

Vgenes clinical and microbiome

Guannan Shen

March 18, 2019

Contents

```
##### Set up workspace
rm(list = ls())
library(edgeR)
library(EDASeq)
library(DESeq2)
library(knitr)
library(tidyverse)
library(magrittr)
library(stats)
library(readxl)
library(openxlsx)
# multiple testing
library(mutoss)
options(stringsAsFactors = F)
options(dplyr.width = Inf)
getwd()

## [1] "/home/guanshim/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataRaw/Vgenes"

## not in function
"%nin%" <- Negate("%in%")

# ##### clean memory ##### rm(list =
# ls()) gc() is(dds) slotNames(dds)

# remapping Hi-seq2 with moderate threshold rna-seq cnts data
# #####
cnts_raw <- read_excel("HIV_Infected_vs_Uninfected_relaxed_mapping_geneCounts.xlsx")
head(cnts_raw)

## # A tibble: 6 x 35
##   Gene_ID      Symbol      Length `138` `178` `255` `278` `361` `404`
##   <chr>         <chr>         <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 ENSG00000275877.1 5_S rRNA      152    0    0    0    0    1    0
## 2 ENSG00000283274.1 5_S rRNA      151    0    0    0    0    0    0
## 3 ENSG00000212595.1 5_S rRNA      112    0    0    0    0    0    0
## 4 ENSG00000274059.1 5_S rRNA       84    0    0    0    1    1    0
## 5 ENSG00000274408.1 5_S rRNA      123    0    0    0    0    0    0
## 6 ENSG00000274663.1 5_S rRNA       87    0    0    0    0    0    0
##   `493` `582` `708` `716` `914` `947` `972` `124` `132` `154` `188` `217`
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1    0    0    0    0    0    0    0    0    1    2    0    0
## 2    0    0    1    1    0    0    0    0    0    1    0    0
## 3    2    0    0    0    0    0    0    0    2    0    0    0
## 4    0    0    0    10   0    0    0    0    0    0    0    0
## 5    0    0    0    0    1    0    0    0    0    0    0    0
```

```
## 6      0      0      0      0      0      0      0      0      0      0      0      0      0
##      `286` `307` `323` `391` `428` `594` `622` `648` `683` `819` `825` `839`
##      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      0      0      0      0      2      0      1      0      0      0      0      0      0
## 2      0      0      0      0      0      0      0      0      0      1      0      0      1
## 3      0      0      0      0      0      0      3      0      0      0      0      0      0
## 4      0      0      0      0      0      0      1      0      0      0      0      0      1
## 5      0      0      0      0      0      0      0      0      0      0      0      0      0
## 6      0      0      0      0      0      0      0      0      0      0      0      0      0
##      `965` `998`
##      <dbl> <dbl>
## 1      0      0
## 2      0      0
## 3      0      0
## 4      0      1
## 5      0      0
## 6      0      1
```

```
##### v genes lists #####
```

```
vgenes <- read_excel("/home/guanshim/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/Vgenes/
head(vgenes)
```

```
## # A tibble: 6 x 3
##   Num Vsymbols Type
##   <dbl> <chr>   <chr>
## 1     1 IGHV1-18 heavy
## 2     2 IGHV1-2  heavy
## 3     3 IGHV1-24 heavy
## 4     4 IGHV1-3  heavy
## 5     5 IGHV1-45 heavy
## 6     6 IGHV1-46 heavy
```

```
##### clinical and microbiome data #####
```

```
clin_micro <- read_excel("/home/guanshim/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/Vgenes/
head(clin_micro)
```

```
