Single cell analysis of Interferon-β-stimulated blood: Focusing on trajectory and differential expression in cMonocytes

Alice Saparov

Dataset: Interferon-beta-stimulated blood: ifnb_14k

Introduction and Description of Dataset

The dataset originated from a research article that was interested in developing new methods for doublet detection (droplets containing two cells) in single cell analysis, as well as determining sample identities in singlets (droplets with a single cell). They evaluated their algorithm, demuxlet, on single cell expression data from eight pooled lupus patients using peripheral blood mononuclear cells (PBMCs). The following experimental steps were taken to prepare the cells for analysis. The cells were Ficoll separated (procedure to isolate mononuclear cells) and frozen (cryopreserved). They were then thawed in a water bath, washed, and suspended in a buffer again. The cells were subsequently treated, incubated, filtered, reconcentrated, and finally pooled (all donor cells into one sample). Before pooling, half the cells from each sample were treated with IFN-β and the rest remained untreated. Samples were stained, then analyzed and sorted using fluorescent tags (BD FACSAria Fusion instrument). They isolated RNA from sorted CD4+ T cells and performed qPCR. Using a 10x Chromium instrument (10x Genomics) they loaded and sequenced the cellular suspensions and used a custom program which yielded >90% transcript capture [1]. The aims of this project are now to investigate the single cell dataset generated by the research article for further analysis.

Methodology

Before beginning any quality control, there were 13,999 cells and 14,053 genes. As noted in the readme file provided, all the mitochondria were already removed so they will not be used for QC. Based on the nFeature_RNA plot (Fig. 1), cells should be QCed to have the expected amount of RNA within them so the cells were filtered to have nFeature_RNA greater than 500. There was only one metric (orig.ident) that was provided in the meta data for potential batch effects. Based on the plot of cells grouped by their origin identity (stimulated v. control) (Fig. 2), it will be necessary to use Harmony to perform batch correction. Even though this is a factor of interest to investigate and analyze in downstream analysis, the batch effects appear quite disruptive. While two of the clusters seem to primarily only contain one type of orig.ident, the other clusters have clear and large separations of the types and the batch correction would help mix them. However, performing batch effect correction on orig.ident may influence downstream analyses and potentially diminish some of the true effects or differences between the stimulated and control groups. It is still useful to try and correct for some of the technical variation that is observed though.

Next, I performed log normalization of the data with a scale factor of 10,000, selected 2,000 variable features using a variance stabilizing transformation (vst) selection method, and then I scaled and centered the data. In order to reduce the dimensionality of the data, I performed PCA and plotted an elbow plot to determine the best number of dimensions to use. Based on the elbow plot (Fig. 3), the (explained) standard deviation appeared to stop decreasing between 20-25 principal components so I chose to proceed using 20 principal components in subsequent analysis. I then generated my UMAP with the dimension parameter

set to 20 and plotted a UMAP of orig.ident (Fig. 2) where I evaluated for potential batch effects. After determining the need for batch correction, Harmony was run on orig.ident to perform the correction and another UMAP was plotted grouping by orig.ident (Fig. 4) to ensure sufficient correction had occurred. Using the batch corrected data, the neighbors and clusters were calculated using 20 dimensions again. A resolution of 0.1 was used for the clustering using the original Louvain algorithm. A UMAP plot was generated (Fig. 5), grouped by cluster, to confirm the number of clusters produced at the given resolution was consistent with the number of clusters found in the original research article (8 clusters).

In order to identify gene markers for each cluster a Wilcoxon Rank Sum test was used to identify differentially expressed genes between the cells. The test was conducted with a log fold-change threshold of 0.5 and set to return only positive gene markers. Next, a table was created with the top 10 ranked marker genes by average log fold-change for each cluster in order to classify the cell type present in each cluster. There were some difficulties in assigning the proper cell types to each cluster. Originally, I attempted to use Supplementary Table 1 from the original study which provided a list of marker genes found for each of their clusters [1]. This did not appear to be sufficient to classify the clusters I found since occasionally the genes were missing or present in multiple clusters. I resorted to using the UCSC Cell Browser, specifically the Global Immune Cell Atlas which is a reference atlas for various blood cell types [2] since I am also analyzing PBMCs (a blood cell).

Plugging in the top ranked marker genes for each cluster I found into the atlas, and then occasionally trying to verify my cluster assignments using the Supplementary Table 1, I was able to assign each cluster to a cell type. Cluster 0 contained markers for CD3D, CD3E, and GIMAP4 which were also found to be present in CD8+ T cells (Tc). Cluster 1 had APOBEC3B, LYZ, and IDO1 which the cell atlas reported as being expressed in CD14+CD16- monocytes (cMono). Cluster 2 contained gene markers for GNLY, NKG7, and GZMB which were consistent with natural killer cells (NK). Cluster 3 was classified as B cells since it had high expression of CD79A, MS4A1, and CD83. Cluster 4 contained high expression for CXCL11, DEFB1, and CH25H so it was deemed to be CD14+CD16+ monocytes (ncMono). Cluster 5 was classified as megakaryocytes (Mkc) due to expression of PPBP, PF4, and GNG11. Cluster 6 contained TSPAN13, CD74, and IRF8 which appeared to be associated with dendritic cells (DC) in the cell atlas. Finally, cluster 7 was very small and contained genes (HBB, BLVRB, etc.) that the atlas classified as erythroid cells, while this cell type was not present in the original research article I was unable to find many markers for CD4+ T cells in my clusters. It appears the CD4+ T cell type may have accidentally become aggregated into another cluster/cell type due to different QC, preprocessing, or clustering approaches taken compared to the original research article.

Once the cluster cell types were identified, a UMAP was created with the cluster names (Fig. 6), and the top ten markers for each cluster based on their average log fold-change were plotted on a heatmap colored by their expression values (Fig. 7). The top 20 ranked marker genes by average log fold-change for each cell type are also included in a table where the genes are sorted by their average log-fold change within each respective cell type (Table 1).

Further Analyses

After cluster cell types were identified, trajectory analysis was performed in an attempt to quantify and observe differences between stimulated vs control cells since stimulated cells are

expected to undergo genome level transcriptional changes. Using monocle3 the seurat dataset was converted to a cell data set and the data was once again preprocessed, batch corrected, and the dimensionality was reduced to be ready for trajectory analysis. Two UMAPs were created to observe the distribution of cells by cell type (Fig. 8) and by orig.iden (stimulated v control). It was observed that cMono cells contained a separation between stimulated and control cells within the cluster (Fig. 9). Focusing on conducting trajectory analysis of cells within the cMono type, the cluster was subset for subsequent analysis. This step must be done manually as it prompts cluster selection through an interactive window (Fig. 10) (file fails to knit if this step is left included). A trajectory analysis was performed on the isolated cMono cells where a starting point was selected in the control cells, and the resultant trajectory was plotted and colored by pseudotime (Fig. 11). A correlation was performed to evaluate the association between pseudotime and stimulation status and it was found to be 0.89. This indicates that a positive correlation exists between stimulation status within cMono cells relative to pseudotime. While it is not yet possible to show that changes in certain gene expression may cause the control cells to progress to the identity of stimulated cells over time, it is possible to further investigate what gene expression changes are occurring within cMono cells.

