Bladder cohort - R Notebook

Adam Cankaya
acankaya2017@fau.edu
Based on tutorial found at https://www.costalab.org/wp-content/uploads/2020/11/R_class_D3.htm

First install BiocManager, edgeR, and TCGAbiolinks

```
if (!require("BiocManager", quietly = TRUE))
  install.packages("BiocManager")

BiocManager::install("edgeR")
```

```
## Bioconductor version 3.19 (BiocManager 1.30.23), R 4.4.0 (2024-04-24 ucrt)

## Warning: package(s) not installed when version(s) same as or greater than current; use
##   'force = TRUE' to re-install: 'edgeR'

## Installation paths not writeable, unable to update packages
##   path: C:/Program Files/R/R-4.4.0/library
##   packages:
##     KernSmooth, nlme, survival

## Old packages: 'evaluate', 'leaps'
```

```
BiocManager::install("TCGAbiolinks")
```

```
## Bioconductor version 3.19 (BiocManager 1.30.23), R 4.4.0 (2024-04-24 ucrt)

## Warning: package(s) not installed when version(s) same as or greater than current; use
##   'force = TRUE' to re-install: 'TCGAbiolinks'

## Installation paths not writeable, unable to update packages
##   path: C:/Program Files/R/R-4.4.0/library
##   packages:
##     KernSmooth, nlme, survival
## Old packages: 'evaluate', 'leaps'
```

```
BiocManager::install("genefilter")
```

```
## Bioconductor version 3.19 (BiocManager 1.30.23), R 4.4.0 (2024-04-24 ucrt)
```

```
## Warning: package(s) not installed when version(s) same as or greater than current; use
##   'force = TRUE' to re-install: 'genefilter'

## Installation paths not writeable, unable to update packages
##   path: C:/Program Files/R/R-4.4.0/library
##   packages:
##     KernSmooth, nlme, survival
## Old packages: 'evaluate', 'leaps'
```

Step 1 - Load packages, download data from TCGA, and prepare it for DEGList

```
library("TCGAbiolinks")
library("limma")
library("edgeR")
library("glmnet")
```

```
## Loading required package: Matrix

## Loaded glmnet 4.1-8
```

```
library("factoextra")
```

```
## Loading required package: ggplot2

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library("FactoMineR")
library("caret")
```

```
## Loading required package: lattice
```

```
library("SummarizedExperiment")
```

```
## Loading required package: MatrixGenerics

## Loading required package: matrixStats

##
## Attaching package: 'MatrixGenerics'

## The following objects are masked from 'package:matrixStats':
##
##     colAlls, colAnyNAs, colAnys, colAvgsPerRowSet, colCollapse,
##     colCounts, colCummaxs, colCummins, colCumprods, colCumsums,
##     colDiffs, colIQRDiffs, colIQRs, colLogSumExps, colMadDiffs,
##     colMads, colMaxs, colMeans2, colMedians, colMins, colOrderStats,
##     colProds, colQuantiles, colRanges, colRanks, colSdDiffs, colSds,
```

```
##     colSums2, colTabulates, colVarDiffs, colVars, colWeightedMads,
##     colWeightedMeans, colWeightedMedians, colWeightedSds,
##     colWeightedVars, rowAlls, rowAnyNAs, rowAnys, rowAvgsPerColSet,
##     rowCollapse, rowCounts, rowCummaxs, rowCummins, rowCumprods,
##     rowCumsums, rowDiffs, rowIQRDiffs, rowIQRs, rowLogSumExps,
##     rowMadDiffs, rowMads, rowMaxs, rowMeans2, rowMedians, rowMins,
##     rowOrderStats, rowProds, rowQuantiles, rowRanges, rowRanks,
##     rowSdDiffs, rowSds, rowSums2, rowTabulates, rowVarDiffs, rowVars,
##     rowWeightedMads, rowWeightedMeans, rowWeightedMedians,
##     rowWeightedSds, rowWeightedVars

## Loading required package: GenomicRanges

## Loading required package: stats4

## Loading required package: BiocGenerics

##
## Attaching package: 'BiocGenerics'

## The following object is masked from 'package:limma':
##
##     plotMA

## The following objects are masked from 'package:stats':
##
##     IQR, mad, sd, var, xtabs

## The following objects are masked from 'package:base':
##
##     anyDuplicated, aperm, append, as.data.frame, basename, cbind,
##     colnames, dirname, do.call, duplicated, eval, evalq, Filter, Find,
##     get, grep, grepl, intersect, is.unsorted, lapply, Map, mapply,
##     match, mget, order, paste, pmax, pmax.int, pmin, pmin.int,
##     Position, rank, rbind, Reduce, rownames, sapply, setdiff, table,
##     tapply, union, unique, unsplit, which.max, which.min

## Loading required package: S4Vectors

##
## Attaching package: 'S4Vectors'

## The following objects are masked from 'package:Matrix':
##
##     expand, unname

## The following object is masked from 'package:utils':
##
##     findMatches
```

```
## The following objects are masked from 'package:base':
##
##     expand.grid, I, unname

## Loading required package: IRanges

##
## Attaching package: 'IRanges'

## The following object is masked from 'package:grDevices':
##
##     windows

## Loading required package: GenomeInfoDb

## Loading required package: Biobase

## Welcome to Bioconductor
##
##     Vignettes contain introductory material; view with
##     'browseVignettes()'. To cite Bioconductor, see
##     'citation("Biobase")', and for packages 'citation("pkgname")'.

##
## Attaching package: 'Biobase'

## The following object is masked from 'package:MatrixGenerics':
##
##     rowMedians

## The following objects are masked from 'package:matrixStats':
##
##     anyMissing, rowMedians
```

```
library("gplots")
```

```
##
## Attaching package: 'gplots'

## The following object is masked from 'package:IRanges':
##
##     space

## The following object is masked from 'package:S4Vectors':
##
##     space

## The following object is masked from 'package:stats':
##
##     lowess
```

```r
library("survival")
```

```
##
## Attaching package: 'survival'

## The following object is masked from 'package:caret':
##
##     cluster
```

```r
library("survminer")
```

```
## Loading required package: ggpubr

##
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':
##
##     myeloma
```

```r
library("RColorBrewer")
library("gProfileR")
library("genefilter")
```

```
##
## Attaching package: 'genefilter'

## The following objects are masked from 'package:MatrixGenerics':
##
##     rowSds, rowVars

## The following objects are masked from 'package:matrixStats':
##
##     rowSds, rowVars
```

```r
setwd('C:/Adam/R/')  # make sure it already exists

# Before we perform a GDC query let's look at the TCGA-BLCA data
# As of June 2024 we should see a case count of 412
TCGAbiolinks:::getProjectSummary("TCGA-BLCA")
```

```
## $file_count
## [1] 23394
##
## $data_categories
##    file_count case_count                data_category
## 1        6729        412  Simple Nucleotide Variation
## 2        4285        412             Sequencing Reads
## 3        1760        412                  Biospecimen
```

```
## 4         994        412                     Clinical
## 5        4478        412        Copy Number Variation
## 6        1736        412      Transcriptome Profiling
## 7        1320        412              DNA Methylation
## 8         343        343           Proteome Profiling
## 9          26         12 Somatic Structural Variation
## 10       1723        406         Structural Variation
##
## $case_count
## [1] 412
##
## $file_size
## [1] 4.082979e+14
```

```r
# Download TCGA-BLCA data from GDC
# We want the complete RNA sequencing and raw gene count data
# So we run a query of the Transcriptome Profiling category and RNA-Seq experimental type
# We use the STAR - Counts workflow type because it contains the raw gene counts we need
# We ignore other sample types besides tumor and normal
# The original paper by Wang uses the HTSeq-counts workflow, but this is a legacy version of
#  the new STAR - COUNTS workflow type
query_TCGA = GDCquery(
  project = "TCGA-BLCA",
  data.category = "Transcriptome Profiling",
  data.type="Gene Expression Quantification",
  experimental.strategy = "RNA-Seq",
  workflow.type = "STAR - Counts",
  sample.type = c("Primary Tumor", "Solid Tissue Normal"))
```

```
## ------------------------------------

## o GDCquery: Searching in GDC database

## ------------------------------------

## Genome of reference: hg38

## --------------------------------------------

## oo Accessing GDC. This might take a while...

## --------------------------------------------

## ooo Project: TCGA-BLCA

## --------------------

## oo Filtering results

## --------------------
```

```
## ooo By experimental.strategy

## ooo By data.type

## ooo By workflow.type

## ooo By sample.type

## -----------------

## oo Checking data

## -----------------

## ooo Checking if there are duplicated cases

## ooo Checking if there are results for the query

## --------------------

## o Preparing output

## --------------------
```

```r
# Run the query and format it as a table
# The results are a table with 431 rows (because some patients have multiple cases each)
# There are 29 columns with meta data about each case such as sample_type (tumor vs normal)
lihc_res = getResults(query_TCGA)

# We can create a summary table shows there are 412 tumor and 19 normal (412+19=431)
summary(factor(lihc_res$sample_type))
```

```
##       Primary Tumor Solid Tissue Normal
##                 412                  19
```

```r
# Go ahead and download all the data from GDC to our working directory
GDCdownload(query = query_TCGA)
```

```
## Downloading data for project TCGA-BLCA

## Of the 431 files for download 431 already exist.

## All samples have been already downloaded
```

```r
# Now load the RNA-Seq data from the files into R workspace
tcga_data = GDCprepare(query_TCGA)
```

```
## |                                                  |   0%                    |
```

```
## Starting to add information to samples

##  => Add clinical information to samples

##  => Adding TCGA molecular information from marker papers

##  => Information will have prefix 'paper_'

## blca subtype information from:doi:10.1016/j.cell.2017.09.007

## Available assays in SummarizedExperiment :
##   => unstranded
##   => stranded_first
##   => stranded_second
##   => tpm_unstrand
##   => fpkm_unstrand
##   => fpkm_uq_unstrand
```

```
# This data object has 60660 rows and 431 columns
# This indicates there are 60660 different genes found throughout all the cases
# The object contains both clincal and expression data
dim(tcga_data)
```

```
## [1] 60660   431
```

```
# We can access the data in the object like this which verifies 412 tumor and 19 normal
table(tcga_data@colData$definition)
```

```
##
## Primary solid Tumor Solid Tissue Normal
##                 412                  19
```

```
# Or see the gender data of 117 female and 314 male
table(tcga_data@colData$gender)
```

```
##
## female   male
##    117    314
```

```
# let's look at the various names of the first 6 genes...
head(rowData(tcga_data))
```

```
## DataFrame with 6 rows and 10 columns
##                       source     type     score     phase            gene_id
##                     <factor> <factor> <numeric> <integer>        <character>
## ENSG00000000003.15   HAVANA     gene        NA        NA ENSG00000000003.15
## ENSG00000000005.6    HAVANA     gene        NA        NA  ENSG00000000005.6
## ENSG00000000419.13   HAVANA     gene        NA        NA ENSG00000000419.13
## ENSG00000000457.14   HAVANA     gene        NA        NA ENSG00000000457.14
## ENSG00000000460.17   HAVANA     gene        NA        NA ENSG00000000460.17
```

```
## ENSG00000000938.13   HAVANA     gene        NA        NA ENSG00000000938.13
##                         gene_type   gene_name       level     hgnc_id
##                       <character> <character> <character> <character>
## ENSG00000000003.15 protein_coding      TSPAN6           2  HGNC:11858
## ENSG00000000005.6  protein_coding        TNMD           2  HGNC:17757
## ENSG00000000419.13 protein_coding        DPM1           2   HGNC:3005
## ENSG00000000457.14 protein_coding       SCYL3           2  HGNC:19285
## ENSG00000000460.17 protein_coding    C1orf112           2  HGNC:25565
## ENSG00000000938.13 protein_coding         FGR           2   HGNC:3697
##                             havana_gene
##                             <character>
## ENSG00000000003.15 OTTHUMG00000022002.2
## ENSG00000000005.6  OTTHUMG00000022001.2
## ENSG00000000419.13 OTTHUMG00000032742.2
## ENSG00000000457.14 OTTHUMG00000035941.6
## ENSG00000000460.17 OTTHUMG00000035821.9
## ENSG00000000938.13 OTTHUMG00000003516.3
```

