Data Analysis Workflow

1. Table of Contents

[2 Environment Setup 4](#_Toc75357157)

[2.1 Singularity container 4](#_Toc75357158)

[3 Titers and Pattern Profiles of Clinical CBC and Luminex Data 4](#_Toc75357159)

[3.1 Figure 1 4](#_Toc75357160)

[3.2 Supplemental Figure 9 4](#_Toc75357161)

[4 Gene Expression (PBMC) Data Processing 5](#_Toc75357162)

[4.1 Expression Set Creation from APT Output 5](#_Toc75357163)

[4.2 Probesets to Genes Mapping 5](#_Toc75357164)

[4.2.1 Convert the Table to Probeset-Gene Mapping 5](#_Toc75357165)

[4.2.2 Select the Best Probeset for a Gene 5](#_Toc75357166)

[4.3 Data post processing 5](#_Toc75357167)

[4.3.1 Correct the Switched Samples 5](#_Toc75357168)

[4.3.2 Apply Different Filtering to Samples and Genes 5](#_Toc75357169)

[4.3.3 Calculate Fold Change from Day 0 5](#_Toc75357170)

[5 Pattern Discovery in Post-Vaccination Profiles of PBMC Gene Expression Data 6](#_Toc75357171)

[5.1 Profiles Clustering with DIANA 6](#_Toc75357172)

[5.2 Stable Cluster/Pattern Detection by Tree Cutting 6](#_Toc75357173)

[5.3 Pattern Filtering 6](#_Toc75357174)

[5.4 Patterns Stats Summary 6](#_Toc75357175)

[5.5 Figure 2A 6](#_Toc75357176)

[5.6 Figure 2B 6](#_Toc75357177)

[5.7 Expanding the List of Pattern Signature Genes 6](#_Toc75357178)

[5.8 Compute Correlations and Clean up Gene List 6](#_Toc75357179)

[5.8.1 Supplemental Figure 1A 6](#_Toc75357180)

[5.9 Compute Subject Scores for Each Pattern 7](#_Toc75357181)

[5.10 Table of Genes (with annotations) for Each Pattern 7](#_Toc75357182)

[5.11 Figure 2C 7](#_Toc75357183)

[5.12 Add Data for Subject s10 and Update the Score Matrix 7](#_Toc75357184)

[6 Pattern Discovery in Post-Vaccination Profiles of Flow Cytometry Data 7](#_Toc75357185)

[6.1 Generate Trajectory Matrix 7](#_Toc75357186)

[6.2 Generate Trajectory Clusters 7](#_Toc75357187)

[6.3 Figure 2D, 2E, and 2F 7](#_Toc75357188)

[6.4 Supplemental Figure 2 8](#_Toc75357189)

[7 Find Signature for Adjuvant Status Prediction using PBMC Data 8](#_Toc75357190)

[7.1 Figure 3C 8](#_Toc75357191)

[7.2 Figure 3D 8](#_Toc75357192)

[7.3 Supplemental Figure 3B 8](#_Toc75357193)

[7.4 Supplemental Figure 3D 8](#_Toc75357194)

[7.5 Figure 3E 8](#_Toc75357195)

[7.6 Elastic Net Models 8](#_Toc75357196)

[7.6.1 Generate Input Data 8](#_Toc75357197)

[7.6.2 Run eNetXplorer 8](#_Toc75357198)

[7.6.3 Figure 3B 9](#_Toc75357199)

[7.6.4 Figure 5A, and 5b 9](#_Toc75357200)

[8 Gene Expression (Whole Bloood/PAXgene) Data Processing 9](#_Toc75357201)

[8.1 Data Post Processing 9](#_Toc75357202)

[8.2 Apply Different Filtering to Samples and Genes 9](#_Toc75357203)

[8.3 Calculate Fold Change from Day 0 9](#_Toc75357204)

[9 Baseline Data Analysis 9](#_Toc75357205)

[9.1 Preparing PBMC Day 0 Samples 9](#_Toc75357206)

[9.2 WGCNA Clustering of PBMC Samples 9](#_Toc75357207)

[9.3 BTM Enrichment Analysis of Data from PBMC Samples 10](#_Toc75357208)

[9.4 Preparing Whole Blood (PAXgene) Day 0 Samples 10](#_Toc75357209)

[9.5 WGCNA Clustering of Whole Blood Samples 10](#_Toc75357210)

[9.6 BTM Enrichment Analysis of Data from Whole Blood Samples 10](#_Toc75357211)

[9.7 FIgure 4C 10](#_Toc75357212)

[9.8 Elastic Net Models for Baseline Prediction 10](#_Toc75357213)

[9.8.1 Generate Input Data 10](#_Toc75357214)

[9.8.2 Run eNetXplorer 10](#_Toc75357215)

[9.8.3 Supplemental Figure 5C 10](#_Toc75357216)

[9.8.4 Figure 4E 10](#_Toc75357217)

[10 Unblinding Adjuvant Status 11](#_Toc75357218)

[10.1 Supplemental Figure 4A 11](#_Toc75357219)

[10.2 Figure 4F 11](#_Toc75357220)

[11 Blindly Predicting Adjuvant Status using Data from Emory University 11](#_Toc75357221)

[11.1 Process Data to Generate Espression Set 11](#_Toc75357222)

