# H5N1 Data Analysis Workflow

# 

# Table of Content

1. [Environment Setup](https://github.niaid.nih.gov/chi/h5n1#environment-setup)
2. [Titers](https://github.niaid.nih.gov/chi/h5n1#titers)
   1. [Figure 1A](https://github.niaid.nih.gov/chi/h5n1#figure-1a)
   2. [Figure 1B](https://github.niaid.nih.gov/chi/h5n1#figure-1a)
   3. [Supplemental Figure 1](https://github.niaid.nih.gov/chi/h5n1#supplemental-figure-1)
   4. [Figure 1C,D,E](https://github.niaid.nih.gov/chi/h5n1#figures-1-cde-figuresprofiles)
3. [Pattern profiles of clinical CBC data and Luminex](https://github.niaid.nih.gov/chi/h5n1#pattern-profiles-of-clinical-cbc-data-and-luminex)
4. [Pattern simulation (Figure 2A)](https://github.niaid.nih.gov/chi/h5n1#pattern-simulation-figure-2A)
5. [Gene Expression PBMC data processing](https://github.niaid.nih.gov/chi/h5n1#gene-expression-pbmc-data-processing)
6. [Pattern discovery in post-vaccination profiles of gene expression](https://github.niaid.nih.gov/chi/h5n1#pattern-discovery-in-post-vaccination-profiles-of-gene-expression)
   1. [Figure 2B](https://github.niaid.nih.gov/chi/h5n1#figure-2b---patterns-profile-plot)
   2. [Supplemental Figure 2A](https://github.niaid.nih.gov/chi/h5n1#supplemental-figure-2a)
   3. [Figure 2C](https://github.niaid.nih.gov/chi/h5n1#btm-enrichment-in-patterns-genes-figure-2c)
7. [Pattern discovery in post-vaccination profiles of flow cytometry data](https://github.niaid.nih.gov/chi/h5n1#pattern-discovery-in-post-vaccination-profiles-of-flow-cytometry-data)
   1. [Figure 2D](https://github.niaid.nih.gov/chi/h5n1#figure-2d)
   2. [Figure 2E](https://github.niaid.nih.gov/chi/h5n1#figure-2e)
   3. [Figure 2F](https://github.niaid.nih.gov/chi/h5n1#figure-2f)
   4. [Supplemental Figure 3](https://github.niaid.nih.gov/chi/h5n1#supplemental-figure-3)
8. [Find signature for adjuvant status prediction](https://github.niaid.nih.gov/chi/h5n1#find-signature-for-adjuvant-status-prediction)
   1. [Figure 3C](https://github.niaid.nih.gov/chi/h5n1#figure-3c)
   2. [Figure 3D](https://github.niaid.nih.gov/chi/h5n1#figure-3d)
   3. [Supplemental Figure 4B](https://github.niaid.nih.gov/chi/h5n1#supplemental-figure-4b)
   4. [Supplemental Figure 4C](https://github.niaid.nih.gov/chi/h5n1#supplemental-figure-4c)
   5. [Figure 3E](https://github.niaid.nih.gov/chi/h5n1#figure-3e)
   6. [Figure 3B](https://github.niaid.nih.gov/chi/h5n1#figure-3b)
   7. [Figure 3F&G](https://github.niaid.nih.gov/chi/h5n1#figure-3f-and-3g)
9. [Gene Expression PAXgene data processing](https://github.niaid.nih.gov/chi/h5n1#gene-expression-paxgene-data-processing)
10. [Baseline Data Analysis](https://github.niaid.nih.gov/chi/h5n1#baseline-data-analysis)
    1. [Figure 4A](https://github.niaid.nih.gov/chi/h5n1#figure-4a-combining-btm-enrichment-results-from-pbmc-and-whole-blood-samples)
    2. [Figure 4B](https://github.niaid.nih.gov/chi/h5n1#figure-4b)
    3. [Figure 4C](https://github.niaid.nih.gov/chi/h5n1#figure-4c)
11. [Unbliding Results](https://github.niaid.nih.gov/chi/h5n1#unblinding-results-figure-5-suppl-figure-5)
    1. [Figure 5A](https://github.niaid.nih.gov/chi/h5n1#figure-5a)
    2. [Figure 5B](https://github.niaid.nih.gov/chi/h5n1#figure-5b)
    3. [Figure 5C](https://github.niaid.nih.gov/chi/h5n1#figure-5c-figuresifn_signature)
12. [Emory data analysis blindly predicting adjuvant status](https://github.niaid.nih.gov/chi/h5n1#emory-data-analysis-blindly-predicting-adjuvant-status)
    1. [Supplemental Figure 5D](https://github.niaid.nih.gov/chi/h5n1#supplemental-figure-5d)
13. [Tfh cells data analysis](https://github.niaid.nih.gov/chi/h5n1#tfh-cells-data-analysis)
    1. [Figure 6A](https://github.niaid.nih.gov/chi/h5n1#figure-6a)
    2. [Figure 6B](https://github.niaid.nih.gov/chi/h5n1#figure-6b)
    3. [Figure 6C](https://github.niaid.nih.gov/chi/h5n1#figure-6c)
    4. [Figure 6D](https://github.niaid.nih.gov/chi/h5n1#figure-6d)
    5. [Supplemental Figure 7D](https://github.niaid.nih.gov/chi/h5n1#supplemental-figure-7d)
14. [SOMAscan Data Analysis](https://github.niaid.nih.gov/chi/h5n1#somascan-data-analysis)
    1. [Figure 7A](https://github.niaid.nih.gov/chi/h5n1#figure-7a)
    2. [Figure 7B](https://github.niaid.nih.gov/chi/h5n1#figure-7b)

