Basic Tutorial
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting Started</td>
<td>2</td>
</tr>
<tr>
<td>Data Input</td>
<td>8</td>
</tr>
<tr>
<td>Data Formats</td>
<td>9</td>
</tr>
<tr>
<td>Tutorial – Getting Data from GEO</td>
<td>10</td>
</tr>
<tr>
<td>ROGUE – Differential Expression Analysis</td>
<td>24</td>
</tr>
<tr>
<td>ROGUE – Create Groups</td>
<td>34</td>
</tr>
<tr>
<td>ROGUE – Groups Differential Expression Analysis</td>
<td>43</td>
</tr>
<tr>
<td>ROGUE – Single Sample Gene Expression Analysis</td>
<td>44</td>
</tr>
<tr>
<td>ROGUE – Groups Gene Expression Analysis</td>
<td>46</td>
</tr>
<tr>
<td>ROGUE – Gene Set Enrichment Analysis</td>
<td>47</td>
</tr>
<tr>
<td>ROGUE – Gene Ontology Analysis</td>
<td>49</td>
</tr>
<tr>
<td>ROGUE – Statistical Group Comparison (Biomarker Discovery)</td>
<td>52</td>
</tr>
<tr>
<td>ROGUE – tSNE</td>
<td>55</td>
</tr>
<tr>
<td>ROGUE – Gene Ontology Expression Comparison</td>
<td>57</td>
</tr>
<tr>
<td>ROGUE – Differentially Expressed Gene Ontologies</td>
<td>60</td>
</tr>
</tbody>
</table>
Getting Started

Rogue G2M & Gene Ontology Graphic User Environment

- View ROGUE Manual
- ROGUE
  - Load Count Data (EdgeR)
  - Load Expression Data
  - Create Groups
  - EdgeR Group Comparison
  - Gene Comparison (Samples)
  - Gene Comparison (Groups)
  - Advanced Analysis
- Download Report
In-app help
Click the ‘?’ for tips on using the app
In-app help: Describes the tools in the app. Click 'Next' to see more tips.

If your RNAseq data is in raw reads or counts, click the 'Load Count Data(EdgeR)' in the left menu bar to convert to RPKM and perform differential gene expression analysis.
Rshiny applications may timeout or crash if there is a disconnect from the server. Session IDs are generated so that an analysis can be recovered in the event of disconnection. The uploaded data and some of the plots will be recovered.

A session ID is generated each time the app is started. The session ID consists of the user’s IP address, the date, and an iterative number (1, 2, 3, ...):

XXX.XX.XXX.XXX_YYYY.MM.DD_I

This format allows the user to figure out the session ID if the previous or crashed session is no longer available or closed.

If the user is behind a firewall that blocks the IP address, a custom IP address is generated for the session ID and will be different each time the app is started. This would make it more difficult to guess a session ID if the previous session page has been closed. The custom IP address will start with ‘CUS’ so the session ID will be in the following format:

CUS.XX.XXX.YYYY.MM.DD_I

The Session IDs are found at the bottom of the left menu pane and on the Restore/Load Session Page.
Getting Started – Restoring a session with a session ID

1. Go to the ‘Restore/Load Session’ page

2. Enter a previous session ID in the ‘Enter Session ID’ field and click the ‘Restore Session’ button. Note: sessions are only stored on the server temporarily and will be removed often.

3. The ‘Sample’ and ‘Gene’ fields in the application will be populated with the restored data. Some analyses and plots will also be restored.
Getting Started – Downloading and uploading with a session

(1) Go to the ‘Restore/Load Session’ page

(2) A session can be downloaded to your local computer by clicking the ‘Download session’ button. The session will be a ‘rdata’ file

(3) If a session file (rdata) has been downloaded to your local computer, it can be uploaded using the upload function.