## # A tibble: 6 x 80
##   num ID      Gender  Age HIV  pid id1      `CD4 count (cells/ul)`
##   <dbl> <chr>   <chr>  <dbl> <chr> <dbl> <chr>                <dbl>
## 1     1 MIHIV138 M      29 no   138 MIHIV138                728
## 2     2 MIHIV178 M      33 no   178 MIHIV178                736
## 3     3 MIHIV255 M      34 no   255 MIHIV255                588
## 4     4 MIHIV278 M      23 no   278 MIHIV278                532
## 5     5 MIHIV361 F      33 no   361 MIHIV361                720
## 6     6 MIHIV404 F      29 no   404 MIHIV404               1071
##   id2      `IL-6 (pg/ml)` `CRP (mg/ml)` `iFABP (pg/ml)` `sCD27 (U/ml)`
##   <chr>          <dbl>      <dbl>      <dbl>      <dbl>
## 1 MIHIV138      2.16        0.25      1863.      36.2
## 2 MIHIV178      0.51        0.65     2363.     43.2
## 3 MIHIV255      0.69        0.7       1517.     29.3
## 4 MIHIV278      0.73        0.44     2800.     32.5
## 5 MIHIV361      0.19        3.36     5352.     27.5
## 6 MIHIV404      1.35       12.4      547.      37.4
##   `CD14 (ng/ml)` `LPS (pg/ml)` `LTA\n(OD)` id3
##   <dbl>          <dbl>      <dbl> <chr>
## 1 1633.          9.52        0.182 MIHIV138
```

```

## 2      2073.      9.80      0.381 MIHIV178
## 3      1605.     14.5      0.104 MIHIV255
## 4      2146.      7.45      0.120 MIHIV278
## 5      1940.      8.68      0.162 MIHIV361
## 6      1184.      9.70      0.192 MIHIV404
## `CD38+ HLA-DR+ CD4 T cells (% of CD4 T cells)`
##                                     <dbl>
## 1                                     1.18
## 2                                     0.828
## 3                                     0.896
## 4                                     0.764
## 5                                     0.637
## 6                                     0.510
## `CD38+ HLA-DR+ CD8 T cells (% of CD8 T cells)` id4
##                                     <dbl> <chr>
## 1                                     2.32 MIHIV138
## 2                                     1.59 MIHIV178
## 3                                     1.61 MIHIV255
## 4                                     1.77 MIHIV278
## 5                                     0.772 MIHIV361
## 6                                     1.75 MIHIV404
## `CD4 T cells (% viable)` `CD8 T cells (% viable)` id5
##                                     <dbl> <dbl> <chr>
## 1                                     35.5      12.4 MIHIV138
## 2                                     35.7      15.9 MIHIV178
## 3                                     34.3       6.79 MIHIV255
## 4                                     27.5      16.1 MIHIV278
## 5                                     34.7      18.1 MIHIV361
## 6                                     27.1       4.08 MIHIV404
## `Tissue HIV RNA (per CD4 T cell)` `Plasma VL` id6      Actinobacteria
##                                     <dbl> <dbl> <chr> <dbl>
## 1                                     NA      NA MIHIV138      0.00829
## 2                                     NA      NA MIHIV178      0.0330
## 3                                     NA      NA MIHIV255      0.00864
## 4                                     NA      NA MIHIV278      0.00924
## 5                                     NA      NA MIHIV361      0.00361
## 6                                     NA      NA MIHIV404      0.0229
## Bacteroidetes Cyanobacteria Firmicutes Fusobacteria Proteobacteria
##                                     <dbl> <dbl> <dbl> <dbl> <dbl>
## 1      0.282      0.000316      0.577      0      0.00732
## 2      0.509      0      0.394      0      0.0236
## 3      0.343      0      0.544      0.00000840      0.0831
## 4      0.224      0      0.720      0      0.0193
## 5      0.192      0.0000461      0.743      0      0.0453
## 6      0.323      0.00285      0.334      0.0000828      0.223
## Spirochaetes Tenericutes Verrucomicrobia id7      Prevotellaceae
##                                     <dbl> <dbl> <dbl> <chr> <dbl>
## 1      0      0      0      MIHIV138      0.0000264
## 2      0      0      0      MIHIV178      0.000300
## 3      0.00190      0.0000168      0.00364 MIHIV255      0.000126
## 4      0      0.00414      0.00924 MIHIV278      0.00163
## 5      0      0      0.000221 MIHIV361      0.000157
## 6      0.0000166      0      0.0413 MIHIV404      0.0000828
## Bacteroidaceae Rikenellaceae Porphyromonadaceae Lachnospiraceae

```

