nature portfolio

Peer Review File

Multi-omics profiling of primary human B cell activation by distinct stimuli

Corresponding Author: Dr Vitor Aguiar

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have provided a detailed point-by-point response to reviewer comments. While I don't agree that Supplementary Figure 14 adequately answers my critique, the rest of the queries were adequately addressed and I think the manuscript is acceptable for publication.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
The authors have addressed my major concerns by highlighting the key limitations of their datasets. They have also toned down the language on the impact of genetic variants and instead focussed their results on gene expression in activated B cell states. Methodological concerns have been clarified with new data and additional analysis. Although higher resolution data with purified B cell subsets would be optimal, the current dataset will be a valuable resource to the scientific community.

The code has been review from 3 bioinformaticians:

1. The authors acknowledge in their GitHub repository that the code consists of custom scripts rather than a standalone tool, and that the associated data is hosted on Zenodo. The code is well-documented and organized. However, there are some concerns:

• No execution order is specified. Although the Zenodo data source is referenced, there are no instructions indicating the order in which the scripts should be run (except for the Bcell_rnaseq folder), making it challenging to replicate the pipeline end-to-end.

• Environment requirements are not documented per script/path. In some cases, specifically in the Slurm scripts, the authors referenced the environment they used without specifying its contents (packages, versions, etc.).

• A general README file is present with broad instructions, but it does not provide an overview of the pipeline or step-by-step details, or inputs/outputs specifications of each step. Adding a descriptive header to each script explaining its purpose would help navigate through it. Ideally, each analysis folder would include its own README documenting the purpose of each script, the execution order, and environment configuration instructions. In the same way, adding a header comment to each script with its purpose, inputs/ outputs would be very helpful.

• Because of this, and as noted by the authors, reproducing their analysis is not that straightforward. It would require downloading the data, determining the correct execution order, identifying the specific inputs for each step and reformatting if needed, and changing any hardcoded path (they can create config files with all parameters needed).

2. The authors provide a comprehensive collection of scripts covering data preparation, basic analyses, workflow execution, and figure generation. The scripts are generally well organized by analysis type, and most are tidy and simple, which

improves readability and accessibility. In addition, dependency requirements are listed on the main GitHub page, allowing users to install the packages needed for the analyses. Overall, the repository is well-structured and provides users with an overview of the analyses performed in the study.

The authors acknowledge the absence of demo datasets, meaning users must download the original data to reproduce the analyses. However, it is often unclear what input file formats, column structures, or metadata are required for individual scripts. Providing simple documentation describing expected file formats, required headers, and example inputs—as commonly included in other GitHub repositories—would substantially improve code usability.

In general, users can follow the overall purpose of most scripts, but the documentation remains limited. More detailed comments and explanations throughout the code would help clarify specific steps and improve understanding of the analysis.

Although the scripts are organized according to analysis type, the repository lacks clear execution order. In several cases, scripts depend on outputs generated by previous steps, but these dependencies are not always obvious, making it difficult to determine the correct execution order.

Overall, the repository contains an extensive set of scripts for a broad range of analyses; however, the limited guidance on execution order, required inputs, and file formatting makes the analyses more challenging to reproduce.

3. Code is organized into analysis-specific folders with clearly named scripts and some inline documentation. However, the README is basic, giving minimal detail regarding the contents of this repository. Additionally, the documentation provided is not sufficient to reliably and reproducibly run scripts to reproduce data shown in the paper in a streamlined and user friendly manner. Whilst individual scripts do contain some documentation, they lack a clear description of their objectives, required inputs and how they should be executed. This makes it difficult for the user to quickly understand and run the script beyond what the script names suggest. Scripts contain both relative and absolute paths, requiring manual edits.
There is a wide breadth of analyses provided in this repository so it is unrealistic for dummy data datasets to be provided in GitHub. However, analysis specific READMEs would drastically improve usability and allow users to reproduce data. Software dependencies are listed and conda environments are called in SLURM scripts but the absence of shared YAML file risks software version conflicts upon attempting to install each program / software into a singular environment. For users who do not have experience with specific analyses, it will be difficult to separate software / packages into analyses of specific environments easily.

A more comprehensive user guide, including input requirements for each script, sequence of script execution and minimal examples of data for less computationally intensive analyses would drastically improve reproducibility for the execution of the scripts published in this repository.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have addressed many of the points raised in the original review, which has substantially improved the manuscript. However, several issues remain and should be addressed prior to publication. Once these points have been satisfactorily addressed, I believe the manuscript will be well-suited for publication in Nature Communications.

- Figure 3 and the accompanying text: several statements on differential gene expression and differential expression of cell surface markers appear to be based primarily on visual inspection of UMAP gene expression patterns, without sufficient quantitative support. For the points listed below, the authors should provide quantitative evidence, ideally through pseudobulk differential gene expression analyses comparing the relevant conditions for each claim, while accounting for inter-individual variation in the differential gene/cell surface marker expression model.
- We observed that BCR engagement in BCRc and DN2c leads to decreased expression of IgD/IgM on the surface of cells in comparison with the Day 0, IL-4c, and TLR7c conditions (Fig. 3c).
- Concordantly, at 24 hours, these conditions also upregulated CD69 protein expression in comparison to IL-4c and TLR7c, indicating high B cell activation after BCR stimulation (Fig. 3c).
- CTSH, which encodes for Cathepsin H and is associated with MS and type 1 diabetes (T1D),39,51 increases at 72 hours after BCR stimulation and is mainly expressed in CD27 positive cells.
- CXCR5 encodes a B cell homing factor, is located in a risk locus for multiple IMDs,32,39,60 and in our data it is expressed within a subset of cells in the DN2c at 24 hours.
- IRF8, another interferon response transcription factor gene that is associated with IMDs such as IBD, PBC, SLE, and SSc,37,52,61–63 is more highly expressed in cells that co-express MYC, CCL3, and CCL4, in BCRc and DN2c at 24 hours.
- RGS1, which is associated with multiple autoimmune diseases,39,53,64,65 is expressed in BCRc at 72 hours, restricted to a small population expressing ITGAX and FCRL5.
- ZBTB38, associated with asthma,32 increases expression in BCRc with time, consistent with the bulk RNA-seq data.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
The authors have successfully addressed all my comments and I believe this manuscript is acceptable for publication at Nature communications.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/