```
# To preview the raw gene counts let's look at the expression levels of the first
#  6 genes in the first 3 cases...
rownames = values(tcga_data)$gene_name[1:6]
firs6genes = head(assay(tcga_data)[,1:3])
rownames(firs6genes) = rownames
colnames(firs6genes) = c("Case 1", "Case 2", "Case 3")
firs6genes
```

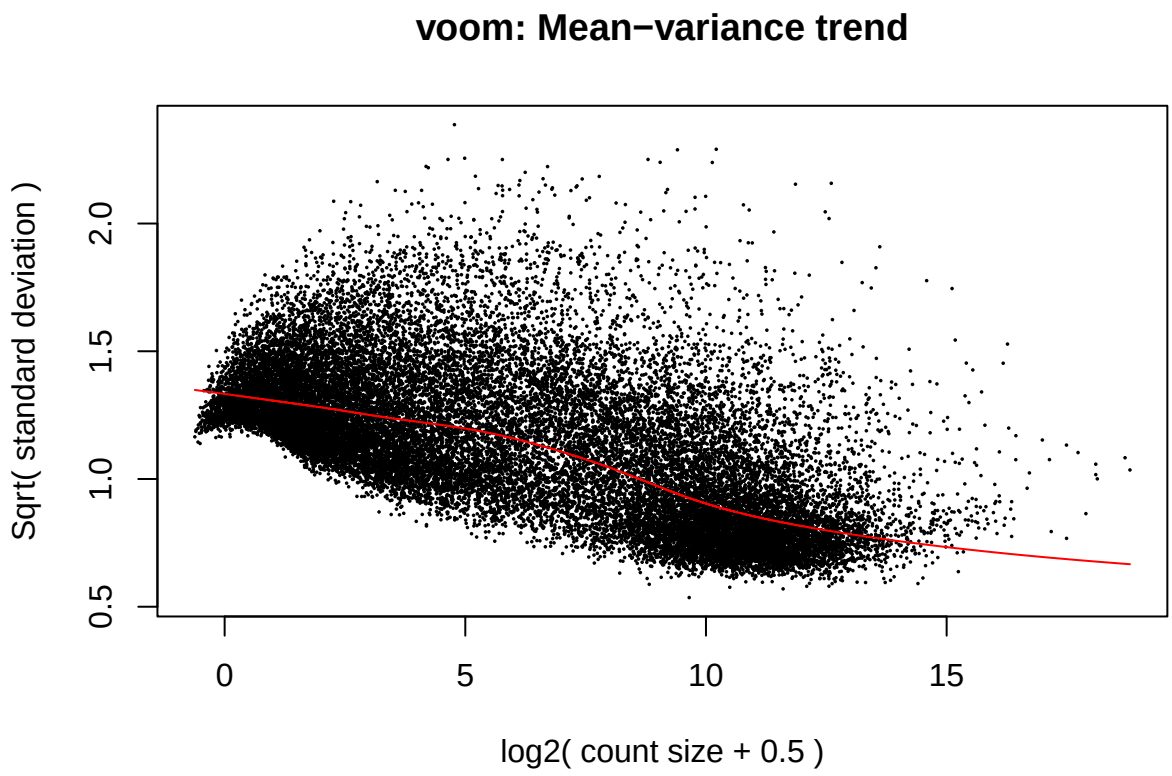
```
##          Case 1 Case 2 Case 3
## TSPAN6     3679  28986    951
## TNMD          0     21      1
## DPM1       4190   2917   2976
## SCYL3       850   1910    705
## C1orf112   1196   1495    655
## FGR         353    905   2282
```

Step 2 - Generate DEGList, filter low counts, and normalize data

```
# Before we can perform DEG analysis we need to normalize the data
# Let's create a limma pipeline to do this...
# The pipeline function will take in three input parameters:
#   tcga_data - the data object we created in Step 1
#   condition_variable - the variable by which we will group patients (tumor vs normal)
#   reference_group - indicates which of the condition variable
#    values is the reference group (no tumors)
# The pipeline will return a list of three objects:
#   voom - the TMM normalized data returned by running voom
#   eBayes - the fitted model returned by running eBayes
#   topTable - a simple table which contains the top 1000 differentially expressed genes
#    sorted by p.value
limma_pipeline = function(
  tcga_data,
  condition_variable,
```

```
reference_group=NULL){

# Create a design matrix
# The factor is the category classifier for the data (tumor vs normal)
#  limma requires it to be a factor object
design_factor = colData(tcga_data)[, condition_variable, drop=T] # definition
group = factor(design_factor) # Solid Normal Tissue

# otherwise just pick the first class as the reference class
if (!is.null(reference_group)) {
  group = relevel(group, ref=reference_group)
}

# make the design matrix
design = model.matrix(~ group)

# generate the DGEList object using the input...
#  counts is the raw gene counts (numericla matrix - rows as genes, columns as cases)
#  samples is the clinical data (data frame)
#  genes is the annotation information (data frame - gene id and names)
# the DGEList object returned is a transformed version of tcga_data
dge = DGEList(counts=assay(tcga_data),
              samples=colData(tcga_data),
              genes=as.data.frame(rowData(tcga_data)))

# filtering - by default genes with less than 10 counts per million reads are removed
# after filtering we have 28087 genes remaining
# no need to filter further by logfc or adjusted p-value because all
#  entries already meet the cutoff criteria
keep = filterByExpr(dge,design) # genes which meet are left after filtering
dge = dge[keep,,keep.lib.sizes=FALSE] # filter the DGEList object, only keep the genes we want
rm(keep) # remove this object from memory because we are done with it

# TODO do we need rpkm() filtering?

# Normalization (TMM followed by voom)
# normalizing - minimize batch effects and variation with the TMM normalization
# TMM - trimmed mean of M-values
# use the voom method to convert the data to have a similar variance as arrays
#  (TODO what is this?)
dge = calcNormFactors(dge)
v = voom(dge, design, plot=TRUE)

# Fit model to data given design
#  fits a series of linear models, one to each probe
#  then pass it to eBayes to rank the differential expression
fit = lmFit(v, design)
fit = eBayes(fit)

# Show top genes
topGenes = topTable(fit, coef=ncol(design), number=1000, sort.by="p")
print(topGenes[10])
```

```
  return(
    list(
      voomObj=v, # normalized data
      fit=fit, # fitted model and statistics
      topGenes=topGenes # the 1000 most differentially expressed genes
    )
  )
}

# TODO only run the pipeline if we didn't already run it before and save the data to a local file
#  tcga_data = readRDS(file = "tcga_data.RDS")
#  saveRDS(object = tcga_data,
#    file = "tcga_data.RDS",
#    compress = FALSE)

# Run the pipeline on the tcga_data from step 1 and normal tissue as the reference
#  "definition" is the column name for the tissue type (tumor vs normal)
#  "Solid Tissue Normal" is our baseline/control/reference class value
# The limma_res object returned is a list of 3 objects - voomObj, fit, topGenes
limma_res = limma_pipeline(
  tcga_data=tcga_data,
  condition_variable="definition",
  reference_group="Solid Tissue Normal"
)
```



```
##                            havana_gene
## ENSG00000164530.15  OTTHUMG00000014611.4
## ENSG00000168079.17  OTTHUMG00000132172.4
## ENSG00000163815.6   OTTHUMG00000133087.3
## ENSG00000153446.16  OTTHUMG00000159314.4
## ENSG00000196616.14  OTTHUMG00000161413.4
## ENSG00000168309.18  OTTHUMG00000159159.5
## ENSG00000108924.14  OTTHUMG00000177840.3
## ENSG00000018625.15  OTTHUMG00000024080.4
## ENSG00000224958.6   OTTHUMG00000019962.6
## ENSG00000197766.8   OTTHUMG00000181840.6
## ENSG00000126218.12  OTTHUMG00000017374.8
## ENSG00000168477.19 OTTHUMG00000031088.12
## ENSG00000068976.14  OTTHUMG00000066835.3
## ENSG00000123560.14  OTTHUMG00000022111.6
## ENSG00000034971.17  OTTHUMG00000034789.4
## ENSG00000241158.7   OTTHUMG00000158723.8
## ENSG00000168497.5   OTTHUMG00000154309.3
## ENSG00000004776.13  OTTHUMG00000048122.5
## ENSG00000077943.8   OTTHUMG00000017733.2
## ENSG00000154330.13  OTTHUMG00000019966.5
## ENSG00000106809.11  OTTHUMG00000020224.4
## ENSG00000119147.10  OTTHUMG00000130921.4
## ENSG00000181856.15  OTTHUMG00000102181.6
## ENSG00000232855.7   OTTHUMG00000078747.6
## ENSG00000144218.19 OTTHUMG00000153011.13
## ENSG00000108018.15  OTTHUMG00000019018.5
## ENSG00000167281.19 OTTHUMG00000150183.10
## ENSG00000141052.18  OTTHUMG00000058767.6
## ENSG00000119508.18  OTTHUMG00000021030.3
## ENSG00000101605.13  OTTHUMG00000178209.7
## ENSG00000171368.12  OTTHUMG00000131011.5
## ENSG00000163145.13  OTTHUMG00000097095.5
## ENSG00000179915.24 OTTHUMG00000129263.23
## ENSG00000125851.10  OTTHUMG00000031941.6
## ENSG00000112936.19  OTTHUMG00000150340.4
## ENSG00000136546.16  OTTHUMG00000154078.6
## ENSG00000182253.15  OTTHUMG00000171887.6
## ENSG00000205221.12  OTTHUMG00000152149.3
## ENSG00000179388.9   OTTHUMG00000097825.3
## ENSG00000123358.20  OTTHUMG00000150393.8
## ENSG00000189129.14  OTTHUMG00000018596.3
## ENSG00000118526.7   OTTHUMG00000015608.2
## ENSG00000123243.15  OTTHUMG00000017635.7
## ENSG00000225398.3   OTTHUMG00000047819.3
## ENSG00000268926.3   OTTHUMG00000187223.2
## ENSG00000149294.17  OTTHUMG00000167196.8
## ENSG00000172403.11  OTTHUMG00000161165.7
## ENSG00000127528.6   OTTHUMG00000182330.1
## ENSG00000172348.15  OTTHUMG00000014782.2
## ENSG00000004799.8   OTTHUMG00000153977.3
## ENSG00000206579.9   OTTHUMG00000164288.3
## ENSG00000172260.15  OTTHUMG00000009698.6
## ENSG00000181072.11  OTTHUMG00000155658.3
```

```
## ENSG00000231943.9   OTTHUMG00000047830.7
## ENSG00000065325.13 OTTHUMG00000130269.12
## ENSG00000186642.16  OTTHUMG00000102045.6
## ENSG00000111452.13  OTTHUMG00000168339.5
## ENSG00000268388.6   OTTHUMG00000183870.4
## ENSG00000181234.9   OTTHUMG00000163736.4
## ENSG00000173175.15  OTTHUMG00000159517.6
## ENSG00000163431.13  OTTHUMG00000035802.3
## ENSG00000156218.13  OTTHUMG00000147363.4
## ENSG00000176533.13  OTTHUMG00000180435.3
## ENSG00000103241.7   OTTHUMG00000137651.5
## ENSG00000140538.16 OTTHUMG00000148677.12
## ENSG00000141338.14  OTTHUMG00000180192.5
## ENSG00000121671.12  OTTHUMG00000153225.5
## ENSG00000280429.1   OTTHUMG00000177388.1
## ENSG00000149090.12  OTTHUMG00000166328.4
## ENSG00000070193.5   OTTHUMG00000131153.5
## ENSG00000144655.15  OTTHUMG00000131293.3
## ENSG00000118407.15  OTTHUMG00000015056.4
## ENSG00000149451.18  OTTHUMG00000031758.4
## ENSG00000254510.2   OTTHUMG00000166924.2
## ENSG00000147588.7   OTTHUMG00000164600.2
## ENSG00000166091.21  OTTHUMG00000028751.9
## ENSG00000108405.4   OTTHUMG00000177673.2
## ENSG00000154175.18  OTTHUMG00000159094.7
## ENSG00000135472.9   OTTHUMG00000169808.5
## ENSG00000153234.15 OTTHUMG00000131950.10
## ENSG00000267505.1   OTTHUMG00000179859.1
## ENSG00000174576.10  OTTHUMG00000167045.2
## ENSG00000100307.13  OTTHUMG00000150418.4
## ENSG00000198932.13  OTTHUMG00000022061.6
## ENSG00000154721.15  OTTHUMG00000078441.4
## ENSG00000077157.22  OTTHUMG00000041393.7
## ENSG00000153823.19  OTTHUMG00000133191.5
## ENSG00000022267.19 OTTHUMG00000022504.13
## ENSG00000154734.16  OTTHUMG00000078688.5
## ENSG00000059915.17  OTTHUMG00000018954.5
## ENSG00000151892.15  OTTHUMG00000019097.4
## ENSG00000143171.13  OTTHUMG00000034626.5
## ENSG00000132840.10  OTTHUMG00000108158.5
## ENSG00000133392.18  OTTHUMG00000129935.9
## ENSG00000138356.14  OTTHUMG00000154536.6
## ENSG00000164736.6   OTTHUMG00000164377.3
## ENSG00000241684.6   OTTHUMG00000158725.3
## ENSG00000108381.11  OTTHUMG00000090655.5
## ENSG00000188729.6   OTTHUMG00000156190.2
## ENSG00000179796.12  OTTHUMG00000130572.8
## ENSG00000106034.18  OTTHUMG00000156982.6
## ENSG00000101938.15  OTTHUMG00000022199.4
## ENSG00000170271.11  OTTHUMG00000164141.3
## ENSG00000258274.1   OTTHUMG00000170323.1
## ENSG00000130176.8   OTTHUMG00000182033.3
## ENSG00000148339.12  OTTHUMG00000020736.2
## ENSG00000162706.13  OTTHUMG00000037177.3
```

