

VEuPathDB RNA-Seq BigWig & expression value export & integration tool

Note: using VEuPathDB *My Datasets* and the VEuPathDB Galaxy workspace requires you to log in to the site since the data is stored privately under your account.

Currently this tool works on workflows that use Cufflinks to generate expression values. It is recommended that all data files used are grouped using the collections method in Galaxy. For more information about collections please see the [Galaxy Collections help pages](#). An example workflow is available on the VEuPathDB Galaxy front page for you to use and try out.

Once a workflow has run successfully you can export and integrate your data privately into VEuPathDB by following these steps:

1. Click on the “VEuPathDB Export Tools” menu item in the left-hand panel, then click on the “RNA-Seq to VEuPathDB” export tool.

The screenshot shows the VEuPathDB Galaxy interface. The left-hand panel contains a 'Tools' section with a search bar and a list of 'VEUPATHDB APPLICATIONS'. The 'VEuPathDB Export Tools' item is highlighted with a red box and a red arrow. Below it, the 'RNA-Seq to VEuPathDB Export an RNA-Seq result to VEuPathDB' tool is listed. The main panel displays the 'Welcome to the VEuPathDB Galaxy Site' message, a list of 'With VEuPathDB Galaxy you can:' actions, and a 'Get started with VEuPathDB pre-configured workflows:' section. The right-hand panel shows a 'History' section with a list of datasets, including 'Sexual differentiation' and 'Cufflinks on collection 130: total map mass'.

2. The RNA-Seq export tool parameters will appear in the middle section of the Galaxy page. Fill out the parameters completely:
 - a. Enter a meaningful name for your dataset (you can always modify this).
 - b. Select whether your RNA-Seq dataset is strand specific or not.
 - c. Select the BigWig collection from the list.
 - d. Select the Cufflinks **gene** expression collection.
 - e. Enter a summary for your dataset (you can always modify this).
 - f. Enter a description for your dataset (you can always modify this).
 - g. Click on the “Execute” button.

RNA-Seq to VEuPathDB Export an RNA-Seq result to VEuPathDB (Galaxy Version 1.0.0) Options

My Data Set name:
 A
 specify a name for the new dataset

Is your dataset strand-specific?
 B
 Is this a strand-specific dataset?

BigWig collection:
 C
 Select the BigWig collection to include in the new VEuPathDB My Data Set. The bigwig collection you select here must be mapped to the reference genome that you select below.

FPKM collection:
 D
 Select the FPKM collection. For an unstranded dataset, its name should include the phrase 'gene expression'.

My Data Set summary:
 E

My Data Set description:
 F

G

3. After you click on the execute button a new step is initiated in the right hand history pannel: ExportToVEuPathDBInfo.html. The step will initially be grey, then yellow when running and green when completed. This should not take a long time.

✓ 1 job has been successfully added to the queue – resulting in the following datasets:

92: exportToEuPathDBInfo.html

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

The History panel shows the job '92: exportToEuPathDBInfo.html' in three states: grey (queued), yellow (running), and green (finished). Red arrows indicate the progression from grey to yellow to green.

- Once the export is complete, go to the VEuPathDB site where your data will be visualized. For example, if your data is *Plasmodium berghei* RNA-Seq data go to PlasmoDB.org
- Click on the “My Workspace” menu in the header then select the *My data sets* link.

- The “My Datasets” page includes a list of all datasets that you have exported to VEuPathDB with the most recent export at the top of the list.

My Data Sets

Search Datasets Showing 1 of 1 data set ☐ Only show data sets related to ToxoDB 295.20 M (0.02%) of 10.00 G used

☐	Name / ID	Summary	Type	EuPathDB Websites	Status	Owner	Created	File Count	Size	Quota Usage
☐	Experiment A1 (12640)	Experiment A1	RNASeq (1.0)	ToxoDB		Me	7 minutes ago	9	295.20 M	3.09%

- Click on the experiment name in the Name/ID column to enter that dataset’s page. If the experiment is still being installed into VEuPathDB you should notice the status line indicating that the dataset is currently being installed.

My Dataset: Experiment A1

Status: This data set is currently being installed in ToxoDB. Please check again soon.

Owner:

Description: samples from experiment A1

ID: 12640

Data Type: RNASeq (RNASeq 1.0)

Summary: Experiment A1

Created: 7 minutes ago

Dataset Size: 295.20 M

Quota Usage: 3.09% of 10.00 G

Use This Dataset in ToxoDB

Compatibility Information

EuPathDB Website	Required Resource	Required Resource Release	Installed Resource Release
ToxoDB	TgondME49 Genome	29	29

This dataset is compatible with the current release, build 40, of ToxoDB. It is installed for use.

Data Files

- Once the dataset is installed, it is ready for use (note the status line indicates it is ready with a green checkmark). You can access the results in two ways:

- Access available searches against your data
- Access GBrowse tracks representing your data

My Dataset: Experiment_A1

 **Status:**  This data set is installed and ready for use in ToxoDB.

Owner: Me

Description: samples from experiment A1 

ID: 12993

Data Type: **RNASeq** (RnaSeq 1.0)

Summary: Experiment A1 

Created: an hour ago

Dataset Size: 295.20 M

Quota Usage: 3.09% of 10.00 G

 **Available Searches:** • [genes by RNA-Seq user dataset \(fold change\)](#)

Use This Dataset in ToxoDB

Compatibility Information

EuPathDB Website	Required Resource	Required Resource Release	Installed Resource Release
ToxoDB	TgondiiME49 Genome	29	29

This dataset is compatible with the current release, build 40, of ToxoDB. It is installed for use.