##	<dbl>	<dbl>	<dbl>	<dbl>
## 1	0.227	0.0124	0.0329	0.239
## 2	0.468	0.00252	0.0232	0.304
## 3	0.209	0.0150	0.0307	0.374
## 4	0.174	0.0155	0.0203	0.216
## 5	0.145	0.0312	0.00193	0.488
## 6	0.256	0.0110	0.0471	0.119
##	Christensenellaceae	Ruminococcaceae	Xanthomonadaceae	Moraxellaceae
##	<dbl>	<dbl>	<dbl>	<dbl>
## 1	0	0.123	0.00105	0.000626
## 2	0	0.0491	0.0000364	0.00679
## 3	0.0111	0.111	0.00000840	0.00499
## 4	0.152	0.293	0.0000739	0.00466
## 5	0.000267	0.187	0.000129	0.00112
## 6	0.0119	0.102	0.000927	0.0424
##	Brucellaceae	Rhodospirillaceae	id8	Bacteroides
##	<dbl>	<dbl>	<chr>	<dbl>
## 1	0	0.000185	MIHIV138	0.227
## 2	0.0000545	0	MIHIV178	0.468
## 3	0	0.00386	MIHIV255	0.209
## 4	0	0	MIHIV278	0.174
## 5	0	0.000101	MIHIV361	0.145
## 6	0	0.00520	MIHIV404	0.256
##	Prevotella	Alistipes	Blautia	Coprococcus
##	<dbl>	<dbl>	<dbl>	<dbl>
## 1	0.0000264	0.0124	0.0350	0.00876
## 2	0.000291	0.00252	0.0452	0.0205
## 3	0.000126	0.0150	0.0419	0.00238
## 4	0.00155	0.0154	0.0733	0.0205
## 5	0.000147	0.0312	0.0612	0.0279
## 6	0.0000662	0.0110	0.0275	0.000977
##	id9	`P. copri`	`P. stercora`	`P. oris`
##	<chr>	<dbl>	<dbl>	<dbl>
## 1	MIHIV138	0.0000221	0	0
## 2	MIHIV178	0.000188	0.0000659	0.00000942
## 3	MIHIV255	0.0000898	0.0000539	0
## 4	MIHIV278	0.000921	0.000537	0.0000768
## 5	MIHIV361	0.0000935	0.0000374	0.0000187
## 6	MIHIV404	0.0000528	0.0000176	0.0000176
##	`Acidaminococcus`	nintestini`	`Acinetobacter`	njunii`
##		<dbl>		<dbl>
## 1		0		0.000251
## 2		0		0.00107
## 3		0.00000898		0.00444
## 4		0		0.00130
## 5		0		0.0000374
## 6		0.0651		0.00772
##	`Schlegelella`	nthermopolymers`	id10	`Bacteroides`
##		<dbl>	<chr>	<dbl>
## 1		0	MIHIV138	0
## 2		0	MIHIV178	0.104
## 3		0	MIHIV255	0.0351
## 4		0	MIHIV278	0.0291
## 5		0	MIHIV361	0.0000374

```
## 6          0.000123 MIHIV404          0.00799
## `Bacteroides\nthetaiotaomicron` `Bacteroides\ndorei`
##          <dbl>          <dbl>
## 1          0.00584          0.000450
## 2          0.0160          0.000320
## 3          0.00600          0.00668
## 4          0.00883          0.00791
## 5          0.00713          0.000355
## 6          0.0607          0.00827
## `Bacteroides\nacidifaciens` `Barnesiella\nintestinihominis`
##          <dbl>          <dbl>
## 1          0.000111          0.0128
## 2          0.0000942          0.00000942
## 3          0.000108          0.00379
## 4          0          0.0104
## 5          0.00160          0.00000935
## 6          0.0000704          0
## `Blautia\nluti` `Blautia\nglucerasei` `Blautia\nschinkii`
##          <dbl>          <dbl>          <dbl>
## 1          0.0125          0.0171          0.000177
## 2          0.0210          0.000443          0.000122
## 3          0.00568          0.0000180          0
## 4          0.0214          0          0.0139
## 5          0.0335          0.0187          0.00433
## 6          0.000880          0.000862          0.0000176
## `Ruminococcus\nbromii` `Clostridium\nsaccharogumia` ids1 iga_smu
##          <dbl>          <dbl> <chr> <dbl>
## 1          0.00000738          0.00739 138_IgA 24.3
## 2          0.0000188          0      178_IgA 24.9
## 3          0.00217          0.00941 255_IgA 23.2
## 4          0.0625          0.00376 278_IgA 24.7
## 5          0.0124          0.00639 361_IgA 22.3
## 6          0          0.00174 404_IgA 18.8
## ids2  igk_smu ids3  igl_smu
## <chr>  <dbl> <chr>  <dbl>
## 1 138_VK 10.9 138_V 18.1
## 2 178_VK 11.2 178_V 13.6
## 3 255_VK 11.3 255_V 15
## 4 278_VK 15.5 278_V NA
## 5 361_VK 9.22 361_V NA
## 6 404_VK 14.2 404_V NA
```