```
## ENSG00000136842.14  OTTHUMG00000020325.3
## ENSG00000145936.9   OTTHUMG00000130439.4
## ENSG00000118729.12  OTTHUMG00000011970.1
## ENSG00000122121.11  OTTHUMG00000022373.2
## ENSG00000109906.14  OTTHUMG00000168243.4
## ENSG00000100146.18  OTTHUMG00000149913.7
## ENSG00000163710.9   OTTHUMG00000159312.4
## ENSG00000122367.20  OTTHUMG00000018655.7
## ENSG00000112183.15  OTTHUMG00000014306.5
## ENSG00000172987.13  OTTHUMG00000018880.3
## ENSG00000229619.4   OTTHUMG00000186506.2
## ENSG00000196666.6   OTTHUMG00000167871.2
## ENSG00000102683.8   OTTHUMG00000016563.1
## ENSG00000197561.7   OTTHUMG00000181839.6
## ENSG00000076555.15  OTTHUMG00000169250.5
## ENSG00000267405.1   OTTHUMG00000179860.1
## ENSG00000273348.1   OTTHUMG00000186239.1
## ENSG00000159251.8   OTTHUMG00000129675.6
## ENSG00000162761.14  OTTHUMG00000034575.3
## ENSG00000164616.16  OTTHUMG00000155971.5
## ENSG00000128849.11  OTTHUMG00000166485.3
## ENSG00000172935.9   OTTHUMG00000167897.2
## ENSG00000100628.12  OTTHUMG00000171298.3
## ENSG00000187479.8   OTTHUMG00000166555.7
## ENSG00000153902.14  OTTHUMG00000044558.5
## ENSG00000253105.6   OTTHUMG00000164715.3
## ENSG00000078596.11  OTTHUMG00000021900.2
## ENSG00000141404.17  OTTHUMG00000131660.5
## ENSG00000132326.12  OTTHUMG00000152884.4
## ENSG00000184347.15  OTTHUMG00000130409.4
## ENSG00000124440.16  OTTHUMG00000141296.6
## ENSG00000131831.18  OTTHUMG00000021209.6
## ENSG00000188783.6   OTTHUMG00000035911.2
## ENSG00000128591.16  OTTHUMG00000023627.3
## ENSG00000072163.20  OTTHUMG00000131529.6
## ENSG00000082175.15  OTTHUMG00000167531.4
## ENSG00000127241.18  OTTHUMG00000156461.4
## ENSG00000229019.1   OTTHUMG00000019965.1
## ENSG00000163827.13  OTTHUMG00000133479.3
## ENSG00000138944.8   OTTHUMG00000030991.3
## ENSG00000198838.14  OTTHUMG00000172253.8
## ENSG00000152583.12  OTTHUMG00000130605.5
## ENSG00000095637.22 OTTHUMG00000018812.10
## ENSG00000158445.10  OTTHUMG00000033051.8
## ENSG00000131471.7   OTTHUMG00000180656.3
## ENSG00000115556.14  OTTHUMG00000154743.4
## ENSG00000165124.18  OTTHUMG00000020482.3
## ENSG00000103710.11  OTTHUMG00000133115.4
## ENSG00000187955.12  OTTHUMG00000149877.5
## ENSG00000130558.20  OTTHUMG00000020897.4
## ENSG00000162951.11  OTTHUMG00000130021.9
## ENSG00000108823.16  OTTHUMG00000162024.3
## ENSG00000162407.9   OTTHUMG00000008160.3
## ENSG00000112320.12  OTTHUMG00000015312.4
```

```
## ENSG00000174403.16  OTTHUMG00000048002.3
## ENSG00000165966.16  OTTHUMG00000169364.4
## ENSG00000147724.12  OTTHUMG00000149864.6
## ENSG00000183963.19  OTTHUMG00000151203.8
## ENSG00000151623.15  OTTHUMG00000161455.3
## ENSG00000125740.14  OTTHUMG00000182122.3
## ENSG00000237125.11  OTTHUMG00000160783.9
## ENSG00000196557.13  OTTHUMG00000172992.5
## ENSG00000112658.8   OTTHUMG00000014722.3
## ENSG00000161281.11  OTTHUMG00000048144.7
## ENSG00000168490.14  OTTHUMG00000163776.4
## ENSG00000183036.11  OTTHUMG00000086731.4
## ENSG00000177363.5   OTTHUMG00000167581.2
## ENSG00000165072.10  OTTHUMG00000019990.2
## ENSG00000147408.14  OTTHUMG00000130827.5
## ENSG00000079308.19 OTTHUMG00000133056.10
## ENSG00000161649.13  OTTHUMG00000181796.2
## ENSG00000080854.16  OTTHUMG00000167124.4
## ENSG00000152580.8   OTTHUMG00000159856.1
## ENSG00000111077.18 OTTHUMG00000169730.13
## ENSG00000159176.14  OTTHUMG00000035773.6
## ENSG00000150764.14  OTTHUMG00000166912.4
## ENSG00000165379.13  OTTHUMG00000140261.4
## ENSG00000178878.13  OTTHUMG00000153561.2
## ENSG00000164949.8   OTTHUMG00000164392.5
## ENSG00000255248.9  OTTHUMG00000166035.32
## ENSG00000180875.5   OTTHUMG00000039909.2
## ENSG00000134917.10  OTTHUMG00000165656.3
## ENSG00000127951.8   OTTHUMG00000130681.3
## ENSG00000165633.13  OTTHUMG00000018184.4
## ENSG00000170153.11  OTTHUMG00000161380.8
## ENSG00000181778.5   OTTHUMG00000019967.1
## ENSG00000145362.20 OTTHUMG00000132912.13
## ENSG00000154556.18 OTTHUMG00000157215.10
## ENSG00000143196.5   OTTHUMG00000034554.3
## ENSG00000185739.13  OTTHUMG00000177617.1
## ENSG00000197380.11  OTTHUMG00000153070.4
## ENSG00000162882.15  OTTHUMG00000152348.6
## ENSG00000261292.2   OTTHUMG00000175728.2
## ENSG00000171819.5   OTTHUMG00000002002.4
## ENSG00000120129.6   OTTHUMG00000130523.4
## ENSG00000106018.14  OTTHUMG00000151446.4
## ENSG00000239474.7   OTTHUMG00000132205.3
## ENSG00000090006.18 OTTHUMG00000182635.12
## ENSG00000175084.12  OTTHUMG00000058924.4
## ENSG00000165410.15  OTTHUMG00000029536.9
## ENSG00000111404.7   OTTHUMG00000168820.2
## ENSG00000163017.14  OTTHUMG00000129813.6
## ENSG00000120738.8   OTTHUMG00000129197.3
## ENSG00000103175.11  OTTHUMG00000137641.5
## ENSG00000165197.5   OTTHUMG00000021175.2
## ENSG00000249669.10 OTTHUMG00000163464.18
## ENSG00000198624.13  OTTHUMG00000130127.3
## ENSG00000109846.9  OTTHUMG00000166885.12
```

```
## ENSG00000231246.2   OTTHUMG00000012183.5
## ENSG00000206052.11  OTTHUMG00000178688.2
## ENSG00000134201.12  OTTHUMG00000011644.5
## ENSG00000138685.17  OTTHUMG00000039506.5
## ENSG00000048540.15  OTTHUMG00000168837.5
## ENSG00000069535.14  OTTHUMG00000021388.5
## ENSG00000171385.10  OTTHUMG00000011989.1
## ENSG00000129244.9   OTTHUMG00000178018.3
## ENSG00000166482.12  OTTHUMG00000059585.6
## ENSG00000170419.11  OTTHUMG00000129271.3
## ENSG00000269186.3   OTTHUMG00000183863.4
## ENSG00000106772.18  OTTHUMG00000020047.5
## ENSG00000131018.24 OTTHUMG00000015841.14
## ENSG00000072952.19  OTTHUMG00000165773.7
## ENSG00000182916.8   OTTHUMG00000022096.2
## ENSG00000164330.17  OTTHUMG00000130304.5
## ENSG00000240654.6   OTTHUMG00000016576.2
## ENSG00000228877.2   OTTHUMG00000020886.2
## ENSG00000101955.15  OTTHUMG00000021362.2
## ENSG00000240583.13  OTTHUMG00000023944.6
## ENSG00000139910.20  OTTHUMG00000029385.9
## ENSG00000282556.3   OTTHUMG00000191408.4
## ENSG00000188778.6   OTTHUMG00000164028.5
## ENSG00000184905.9   OTTHUMG00000022048.4
## ENSG00000150672.17  OTTHUMG00000134309.9
## ENSG00000175785.13  OTTHUMG00000141313.3
## ENSG00000185274.12  OTTHUMG00000129783.5
## ENSG00000154678.18 OTTHUMG00000023836.10
## ENSG00000067840.12  OTTHUMG00000024209.4
## ENSG00000115252.18  OTTHUMG00000132596.2
## ENSG00000277494.2   OTTHUMG00000164950.4
## ENSG00000068615.19  OTTHUMG00000130205.9
## ENSG00000036448.10  OTTHUMG00000129175.8
## ENSG00000108231.14  OTTHUMG00000018777.8
## ENSG00000075073.15  OTTHUMG00000018377.2
## ENSG00000166509.12  OTTHUMG00000137620.5
## ENSG00000116604.18  OTTHUMG00000033095.9
## ENSG00000101333.17  OTTHUMG00000031853.4
## ENSG00000112562.19  OTTHUMG00000016050.5
## ENSG00000095303.17  OTTHUMG00000020605.5
## ENSG00000132837.15  OTTHUMG00000108159.5
## ENSG00000156113.24 OTTHUMG00000018543.33
## ENSG00000163531.15  OTTHUMG00000151697.3
## ENSG00000279484.2   OTTHUMG00000152904.1
## ENSG00000112425.16  OTTHUMG00000015747.9
## ENSG00000198947.16  OTTHUMG00000021336.7
## ENSG00000104332.12  OTTHUMG00000164074.3
## ENSG00000143772.9   OTTHUMG00000037582.1
## ENSG00000141448.11  OTTHUMG00000131767.4
## ENSG00000119138.4   OTTHUMG00000019991.2
## ENSG00000121440.15  OTTHUMG00000158865.6
## ENSG00000133800.9   OTTHUMG00000165713.2
## ENSG00000214575.9   OTTHUMG00000172877.6
## ENSG00000165424.7   OTTHUMG00000018561.2
```

