Direct inhibition of the NOTCH TF. Differenctial expression analysis

Fernando Freire

May 25, 2019

Contents

1 Array expression profiling

Summary

NOTCH proteins regulate signaling pathways involved in cellular differentiation, proliferation and death. Overactive Notch signaling as been observed in numerous cancers and has been extensively studied in the context of T-cell acute lymphoblastic leukemia (T-ALL) where more than 50% of pateints harbour mutant NOTCH1. Small molecule modulators of these proteins would be important for understanding the role of NOTCH proteins in malignant and normal biological processes.

We were interested to measure the global changes in gene expression upon treatment of the human T-ALL cell lines HPB-ALL and KOPT-K1 with either vehicle alone (DMSO) or SAHM1, an alpha-helical hydrocarbon stapled peptide derived from the MAML1 co-activator protein.

Overall design Triplicate cultures of KOPT-K1 or HPB-ALL cells were treated with either DMSO alone or SAHM1 (20 uM) for 24 hours. Total RNA was extracted and hybridized to Affymetrix human U133 plus 2.0 microarrays (three arrays per treatment per cell line for a total of 12 arrays).

1.1 Pipeline

1.2 Python imports

Script 1.2.1 (python)

```
import rpy2.rinterface
%reload_ext rpy2.ipython
```

1.3 R imports

Script 1.3.1 (R)

```
%%R
##1. Load libraries
library("affy")
library("limma")
library("genefilter")
library(hgu133plus2.db)
wd <- "/Users/nandoide/misc_work/Desktop/uni/TRREP"
setwd(wd)
```

1.4 Functions

Script 1.4.1 (R)

```
%%R

import_CEL <- function(pattern) {
    # Import CEL files into affiBatch object
    files <- list.files(pattern = pattern)
    names <- gsub(".CEL.gz", "", files)
```

```
    abatch <- ReadAffy(filenames = files,  compress = TRUE, sampleNames = names)
    return(abatch)
}

create_eset <- function(affyBatch) {
    # Generates object eset (class ExprSet),
    # expresso function provides intensities in log scale
    return(expresso(affyBatch,
              bg.correct = TRUE,
              bgcorrect.method="rma",
              normalize = TRUE,
              normalize.method="quantiles",
              pmcorrect.method="pmonly",
              summary.method="medianpolish",
              verbose = TRUE))
}

boxplots <- function(affyBatch, eset, title) {
    # Generate BOXPLOTS before and after normalization
    boxplot(affyBatch,
        main=paste0("Boxplot Before Normalization ", title),
        col = "lightgrey")
    df_eset <- as.data.frame(exprs(eset))

    boxplot(data.frame(df_eset),
        main=paste0("Boxplot After Normalization (log scale) ", title), col = "white")
}

create_TopTable <- function(eset, names) {
    # Generate Toptable with limma

    # Data filtering using IQR.
    esetIQR <- varFilter(eset, var.func=IQR, var.cutoff=0.5, filterByQuantile=TRUE)

    # Differential expression analysis.#######
    design <- cbind(DMSO=c(1,1,1,0,0,0), SAHM1=c(0,0,0,1,1,1))

    rownames(design) <- names

    #7. Contrasts matrix.
    cont.matrix <- makeContrasts(DMSO_SAHM1 = SAHM1 - DMSO, levels = design)

    #8. Obtaining differentially expressed genes (DEGs)
    #Linear model and eBayes
    fit <- lmFit(esetIQR, design)  ##getting DEGs from IQR
    fit2 <- contrasts.fit(fit, cont.matrix)
    fit2 <- eBayes(fit2)

    #Table with DEGs results
    toptableIQR <- topTable(fit2, number=dim(exprs(esetIQR))[1], adjust.method="BH",
    ↪  sort.by="p")
    return(toptableIQR)
```

```
}

anotate_TopTable <- function(toptable) {
    # Obtain gene names from probe names and chip symbol dataset
    probenames_toptable <- as.character(rownames(toptable ))
    genesymbols_toptable <- as.character(mget(probenames_toptable, hgu133plus2SYMBOL))
    # Annotated gene table
    toptable_anot <- cbind(Symbol = genesymbols_toptable, toptable)
    return(toptable_anot)
}

generank_table <- function(toptable, rnk.file) {
    # Generate rank of table top 50 upregulared and top 50 downregulated from 250 better
    # adjustes p-values
    more_significant = toptable[order(toptable$adj.P.Val, decreasing = FALSE),][1:250,]
    up_50 = more_significant[which(toptable$logFC > 0), ] [1:50,] # up reg top 50
    down_50 = more_significant[which(toptable$logFC < 0), ] [1:50,] # down reg top 50

    print("Down-regulated genes")
    print(down_50[order(down_50$logFC), c(1,2,5)])

    print("Up-regulated genes")
    print(up_50[order(up_50$logFC), c(1,2,5)])

    d <- rbind(down_50[order(down_50$logFC), c(1,2,5)], up_50[order(up_50$logFC), c(1,2,5)])

    #df <- data.frame(paste0(dSymbol,'|',row.names(d)),dlogFC)
    df <- data.frame(row.names(d), d$logFC)
    write.table(df,row.names=FALSE,col.names=FALSE,
                quote=FALSE,sep="\t",file=paste0(rnk.file, ".rnk"))
}
```

