

10xgenomics_Mousedata

```
library(dplyr)

library(tidyr)

data1 = read.csv
('/home/sedreh/university_files/ITM0/semester2/Bcellsproject/third
session/filteredreddata.csv')
summary(data1)
```

```
##          barcode      is_cell
## AGAGCGAGTTTCCACC-1:      9 Mode:logical
## AGGGTGAAGAACAATC-1:      9 TRUE:16036
## AGGGTGACATGAAGTA-1:      9
## GTCAAGTTCTCCAGGG-1:      9
## TAAGAGACATGCCCGA-1:      9
## AACTTTCAGCCTCGTG-1:      8
## (Other)                  :15983
##          contig_id      high_confidence
length
## AAACCTGAGACCTTTG-1_contig_1:      1 Mode:logical   Min.    :
250.0
## AAACCTGAGACCTTTG-1_contig_2:      1 TRUE:16036     1st Qu.:
499.0
## AAACCTGAGACCTTTG-1_contig_5:      1                Median :
533.0
## AAACCTGAGACCTTTG-1_contig_6:      1                Mean    :
568.7
## AAACCTGAGACCTTTG-1_contig_7:      1                3rd Qu.:
680.0
## AAACCTGAGCAACGGT-1_contig_1:      1
Max.    :1130.0
## (Other)                  :16030

##          chain          v_gene          d_gene          j_gene
## IGK      :5673   None      :5170   None      :11614   IGKJ1   :1802
## IGH      :5385   IGLV1     : 377   IGHD1-1: 1453   IGHJ2   :1738
## Multi    :2717   IGKV1-135: 292   IGHD2-4:  609   IGKJ2   :1559
## IGL      :1873   IGKV10-96: 291   IGHD2-3:  561   IGHJ4   :1503
## TRA      : 274   IGHV1-26 : 268   IGHD4-1:  491   TRAJ9   :1380
## None     :  65   IGKV1-117: 214   IGHD2-6:  400   IGKJ5   :1353
```

```

## (Other): 49 (Other) :9424 (Other): 908 (Other):6701
## c_gene full_length productive cdr3
## IGKC :5685 Mode :logical FALSE: 389 None :6153
## IGHM :5570 FALSE:5199 None :6154 CALWYSNHWVF: 144
## IGHD :1876 TRUE :10837 TRUE :9493 CALWYSTHYVF: 114
## IGLC2 :1357 CQQGNTLPWTF: 100
## IGLC1 : 841 CQQHYSTPYTF: 78
## None : 366 CQQWSSNPPTF: 77
## (Other): 341 (Other) :9370

## cdr3_nt reads
umis
## None :6153 Min. : 20
Min. : 1
## TGTGCTCTATGGTACAGCAACCATTGGGTGTTC: 117 1st Qu.: 430 1st
Qu.: 3
## TGTGCTCTATGGTACAGCACCCATTATGTTTTC: 114 Median : 1128
Median : 9
## TGCCAACAGGGTAATACGCTTCCGTGGACGTTC: 99 Mean : 2682
Mean : 29
## TGCCAGCAGTGGAGTAGTAACCCACCCACGTTC: 72 3rd Qu.: 2934 3rd
Qu.: 20
## TGCTTTCAAGGTTACATGTTCCGTGGACGTTC: 66 Max. :180397
Max. :13527
## (Other) :9415

## raw_clonotype_id raw_consensus_id
## clonotype3 : 12 None :6546
## clonotype9 : 11 clonotype1_consensus_1: 5
## clonotype28: 10 clonotype2_consensus_1: 4
## clonotype38: 10 clonotype3_consensus_1: 4
## clonotype6 : 10 clonotype3_consensus_2: 4
## clonotype10: 9 clonotype4_consensus_1: 4
## (Other) :15974 (Other) :9469

```

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

```

barcode_summary <- function(data1)
{

