Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

April, 18, 2022

Ranjit Pelia Date

Cellular Transcriptional Profiling Reveals Multiple B Cell Differentiation Branching Points in Mus musculus Spleen Exposed to Lipopolysaccharides (LPS) and NP-4-Hydroxy-3-Nitrophenylacetic (NPF)

By

Ranjit Pelia

Qiang Zhang, M.D, Ph.D.

Committee Chair

Cellular Transcriptional Profiling Reveals Multiple B Cell Differentiation Branching Points in Mus musculus Spleen Exposed to Lipopolysaccharides (LPS) and NP-4-Hydroxy-3-Nitrophenylacetic (NPF)

By

Ranjit Pelia

Bachelor of Science
Emory University
2018

Thesis Committee Chair: Qiang Zhang, M.D, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master in Public Health
in Global and Environmental Health
2022

Abstract

Cellular Transcriptional Profiling Reveals Multiple B Cell Differentiation Branching Points in Mus musculus Spleen Exposed to Lipopolysaccharides (LPS) and NP-4-Hydroxy-3-Nitrophenylacetic (NPF)

By

Ranjit Pelia

Background: Environmental exposure of toxicants has been shown to be detrimental to health. The humoral immune response is an integral component of pathogen recognition and elimination. In which, B cell production is increased and cascades into antibody secreting cells differentiation. These cells are one of the primary defense mechanisms of multicellular organisms, as they can bind, eliminate, decoy, or neutralize the pathogen. There is a paucity in our current understanding of immunological dynamics at a cellular level. The duality of single-cell RNA sequencing (scRNA-seq) and mouse immunity models have enabled researchers to decipher genetic, epigenetic, and transcriptomic mechanisms of action on a molecular scale.

Objective: The transcriptional program driving immune responses and the humoral homeostasis dynamics remains elusive. We aimed to characterize transcriptional profiling during B-cell differentiation into antibody secreting cells (ASC) using scRNA-seq in order to gain deeper insight into the mechanisms of immunological regulation.

Method: Using the datasets provided by the Scharer lab and published in *Nature Communications* (2020:11:3989), we employed integrated bioinformatics to assess mouse (*Mus musculus*) spleen exposed to Lipopolysaccharides (LPS), n=2, and NP-4-Hydroxy-3-Nitrophenylacetic (NPF), n=2, and sequenced using 10X 3-prime scRNA-seq. R packages Seurat, VelocytoR, and the methods described in *Nature* by Le Manno et al (2018) were utilized for quality control, cell type clustering, and differential kinetics analysis. Lastly, Python based ScVelo was performed on the integrated LPS and NPF samples to calculate latent time, RNA velocity, and gene-cell-type specific trajectories.

Results: Upon sequencing, the LPS samples contained 3312 and 3164 cells and NPF samples contained 1893 and 3330 cells. Upon integration, LPS and NPF were clustered and annotated into three main groups: activated B cells, naïve B cells, and ASCs. Using the ROC for differential expression analysis, there were 4974 in LPS and 2037 in NPF genes specific for the overall subclusters: naïve B, activated B non-ASC, activated B ASC, non-ASC, and ASCs. Trajectory analysis reveals two distinct B-cell lineages in both LPS and NPF. Latent time showed ASCs to be at a terminal state where B-cells were shown to be proliferating.

Conclusion: Here, we characterized the transcriptomic nature of immune cells in response to pathogens and the consequential differentiation lineages of B-cells. Our results demonstrate insights into the dynamics of immune response and may be translational in the realms of autoimmune illnesses, environmental detriments to health, and our global understanding of pathogen antagonization.

Cellular Transcriptional Profiling reveals multiple B Cell Differentiation Branching Points in Mus musculus spleen exposed to Lipopolysaccharides (LPS) and NP-4-Hydroxy-3-Nitrophenylacetic (NPF)

By

Ranjit Pelia

Bachelor of Science
Emory University
2018

Thesis Committee Chair: Qiang Zhang, M.D, Ph.D.

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master in Public Health
in Global and Environmental Health
2022

Acknowledgements

I am thankful for Emory University and Rollins School of Public Health in providing me with guidance, resources, and the tools to conduct research in the betterment of public health andglobal environment. Specifically, Dr. Qiang Zhang, for demonstrating confidence in my capabilities and allowing me the opportunity to conduct transcriptomic research.

My work would not have been possible without the support of Dr. Chris Scharer and his team. They allowed me the privilege to use their single-cell RNAseq dataset, alongside sharing scripts, tutorials, and other helpful advice. Also would like to thank Yajun He for her advice andprogramming support in conducting single-cell analysis.

With my deepest gratitude to Dr. Michael Caudle, Dr. Arri Eisen, Dr. Subra Kugathasan, Dr. Jason Matthews, and Dr. Suresh Venkateswaran for their continuous guidance. Special thankyou to Ariadne Swichtenberg, my journey at Rollins would not have been possible without you.

Table of Contents

Introduction

The mammalian immune system is characterized into adaptive and innate immunities. An integral component of the adaptive network is the humoral immune response, a dynamic system maintaining homeostasis and modulating controlled immune response[1]. The fundamental principles underlying biological systems reaction to external stimuli have been highlighted by the works of Ehrlich, Kitasato, Pasteur, von Behring, and others[2]. Generally, in reaction to external stimuli or foreign invaders, a plethora of naturally occurring antibodies, pentraxins, and other gene-regulatory cascades are produced[3]. There are three main types of B cells involved in the humoral immune response; naïve B (nB), activated B (actB), and antibody secreting cells (ASCs). Upon stimulation, B cells either naïve or memory-B, progress to become actB or directly differentiate into ASCs[4]. Although much is known about the systems immunology, the cellular, molecular, and transcriptomic layers remain undeciphered.

In complex organisms, B-cell differentiation is a quintessential part of pathogen recognition and appropriate immunological regulation. In the context of humans, early B-cells are generated by the hematopoietic stem cells and depart the bone marrow, in one of its various forms: lymphoid progenitor cell, pro-B cell, pre-B cell, or immature B cell[5]. In some instances, B cells rapidly differentiate into ASCs. There are a multitude of factors that may affect an organism's capacity, extent, and diversity of B-cell generation throughout a lifetime. *Ex-vivo* experiments using fluorescence activated cell sorting (FACS), mass cytometry, and computational tools, have shown putative B-lineage cell trajectories in both human and mouse models[6]. Specifically, the underlying mechanisms of molecular cellular programming taking place during the germinal center reaction and the consequences thereof have been thoroughly investigated.[7] Nonetheless, there is a paucity in our understanding of pre-germinal center responses and the cellular reprogramming occurring to prime an immune response.

The duality of *Mus musculus* models and scRNAseq technology has demonstrated insightful capabilities in deciphering transcriptomic signatures at a molecular, gene-by-gene, level. Although we may not be able to exhibit all the complexities of humoral immune responses through stimulation, we can use known toxicants to elicit and observe these behaviors. Specifically, lipopolysaccharides (LPS) and NP-4-Hydroxy-3-Nitrophenylacetic (NPF) are known to induce distinct cytokine releases thus signaling inflammation[8]. In mammals, these toxicants lead to a very specific pathogen-associated molecular pattern. Using scRNA-seq, the underlying transcriptional levels of RNAs, or genes, may be reflective of the defensive innate immune responses[9].

Previously, Scharer et al[10] reported on B cell differentiation division kinetics, *in-vivo*, by identifying molecular trajectories of B cell fate and plasma or ASC production. They used an adoptive mouse model system with CTV-labeled $Cd19^{Cre/+}$ $Prdm1^{fl/fl}$ (BcKO) or $Prdm1^{fl/fl}$, as control, B cells from spleen region. These captured B cells were stimulated with LPS demonstrated consistent distributions of divisions, compared with Control B cells. To assess the transcriptomic nature of these cells, they were FACS isolated and sequenced with scRNAseq[10]. Here, we build upon these results by using the LPS scRNAseq dataset (n=2), in combination with NPF scRNAseq data (n=2) of mouse splenic tissue to characterize potential transcriptomic programming occurring upon B-cell activation and consequential differentiation lineages.

