

# Vgenes Mutation Correlation

Guannan Shen

March 19, 2019

## Contents

```
##### Set up workspace
rm(list = ls())
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/DataProcessed/Vgenes"
## not in function
"%nin%" <- Negate("%in%")

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

##### all V genes ##### rnaseq
rnaseq <- read_excel("vgenes_rnaseq.xlsx")
head(rnaseq)

## # A tibble: 6 x 47
##   IGHV1.18 IGHV1.2 IGHV1.24 IGHV1.3 IGHV3.15 IGHV3.21 IGHV3.23 IGHV3.30
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1.13 1.37 2.91 2.50 5.17 1.19 3.38 4.20
## 2 0.856 1.56 3.18 2.64 4.74 1.38 3.61 4.46
## 3 0.729 1.94 3.17 2.55 5.17 1.14 4.14 3.79
## 4 0.848 1.16 3.46 2.09 4.73 1.52 2.84 3.74
## 5 0.832 2.09 3.01 2.39 5.15 0.985 4.26 4.12
## 6 1.11 1.75 3.65 2.34 5.38 1.02 3.89 3.05
##   IGHV3.33 IGHV3.48 IGHV3.49 IGHV3.53 IGHV3.7 IGHV3.72 IGHV3.73 IGHV3.74
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 4.90 1.13 2.10 2.02 6.23 3.12 1.26 3.54
## 2 5.50 1.52 2.20 1.74 6.38 3.17 1.23 3.27
## 3 5.30 1.49 1.87 1.64 6.20 3.87 1.61 2.75
## 4 5.07 0.947 1.75 1.50 5.58 2.40 1.06 2.61
## 5 5.37 1.05 1.62 1.99 6.47 2.80 1.47 3.64
## 6 5.53 1.75 2.16 1.58 6.89 3.40 1.31 3.69
##   IGHV4.34 IGHV4.39 IGHV4.59 IGHV7.81 IGKV1.39 IGKV1.5 IGKV1D.39 IGKV2.28
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
```

```
## 1 1.08 0.906 1.37 4.80 0.796 1.65 1.22 1.79
## 2 1.05 0.881 1.34 4.28 0.773 2.01 1.19 1.52
## 3 1.26 0.750 2.01 4.52 1.01 1.86 1.05 1.94
## 4 1.14 0.742 1.40 4.40 0.886 1.64 1.26 1.66
## 5 1.38 0.854 1.15 4.06 0.641 2.26 1.14 1.54
## 6 0.943 0.936 1.59 4.59 1.14 1.87 1.21 1.74
## IGV2D.28 IGV3.11 IGV3.15 IGV3.20 IGV4.1 IGLV1.40 IGLV1.44 IGLV1.47
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 3.16 1.22 1.60 1.92 5.20 1.03 2.00 1.82
## 2 1.58 1.19 1.54 2.43 4.61 1.00 1.95 0.970
## 3 1.38 1.05 1.11 1.78 5.11 1.31 2.32 0.836
## 4 1.49 1.45 1.21 1.87 4.68 0.867 2.28 1.07
## 5 2.20 1.42 1.37 2.10 5.01 0.972 1.79 0.937
## 6 1.45 1.39 1.40 2.61 5.22 1.18 2.20 0.772
## IGLV2.11 IGLV2.14 IGLV2.23 IGLV2.8 IGLV3.1 IGLV3.10 IGLV3.19 IGLV3.21
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 1.52 2.87 1.01 1.41 3.42 2.64 1.04 3.35
## 2 1.49 2.22 0.989 2.10 3.86 2.77 1.02 2.74
## 3 1.44 2.69 1.54 1.64 3.41 3.10 0.870 3.22
## 4 1.42 2.65 1.08 1.53 3.82 2.44 0.990 3.05
## 5 1.68 2.22 0.839 1.40 3.31 3.30 0.851 2.72
## 6 1.53 2.36 0.960 1.29 4.02 3.51 1.12 3.06
## IGLV3.25 IGLV3.9 IGLV4.69 IGLV5.48 IGLV6.57 HIV ID
## <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr>
## 1 1.32 1.63 0.744 1.30 6.20 no MIHIV138
## 2 1.28 1.59 1.15 1.04 5.95 no MIHIV178
## 3 1.80 1.74 0.608 1.14 6.69 no MIHIV255
## 4 1.45 1.81 0.601 1.40 5.92 no MIHIV278
## 5 1.59 1.78 1.16 1.29 6.60 no MIHIV361
## 6 1.54 1.60 0.641 1.21 6.50 no MIHIV404
```

```
dim(rnaseq)
```

