

Analysis of Nasal Epithelium

2024-09-27

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com> (<http://rmarkdown.rstudio.com>).

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
#install.packages("limma")
library(limma)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ forcats   1.0.0   ✓ readr     2.1.5
## ✓ ggplot2   3.5.1   ✓ stringr   1.5.1
## ✓ lubridate 1.9.3   ✓ tibble    3.2.1
## ✓ purrr     1.0.2   ✓ tidyr     1.3.1
```

```
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(GEOquery)
```

```
## Loading required package: Biobase
## Loading required package: BiocGenerics
##
## Attaching package: 'BiocGenerics'
##
## The following objects are masked from 'package:lubridate':
##
##   intersect, setdiff, union
##
## The following objects are masked from 'package:dplyr':
##
##   combine, intersect, setdiff, union
##
## 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, sort,
##   table, tapply, union, unique, unsplit, which.max, which.min
##
## Welcome to Bioconductor
##
##   Vignettes contain introductory material; view with
##   'browseVignettes()'. To cite Bioconductor, see
##   'citation("Biobase)", and for packages 'citation("pkgname)".
##
## Setting options('download.file.method.GEOquery'='auto')
## Setting options('GEOquery.inmemory.gpl'=FALSE)
```

```
library(pheatmap)
```

```
file_path <- "~/Downloads/GDS3309_full.soft"
raw_data <- readLines(file_path)
head(raw_data, 100)
```

```

## [1] "^DATABASE = Geo"
## [2] "!!Database_name = Gene Expression Omnibus (GEO)"
## [3] "!!Database_institute = NCBI NLM NIH"
## [4] "!!Database_web_link = http://www.ncbi.nlm.nih.gov/geo"
## [5] "!!Database_email = geo@ncbi.nlm.nih.gov"
## [6] "!!Database_ref = Nucleic Acids Res. 2005 Jan 1;33 Database Issue:D562-6"
## [7] "^DATASET = GDS3309"
## [8] "!!dataset_title = Cigarette smoking effect on the nasal epithelium"
## [9] "!!dataset_description = Analysis of nasal epithelia from cigarette smokers. Cigarette smoke creates a field of injury in epithelial cells lining the respiratory tract. Results extend the concept of a smoking-induced field of injury beyond intrathoracic (bronchial) epithelia to extrathoracic epithelia that line the nose."
## [10] "!!dataset_type = Expression profiling by array"
## [11] "!!dataset_pubmed_id = 18513428"
## [12] "!!dataset_platform = GPL571"
## [13] "!!dataset_platform_organism = Homo sapiens"
## [14] "!!dataset_platform_technology_type = in situ oligonucleotide"
## [15] "!!dataset_feature_count = 22277"
## [16] "!!dataset_sample_organism = Homo sapiens"
## [17] "!!dataset_sample_type = RNA"
## [18] "!!dataset_channel_count = 1"
## [19] "!!dataset_sample_count = 15"
## [20] "!!dataset_value_type = count"
## [21] "!!dataset_reference_series = GSE8987"
## [22] "!!dataset_order = none"
## [23] "!!dataset_update_date = Sep 10 2008"
## [24] "^SUBSET = GDS3309_1"
## [25] "!!subset_dataset_id = GDS3309"
## [26] "!!subset_description = control"
## [27] "!!subset_sample_id = GSM227868,GSM227870,GSM227871,GSM227874,GSM227876,GSM227877,GSM227878,GSM227880"
## [28] "!!subset_type = agent"
## [29] "^SUBSET = GDS3309_2"
## [30] "!!subset_dataset_id = GDS3309"
## [31] "!!subset_description = cigarette smoke"
## [32] "!!subset_sample_id = GSM227869,GSM227872,GSM227873,GSM227875,GSM227879,GSM227881,GSM227882"
## [33] "!!subset_type = agent"
## [34] "^Annotation"
## [35] "!!Annotation_date = Aug 09 2016"
## [36] "!!Annotation_platform = GPL571"
## [37] "!!Annotation_platform_title = [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array"
## [38] "!!Annotation_platform_organism = Homo sapiens"
## [39] "^DATASET = GDS3309"
## [40] "#ID_REF = Platform reference identifier"
## [41] "#IDENTIFIER = identifier"
## [42] "#GSM227868 = Value for GSM227868: Nose10 (Never Smoker); src: Nasal epithelial samples from Never Smoker"
## [43] "#GSM227870 = Value for GSM227870: Nose12 (Never Smoker); src: Nasal epithelial samples from Never Smoker"
## [44] "#GSM227871 = Value for GSM227871: Nose13 (Never Smoker); src: Nasal epithelial

```

