

Apr 19, 2023

LEGACY01: STATISTICS AND DATA ANALYSIS

 In 1 collection

DOI

dx.doi.org/10.17504/protocols.io.5jyl8jbedg2w/v1

Katrina M Pollock¹, Calliope Dendrou²

¹NIHR Imperial Clinical Research Facility, ICTEM Building, Hammersmith Hospital Campus, Du Cane Road, London W12 0HS, UK;

²Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN

Katrina M Pollock: Chief Investigator;

Calliope Dendrou: Scientific Lead

Human Cell Atlas Metho...



Katrina M Pollock

University of Oxford , Imperial College London

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.5jyl8jbedg2w/v1>

Protocol Citation: Katrina M Pollock, Calliope Dendrou 2023. LEGACY01: STATISTICS AND DATA ANALYSIS. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.5jyl8jbedg2w/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: January 11, 2023

Last Modified: April 19, 2023

Protocol Integer ID: 75104

Keywords: Single-cell sequencing, lymph node, influenza vaccine, ancestry, seasonal influenza vaccination responses in lymph node, cell genomics in ancestry, seasonal influenza vaccination response, experimental medicine study of seasonal influenza vaccination response, cell genomic, data analysis in an experimental medicine study, legacy01

Funders Acknowledgements:

Chan Zuckerberg Initiative

Grant ID: DAF2022-239944

Disclaimer

DISCLAIMER – FOR INFORMATIONAL PURPOSES ONLY; USE AT YOUR OWN RISK

The protocol content here is for informational purposes only and does not constitute legal, medical, clinical, or safety advice, or otherwise; content added to **protocols.io** is not peer reviewed and may not have undergone a formal approval of any kind. Information presented in this protocol should not substitute for independent professional judgment, advice, diagnosis, or treatment. Any action you take or refrain from taking using or relying upon the information presented here is strictly at your own risk. You agree that neither the Company nor any of the authors, contributors, administrators, or anyone else associated with **protocols.io**, can be held responsible for your use of the information contained in or linked to this protocol or any of our Sites/Apps and Services.

Abstract

This protocol details about statistics and data analysis in an experimental medicine study of seasonal influenza vaccination responses in Lymph node single-cell Genomics in AnCestrY (LEGACY01).

Attachments



[602-1266.docx](#)

631KB

Guidelines

STATISTICS AND DATA ANALYSIS

Thirty volunteers will be enrolled.

Ultrasound images will be collected using the software provided with the US machine operating system. These may be securely shared for storage on a secure password protected computer as anonymised images using the US applications such as ApliGate, developed by Canon for the secure sharing of images. Ultrasound images will be stored with the unique study identifier for each participant, the date of the scan and the initials of the person performing the examination.

For single cell RNA-seq experiments, we anticipate capturing ~3,000-5,000 cells per sample after quality control (QC) and filtering. Sequencing, batch correction, filtering, QC, doublet removal, ambient RNA correction and correction for multiple testing will be performed according to single-cell RNA-seq best practices using established pipelines and tools (e.g. <https://github.com/DendrouLab>; COMBAT Consortium 2022). Statistical and outcome measures will typically be assessed through ANOVA, Wilcoxon rank, and Fisher exact tests. Dimensionality reduction processes will be used to understand data structure and to identify clusters and cell types for statistical testing. Differential cluster abundance and differential gene expression using pseudobulk counts will be performed using relevant edgeR packages and applying a Benjamini-Hochberg multiple testing correction. Power calculations for single-cell RNA-seq data analysis are based on relevant available data (e.g., Turner et al. 2020), and for assessing differential abundance and gene expression, for instance, we estimate that n=10 individuals per group are required to have >80% power to detect 30% differences between longitudinal samples (pre and post vaccination), with a false discovery rate of 5%. Given that there is expected variation in sample yield after fine needle aspiration which may be anatomically determined, there is a 50% overage in each group to allow for this and for any participants lost to follow up bringing the total to n=30.

Single-cell repertoire analyses will be performed as previously described^{22,23,24}. Briefly, T cell receptor (TCR) and B cell receptor (BCR) outputs from CellRanger will be further filtered and processed to remove homotypic doublets and low-quality droplets and annotated using IMGT. Single-cell TCR clonality measurements will include Shannon diversity calculation using the entropy R package, and mean clone size estimation by bootstrapped down-sampling. Single-cell BCR clonality measurements will include clonal expansion index calculation based on the Gini index of the number of total BCRs per clone, whilst the clonal diversification index is calculated as the Gini index of the number of unique BCRs per clone.

For bulk sequencing, standard pipelines will be used for data processing (COMBAT Consortium 2022). Principal component analysis will be used for initial analyses of the normalised and filtered data. Differential expression analysis of the normalised data will be performed using the limma R package for example, with pathway enrichment analysis being performed using the Reactome pathways via the XGR R package with Fisher's exact test. Weighted gene correlation network analysis will be applied to identify modules of highly correlated genes. Bulk repertoire analyses will include filtering for base quality amongst other quality control checks, followed by TCR chain/Ig isotype frequency estimates and usage, CDR3 length characterisation, and somatic hypermutation and BCR class-switching analyses.



For flow cytometric analyses, data will be analysed using FlowJo, with appropriate control/validation with non-specific isotype controls and beads to ensure stable equipment calibration over time, as well as use of gating templates. For functional/serological analyses, assays will be performed in duplicate or triplicate with appropriate repeat experiments to ensure reproducibility of findings. Appropriate tests (e.g., t-tests, Wilcoxon rank-sum tests) for assessing inter-group differences will be performed using R, python or GraphPad Prism as required.

Data generated and all appropriate documentation will be stored for a minimum of 10 years after the completion of the study, including the follow-up period.

REFERENCES

22. Corridoni D, Antanaviciute A, Gupta T, et al. (2020) Single-cell atlas of colonic CD8+ T cells in ulcerative colitis. *Nat Med* 26: 1480-1490.
23. Huang B, Chen Z, Geng L, et al. (2019) Mucosal profiling of pediatric-onset colitis and IBD reveals common pathogenics and therapeutic pathways. *Cell* 179: 1160-1176.
24. COvid-19 Multi-omics Blood ATlas (COMBAT) Consortium. (2022) A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. *Cell* 185: 916-938.

Troubleshooting

