Codes:

```
library(DOSE)
library(enrichplot)
library(clusterProfiler)
library(org.Hs.eg.db)
https://www.jianshu.com/p/3c5de0c4ca7
ego <- enrichGO( gene    = a$ENTREZID,
keyType ="ENTREZID",
OrgDb      = org.Hs.eg.db,
ont          = "CC",
pAdjustMethod = "BH",
pvalueCutoff    = 0.05,
qvalueCutoff    = 0.05,
readable    = TRUE)
ekegg <- enrichKEGG(gene = a$ENTREZID)
edo<- enrichDO(gene = a$ENTREZID,readable = TRUE)
cnetplot(ego, foldChange=a$LFC, circular = TRUE, colorEdge = TRUE,showCategory = 10)
cell_markers                                                                                    <-
vroom::vroom('http://bio-bigdata.hrbmu.edu.cn/CellMarker/download/Human_cell_markers.txt'
) %>%
   tidyr::unite("cellMarker", tissueType, cancerType, cellName, sep=", ") %>%
   dplyr::select(cellMarker, geneID) %>%
   dplyr::mutate(geneID = strsplit(geneID, ', '))
cell_markers
eTC <- enricher(a$ENTREZID, TERM2GENE=cell_markers, minGSSize=1)

library(ggplot2)
library(limma)
library(pheatmap)
logFCfilter=1
adjPfilter=0.05
expFile="geneMatrix.txt"
conFile="sample1.txt"
treatFile="sample2.txt"

rt=read.table(expFile,sep="\t",header=T,check.names=F)
rt=as.matrix(rt)
rownames(rt)=rt[,1]
exp=rt[,2:ncol(rt)]
dimnames=list(rownames(exp),colnames(exp))
data=matrix(as.numeric(as.matrix(exp)),nrow=nrow(exp),dimnames=dimnames)
data=avereps(data)
#data=log2(data+1)
data=normalizeBetweenArrays(data)
```

```
library(plyr)
library(ggplot2)
library(grid)
library(gridExtra)
files=grep(".xls",dir(),value=T)
data = lapply(files,read.delim)
names(data) = files
dataSet = ldply(data, data.frame)
dataSet$pathway = gsub(".xls","",dataSet$.id)
gseaCol=c("#58CDD9","#7A142C","#5D90BA","#431A3D","#91612D","#6E568C","#E0367A","#D8
D155","#64495D","#7CC767","#223D6C","#D20A13","#FFD121","#088247","#11AA4D")
pGsea=ggplot(dataSet,aes(x=RANK.IN.GENE.LIST,y=RUNNING.ES,colour=pathway,group=pathway))
+
   geom_line(size = 1.5) + scale_color_manual(values = gseaCol[1:nrow(dataSet)]) +
   labs(x = "", y = "Enrichment Score", title = "") + scale_x_continuous(expand = c(0, 0)) +
   scale_y_continuous(expand = c(0, 0),limits = c(min(dataSet$RUNNING.ES - 0.02),
max(dataSet$RUNNING.ES + 0.02))) +
   theme_bw() + theme(panel.grid = element_blank()) + theme(panel.border = element_blank())
+ theme(axis.line = element_line(colour = "black")) + theme(axis.line.x =
element_blank(),axis.ticks.x = element_blank(),axis.text.x = element_blank()) +
   geom_hline(yintercept = 0) +
   guides(colour = guide_legend(title = NULL)) + theme(legend.background = element_blank()) +
theme(legend.key = element_blank())+theme(legend.key.size=unit(0.5,'cm'))
pGene=ggplot(dataSet,aes(RANK.IN.GENE.LIST,pathway,colour=pathway))+geom_tile()+
   scale_color_manual(values = gseaCol[1:nrow(dataSet)]) +
   labs(x = "High expression<---------->Low expression", y = "", title = "") +
   scale_x_discrete(expand = c(0, 0)) + scale_y_discrete(expand = c(0, 0)) +
   theme_bw() + theme(panel.grid = element_blank()) + theme(panel.border = element_blank())
+ theme(axis.line = element_line(colour = "black"))+
   theme(axis.line.y = element_blank(),axis.ticks.y = element_blank(),axis.text.y =
element_blank())+ guides(color=FALSE)

gGsea = ggplot_gtable(ggplot_build(pGsea))
gGene = ggplot_gtable(ggplot_build(pGene))
maxWidth = grid::unit.pmax(gGsea$widths, gGene$widths)
gGsea$widths = as.list(maxWidth)
gGene$widths = as.list(maxWidth)
dev.off()

pdf('multipleGSEA.pdf',
       width=10,
       height=6)
par(mar=c(5,5,2,5))
```

```
grid.arrange(arrangeGrob(gGsea,gGene,nrow=2,heights=c(.8,.3)))
dev.off()

sample1=read.table(conFile,sep="\t",header=F,check.names=F)
