

# EuPathDB RNAseq BigWig & expression value export & integration tool

**Note:** using EuPathDB *My Datasets* and the EuPathDB 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 EuPathDB 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 EuPathDB by following these steps:

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

The screenshot shows the EuPathDB Galaxy site interface. The left-hand panel contains a 'Tools' menu with various categories. The 'EuPathDB Export Tools' category is highlighted, and a red box is drawn around the 'RNA-Seq to EuPathDB Export an RNA-Seq result to EuPathDB' tool. The main section of the page displays a welcome message and a list of pre-configured workflows. The workflows are organized into sections: 'With EuPathDB Galaxy you can:', 'Get started with pre-configured workflows:', and 'OrthoMCL'. The 'RNA Sequencing' section lists several workflows for different NGS data types and analysis goals. The 'NGS APPLICATIONS' section lists various tools for different NGS data types and analysis goals.

2. The RNAseq 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 the BigWig collection from the list.
  - c. Select the Cufflinks **gene** expression collection.
  - d. Enter a summary for your dataset (you can always modify this).

- e. Enter a description for your dataset (you can always modify this).
- f. Click on the “Execute” button.

**RNA-Seq to EuPathDB** Export an RNA-Seq result to EuPathDB (Galaxy Tool Version 1.0.0) Options

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

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

**C** **FPKM collection:**  
  
 Select the FPKM collection. Its name should include the phrase 'gene expression'.

**D** **My Data Set summary:**

**E** **My Data Set description:**

**F**

**What it does** (check this [Tutorial!](#))

This tool exports an RNA-Seq result to a EuPathDB site creating a new My Data Set to contain them.

3. After you click on the execute button a new step is initiated in the right hand history pannel: ExportToEuPathDBInfo.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.

**History**

search datasets

**Experiment A1**  
 27 shown, 33 deleted, 32 hidden  
 62.68 GB

92: exportToEuPathDBInfo.html

92: exportToEuPathDBInfo.html

92: exportToEuPathDBInfo.html

Red arrows indicate the job status progression: Grey (queued) → Yellow (running) → Green (finished).

4. Once the export is complete, go to the EuPathDB site where your data will be visualized. For example, if your data is *Plasmodium berghei* RNAseq data go to PlasmoDB.org

5. Click on the “My Datasets” link in the grey menu bar.

The screenshot shows the ToxoDB website interface. At the top, there's a header with the ToxoDB logo and 'Release 39 30 Aug 2018'. Below the header is a grey menu bar with links: Home, New Search, My Strategies, My Basket, **My Data Sets** (circled in red), Tools, Data Summary, Downloads, Community, and Analyze My Experiment. The main content area is divided into four columns: Data Summary, Search for Genes, Search for Other Data Types, and Tools. The 'My Data Sets' link is highlighted with a red circle.

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

The screenshot shows the 'My Data Sets' page. At the top, there's a search bar and a status bar indicating 'Showing 1 of 1 data set'. Below this is a table with columns: Name / ID, Summary, Type, EuPathDB Websites, Status, Owner, Created, File Count, Size, and Quota Usage. The table contains one entry: 'Experiment A1 (12640)' with a status of 'Installing'.

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

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

The screenshot shows the 'My Dataset: Experiment A1' page. The status line is highlighted in red: 'Status: This data set is currently being installed in ToxoDB. Please check again soon.' Below this, there's a description of the dataset, including its owner, ID, type, summary, creation time, size, and quota usage.

**My Dataset: Experiment A1**

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

**Owner:** Me

**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**

8. 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:
  - a. Access available searches against your data
  - b. 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	<input type="radio"/> This file has not been added to <b>GBrowse</b> . <a href="#">Send To GBrowse</a> 
day3_rep1 .bw	<input type="radio"/> This file has not been added to <b>GBrowse</b> . <a href="#">Send To GBrowse</a> 
Display a menu	<input type="radio"/> This file has not been added to <b>GBrowse</b> . <a href="#">Send To GBrowse</a> 

9. 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 dataset (fold change) BETA

## 2 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

### Example showing one gene that would meet search criteria

(Dots represent this gene's expression values for selected samples)


The example shows two graphs. The left graph, titled 'Up or down regulated', shows a gene with expression values (dots) that are significantly higher or lower than the baseline across the selected samples. The right graph shows a gene with expression values that are not significantly different from the baseline.

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.

