SCE assembly for the 2016-17 cohort

EJC

25/06/2020, based on code from May 2019

Setup

The Babraham compute cluster does not contain a global tex installation, so a local tex is added to $PATH to allow knitting to pdf.

```
Sys.setenv(PATH=paste(Sys.getenv("PATH"),
                      "/bi/home/carre/texlive/2017/bin/x86_64-linux/",sep=":"))
```

Load QC-pass SCE

```
load(file = "data/SCE_QC_pass.RData")
```

Normalise by deconvolution, PCA, UMAP, find PB cluster and remove

For convenience PCA + UMAP on the final SCE is re-run (to avoid needing to re-run it for each figure.Rmd).

```
library(SingleCellExperiment)
library(scran)
library(scater)

# Normalisation by deconvolution

set.seed(100)
clust <- quickCluster(sce)
# table(clust)
deconv.sf <- calculateSumFactors(sce, cluster = clust)
# summary(deconv.sf)

sce <- computeSumFactors(sce, cluster = clust, min.mean = 0.1)
sce <- logNormCounts(sce)

# Select 25% of genes with largest variance
dec <- modelGeneVar(sce)
hvg <- getTopHVGs(dec, prop = 0.25)

set.seed(10000)
sce <- runPCA(sce, ncomponents = 50, subset_row = hvg)
```

```
set.seed(1e+05)
sce <- runUMAP(sce, dimred = "PCA", external_neighbors = TRUE)

# Clustering.
set.seed(1e+06)
g <- buildSNNGraph(sce, use.dimred = "PCA")
set.seed(1e+07)
sce$clusters <- factor(igraph::cluster_louvain(g)$membership)


# So there is always 1 cluster whose row has 9 zeros in Let's
# select that programmatically.
cluster.selector <- table(Cluster = sce$clusters, Batch = sce$library)
cluster.to.discard <- c(1:nlevels(sce$clusters))[apply(cluster.selector,
    1, function(x) sum(x == 0)) == 9]

# Discard the PB cluster, then re-do analysis: The PB cluster
# is so different, that % hvg or # PCs does not influence
# their clustering apart
sce <- sce[, !sce$clusters == cluster.to.discard]

set.seed(100)
clust <- quickCluster(sce)
# table(clust)
deconv.sf <- calculateSumFactors(sce, cluster = clust)
# summary(deconv.sf)

sce <- computeSumFactors(sce, cluster = clust, min.mean = 1)
sce <- logNormCounts(sce)

# Select 10% of genes with largest variance
dec <- modelGeneVar(sce)
hvg <- getTopHVGs(dec, prop = 0.1)

set.seed(10000)
sce <- runPCA(sce, ncomponents = 40, subset_row = hvg)
set.seed(1e+05)
sce <- runUMAP(sce, dimred = "PCA", external_neighbors = TRUE)

# Clustering.
set.seed(1e+06)
g <- buildSNNGraph(sce, use.dimred = "PCA")
set.seed(1e+07)
sce$clusters <- factor(igraph::cluster_louvain(g)$membership)
```

18 cells within the plasmablast cluster are removed - see the supplementary figure Rmd/pdf to see feature plots supporting their identification as plasmablasts.

Plot SCE

```
library(ggpubr)
```

```
## Loading required package: magrittr
##
## Attaching package: 'magrittr'
## The following object is masked from 'package:AnnotationFilter':
##
##     not
```

```
cowplot::plot_grid(
plotUMAP(sce, colour_by="clusters") +
     theme_pubr(legend = "right") +
     guides(fill = guide_legend(title = "Louvain\ncluster",
                                title.position = "top")) +
     theme(legend.key.size = unit(0, 'lines'),
           legend.margin = margin(0,0,0,0, 'lines'),
           aspect.ratio = 1) + rotate_x_text(),

cowplot::plot_grid(
  # UMAP both day 0 and day 42
  # Colours from index flow IgD
  plotUMAP(sce, colour_by="IgD.BUV737") +
    viridis::scale_fill_viridis(option="inferno") +
    labs(fill = "Surface\nIgD") +
    theme_pubr() +
    theme(legend.position = "right",
          axis.ticks = element_blank(),
          axis.text = element_blank(),
          legend.key.width = unit(6,"points"),
          legend.key.height = unit(12,"points"),
          legend.margin = margin(0,0,0,0, 'lines'),
          aspect.ratio = 1),
  plotUMAP(sce, colour_by="CD27.BV711") +
    viridis::scale_fill_viridis(option="inferno") +
     labs(fill = "Surface\nCD27") +
     theme_pubr() +
     theme(legend.position = "right",
           axis.ticks = element_blank(),
           axis.text = element_blank(),
           legend.key.width = unit(6,"points"),
           legend.key.height = unit(12,"points"),
           legend.margin = margin(0,0,0,0, 'lines'),
           aspect.ratio = 1),

  # UMAP both day 0 and day 42
  # Colours from index flow CD21
  plotUMAP(sce, colour_by="CD21.PE.cy7") +
    viridis::scale_fill_viridis(option="inferno") +
    labs(fill = "Surface\nCD21") +
    theme_pubr() +
    theme(legend.position = "right",
          axis.ticks = element_blank(),
          axis.text = element_blank(),
          legend.key.width = unit(6,"points"),
          legend.key.height = unit(12,"points"),
```