1.5 Load raw data

Script 1.5.1 (R)

```
%%R
setwd("GSE18198_data")
affyBatch_HPB = import_CEL("HPB*")
affyBatch_KOPT = import_CEL("KOPT*")
setwd(wd)
```

1.6 Create expression sets

Script 1.6.1 (R)

```
%%R
eset_HPB <- create_eset(affyBatch_HPB)
eset_KOPT <- create_eset(affyBatch_KOPT)
```

```
background correction: rma
normalization: quantiles
PM/MM correction : pmonly
expression values: medianpolish
background correcting...done.
normalizing...done.
54675 ids to be processed
|                    |
|####################|
background correction: rma
normalization: quantiles
PM/MM correction : pmonly
expression values: medianpolish
background correcting...done.
normalizing...done.
54675 ids to be processed
|                    |
|####################|
```

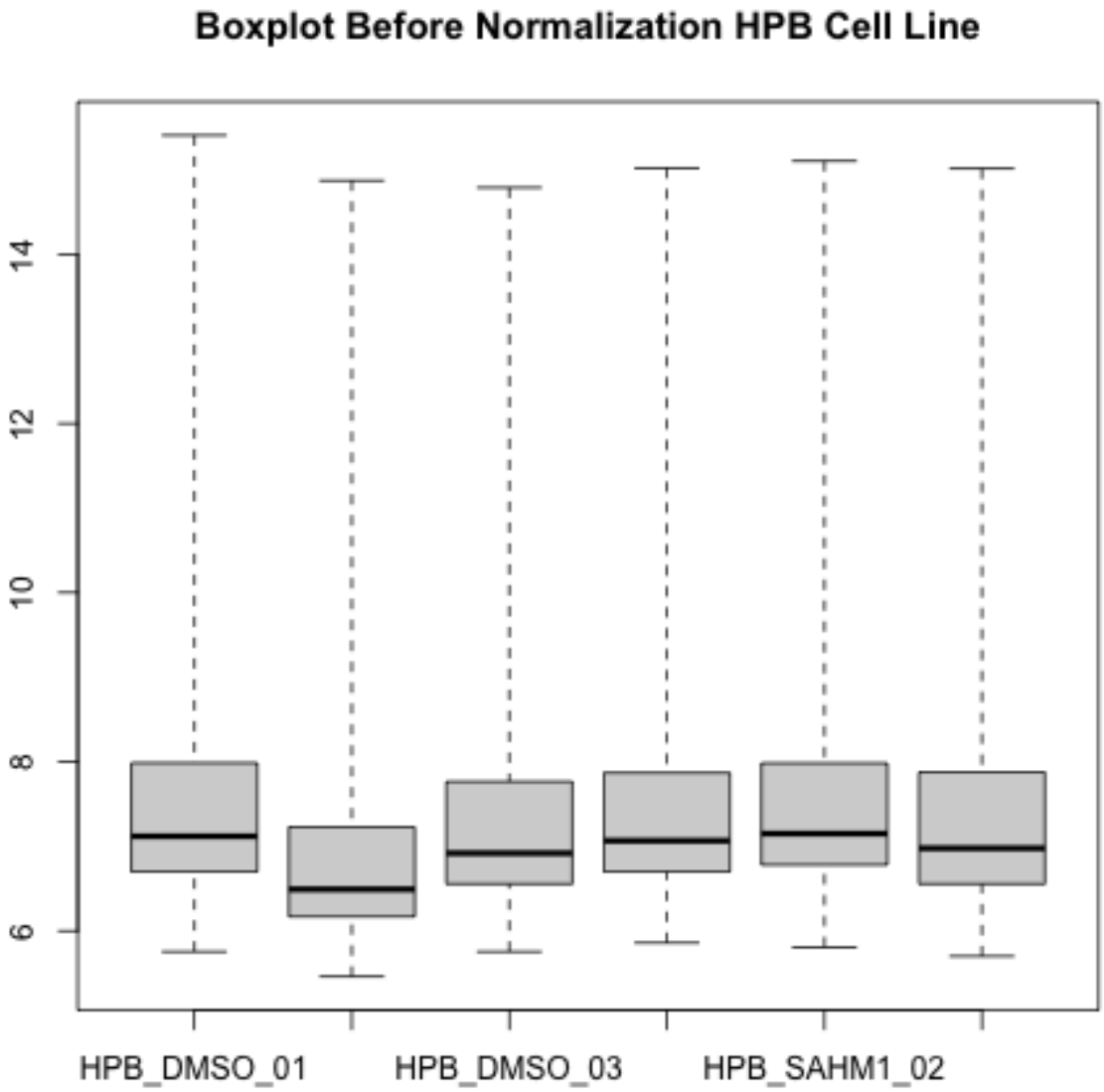
Script 1.6.2 (R)

