H5N1 Data Analysis Workflow

Table of Contents

[1 Table of Contents 1](#_Toc20482543)

[2 Environment Setup 4](#_Toc20482544)

[2.1 Data and Scripts 4](#_Toc20482545)

[2.2 Build and Activate Conda Environment 4](#_Toc20482546)

[2.3 Singularity container 4](#_Toc20482547)

[3 Titers and Pattern Profiles of Clinical CBC and Luminex Data 5](#_Toc20482548)

[3.1 Figures 1B, C, D, and E 5](#_Toc20482549)

[3.2 Supplemental Figure 1 5](#_Toc20482550)

[4 Gene Expression (PBMC) Data Processing 5](#_Toc20482551)

[4.1 Expression Set Creation from APT Output 6](#_Toc20482552)

[4.2 Probesets to Genes Mapping 6](#_Toc20482553)

[4.2.1 Convert the Table to Probeset-Gene Mapping 6](#_Toc20482554)

[4.2.2 Select the Best Probeset for a Gene 6](#_Toc20482555)

[4.3 Data post processing 6](#_Toc20482556)

[4.3.1 Correct the Switched Samples 6](#_Toc20482557)

[4.3.2 Apply Different Filtering to Samples and Genes 6](#_Toc20482558)

[4.3.3 Calculate Fold Change from Day 0 6](#_Toc20482559)

[5 Pattern Discovery in Post-Vaccination Profiles of PBMC Gene Expression Data 6](#_Toc20482560)

[5.1 Profiles Clustering with DIANA 6](#_Toc20482561)

[5.2 Stable Cluster/Pattern Detection by Tree Cutting 6](#_Toc20482562)

[5.3 Pattern Filtering 7](#_Toc20482563)

[5.4 Patterns Stats Summary 7](#_Toc20482564)

[5.5 Figure 2A 7](#_Toc20482565)

[5.6 Figure 2B 7](#_Toc20482566)

[5.7 Expanding the List of Pattern Signature Genes 7](#_Toc20482567)

[5.8 Compute Correlations and Clean up Gene List 7](#_Toc20482568)

[5.8.1 Supplemental Figure 2A 7](#_Toc20482569)

[5.9 Compute Subject Scores for Each Pattern 7](#_Toc20482570)

[5.10 Table of Genes (with annotations) for Each Pattern 7](#_Toc20482571)

[5.11 Figure 2C 7](#_Toc20482572)

[5.12 Add Data for Subject s10 and Update the Score Matrix 8](#_Toc20482573)

[6 Pattern Discovery in Post-Vaccination Profiles of Flow Cytometry Data 8](#_Toc20482574)

[6.1 Generate Trajectory Matrix 8](#_Toc20482575)

[6.2 Genetrate Trajectory Clusters 8](#_Toc20482576)

[6.3 Figure 2D, 2E, and 2F 8](#_Toc20482577)

[6.4 Supplemental Figure 3 8](#_Toc20482578)

[7 Find Signature for Adjuvant Status Prediction using PBMC Data 8](#_Toc20482579)

[7.1 Figure 3C 8](#_Toc20482580)

[7.2 Figure 3D 9](#_Toc20482581)

[7.3 Supplemental Figure 4B 9](#_Toc20482582)

[7.4 Supplemental Figure 4C 9](#_Toc20482583)

[7.5 Figure 3E 9](#_Toc20482584)

[7.6 Elastic Net Models 9](#_Toc20482585)

[7.6.1 Generate Input Data 9](#_Toc20482586)

[7.6.2 Run eNetXplorer 9](#_Toc20482587)

[7.6.3 Figure 3B 9](#_Toc20482588)

[7.6.4 Figure 3F, and 3G 9](#_Toc20482589)

[8 Gene Expression (Whole Bloood/PAXgene) Data Processing 10](#_Toc20482590)

[8.1 Data Post Processing 10](#_Toc20482591)

[8.2 Apply Different Filtering to Samples and Genes 10](#_Toc20482592)

[8.3 Calculate Fold Change from Day 0 10](#_Toc20482593)

[9 Baseline Data Analysis 10](#_Toc20482594)

[9.1 Preparing PBMC Day 0 Samples 10](#_Toc20482595)

[9.2 WGCNA Clustering of PBMC Samples 10](#_Toc20482596)

[9.3 BTM Enrichment Analysis of Data from PBMC Samples 10](#_Toc20482597)

[9.4 Preparing Whole Blood (PAXgene) Day 0 Samples 10](#_Toc20482598)

[9.5 WGCNA Clustering of Whole Blood Samples 10](#_Toc20482599)

