PanGraphRNA User Manual

(version 1.0)

- PanGraphRNA is an efficient, flexible and web-based Galaxy platform that can be easily used to construct graph pangenomes from genetic variations at individual, subpopulation, and population levels. It can assist researchers to select appropriate graph pangenomes using various performance metrics for both real and simulation experiments.
- Currently, PanGraphRNA is composed of four functional modules: **Graph Pangenome Preparation Module, Construction Module, Evaluation Module, and Application Moudule**.
- PanGraphRNA was powered with an advanced packaging technology, which enables compatibility and portability.
- PanGraphRNA project is hosted on https://github.com/cma2015/PanGraphRNA
- PanGraphRNA docker image is available at https://hub.docker.com/r/malab/pangraphrna

Graph Pangenome Application Module

This module performs several applications of graph pangenomes, including differential expression analysis and expression-based QTL analysis.

Tools	Description	Input	Output	Time (test data)	Reference
Expression Quantification	Perform expression quantification for read-genome alignment result	Alignment result in BAM format and annotation file in GTF format	Gene read count matrix in TXT format and gene expression quantification matrix in TXT format	~5 mins	StringTie
Differential Expression Analysis	Perform differential expression analysis	Gene read count matrix in CSV format	Information of differentially expressed genes in CSV format	~5 mins	DESeq2

Tools	Description	Input	Output	Time (test data)	Reference
Quantitative Trait Locus Identification	Perform identification of expression quantitative trait locus	Variation files and gene expression matrix in TXT format	Identification of expression quantitative trait locus in RData and TXT format	~10 mins	MatrixEQTL

Expression Quantification

In this function, an fast and highly efficient tool **StringTie** (Pertea, M., *et al*., 2015) is used for gene expression quantification. See https://github.com/gpertea/stringtie for details.

Input

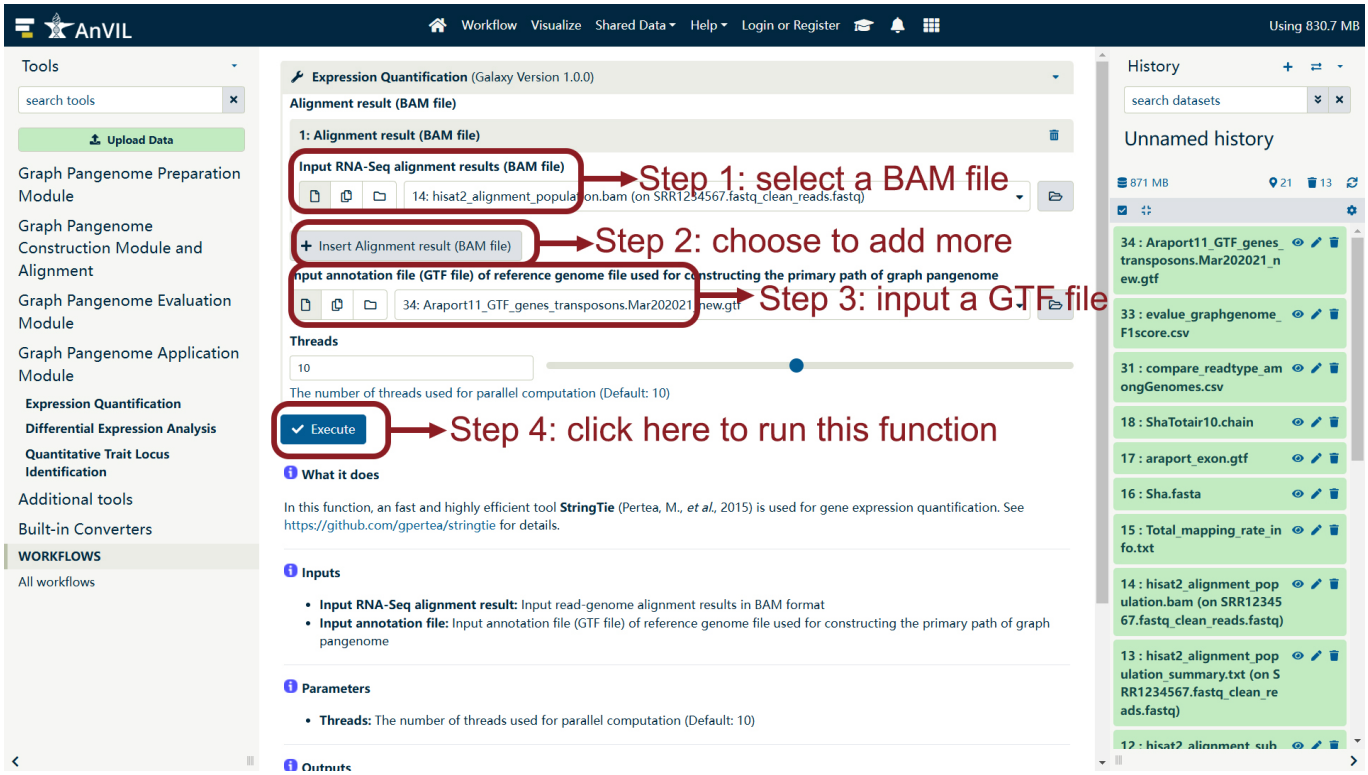
- **Input RNA-Seq alignment result:** Input read-genome alignment results in BAM format
- **Input annotation file:** Input annotation file (GTF file) of reference genome file used for constructing the primary path of graph pangenome