```

samples from Never Smoker"
## [45] "#GSM227874 = Value for GSM227874: Nose16 (Never Smoker); src: Nasal epithelial
samples from Never Smoker"
## [46] "#GSM227876 = Value for GSM227876: Nose31 (Never Smoker); src: Nasal epithelial
samples from Never Smoker"
## [47] "#GSM227877 = Value for GSM227877: Nose32 (Never Smoker); src: Nasal epithelial
samples from Never Smoker"
## [48] "#GSM227878 = Value for GSM227878: Nose33 (Never Smoker); src: Nasal epithelial
samples from Never Smoker"
## [49] "#GSM227880 = Value for GSM227880: Nose35 (Never Smoker); src: Nasal epithelial
samples from Never Smoker"
## [50] "#GSM227869 = Value for GSM227869: Nose11 (Current Smoker); src: Nasal epitheli
al samples from Current Smoker"
## [51] "#GSM227872 = Value for GSM227872: Nose14 (Current Smoker); src: Nasal epitheli
al samples from Current Smoker"
## [52] "#GSM227873 = Value for GSM227873: Nose15 (Current Smoker); src: Nasal epitheli
al samples from Current Smoker"
## [53] "#GSM227875 = Value for GSM227875: Nose17 (Current Smoker); src: Nasal epitheli
al samples from Current Smoker"
## [54] "#GSM227879 = Value for GSM227879: Nose34 (Current Smoker); src: Nasal epitheli
al samples from Current Smoker"
## [55] "#GSM227881 = Value for GSM227881: Nose36 (Current Smoker); src: Nasal epitheli
al samples from Current Smoker"
## [56] "#GSM227882 = Value for GSM227882: Nose37 (Current Smoker); src: Nasal epitheli
al samples from Current Smoker"
## [57] "#Gene title = Entrez Gene name"
## [58] "#Gene symbol = Entrez Gene symbol"
## [59] "#Gene ID = Entrez Gene identifier"
## [60] "#UniGene title = Entrez UniGene name"
## [61] "#UniGene symbol = Entrez UniGene symbol"
## [62] "#UniGene ID = Entrez UniGene identifier"
## [63] "#Nucleotide Title = Entrez Nucleotide title"
## [64] "#GI = GenBank identifier"
## [65] "#GenBank Accession = GenBank accession"
## [66] "#Platform_CLONEID = CLONE_ID from Platform data table"
## [67] "#Platform_ORF = ORF from Platform data table"
## [68] "#Platform_SPOTID = SPOT_ID from Platform data table"
## [69] "#Chromosome location = Entrez gene chromosome and location"
## [70] "#Chromosome annotation = Entrez gene chromosome annotation"
## [71] "#GO:Function = Gene Ontology Function term"
## [72] "#GO:Process = Gene Ontology Process term"
## [73] "#GO:Component = Gene Ontology Component term"
## [74] "#GO:Function ID = Gene Ontology Function identifier"
## [75] "#GO:Process ID = Gene Ontology Process identifier"
## [76] "#GO:Component ID = Gene Ontology Component identifier"
## [77] "!dataset_table_begin"
## [78] "ID_REF\tIDENTIFIER\tGSM227868\tGSM227870\tGSM227871\tGSM227874\tGSM227876\tGSM
227877\tGSM227878\tGSM227880\tGSM227869\tGSM227872\tGSM227873\tGSM227875\tGSM227879\tGSM
227881\tGSM227882\tGene title\tGene symbol\tGene ID\tUniGene title\tUniGene symbol\tUniG
ene ID\tNucleotide Title\tGI\tGenBank Accession\tPlatform_CLONEID\tPlatform_ORF\tPlatfor
m_SPOTID\tChromosome location\tChromosome annotation\tGO:Function\tGO:Process\tGO:Compon
ent\tGO:Function ID\tGO:Process ID\tGO:Component ID"

```

[79] "1007_s_at\tMIR4640\t5236.48\t6183.2\t5171.48\t4553.4\t6094.66\t5976.88\t7713.88\t6781.24\t4948.99\t5029.57\t5892.17\t6534.86\t8393.72\t6367.62\t8266.48\tmicroRNA 4640///discoidin domain receptor tyrosine kinase 1\tMIR4640///DDR1\t100616237///780\t\t\t\tHuman receptor tyrosine kinase DDR gene, complete cds\t1753221\tU48705\t\t\t\t6p21.3\tChromosome 6, NC_000006.12 (30890883..30890972)///Chromosome 6, NC_000006.12 (30880909..30900156)\tATP binding///collagen binding///collagen binding///metal ion binding///protein binding///protein tyrosine kinase collagen receptor activity///transmembrane receptor protein tyrosine kinase activity\tbranching involved in mammary gland duct morphogenesis///cell adhesion///collagen-activated tyrosine kinase receptor signaling pathway///collagen-activated tyrosine kinase receptor signaling pathway///ear development///embryo implantation///extracellular matrix organization///lactation///mammary gland alveolus development///negative regulation of cell proliferation///organ regeneration///peptidyl-tyrosine autophosphorylation///protein autophosphorylation///regulation of cell growth///regulation of cell-matrix adhesion///regulation of extracellular matrix disassembly///skin development///smooth muscle cell migration///smooth muscle cell-matrix adhesion///wound healing, spreading of cells\tbasolateral plasma membrane///extracellular exosome///extracellular space///integral component of plasma membrane///plasma membrane///receptor complex\tG0:0005524///G0:0005518///G0:0005518///G0:0046872///G0:0005515///G0:0038062///G0:0004714\tG0:0060444///G0:0007155///G0:0038063///G0:0038063///G0:0043583///G0:0007566///G0:0030198///G0:0007595///G0:0060749///G0:0008285///G0:0031100///G0:0038083///G0:0046777///G0:0001558///G0:0001952///G0:0010715///G0:0043588///G0:0014909///G0:0061302///G0:0044319\tG0:0016323///G0:0070062///G0:0005615///G0:0005887///G0:0005886///G0:0043235"

[80] "1053_at\tRFC2\t237.517\t225.653\t225.017\t134.267\t228.18\t235.723\t262.208\t254.174\t196.749\t188.746\t202.167\t174.622\t178.976\t220.643\t215.606\treplication factor C subunit 2\tRFC2\t5982\t\t\t\tHuman replication factor C, 40-kDa subunit (A1) mRNA, complete cds\t1590810\tM87338\t\t\t\t7q11.23\tChromosome 7, NC_000007.14 (74231502..74254458, complement)\tATP binding///contributes_to DNA clamp loader activity///enzyme binding///protein binding///contributes_to single-stranded DNA-dependent ATPase activity\tDNA damage response, detection of DNA damage///DNA replication///error-free translesion synthesis///error-prone translesion synthesis///nucleotide-excision repair, DNA gap filling///nucleotide-excision repair, DNA incision///nucleotide-excision repair, DNA incision, 5'-to lesion///positive regulation of DNA-directed DNA polymerase activity///regulation of signal transduction by p53 class mediator///telomere maintenance via recombination///transcription-coupled nucleotide-excision repair///translesion synthesis\tCtf18 RFC-like complex///DNA replication factor C complex///nucleoplasm\tG0:0005524///contributes_to G0:0003689///G0:0019899///G0:0005515///contributes_to G0:0043142\tG0:0042769///G0:0006260///G0:0070987///G0:0042276///G0:0006297///G0:0033683///G0:0006296///G0:1900264///G0:1901796///G0:0000722///G0:0006283///G0:0019985\tG0:0031390///G0:0005663///G0:0005654"

[81] "117_at\tHSPA6\t100.439\t159.248\t136.267\t45.6665\t166.132\t132.373\t47.6219\t127.556\t160.135\t125.743\t163.965\t51.9517\t140.141\t325.339\t140.865\theat shock protein family A (Hsp70) member 6\tHSPA6\t3310\t\t\t\tHuman heat-shock protein HSP70B' gene\t35221\tX51757\t\t\t\t1q23\tChromosome 1, NC_000001.11 (161524540..161526897)\tATP binding///ATPase activity, coupled///enzyme binding///heat shock protein binding///protein binding///unfolded protein binding\tNOT cellular heat acclimation///cellular response to heat///cellular response to heat///protein refolding///response to unfolded protein\tcolocalizes_with COP9 signalosome///blood microparticle///centriole///cytoplasm///cytosol///extracellular exosome\tG0:0005524///G0:0042623///G0:0019899///G0:0031072///G0:0005515///G0:00051082\tNOT G0:0070370///G0:0034605///G0:0034605///G0:0042026///G0:0006986\tcolocalizes_with G0:0008180///G0:0072562///G0:0005814///G0:0005737///G0:0005829///G0:0070062"

[82] "121_at\tPAX8\t1123.99\t1258\t923.859\t619.669\t1369.28\t1215.57\t1102.9\t1377.01\t1195.4\t1176.89\t959.101\t926.235\t984.344\t1850.01\t1200.36\tpaired box 8\tPAX8\t7849\t\t\t\tH.sapiens Pax8 mRNA\t38425\tX69699\t\t\t\t2q13\tChromosome 2, NC_000002.12 (11

3215997..113278950, complement)\tDNA binding///DNA binding///RNA polymerase II core promoter proximal region sequence-specific DNA binding///RNA polymerase II core promoter sequence-specific DNA binding///protein binding///thyroid-stimulating hormone receptor activity///transcription factor activity, sequence-specific DNA binding///transcription regulatory region DNA binding///transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding\tanatomical structure morphogenesis///branching involved in ureteric bud morphogenesis///cellular response to gonadotropin stimulus///central nervous system development///inner ear morphogenesis///kidney development///mesenchymal to epithelial transition involved in metanephros morphogenesis///mesonephros development///metanephric S-shaped body morphogenesis///metanephric comma-shaped body morphogenesis///metanephric distal convoluted tubule development///metanephric epithelium development///metanephric nephron tubule formation///negative regulation of apoptotic process involved in metanephric collecting duct development///negative regulation of apoptotic process involved in metanephric nephron tubule development///negative regulation of mesenchymal cell apoptotic process involved in metanephric nephron morphogenesis///negative regulation of mesenchymal cell apoptotic process involved in metanephros development///otic vesicle development///positive regulation of branching involved in ureteric bud morphogenesis///positive regulation of mesenchymal to epithelial transition involved in metanephros morphogenesis///positive regulation of metanephric DCT cell differentiation///positive regulation of thyroid hormone generation///positive regulation of transcription from RNA polymerase II promoter///positive regulation of transcription, DNA-templated///positive regulation of transcription, DNA-templated///pronephric field specification///pronephros development///regulation of apoptotic process///regulation of metanephric nephron tubule epithelial cell differentiation///regulation of thyroid-stimulating hormone secretion///sulfur compound metabolic process///thyroid gland development///thyroid gland development///thyroid-stimulating hormone signaling pathway///transcription from RNA polymerase II promoter///transcription, DNA-templated///urogenital system development\t nucleoplasm///nucleoplasm///nucleus\tG0:0003677///G0:0003677///G0:0000978///G0:0000979///G0:0005515///G0:0004996///G0:0003700///G0:0044212///G0:0001077\tG0:0009653///G0:0001658///G0:0071371///G0:0007417///G0:0042472///G0:0001822///G0:0003337///G0:0001823///G0:0072284///G0:0072278///G0:0072221///G0:0072207///G0:0072289///G0:1900215///G0:1900218///G0:0072305///G0:1900212///G0:0071599///G0:0090190///G0:0072108///G0:2000594///G0:2000611///G0:0045944///G0:0045893///G0:0045893///G0:0039003///G0:0048793///G0:0042981///G0:0072307///G0:2000612///G0:0006790///G0:0030878///G0:0030878///G0:0038194///G0:0006366///G0:0006351///G0:0001655\tG0:0005654///G0:0005654///G0:0005634"