[9.6 BTM Enrichment Analysis of Data from Whole Blood Samples 10](#_Toc20482600)

[9.7 FIgure 4A 11](#_Toc20482601)

[9.8 Elastic Net Models for Baseline Prediction 11](#_Toc20482602)

[9.8.1 Generate Input Data 11](#_Toc20482603)

[9.8.2 Run eNetXplorer 11](#_Toc20482604)

[9.8.3 Figure 4B 11](#_Toc20482605)

[9.8.4 Figure 4C 11](#_Toc20482606)

[10 Unblinding Adjuvant Status 11](#_Toc20482607)

[10.1 Figure 5B 11](#_Toc20482608)

[10.2 Figure 5C 11](#_Toc20482609)

[11 Blindly Predicting Adjuvant Status using Data from Emory University 11](#_Toc20482610)

[11.1 Process Data to Generate Espression Set 11](#_Toc20482611)

[11.2 Get Annotations 11](#_Toc20482612)

[11.3 Map Probes to Genes 12](#_Toc20482613)

[11.4 Claculate 2 Peak Scores 12](#_Toc20482614)

[11.5 Supplemental Figure 5 12](#_Toc20482615)

[12 SOMAscan Data Analysis 12](#_Toc20482616)

[12.1 DATA Preparation 12](#_Toc20482617)

[12.2 Figure 6A 12](#_Toc20482618)

[12.3 Figure 6B 12](#_Toc20482619)

# Environment Setup

## Data and Scripts

Original and processed data, results, figures, and scripts can be downloaded from <https://nih.box.com/v/chi-010-h5n1-tsang>.

## 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:

$ 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 and Pattern Profiles of Clinical CBC and Luminex Data

## Figures 1B, C, D, and E

Fig 1B.

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

**Fig 1C.**

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

**Fig 1D.**

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

**Fig 1E.**

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

# Gene Expression (PBMC) Data Processing

First, the CEL files were processed using Affymetrix power tools.

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.*

## Expression Set Creation 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

### Correct the Switched Samples

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

### Apply Different Filtering to Samples and 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 PBMC Gene Expression Data

## Profiles Clustering with DIANA

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

## Stable Cluster/Pattern Detection by Tree Cutting

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

## Pattern Filtering

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

## Patterns Stats Summary

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

## Figure 2A

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

## Figure 2B

**Patterns Profile Plot**

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

## Expanding the List of Pattern Signature Genes

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

## Compute Correlations and Clean up Gene List

### 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

## Table of Genes (with annotations) for Each Pattern

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

## Figure 2C

**BTM Enrichment In Patterns Genes**

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, 2E, and 2F

Fig 2D.

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

Fig 2E.

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

Fig 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 using PBMC Data

## 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

## 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

**Fig 3F.**

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

**Fig 3G.**

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

# Gene Expression (Whole Bloood/PAXgene) Data Processing

## Data Post Processing

*Note: 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

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 Adjuvant Status

## 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

# Blindly Predicting Adjuvant Status using Data from Emory University

## Process Data to Generate Espression Set

Rscript --vanilla SCRIPTS/Emory/emory\_data.r

## Get Annotations

Rscript --vanilla SCRIPTS/Emory/get\_ann.r

## Map Probes to Genes

Rscript --vanilla SCRIPTS/Emory/probe2gene.r

## Claculate 2 Peak Scores

Rscript --vanilla SCRIPTS/Emory/2peak\_scores.r

## Supplemental Figure 5

Rscript --vanilla SCRIPTS/Emory/adjuvant\_prediction.r  
Rscript --vanilla SCRIPTS/MA/baseline/analyze\_IFNg\_genes\_emory\_171208.r  
Rscript --vanilla SCRIPTS/Emory/emory\_pattern\_scores\_wrong\_subjects.r  
Rscript --vanilla SCRIPTS/Emory/ emory\_ip10\_wrong\_sbujects.r

# SOMAscan Data Analysis

## DATA Preparation

Rscript --vanilla SCRIPTS/SOMAscan/Data\_Processing.R  
Rscript --vanilla SCRIPTS/SOMAscan/Data\_Normalization.R

## Figure 6A

Rscript --vanilla SCRIPTS/eNetXplorer/eNet\_input\_r8.r  
Rscript --vanilla SCRIPTS/eNetXplorer/eNetXplorer\_R8\_181022.r  
Rscript --vanilla SCRIPTS/eNet\_figures/enet\_plots\_R8.r

## Figure 6B

Rscript --vanilla SCRIPTS/SOMAscan/soma\_BTM\_enrichment.r  
Rscript --vanilla SCRIPTS/SOMAscan/soma\_enrichment\_heatmap.r