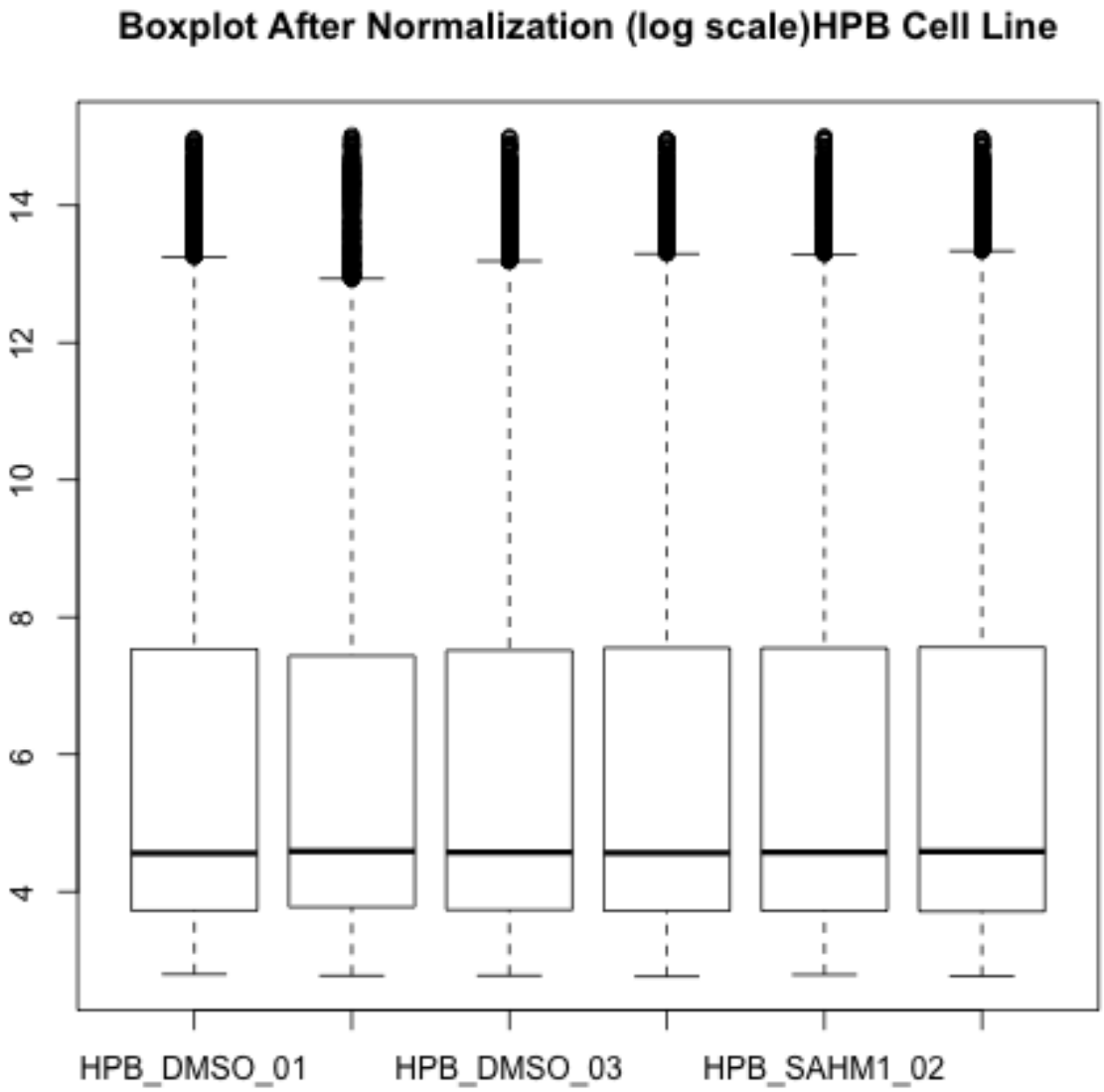
```
%%R
save(eset_HPB, file="eset_HPB.RData")
save(eset_KOPT, file="eset_KOPT.RData")
```