# Environment Setup

## 

## Data and Scripts

Original and processed data, results, figures, and scripts can be downloaded from <https://nih.box.com/s/w2tubby2kz31ssesxna90o625qudz1gd>

## Build and Activate Conda Environment

conda env create --file=env.yaml  
conda activate h5n1  
  
# Install eNetXlporer through install.packages()  
install.packages("eNetXplorer")

Most of the scripts initialize the environment at the very beginning however, if not then run:  
source("SCRIPTS/0\_initialize.r")

To run a script using singularity container:

$ Rscript --vanilla <script path>

## Singularity container

Singularity container with the R packages that were used for the analysis are available in the data folder mentioned above however, it can also be downloaded from:

<https://cloud.sylabs.io/library/_container/5d728283cb093ee1142b8810>

or

$ singularity pull library://rohitfarmer/default/h5n1\_workflow:sha256.240542df22ff14f6e49b58f43065e26b1d2d8a1103840d869f6034a5426eaeca

To run a script using singularity container:

$ singularity exec h5n1-container.sif Rscript –vanilla <script path>

*Note: Singularity container also has Neovim installed that can be used with Nvim-R plugin installed on the host system. Recipe file that was used to build the singularity container is available at* <https://github.com/rohitfarmer/singularity-defs/tree/master/H5N1>

*Note: For both conda environment and Singularity container project folder is the working directory. Therefore, all the scripts executes relative to the project folder path.*

# Titers

## 

## Figure 1A

URL to the Shinyapp that produces figure 1A H5N1/MANUSCRIPT/Figures/Figure1\_url\_180322.txt

## Figure 1B

Rscript --vanilla SCRIPTS/titers/mn\_titer\_profiles.r

## Supplemental Figure 1

Fig 1A.

Rscript --vanilla SCRIPTS/titers/titer\_response\_rate.r

Fig 1B.

Rscript --vanilla SCRIPTS/titers/hai\_titer\_profiles.r

Fig 1C.

Rscript --vanilla SCRIPTS/titers/mn\_titer\_peak.r

# 

# Pattern profiles of clinical CBC data and Luminex

## 

## Figures 1 C,D,E

Fig 1C, D.

Rscript --vanilla SCRIPTS/profiles/Monocytes\_figure.r  
Rscript --vanilla SCRIPTS/profiles/Neutrophils\_figure.r

Fig 1E.

Rscript --vanilla SCRIPTS/profiles/IP10\_figure.r

# Pattern simulation (Figure 2A)

Rscript --vanilla SCRIPTS/pattern\_sim/pattern\_simulation.r

## Gene Expression PBMC data processing. First, the CEL files were processed with Power Tools RNA-sketch algorithm.

Rscript --vanilla SCRIPTS/MA/processing\_pbmc/apt.config.r  
Rscript --vanilla SCRIPTS/MA/processing\_pbmc/apt.call.r

*Note: Power Tools are not included in the singularity container or in the conda environment file.*

## 

## Create ExpressionSet from APT output

Rscript --vanilla SCRIPTS/MA/processing\_pbmc/eset.config.r

Rscript --vanilla SCRIPTS/MA/processing\_pbmc/eset.call.r

## 

## Probesets to genes mapping

### 

### Convert the table to probeset-gene mapping

Rscript --vanilla SCRIPTS/MA/annotation/affy\_hugene-2\_1-st\_annotation.r

### 

### Select the best probeset for a gene

Rscript --vanilla SCRIPTS/MA/annotation/generate\_ps2gene\_map.config.r  
Rscript --vanilla SCRIPTS/MA/annotation/generate\_ps2gene\_map.call.r