```
sample_size <- length(clin_micro$pid)
sample_size
```

```
## [1] 32
```

```
# global regular expression test all ids in the clin_micro
idtest <- 1:sample_size
for (i in 1:sample_size) {
  idtest[i] <- (grepl(clin_micro$pid[i], clin_micro$ID[i],
    fixed = T) == T)
}
idtest
```

```
## [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```

ids <- paste("id", as.character(1:10), sep = "")
for (i in ids) {
  idtest[i] <- sum(clin_micro$ID != clin_micro[, colnames(clin_micro) ==
    i])
}
idtest

##
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##      id1 id2 id3 id4 id5 id6 id7 id8 id9 id10
##  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0

for (i in 1:sample_size) {
  idtest[i] <- (grepl(clin_micro$pid[i], clin_micro$ids1[i],
    fixed = T) == F)
}
idtest

##
##  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##
##  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      id1 id2 id3 id4 id5 id6 id7 id8 id9 id10
##  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

for (i in 1:sample_size) {
  idtest[i] <- (grepl(clin_micro$pid[i], clin_micro$ids2[i],
    fixed = T) == T)
}
idtest

##
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
##      id1 id2 id3 id4 id5 id6 id7 id8 id9 id10
##  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0

for (i in 1:sample_size) {
  idtest[i] <- (grepl(clin_micro$pid[i], clin_micro$ids3[i],
    fixed = T) == F)
}
idtest

##
##  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##
##  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      id1 id2 id3 id4 id5 id6 id7 id8 id9 id10
##  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0

## test the ids between rna-seq and clin microbiome data
sum(colnames(cnts_raw)[-c(1:3)] != clin_micro$pid)

## [1] 0

```

```
# The values are relative abundance (%) of total bacteria for
# everything except the species which is % of classified
# species.
```

```
## filter to guarantee cnts larger than 1
cnts_fsym <- cnts_raw[rowSums(as.matrix(cnts_raw[, -c(1:3)])) >=
  (3 * ncol(cnts_raw[, 1:32])), ]
## processing of rnaseq data
cnts <- cnts_fsym %>% dplyr::select(-c(Symbol, Length)) %>% tibble::column_to_rownames("Gene_ID")
cnts <- as.matrix(cnts)
dim(cnts)
```

```
## [1] 22950    32
```

```
pheno <- data.frame(pid = colnames(cnts), txt = as.factor(c(rep("Control",
  13), rep("HIV", 19))))
pheno$txt %<>% relevel("Control")
## create EDaseq set
set <- newSeqExpressionSet(as.matrix(cnts), phenoData = data.frame(condition = as.factor(pheno$txt),
  row.names = colnames(cnts)))
## deseq2 and rlog transfer now using deseq2
register(MulticoreParam(6))
dds <- DESeqDataSetFromMatrix(countData = counts(set), colData = pData(set),
  design = ~condition)
dds <- estimateSizeFactors(dds)
## normalization factors
sizeFactors(dds)
```

```
##      138      178      255      278      361      404      493
## 0.2899319 0.3711273 1.2003938 1.2960354 1.4202606 2.2946684 0.7927622
##      582      708      716      914      947      972      124
## 1.9049499 1.1110458 3.6075976 1.9497161 0.6619582 2.3450261 0.8412778
##      132      154      188      217      286      307      323
## 1.1734326 1.1096548 1.3694388 0.9045053 0.4993808 1.1438402 1.0644692
##      391      428      594      622      648      683      819
## 0.7256593 0.3921047 3.8447311 0.8909101 1.1798187 0.3615224 0.8283123
##      825      839      965      998
## 0.6102792 0.8476626 0.2287530 5.6680177
```

