

# add best lineage barcodes

Andy Sposato

The purpose of this notebook is to add the best lineage barcode to each cell for which there is a high confidence lineage barcode. We generated a table with cell id and barcode information in the python notebook called find-best-lineage-barcode.ipynb. If you have not yet worked through that script, go back and do so.

```
# load required packages
```

```
library(dplyr)
library(Seurat)
library(ggplot2)
library(stringr)
library(cowplot)
```

## Connect lineage and transcriptome for unedited

```
# load the Seurat objects
```

```
unedited <- readRDS("data/objects/unedited_w_diffscores.rds")
```

```
# load lineage barcode data
```

```
# These tables were written as outputs from find-best-lineage-barcode.ipynb
```

```
un1_data <- read.table('data/tables/unedited1_bestlineagebarcode.tsv',
                      sep = '\t', header = T)
```

```
un2_data <- read.table('data/tables/unedited2_bestlineagebarcode.tsv',
                      sep = '\t', header = T)
```

```
un_meta <- unedited@meta.data
```

```
# add a column that holds cell.id (10X barcode)
```

```
un_meta$cell.id <- row.names(un_meta)
```

```
# we have a data frame of barcodes associated with unedited1 or unedited2
```

```
# (un1_data and un2_data)
```

```
# let's treat these metadata frames separately and then combine them later with rbind
```

```
# before adding back as meta data to the Seurat object
```

```
un_meta_1 <- subset(un_meta, orig.ident == "mo22_A")
```

```
un_meta_2 <- subset(un_meta, orig.ident == "mo22_B")
```

```
# get list of cell ids as they appear in each meta data frame
```

```
un_cells_1 <- as.data.frame(un_meta_1$cell.id)
```

```
names(un_cells_1)[names(un_cells_1) == "un_meta_1$cell.id"] <- "cell.id"
```

```
un_cells_2 <- as.data.frame(un_meta_2$cell.id)
```

```
names(un_cells_2)[names(un_cells_2) == "un_meta_2$cell.id"] <- "cell.id"
```

```
source("bin/functions_by_andy.R")
```

```
# check the cell.id in your seurat objects to verify the correct suffix
```

```
# is being added to your barcode dataframes
```

```
head(un1_data$cell.id, n = 2)
```

```

## [1] "CTCAACCGTGACTCGC" "GGAAGTGAGTAGGGTC"
un1_data$cell.id <- add_chars(un1_data$cell.id, '-1_1')
head(un1_data$cell.id, n = 2)

## [1] "CTCAACCGTGACTCGC-1_1" "GGAAGTGAGTAGGGTC-1_1"
head(un2_data$cell.id, n = 2)

## [1] "TGTTCCGGTCGTA" "AAGTACCGTTCTCAGA"
un2_data$cell.id <- add_chars(un2_data$cell.id, '-1_2')
head(un2_data$cell.id, n = 2)

## [1] "TGTTCCGGTCGTA-1_2" "AAGTACCGTTCTCAGA-1_2"
# using left_join from dplyr maintains the order of the cell_ids according to the Seurat
# object's metadata (stored those in un_cells_1 and un_cells_2) and only uses barcodes
# for cells which are in the transcriptome (Seurat object)
un_cells_1_barcodes <- left_join(un_cells_1, un1_data)

## Joining with `by = join_by(cell.id)`
un_cells_2_barcodes <- left_join(un_cells_2, un2_data)

## Joining with `by = join_by(cell.id)`
# add category to differentiate sub samples
# this is likely already stored in your metadata under orig.ident
un_cells_1_barcodes$condition <- "unedited1"
un_cells_2_barcodes$condition <- "unedited2"

# recopy unedited@meta.data
# don't want extra cell.id column this time
# subset for just unedited1 (mo22_A) cells
# and subset for just unedited2 (mo22_B) cells
Cells = unedited@meta.data
Cells1 <- subset(Cells, orig.ident == "mo22_A")
Cells2 <- subset(Cells, orig.ident == "mo22_B")

# we define all the lineage barcode information we want to add
barcode_count <- un_cells_1_barcodes[['barcode.count']]
barcode <- un_cells_1_barcodes[['top.barcode']]
site.1 <- un_cells_1_barcodes[['site1']]
site.2 <- un_cells_1_barcodes[['site2']]
site.3 <- un_cells_1_barcodes[['site3']]
site.4 <- un_cells_1_barcodes[['site4']]
site.5 <- un_cells_1_barcodes[['site5']]
site.6 <- un_cells_1_barcodes[['site6']]
site.7 <- un_cells_1_barcodes[['site7']]
site.8 <- un_cells_1_barcodes[['site8']]
site.9 <- un_cells_1_barcodes[['site9']]
site.10 <- un_cells_1_barcodes[['site10']]

# modify Cells1 by adding in those columns
Cells1[['barcode.count']] <- barcode_count
Cells1[['barcode']] <- barcode
Cells1[['site1']] <- site.1
Cells1[['site2']] <- site.2

