Re: Booeshaghi & Pachter, BioRxiv, October 2020

Christoph Hafemeister

2021-03-17

Data is taken from https://github.com/pachterlab/BP_2020_2/tree/master/data/raw

Load raw matrix, round to integers, create Seurat object

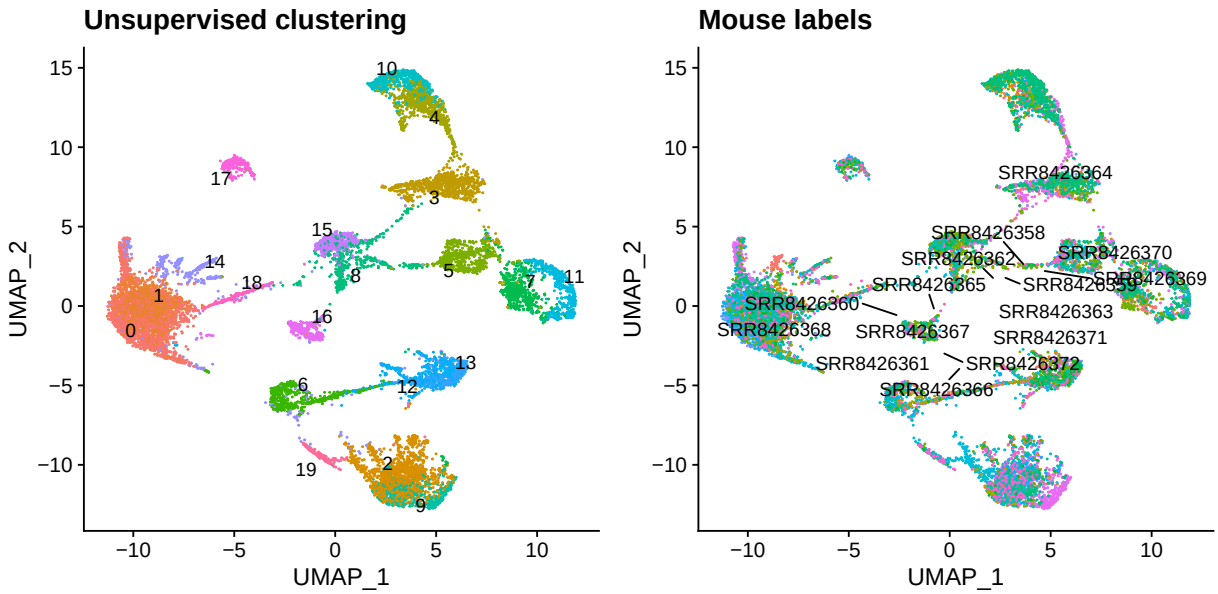
```
raw <- Read10X(data.dir = data_dir)
raw@x <- round(raw@x)
s <- CreateSeuratObject(counts = raw, names.delim = "-")
print(s)
#> An object of class Seurat
#> 19632 features across 12185 samples within 1 assay
#> Active assay: RNA (19632 features, 0 variable features)
```

Add age meta data

```
age_df <- data.frame(orig.ident = c("SRR8426358", "SRR8426359", "SRR8426362",
    "SRR8426363", "SRR8426364", "SRR8426369", "SRR8426370", "SRR8426360",
    "SRR8426361", "SRR8426365", "SRR8426366", "SRR8426367", "SRR8426368",
    "SRR8426371", "SRR8426372"), age = c(rep(24, 7), rep(3, 8)))
md <- left_join(s@meta.data, age_df)
#> Joining, by = "orig.ident"
s <- AddMetaData(s, metadata = md$age, col.name = "age")
```