Parameters

- **Threads:** The number of threads used for parallel computation (Default: 10)

Output

- **Gene read count matrix in TXT format**
- **Gene expression quantification matrix in TXT format**



Differential Expression Analysis

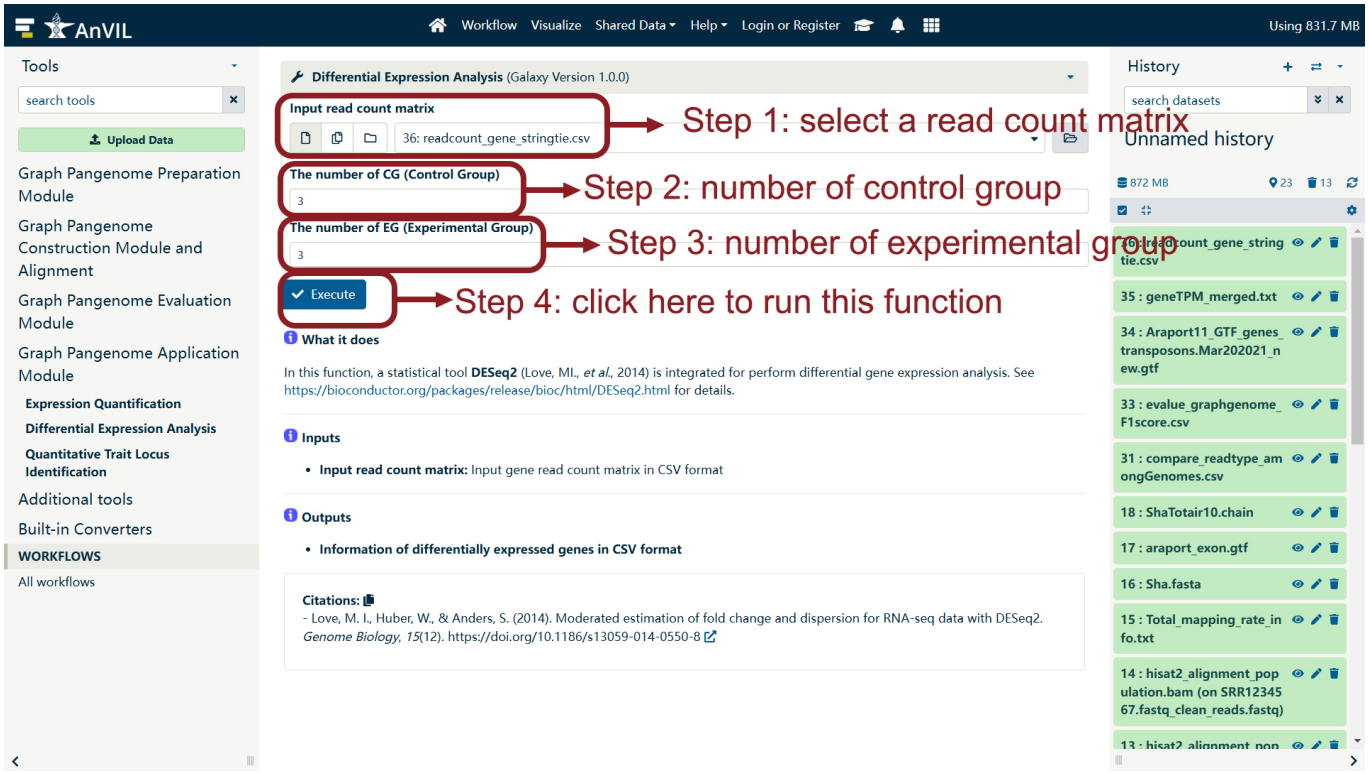
In this function, a statistical tool **DESeq2** (Love, MI., *et al.*, 2014) is integrated for perform differential gene expression analysis.
See https://bioconductor.org/packages/release/bioc/html/DESeq2.html for details.

