scRNAseq analysis of Gstt2-KO mouse implanted with bladder tumors

Kane Toh*

2023-08-09

Contents

1 Methods

Differences in cell type abundances and imputed cell cycle phases between the KO and WT conditions were carried out with Wilcoxon tests as well as a generalised linear model-based method called Milo (Dann et al., 2021). The adjusted p value and spatial false discovery rate thresholds for both tests were set to 0.1. In brief, Milo is a cluster-free method that embeds cells into a k-nearest neighbor graph, models cell states as overlapping neighborhoods on the knn graph, counts cell numbers within each neighborhood per sample and then carries out differential abundance testing between the experimental conditions. In Milo, neighborhoods were defined as "Mixed" if thecell-type fraction or cell-cycle fraction of that neighborhood is less than 0.7.

Subclustering of the granulocyte/neutrophils cluster was run with the Seurat `FindSubCluster` function, using the `RNA.snn` graph. A total of 3 subclusters was identified at a resolution of 0.5.

To investigate the differences in gene expression patterns between the GSTT2-KO and WT samples, we used the *distinct* approach for differential analysis (Tiberi et al., 2022: preprint). Distinct takes a full distribution approach to differential analysis, and thus can detect differences in gene expression such as differential variability and differential modality, on top of shifts in mean abundance between the two sample conditions. With the exception ofthe `P_4` parameter which we set to 20000 following the author's recommendation, all other parameters were set to the default, along with an adjusted p-value of 0.01.

*Genomics and Data Analytics Core (GeDaC), Cancer Science Institute of Singapore, National University of Singapore; kanetoh@nus.edu.sg

2 Results

2.1 Sample swap between KO_BL and WT_BR

When reanalysing the data for differences between KO an, we identified a potential sample swap in the dataset between specifically between the KO-BL and WT-BR samples. We double checked the sample-well loading table that was input to the Parse Biosciences pipeline from the log file, and this was consistent with what was provided by the experimental investigators (screenshots titled `split_pipe_log.png` and `sample_loading_table.png` in the results folder). This is evident in the UMAP plots (Figure 1 where cells in the proliferative CAFs and tumorigenic urothelial cells clusters were absent/present in the KO_BL/WT_BR samples, which is the opposite of what is expected for the KO and WT samples based on comparison against the other samples. Therefore, we swapped the labels for both samples for subsequent analyses.

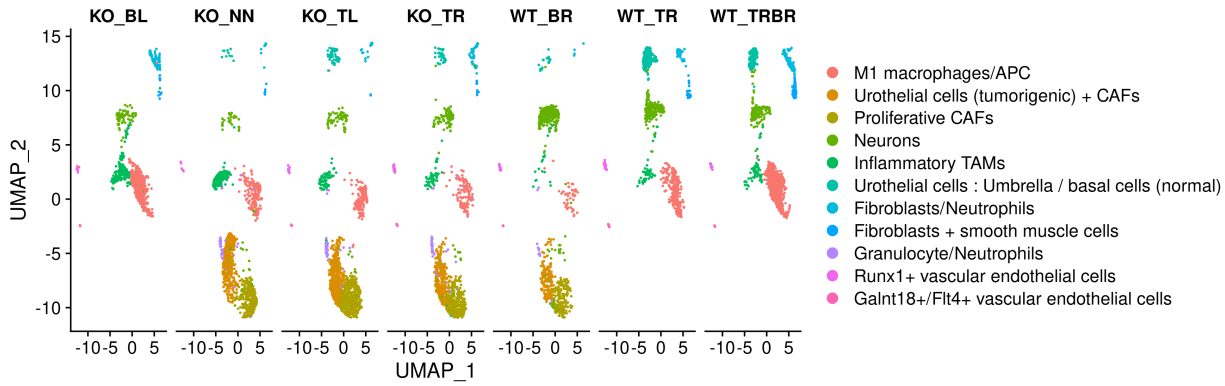

Figure 1: UMAP of sample labels showing possible sample label swap between KOBL and WTBR samples

2.2 Differential abundance analysis

Testing for differences in cell type abundance between the WT and KO conditions showed that the fibroblasts and M1 macrophages are more abundant in WT condition, whereas the tumorigenic urothelial cells + CAFs, proliferative CAFs and granulocyte/neutrophil clusters are enriched in the KO condition (Figure 2 A-C). In fact, we find that these 3 clusters contain only cells from the KO samples (`DA_table.txt`). At a significance threshold of 0.1, these results are consistent across both the Wilcoxon test and the GLM method employed by Milo.

We repeated the same analyses for differences in cell cycle phases across the WT and KO conditions. Predictably, as the KO condition is composed of more tumorigenic cells, the KO condition is enriched for cells in the G2/M phase whereas the WT condition has more cells in the G1 phase (Figure 3 A-B). Cell cycle imputation was carried out in Seurat with the `cell_cycle_markers` function (For more details, see results_summary_091222.pdf).