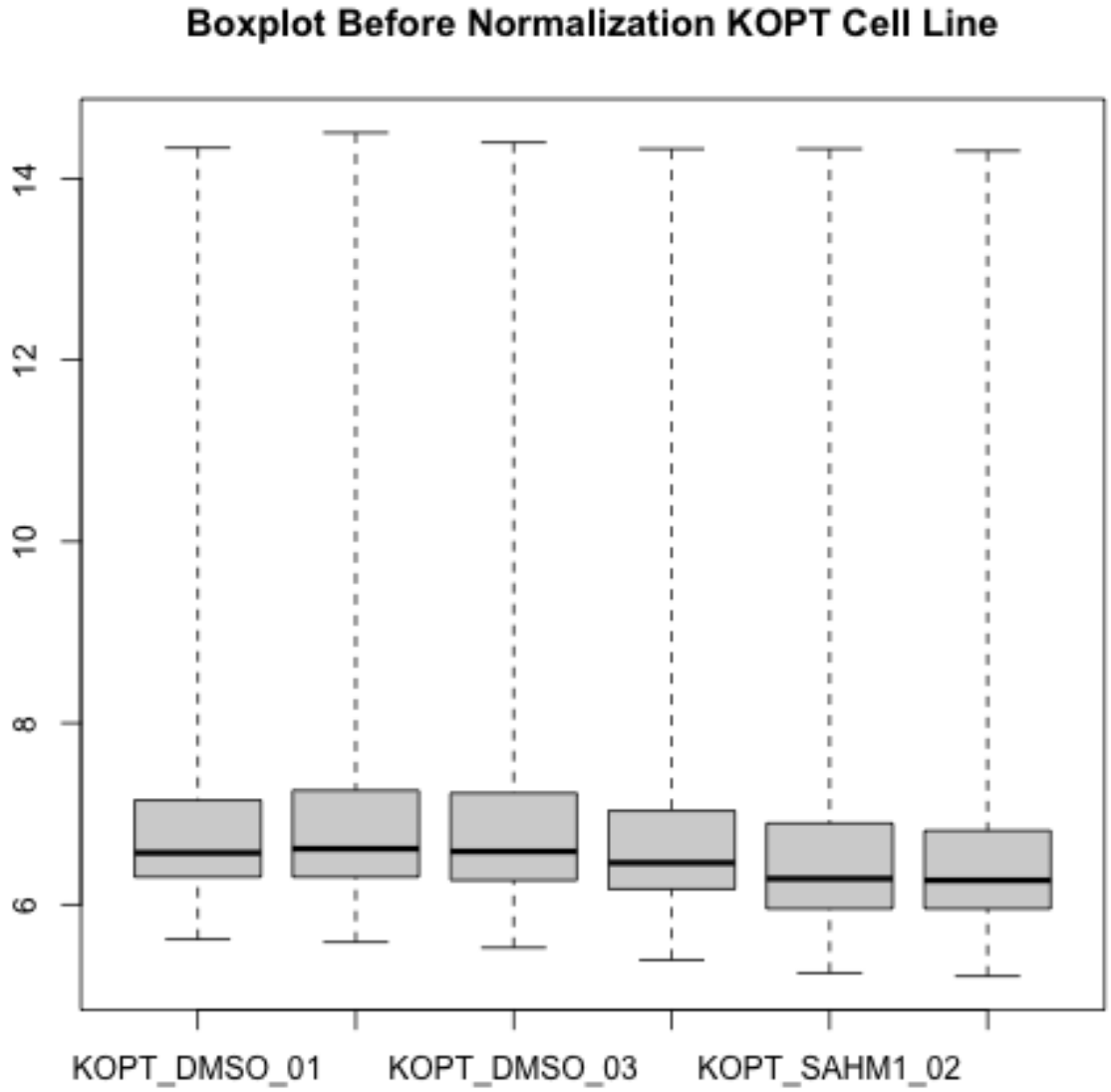
1.7 Quality plots

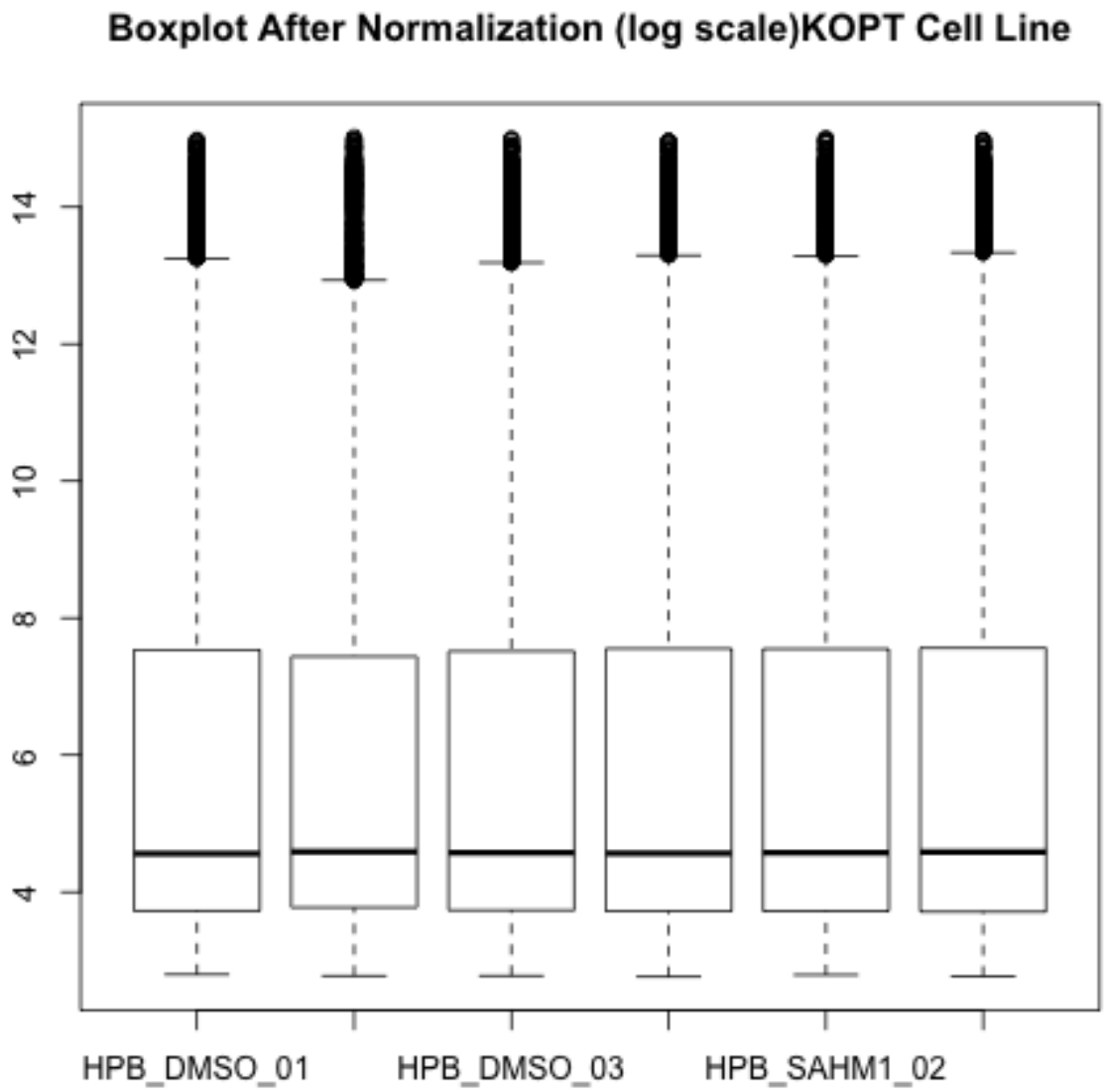
Script 1.7.1 (R)

```
%%R
boxplots(affyBatch_HPB, eset, "HPB Cell Line")
boxplots(affyBatch_KOPT, eset, "KOPT Cell Line")
```


Boxplot Before Normalization HPB Cell Line
14
12
10
8
6
HPB_DMSO_01
HPB_DMSO_03
HPB_SAHM1_02


Boxplot After Normalization (log scale)HPB Cell Line
14
12
10
8
6
4
HPB_DMSO_01
HPB_DMSO_03
HPB_SAHM1_02


Boxplot Before Normalization KOPT Cell Line
14
12
10
8
6
KOPT_DMSO_01
KOPT_DMSO_03
KOPT_SAHM1_02