[83] "1255_g_at\tGUCA1A\t34.7732\t9.23039\t8.90546\t29.3123\t60.9262\t55.368\t61.0796\t83.12\t37.0302\t30.8798\t31.6348\t4.08673\t64.7151\t15.3595\t29.2318\tguanylate cyclase activator 1A\tGUCA1A\t2978\t\t\t\t\tHomo sapiens guanylate cyclase activating protein (GCAP) gene exons 1-4, complete cds\t623404\tL36861\t\t\t\t\t6p21.1\tChromosome 6, NC_000006.12 (42155377..42180083)\tcalcium ion binding///calcium sensitive guanylate cyclase activator activity///guanylate cyclase regulator activity\tcellular response to calcium ion///phototransduction///positive regulation of guanylate cyclase activity///regulation of rhodopsin mediated signaling pathway///signal transduction///visual perception\tphotorceptor disc membrane///photoreceptor inner segment///plasma membrane\tG0:0005509///G0:0008048///G0:0030249\tG0:0071277///G0:0007602///G0:0031284///G0:0022400///G0:0007165///G0:0007601\tG0:0097381///G0:0001917///G0:0005886"

[84] "1294_at\tMIR5193\t348.855\t399.54\t429.155\t342\t1161.05\t486.014\t754.339\t760.1\t403.218\t478.731\t475.194\t393.897\t527.179\t469.237\t499.791\tmicroRNA 5193///ubiquitin like modifier activating enzyme 7\tMIR5193///UBA7\t100847079///7318\t\t\t\t\tHomo sapiens ubiquitin-activating enzyme E1 related protein mRNA, complete cds\t520832\tL13852\t\t\t\t\t3p21\tChromosome 3, NC_000003.12 (49806137..49806245, complement)///Chromosome 3, NC_000003.12 (49805205..49813958, complement)\tATP binding///ISG15 activating enzyme

activity///protein binding///ubiquitin activating enzyme activity///ubiquitin-protein transferase activity///ubiquitin-protein transferase activity\tISG15-protein conjugation///cellular protein modification process///modification-dependent protein catabolic process///negative regulation of type I interferon production///protein ubiquitination///translesion synthesis\tcytosol///cytosol///nucleoplasm///nucleus\tG0:0005524///G0:0019782///G0:0005515///G0:0004839///G0:0004842///G0:0004842\tG0:0032020///G0:0006464///G0:0019941///G0:0032480///G0:0016567///G0:0019985\tG0:0005829///G0:0005829///G0:0005654///G0:0005634"

[85] "1316_at\tTHRA\t118.301\t126.449\t116.141\t93.1706\t159.653\t128.066\t101.399\t116.583\t118.106\t85.1356\t147.734\t180.755\t83.9113\t314.811\t94.9472\tthyroid hormone receptor, alpha\tTHRA\t7067\t\t\t\t\tHomo sapiens mRNA for thyroid hormone receptor alpha 1 THRA1, (c-erbA-1 gene)\t29878\tX55005\t\t\t\t\t17q11.2\tChromosome 17, NC_000017.11 (40062193..40093867)\tTBP-class protein binding///chromatin DNA binding///intronic transcription regulatory region DNA binding///protein binding///protein complex binding///protein domain specific binding///protein heterodimerization activity///protein homodimerization activity///steroid hormone receptor activity///steroid receptor RNA activator RNA binding///thyroid hormone binding///thyroid hormone binding///thyroid hormone receptor activity///transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding///transcription factor activity, sequence-specific DNA binding///transcription factor binding///transcription regulatory region DNA binding///transcription regulatory region sequence-specific DNA binding///zinc ion binding\tType I pneumocyte differentiation///adrenal gland development///brain development///cartilage condensation///cytoplasmic sequestering of transcription factor///digestive tract development///embryonic organ development///erythrocyte differentiation///female courtship behavior///hormone-mediated signaling pathway///intracellular receptor signaling pathway///kidney development///learning or memory///liver development///negative regulation of DNA-templated transcription, initiation///negative regulation of RNA polymerase II transcriptional preinitiation complex assembly///negative regulation of sequence-specific DNA binding transcription factor activity///negative regulation of transcription, DNA-templated///ossification///positive regulation of female receptivity///positive regulation of myotube differentiation///positive regulation of transcription from RNA polymerase II promoter///regulation of heart contraction///regulation of lipid catabolic process///regulation of myeloid cell apoptotic process///regulation of thyroid hormone mediated signaling pathway///regulation of transcription from RNA polymerase II promoter///response to cold///response to drug///response to nutrient levels///steroid hormone mediated signaling pathway///thyroid gland development///transcription from RNA polymerase II promoter///transcription initiation from RNA polymerase II promoter\tcytosol///mitochondrion///nucleoplasm///nucleus\tG0:0017025///G0:0031490///G0:0044213///G0:0005515///G0:0032403///G0:0019904///G0:0046982///G0:0042803///G0:0003707///G0:0002153///G0:0070324///G0:0070324///G0:0004887///G0:0003705///G0:0003700///G0:0008134///G0:0044212///G0:0000976///G0:0008270\tG0:0060509///G0:0030325///G0:0007420///G0:0001502///G0:0042994///G0:0048565///G0:0048568///G0:0030218///G0:0008050///G0:0009755///G0:0030522///G0:0001822///G0:0007611///G0:0001889///G0:2000143///G0:0017055///G0:0043433///G0:0045892///G0:0001503///G0:0045925///G0:0010831///G0:0045944///G0:0008016///G0:0050994///G0:0033032///G0:0002155///G0:0006357///G0:0009409///G0:0042493///G0:0031667///G0:0043401///G0:0030878///G0:0006366///G0:0006367\tG0:0005829///G0:0005739///G0:0005654///G0:0005634"

[86] "1320_at\tPTPN21\t92.2018\t69.4191\t53.3477\t6.84072\t16.3806\t115.919\t23.3512\t108.671\t44.9332\t88.9449\t86.3159\t20.3402\t86.3494\t121.366\t74.5479\tprotein tyrosine phosphatase, non-receptor type 21\tPTPN21\t11099\t\t\t\t\tH.sapiens mRNA for protein-tyrosine-phosphatase D1\t532055\tX79510\t\t\t\t\t14q31.3\tChromosome 14, NC_000014.9 (88465778..88554779, complement)\tprotein binding///protein tyrosine phosphatase activity\tpeptidyl-tyrosine dephosphorylation///protein dephosphorylation\tcytoplasm///cytoskeleton\tG

0:0005515///G0:0004725\tG0:0035335///G0:0006470\tG0:0005737///G0:0005856"

[87] "1405_i_at\tCCL5\t409.241\t190.261\t881.393\t656.464\t379.966\t1109.02\t1017.92\t398.593\t1016.55\t1099.13\t316.669\t138.23\t864.094\t115.858\t773.374\tC-C motif chemo
kine ligand 5\tCCL5\t6352\t\t\t\t\tHuman T cell-specific protein (RANTES) mRNA, complete c
ds\t339420\tM21121\t\t\t\t\t17q12\tChromosome 17, NC_000017.11 (35871491..35880373, comple
ment)\tCCR1 chemokine receptor binding///CCR1 chemokine receptor binding///CCR1 chemokin
e receptor binding///CCR4 chemokine receptor binding///CCR5 chemokine receptor bindin
g///chemoattractant activity///chemokine activity///chemokine activity///chemokine recep
tor antagonist activity///chemokine receptor binding///heparin binding///phosphatidylinositol
phospholipase C activity///phospholipase activator activity///protein binding///pr
otein homodimerization activity///protein kinase activity///protein self-association///r
eceptor signaling protein tyrosine kinase activator activity\tG-protein coupled receptor
signaling pathway///MAPK cascade///activation of phospholipase D activity///aging///calc
ium ion transport///cell-cell signaling///cellular calcium ion homeostasis///cellular pr
otein complex assembly///cellular response to alkyl hydroperoxide///cellular response to
amino acid stimulus///cellular response to ethanol///cellular response to fibroblast gro
wth factor stimulus///cellular response to high density lipoprotein particle stimulus///
cellular response to interferon-gamma///cellular response to interleukin-1///cellular re
sponse to morphine///cellular response to organic cyclic compound///cellular response to
transforming growth factor beta stimulus///cellular response to tumor necrosis factor///
cellular response to vitamin K///chemokine-mediated signaling pathway///chemokine-medi
ated signaling pathway///chemotaxis///chronic inflammatory response///dendritic cell chemo
taxis///dibenzo-p-dioxin metabolic process///eosinophil chemotaxis///exocytosis///inflam
matory response///leukocyte cell-cell adhesion///lipopolysaccharide-mediated signaling p
athway///lymphocyte chemotaxis///macrophage chemotaxis///monocyte chemotaxis///negative
regulation by host of viral transcription///negative regulation of G-protein coupled rec
eptor protein signaling pathway///negative regulation of T cell apoptotic process///nega
tive regulation of chemokine-mediated signaling pathway///negative regulation of macroph