```

```

Cells1['site3'] <- site.3
Cells1['site4'] <- site.4
Cells1['site5'] <- site.5
Cells1['site6'] <- site.6
Cells1['site7'] <- site.7
Cells1['site8'] <- site.8
Cells1['site9'] <- site.9
Cells1['site10'] <- site.10

# now subset the columns that are going into the Seurat object for unedited1
Cells1.info <- subset(Cells1, select = c('barcode.count', 'barcode', 'site1',
                                         'site2', 'site3', 'site4', 'site5', 'site6',
                                         'site7', 'site8', 'site9', 'site10'))

# we define all the lineage barcode information we want to add for unedited2
barcode_count <- un_cells_2_barcode[["barcode.count"]]
barcode <- un_cells_2_barcode[["top.barcode"]]
site.1 <- un_cells_2_barcode[["site1"]]
site.2 <- un_cells_2_barcode[["site2"]]
site.3 <- un_cells_2_barcode[["site3"]]
site.4 <- un_cells_2_barcode[["site4"]]
site.5 <- un_cells_2_barcode[["site5"]]
site.6 <- un_cells_2_barcode[["site6"]]
site.7 <- un_cells_2_barcode[["site7"]]
site.8 <- un_cells_2_barcode[["site8"]]
site.9 <- un_cells_2_barcode[["site9"]]
site.10 <- un_cells_2_barcode[["site10"]]

# modify Cells2 by adding in those columns
Cells2['barcode.count'] <- barcode_count
Cells2['barcode'] <- barcode
Cells2['site1'] <- site.1
Cells2['site2'] <- site.2
Cells2['site3'] <- site.3
Cells2['site4'] <- site.4
Cells2['site5'] <- site.5
Cells2['site6'] <- site.6
Cells2['site7'] <- site.7
Cells2['site8'] <- site.8
Cells2['site9'] <- site.9
Cells2['site10'] <- site.10

# now subset the columns that are going into the Seurat object for unedited2
Cells2.info <- subset(Cells2, select = c('barcode.count', 'barcode', 'site1',
                                         'site2', 'site3', 'site4', 'site5', 'site6',
                                         'site7', 'site8', 'site9', 'site10'))

# stack these two data frames on top of each other with rbind()
Cells.both_un <- rbind(Cells1.info, Cells2.info)

# for plotting purposes later, we want to change 'NA' barcode.count values to 0
Cells.both_un$barcode.count[is.na(Cells.both_un$barcode.count)] <- 0

# add back to Seurat object unedited
unedited <- AddMetaData(unedited, Cells.both_un)