GBrowse Tracks

Filename	GBrowse Status
day3_rep2.bw	☐ This file has not been added to GBrowse. <button>Send To GBrowse </button>
day3_rep1.bw	☐ This file has not been added to GBrowse. <button>Send To GBrowse </button>
Display a menu	☐ This file has not been added to GBrowse. <button>Send To GBrowse </button>

- To run a search against your data, click on the “genes by RNA-Seq user dataset (fold change)” link.

10. The next page provides you with customizable search parameters with you analyzed samples available for selection.

Identify Genes based on genes by RNA-Seq user dataset (fold change) BETA

Your RNA-Seq Dataset

Experiment A1

For the Experiment **unstranded** ?

return **protein coding** ? Genes

that are **up or down regulated** ?

with a **Fold change** $\geq$ **2** ?

between each gene's expression value ?

in the following **Reference Samples** ?

day3_rep1

day3_rep2

day4_rep1

day4_rep2

[select all](#) | [clear all](#)

and its expression value ?

in the following **Comparison Samples** ?

day3_rep1

day3_rep2

day4_rep1

day4_rep2

[select all](#) | [clear all](#)

[Get Answer](#)

Example showing one gene that would meet search criteria
(Dots represent this gene's expression values for selected samples)

Up or down regulated

This graphic will help you visualize the parameter choices you make at the left. It will begin to display when you choose a **Reference Sample** or a **Comparison Sample**.

See the [detailed help](#) for this search.

11. Once you configure the search parameters, you can click on the “Get Answer” button to get your results.

Identify Genes based on genes by RNA-Seq user dataset (fold change) BETA

Your RNA-Seq Dataset

Experiment A1

For the Experiment **unstranded** ?

return **protein coding** ? Genes

that are **up-regulated** ?

with a **Fold change** $\geq$ **2** ?

between each gene's **expression value** ?

in the following **Reference Samples** ?

day3_rep1

day3_rep2

day4_rep1

day4_rep2

[select all](#) | [clear all](#)

and its **expression value** ?

in the following **Comparison Samples** ?

day3_rep1

day3_rep2

day4_rep1

day4_rep2

[select all](#) | [clear all](#)

[Get Answer](#)

Example showing one gene that would meet search criteria
(Dots represent this gene's expression values for selected samples)

Up-regulated

You are searching for genes that are up-regulated between at least two reference samples and at least two comparison samples.

For each gene, the search calculates:

Fold change = $\frac{\text{Maximum expression level in comparison samples}}{\text{Minimum expression level in reference samples}}$

and returns genes when fold change ≥ 2 .

This calculation involves the maximum value of expression values in which to look for genes that meet your fold change cutoff. To broaden the window, use the average or minimum reference value, or average or maximum comparison value.

See the [detailed help](#) for this search.

183 Genes (166 ortholog groups)

Gene ID	Transcript ID	Product Description	Fold Change	Chosen Ref	Chosen Comp	Profile
TSME49_273290	TSME49_273290.tSL.1	hypothetical protein	265.5	1.31	483.15	
TSME49_220560	TSME49_220560.tSL.1	hypothetical protein	339.1*	0	339.11	

12. The results page includes custom columns based on your data including: fold change, expression values for the reference and comparator (FPKM) and a profile graph. The results can be further intersected with results from other searches in VEuPathDB or analyzed using one of the available analysis tools such as GO enrichment.

13. The data is also represented on gene pages in the Transcriptomics section under the heading “User Dataset Transcriptomics Graphs”.

Contents <<

expand all | collapse all

Search section names...

- 1 Gene models [↗](#)
- 2 Annotation, curation and identifiers [↗](#)
- 3 Link outs [↗](#)
- 4 Genomic Location [↗](#)
- 5 Literature [↗](#)
- 6 Taxonomy [↗](#)
- 7 Orthology and synteny [↗](#)
- 8 Phenotype [↗](#)
- 9 Genetic variation [↗](#)
- 10 Transcriptomics [↗](#)
- 11 Sequence analysis [↗](#)
- 12 Sequences [↗](#)
- 13 Structure analysis [↗](#)
- 14 Protein features and properties [↗](#)
- 15 Function prediction [↗](#)
- 16 Pathways and interactions [↗](#)
- 17 Proteomics [↗](#)
- 18 Immunology [↗](#)

expand all | collapse all

Synonymous SNPs All Strains 4

Total SNPs All Strains 28

10 Transcriptomics

► Transcript Expression [Data sets](#)

▼ User Dataset Transcriptomics Graphs

Search this table... Showing 2 rows

Preview	Name	Assay Type
	Experiment_A1	RNA-seq

TGME49_263310 - UserDataset 12993

Full Dataset Description

► Data table

X-axis
StageSample

Y-axis
Transcript levels of fragments per kilobase of exon model per million mapped reads (FPKM).

Choose gene for which to display graph

● TGME49_263310

Choose graph(s) to display

■ fpkm

Graph options

☐ Show log Scale (not applicable for log(ratio) graphs, percentile graphs or data tables)

14. To view RNA-Seq coverage graphs in the genome browser, go back to the “My Datasets” section, select the dataset you want to view and scroll down to the Genome Browser Tracks section.

Genome Browser Tracks

Filename	Genome Browser Link
Cyst.bw	View in Genome Browser
Tachy.bw	View in Genome Browser

15. Now you can click on the “View in Genome Browser” button to visualize your data in the genome browser.

