周总结

1 PR 与 RF 的细胞类型甲基化谱的 MAE 和相关系数

1.1 代码实现

```
#计算 PR 与 RF 得出的细胞类型的 DNA 甲基化矩阵与真实甲基化矩阵的相关性和 MAE
source("functions0725.R")

# 导入全血样本数据
load("WB.Rdata")

# 导入真实比例
load("WB_Truepro.Rdata")

# 取消科学计数法
options(scipen = 200) #scipen 表示在 200 个数字以内都不使用科学计数法

CD4T = WB[,7:12]
CD8T = WB[,13:18]
Monocyte = WB[,19:24]
B = WB[,25:30]
NK = WB[,31:36]
Neutrophil = WB[,37:42]
Eosinophil = WB[,43:48]

#得到处理后各细胞类型的数据
CD4T_dell <- as.matrix(apply(CD4T,1,mean))
CD8T_dell <- as.matrix(apply(CD8T,1,mean))
Monocyte_dell <- as.matrix(apply(Monocyte,1,mean))
B_dell <- as.matrix(apply(B,1,mean))
NK_dell <- as.matrix(apply(NK,1,mean))
Neutrophil_dell <- as.matrix(apply(Neutrophil,1,mean))
Eosinophil_dell <- as.matrix(apply(Eosinophil,1,mean))

WB_Y = WB[,1:6]
WB_W =
cbind(CD4T_dell,CD8T_dell,Monocyte_dell,B_dell,NK_dell,Neutrophil_dell,Eosinophil_dell)
colnames(WB_W) = c("CD4T","CD8T","Monocyte","B","NK","Neutrophil","Eosinophil")
rownames(WB_W) = rownames(WB_Y)

MAEout = c()
Corout = c()
for(featsize in c(1000,2000,3000,4000,5000)){
  #值的初始化
  Y = WB_Y
```

```
  W = WB_W
  #W1.index = 1:5
  W1.index = 1
  #H = WB_Truepro
  #  提取特征位点
  feat = select_feature(mat = as.matrix(Y),method = "cv",nmarker = featsize,startn = 0)

  Y = as.matrix(Y[feat,])
  W = as.matrix(W[feat,])
  W1 = as.matrix(W[feat,W1.index])

  result_CellTypeMethyMAE=list()#PR 与 RF 得到的预测的细胞类型甲基化矩阵与真实矩阵之间的 MAE
  result_CellTypeMethyCor=list()#PR 与 RF 得到的预测的细胞类型甲基化矩阵与真实矩阵之间的相关性

  #----test RF----
  #RF 是无参考的方法
  #输入：
  #W1 = as.matrix(W[feat,W1.index])

  cat("\nRunning RF ...\n")
  #RF 中参数 W 的值为 W 时，得出的预测结果就与 W 同样大小，除非 K 取 1 时，得到的预测结果为 3 列
  #这是无参考方法，但是 K 的值不能取所有细胞类型的个数，必须小于等于 Y 的列数
  RFout <- RF(Y = Y,W = NULL,K = 6,type = "ME",iters = 1000)
  #由于 RF 是无参考的，K 最多只能取 6，得到的预测 W 只有 6 列，因此需要给最后的结果增加一列值
  #在这里，我们取最后两列的平均值作为最后一列
  addRFout_Wcol <- apply(RFout$W[,5:6], 1, mean)
  RFout$W <- cbind(RFout$W,addRFout_Wcol)
  RFout_CellTypeMethyMAE = mean(colSums(abs(W-RFout$W)))#RF 细胞类型甲基化矩阵与真实矩阵的 MAE
  result_CellTypeMethyMAE$RFout_CellTypeMethyMAE = RFout_CellTypeMethyMAE
  RFout_CellTypeMethyCor = mean(diag(cor(W,RFout$W)))#RF 细胞类型甲基化矩阵与真实矩阵的 MAE
  result_CellTypeMethyCor$RFout_CellTypeMethyCor = RFout_CellTypeMethyCor
  #-----end-----

  #-----test PR-----
  cat("\nRunning PR ...\n")
  PRout <- PR(Y = Y,W = W,W1 = W1,type = "ME",K = ncol(W),iters = 1000)