## Data post processing

### 

### We found that two samples were switched. This is to correct it.

Rscript --vanilla SCRIPTS/MA/filtering\_pbmc/switch.samples/switch.samples.call.r

### Apply different filtering to samples and genes. The probesets mapped to genes.

Rscript --vanilla SCRIPTS/MA/filtering\_pbmc/samples.clean\_genes.all/filtering.r  
Rscript --vanilla SCRIPTS/MA/filtering\_pbmc/samples.clean\_genes.iqr/filtering.r  
Rscript --vanilla SCRIPTS/MA/filtering\_pbmc/samples.all\_genes.all/filtering.r  
Rscript --vanilla SCRIPTS/MA/filtering\_pbmc/samples.all\_genes.iqr/filtering.r

### 

### Calculate fold change from day 0

Rscript --vanilla SCRIPTS/MA/calculate\_d0\_fc/calculate\_d0\_fc\_pbmc.r

# Pattern discovery in post-vaccination profiles of gene expression

## 

## Profiles clustering with DIANA

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_discovery.r

## Cut the dendrogram tree at different levels and detect stable clusters

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/patterns\_cutTree\_stable.r

## 

## Filter the patterns

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_filter.r

## 

## Summarize patterns stats

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/patterns\_stats.r

## 

## Figure 2B - patterns profile plot

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/plot\_patterns.r

## 

## Expand list of pattern signature genes

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_filter\_expanded.r

## 

## Compute correlations and clean up the genes

### 

### Supplemental Figure 2A

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_expanded\_genes\_cor.r  
Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_expanded\_genes\_clean.r

## 

## Compute subject scores for each pattern

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/patterns\_to\_subjects.r

## Output the table of genes (with annotations) for each pattern

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_genes\_output.r

## 

## BTM enrichment in patterns genes.

## Figure 2C

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_BTM\_enrichment.r

## 

## Add data for subject s10 and update the score matrix

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/s10\_peaks\_assessment.r  
Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_scores\_in\_samples\_GE\_incl.s10.r

# 

# Pattern discovery in post-vaccination profiles of flow cytometry data

FlowJo software was used to export flow data.

## 

## Generate Trajectory Matrix

Rscript --vanilla SCRIPTS/Flow\_10c/Flow\_10c\_TrajMatrix\_v2.R  
Rscript --vanilla SCRIPTS/Flow\_10c/Flow\_10c\_TrajMatrix\_v3.R

## Genetrate Trajectory Clusters

Rscript --vanilla SCRIPTS/Flow\_10c/Flow\_10c\_TrajCluster\_v2.R  
Rscript --vanilla SCRIPTS/Flow\_10c/Flow\_10c\_TrajCluster\_v3.R

## Figure 2D

Rscript --vanilla SCRIPTS/Flow/pattern\_figures/plot\_flow\_patters\_only.r

## 

## Figure 2E

Rscript --vanilla SCRIPTS/Flow/pattern\_figures/pattern\_flow\_ann\_heatmap.r

## Figure 2F

Rscript --vanilla SCRIPTS/MA/pattern\_discovery/pattern\_scores\_GE\_flow\_heatmap.r

## 

## Supplemental Figure 3

Rscript --vanilla SCRIPTS/Flow\_10c/Flow\_10c\_TrajCluster\_QM.R

# 

# Find signature for adjuvant status prediction

## 

## Figure 3C

Rscript --vanilla SCRIPTS/adjuvant\_prediction/2peaks\_pca\_2clusters.r

## 

## Figure 3D

Rscript --vanilla SCRIPTS/adjuvant\_prediction/ip10\_2clusters\_compare.r

## 

## Supplemental Figure 4B

Rscript --vanilla SCRIPTS/adjuvant\_prediction/cytokines\_2clusters\_compare.r

## 

## Supplemental Figure 4C

Rscript --vanilla SCRIPTS/MA/baseline/Gb13\_vs\_GbWB11.r

## 

## Figure 3E

Rscript --vanilla SCRIPTS/adjuvant\_prediction/ip10\_2peak\_scores.r  
Rscript --vanilla SCRIPTS/adjuvant\_prediction/2peaks\_pca\_final\_heamap.r