```
## ENSG00000137726.17  OTTHUMG00000166991.5
## ENSG00000166831.9   OTTHUMG00000172423.3
## ENSG00000187824.9   OTTHUMG00000178043.4
## ENSG00000128016.7   OTTHUMG00000182970.4
## ENSG00000061455.11  OTTHUMG00000150469.4
## ENSG00000148180.21  OTTHUMG00000020584.8
## ENSG00000164107.9   OTTHUMG00000160775.3
## ENSG00000135298.14  OTTHUMG00000014982.4
## ENSG00000171873.8   OTTHUMG00000031779.5
## ENSG00000179094.16  OTTHUMG00000178317.5
## ENSG00000258537.5   OTTHUMG00000171076.1
## ENSG00000198523.6   OTTHUMG00000015462.2
## ENSG00000183580.10  OTTHUMG00000161798.2
## ENSG00000248174.6   OTTHUMG00000160745.7
## ENSG00000065534.19 OTTHUMG00000141304.13
## ENSG00000138735.16  OTTHUMG00000132971.7
## ENSG00000244953.1   OTTHUMG00000166556.1
## ENSG00000277631.5   OTTHUMG00000019423.7
## ENSG00000187942.12  OTTHUMG00000002675.2
## ENSG00000173641.18  OTTHUMG00000009531.2
## ENSG00000116741.8   OTTHUMG00000035600.3
## ENSG00000044524.11  OTTHUMG00000159040.2
## ENSG00000144891.18  OTTHUMG00000159503.1
## ENSG00000249082.2   OTTHUMG00000163077.2
## ENSG00000155970.12  OTTHUMG00000096965.4
## ENSG00000078804.13  OTTHUMG00000032310.4
## ENSG00000167641.11  OTTHUMG00000181892.2
## ENSG00000213088.12  OTTHUMG00000037182.8
## ENSG00000164056.11  OTTHUMG00000133071.4
## ENSG00000205037.2   OTTHUMG00000175541.1
## ENSG00000130700.7   OTTHUMG00000032919.3
## ENSG00000164122.9   OTTHUMG00000160793.4
## ENSG00000132554.20  OTTHUMG00000164802.7
## ENSG00000122420.10  OTTHUMG00000009644.7
## ENSG00000280323.1   OTTHUMG00000150115.1
## ENSG00000050555.19  OTTHUMG00000020819.5
## ENSG00000237179.6   OTTHUMG00000152839.4
## ENSG00000160801.14  OTTHUMG00000133515.6
## ENSG00000143878.10  OTTHUMG00000090755.4
## ENSG00000182575.8   OTTHUMG00000161491.2
## ENSG00000082293.13  OTTHUMG00000014987.5
## ENSG00000087258.16 OTTHUMG00000133241.11
## ENSG00000148357.16  OTTHUMG00000140096.5
## ENSG00000163681.16  OTTHUMG00000133764.5
## ENSG00000268505.1   OTTHUMG00000183862.1
## ENSG00000162630.6   OTTHUMG00000035687.2
## ENSG00000197256.11  OTTHUMG00000180799.4
## ENSG00000137727.12  OTTHUMG00000166590.2
## ENSG00000174059.17  OTTHUMG00000036565.3
## ENSG00000198597.9   OTTHUMG00000182112.9
## ENSG00000170345.10  OTTHUMG00000171774.4
## ENSG00000275088.1   OTTHUMG00000187070.1
## ENSG00000168306.13  OTTHUMG00000159154.4
## ENSG00000232057.1   OTTHUMG00000151406.1
```

```
## ENSG00000249797.2   OTTHUMG00000162888.6
## ENSG00000230393.1   OTTHUMG00000153210.1
## ENSG00000196569.13  OTTHUMG00000015545.2
## ENSG00000166086.13  OTTHUMG00000167130.7
## ENSG00000166292.12  OTTHUMG00000177858.5
## ENSG00000188215.10  OTTHUMG00000131553.4
## ENSG00000163380.16  OTTHUMG00000158774.1
## ENSG00000197361.8   OTTHUMG00000132905.6
## ENSG00000048740.18  OTTHUMG00000017668.8
## ENSG00000138615.6   OTTHUMG00000133140.3
## ENSG00000180440.4   OTTHUMG00000016735.3
## ENSG00000130037.5   OTTHUMG00000168232.3
## ENSG00000142871.18  OTTHUMG00000010577.2
## ENSG00000223382.5  OTTHUMG00000003749.16
## ENSG00000187239.17  OTTHUMG00000020800.3
## ENSG00000146151.14  OTTHUMG00000014902.6
## ENSG00000105270.15  OTTHUMG00000181747.4
## ENSG00000135333.14  OTTHUMG00000015228.2
## ENSG00000116194.13  OTTHUMG00000035075.5
## ENSG00000187922.14  OTTHUMG00000150428.4
## ENSG00000068831.19 OTTHUMG00000045420.38
## ENSG00000134245.18  OTTHUMG00000011157.3
## ENSG00000162614.19  OTTHUMG00000040533.4
## ENSG00000112964.14  OTTHUMG00000094791.3
## ENSG00000125868.16 OTTHUMG00000031947.10
## ENSG00000163873.10  OTTHUMG00000004189.2
## ENSG00000071991.9   OTTHUMG00000132802.4
## ENSG00000134769.21  OTTHUMG00000132309.6
## ENSG00000150995.20 OTTHUMG00000154996.16
## ENSG00000145147.20  OTTHUMG00000128551.5
## ENSG00000136826.15  OTTHUMG00000020449.4
## ENSG00000126785.13  OTTHUMG00000140342.6
## ENSG00000139946.10  OTTHUMG00000152336.4
## ENSG00000189056.14  OTTHUMG00000157247.2
## ENSG00000174944.9   OTTHUMG00000159859.4
## ENSG00000184205.14  OTTHUMG00000021597.4
## ENSG00000047648.23  OTTHUMG00000021134.7
## ENSG00000135063.19  OTTHUMG00000019979.4
## ENSG00000274624.1   OTTHUMG00000187113.1
## ENSG00000184113.9   OTTHUMG00000150441.5
## ENSG00000148411.8   OTTHUMG00000020921.3
## ENSG00000164035.10  OTTHUMG00000131051.7
## ENSG00000183801.8   OTTHUMG00000165527.3
## ENSG00000185345.23 OTTHUMG00000015970.14
## ENSG00000104490.18  OTTHUMG00000164876.8
## ENSG00000250007.7   OTTHUMG00000172256.3
## ENSG00000112175.8   OTTHUMG00000014903.2
## ENSG00000107796.13  OTTHUMG00000018700.2
## ENSG00000151320.11  OTTHUMG00000140207.6
## ENSG00000198961.10  OTTHUMG00000128750.3
## ENSG00000101335.10  OTTHUMG00000032387.4
## ENSG00000162989.5   OTTHUMG00000131937.5
## ENSG00000287281.1   OTTHUMG00000194850.1
## ENSG00000248309.7   OTTHUMG00000162630.6
```

```
## ENSG00000055118.16  OTTHUMG00000158341.5
## ENSG00000266524.3   OTTHUMG00000188319.2
## ENSG00000165349.12  OTTHUMG00000021780.2
## ENSG00000166501.14  OTTHUMG00000131615.9
## ENSG00000158887.19  OTTHUMG00000034341.6
## ENSG00000164188.9   OTTHUMG00000131110.3
## ENSG00000187151.7   OTTHUMG00000167318.2
## ENSG00000068024.17  OTTHUMG00000133344.7
## ENSG00000181804.15  OTTHUMG00000159373.4
## ENSG00000161835.11  OTTHUMG00000169595.4
## ENSG00000102678.7   OTTHUMG00000017412.3
## ENSG00000164764.11  OTTHUMG00000157144.3
## ENSG00000114200.10  OTTHUMG00000158131.3
## ENSG00000174099.12  OTTHUMG00000168866.6
## ENSG00000128482.16  OTTHUMG00000059601.9
## ENSG00000137094.15  OTTHUMG00000019840.8
## ENSG00000162062.15  OTTHUMG00000128859.6
## ENSG00000138741.11  OTTHUMG00000133069.5
## ENSG00000136244.12  OTTHUMG00000023178.7
## ENSG00000122035.7   OTTHUMG00000016627.2
## ENSG00000276365.1                   <NA>
## ENSG00000177098.9   OTTHUMG00000166994.1
## ENSG00000145012.13  OTTHUMG00000156387.5
## ENSG00000243244.7   OTTHUMG00000129169.7
## ENSG00000168398.7   OTTHUMG00000171408.1
## ENSG00000025423.11  OTTHUMG00000170854.5
## ENSG00000172915.18  OTTHUMG00000016724.2
## ENSG00000113448.19  OTTHUMG00000162218.9
## ENSG00000138135.7   OTTHUMG00000018705.2
## ENSG00000282840.1   OTTHUMG00000191601.1
## ENSG00000144642.22 OTTHUMG00000155699.10
## ENSG00000103316.12  OTTHUMG00000090707.4
## ENSG00000162772.17  OTTHUMG00000036747.9
## ENSG00000234638.1   OTTHUMG00000058923.1
## ENSG00000140092.14  OTTHUMG00000171122.3
## ENSG00000159433.12  OTTHUMG00000175799.3
## ENSG00000018236.15  OTTHUMG00000169362.4
## ENSG00000204241.8   OTTHUMG00000167125.4
## ENSG00000131711.15 OTTHUMG00000100952.13
## ENSG00000124212.6   OTTHUMG00000033077.5
## ENSG00000180354.16  OTTHUMG00000128559.5
## ENSG00000255085.8   OTTHUMG00000165188.3
## ENSG00000117600.12  OTTHUMG00000010763.2
## ENSG00000111860.14  OTTHUMG00000015465.6
## ENSG00000135424.18 OTTHUMG00000170699.17
## ENSG00000261616.1   OTTHUMG00000177238.1
## ENSG00000163637.13  OTTHUMG00000158789.6
## ENSG00000154553.16  OTTHUMG00000160412.8
## ENSG00000162733.19  OTTHUMG00000034423.9
## ENSG00000100490.9   OTTHUMG00000140290.6
## ENSG00000213144.2   OTTHUMG00000168933.1
## ENSG00000198848.13  OTTHUMG00000176718.4
## ENSG00000241644.2   OTTHUMG00000167163.2
## ENSG00000004846.16  OTTHUMG00000094789.7
```

```