sample2=read.table(treatFile,sep="\t",header=F,check.names=F)
conData=data[,as.vector(sample1[,1])]
treatData=data[,as.vector(sample2[,1])]
rt=cbind(conData,treatData)
conNum=ncol(conData)
treatNum=ncol(treatData)

Type=c(rep("con",conNum),rep("treat",treatNum))
design <- model.matrix(~0+factor(Type))
colnames(design) <- c("con","treat")
fit <- lmFit(rt,design)
cont.matrix<-makeContrasts(treat-con,levels=design)
fit2 <- contrasts.fit(fit, cont.matrix)
fit2 <- eBayes(fit2)

allDiff=topTable(fit2,adjust='fdr',number=200000)
write.table(allDiff,file="GEO_all.xls",sep="\t",quote=F)

diffSig=allDiff[with(allDiff, (abs(logFC)>logFCfilter & adj.P.Val < adjPfilter )), ]
diffSigOut=rbind(id=colnames(diffSig),diffSig)
write.table(diffSigOut,file="GEO_diff.xls",sep="\t",quote=F,col.names=F)
write.table(diffSigOut,file="GEO_diff.txt",sep="\t",quote=F,col.names=F)

geneNum=50
diffSig=diffSig[order(as.numeric(as.vector(diffSig$logFC))),]
diffGeneName=as.vector(rownames(diffSig))
diffLength=length(diffGeneName)
hmGene=c()
if(diffLength>(geneNum*2)){
    hmGene=diffGeneName[c(1:geneNum,(diffLength-geneNum+1):diffLength)]
}else{
    hmGene=diffGeneName
}
hmExp=rt[hmGene,]
Type=c(rep("N",conNum),rep("T",treatNum))
names(Type)=colnames(rt)
Type=as.data.frame(Type)
pdf(file="GEO_heatmap.pdf",height=8,width=10)
pheatmap(hmExp,
        annotation=Type,
```

```
              color = colorRampPalette(c("green", "black", "red"))(50),
              cluster_cols =F,
              show_colnames = F,
              scale="row",
              fontsize = 8,
              fontsize_row=6,
              fontsize_col=8)
dev.off()

Significant=ifelse((allDiff$adj.P.Val<adjPfilter & abs(allDiff$logFC)>logFCfilter),
ifelse(allDiff$logFC>logFCfilter,"Up","Down"), "Not")
p = ggplot(allDiff, aes(logFC, -log10(adj.P.Val)))+
    geom_point(aes(col=Significant))+
    scale_color_manual(values=c("green", "black", "red"))+
    labs(title = " ")+
    theme(plot.title = element_text(size = 16, hjust = 0.5, face = "bold"))
p=p+theme_bw()

pdf("GEO_vol.pdf",width=5.5,height=5)
print(p)
dev.off()

library(pROC)
inputFile="input.txt"
outFile="ROC.pdf"
rt=read.table(inputFile,header=T,sep="\t",check.names=F,row.names=1)
y=colnames(rt)[1]
rocobj1=roc(rt[,y], as.vector(rt[,x]))
pdf(file=outFile,width=5,height=5)
plot(rocobj1, print.auc=TRUE, col="red")
dev.off()
inputFile="1.txt"
gmtFile="immune.gmt"#GMT 文件

library(GSVA)
library(limma)
library(GSEABase)

rt=read.table(inputFile,sep="\t",header=T,check.names=F)
rt=as.matrix(rt)
rownames(rt)=rt[,1]
exp=rt[,2:ncol(rt)]
dimnames=list(rownames(exp),colnames(exp))
mat=matrix(as.numeric(as.matrix(exp)),nrow=nrow(exp),dimnames=dimnames)
```

```
mat=avereps(mat)
mat=mat[rowMeans(mat)>0,]
geneSet=getGmt(gmtFile,geneIdType=SymbolIdentifier())

ssgseaScore=gsva(mat, geneSet, method='ssgsea', kcdf='Gaussian', abs.ranking=TRUE)
normalize=function(x){return((x-min(x))/(max(x)-min(x)))}
ssgseaOut=normalize(ssgseaScore)
ssgseaOut=rbind(id=colnames(ssgseaOut),ssgseaOut)
write.table(ssgseaOut,file="ssgseaOut.txt",sep="\t",quote=F,col.names=F)

exp = read.delim("1.txt", row.names = 1, header = T)
gene_set<-read.csv("mmc3.csv")[, 1:2]
head(gene_set)
list<- split(as.matrix(gene_set)[,1], gene_set[,2])
gsva_matrix<- gsva(as.matrix(exp), list,method='ssgsea',kcdf='Gaussian',abs.ranking=TRUE)
gsva_matrix1<- t(scale(t(gsva_matrix)))
gsva_matrix1[gsva_matrix1< -2] <- -2
gsva_matrix1[gsva_matrix1>2] <- 2
normalization<-function(x){
   return((x-min(x))/(max(x)-min(x)))}
nor_gsva_matrix1 <- normalization(gsva_matrix1)
write.csv(nor_gsva_matrix1,"nor_gsva_matrix.csv")
```