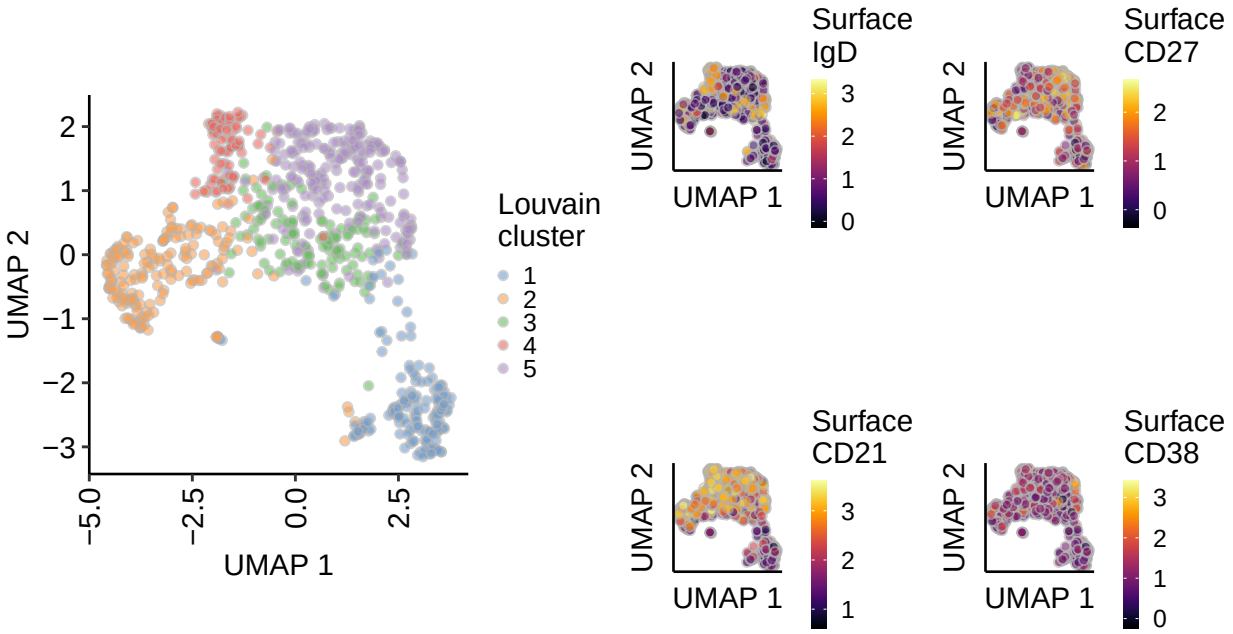
```
          legend.margin = margin(0,0,0,0, 'lines'),
          aspect.ratio = 1),

  # UMAP both day 0 and day 42
  # Colours from index flow CD28
  plotUMAP(sce, colour_by="CD38.BV421") +
    viridis::scale_fill_viridis(option="inferno") +
    labs(fill = "Surface\nCD38") +
    theme_pubr() +
    theme(legend.position = "right",
          axis.ticks = element_blank(),
          axis.text = element_blank(),
          legend.key.width = unit(6,"points"),
          legend.key.height = unit(12,"points"),
          legend.margin = margin(0,0,0,0, 'lines'),
          aspect.ratio = 1),

  ncol = 2),
ncol = 2)
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```

```
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
## Scale for 'fill' is already present. Adding another scale for 'fill', which
## will replace the existing scale.
```



Save final SCE: NTC removed, QC passing cells, plasmablasts removed, with dimensionalty reduction pre-calculated

