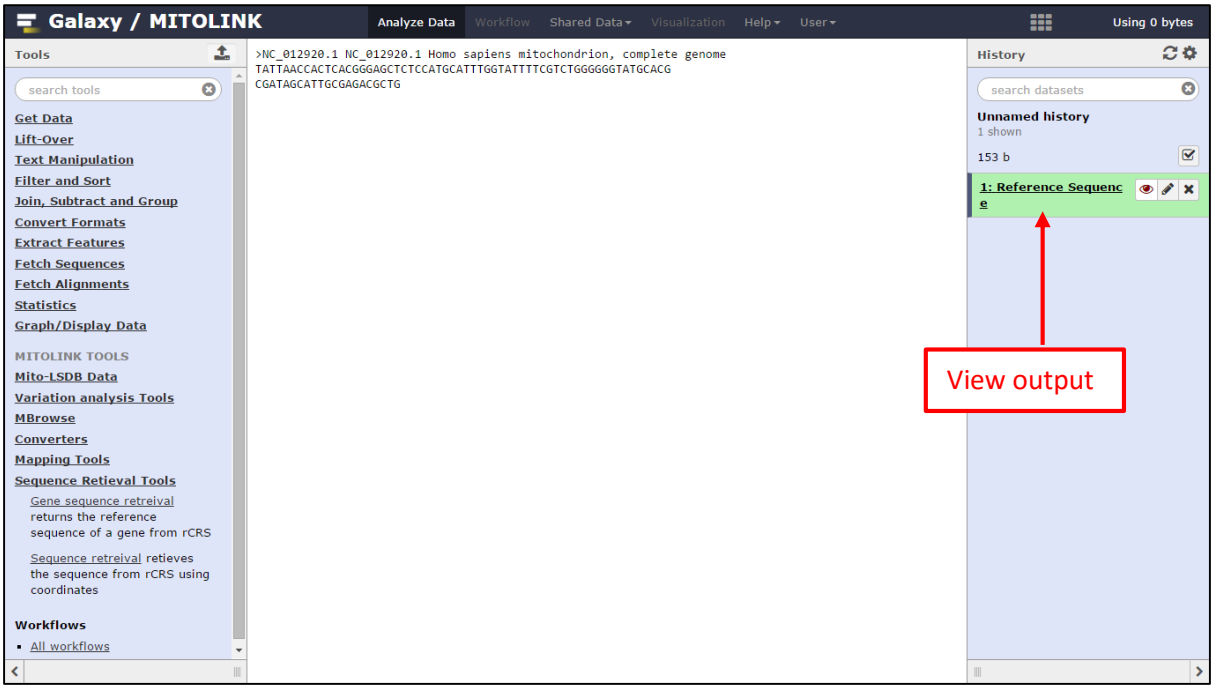

Figure 32: Sequence retrieval output

III. Detail Panel

This panel displays the interface of all the tools along with Input Parameters required to run a tool. It also provides help and examples to run a tool. This panel also displays the Output of a tool after its execution when user clicks on the eye  icon show in **History Panel**.

IV. History Panel

When data is uploaded from your computer or analysis is done on existing data using Galaxy, each output from those steps generates a dataset. These datasets (and the output datasets from later analysis on them) are stored by Galaxy in **Histories**.
Users that have registered an account and logged in can have many histories and the history panel allows switching between them and creating new ones.

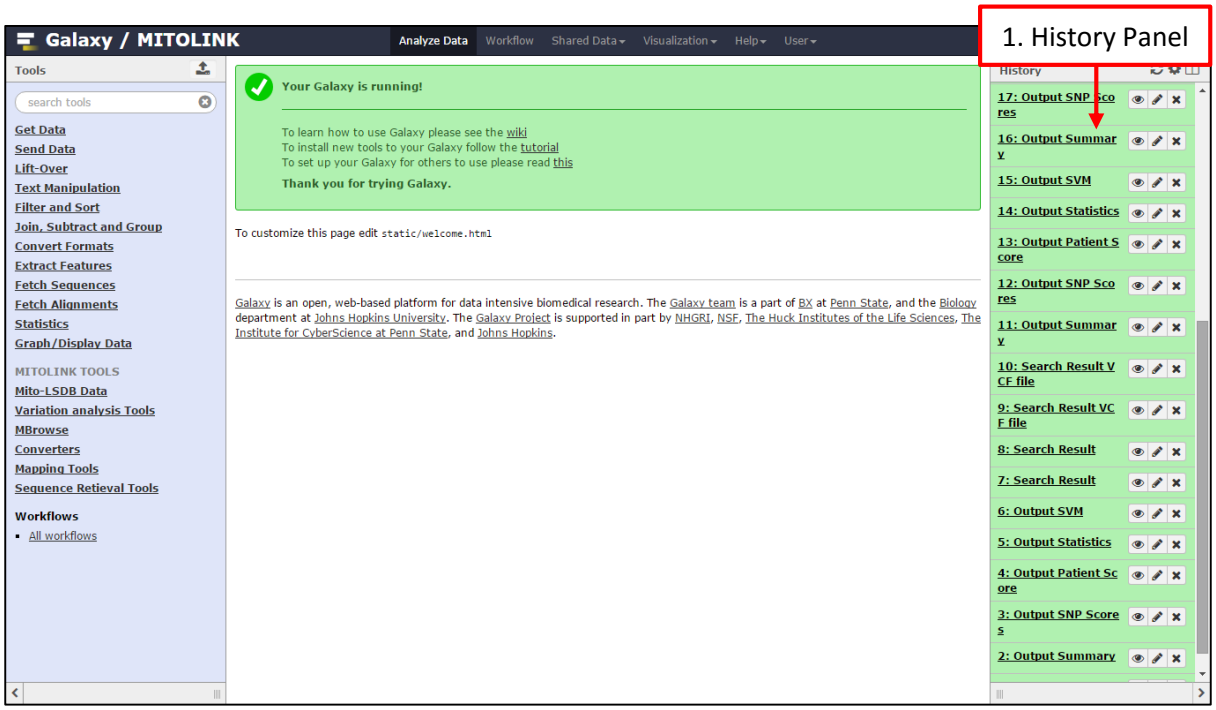

Figure 33: History Panel view