age apoptotic process///negative regulation of neuron death///negative regulation of vir
al genome replication///neutrophil activation///neutrophil chemotaxis///positive chemota
xis///positive regulation of ERK1 and ERK2 cascade///positive regulation of GTPase activ
ity///positive regulation of JAK-STAT cascade///positive regulation of T cell apoptotic
process///positive regulation of T cell chemotaxis///positive regulation of T cell migra
tion///positive regulation of T cell proliferation///positive regulation of activation o
f JAK2 kinase activity///positive regulation of angiogenesis///positive regulation of ca
lcium ion transport///positive regulation of cell adhesion///positive regulation of cell
migration///positive regulation of cell-cell adhesion mediated by integrin///positive re
gulation of cellular biosynthetic process///positive regulation of epithelial cell proli
feration///positive regulation of fever generation///positive regulation of homotypic ce
ll-cell adhesion///positive regulation of inflammatory response///positive regulation of
innate immune response///positive regulation of macrophage chemotaxis///positive regulat
ion of mast cell chemotaxis///positive regulation of monocyte chemotaxis///positive regu
lation of natural killer cell chemotaxis///positive regulation of neuron differentiatio
n///positive regulation of osteoclast differentiation///positive regulation of phosphati
dylinositol 3-kinase signaling///positive regulation of phosphorylation///positive regul
ation of protein tyrosine kinase activity///positive regulation of smooth muscle cell mi
gration///positive regulation of smooth muscle cell proliferation///positive regulation
of translational initiation///positive regulation of tyrosine phosphorylation of STAT pr
oteins///positive regulation of viral genome replication///protein kinase B signaling///p
rotein tetramerization///regulation of T cell activation///regulation of chronic inflamm
atory response///regulation of insulin secretion///regulation of neuron death///response
to activity///response to cholesterol///response to drug///response to estrogen///respon

se to glucocorticoid///response to insulin///response to salt stress///response to toxic substance///response to virus\tcytoplasm///extracellular region///extracellular space\tG0:0031726///G0:0031726///G0:0031726///G0:0031729///G0:0031730///G0:0042056///G0:0008009///G0:0008009///G0:0046817///G0:0042379///G0:0008201///G0:0004435///G0:0016004///G0:0005515///G0:0042803///G0:0004672///G0:0043621///G0:0030298\tG0:0007186///G0:0000165///G0:0031584///G0:0007568///G0:0006816///G0:0007267///G0:0006874///G0:0043623///G0:0071448///G0:0071230///G0:0071361///G0:0044344///G0:0071403///G0:0071346///G0:0071347///G0:0071315///G0:0071407///G0:0071560///G0:0071356///G0:0071307///G0:0070098///G0:0070098///G0:0006935///G0:0002544///G0:0002407///G0:0018894///G0:0048245///G0:0006887///G0:0006954///G0:0007159///G0:0031663///G0:0048247///G0:0048246///G0:0002548///G0:0043922///G0:0045744///G0:0070233///G0:0070100///G0:2000110///G0:1901215///G0:0045071///G0:0042119///G0:0030593///G0:0050918///G0:0070374///G0:0043547///G0:0046427///G0:0070234///G0:0010820///G0:2000406///G0:0042102///G0:0010535///G0:0045766///G0:0051928///G0:0045785///G0:0030335///G0:0033634///G0:0031328///G0:0050679///G0:0031622///G0:0034112///G0:0050729///G0:0045089///G0:0010759///G0:0060754///G0:0090026///G0:2000503///G0:0045666///G0:0045672///G0:0014068///G0:0042327///G0:0061098///G0:0014911///G0:0048661///G0:0045948///G0:0042531///G0:0045070///G0:0043491///G0:0051262///G0:0050863///G0:0002676///G0:0050796///G0:1901214///G0:0014823///G0:0070723///G0:0042493///G0:0043627///G0:0051384///G0:0032868///G0:0009651///G0:0009636///G0:0009615\tG0:0005737///G0:0005576///G0:0005615"

[88] "1431_at\tCYP2E1\t107.18\t63.6089\t54.3381\t29.4182\t72.1769\t57.7296\t50.9251\t158.483\t64.3109\t40.7661\t104.083\t49.638\t74.0772\t134.333\t31.5311\tcytochrome P450 family 2 subfamily E member 1\tCYP2E1\t1571\t\t\t\tHuman cytochrome P450IIE1 (ethanol-inducible) gene, complete cds\t181355\tJ02843\t\t\t\t10q26.3\tChromosome 10, NC_000010.11 (133527363..133539116)\tarachidonic acid epoxygenase activity///enzyme binding///heme binding///iron ion binding///monooxygenase activity///oxidoreductase activity///oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, NAD(P)H as one donor, and incorporation of one atom of oxygen///oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen///oxygen binding///steroid hydroxylase activity\tbenzene metabolic process///carbon tetrachloride metabolic process///drug metabolic process///drug metabolic process///drug metabolic process///epoxygenase P450 pathway///halogenated hydrocarbon metabolic process///heterocycle metabolic process///monoterpenoid metabolic process///oxidation-reduction process///response to drug///response to ethanol///response to organonitrogen compound///response to ozone///steroid metabolic process///triglyceride metabolic process///xenobiotic metabolic process\tGolgi membrane///endoplasmic reticulum membrane///intrinsic component of endoplasmic reticulum membrane///mitochondrion\tG0:0008392///G0:0019899///G0:0020037///G0:0005506///G0:0004497///G0:0016491///G0:0016709///G0:0016712///G0:0019825///G0:0008395\tG0:0018910///G0:0018885///G0:0017144///G0:0017144///G0:0017144///G0:0019373///G0:0042197///G0:0046483///G0:0016098///G0:0055114///G0:0042493///G0:0045471///G0:0010243///G0:0010193///G0:0008202///G0:0006641///G0:0006805\tG0:0000139///G0:0005789///G0:0031227///G0:0005739"

[89] "1438_at\tEPHB3\t495.203\t476.729\t322.715\t38.7511\t324.719\t400.486\t425.511\t473.014\t526.462\t323.132\t418.648\t226.17\t730.825\t366.646\t410.78\tEPH receptor B3\tEPHB3\t2049\t\t\t\tH.sapiens HEK2 mRNA for protein tyrosine kinase receptor\t406867\tX75208\t\t\t\t3q27.1\tChromosome 3, NC_000003.12 (184561799..184582408)\tATP binding///axon guidance receptor activity///ephrin receptor activity\tangiogenesis///axon guidance///axonal fasciculation///cell migration///central nervous system projection neuron axonogenesis///corpus callosum development///dendritic spine development///dendritic spine morphogenesis///digestive tract morphogenesis///ephrin receptor signaling pathway///ephrin receptor signaling pathway///palate development///peptidyl-tyrosine phosphorylation///positive regulation of synapse assembly///protein autophosphorylation///regulation o

f GTPase activity///regulation of axonogenesis///regulation of cell-cell adhesion///retinal ganglion cell axon guidance///substrate adhesion-dependent cell spreading///thymus development///urogenital system development\tcytosol///dendrite///extracellular region///integral component of plasma membrane///plasma membrane\tGO:0005524///GO:0008046///GO:0005003\tGO:0001525///GO:0007411///GO:0007413///GO:0016477///GO:0021952///GO:0022038///GO:0060996///GO:0060997///GO:0048546///GO:0048013///GO:0048013///GO:0060021///GO:0018108///GO:0051965///GO:0046777///GO:0043087///GO:0050770///GO:0022407///GO:0031290///GO:0034446///GO:0048538///GO:0001655\tGO:0005829///GO:0030425///GO:0005576///GO:0005887///GO:0005886"

[90] "1487_at\tESRRA\t1027.78\t959.232\t858.273\t391.557\t1252.2\t769.115\t755.587\t1106.29\t871.631\t626.426\t842.376\t862.349\t884.817\t1004.41\t660.146\ttestosterone related receptor alpha\tESRRA\t2101\t\t\t\t\tHuman estrogen receptor-related protein (hERRa1) mRNA, 3' end, partial cds\t1239956\tL38487\t\t\t\t\t11q13\tChromosome 11, NC_000011.10 (64305528..64316743)\tDNA binding///RNA polymerase II core promoter proximal region sequence-specific DNA binding///RNA polymerase II transcription factor activity, ligand-activated sequence-specific DNA binding///protein binding///protein domain specific binding///sequence-specific DNA binding///steroid binding///steroid hormone receptor activity///transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding///transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding///zinc ion binding\tcartilage development///intracellular receptor signaling pathway///mitochondrion organization///negative regulation of transcription from RNA polymerase II promoter///positive regulation of cellular response to insulin stimulus///positive regulation of transcription from RNA polymerase II promoter///regulation of cell proliferation///regulation of osteoblast differentiation///regulation of osteoclast differentiation///regulation of transcription, DNA-templated///response to estradiol///steroid hormone mediated signaling pathway///transcription initiation from RNA polymerase II promoter\tintercellular bridge///microtubule cytoskeleton///nucleolus///nucleoplasm///nucleus\tGO:0003677///GO:0000978///GO:0004879///GO:0005515///GO:0019904///GO:0043565///GO:0005496///GO:0003707///GO:0001077///GO:0001078///GO:0008270\tGO:0051216///GO:0030522///GO:0007005///GO:0000122///GO:1900078///GO:0045944///GO:0042127///GO:0045667///GO:0045670///GO:0006355///GO:0032355///GO:0043401///GO:0006367\tGO:0045171///GO:0015630///GO:0005730///GO:0005654///GO:0005634"