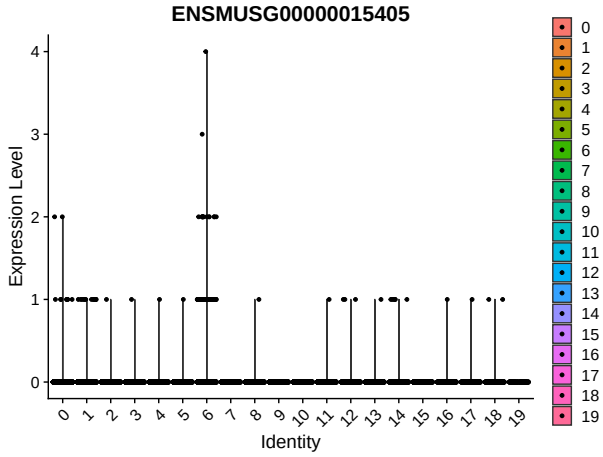
Run 'standard' unsupervised analysis

```
s <- SCTransform(s, verbose = FALSE, method = "qpoisson")
s <- RunPCA(s, verbose = FALSE)
s <- RunUMAP(s, dims = 1:10, verbose = FALSE)
s <- FindNeighbors(s, dims = 1:10, verbose = FALSE)
s <- FindClusters(s, verbose = FALSE)
p1 <- DimPlot(s, label = TRUE, repel = TRUE) + NoLegend() + ggtitle("Unsupervised clustering")
p2 <- DimPlot(s, label = TRUE, repel = TRUE, group.by = "orig.ident") +
    NoLegend() + ggtitle("Mouse labels")
show(p1 | p2)
```



Show Ace2 expression per cluster

```
ace2_gene = "ENSMUSG00000015405"
VlnPlot(s, features = ace2_gene, assay = "RNA")
```



Ace2 is very lowly expressed. Look at detection rate per cluster. We use the raw counts, as well as the sctransform corrected counts (to simulate identical sequencing depth for all cells).

```
data.frame(cluster = Idents(s), age = s$age, ace2_raw = s$RNA@counts[ace2_gene,
    ], ace2_corrected = s$SCT@counts[ace2_gene, ]) %>% group_by(cluster) %>%
    summarise(cells = n(), pct_age3 = round(mean(age == 3) * 100), pct_age24 = round(mean(age ==
        24) * 100), dr_raw = mean(ace2_raw > 0), dr_corrected = mean(ace2_corrected >
        0)) %>% arrange(-dr_raw)
```

cluster	cells	pct_age3	pct_age24	dr_raw	dr_corrected
6	562	74	26	0.1512456	0.0765125
14	448	61	39	0.0133929	0.0133929
18	163	66	34	0.0122699	0.0122699
1	1420	85	15	0.0098592	0.0098592
12	462	55	45	0.0064935	0.0064935
0	1564	72	28	0.0063939	0.0063939
17	257	63	37	0.0038911	0.0038911
16	309	45	55	0.0032362	0.0032362
13	453	39	61	0.0022075	0.0022075
11	464	38	62	0.0021552	0.0021552
8	525	49	51	0.0019048	0.0019048
5	603	56	44	0.0016584	0.0016584
4	691	27	73	0.0014472	0.0014472
3	818	15	85	0.0012225	0.0012225
2	1314	83	17	0.0007610	0.0007610
7	561	20	80	0.0000000	0.0000000
9	511	54	46	0.0000000	0.0000000
10	486	7	93	0.0000000	0.0000000
15	420	29	71	0.0000000	0.0000000
19	154	94	6	0.0000000	0.0000000

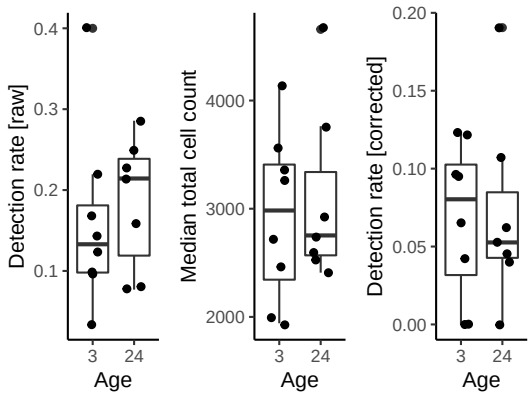
Cluster 6 seems to be the major ACE2 expressing cell type (ca. 15% detection frequency). Not sure about the other clusters with less than 1.5% ACE2 detection. 74% of the cells in cluster 6 come from age 3 samples.