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

```
# rnaseq[rnaseq$HIV == 'no', 1:45] rnaseq[rnaseq$HIV ==
# 'yes', 1:45] miseq
miseq <- read_excel("miseq.xlsx")
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
```

```
# miseq[miseq$HIV == 'no', 3:5] miseq[miseq$HIV == 'yes',
# 3:5] Xs
```

```
Clin <- read_excel("clin_microbiome.xlsx")
```

```
# covariates vgenes
```

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

```
## # A tibble: 6 x 36
##   Vsymbols   Num Type Gene_ID      MIHIV138 MIHIV178 MIHIV255
##   <chr>     <dbl> <chr> <chr>      <dbl>      <dbl>      <dbl>
## 1 IGHV1-18     1 heavy ENSG000000211945.2    1.13    0.856    0.729
## 2 IGHV1-2      2 heavy ENSG000000211934.3    1.37    1.56    1.94
## 3 IGHV1-24     3 heavy ENSG000000211950.2    2.91    3.18    3.17
## 4 IGHV1-3      4 heavy ENSG000000211935.3    2.50    2.64    2.55
## 5 IGHV3-15    19 heavy ENSG000000211943.2    5.17    4.74    5.17
## 6 IGHV3-21    22 heavy ENSG000000211947.2    1.19    1.38    1.14
##   MIHIV278 MIHIV361 MIHIV404 MIHIV493 MIHIV582 MIHIV708 MIHIV716 MIHIV914
##   <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1  0.848    0.832     1.11     1.33     0.685    0.875     1.01     0.954
## 2  1.16     2.09      1.75     1.39     1.38     1.44     1.42     1.38
## 3  3.46     3.01      3.65     3.86     3.69     3.46     3.34     2.72
## 4  2.09     2.39      2.34     3.19     2.55     2.91     3.09     2.48
## 5  4.73     5.15      5.38     5.75     4.79     5.34     5.92     4.72
## 6  1.52     0.985     1.02     1.50     1.23     1.15     1.89     1.22
##   MIHIV947 MIHIV972 MIHIV124 MIHIV132 MIHIV154 MIHIV188 MIHIV217 MIHIV286
##   <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1  0.794    0.909     0.767     1.44     1.21     1.05     0.913     0.824
## 2  1.43     1.33      1.22     2.26     1.31     1.88     1.83     1.67
## 3  3.27     2.95      2.94     3.97     4.36     3.27     3.10     3.05
## 4  3.38     2.39      2.21     2.36     2.49     2.30     3.57     3.30
## 5  5.15     5.11      4.88     6.19     6.35     5.63     5.33     5.71
## 6  1.26     1.48      1.35     1.26     1.02     1.11     1.57     1.12
##   MIHIV307 MIHIV323 MIHIV391 MIHIV428 MIHIV594 MIHIV622 MIHIV648 MIHIV683
##   <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1  0.870    1.41      0.954     0.850     0.750     1.18     1.58     0.858
## 2  1.17     2.26      1.56     1.91     1.21     1.62     1.17     1.34
## 3  3.77     4.66      3.55     4.12     3.71     3.27     5.02     3.03
## 4  3.15     4.49      2.99     2.79     2.73     3.43     3.99     3.50
## 5  5.64     6.55      5.55     5.79     4.62     5.98     6.45     6.64
## 