

# 10xgenomics-Humandata

```
library(dplyr)

library(tidyr)

data = read.csv
('/home/sedreh/university_files/ITM0/semester2/Bcellsproject/fift
h_session/Human_dataset/human_data.csv')
filtered_data = read.csv
('/home/sedreh/university_files/ITM0/semester2/Bcellsproject/fift
h_session/Human_dataset/human_filtered_data.csv')

#Number of cells (As we have several copies of each barcode,we
need to count just one copy of unique barcoe)

barcode_summary <- function(data)
{
  number_of_cells <- data %>%
    distinct(barcode) %>%
    count()
  number_of_cells$n
}
barcode_summary(data)

## [1] 17057

barcode_summary(filtered_data)

## [1] 9465

#look at to the quality of chains

count_table <- function(data)
{
  v_gene_and_j_gene <- filter(data, v_gene != 'None' & j_gene !=
'None')
  v_gene_and_j_gene <- count(v_gene_and_j_gene)$n

  no_v_gene_and_no_j_gene <- filter(data, v_gene == 'None' & j_gene
== 'None')
  no_v_gene_and_no_j_gene <- count(no_v_gene_and_no_j_gene)$n

  v_gene_or_j_gene <- filter(data, v_gene != 'None' | j_gene !=
'None')
  v_gene_or_j_gene <- count(v_gene_or_j_gene)$n

  stats <- data.frame(v_gene_and_j_gene, v_gene_or_j_gene,
no_v_gene_and_no_j_gene)
```

```

stats
}
count_table(data)

##   v_gene_and_j_gene v_gene_or_j_gene no_v_gene_and_no_j_gene
## 1                31861                43926                10717

count_table(filtered_data)

##   v_gene_and_j_gene v_gene_or_j_gene no_v_gene_and_no_j_gene
## 1                22044                27575                5763

#How many IGH, IGH, IGL and Multi we have in the data?
Summary <- function(data)
{
  chain_summary <- group_by(data, chain)
  chain_summary <- summarize(chain_summary, count=n())
  chain_summary
}
Summary (data)

## # A tibble: 9 x 2
##   chain count
##   <fct> <int>
## 1 IGH    19351
## 2 IGK    12270
## 3 IGL     8137
## 4 Multi  11250
## 5 None    1279
## 6 TRA     1471
## 7 TRB      795
## 8 TRD       27
## 9 TRG       63

Summary (filtered_data)

## # A tibble: 9 x 2
##   chain count
##   <fct> <int>
## 1 IGH    11380
## 2 IGK     7209
## 3 IGL     5171
## 4 Multi   6401
## 5 None    1087
## 6 TRA     1282
## 7 TRB      738
## 8 TRD       21
## 9 TRG       49

#Number of copies of each cell

barcode_summary <- function(data)

```

```

{
  barcode_summary <- group_by(data, barcode)
  barcode_summary <- summarize(barcode_summary, count=n())
  barcode_summary
}
barcode_summary(data)

## # A tibble: 17,057 x 2
##   barcode          count
##   <fct>          <int>
## 1 AAACCTGAGACCACGA-1      5
## 2 AAACCTGAGAGGGATA-1      1
## 3 AAACCTGAGCGTAATA-1      5
## 4 AAACCTGAGGCATTGG-1      6
## 5 AAACCTGAGTGAAGTT-1      6
## 6 AAACCTGCAAGTCTAC-1      7
## 7 AAACCTGCAATGCCAT-1      1
## 8 AAACCTGCAATGGACG-1      6
## 9 AAACCTGCACCGCTAG-1      1
## 10 AAACCTGCACTGTTAG-1     2
## # ... with 17,047 more rows

barcode_summary(filtered_data)

## # A tibble: 9,465 x 2
##   barcode          count
##   <fct>          <int>
## 1 AAACCTGAGACCACGA-1      4
## 2 AAACCTGAGCGTAATA-1      3
## 3 AAACCTGAGGCATTGG-1      4
## 4 AAACCTGAGTGAAGTT-1      5
## 5 AAACCTGCAAGTCTAC-1      5
## 6 AAACCTGCAATGGACG-1      4
## 7 AAACCTGCACTGTTAG-1      2
## 8 AAACCTGCACTTACGA-1      3
## 9 AAACCTGCATCGTCGG-1      3
## 10 AAACCTGCATGACGGA-1      4
## # ... with 9,455 more rows

#show the number of occurrence of each copy

barcode_summary <- function(data)
{
  barcode_summary <- group_by(data, barcode)
  barcode_summary <- summarize(barcode_summary, count=n())
  barcode_summary <- group_by(barcode_summary, count)
  #barcode_summary <- summarize(barcode_summary, count_total=n())
  barcode_summary
}
result_row <- barcode_summary(data)
result_row