Input

- **Input read count matrix:** Input gene read count matrix in CSV format

Output

- **Information of differentially expressed genes in CSV format**



Quantitative Trait Locus Identification

In this function, a fast eQTL analysis tool **MatrixEQTL** (Shabalin, AA., *et al.*, 2012) is integrated for perform identification of expression quantitative trait locus.
See http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL for details.

Input

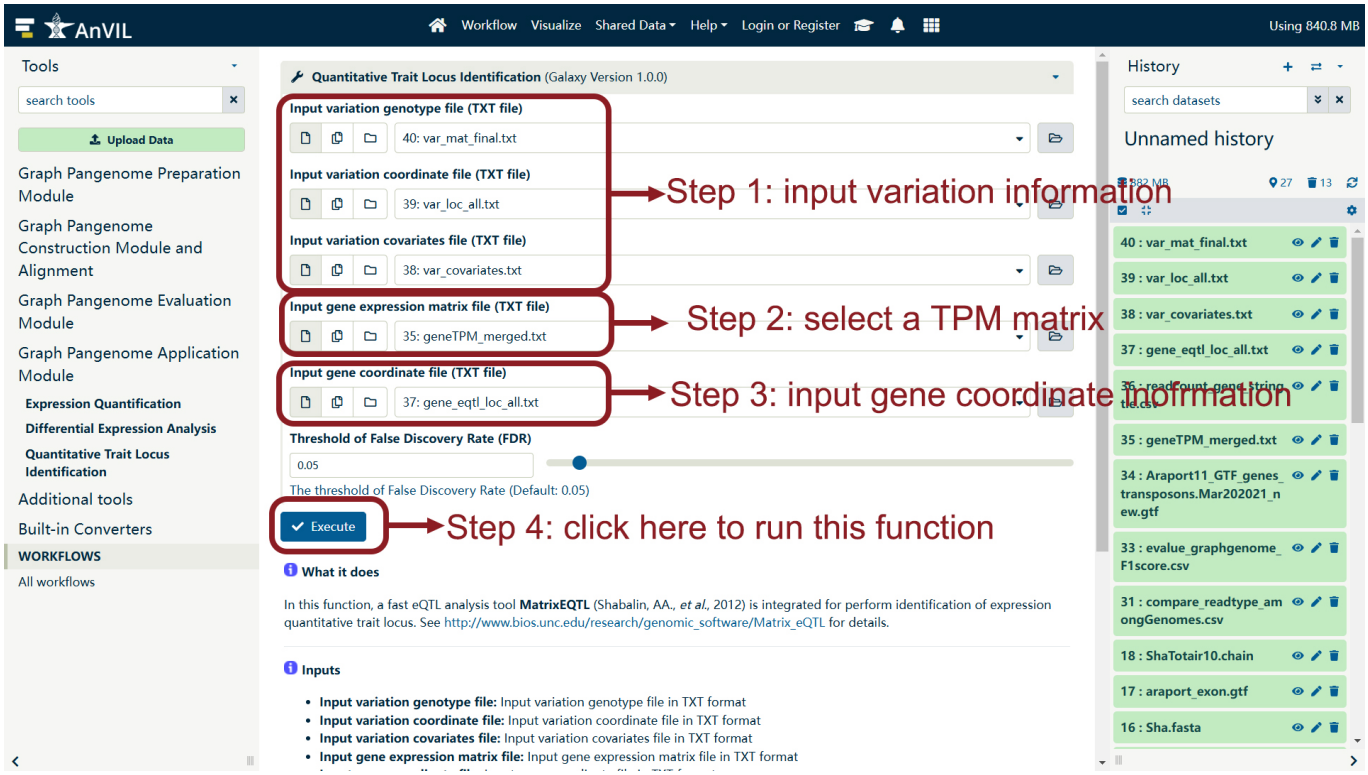
- **Input variation genotype file:** Input variation genotype file in TXT format
- **Input variation coordinate file:** Input variation coordinate file in TXT format
- **Input variation covariates file:** Input variation covariates file in TXT format
- **Input gene expression matrix file:** Input gene expression matrix file in TXT format
- **Input gene coordinate file:** Input gene coordinate file in TXT format

