2023 SEATRAC Hackday

A purple circle with text and lungs

Description automatically generated

Andrew Fiore-Gartland, Fred Hutchinson Cancer Center

Fergal Duffy, Seattle Children’s Research Institute

Kim Dill-McFarland, U of Washington

Shuyi Ma, Seattle Children's Research Institute

<https://github.com/FredHutch/seatrac-hackday-2023>

TAble of Contents

General 1

Goal 1

Code of conduct 1

Agenda 1

Datasets 1

Foreman *et al* 2023: bulk RNAseq from sorted cells in NHP Mtb challenge model 2

Gideon *et al* 2022: scRNAseq from granulomas in NHP Mtb challenge model 2

Darrah *et al* 2023: scRNAseq from BAL in NHP Mtb challenge (BCG route, correlates of protection) 3

Liu *et al* 2023: whole-blood bulk RNAseq from NHP Mtb challenge (BCG route and IV BCG dose) 4

TUTORIALS 4

Bulk RNAseq differential expression 4

scRNAseq ordinations 4

Network analysis 5

Results 5

Liu et al modules by sex 5

Darrah et al, 2020. B cell subset (downsampled) 6

Treg cell pseudo-bulk DEGs 6

# General

Wednesday, December 13, 2023, 9 AM to 3 PM (Pacific Time)

## Goal

Collaborating in small groups to develop new analytical insights from 4 published bulk and single-cell RNAseq datasets that profile NHP host gene expression during Mtb infection and BCG vaccination.

## Code of conduct

SEATRAC is dedicated to providing a harassment-free learning experience for everyone regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion. We do not tolerate harassment of participants in any form. Please reference this [Code of Conduct](https://oceanhackweek.org/about/code-of-conduct.html) adopted by Ocean Hackweek that we will use for our TB Hackday.

## Agenda

8:30 Gather with coffee and light breakfast

9:05 Opening remarks (Shuyi Ma, Andrew Fiore-Gartland)

9:15 Introduction to the datasets (Andrew Fiore-Gartland)

9:45 Small group session: devise questions, hypotheses and analysis plans

10:30 Hands-on tutorial: Bulk RNA seq differential expression and single-cell Seurat workflow (Kim Dill-McFarland)

11:30 Working lunch

12:30 Hands-on tutorial: Gene/cell clustering and network analysis (Shuyi Ma)

1:30 Open hack time

2:30 Small group readouts: Each group shares preliminary figures, analysis plans, code chunks

## Datasets

* Link to aggregated datasets on [Figshare](https://figshare.com/articles/dataset/SEATRAC_TB_Hackday_2023/24425053)
* NOTE: For the Gideon et al. and Darrah et al. single-cell datasets there is also a pseudo-bulk dataset that was created by summing raw counts across cells within a sample and an annotated cell type. Cell annotations from the original data were used for the summation. The resulting CSV contains counts for each sample, cell type and gene.

### Foreman *et al* 2023: bulk RNAseq from sorted cells in NHP Mtb challenge model

Foreman TW, Nelson CE, Sallin MA, Kauffman KD, Sakai S, Otaizo-Carrasquero F, Myers TG, Barber DL. CD30 co-stimulation drives differentiation of protective T cells during Mycobacterium tuberculosis infection. J Exp Med. 2023 Aug 7;220(8):e20222090. doi: 10.1084/jem.20222090. Epub 2023 Apr 25. PMID: 37097292; PMCID: PMC10130742.