Methods

10X 3'(prime) scRNAseq of LPS and NPF stimulated Mus musculus spleen tissue:

Cells were FACS sorted from mouse spleen tissue exposed to LPS and NPF by Scharer et al group at Emory University School of Medicine[10]. Sequencing of these captured cells was conducted using 10X Genomics 3-prime platform with an estimated n=17,400 cells as input for GEM generation and library preparation[11]. The results were processed into Illumina BCL files to be mapped to the mm10 genome using CellRanger v.2.1.1[12]. The methods were consistent for all n=4 samples to mitigate batch effects.

Loom Formatting and Generation:

The output files from CellRanger were converted into Loom files, to analyze transcriptomic changes by comparing spliced and unspliced reads of genes. This was performed by using Velocyto[13] in Linux. The original script was provided by Yajun He and was modified to incorporate the NPF samples, alongside the LPS samples. The sequenced and aligned FAST-Q files were analyzed for further quality control. Our generalized analytical framework is shown in Figure 1.

Quality Control and Batch Effects Assessment:

We used R v.3.7.1 and Python v.3.1 to assess the number of reads sequenced and aligned and the abundance of those gene transcripts to confer biological plausibility. The standard Seurat (v.3.2.2) quality control pipeline was optimized and conducted for each sample individually[14]. Upon sequencing and prior to quality control, there were n=3407, 3336, 1941, and 3394 cells with n=12892, 12722, 11754, and 12865 genes being expressed with a minimum average of n=1 for LPS 1, LPS 2, NPF 1, and NPF 2, respectively (Table 1). The samples showed consistent levels of gene for a given cell (Figure 2). To consider any cells that may be undergoing apoptosis or serving as possible confounders, we filtered out cells with mitochondrial (mtRNA) percentages of 25% or higher (Table 2). This resulted in our filtered dataset of n=3312, 3164, 1893, and 3330 cells for LPS 1, LPS 2, NPF 1, and NPF 2, respectively (Table 3).

In dealing with 3-prime 10X scRNAseq datasets, it is vital to consider the spliced and unspliced assays or RNA gene expression generated by CellRanger, as only one may be chosen for downstream analysis. Previous literature suggests this ratio of unspliced to spliced to be roughly 20%[13]. We used an optimized version of DropletUtils[15] to assess the quality of our transcriptomic data. Our datasets showed LPS 1 ~ 0.18 (Figure 3a), LPS 2 ~ 0.18 (Figure 3b), NPF 1 ~ 0.22 (Figure 3c), and NPF 2 ~ 0.21 (Figure 3d) thus representing a well-captured cohort to be processed further. The datasets were then normalized using log normalization, scaled with scale factor set to 10,000, and then the natural log was taken to perform principal component analysis (PCA). Each sample has its unique gene expression signature; therefore, a set of principal components (PCs) need to be chosen for each sample that capture its true biological transcriptome before integration. Cell cycle genes may influence these principal components and sway away from our genes of interest, i.e., immunologically related ones. Cell Cycle Scoring, a feature of Seurat was used for cell cycle specific assessments, based on pre-defined lists of genes for each state[14]. We assessed for G1, G2/M, and S cell cycle states for each sample (Table 4). There were consistent proportions of G1 and G2/M for both LPS 1 (Figure 4a), LPS 2 (Figure 4b), NPF 1 (Figure 4c), and NPF 2 (Figure 4d).

Statistical Parameters for Integration:

To consolidate our datasets, the two LPS samples were integrated into one LPS and the two NPF samples were also integrated into one NPF. We did not combine all four samples into one as the quality control showed varying degrees of transcription rate and UMI ranks for LPS to be starkly different to NPF, thus this may lead to potential confounders or biological misinterpretation. SCTransform[16] algorithms were used for both LPS and NPF with specific parameters. For LPS, we integrated LPS 1 (n=2795 cells) and LPS 2 (n=2384 cells) using dimensions 1 to 20, canonical correlation analysis (CCA) for neighbors' identification, and reliability-aware network modelling (RANN) for final anchoring. Clustering of the anchored dataset was performed by Louvain algorithm with multilevel refinement[17]. Likewise, NPF 1 (n=1459 cells) and NPF 2 (n=2171 cells) were anchored with CCA and RANN using dimensions 1 to 20.

Results

Integrative Clustering: Clustering was performed on the integrated LPS and NPF Seurat objectswhich contained combined cells of n=5026, n=3630, respectively. Using PCs 1 to 20, we visualized cell clustering with UMAP and TSNE showing n=12 clusters for LPS (Figure 6a, b) and n=11 clusters for NPF (Figure 6c, d). These showed the capability of cells to be assigned tosimilar topological groups denoting similar types, quantities, and varieties of mRNAs.

Cell Type Annotations: Scharer et al., provided markers for commonly known cell types; naïve B, activated B, and ASCs. Acknowledging that most of this data has been annotated previously by Scharer et al., we sequestered the n=12 clusters in LPS and n=11 clusters in NPF into 5 main groups: Naïve B, Activated B Non-ASC Destined, Activated B ASC Destined, Antibody Secreting Cells, and Non-ASC Cells. The genes: *Ebfl1* (+), *Pax5* (+), *Cd19* (+), *Zbtb32* (~+), and *Ezh2* (-) were denoted for Naïve B cells. For activated B non-ASC positive gene expression of *Sell* (+), *Ezh2* (~+), *Zbtb32* (+), and *Tbx21* (+) were tested. Whereas, for activated B ASC categorization *Srm* (+), *Ezh2* (+), *Irf4*(+), *Batf* (+), *Zbtb32* (+), *Tbx21* (~+), and *Sell* (~ -/ ~+) were used. Advanced and non-advanced antibody secreting cells were grouped due to positive expression of *Sdc1* (+), *Srm* (+), *Ezh2* (+), *Xbp1* (+), *Jchain* (+) and negative expression of *Pax5* (-), *Sell* (-), and *Spib* (-). Lastly, non-antibody secreting cells (non-ASC) were grouped based on four positive markers *Sell* (+), *Zbtb32* (+), *Ezh2* (+), and *Tbx21* (+).
Using these markers, our annotated LPS dataset showed n=3312 naïve B, n=1335 ActB Non-ASC, n=355 Asc, n=228 Non-ASC, and n=142 ActB ASC. PC1 and PC2 (Figure 7a) showed separation by annotated cell types thus providing confidence in variance-driven biology clustering (Figure 7a). Likewise, PC 1 vs PC2 showed cell type clustering of NPF dataset n=4229 cells comprised of n=2304 naïve B, n=838 Non-ASC, n=385 ActB ASC, n=339 ActB Non ASC, and n=363 ASC (Figure 7b). The proportions of cell types were consistent amongst both LPS and NPF with naïve B having the greatest proportion and ASC have the least. Both the TSNE and UMAP showed ASC to be topologically farthest from all the other cells thus representing gene variance differences for LPS (Figure 8a,c) and NPF (Figure 8b,d).

RNA Velocity and Pseudotime Analysis: Using a principled probabilistic model, accompanied with Bayesian inferences, variation in temporal dimensions may be calculated with static data,

deemed pseudotime[18]. Our parameters for both LPS and NPF were consistent for conducting RNA velocity analysis using ScVelo[19] with PCs 1 to 20 and a Dynamical Inference model. Here, for a given gene, the differentiation rate of unspliced to spliced ratios is used for calculating differential kinetics per cell. The trajectories implied the cells in LPS to be migrating from Naïve B to two different lineages (Figure 9a, b). Similarly, we also observed the NPF cells, with increased resolution due to more defined topological patterns, to be probabilistically differentiating from Naïve B to binary differentiation points (Figure 10a, b). As expected, we observe the same lineage differentiation patterns of Naïve B cells becoming A) directly antibody secreting cells, B) activated B ASC destined, or C) activated B non-ASC destined.