Zoom in on cluster 6

Look at ACE2 detection rate per animal

```
df <- cbind(s@meta.data, data.frame(ace2_raw = s$RNA@counts[ace2_gene,
    ], ace2_corrected = s$SCT@counts[ace2_gene, ])) %>% filter(seurat_clusters ==
    "6") %>% group_by(orig.ident, age) %>% summarise(cells = n(), median_depth = median(nCount_RNA),
    dr_raw = mean(ace2_raw > 0), dr_corrected = mean(ace2_corrected > 0)) %>%
    arrange(-dr_raw)
show(df)
#> Registered S3 method overwritten by 'cli':
#>   method     from
#>   print.boxx spatstat
#> # A tibble: 15 x 6
#> # Groups:   orig.ident [15]
#>    orig.ident   age cells median_depth dr_raw dr_corrected
#>    <fct>      <dbl> <int>        <dbl>  <dbl>        <dbl>
#>  1 SRR8426371     3    10        4146  0.4          0
#>  2 SRR8426359    24    21        2409  0.286        0.190
#>  3 SRR8426364    24    16        4658  0.25         0.0625
#>  4 SRR8426362    24    22        2916. 0.227        0.0455
#>  5 SRR8426365     3    41        2712  0.220        0.122
#>  6 SRR8426369    24    28        2597  0.214        0.107
#>  7 SRR8426366     3   107        3357  0.168        0.0654
#>  8 SRR8426370    24    19        3763  0.158        0.0526
#>  9 SRR8426372     3    42        1941  0.143        0.0952
#> 10 SRR8426361     3    65        2002  0.123        0.123
#> 11 SRR8426367     3    71        2458  0.0986       0.0423
```

```
#> 12 SRR8426368     3    52        3255  0.0962       0.0962
#> 13 SRR8426358    24    25        2753  0.08         0.04
#> 14 SRR8426363    24    13        2541  0.0769       0
#> 15 SRR8426360     3    30        3562. 0.0333       0
p1 <- ggplot(df, aes(factor(age), dr_raw)) + geom_boxplot() + geom_jitter(width = 0.2) +
    xlab("Age") + ylab("Detection rate [raw]")
p3 <- ggplot(df, aes(factor(age), median_depth)) + geom_boxplot() + geom_jitter(width = 0.2) +
    xlab("Age") + ylab("Median total cell count")
p2 <- ggplot(df, aes(factor(age), dr_corrected)) + geom_boxplot() + geom_jitter(width = 0.2) +
    xlab("Age") + ylab("Detection rate [corrected]")
show(p1 | p3 | p2)
```



Looks like there is no significant difference in detection rate between the two groups.

Run a differential expression test between the age groups within just cluster 6

```
s6 <- subset(s, idents = "6")
s6 <- SCTransform(s6, verbose = FALSE, method = "qpoisson")
de_res <- sctransform::diff_mean_test(y = s6$SCT@counts, group_labels = s6$age,
    log2FC_th = 0, compare = c("3", "24"), R = 499, mean_th = 0.1, mean_type = "arithmetic")
#> Non-parametric DE test for count data
#> Using arithmetic mean and 499 random permutations
#> Input: 11429 genes, 562 cells; 2 groups
#> Comparing 3 (group1, N = 418) to 24 (group2, N = 144)
#> Keeping 3101 genes after initial filtering
```

Show results for ACE2

```
filter(de_res, gene == ace2_gene) %>% select(mean1, cells1, mean2, cells2,
    log2FC, emp_pval_adj, pval_adj)
```

	mean1	cells1	mean2	cells2	log2FC	emp_pval_adj	pval_adj
ENSMUSG00000015405	0.1507177	48	0.1527778	21	-0.0195858	1	0.9765725

Finally, run the DE test (age 3 vs 24) for all clusters for just ACE2. Clusters where ACE2 is not detected in at least 5 cells in one of the groups will be skipped.

```
tmp <- lapply(levels(Idents(s)), function(cluster) {
    s_tmp <- subset(s, idents = cluster)
    s_tmp <- SCTransform(s_tmp, verbose = FALSE, method = "qpoisson")
    if (ace2_gene %in% rownames(s_tmp$SCT@counts)) {
        tmp_res <- sctransform::diff_mean_test(y = s_tmp$SCT@counts[ace2_gene,
            , drop = FALSE], group_labels = s_tmp$age, compare = c("3",
            "24"), log2FC_th = 0, mean_th = 0, cells_th = 0, R = 999, verbosity = 0)
        tmp_res$cluster = cluster
        rownames(tmp_res) <- NULL
        return(tmp_res)
    }
    return(NULL)
})
de_res2 <- do.call(rbind, tmp)
de_res2$emp_pval_adj <- p.adjust(de_res2$emp_pval)
de_res2$pval_adj <- p.adjust(de_res2$pval)
select(de_res2, cluster, mean1, cells1, mean2, cells2, log2FC, emp_pval_adj,
    pval_adj)
```

cluster	mean1	cells1	mean2	cells2	log2FC	emp_pval_adj	pval_adj
0	0.0030636	5	0.0099721	5	-1.702670	0.164	0.1487241
1	0.0051948	9	0.0161745	5	-1.638586	0.164	0.1302902
6	0.0984106	48	0.1094892	21	-0.153902	0.870	0.6828002
14	0.0126824	5	0.0040147	1	1.659478	0.870	0.5704000

ACE2 is not DE in any of the clusters.