6  1.48     2.50      1.24     1.37     1.01     1.57     1.93     1.16
##   MIHIV819 MIHIV825 MIHIV839 MIHIV965 MIHIV998
##   <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1  0.929    0.802     2.28     0.908     1.29
## 2  1.64     1.27      2.63     1.66     1.35
## 3  3.05     3.21      4.00     3.19     3.56
## 4  3.36     3.09      3.92     3.36     3.05
## 5  5.27     4.89      6.68     6.27     6.31
## 6  1.21     1.10      2.52     1.22     1.33
```

```
dim(vgenes)
```

```
## [1] 45 36
```

```
head(Clin[, 1:4])
```

```
## # A tibble: 6 x 4
##   ID      Gender Age HIV
##   <chr>   <chr>  <dbl> <chr>
## 1 MIHIV138 M      29 no
## 2 MIHIV178 M      33 no
## 3 MIHIV255 M      34 no
## 4 MIHIV278 M      23 no
```

```

## 5 MIHIV361 F          33 no
## 6 MIHIV404 F          29 no

##### linear regression ##### equal
##### length of outcomes and covariates
seq_Reg <- function(gene_matrix, Clin, clin_var_name) {
  # get names ready
  genelistname = base::colnames(gene_matrix)
  ## number of gene to test, also the number of multiple test
  n_gene = ncol(gene_matrix)
  clinical_variable = as.matrix(Clin[, colnames(Clin) == clin_var_name])
  ## outcome lm
  outcome_lm = lapply(1:n_gene, function(i) {
    lm = lm(gene_matrix[, i] ~ clinical_variable + Clin$Age +
      Clin$Gender)
    coef = summary(lm)$coefficients[2, ]
    return(coef)
  })
  outcome_lm = data.frame(matrix(unlist(outcome_lm), ncol = 4,
    byrow = TRUE, dimnames = list(c(colnames(gene_matrix)),
      c("Estimate", "Std.Error", "t.statistic", "p.value"))))

  # adjusted p-value
  outcome_lm = outcome_lm %>% # https://cran.r-project.org/web/packages/mutoss/mutoss.pdf
  # The Benjamini-Liu (BL) step-down procedure
  dplyr::mutate(FDR = if (n_gene > 15) {
    p.adjust(p.value, "BH", n_gene)
  } else if (n_gene <= 10) {
    p.value
  } else {
    BL(p.value, alpha = 0.05, silent = T)$adjPValues
  }, names = colnames(gene_matrix)) %>% dplyr::mutate(Estimate = round(Estimate,
    10), Std.Error = round(Std.Error, 10), t.statistic = round(t.statistic,
    4)) %>% mutate(sig = ifelse(FDR < 0.05, "Sig.", "Non")) %>%
  select(names, everything())
  # sort by p.value
  outcome_lm = outcome_lm[order(outcome_lm$p.value), ]

  ## sample size
  size = sum(!is.na(clinical_variable))

  ## summary table
  return(list(results = data.frame(outcome_lm), size = size,
    clinical = clin_var_name))
}

## hiv or not wrapper
hivornot <- function(gene_matrix, Clin, clin_var_name, HIV) {
  if (missing(HIV)) {
    gene_matrix = gene_matrix %>% select(-HIV, -ID)
    gene_matrix = as.matrix(gene_matrix)

    return(list(seq_Reg(gene_matrix, Clin, clin_var_name),
      HIV = "All Participants"))
  }
}