[91] "1494_f_at\tCYP2A6\t2283.38\t2862.55\t2038.42\t1070.87\t4039.22\t1265.16\t2577.36\t5929.89\t1775.23\t1571.06\t2438.24\t4159.15\t2068.03\t4016.63\t1410.29\tcytochrome P450 family 2 subfamily A member 6\tCYP2A6\t1548\t\t\t\t\tHuman cytochrome P450IIA3 (CYP2A3) mRNA, complete cds\t180986\tM33318\t\t\t\t\t19q13.2\tChromosome 19, NC_000019.10 (40843538..40850447, complement)\tarachidonic acid epoxygenase activity///coumarin 7-hydroxylase activity///coumarin 7-hydroxylase activity///enzyme binding///heme binding///iron ion binding///oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen///steroid hydroxylase activity\tcoumarin catabolic process///coumarin metabolic process///coumarin metabolic process///drug metabolic process///drug metabolic process///epoxygenase P450 pathway///exogenous drug catabolic process///oxidation-reduction process///steroid metabolic process\tcytoplasmic microtubule///endoplasmic reticulum membrane///organelle membrane\tGO:0008392///GO:0008389///GO:0008389///GO:0019899///GO:0020037///GO:0005506///GO:0016712///GO:0008395\tGO:0046226///GO:0009804///GO:0009804///GO:0017144///GO:0017144///GO:0019373///GO:0042738///GO:0055114///GO:0008202\tGO:0005881///GO:0005789///GO:0031090"

[92] "1598_g_at\tGAS6\t783.803\t831.405\t728.331\t421.12\t1105.62\t692.44\t1223.94\t1223.01\t787.175\t538.613\t651.041\t717.075\t803.55\t1373.57\t762.996\tgrowth arrest specific 6\tGAS6\t2621\t\t\t\t\tHomo sapiens growth-arrest-specific protein (gas) mRNA, complete cds\t401766\tL13720\t\t\t\t\t13q34\tChromosome 13, NC_000013.11 (113820549..113864103,

complement)\tbinding, bridging///calcium ion binding///cysteine-type endopeptidase inhibitor activity involved in apoptotic process///phosphatidylserine binding///protein binding///protein tyrosine kinase activator activity///receptor agonist activity///receptor binding///receptor binding///receptor tyrosine kinase binding///voltage-gated calcium channel activity\tB cell chemotaxis///ER to Golgi vesicle-mediated transport///activation of protein kinase B activity///apoptotic cell clearance///apoptotic process///calcium ion transmembrane transport///cell adhesion///cell cycle arrest///cell migration///cell proliferation///cell-substrate adhesion///cellular response to drug///cellular response to glucose stimulus///cellular response to growth factor stimulus///cellular response to interferon-alpha///cellular response to starvation///cellular response to vitamin K///dendritic cell differentiation///enzyme linked receptor protein signaling pathway///extracellular matrix assembly///fusion of virus membrane with host plasma membrane///hematopoietic stem cell migration to bone marrow///leukocyte migration///macrophage cytokine production///negative regulation of apoptotic process///negative regulation of biomineral tissue development///negative regulation of cysteine-type endopeptidase activity involved in apoptotic process///negative regulation of dendritic cell apoptotic process///negative regulation of endothelial cell apoptotic process///negative regulation of fibroblast apoptotic process///negative regulation of interferon-gamma production///negative regulation of interleukin-1 secretion///negative regulation of interleukin-6 production///negative regulation of interleukin-6 secretion///negative regulation of oligodendrocyte apoptotic process///negative regulation of protein import into nucleus, translocation///negative regulation of renal albumin absorption///negative regulation of sequence-specific DNA binding transcription factor activity///negative regulation of transcription, DNA-template d///negative regulation of tumor necrosis factor production///negative regulation of tumor necrosis factor-mediated signaling pathway///neuron migration///organ regeneration///peptidyl-serine phosphorylation///phagocytosis///platelet activation///platelet aggregation///platelet degranulation///positive regulation of ERK1 and ERK2 cascade///positive regulation of TOR signaling///positive regulation of cytokine-mediated signaling pathway///positive regulation of dendritic cell chemotaxis///positive regulation of fibroblast proliferation///positive regulation of gene expression///positive regulation of glomerular filtration///positive regulation of natural killer cell differentiation///positive regulation of peptidyl-serine phosphorylation///positive regulation of phagocytosis///positive regulation of protein export from nucleus///positive regulation of protein kinase B signaling///positive regulation of protein kinase activity///positive regulation of protein phosphorylation///positive regulation of protein phosphorylation///positive regulation of protein tyrosine kinase activity///protein kinase B signaling///protein phosphorylation///protein targeting to plasma membrane///receptor-mediated virion attachment to host cell///regulation of growth///signal peptide processing///signal transduction///signal transduction///viral entry into host cell///viral genome replication\tGolgi lumen///cytoplasm///endoplasmic reticulum lumen///extracellular exosome///extracellular region///extracellular space///platelet alpha granule lumen\tG0:0060090///G0:0005509///G0:0043027///G0:0001786///G0:0005515///G0:0030296///G0:0048018///G0:0005102///G0:0005102///G0:0030971///G0:0005245\tG0:0035754///G0:0006888///G0:0032148///G0:0043277///G0:0006915///G0:0070588///G0:0007155///G0:0007050///G0:0016477///G0:0008283///G0:0031589///G0:0035690///G0:0071333///G0:0071363///G0:0035457///G0:0009267///G0:0071307///G0:0097028///G0:0007167///G0:0085029///G0:0019064///G0:0097241///G0:0050900///G0:0010934///G0:0043066///G0:0070168///G0:0043154///G0:2000669///G0:2000352///G0:2000270///G0:0032689///G0:0050711///G0:0032715///G0:1900165///G0:1900142///G0:0033159///G0:2000533///G0:0043433///G0:0045892///G0:0032720///G0:0010804///G0:0001764///G0:0031100///G0:0018105///G0:0006909///G0:0030168///G0:0070527///G0:0002576///G0:0070374///G0:0032008///G0:0001961///G0:2000510///G0:0048146///G0:0010628///G0:0003104///G0:0032825///G0:0033138///G0:0050766///G0:0046827///G0:0051897///G0:0045860///G0:0001934///G0:0001934///G0:0061098///G0:0043491///G0:0006468///

G0:0072661///G0:0046813///G0:0040008///G0:0006465///G0:0007165///G0:0007165///G0:0046718///G0:0019079\tG0:0005796///G0:0005737///G0:0005788///G0:0070062///G0:0005576///G0:0005615///G0:0031093"

[93] "160020_at\tMMP14\t1052.08\t977.155\t874.718\t523.753\t1554.58\t962.912\t839.85\t766.389\t986.378\t755.154\t734.218\t893.406\t1064.24\t1539.16\t837.68\tmatrix metalloproteinase 14\tMMP14\t4323\t\t\t\tH.sapiens mRNA for membrane-type matrix metalloproteinase 1\t963053\tZ48481\t\t\t\t14q11.2\tChromosome 14, NC_000014.9 (22836533..22847600)\tcalcium ion binding///integrin binding///metalloendopeptidase activity///peptidase activator activity///protein binding///serine-type endopeptidase activity///zinc ion binding\tangiogenesis///astrocyte cell migration///branching morphogenesis of an epithelial tube///chondrocyte proliferation///collagen catabolic process///craniofacial suture morphogenesis///embryonic cranial skeleton morphogenesis///endochondral ossification///endodermal cell differentiation///endothelial cell proliferation///extracellular matrix disassembly///lung development///male gonad development///negative regulation of Notch signaling pathway///negative regulation of focal adhesion assembly///ovarian follicle development///positive regulation of B cell differentiation///positive regulation of cell growth///positive regulation of cell migration///positive regulation of myotube differentiation///positive regulation of peptidase activity///proteolysis///response to estrogen///response to hormone///response to hypoxia///response to mechanical stimulus///response to organic cyclic compound///response to oxidative stress///tissue remodeling///zymogen activation\tGolgi lumen///colocalizes_with cytoplasm///cytoplasmic vesicle///extracellular matrix///focal adhesion///integral component of plasma membrane///macropinosome///melanosome///plasma membrane///plasma membrane\tG0:0005509///G0:0005178///G0:0004222///G0:0016504///G0:0005515///G0:0004252///G0:0008270\tG0:0001525///G0:0043615///G0:0048754///G0:0035988///G0:0030574///G0:0097094///G0:0048701///G0:0001958///G0:0035987///G0:0001935///G0:0022617///G0:0030324///G0:0008584///G0:0045746///G0:0051895///G0:0001541///G0:0045579///G0:0030307///G0:0030335///G0:0010831///G0:0010952///G0:0006508///G0:0043627///G0:0009725///G0:0001666///G0:0009612///G0:0014070///G0:0006979///G0:0048771///G0:0031638\tG0:0005796///colocalizes_with G0:0005737///G0:0031410///G0:0031012///G0:0005925///G0:0005887///G0:0044354///G0:0042470///G0:0005886///G0:0005886"

[94] "1729_at\tTRADD\t435.281\t721.096\t639.106\t794.292\t641.425\t853.654\t821.863\t1196.12\t532.481\t661.108\t797.306\t1353.53\t959.422\t802.189\t931.266\tTNFRSF1A associated via death domain\tTRADD\t8717\t\t\t\tHomo sapiens TNF receptor-1 associated protein (TRADD) mRNA, 3' end of cds\t808914\tL41690\t\t\t\t16q22\tChromosome 16, NC_000016.10 (67154185..67159909, complement)\tbinding, bridging///death domain binding///identical protein binding///kinase binding///protein binding///protein complex binding///signal transducer activity///tumor necrosis factor receptor binding\tI-kappaB kinase/NF-kappaB signaling///activation of cysteine-type endopeptidase activity involved in apoptotic process///activation of cysteine-type endopeptidase activity involved in apoptotic signaling pathway///apoptotic process///death-inducing signaling complex assembly///extrinsic apoptotic signaling pathway///extrinsic apoptotic signaling pathway///extrinsic apoptotic signaling pathway via death domain receptors///negative regulation of extrinsic apoptotic signaling pathway via death domain receptors///positive regulation of I-kappaB kinase/NF-kappaB signaling///positive regulation of NF-kappaB transcription factor activity///positive regulation of apoptotic process///positive regulation of hair follicle development///protein heterooligomerization///regulation of extrinsic apoptotic signaling pathway via death domain receptors///regulation of necrotic cell death///regulation of tumor necrosis factor-mediated signaling pathway///signal transduction///tumor necrosis factor-mediated signaling pathway\tcytoplasm///cytoskeleton///cytosol///death-inducing signaling complex///membrane raft///nucleus///plasma membrane///receptor complex\tG0:0060090///G0:0070513///G0:0042802///G0:0019900///G0:0005515///G0:0032403///G0:0004871///G0:0005164\tG0:0007249///G0:0006919///G0:0097296///G0:0006915///G0:0071550///G0:0097191///G0:009719