2.3 Differential state analysis

Having performed the sample swap, we repeated our differential state analysis with *distinct*. The list of genes that are upregulated in the KO/WT conditions are in the files titled `distinct_log2fc_topresults_log2fcsorted_KOup.csv` and `distinct_log2fc_topresults_log2fcsorted_WTup.csv`respectively.

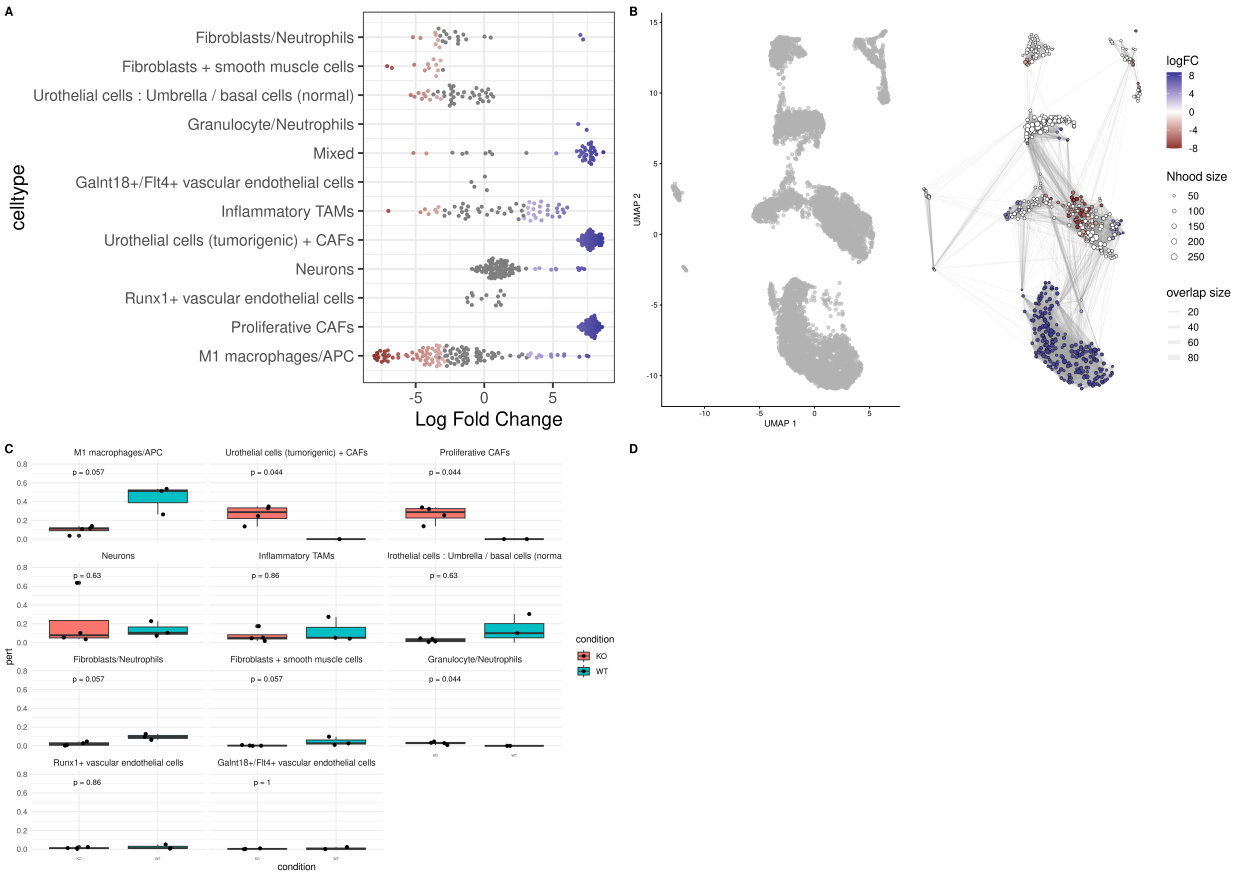

Figure 2: Cell type differential abundance between KO and WT conditions

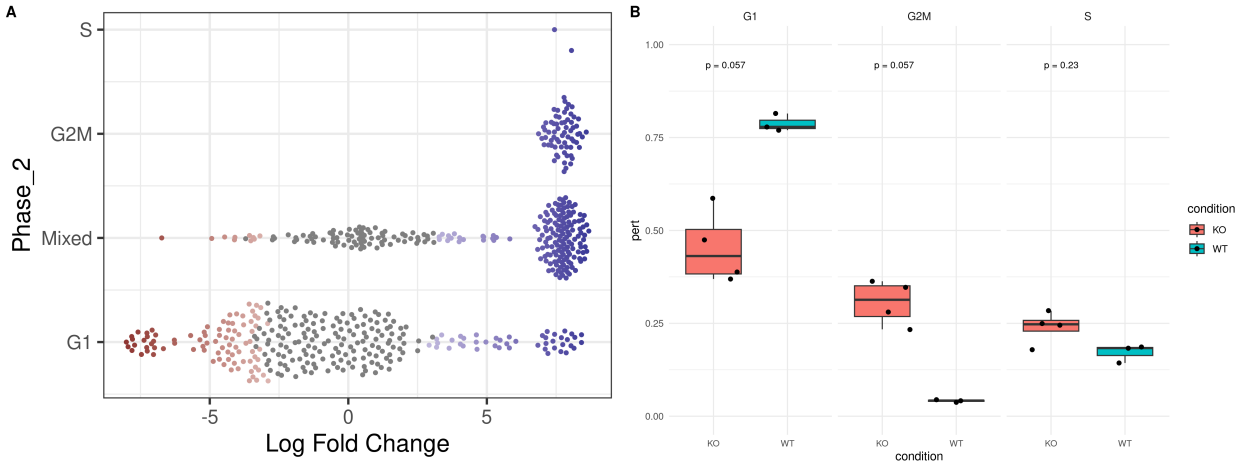

Figure 3: Differential abundance of cell cycle phases between KO and WT conditions

2.4 Subclustering of the granulocyte/neutrophils (g/n) cluster

The g/n cluster consists of around 150 cells. To obtain a finer resolution of the g/n cluster, we subclustered the cells which resulted in 3 additional clusters (Figure 4A). We then obtained the positive markers for each of these 3 clusters, and plotted all the markers with an adjusted p value < 0.05 in