```

```

} else if (HIV == "no") {
  gene_matrix = gene_matrix %>% filter(HIV == "no") %>%
    select(-HIV, -ID)
  gene_matrix = as.matrix(gene_matrix)

  Clin = Clin %>% filter(HIV == "no") %>% select(-HIV,
    -ID)

  return(list(seq_Reg(gene_matrix, Clin, clin_var_name),
    HIV = "Healthy Controls"))
} else if (HIV == "yes") {
  gene_matrix = gene_matrix %>% filter(HIV == "yes") %>%
    select(-HIV, -ID)
  gene_matrix = as.matrix(gene_matrix)

  Clin = Clin %>% filter(HIV == "yes") %>% select(-HIV,
    -ID)

  return(list(seq_Reg(gene_matrix, Clin, clin_var_name),
    HIV = "HIV Infected"))
}
}

## output summary and plot
output_sum <- function(output, genelist) {
  # genelist is the name of the genelist genelist: 'RNAseq' or
  # 'MiSeq' a list of two sublist
  clinical = output[[1]]$clinical
  # find a good file name
  para_name = paste(unlist(strsplit(clinical, " ", fixed = T))[1:3],
    collapse = "_")
  para_name1 = gsub("/", "_", para_name, fixed = T)
  size = output[[1]]$size
  hivstatus = output[[2]]
  input = output[[1]]$results
  # lm results
  sig = sum(input$sig == "Sig.")
  if (sig > 0) {
    write.xlsx(input, paste("~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/DataProcessed/Vgenes/",
      sep = "", Sys.Date(), "_", para_name1, "_", genelist,
      hivstatus, "_", ".xlsx"), sheetName = paste(hivstatus,
      sep = "_", genelist))
    ##### summary table #####
    genes = input$names[input$sig == "Sig."]
    fdr = format.pval(input$FDR[input$sig == "Sig."], digits = 4,
      eps = 1e-04, scientific = F)
    p = format.pval(input$p.value[input$sig == "Sig."], digits = 4,
      eps = 1e-04, scientific = F)
    # neg_log10p = -log10(input$p.value[input$sig == 'Sig.'])
    slope = format(input$Estimate[input$sig == "Sig."], digits = 4,
      scientific = T)
    sum_a = cbind(clinical, hivstatus, genes, fdr, p, slope,
      size)
  }
}

```

```

colnames(sum_a) = c("Parameter", "HIV Status", "Gene",
  "Adjusted P", "Raw P", "Slope", "Sample Size")
##### plot #####
if (genelist == "RNAseq") {
  input1 = input %>% mutate(Chains = vgenes$Type, log10p = ifelse(Estimate >
    0, -log10(p.value), log10(p.value)))
  p = ggplot(input1, aes(Chains, log10p, color = sig)) +
    ylab(expression(paste("-log"[10], "p"))) + theme_minimal() +
    coord_flip() + geom_hline(yintercept = 0) + ggtitle(paste(clinical,
    hivstatus)) + geom_point(aes(shape = Chains),
    position = "jitter", size = 3)
  ggsave(paste(Sys.Date(), "ScatterPlot_RNAseq_", para_name1,
    "_", hivstatus, "_", sep = ""), width = 6, height = 9,
    device = "tiff", path = "~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/Reports/plots",
    dpi = 300)
} else if (genelist == "MiSeq") {
  input2 = input %>% mutate(Directionality = ifelse(Estimate >
    0, "Positive Correlation", "Negative Correlation"))
  p = ggplot(input2, aes(names, -log10(p.value), fill = Directionality)) +
    geom_bar(stat = "identity", width = 0.4) + scale_x_discrete(labels = c("lgA",
    "lgK", expression(paste("lg", sep = "", lambda)))) +
    xlab("Single Mutations per Unique Sequence") +
    ylab(expression(paste("-log"[10], "p"))) + geom_hline(yintercept = -log10(0.05)) +
    annotate(geom = "text", x = 0, y = -log10(0.045),
    label = "p = 0.05") + ggtitle(paste(clinical,
    hivstatus)) + guides(fill = guide_legend(title = NULL)) +
    theme_minimal()
  ggsave(paste(Sys.Date(), "BarPlot_MiSeq_", para_name1,
    "_", hivstatus, "_", sep = ""), width = 6, height = 9,
    device = "tiff", path = "~/Documents/gitlab/Cario_RNASeq_Microbiom_Inte/Reports/plots",
    dpi = 300)
} else {
}
## return the summary table #####
return(sum_a)
} else {
}
}

```

```

# rnaseq clinical parameters
clinical_names <- colnames(Clin)[5:18]
clinical_names

```

```

## [1] "CD4 count (cells/ul)"
## [2] "IL-6 (pg/ml)"
## [3] "CRP (mg/ml)"
## [4] "iFABP (pg/ml)"
## [5] "sCD27 (U/ml)"
## [6] "CD14 (ng/ml)"
## [7] "LPS (pg/ml)"
## [8] "LTA\n(OD)"
## [9] "CD38+ HLA-DR+ CD4 T cells (% of CD4 T cells)"
## [10] "CD38+ HLA-DR+ CD8 T cells (% of CD8 T cells)"