Subsequently, I performed differential gene expression between the stimulated and control cells within the cMono cell type cluster. First, I subset the data to only contain cMono cells and then proceeded to run differential expression analysis between the orig.ident (control v stimulated). The same finding markers procedure was run as before when determining differential gene expression between clusters. The top ten differentially expressed genes were plotted on a heat map (Fig. 12), ranked by their average log fold-change and colored by their expression, for stimulated and control cells. It was found that control cMono cells had higher expression for the following genetic markers: IL8, CD14, CLEC5A, VCAN, IER3, S100A8, CD9, IL1B, and INSIG1. The stimulated cMono cells showed higher expression in these genetic markers: CXCL10, CCL8, ISG15, CXCL11, TNFSF10, IFIT3, IFIT1, RSAD2, IFIT2, and APOBEC3B. A table of these marker genes is also included (Table 2). Based on the UCSC cell atlas, the control cell gene markers were found to be more commonly classified as classical monocytes, and high monocytes are usually associated with an inflammatory response and chronic infection or disease [3]. However, the stimulated gene markers did not show up as much within the cell atlas, or were occasionally more diversely associated (slight associations with various T cells, monocytes, or macrophages) and these cell markers were harder to classify. Further investigation into the stimulated cMono markers could provide future targets for battling inflammation as IFN-β is an immunomodulator and IFN-β stimulation is intended to decrease inflammation [4].

Citations

1. Kang, H., Subramaniam, M., Targ, S. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol* 36, 89–94 (2018). https://doi.org/10.1038/nbt.4042
2. Matthew L Speir, Aparna Bhaduri, Nikolay S Markov, Pablo Moreno, Tomasz J Nowakowski, Irene Papatheodorou, Alex A Pollen, Brian J Raney, Lucas Seninge, W James Kent, Maximilian Haeussler, UCSC Cell Browser: visualize your single-cell data,

Bioinformatics, Volume 37, Issue 23, 1 December 2021, Pages 4578–4580, https://doi.org/10.1093/bioinformatics/btab503
3. Monocytes: A Type of White Blood Cell — What Are Normal Ranges? Cleveland Clinic. Accessed May 5, 2023. https://my.clevelandclinic.org/health/body/22110-monocytes
4. Kasper LH, Reder AT. Immunomodulatory activity of interferon-beta. *Ann Clin Transl Neurol*. 2014;1(8):622-631. doi:10.1002/acn3.84

Figures and Tables

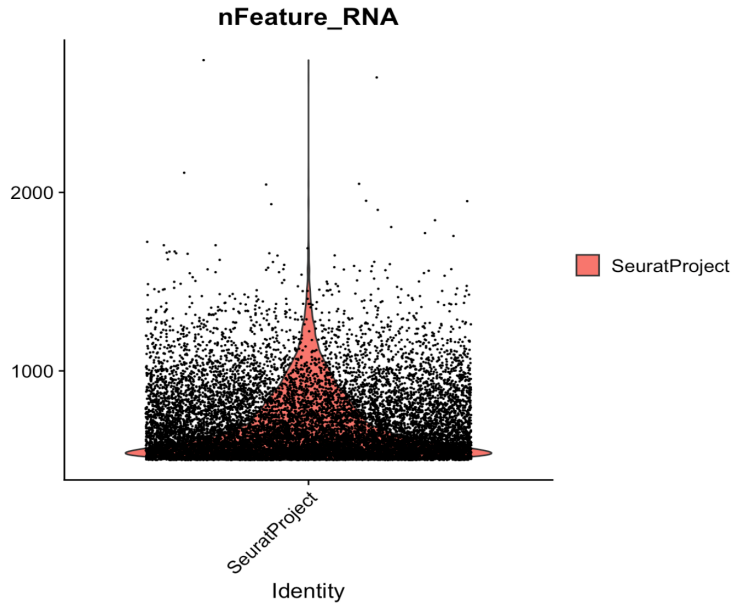

Figure 1. nFeature_RNA plot showing the number of genes that were detected in each cell.

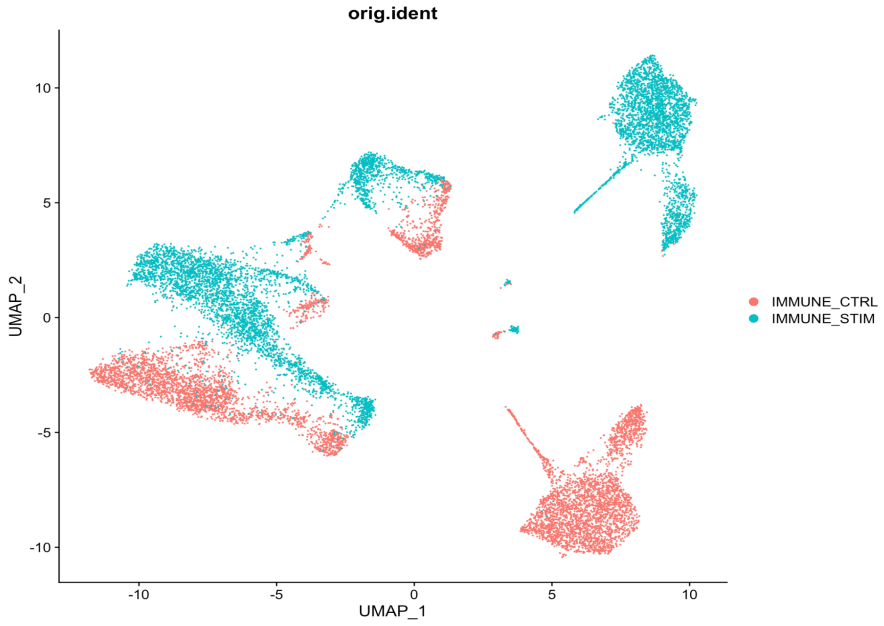

Figure 2. UMAP of cells grouped by metadata variable ‘orig.ident’ prior to batch correction.

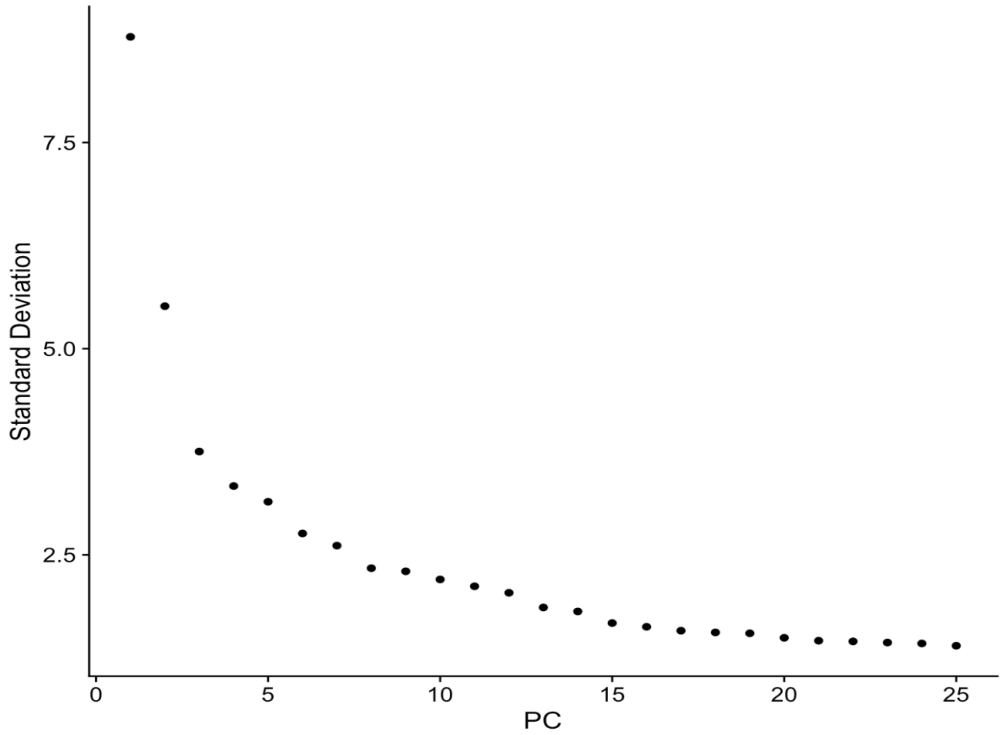

Figure 3. Elbow plot to select the number of dimensions to keep for dimensionality reduction.