```

```

## # A tibble: 17,057 x 2
## # Groups:   count [17]
##   barcode          count
##   <fct>          <int>
## 1 AAACCTGAGACCACGA-1      5
## 2 AAACCTGAGAGGGATA-1      1
## 3 AAACCTGAGCGTAATA-1      5
## 4 AAACCTGAGGCATTGG-1      6
## 5 AAACCTGAGTGAAGTT-1      6
## 6 AAACCTGCAAGTCTAC-1      7
## 7 AAACCTGCAATGCCAT-1      1
## 8 AAACCTGCAATGGACG-1      6
## 9 AAACCTGCACCGCTAG-1      1
## 10 AAACCTGCACTGTTAG-1     2
## # ... with 17,047 more rows

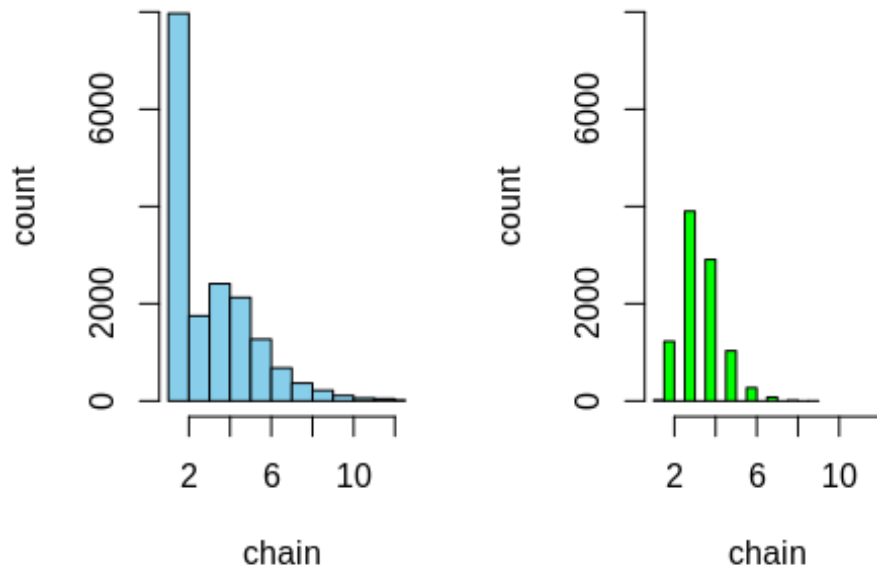
barcode_summary <- function(filtered_data)
{
  barcode_summary <- group_by(filtered_data, barcode)
  barcode_summary <- summarize(barcode_summary, count=n())
  barcode_summary <- group_by(barcode_summary, count)
  #barcode_summary <- summarize(barcode_summary, count_total=n())
  barcode_summary
}
result_filtered <- barcode_summary(filtered_data)
result_filtered

## # A tibble: 9,465 x 2
## # Groups:   count [9]
##   barcode          count
##   <fct>          <int>
## 1 AAACCTGAGACCACGA-1      4
## 2 AAACCTGAGCGTAATA-1      3
## 3 AAACCTGAGGCATTGG-1      4
## 4 AAACCTGAGTGAAGTT-1      5
## 5 AAACCTGCAAGTCTAC-1      5
## 6 AAACCTGCAATGGACG-1      4
## 7 AAACCTGCACTGTTAG-1      2
## 8 AAACCTGCACTTACGA-1      3
## 9 AAACCTGCATCGTCGG-1      3
## 10 AAACCTGCATGACGGA-1     4
## # ... with 9,455 more rows

par(mfrow = c(1,2))
hist(result_row$count, xlim=c(1,12),ylim=c(0,8000), xlab="chain",
ylab="count", col="skyblue")
hist(result_filtered$count, xlim=c(1,12),ylim=c(0,8000),
xlab="chain", ylab="count", col="green")