Branching Points and Trajectories: To further understand the Bayesian inference-based RNA velocity modeling of LPS and NPF, we examined the velocity vector's length and statistical confidence. The length of the velocity vectors represents speed or rate of differentiation. The confidence is calculated using correlations of a given cell to nearby cells by comparing velocities. Interestingly, we observe cell type specific color mapping of both length and confidence in the TSNE LPS (Figure 11a, c) and TSNE NPF (Figure 11 b, d). This is further conveyed onto the UMAP for LPS (Figure 12a, c) and NPF (Figure 12b, d), which often is performed in conjunction to observe any discrepancies. We observed ASCs to have the highestconfidence and length, implying rapid differentiation rate, in both LPS and NPF. Naïve B cellswere the most diffusive and activated B ASC and non-ASC were in between ASC and Naïve, reflective of known biology.

Lastly, we tested our analysis on a holistic perspective by conducting partition-based graph abstraction (PAGA) trajectory analysis using the consolidated, annotated, LPS and NPF datasets[20]. PAGA incorporates diffusion pseudotime (DPT) into its trajectory mapping. DPT is a reconstruction of the developmental differential of cells through diffusion-like random walks[21]. Using principal components 1 to 20, we observed antibody secreting cells and activated B cells with ASC destiny to exemplify the highest pseudotime in LPS (Figure 13a) and NPF (Figure 14a). One of the stark differences observed between LPS and NPF was in the PAGA model having greater decoherence or chaotic nature in NPF. The LPS nodes show a clear, distinct, transition of either A) naïve B cells becoming activated B and then into antibody secreting cells or B) naïve B cells becoming activated B non-ASC destined (Figure 13b). Contrast to this simplification, the NPF PAGA trajectory nodes showed more lineages (Figure 14b). In which activated B cells are proliferating into A) activated B ASC destined, B) activated B to non-ASC destined, or C) activated B directly to ASC (Figure 14b). These discrepancies between LPS and NPF are more likely attributable to quality control and batch effects than biology due to a stark decrease in both quantity of cells captured and the transcriptomic sparsity of the genes within those cells.

Discussion: The advancements of technology, coupled with bioinformatics, have allowed geneticists and biologists to be able to test, previously questioned or experimentally limited, hypothesis. scRNA-seq and mouse models are invaluable in their utility to model and represent holistic dynamics of gene expression, specifically mRNAs. Protein level fluctuations have been well documented in their utility of determining cell fate, timing, and the resulting cell type specificity during varying hallmarks of development[22-24]. scRNAseq is an important tool for the currently occurring paradigm shift of genetic to epigenetic mechanisms of action in the realms of medicine, public health, and exposome detriments of environmental-human interactions[25]. In our

results, we characterized the usages of mammalian model systems to examine immunological consequences to toxicants, through LPS and NPF. Considering the paucity of lack toxicological-based scRNAseq datasets, our *Mus musculus* scRNAseq from Scharer et al was tested for B-cell lineage differentiation with an additional NPF dataset.

Any dataset comprised of sufficient power will contain dense, sparse, connected, disconnected, and ambiguous mathematical constructs, which form complex topology. The limitations of scRNAseq, RNA velocity, and differential kinetics are quintessential when interpreting the biological plausibility of any sequencing data, results, or inferences[20]. RNA velocity and PAGA, these graphical abstractions are based on topological mapping thus not directed by supervised biology. This assumes that gene expression data may be connected due to its generation across cells consistently thus will present with some measurable manifolds. It's important to consider the underlying biology plausibility as these systems are dynamically adaptive. For example, B-cells are known to be highly proliferative so our results may be reflective of inherent biology more so than B-cell lineage differentiation.

In terms of public health, our work helps build on mammalian and multi-cell organism immunological modulations as a consequence of toxicant exposure. Multiple studies have examined neurological burdens of disease by environmental toxicants, but fewer studies have investigated the role of spleen, wheelhouse of immune cell production[26,27]. We propose that our results showing varying lineages of B-cell differentiation into either non-ASC or ASC destinies may reflect the potential of single-cell RNAseq to test toxicant biomonitoring. In which, environmental, occupational or lifestyle toxicants are modeled to be used by public health experts, policy makers, and to predict illness risk and incidence[28].

In conclusion, our analysis characterizes the transcriptomic nature of immune cells in response to a toxicant and the consequential differentiation lineages of B-cells. Our results demonstrate insights into the dynamics of immune response pruning and may be translational in the realms of autoimmune illnesses, environmental detriments to health, and our global understanding of pathogen antagonization.

Tables

Table 1: Number of genes and cells for each mouse spleen sample

Sample	Cells	Features	Counts Spliced median	Features Spliced median	Counts Unspliced median	Features Unspliced median
LPS 1	3407	12892	7557	1989	1859	1005
LPS 2	3336	12722	6784	1880	1713	944
NPF 1	1941	11754	4796	1580	1526	892
NPF 2	3394	12865	4522	1509	1470	863

Table 2: Proportion of mtRNA, mitochondrial RNA percentages per cell per sample

Sample	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
LPS 1	0	0.01258	0.01553	0.03816	0.02054	0.92157
LPS 2	0	0.01286	0.01615	0.03373	0.02126	0.90833
NPF 1	0	0.01563	0.02048	0.04095	0.02748	0.87273
NPF 2	0	0.01414	0.01963	0.03612	0.02616	0.94792

Table 3: Pre- and Post-filtering based on mtRNA percentages

Sample	Cells (pre QC)	mtRNA > 25%	Cells (post QC)
LPS 1	3407	95	3312
LPS 2	3336	62	3164
NPF 1	1941	48	1893
NPF 2	3394	64	3330

Table 4: Cell Cycle specific annotation of probability based on G1, G2, and S scores

Sample	Cells	Features	G1	G2	S
LPS 1	3312	12892	1971	615	726
LPS 2	3164	12722	1880	621	561
NPF 1	1893	11754	1368	248	277
NPF 2	3330	12865	2220	609	501

Table 5: Number of Cells per sample, after final quality control and filtering

Sample	Cells
LPS 1	2642
LPS 2	2384
NPF 1	1459
NPF 2	2171

Figures

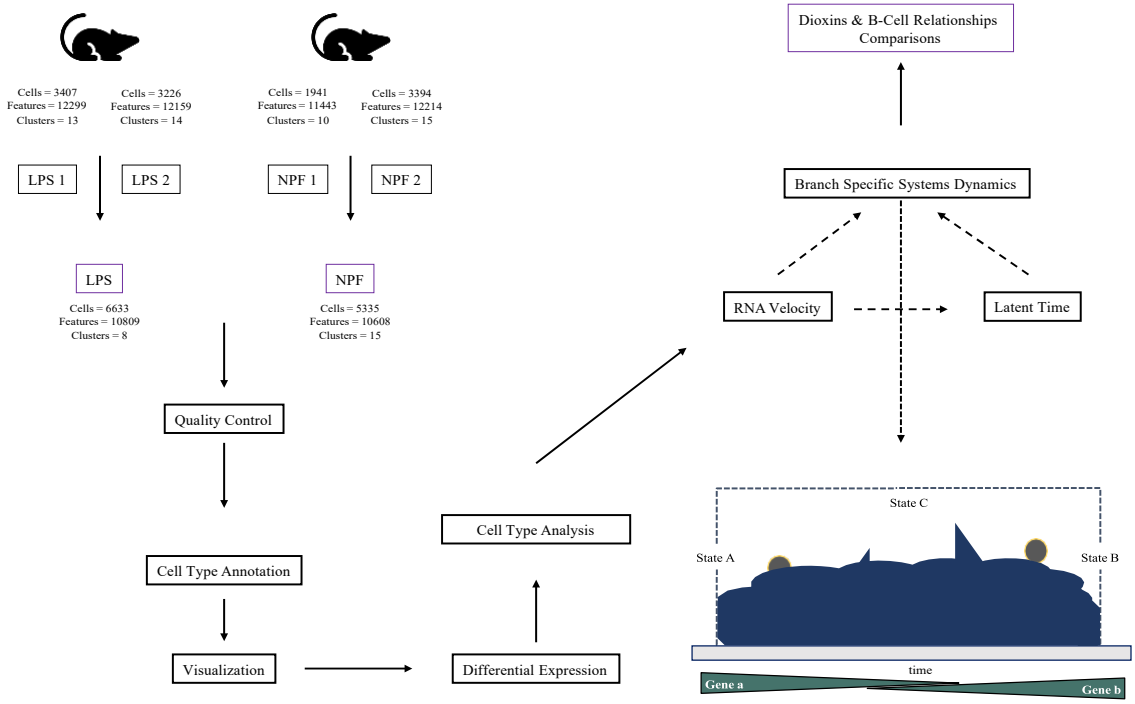

Figure 1: Analytical workflow of LPS (n=2) and NPF (n=2) scRNAseq datasets with consideration for sample vs toxicant specificity. The dynamic homeostatic graph with t=time on the x-axis denotes variable states of a cell with adaptive gene expression.