Supplemental Table 1. Primers used in the study.

Gene symbol	Species		Sequence (5'-3')
AQP9	Human	Forward	CTCTGAGTTCTTGGGCACGT
		Reverse	GCCATTGCAACTGCCATTGA
	Mouse	Forward	AGGGATGGAGCTCAGGTCAT
		Reverse	ATCCCACCAGCCTTTTCTCG
BACH2	Human	Forward	CCCAGCTCCTGAACAGTGAG
		Reverse	CTCCTCCTCTCCTGCAGAGT
	Mouse	Forward	GTTTGCCTACACTGCCAAGC
		Reverse	CTGTTCAGGAGCTGGGTCTG
CD4	Human	Forward	CCTGGTAGTAGCCCCTCAGT
		Reverse	GGCCTTCTGGAAAGCTAGCA
	Mouse	Forward	TGTCACTCAAGGGAAGACGC
		Reverse	ATGCTGCCCCAGAATCTTCC
IL17RE	Human	Forward	CACTGTTCCCGCTGTTTGTG
		Reverse	GCTGGCATCTTGTGTCTCCT
	Mouse	Forward	CCTGTGTACAGCATCAGCCA
		Reverse	TCAAGAGCCCGAGGTGTCTA
S100A9	Human	Forward	GCAAAATGTCGCAGCTGGAA
		Reverse	TGTGTCCAGGTCCTCCATGA
	Mouse	Forward	ACCAGGACAATCAGCTGAGC
		Reverse	ACAGCCTTTGCCATGACTGT
GAPDH	Human	Forward	GGAGCGAGATCCCTCCAAAAT
		Reverse	GGCTGTTGTCATACTTCTCATGG
	Mouse	Forward	AGGTCGGTGTGAACGGATTTG
		Reverse	GGGGTCGTTGATGGCAACA
SREBP1c	Human	Forward	CAGCGTCTACCATAGCCCTG
		Reverse	AAGGAGACGAGCACCAACAG
FAS	Human	Forward	TCTGGTTCTTACGTCTGTTGC
		Reverse	CTGTGCAGTCCCTAGCTTTCC
SCD1	Human	Forward	TCTAGCTCCTATACCACCACCA
		Reverse	TCGTCTCCAACTTATCTCCTCC
ACC	Human	Forward	ATGTCTGGCTTGCACCTAGTA
		Reverse	CCCCAAAGCGAGTAACAAATTCT

Supplemental Table 2. Top 15 significant immunological signature enriched in GSEA.