(4) The ‘Sample’ and ‘Gene’ fields in the application will be populated with the restored data. Some analyses and plots will also be restored.
ROGUE can use count/reads data (raw or normalized) and expression data (FPKM/RPKM) as input.
## Data Formats

### Raw Counts

<table>
<thead>
<tr>
<th>gene_name</th>
<th>symbol</th>
<th>len</th>
<th>Sample_1</th>
<th>Sample_2</th>
<th>Sample_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_177327</td>
<td>Wwp1</td>
<td>5070</td>
<td>687</td>
<td>579</td>
<td>761</td>
</tr>
<tr>
<td>NM_177326</td>
<td>Pak2</td>
<td>4099</td>
<td>5194</td>
<td>4724</td>
<td>5962</td>
</tr>
<tr>
<td>NM_177325</td>
<td>Tsr1</td>
<td>3385</td>
<td>1204</td>
<td>1421</td>
<td>2334</td>
</tr>
</tbody>
</table>

**Note**: Tab delimited files: Raw counts can be submitted genenames, symbols, len (‘len’ column are the gene lengths in bases and needs to be labeled ‘len’), followed by sample columns with raw count values.

### Raw Counts

<table>
<thead>
<tr>
<th>symbol</th>
<th>Sample_1</th>
<th>Sample_2</th>
<th>Sample_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wwp1</td>
<td>687</td>
<td>579</td>
<td>761</td>
</tr>
<tr>
<td>Pak2</td>
<td>5194</td>
<td>4724</td>
<td>5962</td>
</tr>
<tr>
<td>Tsr1</td>
<td>1204</td>
<td>1421</td>
<td>2334</td>
</tr>
</tbody>
</table>

**Note**: Tab delimited files: Raw counts can be submitted with the genename/symbol column followed by the Sample columns with the raw counts. The tool will use pre-determined gen lengths to normalize counts.

### Normalized Counts (by length)

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Sample_1</th>
<th>Sample_2</th>
<th>Sample_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSG000000000419</td>
<td>687.43</td>
<td>579.63</td>
<td>761.19</td>
</tr>
<tr>
<td>ENSG000000000457</td>
<td>5194.64</td>
<td>4724.91</td>
<td>5962.03</td>
</tr>
<tr>
<td>ENSG000000000460</td>
<td>1204.01</td>
<td>1421.63</td>
<td>2334.32</td>
</tr>
</tbody>
</table>

**Note**: Tab delimited files: Normalized counts can be submitted with the genename/symbol column followed by the Sample columns with the normalized counts.
### Data Formats

**Note:** Tab delimited files: Expression values can be submitted with the genename/symbol column followed by the Sample columns with the expression values (eg. FPKM/RPKM).

<table>
<thead>
<tr>
<th>genenames</th>
<th>Sample_1</th>
<th>Sample_2</th>
<th>Sample_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSPAN6</td>
<td>17.84</td>
<td>16.37</td>
<td>18.45</td>
</tr>
<tr>
<td>TNMD</td>
<td>0</td>
<td>0.18</td>
<td>0</td>
</tr>
<tr>
<td>DPM1</td>
<td>23.05</td>
<td>21.24</td>
<td>15.8</td>
</tr>
</tbody>
</table>

**Expression Values**

2. Enter GSE60424 in the search box

3. Click Search

4. You will be redirected to this page

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

GSE60424

Status: Public on Jan 06, 2015
Organism: Homo sapiens
Experiment type: Expression profiling by high throughput sequencing
Summary: This study compared whole transcriptome signatures of 6 immune cell subsets and whole blood from patients with an array of immune-associated diseases. Fresh blood samples were collected from healthy subjects and subjects diagnosed type 1 diabetes, amyotrophic lateral sclerosis, and sepsis, as well as multiple sclerosis patients before and 24 hours after the first treatment with IFN-beta. At the time of blood draw, an aliquot of whole blood was collected into a Tempus tube (Invitrogen), while the remainder of the primary fresh blood sample was processed to highly pure populations of neutrophils, monocytes, B cells, CD8 T cells, CD4 T cells, and natural killer cells. RNA was extracted from each of these cell subsets, as well as the whole blood sample, and processed into RNA sequenc ing (RNAseq) libraries (Illumina TruSeq). Sequencing libraries were analyzed on an Illumina HiScan, with a targeted read depth of ~20M reads. Reads were demultiplexed, mapped to human gene models (ENSGMBL), and tabulated using HTSeq. Read count data were normalized by the TMM procedure (edgeR package).

Overall design: We performed whole genome RNAseq profiling of immune cell subsets and whole blood from subjects with an array of immune-associated diseases.