a heatmap (Figure 4B). Cluster 0 is enriched in the pro-inflammatory cytokine Il1b and receptor Il1r2; cluster 1 is enriched in the long non-coding RNA gene Airn, Tns3 and Oxr1 genes; cluster 2 appears to have more macrophage markers such as Msr1, as well as Lgmn and Ctss for antigen-processing. Indeed, when we reran the cell type abundance analyses and included the g/n subclusters, only g/n cluster 0 was enriched in the KO condition (Figure 4C) whereas the other two clusters were mixed. It is also interesting that cluster 2 is grouped together with the other neutrophils instead of the M1 macrophage cluster, implying that they may have a distinct immune function from the M1 macrophages that is present only in the KO condition. The full list of positive marker genes for these 3 subclusters can be found in the `gn_markers.csv` file.

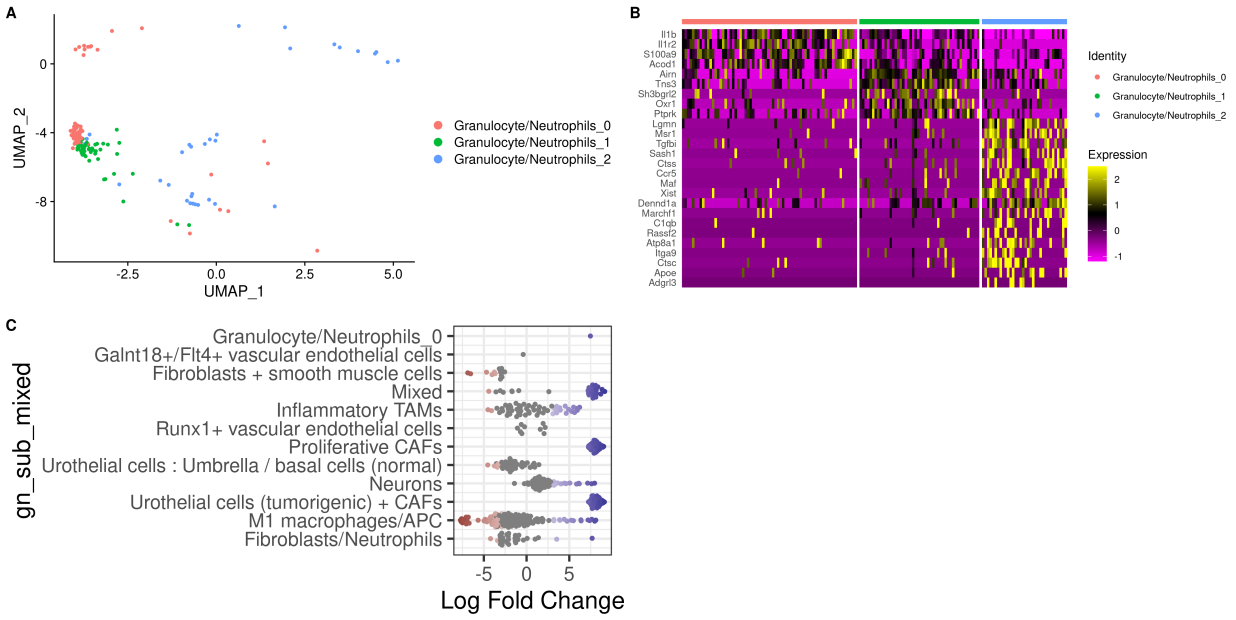

Figure 4: Subclustering of the Granulocyte/Neutrophil cluster