1.8 Differential expressed genes

Script 1.8.1 (R)

```
%%R
toptable_HPB <- create_TopTable(eset_HPB, affyBatch_HPB$names)
toptable_anot_HPB <- anotate_TopTable(toptable_HPB)
generank_table(toptable_anot_HPB, "generank_HPB")
```

```
[1] "Down-regulated genes"
               Symbol      logFC      P.Value
225342_at         AK4 -2.1078858 2.477920e-09
230710_at    MIR210HG -1.9761492 1.124964e-09
227336_at        DTX1 -1.3841443 4.106969e-08
201842_s_at    EFEMP1 -1.3678548 2.196591e-08
204348_s_at       AK4 -1.3491612 3.922276e-07
227347_x_at      HES4 -1.2640712 6.910681e-08
200953_s_at     CCND2 -1.2357909 3.510442e-08
202464_s_at    PFKFB3 -1.1937490 7.190421e-08
202022_at       ALDOC -1.1753644 7.989863e-08
240546_at   LINC01120 -1.1131135 1.103841e-07
227337_at     ANKRD37 -1.1079168 2.161918e-07
200894_s_at     FKBP4 -1.0737202 7.367763e-07
217078_s_at    CD300A -1.0719092 1.507092e-06
201170_s_at   BHLHE40 -1.0402130 2.797321e-07
202934_at         HK2 -1.0377714 2.058505e-07
201848_s_at     BNIP3 -1.0052154 1.108152e-06
218051_s_at    NT5DC2 -1.0049990 1.010083e-05
203394_s_at      HES1 -0.9993973 1.254181e-06
219371_s_at      KLF2 -0.9800918 4.873149e-07
201251_at         PKM -0.9787897 2.113749e-05
201849_at       BNIP3 -0.9724936 9.701724e-07
213746_s_at      FLNA -0.9705676 3.593654e-06
201516_at         SRM -0.9428084 4.346682e-06
203523_at        LSP1 -0.9360851 6.999754e-06
225944_at         NLN -0.9317256 3.431626e-06
214183_s_at     TKTL1 -0.9290362 7.884232e-06
236180_at          NA -0.9287828 1.171221e-05
201194_at     SELENOW -0.9285463 9.512624e-07
231922_at      ZNF276 -0.9269927 1.788903e-05
209933_s_at    CD300A -0.9049911 1.494779e-06
214752_x_at      FLNA -0.9018607 4.491033e-06
226348_at       FUT11 -0.8976507 2.046444e-06
201212_at        LGMN -0.8972886 3.169006e-06
218305_at        IPO4 -0.8783786 5.557955e-06
205544_s_at       CR2 -0.8502424 2.471321e-06
202145_at        LY6E -0.8406636 8.347778e-06
200859_x_at      FLNA -0.8368746 1.053759e-05
202887_s_at     DDIT4 -0.8271607 5.174322e-06
200965_s_at    ABLIM1 -0.8229490 3.927885e-06
203504_s_at     ABCA1 -0.8211483 3.566852e-06
208116_s_at    MAN1A1 -0.7931451 1.709514e-05
202472_at         MPI -0.7831169 1.103785e-05
207543_s_at     P4HA1 -0.7783781 7.120628e-06
201563_at        SORD -0.7663457 2.053337e-05
222150_s_at      GSAP -0.7532244 1.848098e-05
207539_s_at       IL4 -0.7404704 1.875250e-05
```

```
205895_s_at     NOLC1 -0.7357033 2.131309e-05
219389_at       SUSD4 -0.7347418 1.589585e-05
201562_s_at      SORD -0.7343969 1.852132e-05
218984_at        PUS7 -0.7076878 2.067595e-05
[1] "Up-regulated genes"
                 Symbol     logFC      P.Value
204962_s_at          NA 0.8016614 5.641360e-06
222670_s_at        MAFB 0.8314472 3.362702e-06
244075_at            NA 0.8555977 5.711648e-06
205047_s_at        ASNS 0.8566562 2.994643e-06
236153_at            NA 0.8615325 3.057289e-06
228999_at          CHD2 0.8640573 4.876544e-06
202847_at          PCK2 0.8839066 3.064998e-06
242388_x_at       TAGAP 0.8933876 2.243184e-06
243368_at            NA 0.9077303 4.861619e-06
1558212_at           NA 0.9381646 3.234368e-06
212907_at       SLC30A1 0.9572462 7.762742e-07
241505_at            NA 0.9630797 3.611663e-06
230659_at            NA 0.9742286 1.658973e-06
203279_at         EDEM1 0.9785129 5.049724e-07
218923_at          CTBS 0.9839137 2.246790e-06
1558920_at   SLC8A1-AS1 0.9839933 2.671988e-06
215071_s_at   HIST1H2AC 0.9867407 4.424996e-06
217678_at       SLC7A11 1.0002602 1.142148e-06
235795_at          PAX6 1.0066337 1.110147e-06
1556294_at        FXYD2 1.0219109 3.233231e-06
229538_s_at      IQGAP3 1.0315469 8.882998e-07
206864_s_at         HRK 1.0324569 5.907274e-06
243495_s_at      ZNF652 1.0513033 4.009210e-06
218145_at         TRIB3 1.0673820 2.280009e-07
219892_at        TM6SF1 1.0684116 1.312524e-06
244377_at        SLC1A4 1.0880295 1.754737e-07
201010_s_at       TXNIP 1.1062690 1.135917e-07
209921_at       SLC7A11 1.1333321 8.218037e-08
209822_s_at       VLDLR 1.1350680 1.965762e-07
230795_at            NA 1.1699268 2.078133e-07
213931_at            NA 1.1702639 4.592170e-07
201009_s_at       TXNIP 1.2328092 1.338931e-07
244042_x_at          NA 1.2501748 2.095176e-06
218559_s_at        MAFB 1.2553755 7.617753e-07
225957_at        CREBRF 1.2981952 5.931956e-07
222853_at         FLRT3 1.3254635 5.830954e-08
219270_at         CHAC1 1.3362724 9.740672e-09
202672_s_at        ATF3 1.3481402 1.078712e-08
218280_x_at          NA 1.3495011 2.699758e-06
207076_s_at        ASS1 1.4918960 1.202087e-08
201008_s_at       TXNIP 1.4920023 1.281217e-07
243871_at   LOC100130476 1.5055489 5.347338e-07
```

```
201464_x_at          JUN 1.5541848 1.020585e-09
236962_at             NA 1.5966520 3.971628e-06
235412_at         ARHGEF7 1.5978442 3.401600e-07
229541_at             NA 1.6143542 1.792788e-07
229147_at         RASSF6 1.6458323 1.057501e-09
235638_at         RASSF6 1.7804932 3.444189e-08
201466_s_at          JUN 2.1333931 7.714196e-11
201465_s_at          JUN 2.2743608 1.927464e-09
```

