

Host- Pathogen Interactions: WGCNA Study Explorer

What is the VEuPathDB Study Explorer? It is an interactive feature that allows you to

- Learn more about a dataset
- Explore all the variables in a dataset
- Perform **exploratory data analysis** to visualize associations between two or more variables
- Download the data and work with it on your own

Learning objectives: Navigate the study explorer for the WGCNA dataset of metadata from dual (host and parasite) transcriptomic analysis of 46 Gambian children infected with either severe malaria or uncomplicated malaria to

- Browse metadata and use visualization tools
- Conceptualize co-expression networks and choose a host or parasite module of interest

1. Navigate to the WGCNA study explorer:
https://hostdb.org/hostdb/app/workspace/analyses/DS_82dc5abc7f/new/learn
2. **Examine the layout of the page**- there is a diagram at the top (red arrow in image below) and a set of tabs that you can toggle between (red box below). The “Learn” tab explains the basic features of the study explorer.

The screenshot shows the HostDB website interface. At the top, there is a navigation bar with the HostDB logo, release information (6.8, 7 May 2024), a search bar, and links to My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact Us. The main heading is "Study Explorer" with the subtitle "Dual transcriptomes of malaria-infected Gambian children (Study Explorer)". Below this, there is a diagram showing the data structure: "Sample sources" (46 of 46) → "Samples" (46 of 46) → "Assays" (46 of 46). A red arrow points to the "Sample sources" box. Below the diagram is a row of tabs: "View Study Details", "Browse and Subset", "Visualize", "Download", "Record Notes", and "Learn". The "Learn" tab is highlighted with a red box. Below the tabs, there is a section titled "Learn the basic features of the study explorer" with a link to the "WGCNA study explorer tutorial". The "Overview" section explains that the Study Explorer is an interactive feature for learning about datasets, exploring variables, and performing exploratory data analysis. The "The Dataset Diagram" section explains that the diagram at the top depicts the dataset structure and sample size. The "View Study Details" section explains that the "View Study Details" tab provides a summary of the dataset, links to associated publications, and a list of study investigators. The "Browse and Subset" section explains that the "Browse and Subset" tab enables users to browse through a hierarchical tree of study variables, displayed in a list on the left of the page.

3. **Examine the dataset diagram** at the top of the page.
 - a. How many sample sources (i.e., study participants) were in this dataset?
 - b. How many assays are represented in the dataset?
4. **Learn more about the dataset.** Click on the “**View Study Details**” tab and read the study description. How were the blood samples assayed?

Study Explorer
Dual transcriptomes of malaria-infected Gambian children (Study Explorer)

Unnamed Analysis

Sample sources 46 of 46 → Samples 46 of 46 → Assays 46 of 46

[View Study Details](#) [Browse and Subset](#) [Visualize](#) [Download](#) [Record Notes](#) [Learn](#)

Organism (source or reference):
Primary publication: Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. Lee et al. Sci Transl Med 2018 06 27;10(447):1-11.
Primary contact: Aubrey Cunningham, Section of Paediatrics, Imperial College
Source version(s): 2024-01
Release # / date: HostDB rel. 68, 2024-MAY-07
Summary: Clinical and WGCNA analysis variables for the dual transcriptomic (human and parasite) analysis of blood of 46 Gambian children diagnosed with severe or uncomplicated malaria.

1 Data Provenance

Associated Publications

- Integrated pathogen load and dual transcriptome analysis of systemic host-pathogen interactions in severe malaria. Lee et al. Sci Transl Med 2018 06 27;10(447):1-11.