```
## The more the size factors differ, the more residual
## dependence of the variance on the mean will be found in the
## transformed data. rlog is a transformation which can
## perform better in these cases.
rld <- rlog(dds, blind = FALSE, fitType = "parametric")
cnts_rld <- assay(rld)
apply(cnts_rld, 2, function(x) sum(x < 0))
```

```
## 138 178 255 278 361 404 493 582 708 716 914 947 972 124 132 154 188 217
##   3   4   6   5   8   7   5   8   7   5   7   4   7   5   4   4   7   6
## 286 307 323 391 428 594 622 648 683 819 825 839 965 998
##   6   5   5   6   5   7   3   4   5   5   5   6   4   9
```

```
# ## check previous rlg data rlgold <-
# read_excel('/home/guanshim/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/rlog_counts_lin
# apply(as.matrix(rlgold[, -c(1,2)]), 2, function(x)
# sum(is.negative(x)) ) all old rlg data are positive check
```

```

# id
sum(row.names(cnts_rld) != cnts_fsym$Gene_ID)

## [1] 0

sum(colnames(cnts_rld) != clin_micro$pid)

## [1] 0

cnts_rld <- data.frame(cnts_rld) %>% mutate(Symbol = cnts_fsym$Symbol,
  Gene_ID = cnts_fsym$Gene_ID) %>% select(Gene_ID, Symbol,
  everything())
head(cnts_rld)

##           Gene_ID Symbol      X138      X178      X255      X278
## 1 ENSG00000121410.11   A1BG  0.9218465  0.8982751  1.374822  0.7730521
## 2 ENSG00000148584.14   A1CF 10.7326097 10.4811489 10.050048 10.0838572
## 3 ENSG00000175899.14   A2M  8.7590456  9.2356528  8.990610  8.7061008
## 4 ENSG00000245105.2 A2M-AS1  2.1718831  2.4985553  2.528179  2.5572829
## 5 ENSG00000256069.7  A2MP1  3.1709332  2.7190405  3.420555  2.9725446
## 6 ENSG00000128274.15 A4GALT  5.5740961  5.8452511  5.812887  5.3565739
##           X361      X404      X493      X582      X708      X716
## 1  0.8772874  0.9542712  0.8209856  1.137988  0.9174279  0.9597596
## 2 10.1223148 10.3899083 10.1302396 10.222859 10.6826256 10.3355567
## 3  8.9890544  9.4153848  8.8831654  9.117000  9.1304922  9.2201583
## 4  2.2262910  2.5536075  2.3561461  2.777252  2.7037227  2.3533534
## 5  2.4344203  3.1943482  2.7585830  3.510831  2.9818064  3.1012041
## 6  5.5212010  5.8307100  6.0103033  5.928460  5.4522342  5.8318119
##           X914      X947      X972      X124      X132      X154      X188
## 1  0.830085  1.159988  0.9494604  1.523047  1.713261  0.9176396  1.081492
## 2 10.774477 10.691444 10.1328353 10.728724 10.364683 10.4760866  9.927583
## 3  8.877407  8.787015  9.3338557  8.875865  8.852791  8.8135271  9.424338
## 4  2.462705  2.156712  2.3717436  2.536439  2.901440  2.6379854  2.240370
## 5  2.638628  3.651664  2.7779538  3.491803  3.616373  2.9030043  3.756288
## 6  5.096413  6.230956  5.8505945  5.222058  5.191497  5.2614205  6.353300
##           X217      