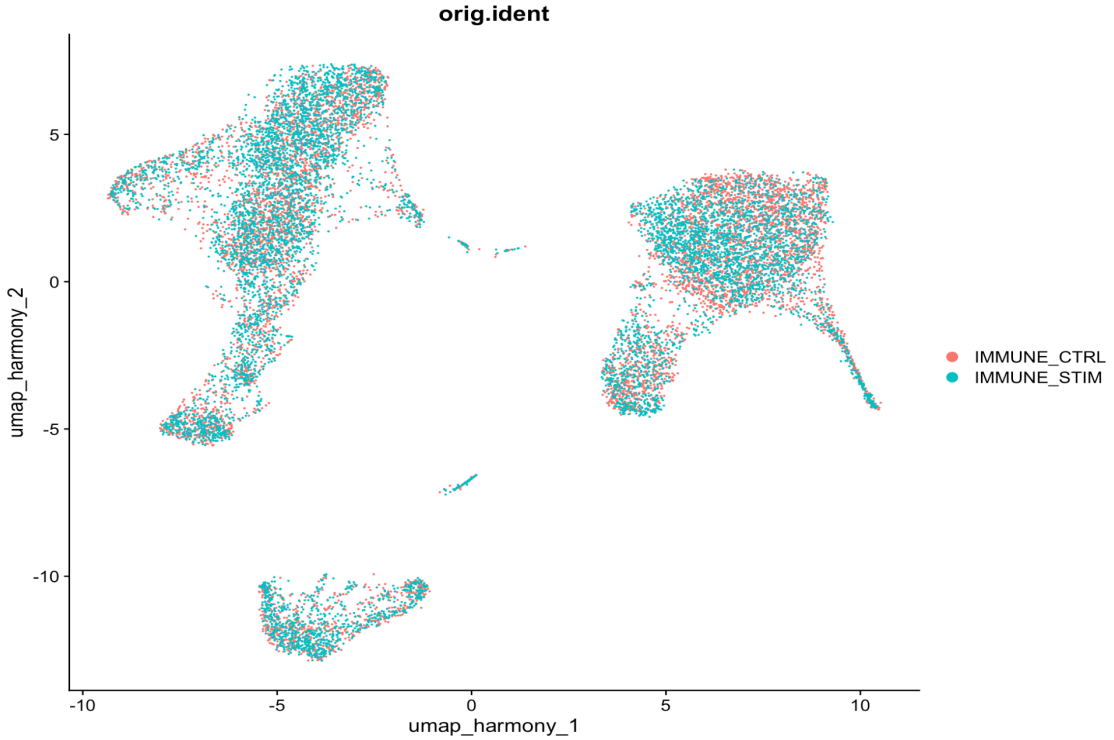

Figure 4. UMAP of cells grouped by ‘orig.ident’ post harmony batch correction.

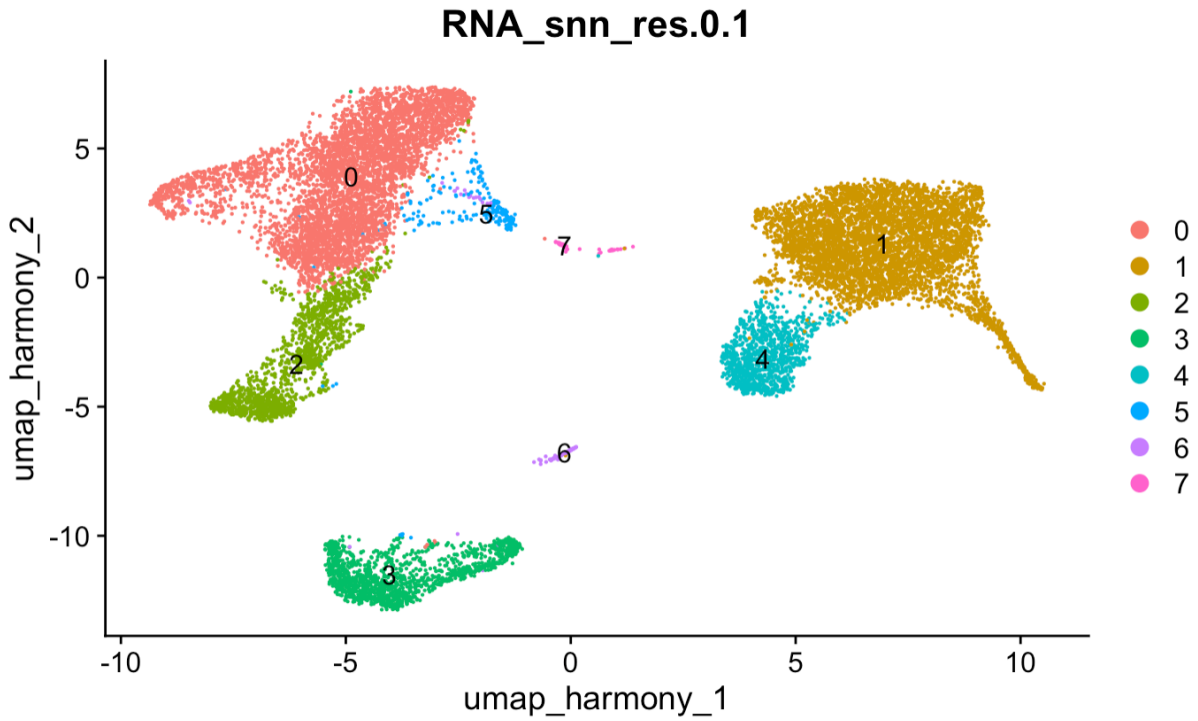

Figure 5. UMAP grouped by cell cluster to check the cluster resolution parameter and to confirm number of clusters.