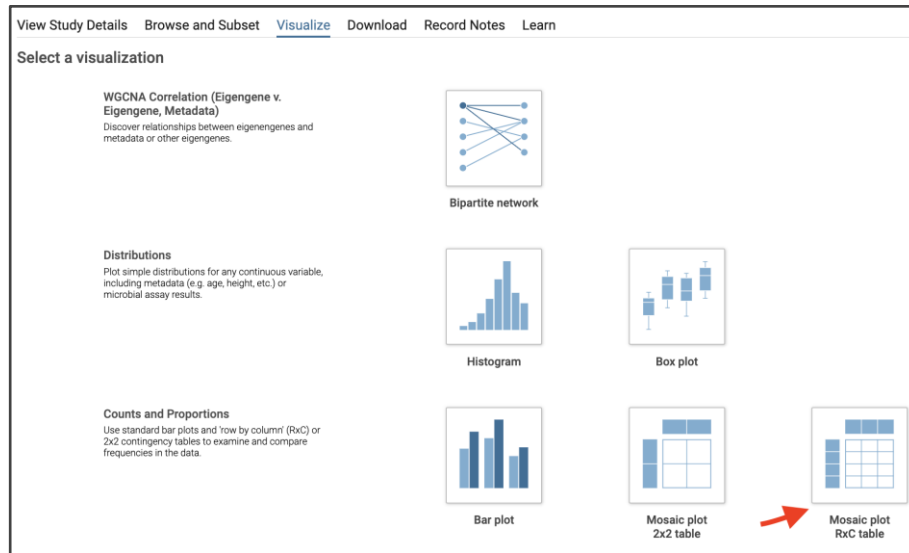
Description

Study Description: Children with uncomplicated (n=21) or severe malaria (n=25) were recruited from a region with relatively low malaria transmission. Consistent with this epidemiology, the children had no cases of severe malaria, hyperlactatemia, or a combination thereof, but no cases of severe anemia. Clinical data was collected from each participant such as: age, ethnicity, Malaria disease (uncomplicated or severe), blood counts, parasitemia and immune profile. Blood samples were also assayed by RNA-sequence analysis.

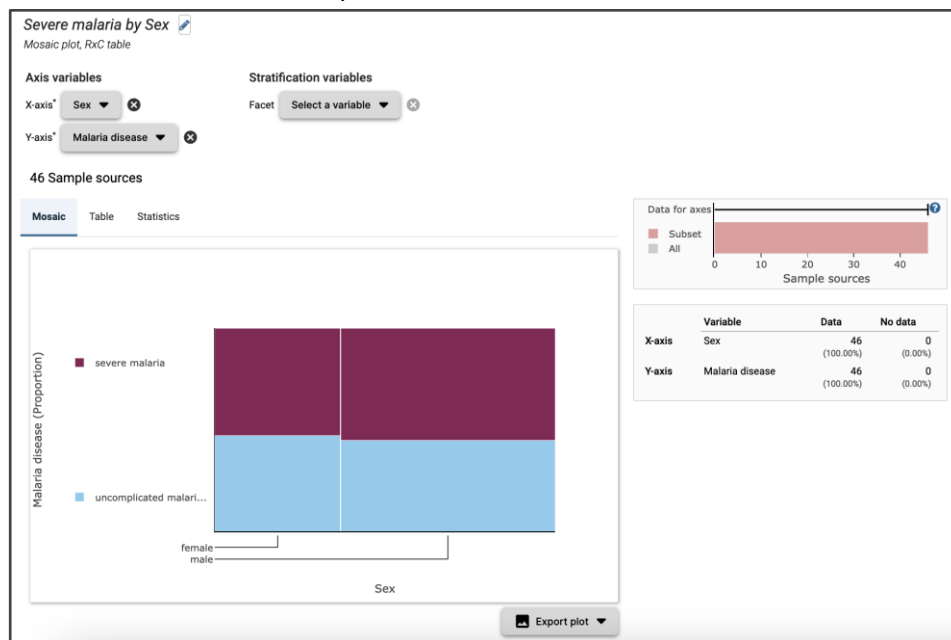
5. **Learn more about the metadata (variables) in this dataset.** Click on the “Browse and Subset” tab. Examine the list of variables on the left. Click on any variable to see its distribution. Note that the distribution of **continuous variables** like *Age* is represented using a histogram, while the distribution of **categorical variables** like *Sex* is represented using a frequency table.

The “Find a variable” search box allows you to search for variables. In this dataset, you can see variables associated with

- a. **Sample Source** (study participants), such as *Age*, *Sex*, and *Malaria disease*
 - i. What is the mean age of the participants?
 - ii. What percentage of participants have severe malaria in this dataset?
 - b. **Sample** (blood sample) such as *Percentage parasitemia*
 - i. How many participants had 5-10% parasitemia? (Hint: you may have to adjust the bin width for the X axis in the controls under the graph)
 - c. **Assay** (WGCNA) such as host and parasite eigengenes, representing co-expression networks/clusters/modules (more on this later)
6. **Explore associations between two or more variables.** Click on the “**Visualize**” tab and click on “new visualization” to open a menu of visualization apps.
 - a. Does the severity of malaria vary in females vs. males in this dataset? What app would you use to answer this question?