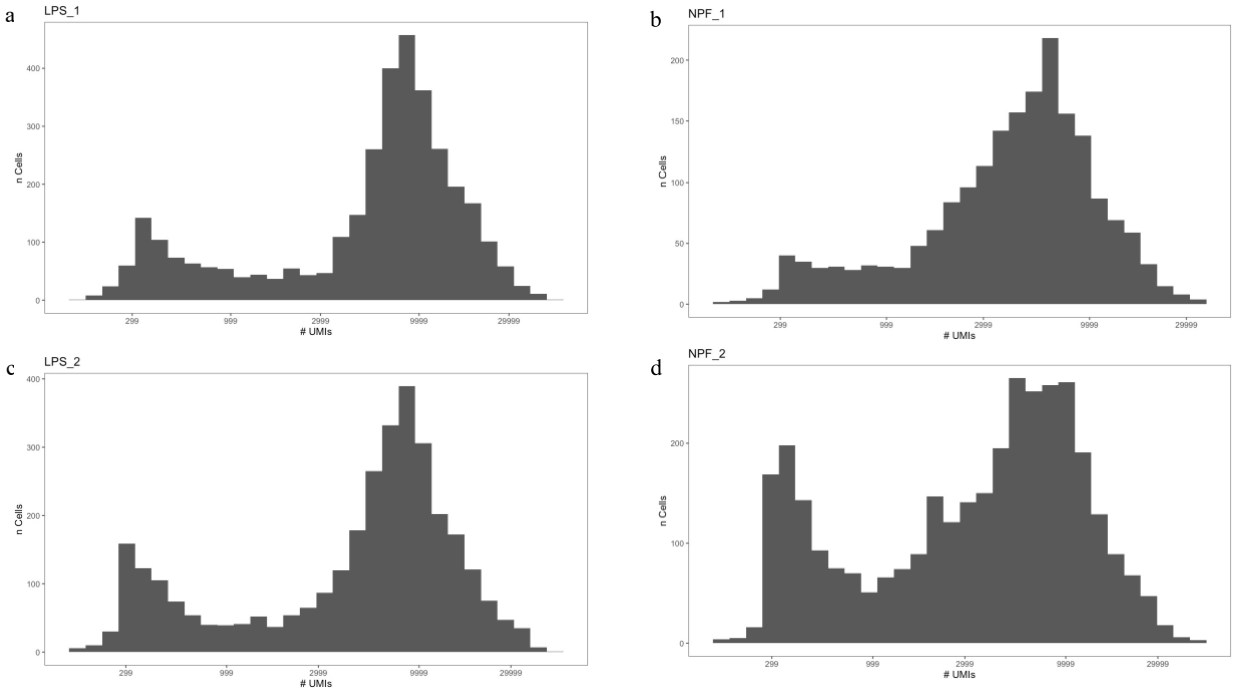

Figure 2: UMI or total number of genes for a given cell for each sample are shown above. LPS 1 (a) and LPS 2 (c) showed extremely consistent values. Whereas, NPF 1 (b) and NPF 2 (d) were slightly inconsistent due to varying degrees in lower ranges of expression.

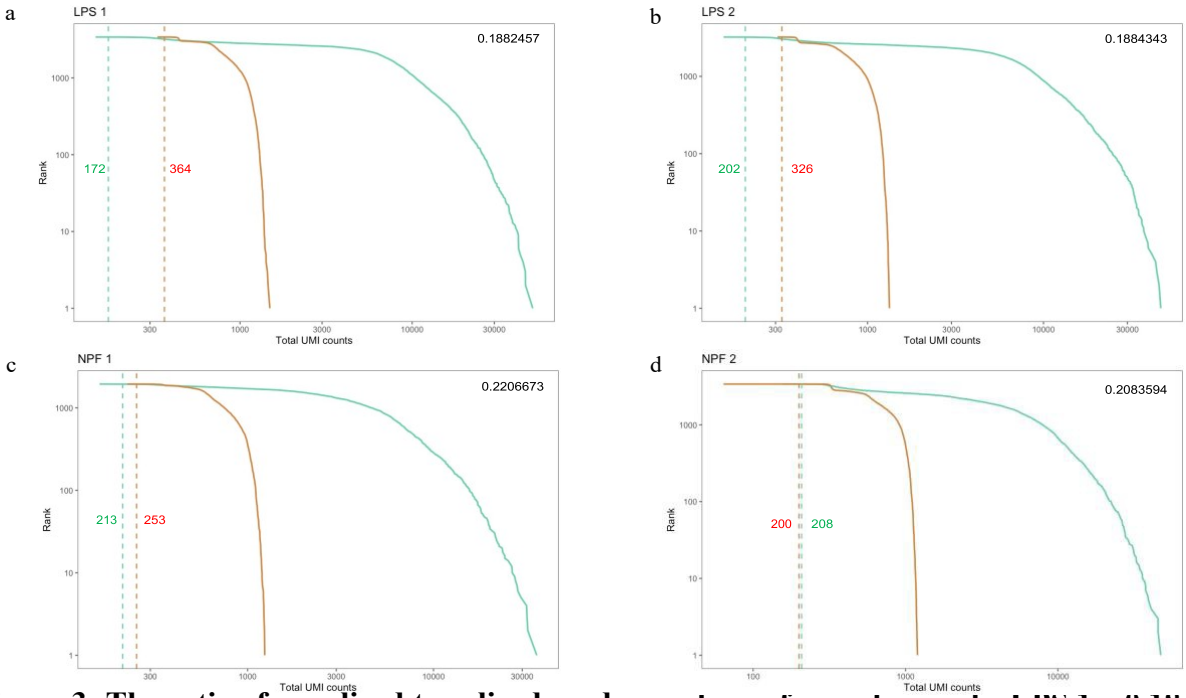

Figure 3: The ratio of unspliced to spliced reads are shown for each sample, LPS 1 ~ 0.18 (a), LPS 2 ~ 0.18 (b), NPF 1 ~ 0.22 (c), and NPF 2 ~ 0.21 (d). The green labeled digit is the unspliced- and the red labeled digit is the spliced average rank of cells with the x axis representing the total UMI counts.