Session info

Session info

```
sessionInfo()
#> R version 4.0.2 (2020-06-22)
#> Platform: x86_64-apple-darwin17.0 (64-bit)
#> Running under: macOS Catalina 10.15.7
#>
#> Matrix products: default
#> BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
#> LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
#>
#> locale:
#> [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
#>
#> attached base packages:
#> [1] stats     graphics  grDevices utils     datasets  methods   base
#>
#> other attached packages:
#> [1] patchwork_1.1.0.9000   ggrepel_0.8.2          dplyr_1.0.2
#> [4] knitr_1.30             Seurat_3.9.9.9008      sctransform_0.3.2.9005
#> [7] reshape2_1.4.4         ggplot2_3.3.2          Matrix_1.2-18
#>
```

```
#> loaded via a namespace (and not attached):
#>   [1] nlme_3.1-149          matrixStats_0.57.0    RcppAnnoy_0.0.16
#>   [4] RColorBrewer_1.1-2    httr_1.4.2            tools_4.0.2
#>   [7] utf8_1.1.4            R6_2.5.0              irlba_2.3.3
#>  [10] rpart_4.1-15          KernSmooth_2.23-17    uwot_0.1.8.9001
#>  [13] mgcv_1.8-33           lazyeval_0.2.2        colorspace_2.0-0
#>  [16] withr_2.3.0           tidyselect_1.1.0      gridExtra_2.3
#>  [19] compiler_4.0.2        cli_2.2.0             formatR_1.7
#>  [22] plotly_4.9.2.1        labeling_0.4.2        scales_1.1.1
#>  [25] lmtest_0.9-38         spatstat.data_1.4-3   ggridges_0.5.2
#>  [28] pbapply_1.4-3         spatstat_1.64-1       goftest_1.2-2
#>  [31] stringr_1.4.0         digest_0.6.27         spatstat.utils_1.17-0
#>  [34] rmarkdown_2.5         pkgconfig_2.0.3       htmltools_0.5.1.1
#>  [37] highr_0.8             fastmap_1.0.1         htmlwidgets_1.5.2
#>  [40] rlang_0.4.9           shiny_1.5.0           farver_2.0.3
#>  [43] generics_0.0.2        zoo_1.8-8             jsonlite_1.7.2
#>  [46] ica_1.0-2             magrittr_2.0.1        fansi_0.4.1
#>  [49] Rcpp_1.0.5            munsell_0.5.0         abind_1.4-5
#>  [52] reticulate_1.16       lifecycle_0.2.0       stringi_1.5.3
#>  [55] yaml_2.2.1            MASS_7.3-53           Rtsne_0.15
#>  [58] plyr_1.8.6            grid_4.0.2            parallel_4.0.2
#>  [61] listenv_0.8.0         promises_1.1.1        crayon_1.3.4.9000
#>  [64] deldir_0.1-29         miniUI_0.1.1.1        lattice_0.20-41
#>  [67] cowplot_1.1.0         splines_4.0.2         tensor_1.5
#>  [70] pillar_1.4.7          igraph_1.2.6          future.apply_1.6.0
#>  [73] codetools_0.2-16      leiden_0.3.3          glue_1.4.2
#>  [76] evaluate_0.14         data.table_1.13.2     vctrs_0.3.5
#>  [79] png_0.1-7             httpuv_1.5.4          gtable_0.3.0
#>  [82] RANN_2.6.1            purrr_0.3.4           polyclip_1.10-0
#>  [85] tidyr_1.1.2           assertthat_0.2.1      future_1.19.1
#>  [88] xfun_0.19             rsvd_1.0.3            mime_0.9
#>  [91] xtable_1.8-4          RSpectra_0.16-0       later_1.1.0.1
#>  [94] survival_3.2-3        viridisLite_0.3.0     tibble_3.0.4
#>  [97] cluster_2.1.0         globals_0.13.1        fitdistrplus_1.1-1
#> [100] ellipsis_0.3.1        ROCR_1.0-11
```