### 

### Generate input data:

Rscript --vanilla SCRIPTS/eNetXplorer/eNet\_input\_r1.r

## Elastic net models

### Generate input data:

Rscript --vanilla SCRIPTS/eNetXplorer/eNet\_input\_r1.r

Rscript --vanilla SCRIPTS/eNetXplorer/eNet\_input\_r2.r

Rscript --vanilla SCRIPTS/eNetXplorer/eNet\_input\_r3.r

### Run eNetXplorer:

Rscript --vanilla SCRIPTS/eNetXplorer/eNetXplorer\_R1\_180530.R

Rscript --vanilla SCRIPTS/eNetXplorer/eNetXplorer\_R2\_180530.R

Rscript --vanilla SCRIPTS/eNetXplorer/eNetXplorer\_R3\_180530.R

### Figure 3B

Rscript --vanilla SCRIPTS/eNet\_figures/enet\_plots\_R1.r

### Figure 3F and 3G

Rscript --vanilla SCRIPTS/eNet\_figures/enet\_plots\_R3.r

Rscript --vanilla SCRIPTS/eNet\_figures/enet\_plots\_R2.r

###### **Gene Expression PAXgene Data Processing**

## Data Post Processing

We found that two samples were switched. This is to correct it.

Rscript --vanilla SCRIPTS/MA/filtering\_pax/switch.samples/switch.samples.call.r

## Apply different filtering to samples and genes. The probesets mapped to genes.

Rscript --vanilla SCRIPTS/MA/filtering\_pax/filtering.r

## Calculate fold change from day 0

Rscript –vanilla SCRIPTS/MA/calculate\_d0\_fc/calculate\_d0\_fc\_pax.r

# Baseline Data Analysis

## Preparing PBMC day 0 samples

Rscript --vanilla SCRIPTS/MA/baseline\_pbmc/d0\_filter.r

## WGCNA clustering of PBMC samples

Rscript --vanilla SCRIPTS/MA/baseline\_pbmc/d0\_wgcna.r

Rscript --vanilla SCRIPTS/MA/baseline\_pbmc/d0\_wgcna\_output.r

## BTM enrichment analysis of data from PBMC samples

Rscript --vanilla SCRIPTS/MA/baseline\_pbmc/d0\_wgcna\_BTM\_enrichment.r

## Preparing whole blood (PAXgene) day 0 samples

Rscript --vanilla SCRIPTS/MA/baseline\_pax/d0\_filter.r

## WGCNA clustering of whole blood samples

Rscript --vanilla SCRIPTS/MA/baseline\_pax/d0\_wgcna.r

Rscript --vanilla SCRIPTS/MA/baseline\_pax/d0\_wgcna\_output.r

## BTM enrichment analysis of data from whole blood samples

Rscript --vanilla SCRIPTS/MA/baseline\_pax/d0\_wgcna\_BTM\_enrichment.r

## FIgure 4A. Combining BTM enrichment results from PBMC and whole blood samples. Also Suppl. Figure 6A

Rscript --vanilla SCRIPTS/MA/baseline/plot\_BTM\_pbmc\_pax.r

## Elastic net models for baseline prediction

### Generate input data:

Rscript --vanilla SCRIPTS/eNetXplorer/eNet\_input\_r6.r

### Run eNetXplorer:

Rscript --vanilla SCRIPTS/eNetXplorer/eNetXplorer\_R6\_181022.R

### Figure 4B

Rscript --vanilla SCRIPTS/eNet\_figures/enet\_plots\_R6.r

### Figure 4C

Rscript --vanilla SCRIPTS/MA/baseline/IFN.gene\_overlap\_figure.r

# Unblinding results

## Figure 5B

Rscript --vanilla SCRIPTS/adjuvant\_prediction/pattern\_gene\_time\_score\_sel.subject.r

Rscript --vanilla SCRIPTS/adjuvant\_prediction/pattern\_flow\_time\_score\_sel.subject.r

Rscript --vanilla SCRIPTS/adjuvant\_prediction/IP10\_time\_score\_sel.subject.r

## Figure 5C

Rscript --vanilla SCRIPTS/MA/baseline/GbWB11.d0\_vs\_MN.d28.r

# Emory data analysis blindly predicting adjuvant status

## Process Data to Generate Espression Set

source("SCRIPTS/Emory/emory\_data.r")

## Get Annotations

source("SCRIPTS/Emory/get\_ann.r")

## Map Probes to Genes

source("SCRIPTS/Emory/probe2gene.r")

## Claculate 2 Peak Scores

source("SCRIPTS/Emory/2peak\_scores.r")

## Supplemental Figure 5D

source("SCRIPTS/Emory/adjuvant\_prediction.r")