## ENSG00000178343.5   OTTHUMG00000161043.2
## ENSG00000144857.14 OTTHUMG00000159305.22
## ENSG00000159182.5   OTTHUMG00000159899.3
## ENSG00000091436.17  OTTHUMG00000132297.8
## ENSG00000138722.10  OTTHUMG00000130947.3
## ENSG00000168386.18  OTTHUMG00000159055.3
## ENSG00000258441.1   OTTHUMG00000170756.1
## ENSG00000172554.12  OTTHUMG00000151370.5
## ENSG00000129682.16  OTTHUMG00000022532.8
## ENSG00000121577.14  OTTHUMG00000159438.3
## ENSG00000262655.4   OTTHUMG00000181576.4
## ENSG00000188906.16 OTTHUMG00000059742.12
## ENSG00000148516.22  OTTHUMG00000017907.5
## ENSG00000102804.15  OTTHUMG00000016838.8
## ENSG00000006071.15 OTTHUMG00000166316.33
## ENSG00000185681.13  OTTHUMG00000020599.4
## ENSG00000100448.4   OTTHUMG00000140182.3
## ENSG00000168546.11  OTTHUMG00000163897.7
## ENSG00000186594.14  OTTHUMG00000132197.7
## ENSG00000116678.20  OTTHUMG00000009115.1
## ENSG00000255007.1   OTTHUMG00000166910.1
## ENSG00000169306.10  OTTHUMG00000021317.2
## ENSG00000134533.6   OTTHUMG00000168745.2
## ENSG00000182950.3   OTTHUMG00000142837.2
## ENSG00000226306.6   OTTHUMG00000163236.2
## ENSG00000204687.4   OTTHUMG00000031089.2
## ENSG00000249249.2   OTTHUMG00000155201.3
## ENSG00000131730.16  OTTHUMG00000119013.5
## ENSG00000268754.4   OTTHUMG00000183865.5
## ENSG00000132561.14  OTTHUMG00000164795.5
## ENSG00000158246.8   OTTHUMG00000004278.4
## ENSG00000162804.14  OTTHUMG00000151803.4
## ENSG00000116176.6   OTTHUMG00000090466.2
## ENSG00000134532.19  OTTHUMG00000169026.8
## ENSG00000124749.17  OTTHUMG00000014907.6
## ENSG00000005249.13  OTTHUMG00000137418.2
## ENSG00000280339.1   OTTHUMG00000167230.1
## ENSG00000150625.16  OTTHUMG00000160760.5
## ENSG00000131409.13  OTTHUMG00000183063.4
## ENSG00000133874.2   OTTHUMG00000163960.2
## ENSG00000234456.8  OTTHUMG00000155465.14
## ENSG00000149575.9                   <NA>
## ENSG00000158402.20  OTTHUMG00000129203.6
## ENSG00000164197.12  OTTHUMG00000162278.3
## ENSG00000185052.12  OTTHUMG00000031993.5
## ENSG00000136732.16  OTTHUMG00000131464.4
## ENSG00000140416.23 OTTHUMG00000132803.19
## ENSG00000247416.4   OTTHUMG00000167662.3
## ENSG00000230500.1   OTTHUMG00000017863.1
## ENSG00000225706.2   OTTHUMG00000019543.3
## ENSG00000119865.9   OTTHUMG00000129565.5
## ENSG00000139874.6   OTTHUMG00000140249.4
## ENSG00000140285.12  OTTHUMG00000131517.5
## ENSG00000227954.7   OTTHUMG00000015607.4
```

```
## ENSG00000115507.10  OTTHUMG00000129454.4
## ENSG00000166963.13  OTTHUMG00000059756.6
## ENSG00000183578.8   OTTHUMG00000172460.5
## ENSG00000145861.9   OTTHUMG00000130323.4
## ENSG00000046889.19  OTTHUMG00000164402.3
## ENSG00000115353.11  OTTHUMG00000129973.4
## ENSG00000092009.10  OTTHUMG00000140181.2
## ENSG00000174611.12  OTTHUMG00000159788.2
## ENSG00000164303.11  OTTHUMG00000160617.3
## ENSG00000152137.8   OTTHUMG00000168932.2
## ENSG00000172236.18  OTTHUMG00000090467.4
## ENSG00000245812.3   OTTHUMG00000163513.2
## ENSG00000203685.10  OTTHUMG00000037583.2
## ENSG00000263513.5   OTTHUMG00000041029.3
## ENSG00000261625.1   OTTHUMG00000172938.1
## ENSG00000074416.14  OTTHUMG00000159641.6
## ENSG00000266256.3  OTTHUMG00000179141.27
## ENSG00000242732.4   OTTHUMG00000021808.2
## ENSG00000162616.9   OTTHUMG00000040905.6
## ENSG00000133121.21  OTTHUMG00000016708.8
## ENSG00000169418.10  OTTHUMG00000037085.4
## ENSG00000081189.16 OTTHUMG00000162634.12
## ENSG00000283175.1   OTTHUMG00000191675.1
## ENSG00000127920.6   OTTHUMG00000023411.5
## ENSG00000136160.17 OTTHUMG00000017111.14
## ENSG00000146966.13  OTTHUMG00000157411.3
## ENSG00000073712.15  OTTHUMG00000140309.5
## ENSG00000046653.15  OTTHUMG00000021162.5
## ENSG00000116117.18  OTTHUMG00000154562.5
## ENSG00000159200.18  OTTHUMG00000086235.5
## ENSG00000175899.15  OTTHUMG00000150267.7
## ENSG00000128573.26 OTTHUMG00000023131.19
## ENSG00000157570.12  OTTHUMG00000163883.5
## ENSG00000173376.14  OTTHUMG00000132973.5
## ENSG00000185532.20 OTTHUMG00000018248.15
## ENSG00000122877.16  OTTHUMG00000018308.7
## ENSG00000170011.14  OTTHUMG00000131392.4
## ENSG00000123119.12  OTTHUMG00000164009.2
## ENSG00000146592.17 OTTHUMG00000097081.11
## ENSG00000258636.2   OTTHUMG00000170691.2
## ENSG00000187391.22 OTTHUMG00000130697.19
## ENSG00000122707.12  OTTHUMG00000019898.2
## ENSG00000173757.10  OTTHUMG00000150724.4
## ENSG00000170381.14  OTTHUMG00000154693.7
## ENSG00000169744.13  OTTHUMG00000048212.5
## ENSG00000185652.12  OTTHUMG00000168649.3
## ENSG00000105784.15  OTTHUMG00000131035.4
## ENSG00000284600.1   OTTHUMG00000191988.1
## ENSG00000174282.12  OTTHUMG00000108137.5
## ENSG00000155760.2   OTTHUMG00000132841.1
## ENSG00000186185.14  OTTHUMG00000179867.2
## ENSG00000287559.1   OTTHUMG00000195323.1
## ENSG00000225472.2   OTTHUMG00000019570.2
## ENSG00000145242.13  OTTHUMG00000129273.4
```

```
## ENSG00000162878.13  OTTHUMG00000152303.5
## ENSG00000116473.14  OTTHUMG00000011959.2
## ENSG00000149596.7   OTTHUMG00000033037.3
## ENSG00000175161.14  OTTHUMG00000158990.3
## ENSG00000143995.20 OTTHUMG00000150714.14
## ENSG00000123570.4   OTTHUMG00000022112.2
## ENSG00000149591.17  OTTHUMG00000167067.3
## ENSG00000130032.17  OTTHUMG00000024170.5
## ENSG00000169760.17  OTTHUMG00000157005.6
## ENSG00000144668.12  OTTHUMG00000130815.5
## ENSG00000277954.1   OTTHUMG00000176937.2
## ENSG00000168734.14  OTTHUMG00000033065.3
## ENSG00000189134.4   OTTHUMG00000014517.2
## ENSG00000163820.15  OTTHUMG00000133447.5
## ENSG00000166780.11  OTTHUMG00000129883.6
## ENSG00000165995.22  OTTHUMG00000017764.9
## ENSG00000256508.2   OTTHUMG00000172940.1
## ENSG00000277010.1   OTTHUMG00000186692.1
## ENSG00000077782.21 OTTHUMG00000147366.18
## ENSG00000265142.9   OTTHUMG00000179284.4
## ENSG00000258987.2   OTTHUMG00000171299.2
## ENSG00000053328.8   OTTHUMG00000015359.1
## ENSG00000205002.4   OTTHUMG00000164961.2
## ENSG00000170312.16  OTTHUMG00000018290.6
## ENSG00000126950.8   OTTHUMG00000022016.2
## ENSG00000154920.15  OTTHUMG00000164974.3
## ENSG00000163788.14  OTTHUMG00000156491.2
## ENSG00000272183.1   OTTHUMG00000185364.1
## ENSG00000166960.16  OTTHUMG00000132279.7
## ENSG00000097007.19  OTTHUMG00000020813.5
## ENSG00000179820.16  OTTHUMG00000060775.8
## ENSG00000135451.13  OTTHUMG00000169486.3
## ENSG00000108551.5   OTTHUMG00000059292.4
## ENSG00000227487.4   OTTHUMG00000167663.2
## ENSG00000281852.1   OTTHUMG00000189559.1
## ENSG00000187720.14  OTTHUMG00000133389.7
## ENSG00000188803.15  OTTHUMG00000154121.7
## ENSG00000230910.4   OTTHUMG00000014981.5
## ENSG00000129810.15  OTTHUMG00000130479.7
## ENSG00000269970.1   OTTHUMG00000183935.1
## ENSG00000169291.10  OTTHUMG00000036072.4
## ENSG00000005102.14  OTTHUMG00000170513.3
## ENSG00000165480.16  OTTHUMG00000016539.5
## ENSG00000171791.14  OTTHUMG00000132791.5
## ENSG00000197410.14  OTTHUMG00000161556.7
## ENSG00000163659.13  OTTHUMG00000158646.3
## ENSG00000010319.7   OTTHUMG00000158570.3
## ENSG00000171303.7   OTTHUMG00000125530.3
## ENSG00000120907.18  OTTHUMG00000099459.3
## ENSG00000189320.9   OTTHUMG00000155537.5
## ENSG00000151834.16  OTTHUMG00000044266.8
## ENSG00000197253.13  OTTHUMG00000155926.4
## ENSG00000107562.16  OTTHUMG00000018054.2
## ENSG00000106617.15 OTTHUMG00000157324.26
```

```
## ENSG00000135842.17  OTTHUMG00000035388.3
## ENSG00000138162.19  OTTHUMG00000019181.8
## ENSG00000180525.13  OTTHUMG00000017529.5
## ENSG00000129116.19 OTTHUMG00000161097.11
## ENSG00000185112.6   OTTHUMG00000156016.2
## ENSG00000170989.10  OTTHUMG00000010835.3
## ENSG00000143248.13  OTTHUMG00000034441.6
## ENSG00000175745.14  OTTHUMG00000119079.5
## ENSG00000207870.1                   <NA>
## ENSG00000196159.14  OTTHUMG00000133100.5
## ENSG00000156298.13  OTTHUMG00000024090.4
## ENSG00000133083.15  OTTHUMG00000016729.3
## ENSG00000137449.17  OTTHUMG00000090669.3
## ENSG00000152484.14  OTTHUMG00000016626.5
## ENSG00000116254.18  OTTHUMG00000000952.7
## ENSG00000169282.18  OTTHUMG00000158552.4
## ENSG00000185271.9   OTTHUMG00000170982.4
## ENSG00000198756.12  OTTHUMG00000035460.4
## ENSG00000184588.18  OTTHUMG00000009088.6
## ENSG00000100884.9   OTTHUMG00000028781.7
## ENSG00000162599.17 OTTHUMG00000008618.11
## ENSG00000130988.13  OTTHUMG00000021435.3
## ENSG00000215784.6   OTTHUMG00000074801.5
## ENSG00000111728.10  OTTHUMG00000169098.3
## ENSG00000172794.20  OTTHUMG00000134282.5
## ENSG00000053254.16  OTTHUMG00000170898.4
## ENSG00000100678.19  OTTHUMG00000152342.3
## ENSG00000185437.13 OTTHUMG00000074113.13
## ENSG00000259291.2   OTTHUMG00000172032.2
## ENSG00000260466.1   OTTHUMG00000176392.1
## ENSG00000131016.17  OTTHUMG00000015833.3
## ENSG00000058272.19  OTTHUMG00000170100.7
## ENSG00000185985.11  OTTHUMG00000022595.3
## ENSG00000071242.12 OTTHUMG00000016007.10
## ENSG00000163106.11  OTTHUMG00000130974.5
## ENSG00000175063.17  OTTHUMG00000033038.3
## ENSG00000205502.4   OTTHUMG00000171959.2
## ENSG00000075218.19  OTTHUMG00000150486.4
## ENSG00000134531.10  OTTHUMG00000168775.6
## ENSG00000144619.15 OTTHUMG00000119031.12
## ENSG00000008441.18  OTTHUMG00000180726.4
## ENSG00000146013.11  OTTHUMG00000129200.3
## ENSG00000271936.1   OTTHUMG00000185752.1
## ENSG00000170891.11  OTTHUMG00000125479.4
## ENSG00000198901.14  OTTHUMG00000171685.8
## ENSG00000115896.16  OTTHUMG00000132750.9
## ENSG00000182534.13  OTTHUMG00000180370.1
## ENSG00000276043.5   OTTHUMG00000180251.5
## ENSG00000185483.12  OTTHUMG00000009022.2
## ENSG00000159899.15  OTTHUMG00000019871.4
## ENSG00000123485.12 OTTHUMG00000059125.12
## ENSG00000112208.11  OTTHUMG00000014919.3
## ENSG00000115275.14  OTTHUMG00000152886.9
## ENSG00000225670.4   OTTHUMG00000037178.4
```