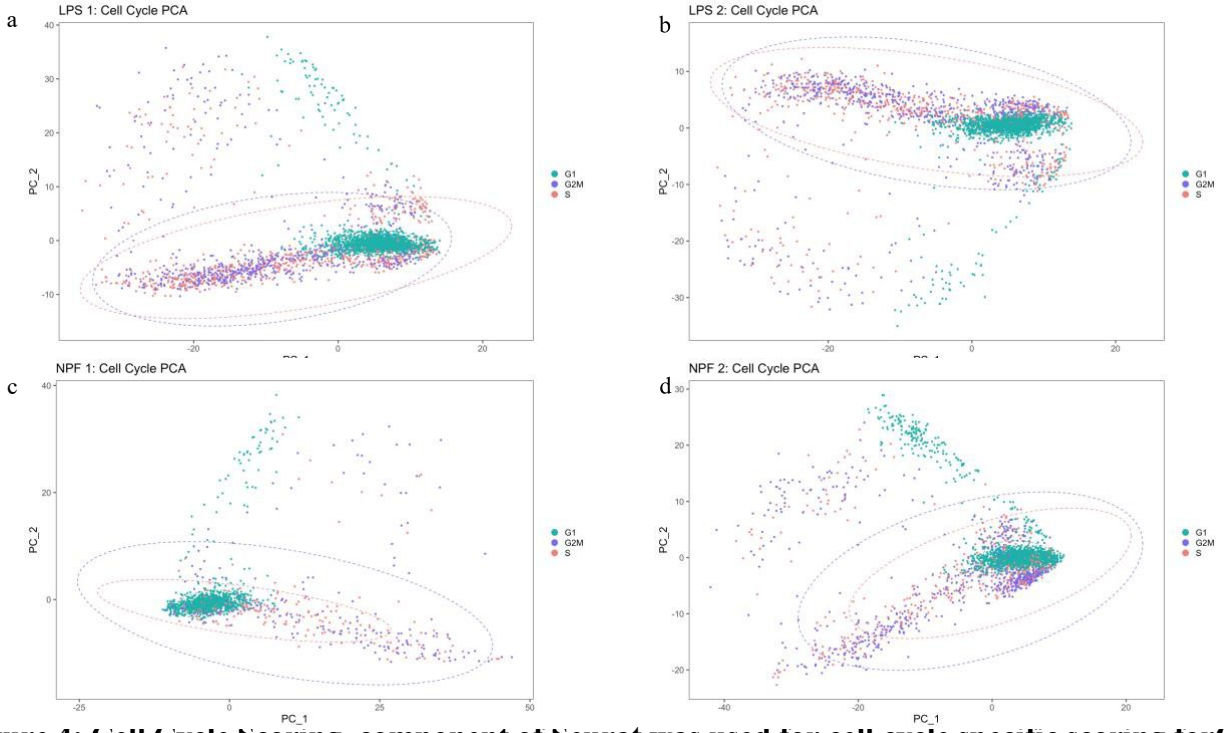

Figure 4: Cell Cycle Scoring, component of Seurat was used for cell cycle specific scoring forG1, G2/M, and S phases for LPS 1 (a), LPS 2 (b), NPF 1 (c), and NPF 2 (d).

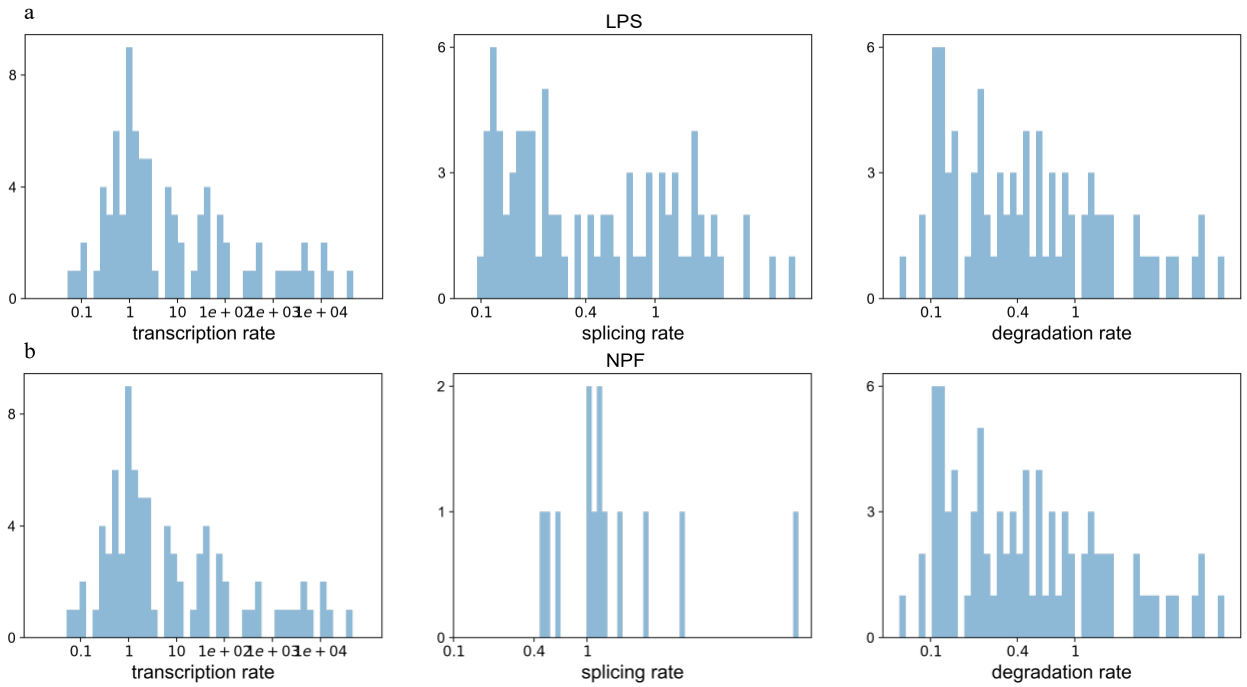

Figure 5: Integrated LPS (a) and NPF (b) scales of transcription rate, splicing rate, and degradation rate for all cells is shown.

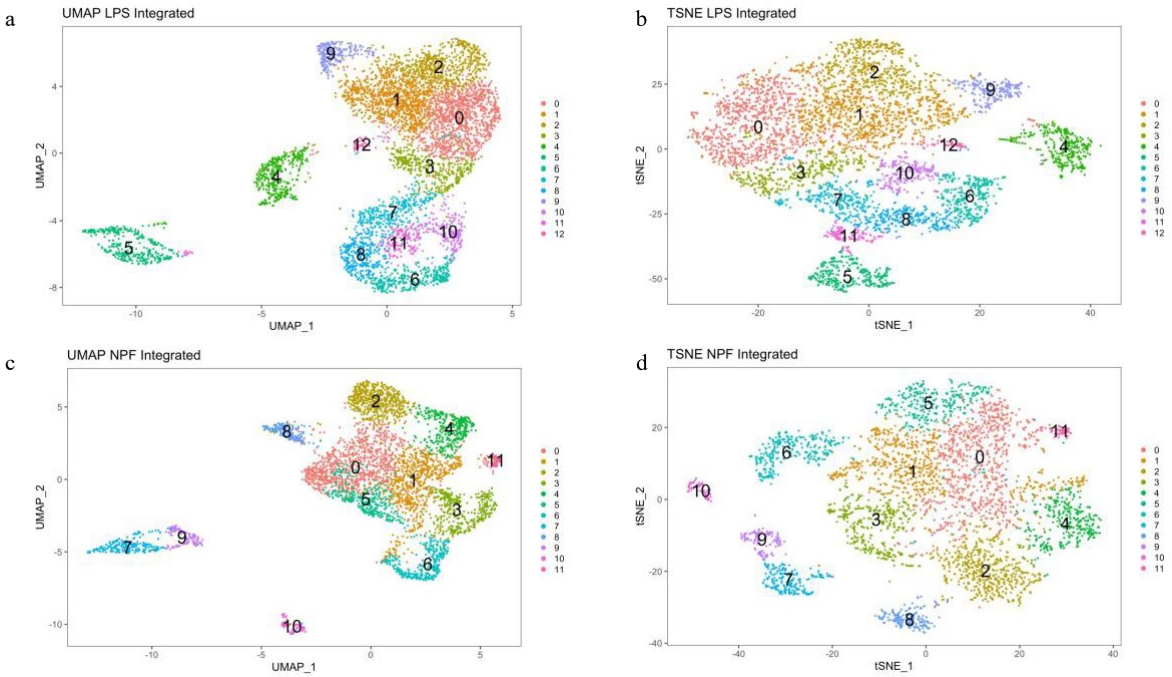

Figure 6: Visualization of LPS and NPF clusters through UMAP (a, c) and TSNE (b, d) performed by Louvain algorithm with multiple level refinement comprised of n=20 principal components from 1 to 20.