```

```
## [11] "CD4 T cells (% viable)"
## [12] "CD8 T cells (% viable)"
## [13] "Tissue HIV RNA (per CD4 T cell)"
## [14] "Plasma VL"

n_clinical <- length(clinical_names)
##### rna seq ##### both
clinical_sum <- matrix(NA, 0, 7)
for (i in clinical_names[1:12]) {
  output = hivornot(rnaseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq")
```

Table 1: Vgenes RNAseq

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
CD4 T cells (% viable)	All Participants	IGKV4.1	0.03217	0.0007148	-2.563e-02	32
CD4 T cells (% viable)	All Participants	IGLV1.44	0.04966	0.0022069	-1.793e-02	32
CD8 T cells (% viable)	All Participants	IGKV4.1	0.03568	0.0007929	1.901e-02	32

```
## hiv yes
clinical_sum <- matrix(NA, 0, 7)
for (i in clinical_names) {
  output = hivornot(rnaseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq HIV Infected")
```

Table 2: Vgenes RNAseq HIV Infected

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Plasma VL	HIV Infected	IGLV2.8	0.04342	0.000965	4.445e-06	19

```
## hiv no
clinical_sum <- matrix(NA, 0, 7)
for (i in clinical_names[1:12]) {
  output = hivornot(rnaseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq Health")
```

Table 3: Vgenes RNAseq Health

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope
CD38+ HLA-DR+ CD4 T cells (% of CD4 T cells)	Healthy Controls	IGHV7.81	0.02084	0.000463	6.496e-01

```
##### miseq ##### both
clinical_sum <- matrix(NA, 0, 7)
for (i in clinical_names[1:12]) {
  output = hivornot(miseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq")
```

Table 4: Vgenes MiSeq

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
CD4 count (cells/ul)	All Participants	iga_smu	0.03359	0.03359	4.928e-03	32
CD4 count (cells/ul)	All Participants	igl_smu	0.03720	0.03720	6.823e-03	32
sCD27 (U/ml)	All Participants	iga_smu	0.001858	0.001858	-9.638e-02	32
CD14 (ng/ml)	All Participants	igl_smu	0.01366	0.01366	-4.469e-03	30
LPS (pg/ml)	All Participants	iga_smu	0.01513	0.01513	-2.757e-01	30
CD38+ HLA-DR+ CD4 T cells (% of CD4 T cells)	All Participants	iga_smu	0.02222	0.02222	-5.492e-01	32
CD38+ HLA-DR+ CD8 T cells (% of CD8 T cells)	All Participants	iga_smu	0.0433	0.0433	-1.044e-01	32
CD4 T cells (% viable)	All Participants	iga_smu	0.001763	0.001763	1.453e-01	32
CD8 T cells (% viable)	All Participants	igl_smu	0.000425	0.000425	-1.511e-01	32
CD8 T cells (% viable)	All Participants	iga_smu	0.001005	0.001005	-1.131e-01	32

```
## yes
clinical_sum <- matrix(NA, 0, 7)
for (i in clinical_names) {
  output = hivornot(miseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq HIV Infected")
```

Table 5: Vgenes MiSeq HIV Infected

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
CD8 T cells (% viable)	HIV Infected	igl_smu	0.04793	0.04793	-1.402e-01	19

```
## no
clinical_sum <- matrix(NA, 0, 7)
for (i in clinical_names[1:12]) {
  output = hivornot(miseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "MiSeq")
}
```



```

    clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq Health")

```

Table 6: Vgenes MiSeq Health

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
CD4 T cells (% viable)	Healthy Controls	igk_smu	0.02005	0.02005	-1.972e-01	13

```

# unlist(strsplit(clinical, ' ', fixed = T))[1:3]

```

```

# classes of microbiome
phylum <- colnames(Clin)[1:9 + 18]
family <- colnames(Clin)[10:20 + 18]
genus <- colnames(Clin)[21:28 + 18]
species <- colnames(Clin)[29:44 + 18]

##### rna seq ##### both phylum
clinical_sum <- matrix(NA, 0, 7)
for (i in phylum) {
  output = hivornot(rnaseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq phylum")