Script 1.8.2 (R)

```
%%R
toptable_KOPT <- create_TopTable(eset_KOPT, affyBatch_KOPT$names)
toptable_anot_KOPT <- anotate_TopTable(toptable_KOPT)
generank_table(toptable_anot_KOPT, "generank_KOPT")
```

```
[1] "Down-regulated genes"
               Symbol     logFC      P.Value
209921_at     SLC7A11 -2.916468 1.310423e-15
205047_s_at      ASNS -2.757747 3.034026e-15
209369_at       ANXA3 -2.661459 1.511951e-15
219270_at       CHAC1 -2.264373 2.401463e-13
226517_at       BCAT1 -2.216157 5.303756e-14
214452_at       BCAT1 -2.166267 1.742595e-12
225285_at       BCAT1 -2.072317 3.848602e-14
217678_at     SLC7A11 -2.058710 2.464934e-13
230748_at     SLC16A6 -2.004791 4.677510e-12
219892_at      TM6SF1 -1.957664 3.702055e-13
220892_s_at     PSAT1 -1.938100 9.078716e-14
204351_at       S100P -1.902944 3.879042e-12
223195_s_at     SESN2 -1.889417 2.491927e-13
214079_at       DHRS2 -1.873374 3.900855e-12
209822_s_at     VLDLR -1.863734 2.466589e-12
212290_at      SLC7A1 -1.860735 2.853913e-11
202847_at        PCK2 -1.828534 5.310992e-13
225520_at          NA -1.790127 4.616209e-13
223062_s_at     PSAT1 -1.786577 5.539870e-13
223196_s_at     SESN2 -1.730910 4.054436e-11
200924_s_at    SLC3A2 -1.698282 2.159071e-11
1553972_a_at      CBS -1.682036 1.973032e-12
207076_s_at      ASS1 -1.675254 1.932857e-11
229787_s_at       OGT -1.662192 2.679095e-10
222632_s_at    LZTFL1 -1.643945 1.265787e-10
212816_s_at       CBS -1.599989 1.558010e-10
224839_s_at      GPT2 -1.553796 1.395913e-11
```

```
240983_s_at      CARS -1.552793 7.189001e-11
215411_s_at  TRAF3IP2 -1.540251 1.930189e-11
221539_at    EIF4EBP1 -1.533442 8.406763e-12
201195_s_at    SLC7A5 -1.511384 1.026657e-11
214390_s_at     BCAT1 -1.459915 4.257738e-10
204999_s_at      ATF5 -1.451634 1.928577e-11
210512_s_at     VEGFA -1.448871 7.446544e-11
200878_at       EPAS1 -1.435966 5.731229e-10
212501_at       CEBPB -1.428290 3.225367e-11
224580_at     SLC38A1 -1.428164 1.563644e-10
204744_s_at      IARS -1.419418 3.173924e-11
208693_s_at      GARS -1.406661 4.937878e-11
223059_s_at   FAM107B -1.392809 6.305787e-11
218437_s_at    LZTFL1 -1.332751 2.784128e-10
1558212_at         NA -1.308392 3.237396e-10
205653_at        CTSG -1.279331 4.485939e-10
217078_s_at    CD300A -1.278830 4.189234e-10
203627_at       IGF1R -1.275358 7.388759e-10
221933_at      NLGN4X -1.272235 8.275829e-10
231894_at        SARS -1.263321 4.165609e-10
214095_at       SHMT2 -1.235111 3.262739e-10
201263_at        TARS -1.197960 7.411469e-10
226181_at       TUBE1 -1.189891 6.361196e-10
[1] "Up-regulated genes"
                         Symbol     logFC      P.Value
224429_x_at                  NA 0.9259213 3.512066e-08
220725_x_at               DNAH3 0.9423323 3.702700e-08
210686_x_at             SLC25A16 0.9631964 2.086660e-08
213605_s_at                  NA 0.9632616 4.049682e-08
1556206_at             LINC00408 0.9653901 1.942815e-08
244114_x_at                  NA 0.9679377 3.950940e-08
241632_x_at                  NA 0.9686426 2.376898e-08
239017_at                    NA 0.9834736 4.450851e-08
1558496_at             LINC02053 0.9853247 2.602389e-08
211585_at                   NPAT 0.9926606 3.603534e-08
236389_x_at                  NA 1.0027588 2.068138e-08
208120_x_at              FKSG49 1.0184438 1.091788e-08
206323_x_at               OPHN1 1.0212889 2.929986e-08
224284_x_at              FKSG49 1.0251308 1.676159e-08
201464_x_at                 JUN 1.0348832 1.442900e-08
220828_s_at            LINC01949 1.0397500 2.077709e-08
81737_at            LOC100505915 1.0462560 2.982880e-08
210800_at                TIMM8A 1.0479516 2.249686e-08
215182_x_at                  NA 1.0559763 2.059300e-08
243489_at                    NA 1.0626878 1.757123e-08
240988_x_at                  NA 1.0629756 2.290259e-08
224288_x_at              FKSG49 1.0813463 3.257173e-09
AFFX-r2-Ec-bioB-5_at         NA 1.1020958 2.764748e-08
```

```
242862_x_at                   NA 1.1113007 3.225720e-09
1563674_at                 FCRL2 1.1127799 4.192285e-08
AFFX-BioC-5_at                NA 1.1199525 3.669763e-08
210718_s_at                   NA 1.1396478 3.879559e-08
1562755_at                    NA 1.1425399 8.188238e-09
232964_at                     NA 1.1460961 5.330929e-09
220232_at                   SCD5 1.1545960 3.507998e-09
1569940_at                SLC6A16 1.1781696 2.322268e-09
211454_x_at                FKSG49 1.1800267 1.323300e-09
AFFX-BioB-M_at                NA 1.1849198 3.522114e-08
209700_x_at               PDE4DIP 1.1871538 2.275537e-09
1566145_s_at                  NA 1.1918754 5.493504e-10
234949_at                  FRG1BP 1.1922760 3.782766e-09
1560144_at                    NA 1.2077566 6.025101e-09
1553185_at                 RASEF 1.2129105 1.543437e-09
231597_x_at                   NA 1.2169048 4.648771e-09
242619_x_at                   NA 1.2257718 9.127723e-10
1553186_x_at               RASEF 1.2330332 1.762221e-09
227952_at                 ZNF595 1.2408024 1.105024e-09
1561754_at                    NA 1.2457988 2.796446e-09
224159_x_at                TRIM4 1.2471076 2.081970e-09
243689_s_at                FRG1BP 1.3748167 1.470956e-09
231598_x_at                   NA 1.4418759 2.204665e-10
228919_at                     NA 1.4589605 5.477094e-09
1562527_at            LOC441666 1.4936054 9.559516e-11
1558048_x_at                  NA 1.5005510 2.143177e-11
211565_at                 SH3GL3 1.5671795 3.136958e-11
```