Contributor(s): Speake C, Lindsey PS, Whelan E, Chauwasebol D, Pressnell SB, Mason MJ, Gersuk VH, O'Brien K, Nguyen O, Greenbaum CJ, Buckner JM, Mathore U

Citation(s): Lindsey PS, Speake C, Whelan E, Chauwasebol D. Copy number loss of the interferon gene cluster in melanomas is linked to reduced T cell infiltration and poor patient prognosis. J Clin Oncol 2014;32(10):e109760. PMID: 25314013

Submission date: Aug 14, 2014
Last update date: May 15, 2019
Contact name: Scott Pressnell
E-mail(s): speake@bennaryresearch.org
Organization name: Benaroya Research Institute
Department: Systems Immunology
Street address: 1201 Ninth Ave.
City: Seattle
State/province: WA
ZIP/Postal code: 98101
Country: USA
Click ‘Series Matrix File(s)’

Download ‘GSE60424_series_matrix.txt.gz’

Click (ftp) to download the normalized counts file.
Note: GSE60424_GEOSubmit_FC1to11_normalized_counts.txt is a normalized counts matrix. The file should look like this:

GSE60424_GEOSubmit_FC1to11_normalized_counts.txt

genenames      lib221 lib222 lib223 lib224 ...... lib355
ENSG00000000003  1   0   1   1
ENSG00000000005  0   0   0   0
ENSG00000000419  23  22  25  28
ENSG00000000457  11  11  14  17
ENSG00000000460  3   3   3   3
ENSG00000000938  1141 872 1068 629
ENSG00000000971  3   1   1   3
ENSG00000001036  17  21  27  17
ENSG00000001084  19  14  19  19
ENSG00000001167  35  33  33  44
...  ...  ...  ...  ...
...  ...  ...  ...
...  ...  ...  ...
...  ...  ...  ...
...  ...  ...  ...
This study compared whole transcriptome signatures of 6 immune cell subsets and whole blood from patients with an array of immune-associated diseases. Fresh blood samples were collected from healthy subjects and subsets diagnosed type 1 diabetes, amyotrophic lateral sclerosis, and sepsis, as well as multiple sclerosis patients before and 24 hours after the first treatment with IFN-beta. At the time of blood draw, an aliquot of whole blood was collected into a Tempus tube (Invitrogen), while the remainder of the primary fresh blood sample was processed to highly pure populations of neutrophils, monocytes, CD4+ and CD8+ T cells, B cells, and natural killer cells. RNA was extracted from each of these cell subsets, as well as the whole blood samples, and processed into mRNAseq libraries (Illumina TruSeq). Sequencing libraries were analyzed on an Illumina HiScan, with a target read depth of ~3M reads. Reads were demultiplexed, mapped to human gene models (ENSEMBL), and tabulated using HTSeq. Read count data were normalized by the TMM procedure (edgeR packe) and differentially expressed using DESeq2.

Note: This is a standard series matrix file that is uploaded with data to GEO. It describes the data using standardized fields defining filenames, sample labels, and sample characteristics, and the related publication.
Tutorial - Getting Data from GEO

Go to ‘ROGUE_Companion’ to prepare files.

Note: The GEO ROGUE Companion tool was created to help prepare a data matrix and group file from data downloaded from GEO for analysis using ROGUE. Both a data matrix and Series matrix file are required.

Follow this tutorial as an example.
Tutorial - Getting Data from GEO

Note: Upload the files downloaded from GSE60424.
Tutorial - Getting Data from GEO

Note: Upload the files downloaded from GSE60424.

When the series_matrix file is uploaded the field headers will populate the ‘Series Matrix field’ dropdown box and the details/members of the selected field will fill the ‘Details’ text box.
When an ID or Title field is selected the ‘Details’ text box is populated with the members of that field as one line per member.

It is important that the user chooses the ID or Title field that has details with the same names as the column headers listed in the ‘Data Matrix Samples Names/headers’ text box.

When a Data matrix is loaded, the column headers fill the ‘Data Matrix Samples Names/headers’ text box.
Enter a group name.

Choose a feature field that is extracted from the "Characteristics" fields from the series matrix file.

Choose the feature details of the members of the group.

The added features will appear in this box. Multiple features can be added. When a group is created, this tool will include the samples that have all the selected characteristics as members of the group.