```

```
saveRDS(unedited, "data/objects/unedited_w_diffscores_lineage.rds")
```

## Connect lineage and transcriptome for edited

```
# load the Seurat objects
edited <- readRDS("data/objects/edited_w_diffscores.rds")

# load lineage barcode data
ed1_data <- read.table('data/tables/inj_heat1_bestlineagebarcode.tsv',
                      sep = '\t', header = T)
ed2_data <- read.table('data/tables/inj_heat2_bestlineagebarcode.tsv',
                      sep = '\t', header = T)

ed_meta <- edited@meta.data
# add a column that holds cell.id (10X barcode)
ed_meta$cell.id <- row.names(ed_meta)

ed_meta_1 <- subset(ed_meta, orig.ident == "mo20_A")
ed_meta_2 <- subset(ed_meta, orig.ident == "mo20_B")

# get list of cell ids as they appear in each meta data frame
ed_cells_1 <- as.data.frame(ed_meta_1$cell.id)
names(ed_cells_1)[names(ed_cells_1) == "ed_meta_1$cell.id"] <- "cell.id"
ed_cells_2 <- as.data.frame(ed_meta_2$cell.id)
names(ed_cells_2)[names(ed_cells_2) == "ed_meta_2$cell.id"] <- "cell.id"

# we added the add_chars function earlier
head(ed1_data$cell.id, n = 2)

## [1] "ATGTCCCCATTACAG" "GTAACCAAGTATTGCC"
ed1_data$cell.id <- add_chars(ed1_data$cell.id, '-1')
head(ed1_data$cell.id, n = 2)

## [1] "ATGTCCCCATTACAG-1" "GTAACCAAGTATTGCC-1"
head(ed2_data$cell.id, n = 2)

## [1] "GTAGAGGCAAATGAGT" "TCAGCCTTCGATTCCC"
ed2_data$cell.id <- add_chars(ed2_data$cell.id, '-1')
head(ed2_data$cell.id, n = 2)

## [1] "GTAGAGGCAAATGAGT-1" "TCAGCCTTCGATTCCC-1"
ed_cells_1_barcode <- left_join(ed_cells_1, ed1_data)

## Joining with `by = join_by(cell.id)`
ed_cells_2_barcode <- left_join(ed_cells_2, ed2_data)

## Joining with `by = join_by(cell.id)`
# this is really just differentiating technical replicates
# same biological samples, different channels of 10X
ed_cells_1_barcode$condition <- "edited1"
ed_cells_2_barcode$condition <- "edited2"
```

```

# recopy edited@meta.data
# don't want extra cell.id column this time
# subset for just edited1 cells
# also subset for just edited2 cells
Cells = edited@meta.data
Cells1 <- subset(Cells, orig.ident == "mo20_A")
Cells2 <- subset(Cells, orig.ident == "mo20_B")

# we define all the lineage barcode information we want to add
barcode_count <- ed_cells_1_barcodes[['barcode.count']]
barcode <- ed_cells_1_barcodes[['top.barcode']]
site.1 <- ed_cells_1_barcodes[['site1']]
site.2 <- ed_cells_1_barcodes[['site2']]
site.3 <- ed_cells_1_barcodes[['site3']]
site.4 <- ed_cells_1_barcodes[['site4']]
site.5 <- ed_cells_1_barcodes[['site5']]
site.6 <- ed_cells_1_barcodes[['site6']]
site.7 <- ed_cells_1_barcodes[['site7']]
site.8 <- ed_cells_1_barcodes[['site8']]
site.9 <- ed_cells_1_barcodes[['site9']]
site.10 <- ed_cells_1_barcodes[['site10']]

# modify Cells1 by adding in those columns
Cells1['barcode.count'] <- barcode_count
Cells1['barcode'] <- barcode
Cells1['site1'] <- site.1
Cells1['site2'] <- site.2
Cells1['site3'] <- site.3
Cells1['site4'] <- site.4
Cells1['site5'] <- site.5
Cells1['site6'] <- site.6
Cells1['site7'] <- site.7
Cells1['site8'] <- site.8
Cells1['site9'] <- site.9
Cells1['site10'] <- site.10

# now subset the columns that are going into the Seurat object for edited1
Cells1.info <- subset(Cells1, select = c('barcode.count', 'barcode', 'site1',
                                         'site2', 'site3', 'site4', 'site5', 'site6',
                                         'site7', 'site8', 'site9', 'site10'))

# we define all the lineage barcode information we want to add for edited2
barcode_count <- ed_cells_2_barcodes[['barcode.count']]
barcode <- ed_cells_2_barcodes[['top.barcode']]
site.1 <- ed_cells_2_barcodes[['site1']]
site.2 <- ed_cells_2_barcodes[['site2']]
site.3 <- ed_cells_2_barcodes[['site3']]
site.4 <- ed_cells_2_barcodes[['site4']]
site.5 <- ed_cells_2_barcodes[['site5']]
site.6 <- ed_cells_2_barcodes[['site6']]
site.7 <- ed_cells_2_barcodes[['site7']]
site.8 <- ed_cells_2_barcodes[['site8']]
site.9 <- ed_cells_2_barcodes[['site9']]
site.10 <- ed_cells_2_barcodes[['site10']]

```

```

# modify Cells2 by adding in those columns
Cells2['barcode.count'] <- barcode_count
Cells2['barcode'] <- barcode
Cells2['site1'] <- site.1
Cells2['site2'] <- site.2
Cells2['site3'] <- site.3
Cells2['site4'] <- site.4
Cells2['site5'] <- site.5
Cells2['site6'] <- site.6
Cells2['site7'] <- site.7
Cells2['site8'] <- site.8
Cells2['site9'] <- site.9
Cells2['site10'] <- site.10

# now subset the columns that are going into the Seurat object for edited2
Cells2.info <- subset(Cells2, select = c('barcode.count', 'barcode', 'site1',
                                         'site2', 'site3', 'site4', 'site5', 'site6',
                                         'site7', 'site8', 'site9', 'site10'))

# stack these two data frames on top of each other with rbind()
Cells.both_ed <- rbind(Cells1.info, Cells2.info)

# for plotting purposes later, we want to change 'NA' barcode.count labels to 0
Cells.both_ed$barcode.count[is.na(Cells.both_ed$barcode.count)] <- 0

# add back to Seurat object unedited
edited <- AddMetaData(edited, Cells.both_ed)

saveRDS(edited, "data/objects/edited_w_diffscores_lineage.rds")

```