[11.2 Get Annotations 11](#_Toc75357223)

[11.3 Map Probes to Genes 11](#_Toc75357224)

[11.4 Calculate 2 Peak Scores 11](#_Toc75357225)

[11.5 Supplemental Figure 4 11](#_Toc75357226)

[12 SOMAscan Data Analysis 12](#_Toc75357227)

[12.1 DATA Preparation 12](#_Toc75357228)

[12.2 Supplemental Figure 6A 12](#_Toc75357229)

[12.3 Supplemental Figure 6B 12](#_Toc75357230)

[13 Day 100 CITE-seq analysis 12](#_Toc75357231)

[13.1 Microarray Differential Expression of Persistence Patterns; day 100 vs. Baseline Random Intercept Model 12](#_Toc75357232)

[13.2 CITE-seq day 100 Normalization and Protein Based Clustering 13](#_Toc75357233)

[13.2.1 CITE-seq Differential Expression Analysis and Enrichment of Persistence Genes Within Cell Types 13](#_Toc75357234)

[13.2.2 Test Leading Edge Genes from Single Cell day 100 H5N1 Vaccines in H1N1 High vs Low Responders 13](#_Toc75357235)

[13.3 Analysis of persistence signals in H1N1 vaccine day 70 vs baseline 13](#_Toc75357236)

# Environment Setup

## Singularity container

Singularity container inside the project folder contains the specific R version along with the packages that were used for data analysis. Folder paths are relative to the project root folder.

To run a script using singularity container:

$ singularity exec h5n1\_workflow.sif Rscript –vanilla <script path>

# Titers and Pattern Profiles of Clinical CBC and Luminex Data

## Figure 1

**B.**

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

**C.**

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

**D.**

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

**E.**

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

## Supplemental Figure 9

**A.**

Rscript --vanillaSCRIPTS/titers/titer\_response\_rate.r

**B.**

Rscript --vanillaSCRIPTS/titers/hai\_titer\_profiles.r

**C.**

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.*

## Expression Set Creation from APT Output

Rscript --vanillaSCRIPTS/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 1A

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.R

## Generate Trajectory Clusters

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

## Figure 2D, 2E, and 2F

**D.**

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

**E.**

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

**F.**

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

## Supplemental Figure 2

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 3B

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

## Supplemental Figure 3D

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 5A, and 5b

**A.**

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

**B.**

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 4C

**Combining BTM enrichment results from PBMC and whole blood samples. Also Suppl. Figure 3C**

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

### Supplemental Figure 5C

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

### Figure 4E

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

# Unblinding Adjuvant Status

## Supplemental Figure 4A

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 4F

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

## Calculate 2 Peak Scores

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

## Supplemental Figure 4

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

## Supplemental 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

## Supplemental Figure 6B

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

# Day 100 CITE-seq analysis

To run analysis, install selected functions from scglmmr R package by installing the included (in the root folder of the repository) package file locally.

install.packages(file.path(path\_to\_package,"scglmmr"), repos = NULL, type = "source")

***Note: for the subsequent scripts change to “H5\_d100\_public-master” folder in the root folder of the repository. These scripts are not part of the singularity container and folder structure mentioned above.***

## Microarray Differential Expression of Persistence Patterns; day 100 vs. Baseline Random Intercept Model

***Note: Microarray analysis section uses R 3.6.1.***

**Format Data**

source("microarray\_analysis\_d100/1\_save\_eset\_object\_as\_dataframe.r")

**Variance Partition and Mixed GLM Workflow**

source("microarray\_analysis\_d100/2\_V4array\_d100\_lme4\_PERSISTENCE\_V4.R")

**Run Separate Models for Adjuvant and Non-Adjuvant Group**

source("microarray\_analysis\_d100/2\_array\_adjonly\_d100\_lme4\_PERSISTENCE\_V3.R")

source("microarray\_analysis\_d100/2\_nonadj\_array\_d100\_lme4\_PERSISTENCE\_V3.R")

**Save Results Table Combined**

source("microarray\_analysis\_d100/3\_make\_combined\_table.r")

## CITE-seq day 100 Normalization and Protein Based Clustering

***Note: Single cell analysis section uses R 3.5.3.***

**Normalize Single Cells from Baseline and Day 100, Cluster Annotate and run UMAP**

source("2\_clustering/3\_umap.r")

## CITE-seq Differential Expression Analysis and Enrichment of Persistence Genes Within Cell Types

**Pseudobulk Workflow**

source("3\_pseudobulk\_de\_workflow/2\_h5\_d0\_vd\_d100\_pseudobulk\_de.r")

source("3\_pseudobulk\_de\_workflow/3\_figure\_generation\_h5d100cite\_scglmmrv2.r")

## Test Leading Edge Genes from Single Cell day 100 H5N1 Vaccines in H1N1 High vs Low Responders

source("3\_pseudobulk\_de\_workflow/4\_h1\_highresponder\_baseline\_persistencesig\_test.r")

source("3\_pseudobulk\_de\_workflow/5\_h1\_highresponder\_baselinepersistence\_figure\_generation.r")

## Analysis of persistence signals in H1N1 vaccine day 70 vs baseline

source("d70\_H1/day70\_h1\_persistencegenes.r")