```

## Histogram of result\_row\$chain vs histogram of result\_filtered\$chain



*#counts number of occurrence of each chain type for every unique barcode how*

```
occurrence_of_each_chain <- function(data)
{
  result_data <- data %>%
    group_by(barcode, chain) %>%
    filter(chain %in% c('IGH', 'IGH', 'IGL')) %>%
    summarize(count=n())
  result_data
}
occurrence_of_each_chain(data)
```

```
## # A tibble: 27,499 x 3
## # Groups:   barcode [16,507]
##   barcode          chain count
##   <fct>          <fct> <int>
## 1 AAACCTGAGACCACGA-1 IGH      2
## 2 AAACCTGAGACCACGA-1 IGL      1
## 3 AAACCTGAGAGGGATA-1 IGL      1
## 4 AAACCTGAGCGTAATA-1 IGH      3
## 5 AAACCTGAGCGTAATA-1 IGL      1
## 6 AAACCTGAGGCATTGG-1 IGH      3
## 7 AAACCTGAGGCATTGG-1 IGL      2
## 8 AAACCTGAGTGAAGTT-1 IGH      2
## 9 AAACCTGAGTGAAGTT-1 IGK      1
```

```

## 10 AAACCTGCAAGTCTAC-1 IGH      2
## # ... with 27,489 more rows

# I have all conditions here

occurance_of_each_chain <- function(data)
{
  result <- data %>%
    group_by(barcode, chain) %>%
    filter(chain %in% c('IGK', 'IGH', 'IGL')) %>%
    summarize(count=n())
  result <- unite(result, 'result_chain', count, chain, remove=F,
sep='')
  result <- summarize(result, type=paste(result_chain,
collapse='_'), count=sum(count))
  result
}
occurance_of_each_chain(data)

## # A tibble: 17,057 x 3
##   barcode      type      count
##   <fct>      <chr>    <int>
## 1 AAACCTGAGACCACGA-1 2IGH_1IGL      3
## 2 AAACCTGAGAGGGATA-1 1IGL           1
## 3 AAACCTGAGCGTAATA-1 3IGH_1IGL      4
## 4 AAACCTGAGGCATTGG-1 3IGH_2IGL      5
## 5 AAACCTGAGTGAAGTT-1 2IGH_1IGK      3
## 6 AAACCTGCAAGTCTAC-1 2IGH_2IGK      4
## 7 AAACCTGCAATGCCAT-1 1IGK           1
## 8 AAACCTGCAATGGACG-1 2IGH_2IGK      4
## 9 AAACCTGCACCGCTAG-1 1IGK           1
## 10 AAACCTGCACTGTTAG-1 1IGH_1IGK      2
## # ... with 17,047 more rows

# I have all conditions here for filtered data

occurance_of_each_chain <- function(filtered_data)
{
  result <- filtered_data %>%
    group_by(barcode, chain) %>%
    filter(chain %in% c('IGK', 'IGH', 'IGL')) %>%
    summarize(count=n())
  result <- unite(result, 'result_chain', count, chain, remove=F,
sep='')
  result <- summarize(result, type=paste(result_chain,
collapse='_'), count=sum(count))
  result
}
occurance_of_each_chain(filtered_data)

```

```
## # A tibble: 9,465 x 3
##   barcode          type      count
##   <fct>          <chr>    <int>
## 1 AAACCTGAGACCACGA-1 1IGH_1IGL      2
## 2 AAACCTGAGCGTAATA-1 1IGH_1IGL      2
## 3 AAACCTGAGGCATTGG-1 2IGH_1IGL      3
## 4 AAACCTGAGTGAAGTT-1 2IGH_1IGK      3
## 5 AAACCTGCAAGTCTAC-1 2IGH_1IGK      3
## 6 AAACCTGCAATGGACG-1 1IGH_1IGK      2
## 7 AAACCTGCACTGTTAG-1 1IGH_1IGK      2
## 8 AAACCTGCACTTACGA-1 1IGH_1IGL      2
## 9 AAACCTGCATCGTCGG-1 1IGH_1IGL      2
## 10 AAACCTGCATGACGGA-1 1IGH_2IGL      3
## # ... with 9,455 more rows
```

*# I added whichever condition that I want from data*

```
Condition <- function (result)
{
  many_conditions = c(
    '1IGL', '1IGH', '1IGK', '1IGH_1IGK_1IGL' , '1IGH_1IGL' ,
    '1IGH_1IGK', '1IGH_2IGK', 'IGH_2IGL')
  result %>%
    filter(type %in% many_conditions) %>%
    group_by(type) %>%
    summarise(total=n())
}
Condition(occurrence_of_each_chain(data))
```

```
## # A tibble: 7 x 2
##   type      total
##   <chr>    <int>
## 1 1IGH      2044
## 2 1IGH_1IGK 1927
## 3 1IGH_1IGK_1IGL 267
## 4 1IGH_1IGL  926
## 5 1IGH_2IGK  566
## 6 1IGK     3140
## 7 1IGL     1363
```

*# I added whichever condition that I want from data*

```
Condition <- function (result)
{
  many_conditions = c(
    '1IGL', '1IGH', '1IGK', '1IGH_1IGK_1IGL' , '1IGH_1IGL' ,
    '1IGH_1IGK', '1IGH_2IGK', 'IGH_2IGL')
  result %>%
    filter(type %in% many_conditions) %>%
```

```

    group_by(type) %>%
    summarise(total=n())
  }
Condition(occurance_of_each_chain(filtered_data))

## # A tibble: 7 x 2
##   type      total
##   <chr>    <int>
## 1 1IGH      4
## 2 1IGH_1IGK 3603
## 3 1IGH_1IGK_1IGL 662
## 4 1IGH_1IGL 2039
## 5 1IGH_2IGK 492
## 6 1IGK     41
## 7 1IGL     19

```