```

Table: Vgenes RNAseq phylum

```

## family
clinical_sum <- matrix(NA, 0, 7)
for (i in family) {
  output = hivornot(rnaseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq family")

```

Table 7: Vgenes RNAseq family

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Prevotellaceae	All Participants	IGKV1.39	0.01111	0.0002468	8.729e-01	27
Brucellaceae	All Participants	IGLV3.19	0.007625	0.0001694	5.264e+02	27
Brucellaceae	All Participants	IGLV1.40	0.043832	0.0019481	3.825e+02	27

```

## genus
clinical_sum <- matrix(NA, 0, 7)
for (i in genus) {
  output = hivornot(rnaseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "RNAseq")
}

```

```

clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq genus")

```

Table 8: Vgenes RNAseq genus

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Prevotella	All Participants	IGKV1.39	0.01966	0.0004368	8.722e-01	27

```

## species
clinical_sum <- matrix(NA, 0, 7)
for (i in species) {
  output = hivornot(rnaseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq species")

```

Table 9: Vgenes RNAseq species

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
P. stercorea	All Participants	IGKV1.39	0.01746	0.0006984	3.173e+00	27
P. stercorea	All Participants	IGHV3.30	0.01746	0.0010978	9.537e+00	27
P. stercorea	All Participants	IGLV3.21	0.01746	0.0011640	5.723e+00	27
P. oris	All Participants	IGHV3.73	0.008818	0.000196	8.854e+00	27

```

##### yes
clinical_sum <- matrix(NA, 0, 7)
for (i in phylum) {
  output = hivornot(rnaseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq phylum")

```

Table 10: Vgenes RNAseq phylum

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Tenericutes	HIV Infected	IGLV3.19	0.005291	0.0001176	1.144e+02	14
Tenericutes	HIV Infected	IGLV2.14	0.017094	0.0007597	1.826e+02	14

```

## family
clinical_sum <- matrix(NA, 0, 7)
for (i in family) {
  output = hivornot(rnaseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "RNAseq")
}

```

```

    clinical_sum = rbind(clinical_sum, sum_a)
  }
kable(clinical_sum, caption = "Vgenes RNAseq family")

```

Table 11: Vgenes RNAseq family

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Christensenellaceae	HIV Infected	IGHV3.21	0.03087	0.0008225	1.060e+02	14
Christensenellaceae	HIV Infected	IGHV4.34	0.03087	0.0013718	3.598e+01	14
Christensenellaceae	HIV Infected	IGHV1.18	0.03251	0.0021671	8.254e+01	14
Christensenellaceae	HIV Infected	IGLV3.9	0.03867	0.0034371	1.155e+02	14
Christensenellaceae	HIV Infected	IGHV1.2	0.04791	0.0056043	8.742e+01	14
Christensenellaceae	HIV Infected	IGHV3.74	0.04791	0.0070997	1.072e+02	14
Christensenellaceae	HIV Infected	IGLV3.25	0.04791	0.0076494	4.774e+01	14
Christensenellaceae	HIV Infected	IGHV3.23	0.04791	0.0085175	1.051e+02	14
Brucellaceae	HIV Infected	IGLV3.19	0.008736	0.0001941	5.286e+02	14

```

## genus
clinical_sum <- matrix(NA, 0, 7)
for (i in genus) {
  output = hivornot(rnaseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq genus")

```

Table: Vgenes RNAseq genus

```

## species
clinical_sum <- matrix(NA, 0, 7)
for (i in species) {
  output = hivornot(rnaseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq species")

```

Table: Vgenes RNAseq species

```

##### no
clinical_sum <- matrix(NA, 0, 7)
for (i in phylum) {
  output = hivornot(rnaseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq phylum")