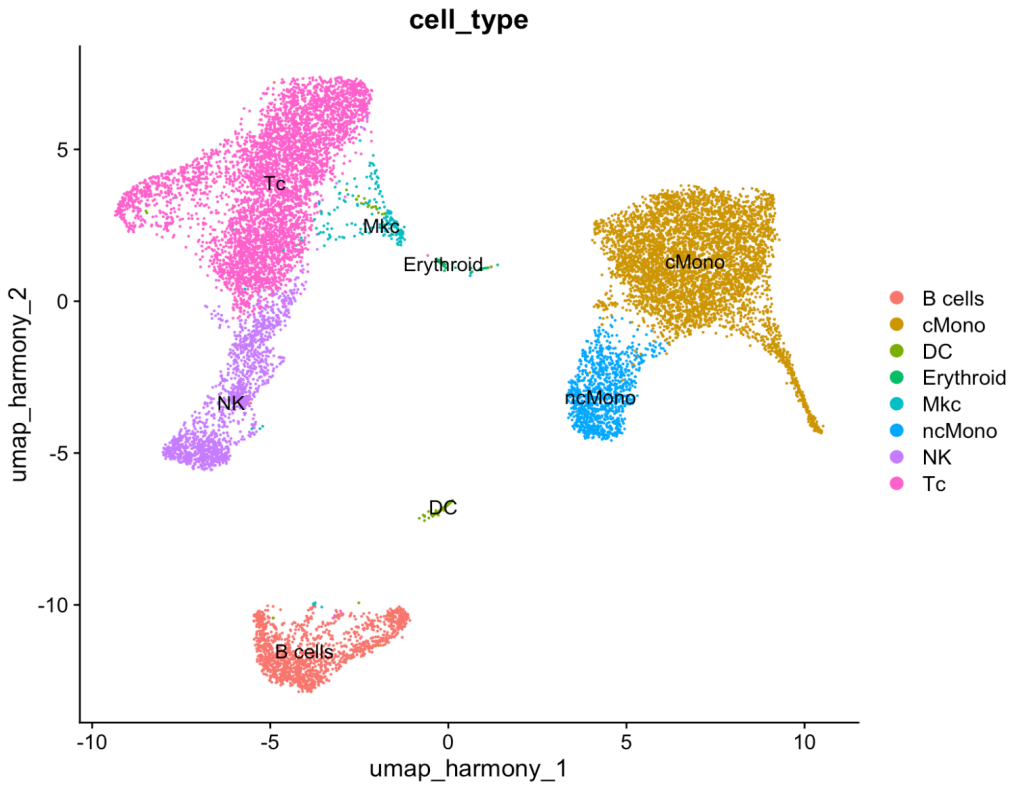

Figure 6. UMAP grouped and labeled by assigned cell type.

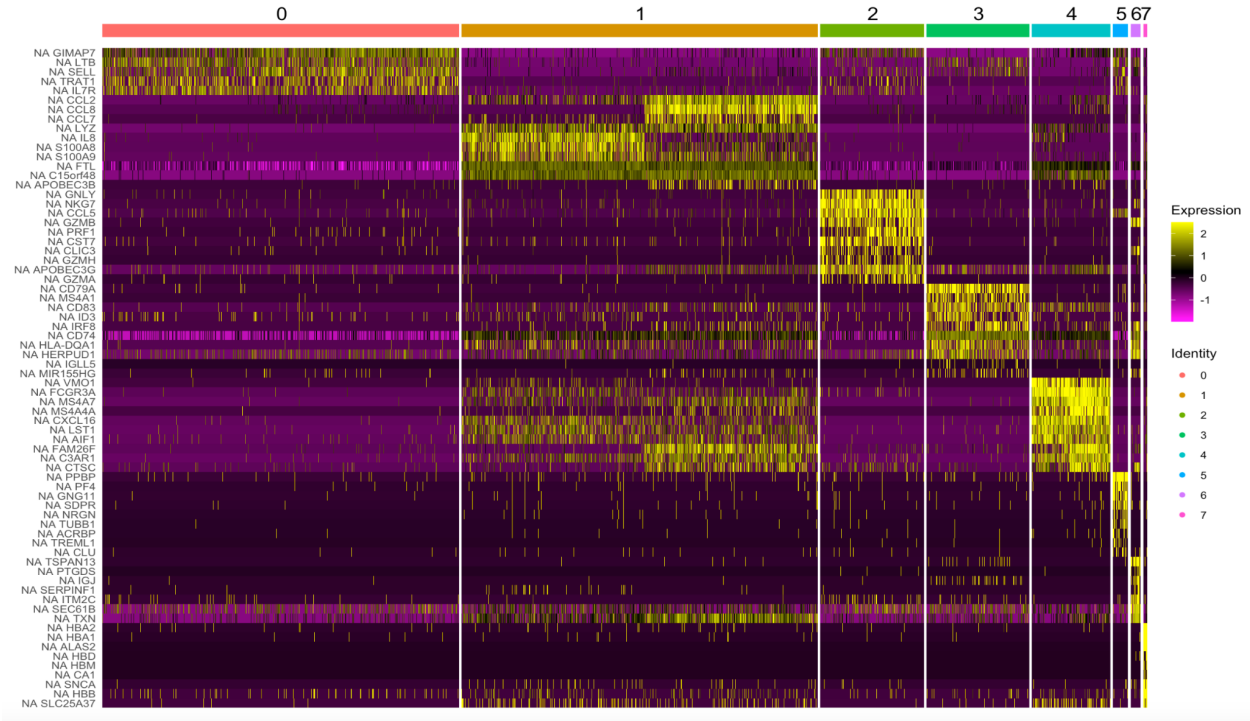

Figure 7. Heatmap of top ten positive marker genes for each cluster based on average log fold-change colored by their gene expression.

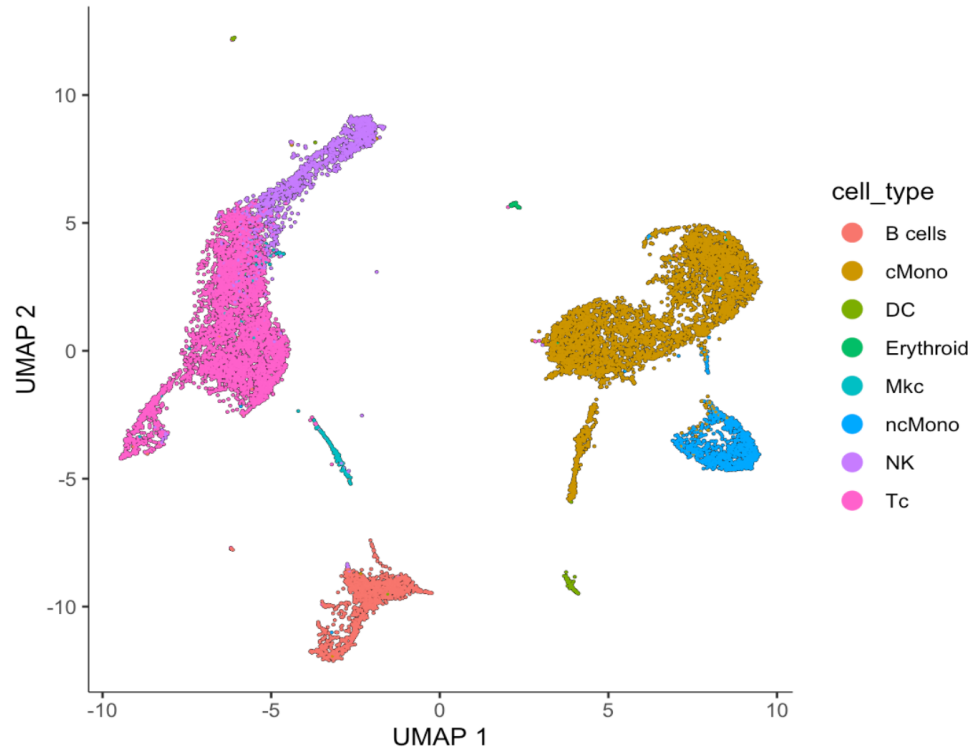

Figure 8. UMAP showing the distribution of cells by cell type after monocle3 preprocessing.