Gene set name	ES	NES	P value
LI PBMC MENOMUNE A C Y W 135 AGE 18 45YO 3DY DN	-0.785	-1.546	< 0.001
GSE46606 UNSTIM VS CD40L IL2 IL5 DAY1 STIMULATED BCELL UP	-0.604	-1.529	< 0.001
GSE45739 NRAS KO VS WT ACD3 ACD28 STIM CD4 TCELL DN	-0.578	-1.486	< 0.001
HARALAMBIEVA PBMC TIV AGE 50 74YO CORRELATED WITH MEMORY B CELL RESPONSE 28DY NEGATIVE	-0.681	-1.465	0.016
GSE9006 TYPE 1 VS TYPE 2 DIABETES PBMC AT DX UP	0.627	1.453	0.004
GAUCHER PBMC YF VAX STAMARIL UNKNOWN AGE 14DY UP	0.663	1.452	0.019
OCONNOR PBMC MENVEO ACWYVAX AGE 30 70YO 7DY AFTER SECOND DOSE VS 7DY AFTER FIRST DOSE UP	0.598	1.441	< 0.001
GSE22886 IGM MEMORY BCELL VS BM PLASMA CELL DN	0.565	1.441	< 0.001
GSE22103 UNSTIM VS LPS STIM NEUTROPHIL DN	0.539	1.429	< 0.001
GSE1432 CTRL VS IFNG 1H MICROGLIA UP	0.535	1.429	< 0.001
GSE8384 CTRL VS B ABORTUS 4H MAC CELL LINE DN	0.515	1.423	0.017
GSE22886 NAIVE BCELL VS BM PLASMA CELL DN	0.564	1.418	< 0.001
GSE9006 TYPE 1 DIABETES AT DX VS 4MONTH POST DX PBMC UP	0.584	1.417	0.023
GSE46606 UNSTIM VS CD40L IL2 IL5 DAY3 STIMULATED BCELL UP	-0.582	-1.414	0.025
GSE40274 CTRL VS FOXP3 AND EOS TRANSDUCED ACTIVATED CD4 TCELL DN	0.579	1.412	0.010

GSEA, gene set enrichment analysis; ES, enrichment score; NES, normalized enrichment score.

Hep G2 cells STR report

Method: An appropriate amount of Hep G2 cells (Cell number PC-H2023011103,1×10^6) were used Microread Genomic DNA Kit to extract DNA, 20 STR loci and gender identification loci were amplified by Microreader™21 ID System, PCR product detection was performed by GenReader 7010 genetic analyzer, detection results were analyzed by GeneMapper Software6(Applied Biosystems), and compared with ATCC, DSMZ, JCRB, ExPASy and other databases.

Experimental result:

1. The results of negative and positive control were correct.
2. The genotyping results of STR locus of Hep G2 cell line is shown in the following table.

Conclusion:

1. The genomic DNA of Hep G2 cell line is clear and the result of genotyping is good.
2. The results of STR typing showed that no cross contamination of human cell was found in the cell line of Hep G2 cell line.
3. The DNA typing of the cell line was 97.06% matched with the cell type in the cell bank, and the cell line name was Hep G2.

Web: www.procell.com.cn
Tel: 400-999-2100; 027-87287608
Email: sales@procell.com.cn
Add: Building C4,3rd Phase of Biomedical Park, #858 Gaoxin Road, Donghu Hi-Tech Development Area.Wuhan.Hubei.China

Appendix I: The genotyping results of STR locus of Hep G2 cell line.