X286      X307      X323      X391      X428      X594
## 1  1.082601  0.8684428  0.9125299  1.641480  1.136863  0.8928487  1.140776
## 2 10.332484 10.9002168 10.8988107 10.321640 10.489439  9.8612794 10.572791
## 3  8.959707  8.6556830  8.5542820  8.753593  8.727761  9.1364957  9.080707
## 4  2.590536  2.6678741  2.4775904  2.244829  2.870334  3.4887474  2.660160
## 5  3.245546  3.4231317  3.6759583  3.643448  3.211979  4.1587073  3.292236
## 6  5.673697  5.7907523  5.1139903  5.812934  5.549143  6.4103179  5.465872
##           X622      X648      X683      X819      X825      X839      X965
## 1  1.086267  1.302032  1.120638  1.225776  0.8478239  0.9645089  0.9483397
## 2 10.243346 10.554936 10.746234 10.450992 10.6781577 10.0359557 10.1247191
## 3  9.211606  8.977503  9.049802  8.891078  9.4709639  9.4833828  9.1624353
## 4  2.311453  2.384131  2.507590  2.543641  2.1798084  2.3304179  2.6619391
## 5  3.391421  3.156013  3.494951  3.728348  3.1059110  4.4432686  3.0702721
## 6  5.781281  5.905446  4.999706  5.829825  5.6728592  5.4911450  6.0455417
##           X998
## 1 1.263315
## 2 9.646036
## 3 9.843177
## 4 2.688318
## 5 3.983223

```



```
## 6 5.929285
rld_pos <- cnts_rld[apply(as.matrix(cnts_rld[, -c(1:2)]), 1,
  function(x) !(any(x <= 0))), ]
dim(cnts_raw)

## [1] 45217    35
dim(rld_pos)

## [1] 22939    34
colnames(rld_pos)[-c(1, 2)] <- clin_micro$ID
write.xlsx(rld_pos, "~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/Vgenes/remapping_allp",
  sheetName = "Positive rld transformed counts")
vgenes_rld <- merge(vgenes, rld_pos, by.x = "Vsymbols", by.y = "Symbol")
kable(vgenes_rld[, c(1, 3)], caption = paste(nrow(vgenes_rld),
  " V genes in the RNAseq data", sep = ""))
```

Table 1: 45 V genes in the RNAseq data

Vsymbols	Type
IGHV1-18	heavy
IGHV1-2	heavy
IGHV1-24	heavy
IGHV1-3	heavy
IGHV3-15	heavy
IGHV3-21	heavy
IGHV3-23	heavy
IGHV3-30	heavy
IGHV3-33	heavy
IGHV3-48	heavy
IGHV3-49	heavy
IGHV3-53	heavy
IGHV3-7	heavy
IGHV3-72	heavy
IGHV3-73	heavy
IGHV3-74	heavy
IGHV4-34	heavy
IGHV4-39	heavy
IGHV4-59	heavy
IGHV7-81	heavy
IGKV1-39	lightK
IGKV1-5	lightK
IGKV1D-39	lightK
IGKV2-28	lightK
IGKV2D-28	lightK
IGKV3-11	lightK
IGKV3-15	lightK
IGKV3-20	lightK
IGKV4-1	lightK
IGLV1-40	lightL
IGLV1-44	lightL
IGLV1-47	lightL
IGLV2-11	lightL
IGLV2-14	lightL

Vsymbols	Type
IGLV2-23	lightL
IGLV2-8	lightL
IGLV3-1	lightL
IGLV3-10	lightL
IGLV3-19	lightL
IGLV3-21	lightL
IGLV3-25	lightL
IGLV3-9	lightL
IGLV4-69	lightL
IGLV5-48	lightL
IGLV6-57	lightL

```
write.xlsx(vgenes_rld, "~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/Vgenes/vgenes_rld.
  sheetName = "Vgenes")
vgenes_rnaseq <- vgenes_rld %>% select(-Num, -Gene_ID, -Type) %>%
  tibble::column_to_rownames("Vsymbols") %>% t(.)