```

```

    number_of_cells <- data1 %>%
      distinct(barcode) %>%
      count()
    number_of_cells$n
  }
barcode_summary(data1)

## [1] 4424

#look at to the quality of chains

count_table <- function(data1)
{
  v_gene_and_j_gene <- filter(data1, 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(data1, 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(data1, 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(data1)

##   v_gene_and_j_gene v_gene_or_j_gene no_v_gene_and_no_j_gene
## 1             10845             15637                 399

#How many IGK,IGH, IGL and Multi we have in the data?
Summary <- function(data1)
{
  chain_summary <- group_by(data1, chain)
  chain_summary <- summarize(chain_summary, count=n())
  chain_summary
}
Summary (data1)

## # A tibble: 8 x 2
##   chain count
##   <fct> <int>
## 1 IGH     5385
## 2 IGK     5673
## 3 IGL     1873
## 4 Multi  2717
## 5 None     65

```

```
## 6 TRA      274
## 7 TRD       2
## 8 TRG      47
```

#Number of copies of each cell

```
barcode_summary <- function(data1)
{
  barcode_summary <- group_by(data1, barcode)
  barcode_summary <- summarize(barcode_summary, count=n())
  barcode_summary
}
barcode_summary(data1)
```

```
## # A tibble: 4,424 x 2
##   barcode      count
##   <fct>      <int>
## 1 AAACCTGAGACCTTTG-1      5
## 2 AAACCTGAGCAACGGT-1      3
## 3 AAACCTGAGTAGCGGT-1      4
## 4 AAACCTGCAAACCCAT-1      3
## 5 AAACCTGCAATCACAC-1      3
## 6 AAACCTGCACTAGTAC-1      4
## 7 AAACCTGCAGCCTTGG-1      4
## 8 AAACCTGCATCTACGA-1      2
## 9 AAACCTGGTTATCCGA-1      3
## 10 AAACCTGGTTGATTG-1      3
## # ... with 4,414 more rows
```

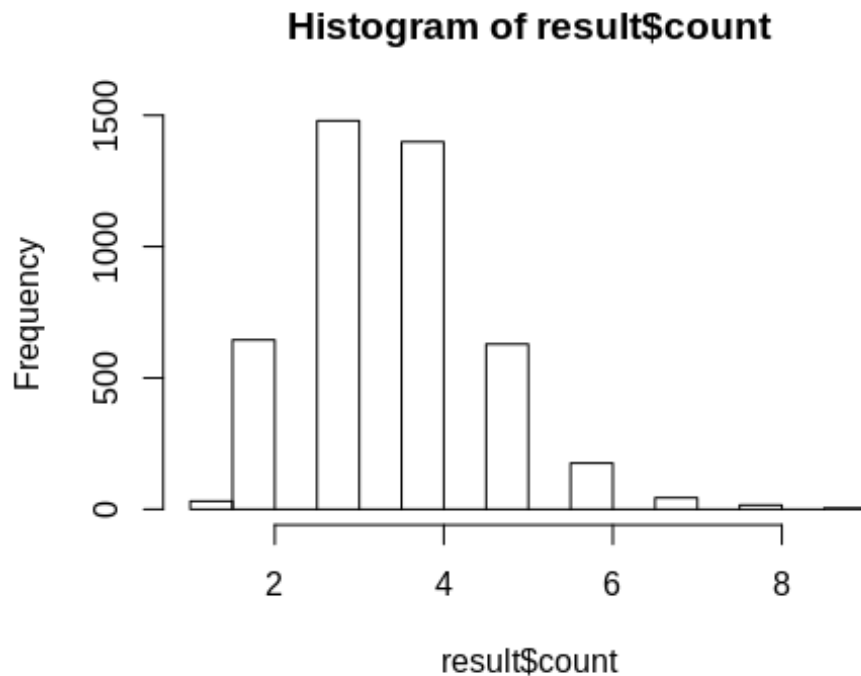
#show the number of occurrence of each copy

```
barcode_summary <- function(data1)
{
  barcode_summary <- group_by(data1, 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 <- barcode_summary(data1)
result
```

```
## # A tibble: 4,424 x 2
## # Groups:   count [9]
##   barcode      count
##   <fct>      <int>
## 1 AAACCTGAGACCTTTG-1      5
## 2 AAACCTGAGCAACGGT-1      3
## 3 AAACCTGAGTAGCGGT-1      4
## 4 AAACCTGCAAACCCAT-1      3
## 5 AAACCTGCAATCACAC-1      3
```

```
## 6 AAACCTGCACTAGTAC-1 4
## 7 AAACCTGCAGCCTTGG-1 4
## 8 AAACCTGCATCTACGA-1 2
## 9 AAACCTGGTTATCCGA-1 3
## 10 AAACCTGGTTGATTCG-1 3
## # ... with 4,414 more rows
```

```
hist(result$count)
```



