|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Supplementary Table 1. Details of public datasets used in this study** | | | | | |
| Dataset | Data type | Platform(s) | Sample type | No. of samples | Reference |
| TCGA | RNA-seq | Illumina HiSeq 2000 | AML | 173 | Ley, TJ, et al (2013) |
| TCGA | miRNA-seq | Illumina Genome Analyzer | AML | 188 | Ley, TJ, et al (2013) |
| GSE63270 | microarray | Affymetrix U133 Plus 2.0 | AML, NBM | 104 | Jung, N, et al (2015) |

1. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ, Jr., Laird PW, Baty JD, et al: Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia**.** *N Engl J Med* 2013; 368**:**2059-2074.

2. Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP: An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis**.** *Nat Commun* 2015; 6**:**8489.

**R script for plotting gene volcano maps co-expressed with *ITGB2-AS1***

library(xlsx)

data <- read.xlsx("～/～/ *ITGB-AS1*\_co-expressed\_genes. xlsx ", 1)

library(ggplot2)

P.Value <- c(data$q.Value)

cor <- c(data$Spearman.s.Correlation)

gene <- as.character(data$Correlated.Gene)

df <- data.frame(P.Value, cor, gene)

df.G <- subset(df, cor < -0.3 & P.Value < 0.05) #define Green

df.G <- cbind(df.G, rep(1, nrow(df.G)))

colnames(df.G)[4] <- "Color"

df.B <- subset(df, (cor >= -0.3 & cor <= 0.3) | P.Value >= 0.05)

df.B <- cbind(df.B, rep(2, nrow(df.B)))

colnames(df.B)[4] <- "Color"

df.R <- subset(df, cor > 0.3 & P.Value < 0.05)

df.R <- cbind(df.R, rep(3, nrow(df.R)))

colnames(df.R)[4] <- "Color"

df.t <- rbind(df.G, df.B, df.R)

df.t$Color <- as.factor(df.t$Color)

rm(list = "gene")

install.packages("ggrepel")

library(ggrepel)

ggplot(df.t,aes(x = cor, y = -log10(P.Value), label = gene)) +

geom\_point(aes(color =Color)) + xlim(-3,3) + ylim(0,100) +

labs(x="correlation coeffient",y="-log10 (p-value)",title="ITGB2-AS1 Association Result") +

scale\_color\_manual(values =c("#377EB8","grey","#E41A1C"),

labels=c("postive", "Not significant", "negative")) +

geom\_hline(yintercept=1.3,linetype=2,size=1,col="black")+

geom\_vline(xintercept=c(-0.3,0.3),linetype=2,size=1,col= "black") +

theme(plot.title = element\_text(hjust = 0.5), legend.position="right", legend.title = element\_blank()) +

geom\_label\_repel(

data = subset(df.t, df.t$P.Value < 0.00001 & abs(df.t$cor) >= 0.7),

aes(x=cor,y=-log10(P.Value),label = gene),

fill="pink",colour = 'black', size = 3,

box.padding=unit(0.35, "lines"), point.padding=unit(0.5, "lines"), segment.size = 1, segment.colour = "grey50")

**Figure1a. ROC curve analysis using *ITGB2-AS1* expression for discriminating AML patients from controls.**



**Figure1b. ROC curve analysis using *ITGB2* expression for discriminating AML patients from controls.**