* PDF: <https://rupress.org/jem/article-pdf/220/8/e20222090/1451373/jem_20222090.pdf>; ([github](https://github.com/FredHutch/seatrac-hackday-2023/blob/main/foreman_etal/Foreman%20et%20al%20CD30%20drives%20differentiation%2C%202023.pdf))
* MS: <https://rupress.org/jem/article/220/8/e20222090/214054/CD30-co-stimulation-drives-differentiation-of>
* DATA: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE227653> (M mulatta data)
* <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE228114> for the superset that contains both the NHP and mouse data for that study
* <https://github.com/FredHutch/seatrac-hackday-2023/tree/main/foreman_etal>
* **DATA SUMMARY**: N=4 Rhesus macaques, Mtb. challenge (40–80 CFU of Mtb-Erdman-mCherry and euthanized 6–7 wk after infection); bulk RNAseq from FACS sorted T cells; [bulk RNAseq count data](https://github.com/FredHutch/seatrac-hackday-2023/blob/main/foreman_etal/GSE227653_TPM_all.csv.gz) is stored for CD4 and CD8 T cells in units of transcript per million (TPM).

### Gideon *et al* 2022: scRNAseq from granulomas in NHP Mtb challenge model

Gideon HP, Hughes TK, Tzouanas CN, Wadsworth MH 2nd, Tu AA, Gierahn TM, Peters JM, Hopkins FF, Wei JR, Kummerlowe C, Grant NL, Nargan K, Phuah JY, Borish HJ, Maiello P, White AG, Winchell CG, Nyquist SK, Ganchua SKC, Myers A, Patel KV, Ameel CL, Cochran CT, Ibrahim S, Tomko JA, Frye LJ, Rosenberg JM, Shih A, Chao M, Klein E, Scanga CA, Ordovas-Montanes J, Berger B, Mattila JT, Madansein R, Love JC, Lin PL, Leslie A, Behar SM, Bryson B, Flynn JL, Fortune SM, Shalek AK. Multimodal profiling of lung granulomas in macaques reveals cellular correlates of tuberculosis control. Immunity. 2022 May 10;55(5):827-846.e10. doi: 10.1016/j.immuni.2022.04.004. Epub 2022 Apr 27. PMID: 35483355; PMCID: PMC9122264.

* PDF: [github](https://github.com/FredHutch/seatrac-hackday-2023/blob/main/gideon_etal/Gideon%20et%20al.%20NHP%20granulomas%20Immunity%202022.pdf)
* MS: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9122264/>
* DATA 4 WKS: [Alexandria](https://singlecell.broadinstitute.org/single_cell/study/SCP1749/cellular-ecology-of-m-tuberculosis-granulomas-4-week-dataset#study-summary); [Github](https://github.com/FredHutch/seatrac-hackday-2023/tree/main/gideon_etal/4week); [Figshare](https://figshare.com/account/articles/24425053)
* DATA 10 WKS
  + <https://singlecell.broadinstitute.org/single_cell/study/SCP257/cellular-ecology-of-m-tuberculosis-granulomas-10-week-dataset#study-summary>
* Single-cell data is on [Figshare](https://figshare.com/account/articles/24425053)
* **DATA SUMMARY**: single-cell RNAseq (SeqWell); 6 granulomas from 2 NHP at 4 wks and 26 granulomas from 4 NHP at 10 weeks post-infection (includes CFU per granuloma)

### Darrah *et al* 2023: scRNAseq from BAL in NHP Mtb challenge (BCG route, correlates of protection)

Darrah PA, Zeppa JJ, Wang C, Irvine EB, Bucsan AN, Rodgers MA, Pokkali S, Hackney JA, Kamath M, White AG, Borish HJ, Frye LJ, Tomko J, Kracinovsky K, Lin PL, Klein E, Scanga CA, Alter G, Fortune SM, Lauffenburger DA, Flynn JL, Seder RA, Maiello P, Roederer M. Airway T cells are a correlate of i.v. Bacille Calmette-Guerin-mediated protection against tuberculosis in rhesus macaques. Cell Host Microbe. 2023 Jun 14;31(6):962-977.e8. doi: 10.1016/j.chom.2023.05.006. Epub 2023 Jun 1. PMID: 37267955; PMCID: PMC10355173.