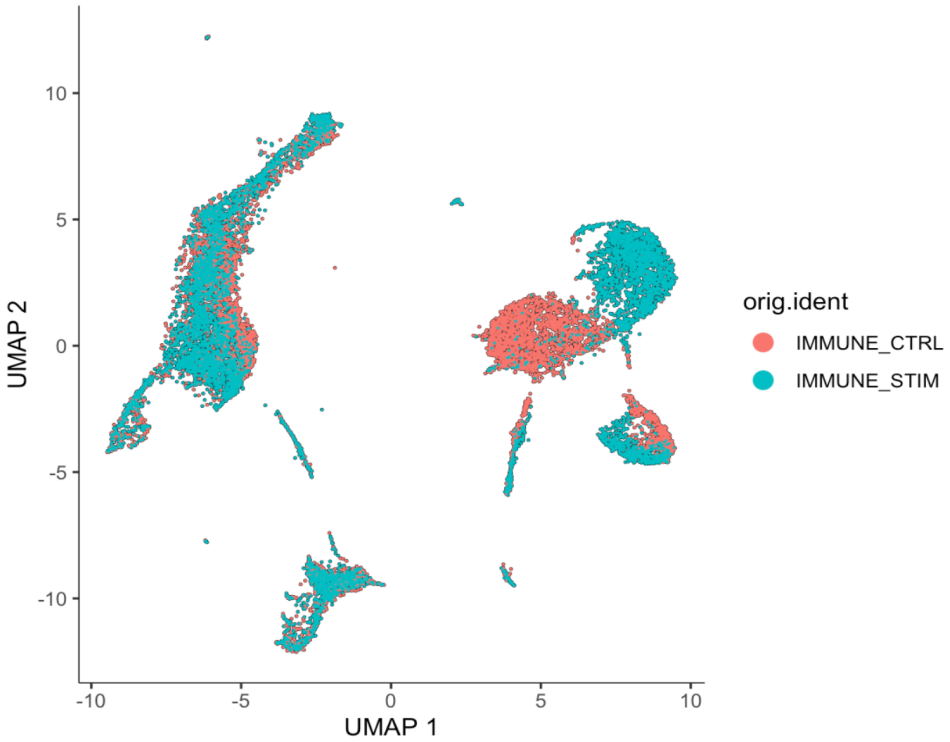

Figure 9. UMAP showing the distribution of cells by origin identity after monocle3 preprocessing.

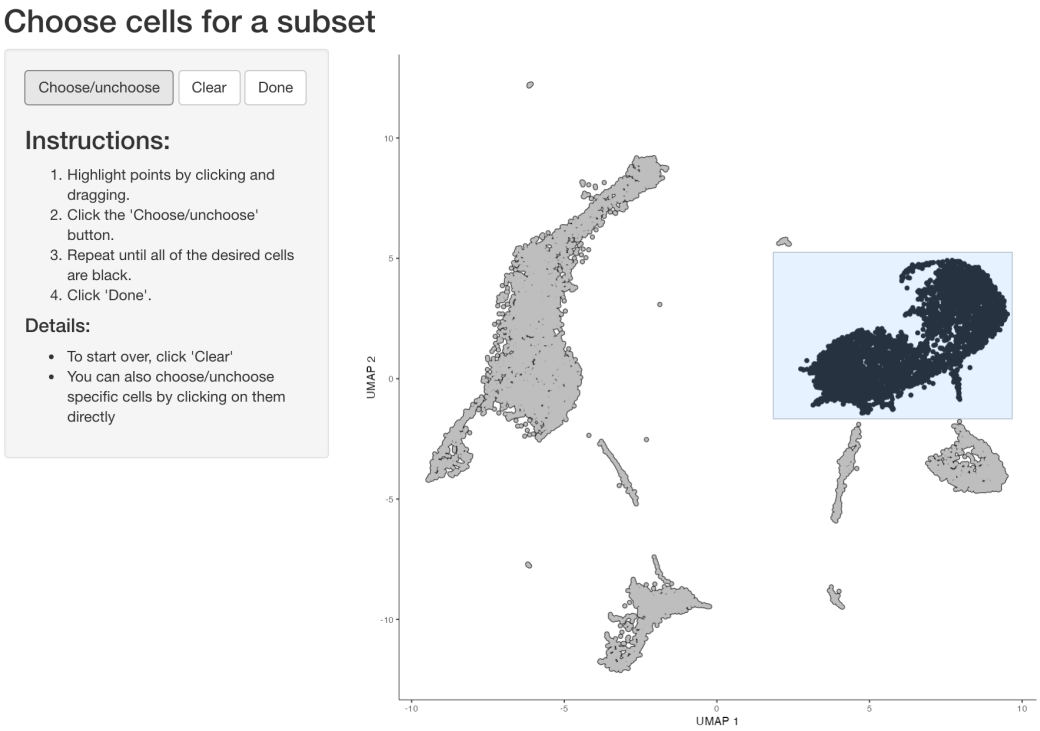

Figure 10. Interactive window for cluster selection to perform trajectory analysis on cMono cells.

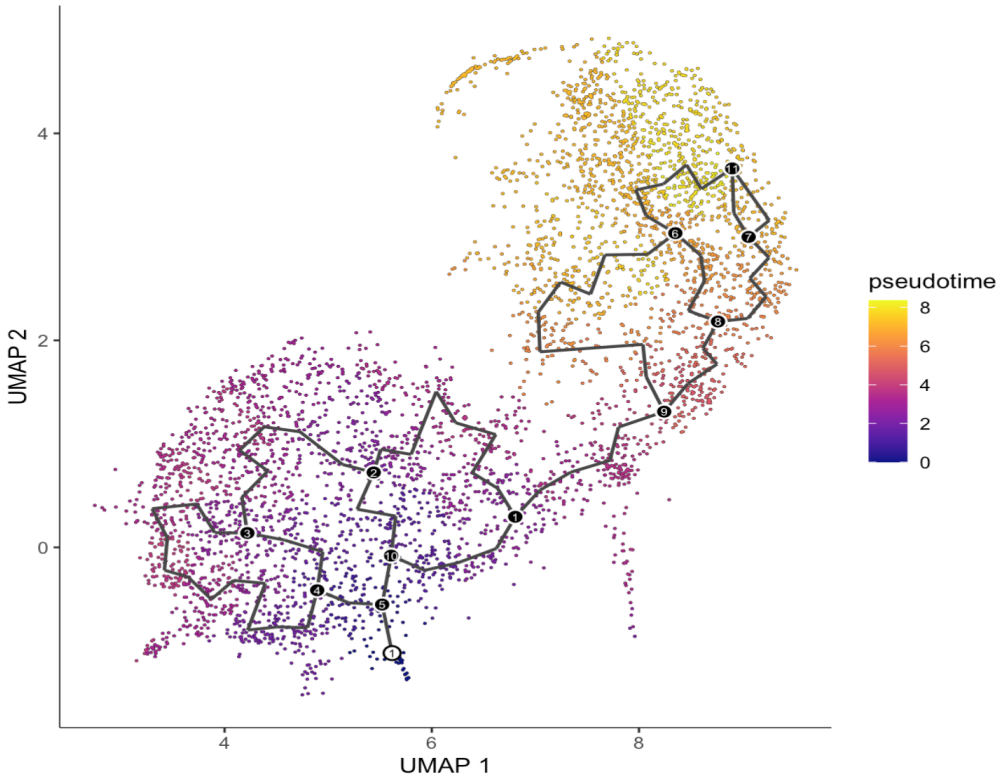

Figure 11. Trajectory plot for selected cMono cells colored by pseudotime.