When 'Create Group' Button is clicked and a group is successfully create, it will appear in the format:
'Group_Name:Member1;Member2;Member3'
This is the same group format required by the main ROGUE tool.
Note: Follow these steps to create groups that will be used in the ROGUE tutorial.

1. Type ‘Healthy_CD4’ as the Group Name
2. Choose ‘celltype’
3. Choose ‘CD4’
4. Click ‘Add Group Feature’
5. Selected feature is listed in the “Group Features” text box.
(6) Choose ‘diseasestatus’

(7) Choose ‘Healthy Control’

(8) Click ‘Add Group Feature’

(9) Selected feature is added to the “Group Features” text box.
Click 'Create_Group' (10)

Groups are created in format 'Group_Name:Member1;Member2;Member3' (11)
Create Healthy CD8 Groups and Healthy NK Groups (This Should be the result).

Select 'Human Gene Symbols' if you want to download a table with gene symbols instead of Ensembl IDs.

Download Data Tables with Gene Symbols and Group Files

Note: Download the Data Table as ‘Data_Matrix.txt’ and the groups file as ‘Groups_File.txt’
ROGUE – Differential Expression Analysis

https://reslnmaris01.research.chop.edu/ROGUE/

- ROGUE
  - Load Count Data: (EdgeR/DEseq2)
  - Load Expression Data
  - Create Groups
  - EdgeR/DEseq2 Group Comparison
  - Gene Comparison (Samples)
  - Gene Comparison (Groups)
  - Advanced Analysis

Click the ‘Load Count Data: (EdgeR)’
Clicking the in-app help button will describe each option and guide the user through the process. Click ‘Next’ to see more tips and be guided through the options.
A tool to load raw or normalized counts appear where a user can perform differentially expressed genes (DEG) analysis and convert the counts to RPKM expression values.
Check the upload radio button and click ‘Browse’ to select a Counts file.
ROGUE – Differential Expression Analysis

Choose Reads File

Browse... No file selected

Select Input Source

○ Database
○ Upload File

Select Data Library

GSE60424

Select Dataset


Select Reads Status

○ Raw ○ Normalized

Load Data

Select Counts File
This file has normalized counts (by gene length). Select the ‘Normalized’ radio button.

Click the ‘Load Data’ Button
ROGUE – Differential Expression Analysis

Load Raw Reads and Perform Differentially Expressed Gene Analysis

An MDS plot with sample names appears when the data is loaded.
Select two Libraries (Sample Names) to perform DEG analysis from the ‘Select Library’ dropdown menu.
Click ‘Compare Libraries’ button to perform DEG analysis and display plots.
Click ‘GeneLists’ tab. User can download the list of genes that are differentially expressed, just the upregulated or downregulated gene lists, or DEG tables with RPKM values.
Click the ‘Create Groups’ tab.

Click ‘?’ button for step by step tips on using this tool.
Create groups by choosing samples/libraries from the ‘Select Group Members’ dropdown list.
Assign a name to the group in the ‘Group Name’ textbox then click the ‘Create Group’ button.
The Group will appear in the ‘Groups’ text box. It is in the format:
Group_Name:Member1;Member2;Member3
Create multiple groups and save them using the ‘Download Groups’ button.
Enter groups manually by typing in the format:
Group_Name:Member1;Member2;Member3
Or upload saved groups by clicking ‘Browse’ button
ROGUE – Create Groups

Select the Groups file
When the Browse box says, 'Upload complete' click the ‘Upload Groups’ button.
Manual group entries by typing and group file upload would not be loaded into the program until the user clicks the ‘Confirm Groups’ button. This checks for errors and removes group members that are not in the loaded dataset. It then loads the groups into the other features of the tool.
Select ‘Healthy_CD8’ from the ‘Select Group 2’ dropdown box. and click ‘Compare Groups’

Plots will be generated on the summary plots page
Click ‘Gene Comparison (Samples)’ tab. User can select genes from the ‘Select Genes’ dropdown list or type/paste genes in the ‘Paste Gene List’ text box. Select Libraries or samples to compare from the ‘Select Samples’ dropdown list.
Click the ‘Compare Gene Expressions’ button to generate bar plots and heatmaps, which will appear when applicable. Change the ‘Select Theme’ option to black and white to alter the view of the plots.
Click on the ‘Gene Comparison (Groups)’ tab. Select Groups from the dropdown list. Enter genes into the ‘Paste Gene List’ textbox or select genes from the dropdown list. Then click the ‘Compare Groups’ Expressions’ button. Boxplots, bar plots, and heatmaps will appear comparing the genes’ expressions.
Click the ‘Advanced Analysis’ tab. Then Click on ‘Gene Set Enrichment Analysis’