  PRout_CellTypeMethyMAE = mean(colSums(abs(W-PRout$W)))#PR 细胞类型甲基化矩阵与真实矩阵的 MAE
```

```
  result_CellTypeMethyMAE$PRout_CellTypeMethyMAE = PRout_CellTypeMethyMAE
  PRout_CellTypeMethyCor = mean(diag(cor(W,PRout$W)))#PR 细胞类型甲基化矩阵与真实矩阵的 MAE
  result_CellTypeMethyCor$PRout_CellTypeMethyCor = PRout_CellTypeMethyCor
  #-----end-------

  #最终输出，用于作图
  row1 = cbind(featsize,RFout_CellTypeMethyMAE,PRout_CellTypeMethyMAE)
  row2 = cbind(featsize,RFout_CellTypeMethyCor,PRout_CellTypeMethyCor)
  MAEout = rbind(MAEout,row1)
  Corout =rbind(Corout,row2)
}
#保存数据
colnames(MAEout) <- c("featsize","RF","PR")
colnames(Corout) <- c("featsize","RF","PR")
save(MAEout,file = "MAEout.Rdata")
save(Corout,file = "Corout.Rdata")

#将 MAEout 和 Corout 的数据类型变为数据帧类型
MAEout = as.data.frame(MAEout)
Corout = as.data.frame(Corout)

#melt 函数根据 featsize 这一列整理原数据
drawMAEout = melt(MAEout,id="featsize")
drawCorout = melt(Corout,id="featsize")

#更改列名
colnames(drawMAEout) = c("featsize","method","value")
colnames(drawCorout) = c("featsize","method","value")

#作图
#RF 与 PR 细胞类型甲基化谱的 MAE 值图
MAEFigure = ggplot(drawMAEout, aes(x=featsize, y=value, group=method,color = method)) +
  geom_line(aes(linetype=method))+
  scale_color_manual(values=brewer.pal(8, "Dark2")[c(2,1)]) +
  theme_classic()+
  geom_point(aes(shape=method)) +
  #geom_errorbar(aes(ymin=corH-sd, ymax=corH+sd), width=.1) +
  xlab("Sample size")+
  ylab("Mean absolute error of cell types methylation profile")+
  labs(fill = "Method")

#RF 与 PR 的细胞类型甲基化谱的相关系数图
CorFigure = ggplot(drawCorout, aes(x=featsize, y=value, group=method,color = method)) +#数据
```

```
导入，aes()用于对数据中的属性进行映射
  geom_line(aes(linetype=method))+#根据 RF 和 PR 画线
  #brewer.pal(8, "Dark2")[c(2,1)]表示取调色板中 Dark2 的 8 种颜色中的两种
  scale_color_manual(values=brewer.pal(8, "Dark2")[c(2,1)]) +#自定义颜色
  theme_classic()+#添加主题
  geom_point(aes(shape=method)) +#给每一条数据添加点，不同类型的数据添加不同形状的点
  #geom_errorbar(aes(ymin=drawCorout$value-sd, ymax=drawCorout$value+sd), width=.1) +
  xlab("Sample size")+
  ylab("Correlation of cell types methylation profile")+
  labs(fill = "Method")

#两图融合为一图
p =ggarrange(MAEFigure,CorFigure,ncol=2,nrow=1,labels=c("A","B"))
ggsave(plot = p,filename="MAEandCor.pdf",width=9, height=4.4)
```

1.2 结果展示

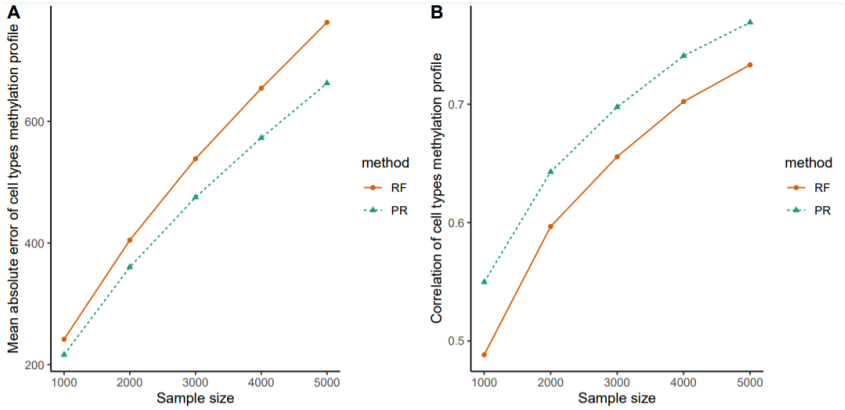

图 1 结果展示