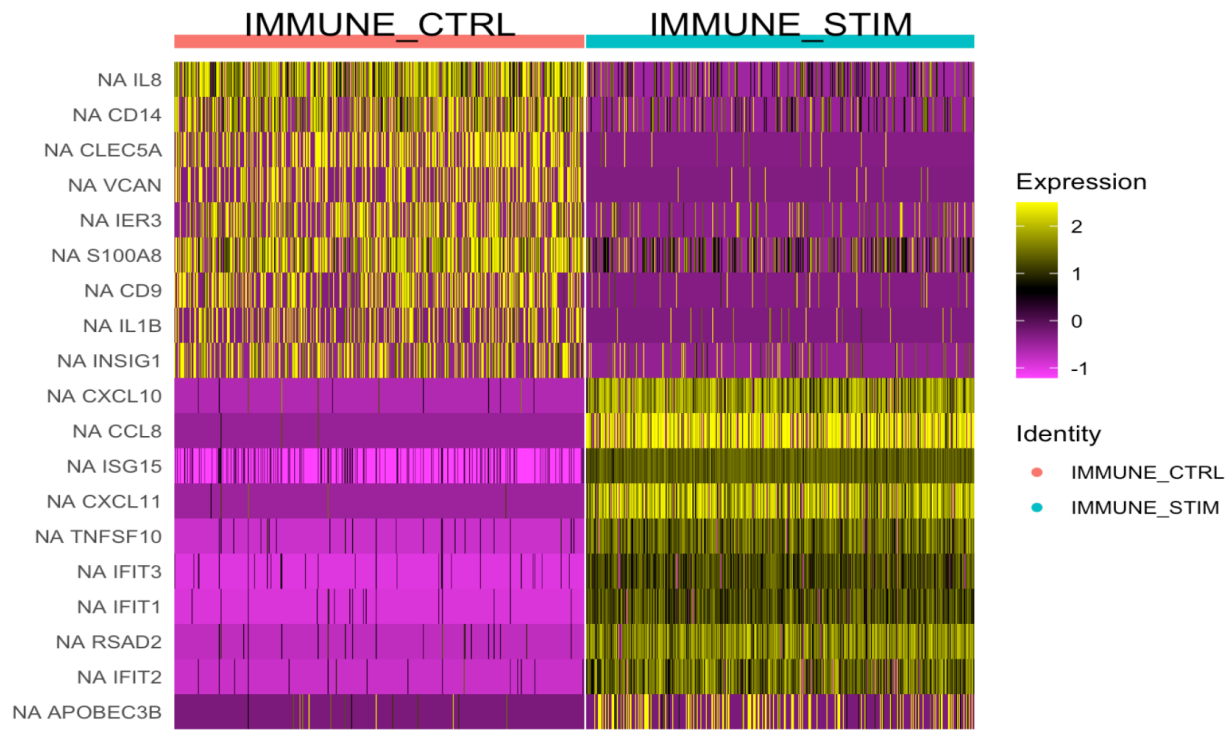

Figure 12. Heatmap of top ten positive marker genes for control v stimulated cells within the cMono cells colored by gene expression.

p_val	avg_log2FC	pct.1	pct.2	p_val_adj	cluster	gene	cell_type
0	2.31377739755873	0.827	0.273	0	0	NA GIMAP7	Tc
0	2.25587010773279	0.704	0.161	0	0	NA LTB	Tc
0	2.18702539435211	0.622	0.189	0	0	NA SELL	Tc
0	1.98331322614968	0.704	0.21	0	0	NA LDHB	Tc
0	1.95894077290277	0.633	0.125	0	0	NA CD3D	Tc
0	1.88960873143539	0.41	0.063	0	0	NA TRAT1	Tc
0	1.83287979259878	0.53	0.145	0	0	NA IL7R	Tc
0	1.77072971678953	0.488	0.137	0	0	NA GIMAP5	Tc
0	1.59941493924462	0.519	0.214	0	0	NA LEPROTL1	Tc
0	1.5984662930091	0.949	0.816	0	0	NA PABPC1	Tc
0	1.59534236101271	0.262	0.026	0	0	NA LEF1	Tc
0	1.5862888471472	0.35	0.069	0	0	NA PIK3IP1	Tc
0	1.54924782833334	0.917	0.656	0	0	NA TMEM66	Tc
0	1.54083697653693	0.493	0.142	0	0	NA OCIAD2	Tc
0	1.47692152279565	0.333	0.063	0	0	NA CD3E	Tc
0	1.46503432013293	0.34	0.07	0	0	NA LCK	Tc

0	1.45969882259484	0.481	0.203	0	0	NA GIMAP4	Tc
0	1.44687056945745	0.56	0.249	0	0	NA ZFP36L2	Tc
0	1.42786710570256	0.777	0.281	0	0	NA CCR7	Tc
1.05101837949326E-296	1.40989377827749	0.208	0.024	1.47699612870188E-292	0	NA PASK	Tc
0	5.12845835603052	0.702	0.098	0	1	NA CCL2	cMono
0	4.96888921212883	0.428	0.084	0	1	NA CCL8	cMono
0	4.06833325076457	0.343	0.019	0	1	NA CCL7	cMono
0	4.01438068613936	0.865	0.085	0	1	NA LYZ	cMono
0	4.01359349668021	0.621	0.049	0	1	NA IL8	cMono
0	4.00174908142069	0.648	0.024	0	1	NA S100A8	cMono
3.01470626612244E-200	3.70741427466942	0.172	0.03	4.23656671578187E-196	1	NA APOBEC3B	cMono
0	3.70517601841883	0.697	0.029	0	1	NA S100A9	cMono
0	3.55284416132631	1	0.87	0	1	NA FTL	cMono
0	3.44096586846942	0.984	0.148	0	1	NA C15orf48	cMono
0	3.43622934021986	0.527	0.072	0	1	NA IDO1	cMono
0	3.16519494379142	0.611	0.117	0	1	NA APOBEC3A	cMono
0	3.07918371762949	0.951	0.199	0	1	NA LGALS3	cMono
0	2.95412335333841	0.883	0.147	0	1	NA CTSB	cMono
0	2.93808983577949	0.976	0.227	0	1	NA LGALS1	cMono
0	2.9143778994178	0.994	0.286	0	1	NA CD63	cMono
0	2.76135820443388	0.674	0.03	0	1	NA PLA2G7	cMono
0	2.74325373832812	0.729	0.128	0	1	NA CTSL	cMono
0	2.72542875904011	0.987	0.284	0	1	NA ANXA5	cMono
0	2.68112971856168	0.444	0.068	0	1	NA IL1RN	cMono
0	5.78807421577339	0.62	0.034	0	2	NA GNLY	NK
0	4.53324857659908	0.785	0.047	0	2	NA NKG7	NK
0	4.44904777423633	0.795	0.089	0	2	NA CCL5	NK
0	4.17016540971768	0.613	0.03	0	2	NA GZMB	NK
0	3.44359731488354	0.482	0.019	0	2	NA PRF1	NK
0	2.91700089152866	0.572	0.041	0	2	NA CST7	NK
0	2.79111011313487	0.372	0.018	0	2	NA CLIC3	NK
0	2.73658439236609	0.356	0.009	0	2	NA GZMH	NK
0	2.72903114408616	0.742	0.218	0	2	NA APOBEC3G	NK
0	2.68077146298959	0.329	0.011	0	2	NA GZMA	NK
0	2.62403393425772	0.392	0.007	0	2	NA KLRD1	NK
0	2.55205453709808	0.268	0.008	0	2	NA FGFBP2	NK