1//G0:0008625//G0:1902042//G0:0043123//G0:0051092//G0:0043065//G0:0051798//G0:0051291//G0:1902041//G0:0010939//G0:0010803//G0:0007165//G0:0033209\tG0:0005737//G0:0005856//G0:0005829//G0:0031264//G0:0045121//G0:0005634//G0:0005886//G0:0043235"

[95] "1773_at\tCHURC1-FNTB\t195.867\t178.351\t215.99\t50.1504\t108.442\t160.581\t13.2086\t69.1939\t230.154\t151.162\t141.482\t143.068\t109.593\t202.974\t38.6318\tCHURC1-FNTB readthrough//farnesyltransferase, CAAX box, beta\tCHURC1-FNTB//FNTB\t100529261//2342\t\t\t\tHuman farnesyl-protein transferase beta-subunit mRNA, complete cds\t292032\tL00635\t\t\t\t14q23//14q23.3\tChromosome 14, NC_000014.9 (64914361..65062655)//Chromosome 14, NC_000014.9 (64986789..65062652)\tdrug binding//farnesyltransferase activity//isoprenoid binding//peptide binding//protein binding//protein farnesyltransferase activity//contributes_to protein farnesyltransferase activity//zinc ion binding\tnegative regulation of cell proliferation//positive regulation of cell cycle//positive regulation of fibroblast proliferation//positive regulation of nitric-oxide synthase biosynthetic process//protein farnesylation//regulation of rhodopsin mediated signaling pathway//response to cytokine//response to inorganic substance//response to organic cyclic compound//wound healing\tcytosol//microtubule associated complex//protein farnesyltransferase complex\tG0:0008144//G0:0004311//G0:0019840//G0:0042277//G0:0005515//G0:0004660//contributes_to G0:0004660//G0:0008270\tG0:0008285//G0:0045787//G0:0048146//G0:0051770//G0:0018343//G0:0022400//G0:0034097//G0:0010035//G0:0014070//G0:0042060\tG0:0005829//G0:0005875//G0:0005965"

[96] "177_at\tPLD1\t138.707\t181.968\t181.743\t134.871\t171.701\t192.754\t128.885\t116.831\t181.198\t183.091\t173.174\t93.3146\t128.687\t244.645\t128.321\tphospholipase D1\tPLD1\t5337\t\t\t\tHuman ARF-activated phosphatidylcholine-specific phospholipase D1a (hPLD1) mRNA, complete cds\t1185462\tU38545\t\t\t\t3q26\tChromosome 3, NC_000003.12 (171600404..171810494, complement)\tN-acylphosphatidylethanolamine-specific phospholipase D activity//phosphatidylinositol binding//phospholipase D activity//protein binding\tRas protein signal transduction//cell motility//chemotaxis//inositol lipid-mediated signaling//lipid catabolic process//phosphatidic acid biosynthetic process//regulation of microvillus assembly//small GTPase mediated signal transduction\tGolgi apparatus//Golgi membrane//apical plasma membrane//endocytic vesicle//endoplasmic reticulum membrane//endosome//late endosome membrane//lysosomal membrane//membrane//perinuclear region of cytoplasm\tG0:0070290//G0:0035091//G0:0004630//G0:0005515\tG0:0007265//G0:0048870//G0:0006935//G0:0048017//G0:0016042//G0:0006654//G0:0032534//G0:0007264\tG0:0005794//G0:0000139//G0:0016324//G0:0030139//G0:0005789//G0:0005768//G0:0031902//G0:0005765//G0:0016020//G0:0048471"

[97] "179_at\tDTX2P1-UPK3BP1-PMS2P11\t1170.57\t1284.9\t1095.1\t474.946\t1703.85\t1047.64\t1136.81\t1105.29\t982.639\t872.834\t1085.9\t1036.73\t877.443\t1608.91\t1123.87\tDTX2P1-UPK3BP1-PMS2P11 readthrough, transcribed pseudogene\tDTX2P1-UPK3BP1-PMS2P11\t441263\t\t\t\t\tHuman PMS2 related (hPMSR6) mRNA, complete cds\t1061425\tU38980\t\t\t\t7q11.23\tChromosome 7, NC_000007.14 (76980822..77023761)\t\t\t\t\t"

[98] "1861_at\tBAD\t267.064\t379.411\t295.774\t135.804\t241.132\t386.543\t479.865\t436.63\t364.257\t299.227\t291.232\t467.637\t384.744\t391.795\t409.989\tBCL2 associated agonist of cell death\tBAD\t572\t\t\t\t\tHuman Bcl-2 binding component 6 (bbc6) mRNA, complete cds\t1683636\tU66879\t\t\t\t11q13.1\tChromosome 11, NC_000011.10 (64269828..64284704, complement)\t14-3-3 protein binding//cysteine-type endopeptidase activator activity involved in apoptotic process//lipid binding//phospholipid binding//protein binding//protein heterodimerization activity//protein kinase B binding//protein kinase binding//protein phosphatase 2B binding\tADP metabolic process//ATP metabolic process//activation of cysteine-type endopeptidase activity//activation of cysteine-type endopeptidase activity involved in apoptotic process//apoptotic process//apoptotic process//cellular process regulating host cell cycle in response to virus//cellular response to chromosome//cellular response to hypoxia//cellular response to lipid//cellular response to mec

hanical stimulus///cellular response to nicotine///cerebral cortex development///cytokine-mediated signaling pathway///extrinsic apoptotic signaling pathway///extrinsic apoptotic signaling pathway in absence of ligand///extrinsic apoptotic signaling pathway via death domain receptors///glucose catabolic process///glucose homeostasis///intrinsic apoptotic signaling pathway///intrinsic apoptotic signaling pathway in response to DNA damage///negative regulation of cytolysis///pore complex assembly///positive regulation of B cell differentiation///positive regulation of T cell differentiation///positive regulation of apoptotic process///positive regulation of apoptotic process///positive regulation of apoptotic process///positive regulation of apoptotic process by virus///positive regulation of autophagy///positive regulation of cysteine-type endopeptidase activity involved in apoptotic process///positive regulation of epithelial cell proliferation///positive regulation of glucokinase activity///positive regulation of insulin secretion///positive regulation of insulin secretion involved in cellular response to glucose stimulus///positive regulation of intrinsic apoptotic signaling pathway///positive regulation of mitochondrial membrane potential///positive regulation of neuron death///positive regulation of protein insertion into mitochondrial membrane involved in apoptotic signaling pathway///positive regulation of proteolysis///positive regulation of release of cytochrome c from mitochondria///positive regulation of type B pancreatic cell development///protein insertion into mitochondrial membrane involved in apoptotic signaling pathway///regulation of mitochondrial membrane permeability///release of cytochrome c from mitochondria///response to amino acid///response to calcium ion///response to drug///response to estradiol///response to ethanol///response to glucocorticoid///response to glucose///response to hydrogen peroxide///response to oleic acid///response to progesterone///response to testosterone///spermatogenesis///suppression by virus of host apoptotic process///type B pancreatic cell proliferation\tcytosol///cytosol///mitochondrial outer membrane///mitochondrial outer membrane///mitochondrion\tG0:0071889///G0:0008656///G0:0008289///G0:0005543///G0:0005515///G0:0046982///G0:0043422///G0:0019901///G0:0030346\tG0:0046031///G0:0046034///G0:0097202///G0:0006919///G0:0006915///G0:0006915///G0:0060154///G0:0071247///G0:0071456///G0:0071396///G0:0071260///G0:0071316///G0:0021987///G0:0019221///G0:0097191///G0:0097192///G0:0008625///G0:0006007///G0:0042593///G0:0097193///G0:0008630///G0:0045918///G0:0046931///G0:0045579///G0:0045582///G0:0043065///G0:0043065///G0:0043065///G0:0060139///G0:0010508///G0:0043280///G0:0050679///G0:0033133///G0:0032024///G0:0035774///G0:2001244///G0:0010918///G0:1901216///G0:1900740///G0:0045862///G0:0090200///G0:2000078///G0:0001844///G0:0046902///G0:0001836///G0:0043200///G0:0051592///G0:0042493///G0:0032355///G0:0045471///G0:0051384///G0:0009749///G0:0042542///G0:0034201///G0:0032570///G0:0033574///G0:0007283///G0:0019050///G0:0044342\tG0:0005829///G0:0005829///G0:0005741///G0:0005741///G0:0005739"

[99] "200000_s_at\tPRPF8\t2118.87\t2581.11\t2310.84\t1743.7\t1936.04\t1778.86\t2794.44\t2439.34\t2390.7\t2341.83\t2177.57\t1045.64\t2509.83\t1704.38\t2204.12\tpre-mRNA processing factor 8\tPRPF8\t10594\t\t\t\t\tHomo sapiens pre-mRNA processing factor 8 (PRPF8), mRNA\t91208425\tNM_006445\t\t\t\t\t17p13.3\tChromosome 17, NC_000017.11 (1650629..1684882, complement)\tK63-linked polyubiquitin binding///U1 snRNA binding///U2 snRNA binding///U5 snRNA binding///U6 snRNA binding///poly(A) RNA binding///pre-mRNA intronic binding///protein binding///second spliceosomal transesterification activity\tRNA splicing///RNA splicing, via transesterification reactions///cellular response to lipopolysaccharide///cellular response to tumor necrosis factor///mRNA processing///mRNA splicing, via spliceosome///mRNA splicing, via spliceosome///spliceosomal tri-snRNP complex assembly\tU5 snRNP///catalytic step 2 spliceosome///membrane///nuclear speck///nucleoplasm///nucleus\tG0:0070530///G0:0030619///G0:0030620///G0:0030623///G0:0017070///G0:0044822///G0:0097157///G0:0005515///G0:0000386\tG0:0008380///G0:0000375///G0:0071222///G0:0071356///G0:0006397///G0:0000398///G0:0000398///G0:0000244\tG0:0005682///G0:0071013///G0:0016020///G0:0016607///G0:0005654///G0:0005634"