* PDF: <https://www.nature.com/articles/s41586-019-1817-8.pdf>; ([github](https://github.com/FredHutch/seatrac-hackday-2023/blob/main/darrah_etal/Darrah%20et%20al.%20IV%20BCG%20correlates%20Cell%20Host%20and%20Microbe%202023.pdf))
* MS: <https://www.nature.com/articles/s41586-019-1817-8>
* DATA: <https://singlecell.broadinstitute.org/single_cell/study/SCP796/prevention-of-mycobacterium-tuberculosis-infection-and-disease-in-nonhuman-primates-following-intravenous-bcg-vaccination?scpbr=the-alexandria-project> Single-cell data is on [Figshare](https://figshare.com/account/articles/24425053)
* **DATA SUMMARY**: single-cell; n=15 Rhesus macaques with BCG vaccination and Mtb challenge; 3 individuals per group for AE, IDhigh, IDlow, IV, Naïve-controls; BAL collected at Weeks 13 and 25 prior to challenge at Week 26; stimulated and unstimulated conditions; 60 samples, 1000 – 5000 cells each.

### Liu *et al* 2023: whole-blood bulk RNAseq from NHP Mtb challenge (BCG route and IV BCG dose)

Liu YE, Darrah PA, Zeppa JJ, Kamath M, Laboune F, Douek DC, Maiello P, Roederer M, Flynn JL, Seder RA, Khatri P. Blood transcriptional correlates of BCG-induced protection against tuberculosis in rhesus macaques. Cell Rep Med. 2023 Jul 18;4(7):101096. doi: 10.1016/j.xcrm.2023.101096. Epub 2023 Jun 29. PMID: 37390827; PMCID: PMC10394165.

* PDF: [github](https://github.com/FredHutch/seatrac-hackday-2023/blob/main/liu_etal/Liu%20et%20al%20IV%20BCG%20NHP%20mRNA%202023.pdf) MS: <https://www.sciencedirect.com/science/article/pii/S266637912300215X?via%3Dihub#sec4.1>
* DATA IV BCG DOSE STUDY (167 samples, 34 NHP): <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE218270>
* DATA BCG ROUTE STUDY (144 samples, 36 NHP): <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE218157>
* **DATA SUMMARY**: Darah et al. IV BCG dose and BCG route study matching; Pre BCG and Day 2, Wk 4, Wk 12 for IV BCG dose study; Pre, Day 2, Wk2 and Wk12 for route study; bulk RNAseq from whole blood; includes protection data for each individual NHP

# TUTORIALS

## Bulk RNAseq differential expression

* 1. <https://fredhutch.github.io/seatrac-hackday-2023/1.rnaseq_tutorial/2.Bulk_RNAseq_deg_analysis.html>
  2. Some other potentially useful tutorials, made by Kim:
     + <https://github.com/BIGslu/tutorials>
     + <https://github.com/BIGslu/workshops>

## scRNAseq ordinations

* 1. <https://fredhutch.github.io/seatrac-hackday-2023/1.rnaseq_tutorial/3.scRNAseq_analysis.html>
  2. Background and help on tSNE
     + <https://distill.pub/2016/misread-tsne/>  
       <https://towardsdatascience.com/t-sne-clearly-explained-d84c537f53a>
     + <https://towardsdatascience.com/t-sne-clearly-explained-d84c537f53a>
  3. Background and help on UMAP
     + <https://pair-code.github.io/understanding-umap/>
  4. Also: in Seurat, a lot of the important data (experimental data, cell type labels) is usually stored in a data frame you can access '<seuratObject>@meta.data'

## Network analysis

* 1. <https://fredhutch.github.io/seatrac-hackday-2023/2.networks_tutorial/gse_and_clustering_1.html>
  2. Another example of WGCNA in R
     + <https://bigslu.github.io/tutorials/RNAseq/4.RNAseq_modules.html>

# Results

## Liu et al modules by sex

A screenshot of a graph

Description automatically generated

## Darrah et al, 2020. B cell subset (downsampled)

A screenshot of a computer screen

Description automatically generatedA screenshot of a graph

Description automatically generated

## Treg cell pseudo-bulk DEGs

A graph of a number of dots

Description automatically generated