0	2.52965839147881	0.382	0.02	0	2	NA CTSW	NK
0	2.00793129098424	0.303	0.025	0	2	NA CXCR3	NK
1.30719935659476E-287	1.92841414456393	0.61	0.237	1.83700725582262E-283	2	NA RARRES3	NK
7.86108013502181E-249	1.84069790156633	0.217	0.025	1.10471759137462E-244	2	NA LAG3	NK
0	1.81878599126285	0.292	0.035	0	2	NA CD8A	NK
1.41833159793295E-254	1.78884583055269	0.428	0.117	1.99318139457518E-250	2	NA CHST12	NK
0	1.77986277969682	0.188	0.002	0	2	NA KLRC1	NK
1.26774133454751E-299	1.76594932300973	0.186	0.012	1.78155689743961E-295	2	NA GZMK	NK
0	3.46961713400528	0.662	0.015	0	3	NA CD79A	B cells
1.03324918713589E-230	3.42278351352971	0.105	0.003	1.45202508268207E-226	3	NA IGLL5	B cells
0	2.70564741622645	0.468	0.007	0	3	NA MS4A1	B cells
0	2.42986348719175	0.586	0.141	0	3	NA CD83	B cells
2.69833099472783E-164	2.39666206809683	0.229	0.047	3.79196454689101E-160	3	NA MIR155HG	B cells
0	2.32590844120043	0.457	0.087	0	3	NA ID3	B cells
0	2.22823794523174	0.452	0.076	0	3	NA IRF8	B cells
0	2.09683429502013	0.992	0.654	0	3	NA CD74	B cells
0	2.06438698677772	0.583	0.149	0	3	NA HLA-DQA1	B cells
0	2.02758657929308	0.72	0.314	0	3	NA HERPUD1	B cells
9.07335285929278E-37	1.72825933915213	0.227	0.119	1.27507827731641E-32	3	NA MYC	B cells
3.31507643431989E-248	1.72589344619638	0.58	0.223	4.65867691314974E-244	3	NA HLA-DQB1	B cells
3.75873515131532E-239	1.5936429318106	0.416	0.113	5.28215050814342E-235	3	NA SYNGR2	B cells
1.62396192745837E-210	1.55007379904852	0.413	0.118	2.28215369665725E-206	3	NA PMAIP1	B cells
0	1.47246267305557	0.198	0.011	0	3	NA CD79B	B cells
0	1.45356432450283	0.232	0.012	0	3	NA BLNK	B cells
0	1.42499421651529	0.227	0.019	0	3	NA HVCN1	B cells
0	1.41522083643412	0.202	0.003	0	3	NA BANK1	B cells
9.05162278119206E-145	1.41269234369206	0.309	0.094	1.27202454944092E-140	3	NA REL	B cells
1.35207658967426E-157	1.39438641173354	0.619	0.349	1.90007323146924E-153	3	NA CD37	B cells
0	4.43169223840653	0.77	0.044	0	4	NA VMO1	ncMono
0	3.77654870358459	0.982	0.166	0	4	NA FCGR3A	ncMono
0	3.27899730006671	0.977	0.214	0	4	NA MS4A7	ncMono
0	2.84731460795712	0.749	0.054	0	4	NA MS4A4A	ncMono
0	2.58587015372884	0.935	0.195	0	4	NA CXCL16	ncMono

0	2.32700870435181	0.91	0.227	0	4	NA LST1	ncMono
0	2.28052333065588	0.845	0.206	0	4	NA AIF1	ncMono
0	2.14034471114249	0.741	0.195	0	4	NA FAM26F	ncMono
0	2.09884833907913	0.878	0.234	0	4	NA C3AR1	ncMono
0	2.02574510742975	0.724	0.201	0	4	NA CTSC	ncMono
0	1.9529874796882	0.77	0.19	0	4	NA CFD	ncMono
0	1.94852724538338	0.742	0.158	0	4	NA SERPINA1	ncMono
0	1.93467697754896	0.652	0.129	0	4	NA HN1	ncMono
0	1.92821669569134	0.857	0.271	0	4	NA ATP1B3	ncMono
0	1.89654072705279	0.794	0.206	0	4	NA ADA	ncMono
0	1.87763143085732	0.828	0.203	0	4	NA PILRA	ncMono
0	1.83969490133037	0.734	0.162	0	4	NA CD86	ncMono
0	1.83479481916198	0.721	0.206	0	4	NA WARS	ncMono
1.00067312071746E-218	1.8226147066651	0.631	0.226	1.40624593654425E-214	4	NA PLAC8	ncMono
0	1.79171959279724	0.619	0.148	0	4	NA GBP5	ncMono
0	4.70440158130872	0.817	0.048	0	5	NA PPBP	Mkc
0	3.51426674542155	0.649	0.016	0	5	NA PF4	Mkc
0	3.27239609370793	0.582	0.022	0	5	NA GNG11	Mkc
0	2.95520315873045	0.524	0.014	0	5	NA SDPR	Mkc
0	2.47861007911554	0.413	0.01	0	5	NA NRGN	Mkc
0	2.25235738629659	0.322	0.007	0	5	NA TUBB1	Mkc
0	2.18892444710569	0.312	0.006	0	5	NA ACRBP	Mkc
6.73228470528686E-50	2.0517876107265	0.471	0.148	9.46087969633962E-46	5	NA NCOA4	Mkc
1.7152428290242E-240	2.02345139595345	0.317	0.013	2.41043074762771E-236	5	NA CLU	Mkc
0	1.79852426017852	0.245	0.004	0	5	NA TREML1	Mkc
1.4792741335291E-69	1.7951949636626	0.332	0.054	2.07882393984845E-65	5	NA MAP3K7CL	Mkc
2.03973737855723E-65	1.75179445840566	0.327	0.054	2.86644293808647E-61	5	NA TUBA4A	Mkc
3.54012835515811E-27	1.73450318082976	0.505	0.244	4.97494237750369E-23	5	NA HIST1H2AC	Mkc
0	1.71886940681554	0.25	0.004	0	5	NA SPARC	Mkc
4.51979185876058E-121	1.7147130846464	0.236	0.016	6.35166349911625E-117	5	NA MMD	Mkc
4.49287463248236E-44	1.53388003038121	0.212	0.034	6.31383672102746E-40	5	NA TSC22D1	Mkc
1.11364874402247E-76	1.53209895567223	0.26	0.03	1.56501057997477E-72	5	NA NGFRAP1	Mkc
4.88246775936548E-165	1.51336460676319	0.212	0.008	6.86133194223631E-161	5	NA RP11-367G6.3	Mkc
1.016504260991E-154	1.45298324988877	0.149	0.004	1.42849343797066E-150	5	NA PTCRA	Mkc