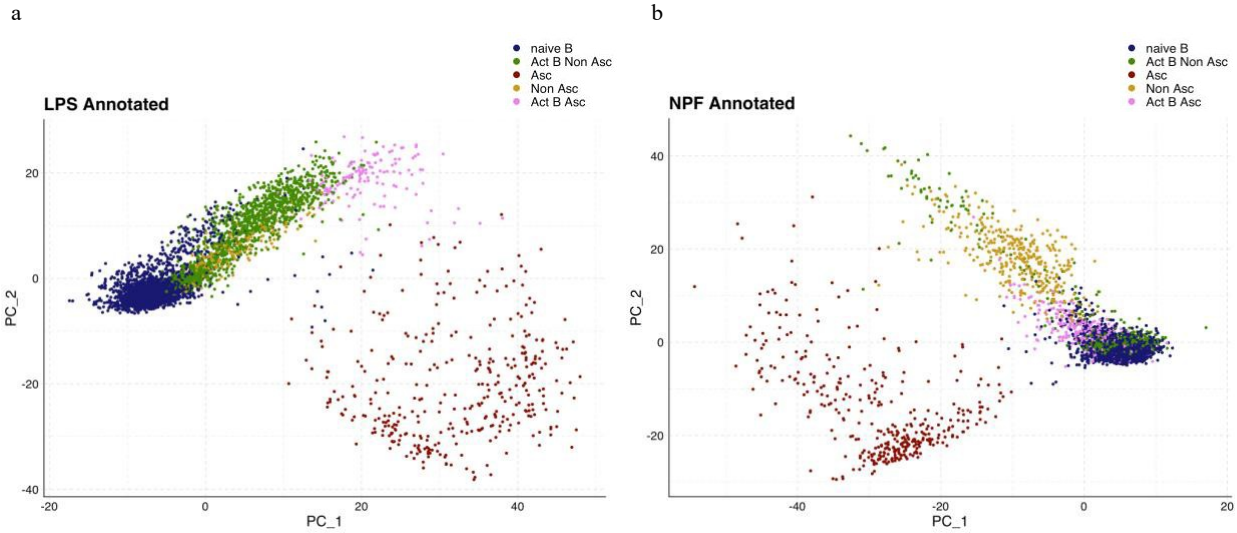

Figure 7: Principal components 1 and 2 showing all of the cells in LPS (a) and NPF (b) clustered based on cell type annotations. LPS dataset of n=5373 cells showed n=3312 naïve B, n=1335 ActB Non-Asc, n=142 ActB Asc, n=355 Asc, and n=228 Non-Asc. NPF dataset of n=4229 cells comprised of n=2304 naïve B, n=339 ActB Non-Asc n=385 ActB Asc, , n=838 Non-Asc, and n=363 Asc.

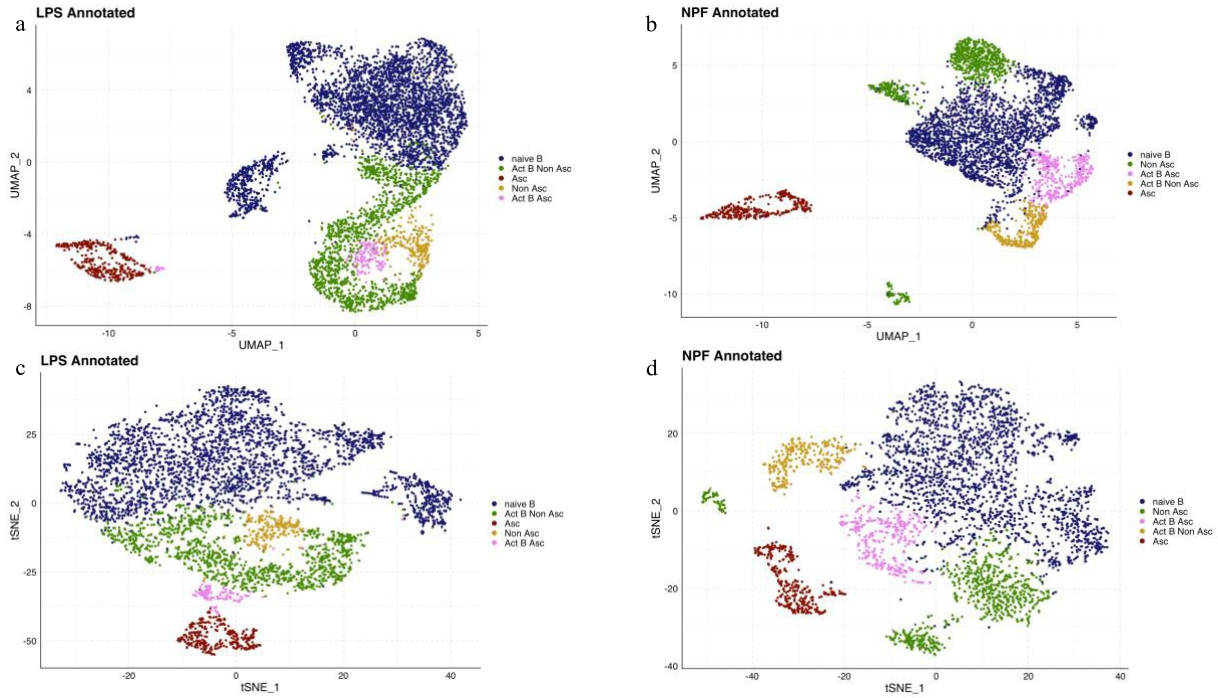

Figure 8: UMAP and TSNE showing all of the cells in LPS (a,c) and NPF (b,d) clustered based on cell type annotations. LPS dataset of n=5373 cells showed n=3312 naïve B, n=1335 ActB Non-Asc, n=355 Asc, n=228 Non-Asc, and n=142 ActB Asc. NPF dataset of n=4229 cells comprised of n=2304 naïve B, n=838 Non Asc, n=385 ActB Asc, n=339 ActB Non Asc, and n=363 Asc.

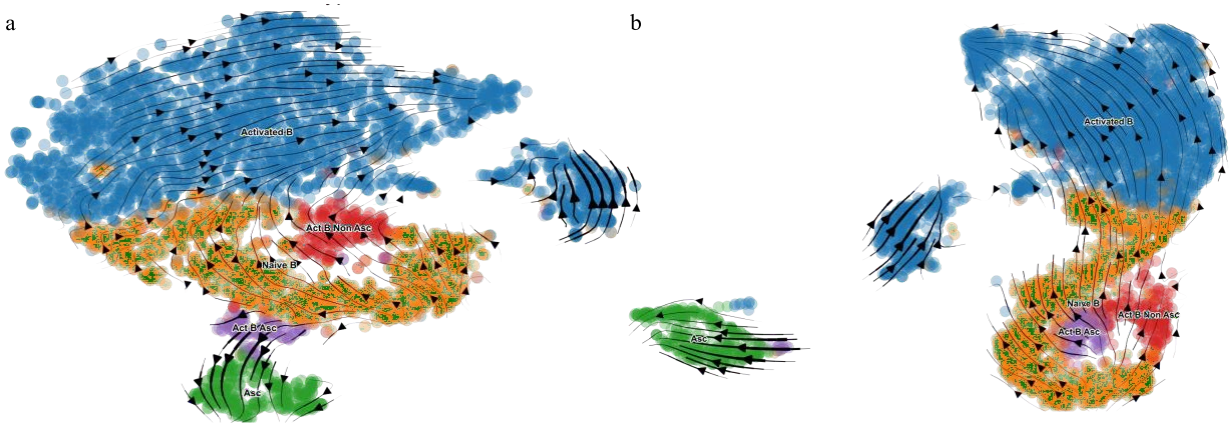

Figure 9: RNA velocity of LPS dataset n=5373 cells conducted with Dynamical Bayesian model inferences using principal components 1 to 20. The left image displays the TSNE and right shows UMAP of probabilistic trajectories.

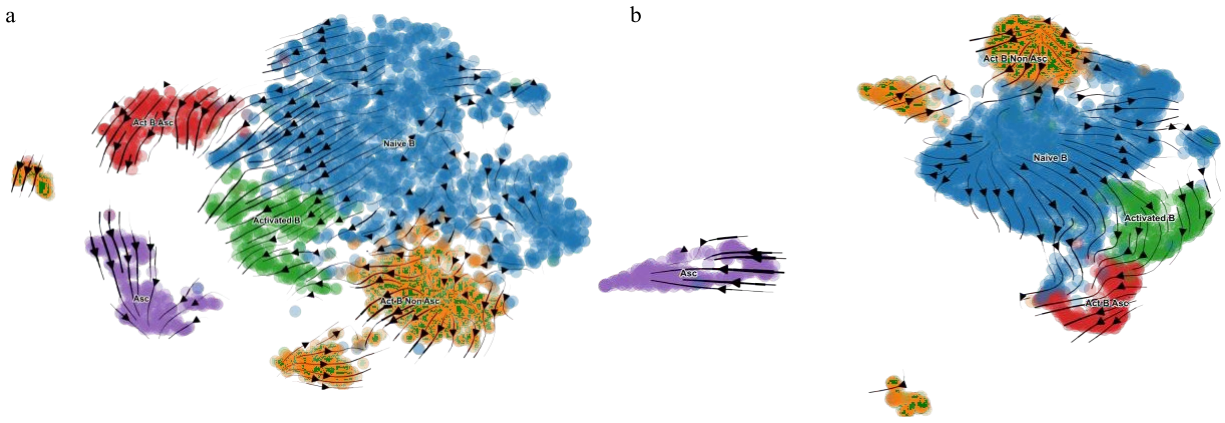

Figure 10: RNA velocity of NPF dataset n=4229 cells conducted with Dynamical Bayesian model inferences using principal components 1 to 20. The left image displays the TSNE and right shows UMAP of probabilistic trajectories.