```
## ENSG00000070808.16  OTTHUMG00000134281.8
## ENSG00000187068.3   OTTHUMG00000156696.2
## ENSG00000196182.11  OTTHUMG00000008238.4
## ENSG00000020181.18  OTTHUMG00000156182.6
## ENSG00000171408.14  OTTHUMG00000015641.2
## ENSG00000196584.3   OTTHUMG00000151470.2
## ENSG00000182168.15  OTTHUMG00000130989.5
## ENSG00000161544.10  OTTHUMG00000180289.2
## ENSG00000180139.11  OTTHUMG00000018703.6
## ENSG00000186198.4   OTTHUMG00000133116.2
## ENSG00000079102.16  OTTHUMG00000164066.6
## ENSG00000118523.6   OTTHUMG00000015573.3
## ENSG00000267304.1   OTTHUMG00000180871.1
## ENSG00000143603.19  OTTHUMG00000037260.4
## ENSG00000130224.15  OTTHUMG00000022233.4
## ENSG00000110422.12  OTTHUMG00000132269.5
## ENSG00000120279.6   OTTHUMG00000015850.4
## ENSG00000136267.14  OTTHUMG00000152477.3
## ENSG00000095794.19 OTTHUMG00000017953.29
## ENSG00000117650.13  OTTHUMG00000037121.3
## ENSG00000078549.15  OTTHUMG00000023884.6
## ENSG00000149534.9   OTTHUMG00000167239.3
## ENSG00000232229.6   OTTHUMG00000018726.3
## ENSG00000230417.12 OTTHUMG00000018552.12
## ENSG00000133895.17 OTTHUMG00000045366.45
## ENSG00000051341.14  OTTHUMG00000159396.2
## ENSG00000166313.19  OTTHUMG00000165454.7
## ENSG00000158270.12  OTTHUMG00000178145.3
## ENSG00000170962.13  OTTHUMG00000165953.4
## ENSG00000011465.18  OTTHUMG00000169998.7
## ENSG00000018408.15 OTTHUMG00000159614.13
## ENSG00000185046.18  OTTHUMG00000170308.8
## ENSG00000285918.1   OTTHUMG00000193214.1
## ENSG00000106144.20  OTTHUMG00000023641.5
## ENSG00000154080.14  OTTHUMG00000179469.5
## ENSG00000112984.12  OTTHUMG00000129195.4
## ENSG00000246379.8   OTTHUMG00000176719.7
## ENSG00000147166.11  OTTHUMG00000021793.2
## ENSG00000085563.15 OTTHUMG00000023393.27
## ENSG00000030304.14  OTTHUMG00000020485.7
## ENSG00000163297.17  OTTHUMG00000151982.4
## ENSG00000159388.6   OTTHUMG00000035834.6
## ENSG00000091651.9   OTTHUMG00000132539.5
## ENSG00000233760.2   OTTHUMG00000152862.6
## ENSG00000146410.12  OTTHUMG00000015643.5
## ENSG00000249679.2   OTTHUMG00000160454.2
## ENSG00000232046.7   OTTHUMG00000152838.3
## ENSG00000169554.22 OTTHUMG00000187624.10
## ENSG00000105011.9   OTTHUMG00000150401.3
## ENSG00000100739.11  OTTHUMG00000171412.2
## ENSG00000166803.14  OTTHUMG00000133017.5
## ENSG00000112742.10  OTTHUMG00000015088.5
## ENSG00000250423.2   OTTHUMG00000162980.2
## ENSG00000163751.4   OTTHUMG00000159526.2
```

```
## ENSG00000159307.19  OTTHUMG00000150679.6
## ENSG00000134198.10  OTTHUMG00000011878.4
## ENSG00000236532.6   OTTHUMG00000078706.2
## ENSG00000162631.19  OTTHUMG00000010965.4
## ENSG00000198835.4   OTTHUMG00000039771.2
## ENSG00000078295.17  OTTHUMG00000090476.6
## ENSG00000131386.19  OTTHUMG00000156893.5
## ENSG00000132329.11  OTTHUMG00000133337.4
## ENSG00000134352.20 OTTHUMG00000097043.14
## ENSG00000115163.15  OTTHUMG00000097073.5
## ENSG00000069431.11  OTTHUMG00000169094.4
## ENSG00000126351.13  OTTHUMG00000133332.3
## ENSG00000072195.15  OTTHUMG00000193740.2
## ENSG00000269067.2   OTTHUMG00000183124.3
## ENSG00000153814.13  OTTHUMG00000128545.8
## ENSG00000227440.1   OTTHUMG00000156866.1
## ENSG00000183856.11  OTTHUMG00000033114.4
## ENSG00000103196.12  OTTHUMG00000137644.7
## ENSG00000135709.12  OTTHUMG00000176597.2
## ENSG00000155093.19  OTTHUMG00000152646.5
## ENSG00000197702.14  OTTHUMG00000165778.6
## ENSG00000127083.7   OTTHUMG00000020225.1
## ENSG00000205363.6   OTTHUMG00000172556.4
## ENSG00000113070.8   OTTHUMG00000129496.5
## ENSG00000197472.15  OTTHUMG00000040707.7
## ENSG00000069702.11  OTTHUMG00000010097.7
## ENSG00000050030.16  OTTHUMG00000021860.4
## ENSG00000174348.14  OTTHUMG00000008934.3
## ENSG00000173258.13  OTTHUMG00000020490.2
## ENSG00000091986.16  OTTHUMG00000159265.4
## ENSG00000121898.13  OTTHUMG00000019206.3
## ENSG00000234336.7   OTTHUMG00000152771.3
## ENSG00000110900.16  OTTHUMG00000168484.2
## ENSG00000154839.10  OTTHUMG00000132685.8
## ENSG00000131747.15  OTTHUMG00000155008.4
## ENSG00000109819.9   OTTHUMG00000097747.3
## ENSG00000116774.12  OTTHUMG00000011980.3
## ENSG00000136895.19 OTTHUMG00000020700.10
## ENSG00000113594.10  OTTHUMG00000131138.4
## ENSG00000204065.3   OTTHUMG00000022092.2
## ENSG00000166265.13  OTTHUMG00000078689.5
## ENSG00000104889.7   OTTHUMG00000180481.3
## ENSG00000140287.11  OTTHUMG00000131644.4
## ENSG00000085999.13 OTTHUMG00000007772.10
## ENSG00000254533.1   OTTHUMG00000165187.1
## ENSG00000091831.24  OTTHUMG00000016103.7
## ENSG00000267466.1   OTTHUMG00000177431.4
## ENSG00000123096.12  OTTHUMG00000169167.4
## ENSG00000105737.9   OTTHUMG00000044573.2
## ENSG00000155511.18  OTTHUMG00000130148.7
## ENSG00000066032.18 OTTHUMG00000152903.13
## ENSG00000189241.8   OTTHUMG00000015427.2
## ENSG00000169224.13  OTTHUMG00000040648.6
## ENSG00000184557.4   OTTHUMG00000177513.5
```

```
## ENSG00000134690.11  OTTHUMG00000004320.2
## ENSG00000258137.5   OTTHUMG00000169930.1
## ENSG00000020577.14  OTTHUMG00000170999.4
## ENSG00000245025.3   OTTHUMG00000163755.2
## ENSG00000234754.2  OTTHUMG00000037427.15
## ENSG00000233304.7   OTTHUMG00000000499.5
## ENSG00000167513.9   OTTHUMG00000173467.3
## ENSG00000162881.6   OTTHUMG00000128643.1
## ENSG00000111665.12  OTTHUMG00000169014.2
## ENSG00000198576.4   OTTHUMG00000134310.4
## ENSG00000167900.12  OTTHUMG00000150674.6
## ENSG00000221829.10  OTTHUMG00000019850.4
## ENSG00000177432.8   OTTHUMG00000130950.3
## ENSG00000223774.5   OTTHUMG00000035808.1
## ENSG00000148773.14  OTTHUMG00000019255.2
## ENSG00000067798.16  OTTHUMG00000170001.5
## ENSG00000179639.10  OTTHUMG00000037176.2
## ENSG00000261026.1   OTTHUMG00000173366.1
## ENSG00000180155.20  OTTHUMG00000191912.4
## ENSG00000105486.14  OTTHUMG00000183187.3
## ENSG00000166341.8   OTTHUMG00000133398.4
## ENSG00000206077.13  OTTHUMG00000162148.5
## ENSG00000233608.4   OTTHUMG00000152836.4
## ENSG00000273162.1   OTTHUMG00000186479.1
## ENSG00000274270.1   OTTHUMG00000187356.1
## ENSG00000170624.13  OTTHUMG00000163445.4
## ENSG00000105509.11  OTTHUMG00000183499.3
## ENSG00000089225.20  OTTHUMG00000166191.3
## ENSG00000206559.8   OTTHUMG00000155705.4
## ENSG00000114853.14  OTTHUMG00000156207.5
## ENSG00000168356.13  OTTHUMG00000048246.6
## ENSG00000003096.14  OTTHUMG00000022252.4
## ENSG00000071967.12  OTTHUMG00000132260.4
## ENSG00000279526.1   OTTHUMG00000151890.1
## ENSG00000153707.17  OTTHUMG00000021005.6
## ENSG00000136717.15  OTTHUMG00000131465.3
## ENSG00000151067.22 OTTHUMG00000150243.13
## ENSG00000165943.5   OTTHUMG00000169184.3
## ENSG00000082482.14  OTTHUMG00000037017.4
## ENSG00000137076.21  OTTHUMG00000019874.5
## ENSG00000080561.14  OTTHUMG00000022171.4
## ENSG00000197245.6   OTTHUMG00000007537.3
## ENSG00000172465.14  OTTHUMG00000022699.2
## ENSG00000180198.16 OTTHUMG00000003645.13
## ENSG00000099860.9   OTTHUMG00000180434.4
## ENSG00000051180.17 OTTHUMG00000130067.10
## ENSG00000112214.10  OTTHUMG00000015239.2
## ENSG00000278934.1   OTTHUMG00000182479.1
## ENSG00000198963.11  OTTHUMG00000020026.3
## ENSG00000145708.11  OTTHUMG00000102133.3
## ENSG00000257943.2   OTTHUMG00000170260.2
## ENSG00000080224.17  OTTHUMG00000159208.8
## ENSG00000104147.9   OTTHUMG00000130251.5
## ENSG00000285336.1   OTTHUMG00000192437.1
```

```