5.07029260120443E-111	1.43154375809692	0.183	0.01	7.12528219247259E-107	5	NA RGS18	Mkc
0	4.20584195864311	0.766	0.013	0	6	NA TSPAN13	DC
1.92537369074333E-52	3.87894259514168	0.927	0.461	2.7057276476016E-48	6	NA TXN	DC
0	3.52333900677176	0.263	0.003	0	6	NA PTGDS	DC
2.7922696898516E-178	3.44547236987142	0.759	0.082	3.92397659514846E-174	6	NA GZMB	DC
7.99205225584831E-69	2.97530532920601	0.905	0.439	1.12312310351436E-64	6	NA SEC61B	DC
1.15746709346852E-250	2.84687491272925	0.613	0.036	1.62658850645131E-246	6	NA ITM2C	DC
0	2.78281073973832	0.599	0.011	0	6	NA IGJ	DC
2.89915621114521E-52	2.44549318019449	0.825	0.35	4.07418422352236E-48	6	NA HERPUD1	DC
2.06137584421662E-260	2.43855127855505	0.401	0.014	2.89685147387762E-256	6	NA SERPINF1	DC
7.46373184235897E-60	2.40545678327637	0.526	0.109	1.04887823580671E-55	6	NA IRF8	DC
2.8173954101335E-158	2.39744399169725	0.606	0.059	3.95928576986061E-154	6	NA P2RY6	DC
1.33140031944194E-74	2.36418865743673	0.212	0.014	1.87101686891175E-70	6	NA TCL1A	DC
6.12746255648096E-78	2.32688527809484	0.401	0.049	8.61092313062269E-74	6	NA TNFRSF4	DC
8.04857638694693E-69	2.22869173569338	0.387	0.05	1.13106643965765E-64	6	NA CLIC3	DC
4.4056447909649E-215	2.06164733568569	0.577	0.038	6.19125262474298E-211	6	NA TCF4	DC
0	1.92217304099831	0.343	0.001	0	6	NA DNASE1L3	DC
7.87769649692808E-47	1.87020239958565	0.971	0.685	1.1070526887133E-42	6	NA CD74	DC
0	1.84667702122302	0.161	0	0	6	NA CCL19	DC
8.75276342126986E-96	1.84657581720707	0.504	0.061	1.23002584359105E-91	6	NA STMN1	DC
1.28894733971696E-48	1.84440250958358	0.774	0.285	1.81135769650424E-44	6	NA COX5A	DC
8.71273238445068E-128	10.6362437334635	1	0.102	1.22440028198685E-123	7	NA HBB	Erythroid
0	9.61917121791266	1	0.022	0	7	NA HBA2	Erythroid
0	8.80772301773642	1	0.013	0	7	NA HBA1	Erythroid
0	3.91388239132043	0.836	0.001	0	7	NA ALAS2	Erythroid
2.3333606385259E-189	3.0573629902733	0.764	0.034	3.27907170532045E-185	7	NA SNCA	Erythroid
0	2.99345791057319	0.455	0.001	0	7	NA HBD	Erythroid
0	2.01457116882182	0.473	0	0	7	NA HBM	Erythroid
0	1.99067159931379	0.436	0	0	7	NA CA1	Erythroid
5.10388311727438E-58	1.79207736607409	0.727	0.09	7.17248694470569E-54	7	NA SLC25A37	Erythroid
1.14772276991851E-50	1.77588150425461	0.709	0.098	1.61289480856649E-46	7	NA SLC25A39	Erythroid
0	1.76368043892094	0.364	0	0	7	NA AHSP	Erythroid
7.16323023434923E-37	1.45022326425597	0.582	0.084	1.0066487448331E-32	7	NA GYPC	Erythroid
1.02697377343237E-35	1.39463609900941	0.418	0.049	1.44320624380451E-31	7	NA BLVRB	Erythroid
7.43583386037184E-97	1.26102903684756	0.418	0.019	1.04495773239805E-92	7	NA STRADB	Erythroid

1.65609310975084E-111	1.22866050301599	0.455	0.019	2.32730764713286E-107	7	NA DCAF12	Erythroid
3.67892536158109E-137	1.20937760592837	0.4	0.012	5.16999381062991E-133	7	NA FAM210B	Erythroid
5.94391243269617E-15	1.16195185589426	0.527	0.146	8.35298014166793E-11	7	NA ADIPOR1	Erythroid
1.46830472546308E-272	1.05995409566219	0.236	0.002	2.06340863069327E-268	7	NA KRT1	Erythroid
5.64151465888796E-13	0.98226815875402	0.509	0.15	7.92802055013526E-09	7	NA BNIP3L	Erythroid
8.78930516521038E-12	0.897700494764219	0.545	0.171	1.23516105486701E-07	7	NA FBXO7	Erythroid

Table 1. Top 20 ranked marker genes based on average log fold-change for each cell type, sorted by their average log-fold change within each respective cell type cluster.

p_val	avg_log2FC	pct.1	pct.2	p_val_adj	cluster	gene
0	2.7153139128886	0.396	0.117	0	IMMUNE_CTRL	NA IL8
0	2.03649211378656	0.34	0.089	0	IMMUNE_CTRL	NA CD14
0	1.38872770208823	0.226	0.01	0	IMMUNE_CTRL	NA CLEC5A
3.21738488731806E-287	1.30623004570458	0.189	0.011	4.52139098214807E-283	IMMUNE_CTRL	NA VCAN
2.28867526034694E-237	1.25421123808995	0.251	0.058	3.21627534336556E-233	IMMUNE_CTRL	NA IER3
9.00566389123385E-236	1.33069645293508	0.473	0.248	1.26556594663509E-231	IMMUNE_CTRL	NA MARCKSL1
9.52342790227185E-227	1.88308570051995	0.181	0.021	1.33832732310626E-222	IMMUNE_CTRL	NA IL1B
1.76968972076197E-216	1.32490914224077	0.196	0.033	2.4869449645868E-212	IMMUNE_CTRL	NA CD9
1.69782653109161E-158	2.06551253989952	0.33	0.162	2.38595562414304E-154	IMMUNE_CTRL	NA S100A8
8.92494608450822E-86	1.30419471647855	0.159	0.058	1.25422267325594E-81	IMMUNE_CTRL	NA CXCL3
0	5.62102118234777	0.589	0.036	0	IMMUNE_STIM	NA CXCL10
0	5.59500500891929	0.376	0.006	0	IMMUNE_STIM	NA CCL8
0	5.44182137489336	0.994	0.286	0	IMMUNE_STIM	NA ISG15
0	4.40748187837566	0.897	0.03	0	IMMUNE_STIM	NA IFIT1
0	4.40541047458625	0.917	0.058	0	IMMUNE_STIM	NA IFIT3
0	3.94896734569654	0.708	0.028	0	IMMUNE_STIM	NA RSAD2
0	3.90206430837784	0.766	0.037	0	IMMUNE_STIM	NA IFIT2
0	3.84651733853949	0.397	0.006	0	IMMUNE_STIM	NA CXCL11
0	3.80743435813626	0.43	0.127	0	IMMUNE_STIM	NA APOBEC3A
4.93692583012709E-138	3.86392683628789	0.132	0.019	6.9378618690776E-134	IMMUNE_STIM	NA APOBEC3B

Table 2. Top 10 ranked marker genes based on average log fold-change for control v stimulated cMono cells.