```
## [100] "200001_at\tCAPNS1\t9604.38\t10668.3\t5951.24\t1981.07\t5611.53\t7958.8\t8330.7
4\t7481.43\t9082.14\t9425.3\t10121.3\t6900.9\t12546.1\t8644.76\t11390.2\tcalpain small s
ubunit 1\tCAPNS1\t826\t\t\t\t\tHomo sapiens calpain small subunit 1 (CAPNS1), transcript v
ariant 1, mRNA\t701461510\tNM_001749\t\t\t\t\t19q13.12\tChromosome 19, NC_000019.10 (36139
926..36150353)\tcalcium ion binding///calcium-dependent cysteine-type endopeptidase acti
vity///protein binding\textracellular matrix disassembly///positive regulation of cell p
roliferation///proteolysis///regulation of macroautophagy\tcytosol///cytosol///extracell
ular exosome///membrane///plasma membrane\tG0:0005509///G0:0004198///G0:0005515\tG0:0022
617///G0:0008284///G0:0006508///G0:0016241\tG0:0005829///G0:0005829///G0:0070062///G0:00
16020///G0:0005886"
```

```
start_line <- grep("!dataset_table_begin", raw_data)
my_data <- read.delim(file_path, skip = start_line, header = TRUE)

expression_data <- my_data %>% select(starts_with("GSM"))

identifier_column <- my_data["IDENTIFIER"]

# Normalize the data
normalized_data <- normalizeBetweenArrays(as.matrix(expression_data))

expression_df <- cbind(identifier_column, expression_data)

# Sample GSM code status mapping as a vector
gsm_status <- c(
  "GSM227868" = "Never Smoker",
  "GSM227870" = "Never Smoker",
  "GSM227871" = "Never Smoker",
  "GSM227874" = "Never Smoker",
  "GSM227876" = "Never Smoker",
  "GSM227877" = "Never Smoker",
  "GSM227878" = "Never Smoker",
  "GSM227880" = "Never Smoker",
  "GSM227869" = "Current Smoker",
  "GSM227872" = "Current Smoker",
  "GSM227873" = "Current Smoker",
  "GSM227875" = "Current Smoker",
  "GSM227879" = "Current Smoker",
  "GSM227881" = "Current Smoker",
  "GSM227882" = "Current Smoker"
)

# Ensure that the GSM codes in expression_df match the order of gsm_status
gsm_codes <- colnames(expression_df) # Exclude the 'IDENTIFIER' column

# Create a new row of smoker status matching the GSM codes
smoker_status <- sapply(gsm_codes, function(gsm) gsm_status[gsm])

# Add the smoker status as a new row to the dataframe
expression_df_with_status <- rbind(smoker_status, expression_df)

# Reshaping data for plotting purposes
long_df <- expression_df %>%
  pivot_longer(cols = starts_with("GSM"),
    names_to = "Sample",
    values_to = "Expression")

#Decided to take top 25 genes for better visualization purposes
#Steps to take top 25 genes:
#One, Calculate the mean expression for each gene
top_genes <- long_df %>%
```