```

Table 12: Vgenes RNAseq phylum

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Bacteroidetes	Healthy Controls	IGKV3.20	0.04994	0.00111	2.528e+00	13
Firmicutes	Healthy Controls	IGKV3.20	0.01173	0.0002608	-1.702e+00	13
Verrucomicrobia	Healthy Controls	IGLV3.1	0.02686	0.0005968	1.39e+01	13

```
## family
clinical_sum <- matrix(NA, 0, 7)
for (i in family) {
  output = hivornot(rnaseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq family")
```

Table 13: Vgenes RNAseq family

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Ruminococcaceae	Healthy Controls	IGHV3.7	0.02533	0.0005628	-4.687e+00	13

```
## genus
clinical_sum <- matrix(NA, 0, 7)
for (i in genus) {
  output = hivornot(rnaseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq genus")
```

Table: Vgenes RNAseq genus

```
## species
clinical_sum <- matrix(NA, 0, 7)
for (i in species) {
  output = hivornot(rnaseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "RNAseq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq species")
```

Table 14: Vgenes RNAseq species

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
P. oris	Healthy Controls	IGHV3.73	0.004916	0.0001092	9.376e+00	13
P. oris	Healthy Controls	IGLV2.23	0.045937	0.0020417	6.885e+00	13

```
##### miseq #####

clinical_sum <- matrix(NA, 0, 7)
for (i in phylum) {
  output = hivornot(miseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes miseq phylum")
```

Table 15: Vgenes miseq phylum

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Cyanobacteria	All Participants	iga_smu	0.00970	0.00970	1.568e+02	27
Cyanobacteria	All Participants	igl_smu	0.01856	0.01856	1.580e+02	27
Firmicutes	All Participants	iga_smu	0.008763	0.008763	8.954e+00	27
Proteobacteria	All Participants	iga_smu	0.03297	0.03297	-8.501e+00	27

```
## family
clinical_sum <- matrix(NA, 0, 7)
for (i in family) {
  output = hivornot(miseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes miseq family")
```

Table 16: Vgenes miseq family

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Lachnospiraceae	All Participants	iga_smu	0.04007	0.04007	1.007e+01	27

```
## genus
clinical_sum <- matrix(NA, 0, 7)
for (i in genus) {
  output = hivornot(miseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq genus")
```

Table 17: Vgenes RNAseq genus

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Blautia	All Participants	iga_smu	0.03105	0.03105	7.152e+01	27
Blautia	All Participants	igl_smu	0.03470	0.03470	1.236e+02	27

```
## species
clinical_sum <- matrix(NA, 0, 7)
for (i in species) {
  output = hivornot(miseq, Clin, i)
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes RNAseq species")
```

Table 18: Vgenes RNAseq species

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Bacteroides dorei All Particip	ants iga_smu 0.	02683	0.02683 1	.283e+02	27	
Bacteroides dorei All Particip	ants igl_smu 0.	03370	0.03370 1	.368e+02	27	
Bacteroides dorei All Particip	ants igk_smu 0.	04955	0.04955 9	.830e+01	27	
Blautia luti All Part	icipants iga_smu	0.03813	0.03813	1.417e+0	2 27	
Blautia luti All Part	icipants igl_smu	0.03847	0.03847	2.128e+0	2 27	
Blautia schinkii All Part	icipants iga_smu	0.001614	0.001614	3.064e+0	2 27	
Ruminococcus bromii All Participa	nts iga_smu 0.0	1074	0.01074 6.	471e+01 2	7	

```
##### yes
clinical_sum <- matrix(NA, 0, 7)
for (i in phylum) {
  output = hivornot(miseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq phylum")
```

Table 19: Vgenes MiSeq phylum

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Tenericutes	HIV Infected	igl_smu	0.01974	0.01974	-6.59e+02	14

```
## family
clinical_sum <- matrix(NA, 0, 7)
for (i in family) {
  output = hivornot(miseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
```

```
kable(clinical_sum, caption = "Vgenes MiSeq family")
```

Table 20: Vgenes MiSeq family

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Rhodospirillaceae	HIV Infected	igl_smu	0.03986	0.03986	-2.048e+04	14

```
## genus
clinical_sum <- matrix(NA, 0, 7)
for (i in genus) {
  output = hivornot(miseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq genus")
```