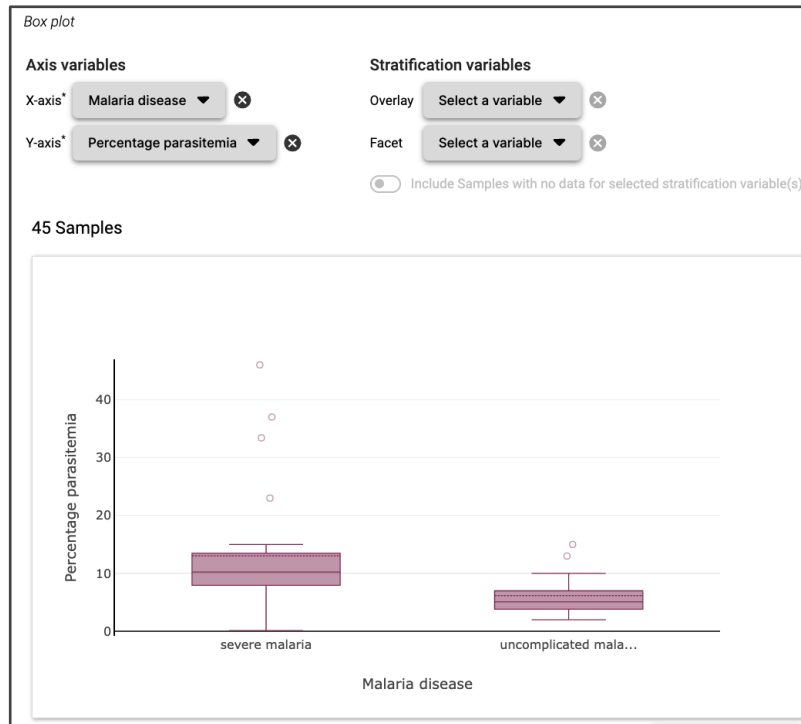


- i. Click on New visualization > Counts and proportions > **Mosaic plot, RxC table**
- ii. Input X-axis= Sex; Y-axis= Malaria disease
- iii. Give the visualization a name
- iv. Hover over the resulting graph to see the proportions of severe/uncomplicated malaria in females/males.



- v. Is there a difference in the proportion of malaria between males and females?

- b. Is there a difference between parasite burden (percentage parasitemia) between children with uncomplicated and severe malaria? What app would you use to answer this question?
- Click on Visualize> New visualization > Counts and proportions > **Box plot**
 - Input X-axis= Malaria disease; Y-axis= Percentage parasitemia
 - Is there a difference in parasite burden between the two groups?