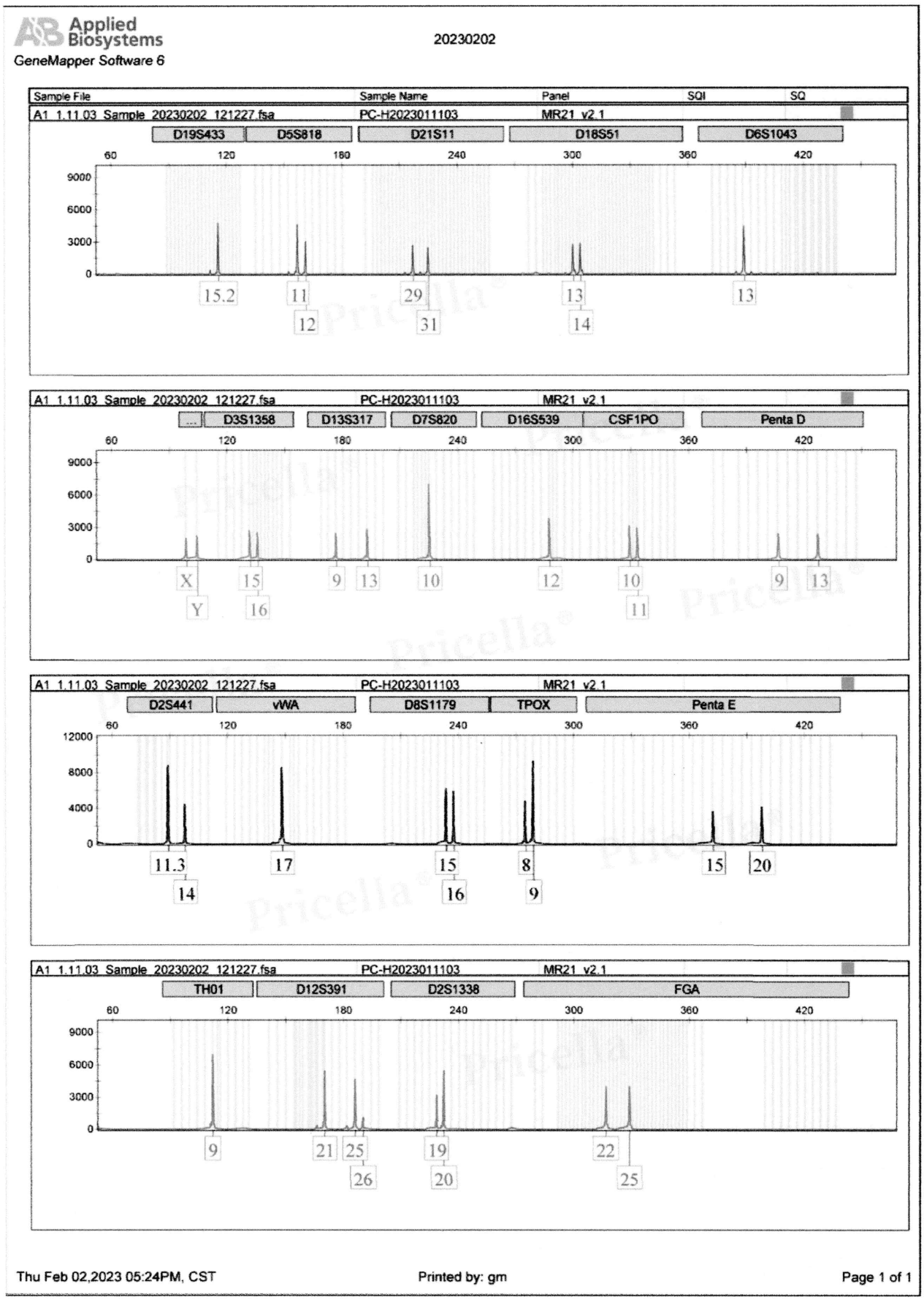
STR Loci	Sample：PC-H2023011103	Database：Hep G2
Amelogenin	X,Y	X,Y
CSF1PO	10,11	10,11
D2S1338	19,20	19,20
D3S1358	15,16	15,16
D5S818	11,12	11,12
D7S820	10	10
D8S1179	15,16	15,16
D13S317	9,13	9,13
D16S539	12	12,13
D18S51	13,14	13,14
D19S433	15.2	15.2
D21S11	29,31	29,31
FGA	22,25	22,25
PentaD	9,13	9,13
PentaE	15,20	15,20
TH01	9	9
TPOX	8,9	8,9
vWA	17	17
D6S1043	13	
D12S391	21,25,26	21,25
D2S411	11.3,14	11.3,14
The ExPASy database has a matching rate of 98.51%，The number of matched bits is 19 (https://web.expasy.org/cellosaurus-str-search/)		

Note:

1. According to the cell STR identification standard established by the International Cell Line Authentication Committee (ICLAC), when the matching degree of cell lines is ≥80%, they are considered to be correlated, that is, derived from common ancestral cells; The matching degree is between 55% and 80%, and the correlation needs to be further verified. Less than 55% indicates no correlation between the two.
2. The effective peak of the map was the real PCR band; Small peaks and nonspecific bands were ignored in the calculation.

Web: www.procell.com.cn
Tel: 400-999-2100; 027-87287608
Email: sales@procell.com.cn
Add: Building C4,3rd Phase of Biomedical Park, #858 Gaoxin Road, Donghu Hi-Tech Development Area,Wuhan,Hubei,China

Appendix II: The genotyping results of STR locus of Hep G2 cell line.



Web: www.procell.com.cn
Tel: 400-999-2100; 027-87287608
Email: sales@procell.com.cn
Add: Building C4,3rd Phase of Biomedical Park, #858 Gaoxin Road,
Donghu Hi-Tech Development Area Wuhan Hubei China