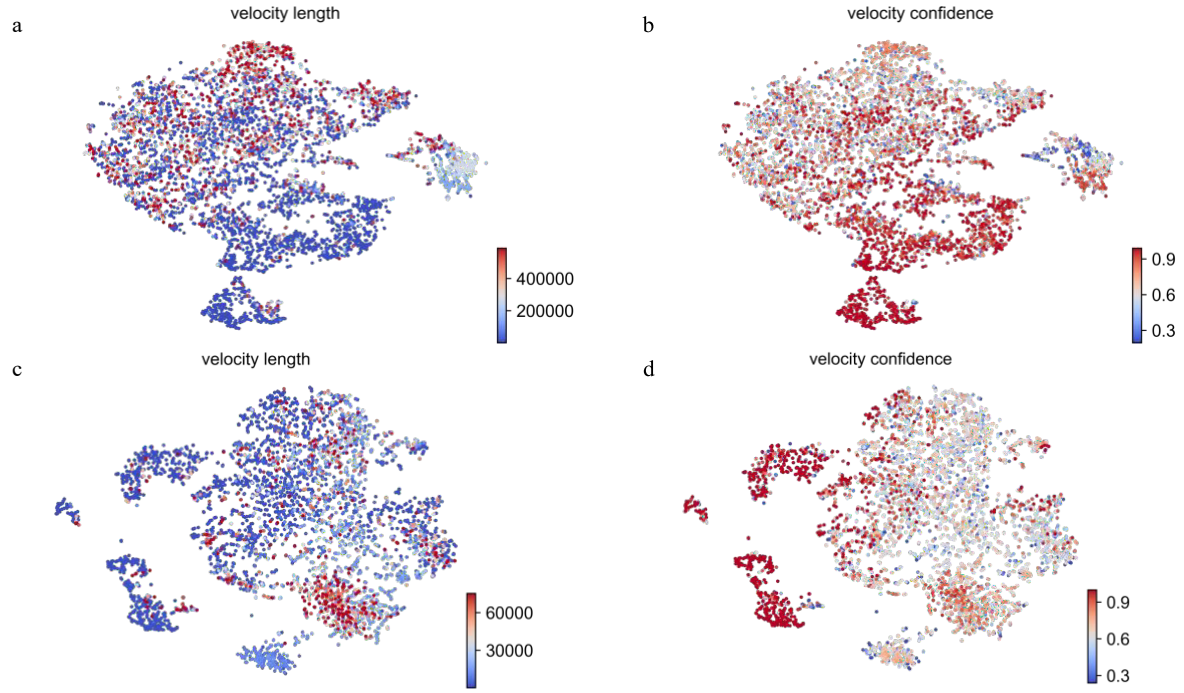

Figure 11: LPS and NPF TSNE based velocity length (a, c) and the confidence (b,d) of the velocities per cell given on a probabilistic scale.

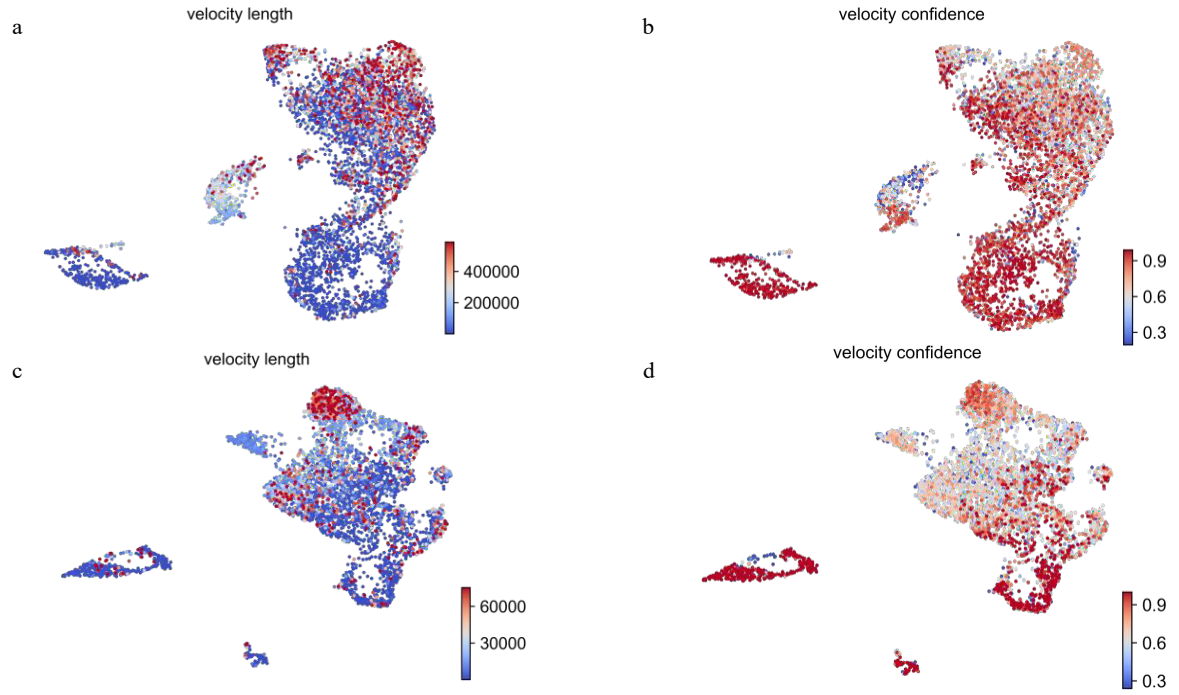

Figure 12: LPS and NPF UMAP based velocity length (a, c) and the confidence (b,d) of the velocities per cell given on a probabilistic scale.

