Pre-processing of sequencing data

Alignment to the reference genome, feature counting and cell calling followed the 10X Genomics CellRanger (v5.0.0) pipeline, using the default mm10 genome supplied by 10X Genomics (refdata-gex-mm10-2020-A). From the output, the filtered matrices were used for downstream analyses. Pre-processing was performed on the University of Edinburgh's compute cluster Eddie.

scRNA-seq analysis

The analysis was performed on a Windows system 10 x64 (build 19041), with R version 4.1.1. Full details to replicate the analysis pipelines described below can be found in code scripts available on GitHub (https://github.com/Anna-Williams/Veronique-Firemice).

Initial quality control and clustering

Cells were filtered using dataset-specific parameters on the basis of genes and UMIs per cell, the ratio between these two parameters and percentage of mitochondrial gene reads per cell. The thresholds were computed with *isOutlier()* function from scuttle(v1.2.1) on batch 3-5, as batch 6 had outlier values. Only genes that were detected in at least two cells were kept. With scran (v1.20.1) the data was normalised by deconvolution and the top 15% highly variable genes were selected. Following principal component analysis (PCA), 25 principal components (PCs) were kept for downstream analysis (cut-off selected by examination of an *Elbowplot*). Non-linear dimensional dimension representation (UMAP and t-SNE) and gene expression variance explained by batch (computed with scater (v1.20.1), revealed the need of batch correction. Batch correction was performed by mutual nearest neighbours with *fastMNN(),* batchelor (v1.8.0). Finally, graph-based clustering approach was used to cluster the cells using *clusterCells()* function from scran, with k = 60.

Analyse Oligodendrocytes

The clusters with highest expression of oligodendrocyte markers (Plp1, Mog, Mag and Mbp) and that did not express other cell type markers (astrocyte, OPC or microglia markers) were subset to analyse separately. Cell and gene quality control were further adjusted setting a stricter minimum UMI count threshold (5000 UMIs) and maximum percentage of mitochondrial gene reads per cell (10 %). A small cluster of cells with lower quality cells was also excluded from the analysis. Ultimately, we included a total of 19506 genes and 13583 cells. The normalisation, feature selection, dimensional reduction and batch correction were repeated with the subset dataset as described above. Clustering was performed with k = 100, and then merged into four clusters; cluster 1 being KO specific. Differential gene expression between cluster 1 and the mean expression of all other cells was performed with *FindMarkers*() from Seurat (v4.1.0).(supplementary file)

**Bullet points:**

1. Alignement, filtering, umi counting and cell calling with CellRanger v5.0.0, aligned to genome refdata-gex-mm10-2020-A
2. Used isOutlier() from scuttle package (v1.2.1) for cell QC on umi counts, detected genes and mithochondrial percentage as well as ratio between umi counts and detected genes. The thresholds were computed on batch 3-5, as batch 6 had outlier values.

Thresholds:

* + 14.04 % mitochondrial genes
  + detected genes 453 7087
  + the sum of umi counts: 470 to 1.8624^{4} umi counts
  + ratio detected genes/ umi counts: 0.04565117 0.72692408

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | umi\_high | genes\_high | umi\_low | genes\_low | mt\_pct |
| Cells filtered | 2294 | 74 | 0 | 3562 | 13179 |
| Threshold | 18623.92 | 7087.371 | 470.3209 | 453.2256 | 14.04298 |

1. The cell QC was performed filtering out any gene not detected in at least two cells
   * The object had 32285 genes , 63430 cells before QC and 23155 genes , 45936 cells after filtering
2. Normalisation by deconvolution with scran (v1.20.1)
3. Feature selection: used modelGeneVar() from scran (v1.20.1), with density.weights=FALSE, to model the variance of the log-expression profiles. Then selected top 15% highly variable genes (2107 HVGs)
4. PCA linear dimensional reduction, keeping 25PCs for downstream analysis
5. Batch correction by mutual nearest neighbours, with fastMNN() from Batchelor (v1.8.0)
6. Graph based clustering with clusterCells() from scran (v1.20.1) with k= 60 - To choose an optimal clustering resolution, a clustering tree was generated with the R package clustree which depicts the movement of cells across clusters as resolution is increased
7. Subset for clusters with highest expression of Oligodendrocyte markers such as Plp1, Mog, Mag and Mbp; and later also subseted for no expression of markers from other celltypes (such as astrocyte or OPC markers)
8. Redo all steps from 2-8 with the oligos only.

2-3b) thresholds kept the same except minum umi 5000 umi counts and mt genes (10%)

before filtering: 23155 genes 16358 cells

after filtering : 19506 genes 13583 cells

4b) same

5b) 1374 HVG

6-7b) same

8b) k=100, merging clusters 6-2 and 4-5 to obtain a total of 4 clusters

1. Differential expression between cluster 1 (KO specific) and the mean expression of all other cells with FindMarkers() from Seurat (v4.1.0).(supplementary file)

R version 4.1.1 (2021-08-10)

Running under: Windows 10 x64 (build 19041)

Others: <https://www.nature.com/articles/s41467-022-29824-1#Sec9> 351 words, - this is the more bio one, should probably aim for that.

<https://www.nature.com/articles/s41593-022-01022-8#Sec11> 1140 words,

<https://www.nature.com/articles/s41593-019-0491-3#Sec12> 911 words.

<https://www.nature.com/articles/s41467-022-29747-x#Sec8> 804 words

References

<https://arxiv.org/abs/1802.03426> UMAP

doi: [10.1186/s13059-016-0947-7](https://dx.doi.org/10.1186%2Fs13059-016-0947-7) scran

doi: [10.1093/bioinformatics/btw777](https://doi.org/10.1093/bioinformatics/btw777). scater