vgenes_rnaseq <- data.frame(vgenes_rnaseq)
vgenes_rnaseq$HIV <- clin_micro$HIV
vgenes_rnaseq$ID <- clin_micro$ID
dim(vgenes_rnaseq)
```

```
## [1] 32 47
```

```
write.xlsx(vgenes_rnaseq, "~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/Vgenes/vgenes_r
  sheetName = "vgenes linear regression")
paste("There are 45 V genes left")
```

```
## [1] "There are 45 V genes left"
```

```
##### miseq data ##### test id
id_matrix <- matrix(NA, sample_size, 3)
for (i in c("ids1", "ids2", "ids3")) {
  k <- which(c("ids1", "ids2", "ids3") %in% i)
  for (j in 1:sample_size) {
    id_matrix[j, k] <- unlist(strsplit(as.character(clin_micro[j,
      colnames(clin_micro) == i]), "_", fixed = T))[1]
  }
}
id_matrix <- as.matrix(id_matrix)
apply(id_matrix, 2, function(x) sum(x != clin_micro$pid))
```

```
## [1] 0 0 0
```

```
miseq <- clin_micro %>% select(ID, HIV, iga_smu, igk_smu, igl_smu)
head(miseq)
```

```
## # A tibble: 6 x 5
##   ID      HIV  iga_smu igk_smu igl_smu
##   <chr>   <chr>   <dbl>   <dbl>   <dbl>
## 1 MIHIV138 no      24.3    10.9    18.1
## 2 MIHIV178 no      24.9    11.2    13.6
## 3 MIHIV255 no      23.2    11.3     15
## 4 MIHIV278 no      24.7    15.5     NA
## 5 MIHIV361 no      22.3     9.22     NA
```

```

## 6 MIHIV404 no      18.8    14.2      NA
write.xlsx(miseq, "~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/Vgenes/miseq.xlsx",
  sheetName = "miseq linear regression")
##### clinical and microbiome covariate #####
Xs <- clin_micro %>% select(-c(ids, "pid", "num", "ids1", "ids2",
  "ids3", "iga_smu", "igk_smu", "igl_smu"))
Xs[Xs$HIV == "yes", 19:62] <- Xs[Xs$HIV == "yes", 19:62]/100
clinical_names <- colnames(Xs)[5:18]
dim(Xs)

## [1] 32 62

# classes of microbiome
phylum <- colnames(Xs)[1:9 + 18]
phylum

## [1] "Actinobacteria" "Bacteroidetes" "Cyanobacteria" "Firmicutes"
## [5] "Fusobacteria" "Proteobacteria" "Spirochaetes" "Tenericutes"
## [9] "Verrucomicrobia"

family <- colnames(Xs)[10:20 + 18]
genus <- colnames(Xs)[21:28 + 18]
species <- colnames(Xs)[29:44 + 18]
write.xlsx(Xs, "~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/Vgenes/clin_microbiome.xlsx",
  sheetName = "clinical and microbiome")

# for(i in 1:n_clinical) { ## number of clinical virable j =
# c(6:18)[i] cliname = base::colnames(clinical_order)[j] ##
# linear regression lin_res_isgs = gene_IFNReg(isgs.05.lin,
# clinical_order[,j], clinical_names[i]) lin_res_genesbeta =
# gene_IFNReg(genesbeta.05.lin, clinical_order[,j],
# clinical_names[i]) ## save data
# write.xlsx(lin_res_isgs$results,
# paste("~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/isgs_",
# cliname, '.xlsx', sep = ''), sheetName= paste('ISGs_',
# cliname, sep = '')) write.xlsx(lin_res_genesbeta$results,
# paste("~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/ifnbeta_",
# cliname, '.xlsx', sep = ''), sheetName= paste('ifnbeta_',
# cliname, sep = '')) ## check basic if(lin_res_isgs$size ==
# lin_res_genesbeta$size){ print('Good') }else stop('sample
# size wrong') if(lin_res_isgs$clinical ==
# lin_res_genesbeta$clinical){ print('Good') }else
# stop('clinical parameter wrong') ## number of genes nisgs =
# nrow(lin_res_isgs$results) ngenesbeta =
# nrow(lin_res_genesbeta$results) ##### summary
# table ##### ### sig prop isgs.sig.no =
# sum(lin_res_isgs$results$FDR <= 0.05) genesbeta.sig.no =
# sum(lin_res_genesbeta$results$FDR <= 0.05) if( (isgs.sig.no
# >= 5) & (genesbeta.sig.no >= 5) ){ prop.test.sig =
# prop.test(x = c(isgs.sig.no, genesbeta.sig.no), n =
# c(nisgs, ngenesbeta), correct = FALSE) sig.p =
# prop.test.sig$p.value }elseif prop.test.sig = prop.test(x =
# c(isgs.sig.no, genesbeta.sig.no), n = c(nisgs, ngenesbeta),
# correct = TRUE) sig.p = prop.test.sig$p.value } }

```

```
# miseq data miseq as the outcome
```