## ENSG00000167874.7   OTTHUMG00000178022.2
## ENSG00000170464.10  OTTHUMG00000129225.6
## ENSG00000158458.21  OTTHUMG00000129241.9
## ENSG00000109610.6   OTTHUMG00000128565.4
## ENSG00000184545.11  OTTHUMG00000133348.5
## ENSG00000086232.13 OTTHUMG00000090689.15
## ENSG00000267532.5   OTTHUMG00000180219.1
## ENSG00000198467.16 OTTHUMG00000019878.10
## ENSG00000087274.17 OTTHUMG00000122080.23
## ENSG00000183023.18  OTTHUMG00000102183.9
## ENSG00000137033.11  OTTHUMG00000019516.1
## ENSG00000198075.10  OTTHUMG00000130958.4
## ENSG00000158301.18  OTTHUMG00000185178.1
## ENSG00000172159.16  OTTHUMG00000020103.4
## ENSG00000169239.13  OTTHUMG00000021183.6
## ENSG00000107758.16  OTTHUMG00000018472.3
## ENSG00000226530.2   OTTHUMG00000021532.2
## ENSG00000122952.17  OTTHUMG00000018261.3
## ENSG00000166165.13  OTTHUMG00000171786.3
## ENSG00000186675.7   OTTHUMG00000021870.2
## ENSG00000175395.16  OTTHUMG00000019344.4
## ENSG00000126803.9   OTTHUMG00000141311.4
## ENSG00000029993.15  OTTHUMG00000024162.3
## ENSG00000234224.3   OTTHUMG00000154762.4
## ENSG00000137601.17  OTTHUMG00000160963.5
## ENSG00000197614.11  OTTHUMG00000168703.4
## ENSG00000185010.15 OTTHUMG00000022688.16
## ENSG00000259236.1   OTTHUMG00000172266.1
## ENSG00000036672.16  OTTHUMG00000166200.3
## ENSG00000153179.13  OTTHUMG00000168812.3
## ENSG00000187140.6   OTTHUMG00000009141.2
## ENSG00000177606.8   OTTHUMG00000008376.2
## ENSG00000137103.20 OTTHUMG00000019885.10
## ENSG00000130653.16 OTTHUMG00000020990.12
## ENSG00000163520.14 OTTHUMG00000155437.10
## ENSG00000129596.5   OTTHUMG00000128891.4
## ENSG00000179528.16  OTTHUMG00000170595.6
## ENSG00000240771.8   OTTHUMG00000152516.4
## ENSG00000154736.6   OTTHUMG00000078686.4
## ENSG00000179051.14  OTTHUMG00000002513.2
## ENSG00000166979.13  OTTHUMG00000064922.6
## ENSG00000146090.16  OTTHUMG00000130914.8
## ENSG00000120057.5   OTTHUMG00000018866.2
## ENSG00000169218.14  OTTHUMG00000004321.5
## ENSG00000144230.16  OTTHUMG00000131533.5
## ENSG00000169946.14  OTTHUMG00000164818.5
## ENSG00000184144.12 OTTHUMG00000037105.39
## ENSG00000012822.16  OTTHUMG00000170095.5
## ENSG00000104936.19  OTTHUMG00000182334.4
## ENSG00000165495.16  OTTHUMG00000165884.5
## ENSG00000144339.12  OTTHUMG00000132723.5
## ENSG00000166405.15  OTTHUMG00000165694.6
## ENSG00000142945.13  OTTHUMG00000008416.7
## ENSG00000118971.9   OTTHUMG00000168123.4
```

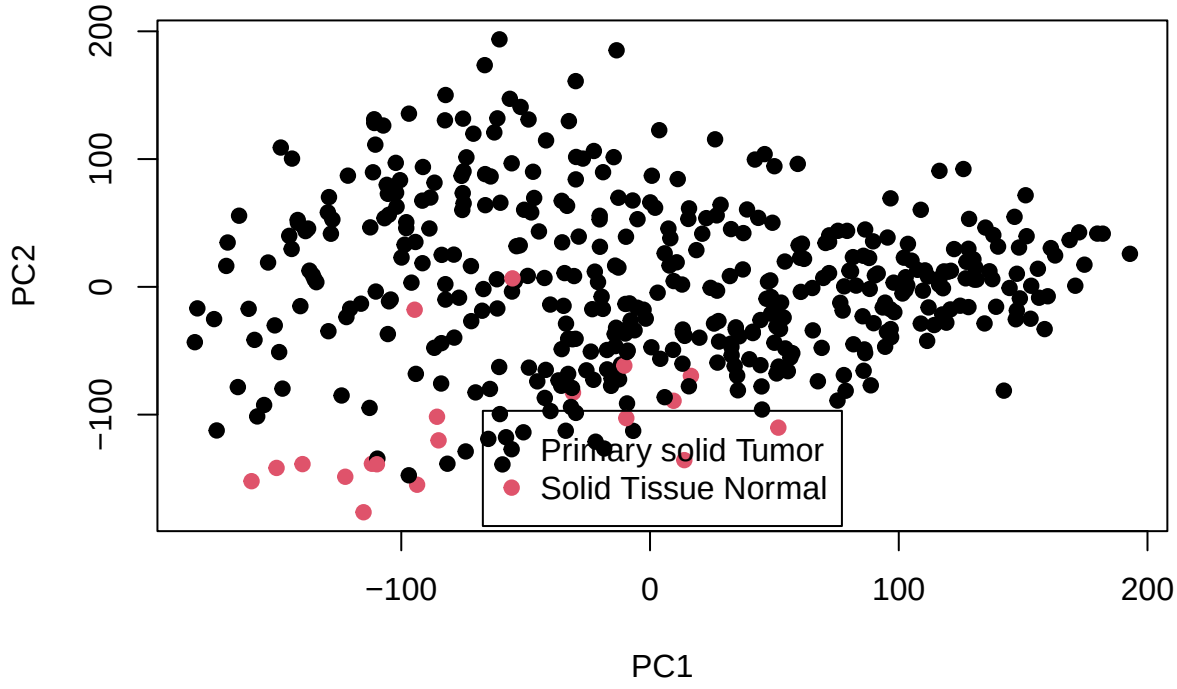
```
## ENSG00000188610.12  OTTHUMG00000185025.1
## ENSG00000035499.13  OTTHUMG00000097083.4
## ENSG00000152253.9  OTTHUMG00000132181.10
## ENSG00000178999.13  OTTHUMG00000108189.6
## ENSG00000187699.10  OTTHUMG00000154361.3
## ENSG00000176697.20  OTTHUMG00000178797.3
## ENSG00000174306.21  OTTHUMG00000032481.7
## ENSG00000109794.13 OTTHUMG00000160565.11
## ENSG00000068383.19  OTTHUMG00000019293.6
## ENSG00000152049.6   OTTHUMG00000133161.4
## ENSG00000113742.14  OTTHUMG00000130541.7
## ENSG00000179921.15  OTTHUMG00000155203.4
## ENSG00000150347.16  OTTHUMG00000018298.3
## ENSG00000121297.8   OTTHUMG00000150184.5
## ENSG00000076382.17 OTTHUMG00000166586.27
## ENSG00000134138.20  OTTHUMG00000129781.7
## ENSG00000166912.17  OTTHUMG00000175662.2
## ENSG00000093009.11  OTTHUMG00000150386.3
## ENSG00000048342.17 OTTHUMG00000160255.16
## ENSG00000225873.4   OTTHUMG00000156179.8
## ENSG00000257222.1   OTTHUMG00000170477.1
## ENSG00000102010.15  OTTHUMG00000021180.2
## ENSG00000007237.19 OTTHUMG00000177945.15
## ENSG00000196843.17  OTTHUMG00000155229.8
## ENSG00000108691.9   OTTHUMG00000132887.3
## ENSG00000088325.16  OTTHUMG00000032190.3
## ENSG00000133687.16  OTTHUMG00000169324.2
## ENSG00000257596.1   OTTHUMG00000169701.1
## ENSG00000185022.12  OTTHUMG00000151163.5
## ENSG00000197321.15 OTTHUMG00000017882.10
## ENSG00000150594.7   OTTHUMG00000019050.3
## ENSG00000089685.15  OTTHUMG00000177505.9
## ENSG00000127564.17  OTTHUMG00000128975.4
## ENSG00000144395.18  OTTHUMG00000154475.6
## ENSG00000065320.9   OTTHUMG00000130257.6
## ENSG00000189184.12  OTTHUMG00000161348.2
## ENSG00000171243.8   OTTHUMG00000090807.2
## ENSG00000102802.10  OTTHUMG00000016679.2
## ENSG00000008710.20  OTTHUMG00000155795.6
## ENSG00000197299.12  OTTHUMG00000149834.7
## ENSG00000126524.10  OTTHUMG00000023165.4
## ENSG00000241399.7   OTTHUMG00000154080.2
## ENSG00000286447.1   OTTHUMG00000195727.1
## ENSG00000278023.7   OTTHUMG00000188399.3
## ENSG00000228063.2   OTTHUMG00000041228.5
## ENSG00000177301.16 OTTHUMG00000011567.10
## ENSG00000121152.10  OTTHUMG00000130451.6
## ENSG00000198673.10  OTTHUMG00000170207.6
## ENSG00000103657.14  OTTHUMG00000172433.6
## ENSG00000169116.11  OTTHUMG00000160827.1
## ENSG00000275569.1   OTTHUMG00000186851.2
## ENSG00000160957.14  OTTHUMG00000165178.7
## ENSG00000144645.14 OTTHUMG00000130672.11
## ENSG00000108342.13  OTTHUMG00000133247.3
```

```
## ENSG00000169607.13  OTTHUMG00000131313.3
## ENSG00000184661.14  OTTHUMG00000099429.7
## ENSG00000140682.19  OTTHUMG00000132467.7
## ENSG00000049540.18 OTTHUMG00000150229.14
## ENSG00000204406.14 OTTHUMG00000150440.15
## ENSG00000168952.15  OTTHUMG00000140186.3
## ENSG00000174371.17  OTTHUMG00000039965.9
## ENSG00000129473.11  OTTHUMG00000028738.5
## ENSG00000271474.1   OTTHUMG00000185100.1
## ENSG00000138092.11  OTTHUMG00000125525.4
## ENSG00000196924.19 OTTHUMG00000022712.33
## ENSG00000132718.9   OTTHUMG00000014105.2
## ENSG00000224743.7   OTTHUMG00000016681.7
## ENSG00000101003.10  OTTHUMG00000032124.2
## ENSG00000239218.2   OTTHUMG00000158605.2
## ENSG00000177374.13 OTTHUMG00000177635.11
## ENSG00000172399.6   OTTHUMG00000132968.2
## ENSG00000179909.15  OTTHUMG00000140375.2
## ENSG00000124831.19  OTTHUMG00000133339.5
## ENSG00000125648.15  OTTHUMG00000180852.5
## ENSG00000182389.20 OTTHUMG00000155091.19
## ENSG00000197043.14  OTTHUMG00000164179.7
## ENSG00000266010.2   OTTHUMG00000179440.2
## ENSG00000141433.13  OTTHUMG00000131479.6
## ENSG00000123219.13  OTTHUMG00000131227.5
## ENSG00000151729.11  OTTHUMG00000134299.5
## ENSG00000204131.9   OTTHUMG00000021807.4
## ENSG00000221818.9   OTTHUMG00000163838.3
## ENSG00000064205.10  OTTHUMG00000033071.2
## ENSG00000184226.15  OTTHUMG00000017040.6
## ENSG00000187741.15  OTTHUMG00000173049.7
## ENSG00000197275.14  OTTHUMG00000133658.7
## ENSG00000166402.9   OTTHUMG00000165690.1
## ENSG00000087586.18  OTTHUMG00000032796.7
## ENSG00000135447.17  OTTHUMG00000169934.4
## ENSG00000174456.15  OTTHUMG00000193487.1
## ENSG00000156486.8   OTTHUMG00000044337.4
## ENSG00000229367.1   OTTHUMG00000039469.1
## ENSG00000151789.12 OTTHUMG00000130484.12
## ENSG00000173068.18 OTTHUMG00000019593.13
## ENSG00000162852.14  OTTHUMG00000040090.4
## ENSG00000166483.11  OTTHUMG00000165863.4
## ENSG00000108001.15  OTTHUMG00000019265.4
## ENSG00000013810.20 OTTHUMG00000089535.25
## ENSG00000189159.16  OTTHUMG00000154521.4
## ENSG00000187079.20  OTTHUMG00000165878.8
## ENSG00000140807.7   OTTHUMG00000133169.7
## ENSG00000164463.12  OTTHUMG00000163322.3
## ENSG00000128266.9   OTTHUMG00000150611.6
## ENSG00000009413.16  OTTHUMG00000016318.5
## ENSG00000184005.11  OTTHUMG00000009615.2
## ENSG00000094804.12  OTTHUMG00000133324.7
## ENSG00000179954.16  OTTHUMG00000180855.5
## ENSG00000146374.14  OTTHUMG00000015521.2
```

```
## ENSG00000196550.10  OTTHUMG00000042552.3
## ENSG00000164161.10  OTTHUMG00000161428.3
## ENSG00000080166.16  OTTHUMG00000017206.8
## ENSG00000110076.19 OTTHUMG00000045214.19
## ENSG00000104313.20 OTTHUMG00000149894.17
## ENSG00000100162.15  OTTHUMG00000151277.3
## ENSG00000197776.8   OTTHUMG00000140295.5
## ENSG00000160808.11  OTTHUMG00000133516.9
## ENSG00000272321.1   OTTHUMG00000185509.1
## ENSG00000163467.12  OTTHUMG00000024060.2
## ENSG00000077152.11  OTTHUMG00000041392.5
## ENSG00000072571.20  OTTHUMG00000130381.8
## ENSG00000143228.13  OTTHUMG00000034275.4
## ENSG00000007908.16  OTTHUMG00000034851.9
## ENSG00000276409.5   OTTHUMG00000188403.3
## ENSG00000185920.16  OTTHUMG00000020280.6
## ENSG00000182397.14  OTTHUMG00000149852.2
## ENSG00000163882.9   OTTHUMG00000156746.2
## ENSG00000258498.8   OTTHUMG00000171682.1
## ENSG00000126947.13  OTTHUMG00000022033.3
## ENSG00000161888.11  OTTHUMG00000180798.2
## ENSG00000111911.7   OTTHUMG00000015514.3
## ENSG00000132514.14  OTTHUMG00000177939.4
## ENSG00000229274.2   OTTHUMG00000086581.3
## ENSG00000101746.15  OTTHUMG00000132291.5
## ENSG00000011304.22 OTTHUMG00000181789.13
## ENSG00000134853.12  OTTHUMG00000128699.6
## ENSG00000164283.13  OTTHUMG00000097010.6
## ENSG00000117724.13  OTTHUMG00000036955.2
```