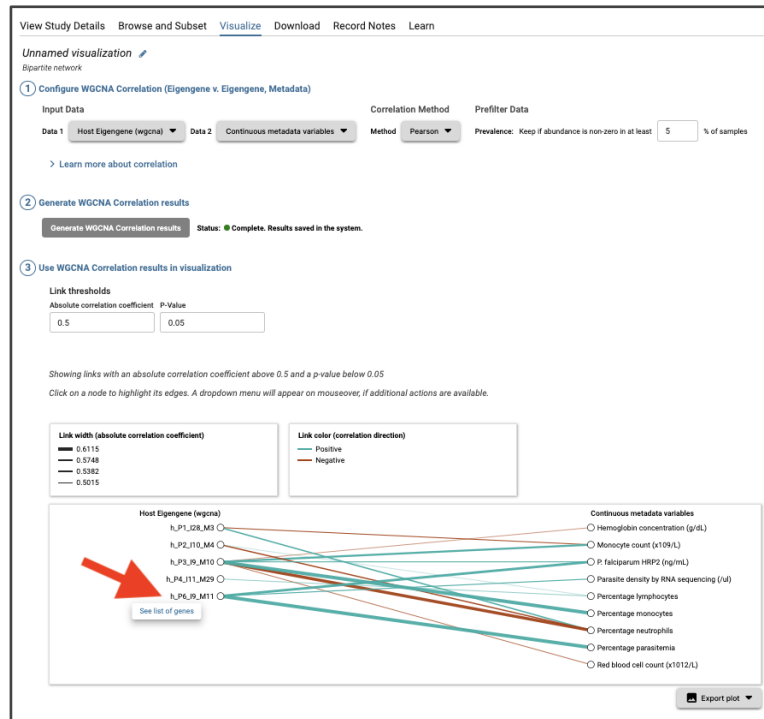


- c. In this dataset, Weighted Gene Coexpression Network Analysis (WGCNA) was performed on the dual transcriptomic data to identify groups of genes that have similar expression across samples. The analysis considers the whole transcriptome data set and clusters coexpressed genes into groups called **modules**. Each module is represented by an **eigengene**, which is created in the WGCNA analysis and is an imaginary gene whose expression profile represents an average gene within the module.

Previous studies have shown a correlation between the expression levels of host genes and circulating parasitemia. In this dataset, parasite load differed between the groups with severe and uncomplicated malaria (as we visualized in step 7 of this tutorial). We were interested in identifying genes correlated with parasite density.

- Click on New visualization > **Bipartite network**

- ii. Input the Data 1, choosing Host eigengene (wgcn) from the drop-down menu. For Data 2, select continuous metadata variables
- lii. Choose a correlation method, e.g., Pearson correlation to identify linear trends
- iv. Click on “Generate correlation results”
- v. Visualize the correlation results. What host eigengene is positively correlated with Percentage parasitemia?



- vi. Click the arrow next to h_P6_I9_M11 to go back to the search strategy and see the list of genes associated with this eigengene. Do you see any interesting genes in this list?

HostDB

Site search, e.g. ENSMUSG000000021403 or "reductase or "binding protein"

My Strategies Searches Tools My Workspace Data About Help Contact Us

My Organism Preferences (10 of 10)

Unnamed Search Strategy *

Heap Gambian children RNA-S... 10 Genes

Step 1

10 Genes (10 ortholog groups) Revise this search

Organism Filter

select all | clear all | expand all | collapse all

☐ Hide zero counts ☐ Reference only

Search organisms...

☐ Aves 0

☐ Mammalia 10

Gene Results Genome View Analyze Results

Genes: 10 Transcripts: 70 (hiding 60) ☒ Show Only One Transcript Per Gene ☐ Show only the Genes in my basket.

Rows per page: 20

Download Send to... Add Col

Gene ID	Transcript ID	Genomic Location (Gene)	Product Description
ENSG00000134827	ENST00000257264	hsapREF_chr11:59,852,800..59,866,489(-)	transcobalamin 1 [Source:HGNC Symbol;Acc:HGNC:11652]
ENSG00000118113	ENST00000236826	hsapREF_chr11:102,711,796..102,727,050(-)	matrix metalloproteinase 8 [Source:HGNC Symbol;Acc:HGNC:7175]
ENSG00000204936	ENST00000618265	hsapREF_chr19:43,353,686..43,363,172(+)	CD177 molecule [Source:HGNC Symbol;Acc:HGNC:30072]
ENSG00000115590	ENST00000332549	hsapREF_chr2:101,991,960..102,028,544(+)	interleukin 1 receptor type 2 [Source:HGNC Symbol;Acc:HGNC:5994]
ENSG0000012223	ENST00000231751	hsapREF_chr3:46,435,645..46,485,234(-)	lactotransferrin [Source:HGNC Symbol;Acc:HGNC:5720]
ENSG00000096006	ENST00000433368	hsapREF_chr6:49,727,376..49,744,437(-)	cysteine rich secretory protein 3 [Source:HGNC Symbol;Acc:HGNC:16904]
ENSG00000272398	ENST00000619133	hsapREF_chr6:106,969,831..106,975,627(+)	CD24 molecule [Source:HGNC Symbol;Acc:HGNC:1645]
ENSG00000118520	ENST00000356962	hsapREF_chr6:131,470,832..131,584,332(+)	arginase 1 [Source:HGNC Symbol;Acc:HGNC:663]
ENSG00000179869	ENST00000435803	hsapREF_chr7:48,171,458..48,647,497(+)	ATP binding cassette subfamily A member 13 [Source:HGNC Symbol;Acc:HGNC:14638]
ENSG00000148346	ENST00000470902	hsapREF_chr9:128,149,071..128,153,453(+)	lipocalin 2 [Source:HGNC Symbol;Acc:HGNC:6526]