Table 21: Vgenes MiSeq genus

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Coproccoccus	HIV Infected	igk_smu	0.004191	0.004191	-3.492e+02	14
Coproccoccus	HIV Infected	iga_smu	0.041544	0.041544	-3.157e+02	14
Thalassospira	HIV Infected	igl_smu	0.02448	0.02448	-2.737e+05	14

```
## species
clinical_sum <- matrix(NA, 0, 7)
for (i in species) {
  output = hivornot(miseq, Clin, i, "yes")
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq species")
```

Table 22: Vgenes MiSeq species

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Acidaminococcus						
intestini HIV Infected i	gk_smu	0.0072	74 0.0	07274 -2.08	6e+02	14
Blautia						
glucerasei HIV Infe	cted iga_smu	0.01136	0.01136	-1.046e+	03	14
Blautia						
glucerasei HIV Infe	cted igl_smu	0.03358	0.03358	-8.168e+	02	14

```
##### no
clinical_sum <- matrix(NA, 0, 7)
for (i in phylum) {
  output = hivornot(miseq, Clin, i, "no")
  # summary table
```

```

sum_a = output_sum(output, "MiSeq")
clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq phylum")

```

Table 23: Vgenes MiSeq phylum

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Bacteroidetes	Healthy Controls	iga_smu	0.03153	0.03153	-1.272e+01	13
Cyanobacteria	Healthy Controls	iga_smu	0.01004	0.01004	1.189e+02	13
Firmicutes	Healthy Controls	iga_smu	0.04295	0.04295	7.761e+00	13
Proteobacteria	Healthy Controls	iga_smu	0.01878	0.01878	-2.186e+01	13
Tenericutes	Healthy Controls	iga_smu	0.02769	0.02769	1.692e+02	13

```

## family
clinical_sum <- matrix(NA, 0, 7)
for (i in family) {
  output = hivornot(miseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq family")

```

Table 24: Vgenes MiSeq family

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Lachnospiraceae	Healthy Controls	igk_smu	0.04028	0.04028	-1.149e+01	13
Christensenellaceae	Healthy Controls	igk_smu	0.022	0.022	3.691e+01	13
Ruminococcaceae	Healthy Controls	igl_smu	0.03989	0.03989	5.568e+01	13
Moraxellaceae	Healthy Controls	iga_smu	0.04008	0.04008	-1.314e+02	13

```

## genus
clinical_sum <- matrix(NA, 0, 7)
for (i in genus) {
  output = hivornot(miseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq genus")

```

Table 25: Vgenes MiSeq genus

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
Coprococcus	Healthy Controls	iga_smu	0.03209	0.03209	1.653e+02	13
Acinetobacter	Healthy Controls	iga_smu	0.04353	0.04353	-1.394e+02	13



```
## species
clinical_sum <- matrix(NA, 0, 7)
for (i in species) {
  output = hivornot(miseq, Clin, i, "no")
  # summary table
  sum_a = output_sum(output, "MiSeq")
  clinical_sum = rbind(clinical_sum, sum_a)
}
kable(clinical_sum, caption = "Vgenes MiSeq species")
```

Table 26: Vgenes MiSeq species

Parameter	HIV Status	Gene	Adjusted P	Raw P	Slope	Sample Size
P. stercorea	Healthy Controls	iga_smu	0.04835	0.04835	2.89e+02	13
Schlegelella						
thermodepolymerans	Healthy Contr	ols iga_smu 0.0	3001	0.03001 -4	.341e+04	13
Bacteroides						
thetaitaomicron	Healthy Cont	rols iga_smu 0.	007971	0.007971 -	1.028e+02	13
Bacteroides						
acidifaciens	Healthy Cont	rols igk_smu 0.	02178	0.02178 -	3.409e+03	13
Blautia						
glucerasei	Healthy	Controls igk_smu	0.04666	0.04666	-1.933e+	02 13
Blautia						
schinkii	Healthy	Controls iga_smu	0.001257	0.001257	2.28e+02	13
Ruminococcus						
bromii	Healthy Contr	ols iga_smu 0.0	2005	0.02005 4.	574e+01	13