Step 3 - Visualize

```r
# make a function to generate a scatter plot to show a separation of tumor vs normal points
plot_PCA = function(voomObj, condition_variable){
  # create a factor
  group = factor(voomObj$targets[, condition_variable])
  # perform a principal component analysis
  pca = prcomp(t(voomObj$E))
  # Take PC1 and PC2 for the plot
  plot(pca$x[,1:2],col=group, pch=19)
  # include a legend for points
  legend("bottom", inset=.01, levels(group), pch=19, col=1:length(levels(group)))
  return(pca)
}

# call the plot function with the voom object and the defintion column
res_pca = plot_PCA(limma_res$voomObj, "definition")
```




```
# create a volcano plot
x = limma_res$topGenes$logFC
y = limma_res$topGenes$adj.P.Val
TCGAVisualize_volcano(
  x,
  y,
  xlab = "logFC",
  title = "Volcano plot of top 1000 genes",)
```

```
## Saving file as: volcano.pdf
```

Step 4 - Classification model training, testing, and evaluation

```
# TODO need to redo this whole step using WGCNA

# use the expression data that has been normalized
# Transpose and make it into a matrix object
d_mat = as.matrix(t(limma_res$voomObj$E))

# and the clinical feature to distinguish cases ("definition")
# Make it a factor
d_resp = as.factor(limma_res$voomObj$targets$definition)
```

```r
# Divide data into training and testing set
# 75% of samples for training and 25% for testing

# Set (random-number-generator) seed so that results are consistent between runs
set.seed(42)

# create a vector of booleans to subset the cases
train_ids = createDataPartition(d_resp, p=0.75, list=FALSE)

# x is the matrix with normalized expression data
# y is the vector with the response variable (tumor vs normal)
x_train = d_mat[train_ids, ]
x_test  = d_mat[-train_ids, ]

y_train = d_resp[train_ids]
y_test  = d_resp[-train_ids]

# do an elastic net model - a generalized linear model that
#  combines lasso and ridge regression, it selects the genes or groups of genes
#  that best predict the condition and uses these to build the model
#  that is then used for classification

# Train model on training dataset using cross-validation
#  alpha can be between 0 (ridge regression) and 1 (lasso)
# the res object here is an object that holds the model coeffiecients and the
#  mean error found during training
res = cv.glmnet(
  x = x_train,
  y = y_train,
  alpha = 0.5,
  family = "binomial")

# Test/Make prediction on test dataset
y_pred = predict(res, newx=x_test, type="class", s="lambda.min")

# confusion matrix shows the TP, TN, FP, and FN
confusion_matrix = table(y_pred, y_test)

# Evaluation statistics
print(confusion_matrix)
```

```
##                      y_test
## y_pred                Primary solid Tumor Solid Tissue Normal
##   Primary solid Tumor                 103                   1
##   Solid Tissue Normal                   0                   3
```

```r
print(paste0("Sensitivity: ",sensitivity(confusion_matrix)))
```

```
## [1] "Sensitivity: 1"
```

```r
print(paste0("Specificity: ",specificity(confusion_matrix)))
```

```
## [1] "Specificity: 0.75"
```

```
print(paste0("Precision: ",precision(confusion_matrix)))
```

```
## [1] "Precision: 0.990384615384615"
```

```
# now we can look at the genes that most contribute for the prediction
res_coef = coef(res, s="lambda.min") # the "coef" function returns a sparse matrix

# ignore zero value coefficients
res_coef = res_coef[res_coef[,1] != 0,]

# remove first coefficient as this is the intercept, a variable of the model itself
res_coef = res_coef[-1]

relevant_genes = names(res_coef) # get names of the (non-zero) variables.
length(relevant_genes) # number of selected genes
```

```
## [1] 83
```

```
# get the Ensembl gene names
head(relevant_genes) # few select genes
```

```
## [1] "ENSG00000034971.17" "ENSG00000078804.13" "ENSG00000081181.8"
## [4] "ENSG00000086991.13" "ENSG00000101057.16" "ENSG00000102683.8"
```

```
# get the common gene names
head(limma_res$voomObj$genes)
```

```
##                     source type score phase            gene_id      gene_type
## ENSG00000000003.15 HAVANA gene    NA    NA ENSG00000000003.15 protein_coding
## ENSG00000000005.6  HAVANA gene    NA    NA  ENSG00000000005.6 protein_coding
## ENSG00000000419.13 HAVANA gene    NA    NA ENSG00000000419.13 protein_coding
## ENSG00000000457.14 HAVANA gene    NA    NA ENSG00000000457.14 protein_coding
## ENSG00000000460.17 HAVANA gene    NA    NA ENSG00000000460.17 protein_coding
## ENSG00000000938.13 HAVANA gene    NA    NA ENSG00000000938.13 protein_coding
##                     gene_name level    hgnc_id          havana_gene
## ENSG00000000003.15    TSPAN6     2 HGNC:11858 OTTHUMG00000022002.2
## ENSG00000000005.6       TNMD     2 HGNC:17757 OTTHUMG00000022001.2
## ENSG00000000419.13      DPM1     2  HGNC:3005 OTTHUMG00000032742.2
## ENSG00000000457.14     SCYL3     2 HGNC:19285 OTTHUMG00000035941.6
## ENSG00000000460.17  C1orf112     2 HGNC:25565 OTTHUMG00000035821.9
## ENSG00000000938.13       FGR     2  HGNC:3697 OTTHUMG00000003516.3
```

```
relevant_gene_names = limma_res$voomObj$genes[relevant_genes,"gene_name"]
head(relevant_gene_names) # few select genes (with readable names now)
```

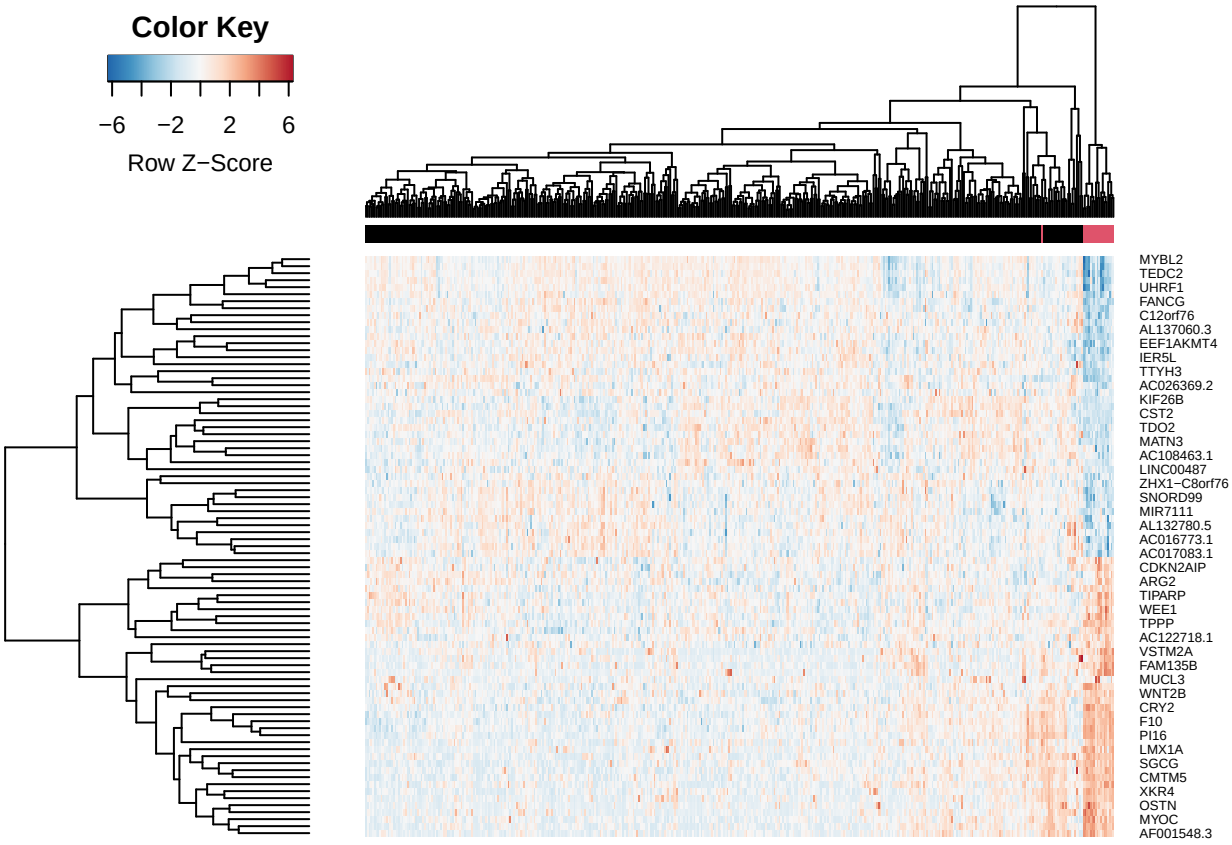
```
## [1] "MYOC"     "TP53INP2" "ARG2"     "NOX4"     "MYBL2"    "SGCG"
```

```r
# did elastic net find the same genes originally found by the limma pipeline?
#  "Of note, we do not expect a high overlap between genes selected by limma and Elastic net.
#   The reason for this is the fact Elastic Net criteria bias the selection of genes,
#   which are not highly correlated against each other, while not such bias is
#   present in limma."
print(intersect(limma_res$topGenes$ensembl_gene_id, relevant_genes))
```

```
## NULL
```

Step 5 - Hierarchical clustering

```r
# we are only considering the elastic net results to cluster genes together
# genes in green are original limma results
# genes in red are normal tissue from the elastic net results
# genes in black are tumor tissue from the elastic net results

 # define the color palette for the plot
hmcol = colorRampPalette(rev(brewer.pal(9, "RdBu")))(256)

# perform complete linkage clustering
clust = function(x) hclust(x, method="complete")
# use the inverse of correlation as distance.
dist = function(x) as.dist((1-cor(t(x)))/2)

# Show green color for genes that also show up in DE analysis
colorLimmaGenes = ifelse(
  # Given a vector of boolean values
  (relevant_genes %in% limma_res$topGenes$ensembl_gene_id),
  "green", # if true, return green for that value
  "white" # if false, return white for that value
)

# generate the heatmap
gene_heatmap = heatmap.2(
  t(d_mat[,relevant_genes]),
  scale="row",          # scale the values for each gene (row)
  density.info="none",  # turns off density plot inside color legend
  trace="none",         # turns off trace lines inside the heat map
  col=hmcol,            # define the color map
  labRow=relevant_gene_names, # use gene names instead of ensembl annotation
  RowSideColors=colorLimmaGenes,
  labCol=FALSE,         # Not showing column labels
  ColSideColors=as.character(as.numeric(d_resp)), # Show colors for each response class
  dendrogram="both",    # Show dendrograms for both axis
  hclust = clust,       # Define hierarchical clustering method
  distfun = dist,       # Using correlation coefficient for distance function
  cexRow=.6,            # Resize row labels
  margins=c(1,5)        # Define margin spaces
)
```


Color Key
−6
−2
2
6
Row Z−Score
MYBL2
TEDC2
UHRF1
FANCG
C12orf76
AL137060.3
EEF1AKMT4
IER5L
TTYH3
AC026369.2
KIF26B
CST2
TDO2
MATN3
AC108463.1
LINC00487
ZHX1−C8orf76
SNORD99
MIR7111
AL132780.5
AC016773.1
AC017083.1
CDKN2AIP
ARG2
TIPARP
WEE1
TPPP
AC122718.1
VSTM2A
FAM135B
MUCL3
WNT2B
CRY2
F10
PI16
LMX1A
SGCG
CMTM5
XKR4
OSTN
MYOC
AF001548.3