Parameters

- **Threshold of False Discovery Rate:** Results exceeding the False Discovery Rate threshold (default: 0.05) will be filtered out

Output

- **Identification of expression quantitative trait locus in RData and TXT format**


AnVIL
Workflow
Visualize
Shared Data
Help
Login or Register
Using 840.8 MB
Tools
search tools
Upload Data
Graph Pangenome Preparation Module
Graph Pangenome Construction Module and Alignment
Graph Pangenome Evaluation Module
Graph Pangenome Application Module
Expression Quantification
Differential Expression Analysis
Quantitative Trait Locus Identification
Additional tools
Built-in Converters
WORKFLOWS
All workflows
Quantitative Trait Locus Identification (Galaxy Version 1.0.0)
Input variation genotype file (TXT file)
40: var_mat_final.txt
Input variation coordinate file (TXT file)
39: var_loc_all.txt
Input variation covariates file (TXT file)
38: var_covariates.txt
Input gene expression matrix file (TXT file)
35: geneTPM_merged.txt
Input gene coordinate file (TXT file)
37: gene_eqtl_loc_all.txt
Threshold of False Discovery Rate (FDR)
0.05
The threshold of False Discovery Rate (Default: 0.05)
Execute
Step 1: input variation information
Step 2: select a TPM matrix
Step 3: input gene coordinate information
Step 4: click here to run this function
What it does
In this function, a fast eQTL analysis tool MatrixEQTL (Shabalin, AA., et al., 2012) is integrated for perform identification of expression quantitative trait locus. See http://www.bios.unc.edu/research/genomic_software/Matrix_eQTL for details.
Inputs
Input variation genotype file: Input variation genotype file in TXT format
Input variation coordinate file: Input variation coordinate file in TXT format
Input variation covariates file: Input variation covariates file in TXT format
Input gene expression matrix file: Input gene expression matrix file in TXT format
History
search datasets
Unnamed history
582 MB
27
13
40 : var_mat_final.txt
39 : var_loc_all.txt
38 : var_covariates.txt
37 : gene_eqtl_loc_all.txt
35 : geneTPM_merged.txt
34 : Araport11_GTF_genes_transposons.Mar202021_new.gtf
33 : evalue_graphgenome_F1score.csv
31 : compare_readtype_amongGenomes.csv
18 : ShaTotair10.chain
17 : araport_exon.gtf
16 : Sha.fasta