Get Answer

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) 

**Your RNA-Seq Dataset**

**Experiment A1: 1**

For the Experiment **Experiment A1** **1** Genes

that are **up-regulated**

with a Fold change **> 2**

between each gene's **Experiment A1** expression value **0**

in the library **Reference Samples**

**exp1\_m01**  
**exp1\_m02**  
**exp1\_m03**  
**exp1\_m04**

**select all** **clear all**

and for the library **Reference Samples** **1**

in the library **Experiment A1** **1**

**exp1\_m01**  
**exp1\_m02**  
**exp1\_m03**  
**exp1\_m04**

**select all** **clear all**

**Example showing one gene that would meet search criteria**

This diagram shows the expression level for selected reference

**Up-regulated**

Expression

Minimum Expression Level

Maximum Expression

Reference Samples

Comparison Samples

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

For each gene, the search calculates:

**fold change =  $\frac{\text{minimum expression level in comparison}}{\text{maximum expression level in reference}}$**

and returns genes when fold change is **2**

The calculation method is the **mean** values of expression values in each library to give genes that meet our fold change search. To simplify the workflow, we'll average or normalize reference value, or average or normalize comparison value.

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

Get Answer

My Strategies: [New](#) | [Quoted \(5\)](#) | [All Genes](#) | [Basket](#) | [Public Strategies \(3\)](#) | [Help](#)

Hide search strategy panel

Strategy: RNASeq (RUC) +

Genes

Refresh Data

Add Stop

NA

45 Genes from Step 1 [Refresh](#)

Strategy: RNASeq (RUC)

Click on a number in this table to breakdown your results																					
		Connectome - Synchrograms				Ethnicity				Hemorrhoids - Shingles				Sarcoidosis				Tryptase			
All Results	Selected Genes	Clopidogrel	Clozapine	Cocaine	Zalcitabine	Zidovudine	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	Zinc	
45	NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Gene Results

[Generate View](#) | [Export Results](#)

Filter: 1 2 3 4 Most Last | [Advanced Filtering](#) | [Download](#) | [Add to Basket](#) | [Add Columns](#)

Gene ID	Transcript ID	Organization	Product Description	Fold Change	Chosen Ref	Chosen Comp	Prefilter %						
TGM4E4_283210	TGM4E4_283210-GS_1	1 genbank NCBI	step 1 replicating factor, putative	41.47	0	41.27	<div><div>TGM4E4_283210 - Use Chosen 12960</div><div><table><thead><tr><th>Category</th><th>Value</th></tr></thead><tbody><tr><td>Chosen</td><td>41.27</td></tr><tr><td>Ref</td><td>41.47</td></tr></tbody></table></div></div>	Category	Value	Chosen	41.27	Ref	41.47
Category	Value												
Chosen	41.27												
Ref	41.47												
TGM4E4_283180	TGM4E4_283180-GS_1	1 genbank NCBI	acid phosphatase	35.47	0	33.30	<div><div>TGM4E4_283180 - Use Chosen 12960</div><div><table><thead><tr><th>Category</th><th>Value</th></tr></thead><tbody><tr><td>Chosen</td><td>33.30</td></tr><tr><td>Ref</td><td>35.47</td></tr></tbody></table></div></div>	Category	Value	Chosen	33.30	Ref	35.47
Category	Value												
Chosen	33.30												
Ref	35.47												

[Display as matrix](#)

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 EuPathDB 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

fpkm

0

20

40

dis0\_rep1

dis0\_rep2

dis4\_rep1

dis4\_rep2

Full Dataset Description

Data table

X-axis

Stage/Sample

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 RNAseq 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 GBrowse Tracks section.

**Data Type:** RNASeq (RnaSeq 1.0)  
**Summary:** Experiment A1 [✎](#)  
**Created:** 2 hours 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	<input type="radio"/> This file has not been added to GBrowse. <a href="#">Send To GBrowse</a> 
day3_rep1.bw	<input type="radio"/> This file has not been added to GBrowse. <a href="#">Send To GBrowse</a> 
day4_rep1.bw	<input type="radio"/> This file has not been added to GBrowse. <a href="#">Send To GBrowse</a> 
day4_rep2.bw	<input type="radio"/> This file has not been added to GBrowse. <a href="#">Send To GBrowse</a> 







15. Click on the “Send To GBrowse” buttons.

### 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	 Sent to GBrowse. <a href="#">View In GBrowse</a> <a href="#">?</a>
day3_rep1.bw	 Sent to GBrowse. <a href="#">View In GBrowse</a> <a href="#">?</a>
day4_rep1.bw	 Sending to GBrowse... <a href="#">Send To GBrowse</a> 
day4_rep2.bw	 Sending to GBrowse... <a href="#">Send To GBrowse</a> 

16. Now you can click on the “View in GBrowse” button to visualize your data in the genome browser.