1.9 Generate GSEA gct files

Script 1.9.1 (R)

```r
%%R

write.table(exprs(eset_HPB),row.names=TRUE,col.names=FALSE,quote=FALSE,file="eset_HPB.tsv",
 ↪  sep = "\t")
write.table(exprs(eset_KOPT),row.names=TRUE,col.names=FALSE,quote=FALSE,file="eset_KOPT.tsv"
 ↪  , sep =
 ↪  "\t")
```

Script 1.9.2 (bash)

```bash
%%bash
echo "#1.2" > gct.head
echo "$(cat eset_HPB.tsv | wc -l) 6" >> gct.head
```

```
echo "GID	NAME	HPB_DMSO_01	HPB_DMSO_02	HPB_DMSO_03	HPB_SAHM1_01	HPB_SAHM1_02	HPB_SAHM1_03" >> gct.head
#head gct.head
cat gct.head eset_HPB.tsv > eset_HPB.gct
cat gct.head eset_KOPT.tsv > eset_KOPT.gct
head eset_KOPT.gct
```

Output

```
#1.2
   54675 6
GID	NAME	HPB_DMSO_01	HPB_DMSO_02	HPB_DMSO_03	HPB_SAHM1_01	HPB_SAHM1_02	HPB_SAHM1_03
1007_s_at	6.39550778914434	6.63361329719407	6.6375785381641	7.006065069446903	6.68384761746578	6.81043796741411
1053_at	8.27291684544428	8.04195281624978	8.13203974449545	8.0715608599753	7.85863717382127	8.02719895633383
117_at	3.14729311345973	2.94262593114145	3.15131758375372	3.05034281709199	3.23612638945269	3.14794105723779
121_at	6.87870865041484	6.88129495627018	6.82820532797191	6.86631576689715	6.95693356773948	6.83081218004268
1255_g_at	2.81967249987976	2.88648735903777	2.88201717529506	2.97379457130539	2.98461193851988	3.01905325883313
1294_at	6.87006027265826	6.80399928959987	6.65695733746297	6.57214844234865	6.65606153751673	6.80098950326962
1316_at	3.77034812760452	3.82602264159359	4.09195191846965	3.72596001077624	4.2350082264062	4.00773283062547
```

1.10 Conclusions