```

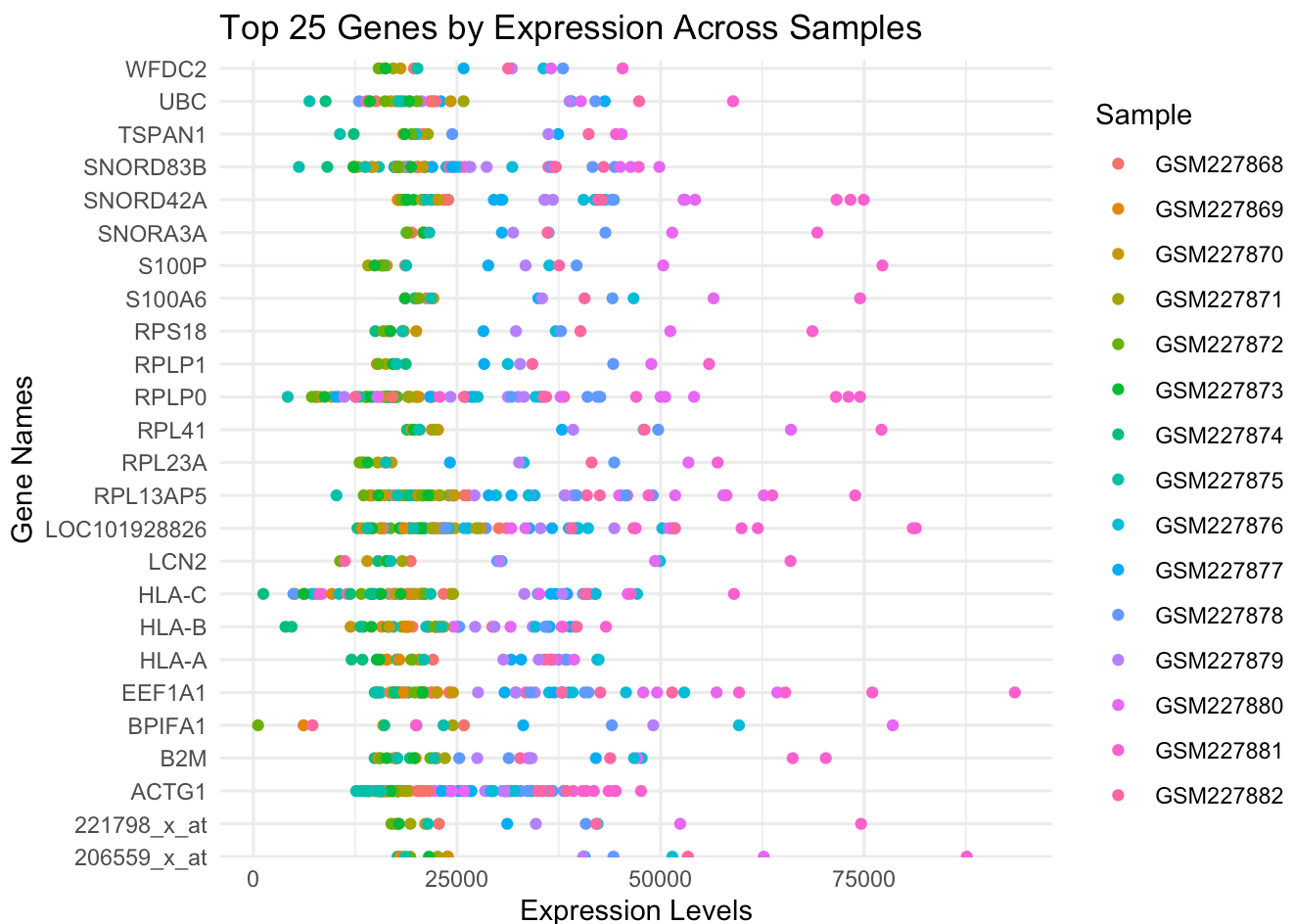
group_by(IDENTIFIER) %>%
  summarize(MeanExpression = mean(Expression, na.rm = TRUE)) %>%
  top_n(25, MeanExpression) %>%
  pull(IDENTIFIER)

#Two, Filter the long dataframe to include only the top genes
filtered_long_df <- long_df %>%
  filter(IDENTIFIER %in% top_genes)

#Three, plot with filtered data
#Scatter plot with geom_point

ggplot(filtered_long_df, aes(x = IDENTIFIER, y = Expression, color = Sample)) +
  geom_point() +
  scale_x_discrete(expand = expansion(mult = c(0.001, 0.01))) + # Adjust the space between labels
  theme(axis.text.x = element_text(angle = 90,
                                     vjust = 0.5,
                                     hjust = 1,
                                     margin = margin(t = 15))) +
  labs(title = "Top 25 Genes by Expression Across Samples",
       x = "Gene Names",
       y = "Expression Levels") +
  coord_flip() + # Optionally flip the coordinates
  theme_minimal()

```



```
#Alternative visualization: Table
#I've commented this out as I prefer the scatter plot
#top_25_genes <- filtered_long_df %>%
#  select(IDENTIFIER, Sample, Expression) %>%
#  arrange(IDENTIFIER, Sample)
#print(top_25_genes)
#save to CSV
#write.csv(top_25_genes_table, "top_25_genes_expression_table.csv", row.names = FALSE)
```

#Principle Component Analysis

```
# Scale the expression data
expression_data_scaled <- scale(expression_data)

# Replace infinite values with 0
expression_data_scaled[is.infinite(expression_data_scaled)] <- 0

# Replace missing (NA) values with 0
expression_data_scaled[is.na(expression_data_scaled)] <- 0

# Check the dimensions of the scaled data to ensure it's still valid
cat("Dimensions of the scaled expression data:", dim(expression_data_scaled), "\n")
```

```
## Dimensions of the scaled expression data: 22278 15
```

```
# Perform PCA
pca_result <- prcomp(expression_data_scaled, center = TRUE, scale. = TRUE)

# Print PCA summary to check the result
summary(pca_result)
```

```
## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation    3.7379 0.67714 0.3529 0.33913 0.2511 0.22946 0.22106
## Proportion of Variance 0.9315 0.03057 0.0083 0.00767 0.0042 0.00351 0.00326
## Cumulative Proportion 0.9315 0.96202 0.9703 0.97799 0.9822 0.98570 0.98896
##              PC8      PC9      PC10     PC11     PC12     PC13     PC14
## Standard deviation    0.20233 0.18110 0.15753 0.14637 0.12975 0.1098 0.09324
## Proportion of Variance 0.00273 0.00219 0.00165 0.00143 0.00112 0.0008 0.00058
## Cumulative Proportion 0.99169 0.99387 0.99553 0.99696 0.99808 0.9989 0.99946
##              PC15
## Standard deviation    0.08979
## Proportion of Variance 0.00054
## Cumulative Proportion 1.00000
```

```
# Step 4: Extract PCA Results and Add Metadata
# Create a data frame with PCA results
pca_df <- as.data.frame(pca_result$x)

colnames(pca_df)[1:15] <- colnames(expression_data)

# Remove the 'Sample' column (which is the 16th column in this case)
#pca_df <- pca_df %>% select(-Sample)

gsm_samples <- colnames(pca_df)[1:15]

# Create a new row corresponding to SmokerStatus using the gsm_status mapping
smoker_status_row <- gsm_status[gsm_samples]

# Combine the SmokerStatus row with the existing pca_df
# We use rbind to add the SmokerStatus row to the top of the dataframe
pca_df <- rbind(SmokerStatus = smoker_status_row, pca_df)

#pca_df <- pca_df %>% slice(-22280)
# Check the result to ensure SmokerStatus is added correctly at the top
#print(pca_df)

# Assuming pca_df already contains your PCA results with SmokerStatus as the first row

# Extract the PCA coordinates (PC1, PC2, etc.) and SmokerStatus
# The SmokerStatus is the first row, so let's separate it from the rest of the PCA data
smoker_status <- pca_df[1, 1:15] # Extract the SmokerStatus row (first row)
pca_data <- pca_df[-1, 1:15] # Remove the SmokerStatus row to keep only PCA values

# Convert smoker_status into a factor to categorize it as "Never Smoker" or "Current Smoker"
# Extract the SmokerStatus as a vector (first row)
smoker_status <- as.vector(as.matrix(pca_df[1, 1:15])) # Convert first row to a vector

# Convert smoker_status into a factor
smoker_status <- as.factor(smoker_status)

# Remove the SmokerStatus row from pca_df to keep only PCA data for plotting
pca_data <- pca_df[-1, 1:15]

# Convert PCA data into numeric for plotting
pca_data <- as.data.frame(lapply(pca_data, as.numeric))

# Transpose the PCA data so that samples are in rows and PCs in columns
pca_plot_df <- as.data.frame(t(pca_data))
```

```
# Add SmokerStatus as a column in the transposed PCA dataframe
pca_plot_df$SmokerStatus <- smoker_status

# Plot PCA with ggplot2, using PC1 and PC2, and color points by SmokerStatus
ggplot(pca_plot_df, aes(x = V1, y = V2, color = SmokerStatus)) +
  geom_point(size = 3) +
  labs(title = "PCA of Gene Expression Data: PC1 vs PC2",
       x = "Principal Component 1",
       y = "Principal Component 2") +
  theme_minimal() +
  scale_color_manual(values = c("Never Smoker" = "blue", "Current Smoker" = "red"))
```

