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)  # PB exclusion do not adjust
```

```
set.seed(10000)
sce <- runPCA(sce, ncomponents = 50, subset_row = hvg)  # PB exclusion do not adjust
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 Remove the PB cluster from a fresh
# SCE without any dim reductions (ensures dim reduction is
# calculated on the orig dataset)
sce.old <- sce
load(file = "data/SCE_QC_pass.RData")  # reload the SCE without any dim reduc
sce <- sce[, !sce.old$clusters == cluster.to.discard]
rm(sce.old)

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)  # NB scran docs recommends min.mean 1 f
sce <- logNormCounts(sce)

# Select 25% of genes with largest variance Select 25 PCs
dec <- modelGeneVar(sce, min.mean = 0.1)
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$clusters <- factor(c(5, 4, 2, 3, 1)[igraph::cluster_louvain(g)$membership])  # change cluster names
```

15 cells within the plasmablast cluster are removed - see the supplementary figure Rmd/pdf to see feature

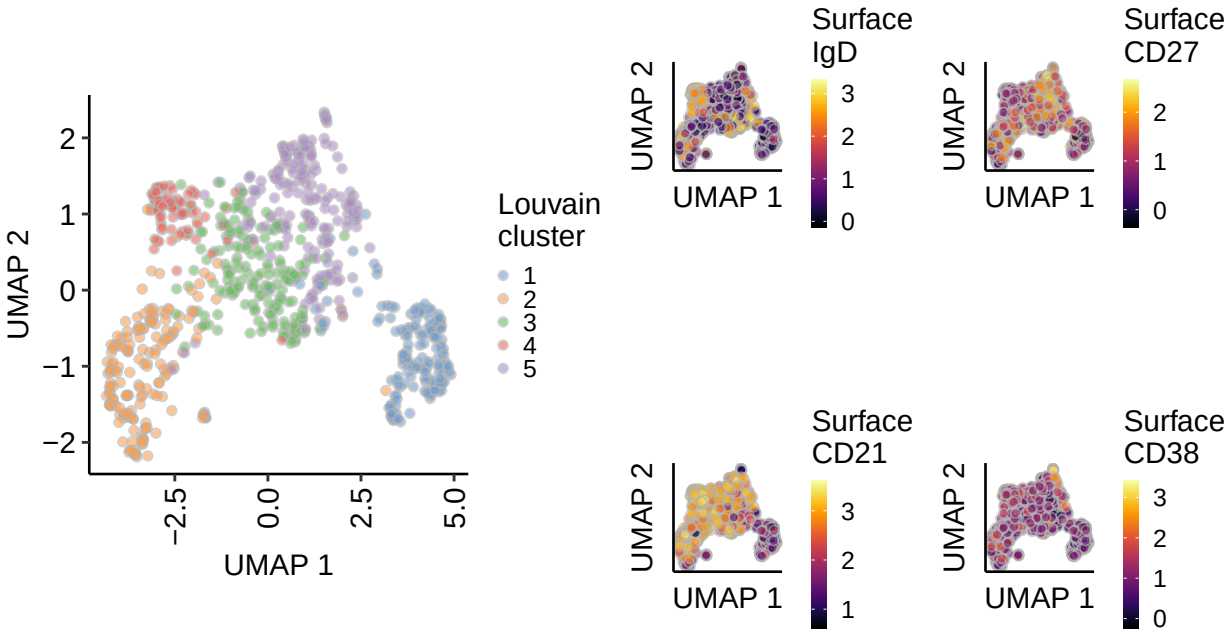
plots supporting their identification as plasmablasts.

Plot SCE

```
library(ggpubr)

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_GB.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_GB.UTF-8        LC_COLLATE=en_GB.UTF-8
##  [5] LC_MONETARY=en_GB.UTF-8    LC_MESSAGES=en_GB.UTF-8
```

```
##  [7] LC_PAPER=en_GB.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_GB.UTF-8 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] scater_1.14.5               ggplot2_3.3.2
##  [5] scran_1.14.5                SingleCellExperiment_1.8.0
##  [7] SummarizedExperiment_1.16.0 DelayedArray_0.12.0
##  [9] BiocParallel_1.20.0         matrixStats_0.55.0
## [11] Biobase_2.46.0              GenomicRanges_1.38.0
## [13] GenomeInfoDb_1.22.0         IRanges_2.20.1
## [15] S4Vectors_0.24.1            BiocGenerics_0.32.0
##
## loaded via a namespace (and not attached):
##  [1] viridis_0.5.1           edgeR_3.28.0             BiocSingular_1.2.0
##  [4] viridisLite_0.3.0       DelayedMatrixStats_1.8.0 RcppParallel_4.4.4
##  [7] statmod_1.4.32          dqrng_0.2.1              GenomeInfoDbData_1.2.2
## [10] vipor_0.4.5             yaml_2.2.0               pillar_1.4.7
## [13] lattice_0.20-38         glue_1.4.2               limma_3.42.0
## [16] digest_0.6.23           XVector_0.26.0           ggsignif_0.6.0
## [19] colorspace_1.4-1        cowplot_1.0.0            htmltools_0.4.0
## [22] Matrix_1.2-17           pkgconfig_2.0.3          zlibbioc_1.32.0
## [25] purrr_0.3.3             scales_1.1.0             RSpectra_0.16-0
## [28] tibble_3.0.4            farver_2.0.1             generics_0.0.2
## [31] ellipsis_0.3.0          withr_2.1.2              crayon_1.3.4
## [34] evaluate_0.14           beeswarm_0.2.3           tools_3.6.1
## [37] formatR_1.7             lifecycle_0.2.0          stringr_1.4.0
## [40] munsell_0.5.0           locfit_1.5-9.1           irlba_2.3.3
## [43] compiler_3.6.1          rsvd_1.0.2               tinytex_0.18
## [46] rlang_0.4.10            grid_3.6.1               RCurl_1.95-4.12
## [49] BiocNeighbors_1.4.1     igraph_1.2.4.2           labeling_0.3
## [52] bitops_1.0-6            rmarkdown_2.0            gtable_0.3.0
## [55] R6_2.4.1                gridExtra_2.3            knitr_1.26
## [58] dplyr_1.0.2             uwot_0.1.5               stringi_1.4.3
## [61] ggbeeswarm_0.6.0        Rcpp_1.0.3               vctrs_0.3.6
## [64] tidyselect_1.1.0        xfun_0.11
```