```
save(sce, file = "data/SCE_QC_pass_finalised.RData")
```

SessionInfo

```
sessionInfo()
```

```
## R version 3.6.1 (2019-07-05)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: CentOS Linux 7 (Core)
##
## Matrix products: default
## BLAS:   /bi/apps/R/3.6.1/lib64/R/lib/libRblas.so
## LAPACK: /bi/apps/R/3.6.1/lib64/R/lib/libRlapack.so
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C         LC_TIME=C
##  [4] LC_COLLATE=C         LC_MONETARY=C        LC_MESSAGES=C
##  [7] LC_PAPER=C           LC_NAME=C            LC_ADDRESS=C
## [10] LC_TELEPHONE=C       LC_MEASUREMENT=C     LC_IDENTIFICATION=C
##
```

```
## attached base packages:
## [1] parallel  stats4    stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] ggpubr_0.2.4                magrittr_1.5
##  [3] rtracklayer_1.46.0          ensembldb_2.10.2
##  [5] AnnotationFilter_1.10.0     GenomicFeatures_1.38.0
##  [7] AnnotationDbi_1.48.0        AnnotationHub_2.18.0
##  [9] BiocFileCache_1.10.2        dbplyr_1.4.2
## [11] scran_1.14.5                scater_1.14.5
## [13] ggplot2_3.3.2               SingleCellExperiment_1.8.0
## [15] SummarizedExperiment_1.16.0 DelayedArray_0.12.0
## [17] BiocParallel_1.20.0         matrixStats_0.55.0
## [19] Biobase_2.46.0              GenomicRanges_1.38.0
## [21] GenomeInfoDb_1.22.0         IRanges_2.20.1
## [23] S4Vectors_0.24.1            BiocGenerics_0.32.0
## [25] dplyr_1.0.2
##
## loaded via a namespace (and not attached):
##  [1] ggbeeswarm_0.6.0              colorspace_1.4-1
##  [3] ggsignif_0.6.0                ellipsis_0.3.0
##  [5] XVector_0.26.0                BiocNeighbors_1.4.1
##  [7] farver_2.0.1                  bit64_0.9-7
##  [9] RSpectra_0.16-0               interactiveDisplayBase_1.24.0
## [11] knitr_1.26                    Rsamtools_2.2.1
## [13] uwot_0.1.5                    shiny_1.4.0
## [15] BiocManager_1.30.10           compiler_3.6.1
## [17] httr_1.4.1                    dqrng_0.2.1
## [19] assertthat_0.2.1              Matrix_1.2-17
## [21] fastmap_1.0.1                 lazyeval_0.2.2
## [23] limma_3.42.0                  later_1.0.0
## [25] BiocSingular_1.2.0            formatR_1.7
## [27] htmltools_0.4.0               prettyunits_1.0.2
## [29] tools_3.6.1                   rsvd_1.0.2
## [31] igraph_1.2.4.2                gtable_0.3.0
## [33] glue_1.4.2                    GenomeInfoDbData_1.2.2
## [35] rappdirs_0.3.1                tinytex_0.18
## [37] Rcpp_1.0.3                    vctrs_0.3.6
## [39] Biostrings_2.54.0             DelayedMatrixStats_1.8.0
## [41] xfun_0.11                     stringr_1.4.0
## [43] mime_0.7                      lifecycle_0.2.0
## [45] irlba_2.3.3                   statmod_1.4.32
## [47] XML_3.98-1.20                 edgeR_3.28.0
## [49] zlibbioc_1.32.0               scales_1.1.0
## [51] BSgenome_1.54.0               hms_0.5.2
## [53] promises_1.1.0                ProtGenerics_1.18.0
## [55] yaml_2.2.0                    curl_4.3
## [57] memoise_1.1.0                 gridExtra_2.3
## [59] biomaRt_2.42.0                stringi_1.4.3
## [61] RSQLite_2.1.4                 BiocVersion_3.10.1
## [63] rlang_0.4.10                  pkgconfig_2.0.3
## [65] bitops_1.0-6                  evaluate_0.14
## [67] lattice_0.20-38               purrr_0.3.3
```

```
## [69] labeling_0.3                GenomicAlignments_1.22.1
## [71] cowplot_1.0.0               bit_1.1-14
## [73] tidyselect_1.1.0            R6_2.4.1
## [75] generics_0.0.2              DBI_1.1.0
## [77] pillar_1.4.7                withr_2.1.2
## [79] RCurl_1.95-4.12             tibble_3.0.4
## [81] crayon_1.3.4                rmarkdown_2.0
## [83] viridis_0.5.1               progress_1.2.2
## [85] locfit_1.5-9.1              grid_3.6.1
## [87] blob_1.2.0                  digest_0.6.23
## [89] xtable_1.8-4                httpuv_1.5.2
## [91] RcppParallel_4.4.4          openssl_1.4.1
## [93] munsell_0.5.0               beeswarm_0.2.3
## [95] viridisLite_0.3.0           vipor_0.4.5
## [97] askpass_1.1
```