#counts number of occurrence of each chain type for every unique barcodehow

```
occurance_of_each_chain <- function(data1)
{
  result <- data1 %>%
    group_by(barcode, chain) %>%
    filter(chain %in% c('IGK', 'IGH', 'IGL')) %>%
    summarize(count=n())
  result
}
occurance_of_each_chain(data1)

## # A tibble: 10,337 x 3
## # Groups:   barcode [4,424]
##   barcode          chain count
##   <fct>           <fct> <int>
## 1 AAACCTGAGACCTTTG-1 IGH      2
```

```
## 2 AAACCTGAGACCTTTG-1 IGK 2
## 3 AAACCTGAGCAACGGT-1 IGH 2
## 4 AAACCTGAGCAACGGT-1 IGK 1
## 5 AAACCTGAGTAGCGGT-1 IGH 2
## 6 AAACCTGAGTAGCGGT-1 IGK 2
## 7 AAACCTGCAAACCCAT-1 IGH 1
## 8 AAACCTGCAAACCCAT-1 IGK 2
## 9 AAACCTGCAATCACAC-1 IGH 2
## 10 AAACCTGCAATCACAC-1 IGK 1
## # ... with 10,327 more rows
```

I have all conditions here

```
occurance_of_each_chain <- function(data1)
{
  result <- data1 %>%
    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(data1)

```
## # A tibble: 4,424 x 3
##   barcode          type          count
##   <fct>          <chr>        <int>
## 1 AAACCTGAGACCTTTG-1 2IGH_2IGK          4
## 2 AAACCTGAGCAACGGT-1 2IGH_1IGK          3
## 3 AAACCTGAGTAGCGGT-1 2IGH_2IGK          4
## 4 AAACCTGCAAACCCAT-1 1IGH_2IGK          3
## 5 AAACCTGCAATCACAC-1 2IGH_1IGK          3
## 6 AAACCTGCACTAGTAC-1 2IGH_1IGL          3
## 7 AAACCTGCAGCCTTGG-1 1IGH_1IGK_1IGL      3
## 8 AAACCTGCATCTACGA-1 1IGH_1IGK          2
## 9 AAACCTGGTTATCCGA-1 1IGH_1IGK          2
## 10 AAACCTGGTTGATTCG-1 1IGH_1IGK_1IGL      3
## # ... with 4,414 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', '1IGH_2IGL')
  result %>%
```

```
  filter(type %in% many_conditions) %>%  
  group_by(type) %>%  
  summarise(total=n())  
}  
Condition(occurance_of_each_chain(data1))
```

```
## # A tibble: 7 x 2  
##   type          total  
##   <chr>        <int>  
## 1 1IGH          3  
## 2 1IGH_1IGK    1442  
## 3 1IGH_1IGK_1IGL 785  
## 4 1IGH_1IGL    115  
## 5 1IGH_2IGK    553  
## 6 1IGK         53  
## 7 1IGL         2
```