Click ‘?’ button for step by step tips on using this tool
Select ‘Groups’ radio button. Then select groups for ‘Select Control’ and ‘Select Subjects’ drop down box. Select GSEA sets from the ‘Select GSEA Collection’ drop down box. Click the ‘Find Enrichment Gene Signatures’ button.
Click ‘Advanced Analysis’ tab. Click on the ‘Gene Ontology tab’. Users can type or paste a list of genes in the ‘Enter Gene List’ textbox. Click ‘Get Ontologies’ button. The initial processing of this function may take a few minutes.

Don’t forget to use the ‘?’ button for tips.

If the gene list is long, it is HIGHLY RECOMMENDED to use an external gene ontology/pathway database.
Gene Ontology bar plots appear. Choose ontology subcategories and click ‘Select GO’ button to explore even deeper subcategories within the selected ontology.
Gene Ontology bar plots appear. Choose ontology subcategories and click ‘Select GO’ button to explore even deeper subcategories within the selected ontology.

Note: This text box lists the genes represented in the bar plot.
Click ‘Group Statistical Comparison’ tab. This tool attempts to identify potential biomarkers by comparing gene expression across all members of each group. Select groups to be compared in the ‘Select Groups’ dropdown list. check ‘Fold Change’ and ‘Mean/SD’ checkboxes. Click ‘Find Genes’ button.
Plots of expression values and most consistently differentially expressed genes across groups will be displayed. Only the top 10 (default) genes will be displayed in the dot and box plots. This value can be changed. All data and plots can be downloaded to a pdf file.
Click on the 'Group_Stats_GeneList' tab.
A list of genes and the log2 fold change will appear in a textbox ranked from highest to lowest fitting the parameters set in the left panel.
Click on 't-SNE' tab

<table>
<thead>
<tr>
<th>Select Group1</th>
<th>Healthy_CD4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select Group2</td>
<td>Healthy_CD8</td>
</tr>
<tr>
<td>Select Tests</td>
<td></td>
</tr>
<tr>
<td>Fold Change</td>
<td></td>
</tr>
<tr>
<td>Mean/SD</td>
<td></td>
</tr>
<tr>
<td>Wilcoxon</td>
<td></td>
</tr>
<tr>
<td>Ttest</td>
<td></td>
</tr>
<tr>
<td>Find Genes</td>
<td></td>
</tr>
<tr>
<td>Select Pvalue</td>
<td></td>
</tr>
<tr>
<td>Min RPKM</td>
<td>5</td>
</tr>
<tr>
<td>Min log2FC</td>
<td>2</td>
</tr>
<tr>
<td>Display Top Genes</td>
<td>10</td>
</tr>
<tr>
<td>Show log2FC in gene list?</td>
<td></td>
</tr>
</tbody>
</table>
Select ‘All Samples’ radio button and change ‘Choose t-SNE max iterations’ sliding bar to 1000. 2D and 3D t-SNE plots will be generated.
Click on ‘Group Expr. Ontologies’ tab.
Select ‘Groups’ radio button
Select one Control and either one or multiple Subjects from the dropdown lists.
Select ‘immune system process’ from ‘Select GO Class’ drop down menu.
Selecting the ‘immune system process’ from the Select GO Class dropdown list populates the ‘Select GO’ dropdown list with all the gene ontologies related to immune system processes. Select the GOs you would like to evaluate (e.g. innate immune response).
Click the ‘Beside’ radio button and the ‘Label points’ check box then click ‘Get GO Gene’s Fold Change’ button. Adjust the ‘Graph Width’ parameter.
Click on ‘Differentially Expressed Gene Ontologies’ tab. Select one Control and one Subject from the dropdown lists. Select GO Class and/or select keywords to include in the search. The tool will look for gene ontologies that are related to the GO class or that include the keywords.
Perform analysis searching for differences in biological regulation