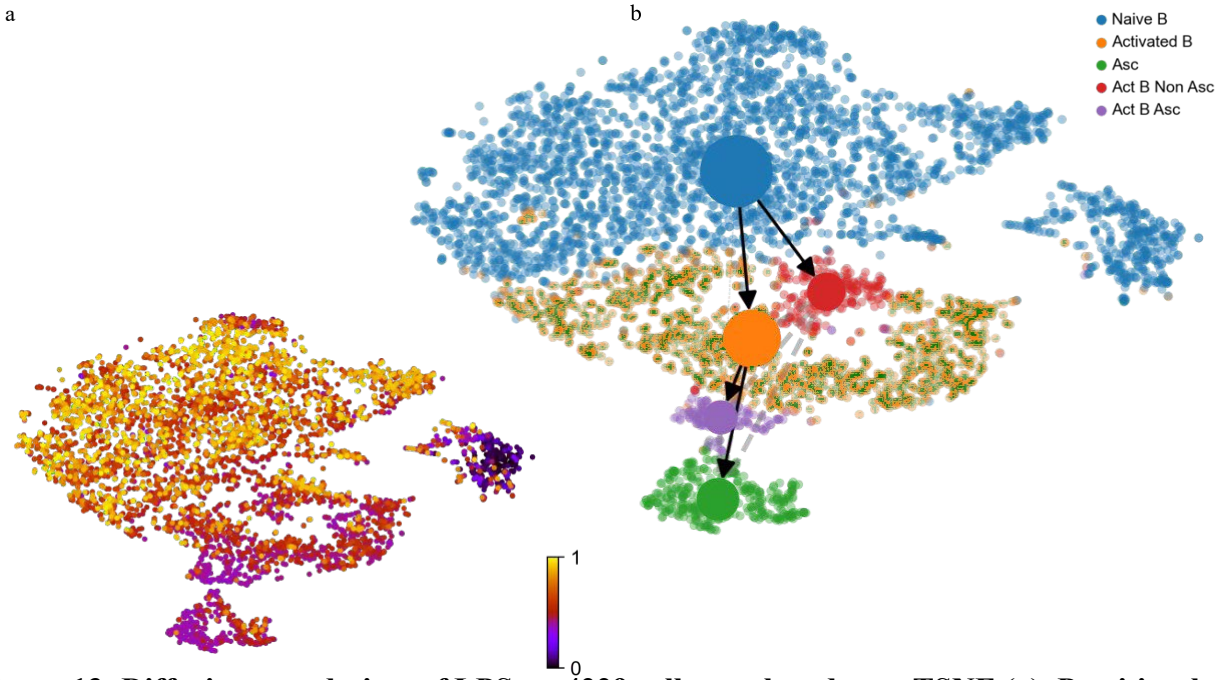

Figure 13: Diffusion pseudotime of LPS, n=4229 cells overlayed onto TSNE (a). Partition-based graphical abstraction (PAGA) of LPS TSNE (b) shows binary lineages from Naïve B to ASC cells or Activated B Cells.

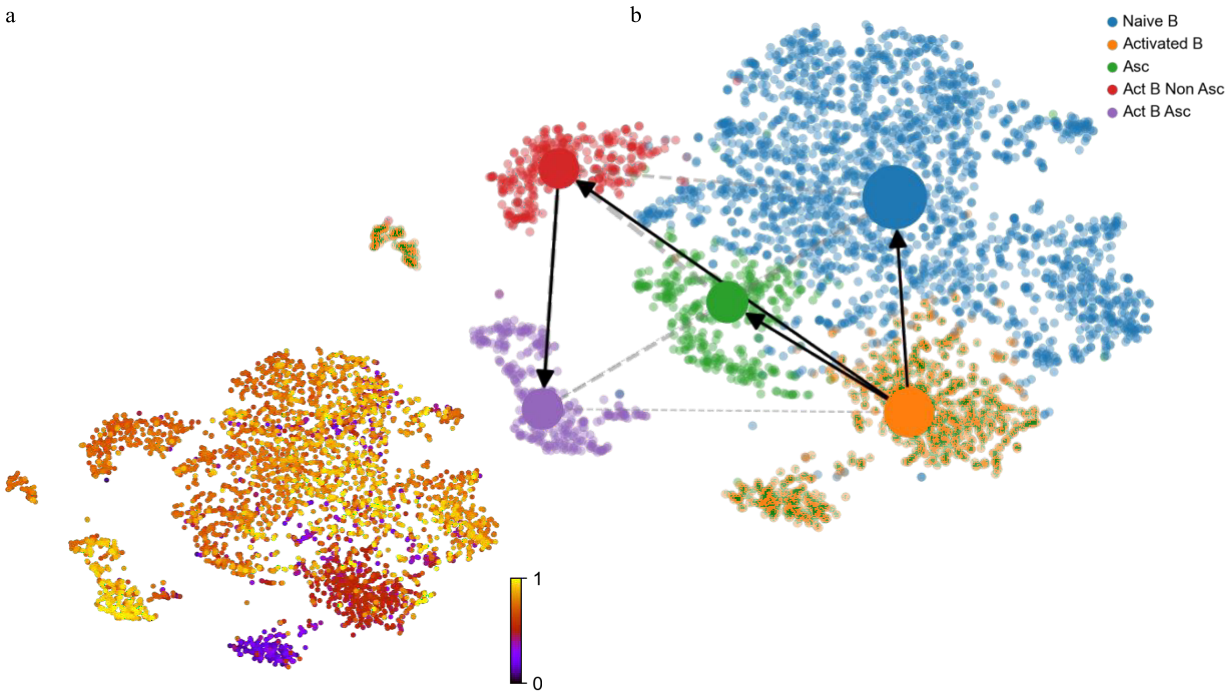

Figure 14: Diffusion pseudotime of NPF, n=5373 cells overlayed onto TSNE (a). Partition-based graphical abstraction (PAGA) of NPF TSNE (b) shows binary lineages from Naïve B to ASC or Activated B Cells.

Bibliography

1. Maloney, B.E., Perera, K.D., Saunders, D.R.D., Shadipeni, N. & Fleming, S.D. Interactions of viruses and the humoral innate immune response. *Clin Immunol* 212, 108351 (2020).
2. Silverstein, A.M. & Bialasiewicz, A.A. History of immunology. A history of theories of acquired immunity. *Cell Immunol* 51, 151-67 (1980).
3. Shishido, S.N., Varahan, S., Yuan, K., Li, X. & Fleming, S.D. Humoral innate immune response and disease. *Clin Immunol* 144, 142-58 (2012).
4. Cancro, M.P. & Tomayko, M.M. Memory B cells and plasma cells: The differentiative continuum of humoral immunity. *Immunol Rev* 303, 72-82 (2021).
5. Jaitin, D.A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343, 776-9 (2014).
6. Bendall, S.C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157, 714-25 (2014).
7. De Silva, N.S. & Klein, U. Dynamics of B cells in germinal centres. *Nat Rev Immunol* 15, 137-48 (2015).
8. Singh, A.K. & Jiang, Y. Lipopolysaccharide (LPS) induced activation of the immune system in control rats and rats chronically exposed to a low level of the organothiophosphate insecticide, acephate. *Toxicol Ind Health* 19, 93-108 (2003).
9. Matsuura, M. Structural Modifications of Bacterial Lipopolysaccharide that Facilitate Gram-Negative Bacteria Evasion of Host Innate Immunity. *Front Immunol* 4, 109 (2013).
10. Scharer, C.D. *et al.* Antibody-secreting cell destiny emerges during the initial stages of B-cell activation. *Nat Commun* 11, 3989 (2020).
11. Zheng, G.X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8, 14049 (2017).
12. Satpathy, A.T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* 37, 925-936 (2019).
13. La Manno, G. *et al.* RNA velocity of single cells. *Nature* 560, 494-498 (2018).
14. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888-1902 e21 (2019).
15. Lun, A.T.L. *et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol* 20, 63 (2019).
16. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 20, 296 (2019).
17. Traag, V.A., Waltman, L. & van Eck, N.J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 9, 5233 (2019).
18. Reid, J.E. & Wernisch, L. Pseudotime estimation: deconfounding single cell time series. *Bioinformatics* 32, 2973-80 (2016).
19. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* 38, 1408-1414 (2020).
20. Wolf, F.A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* 20, 59 (2019).
21. Haghverdi, L., Buttner, M., Wolf, F.A., Buettner, F. & Theis, F.J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* 13, 845-8 (2016).

22. Bhattacharya, S. *et al.* A bistable switch underlying B-cell differentiation and its disruption by the environmental contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin.*Toxicol Sci* 115, 51-65 (2010).
23. Chen, K.C., Wang, T.Y., Tseng, H.H., Huang, C.Y. & Kao, C.Y. A stochastic differentialequation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae. *Bioinformatics* 21, 2883-90 (2005).
24. Zhang, Q. *et al.* All-or-none suppression of B cell terminal differentiation by environmental contaminant 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicol Appl Pharmacol*268, 17-26 (2013).
25. Zhang, Q. *et al.* Embracing Systems Toxicology at Single-Cell Resolution. *Curr OpinToxicol* 16, 49-57 (2019).
26. Nancharaiah, Y.V., Rajadurai, M. & Venugopalan, V.P. Single cell level microalgal ecotoxicity assessment by confocal microscopy and digital image analysis. *Environ SciTechnol* 41, 2617-21 (2007).
27. Sun, R. *et al.* Toxicity in hematopoietic stem cells from bone marrow and peripheral blood in mice after benzene exposure: Single-cell transcriptome sequencing analysis.*Ecotoxicol Environ Saf* 207, 111490 (2021).
28. Anderson, D., Dhawan, A. & Laubenthal, J. The Comet Assay in Human Biomonitoring. *Methods Mol Biol* 20