LUAD肺腺癌

1. 收集TCGA-LUAD数据

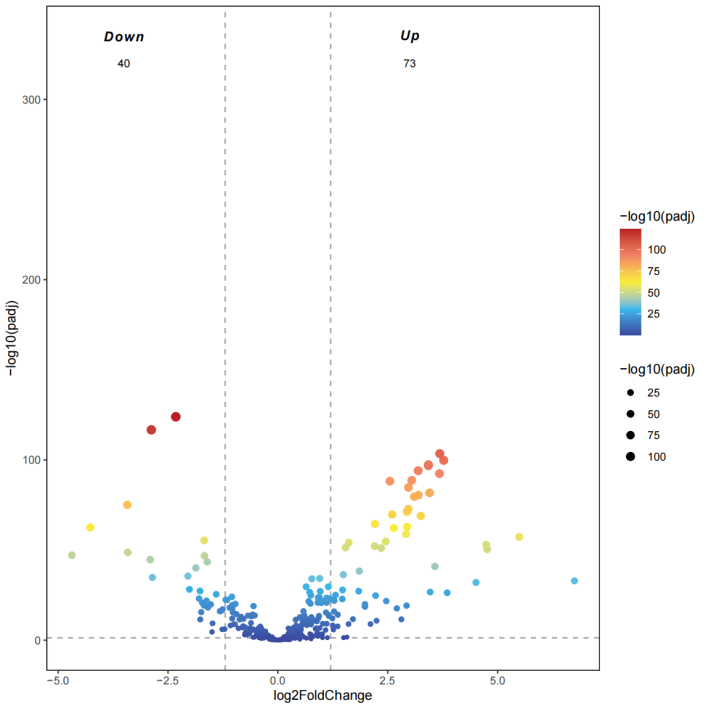
Data（LUAD数据）

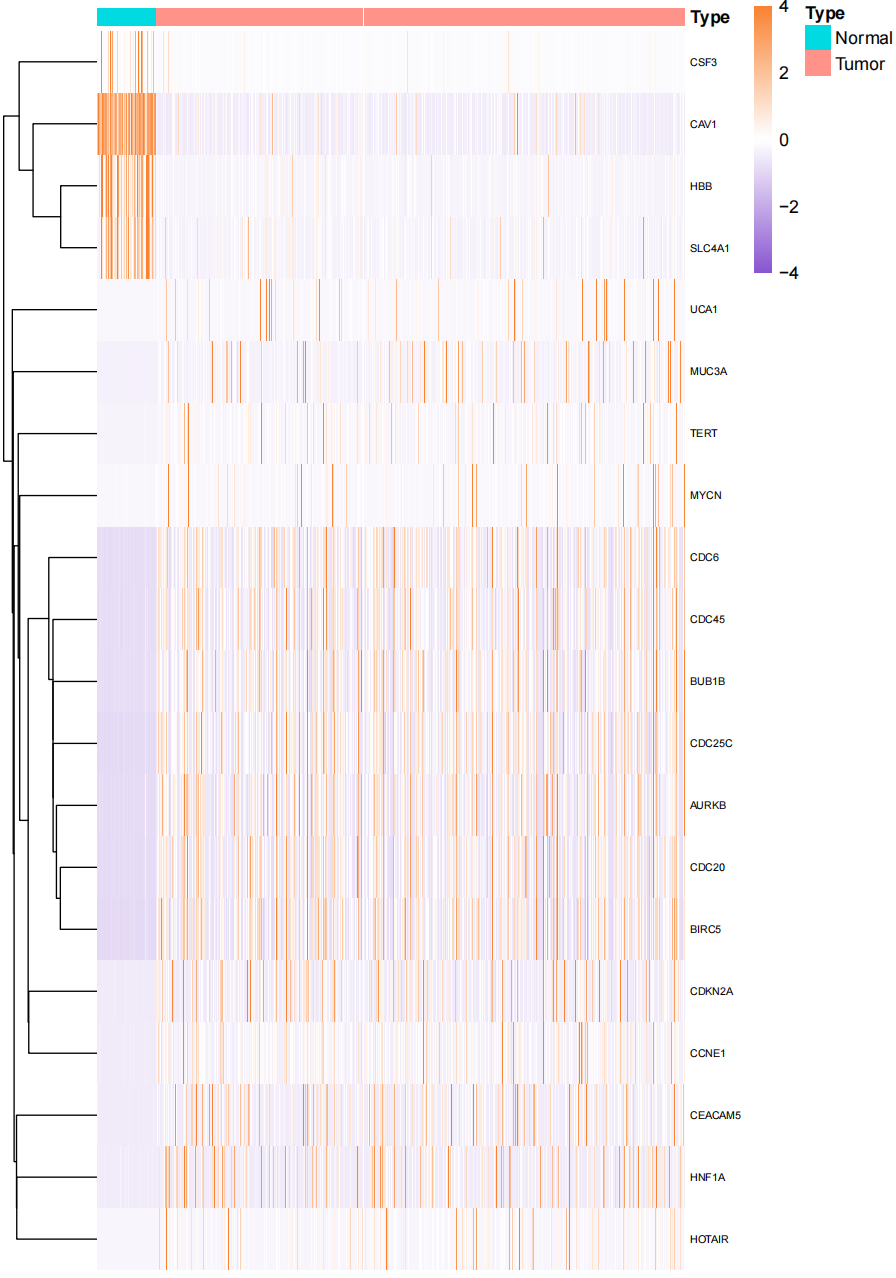
从TCGA中共收集592条转录组数据。

1. 识别与细胞周期相关的差异表达基因

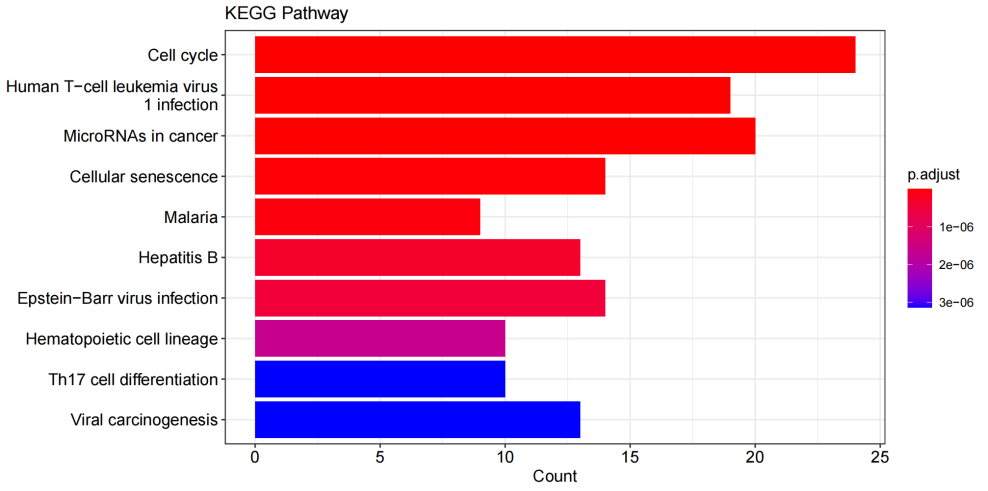
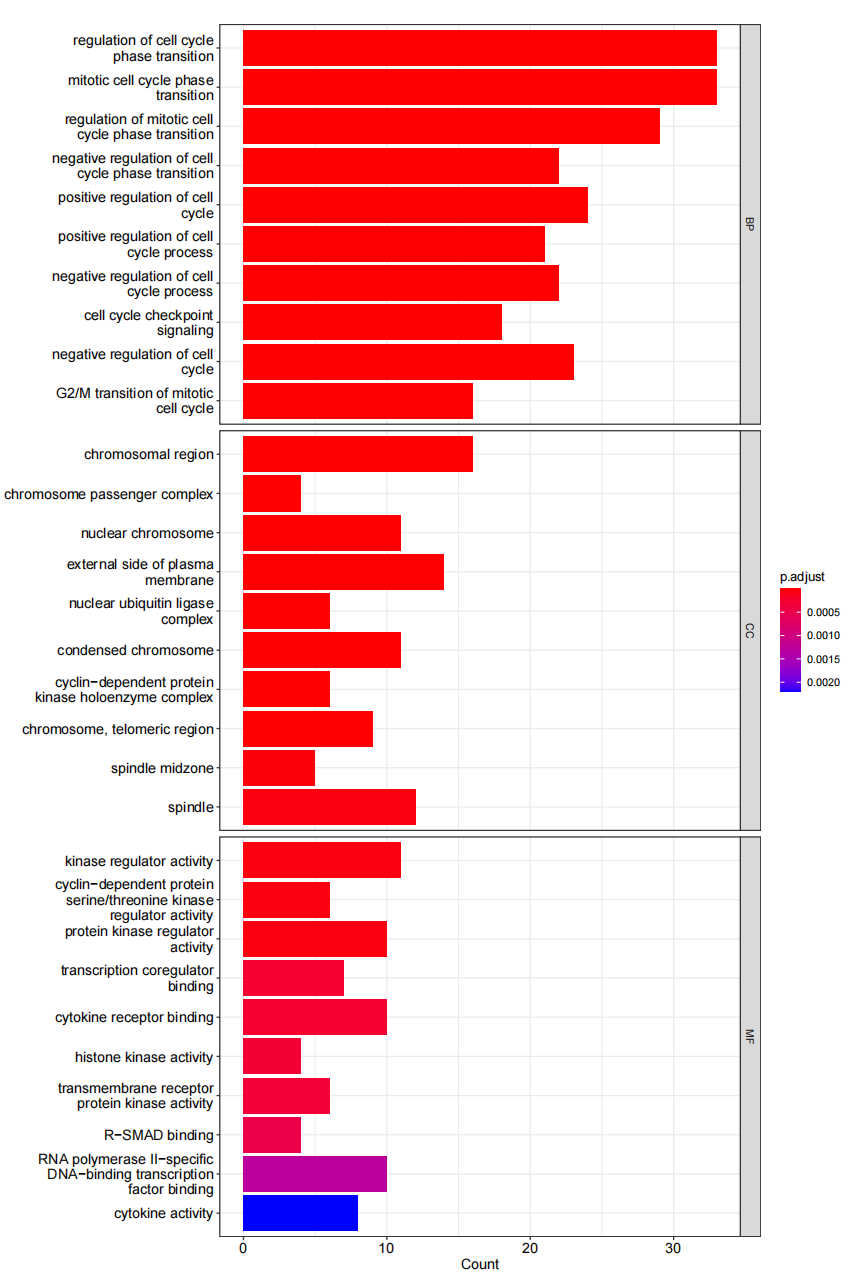
Data（GeneCards-细胞周期相关评分大于30.xlsx、all\_diff.csv、GO\_df.csv、KEGG\_df.csv、gene\_entriz.csv）

Picture（express.pdf、volco.pdf）

1、使用R包DESEQ，用来差异分析人的肿瘤转录组和正常转录组得到的差异基因与细胞周期基因取交集，做火山图以及表达量热图，得到上调的基因有73个，下调的基因有40个。其中我们关注的**CDC20**这个基因为上调基因，它的差异倍数为3.773082。



1. 使用R包clusterProfiler根据差异基因和周期基因得到的交集基因做GO和KEGG功能富集分析(picture:go\_bars.pdf、kegg\_bar.pdf)

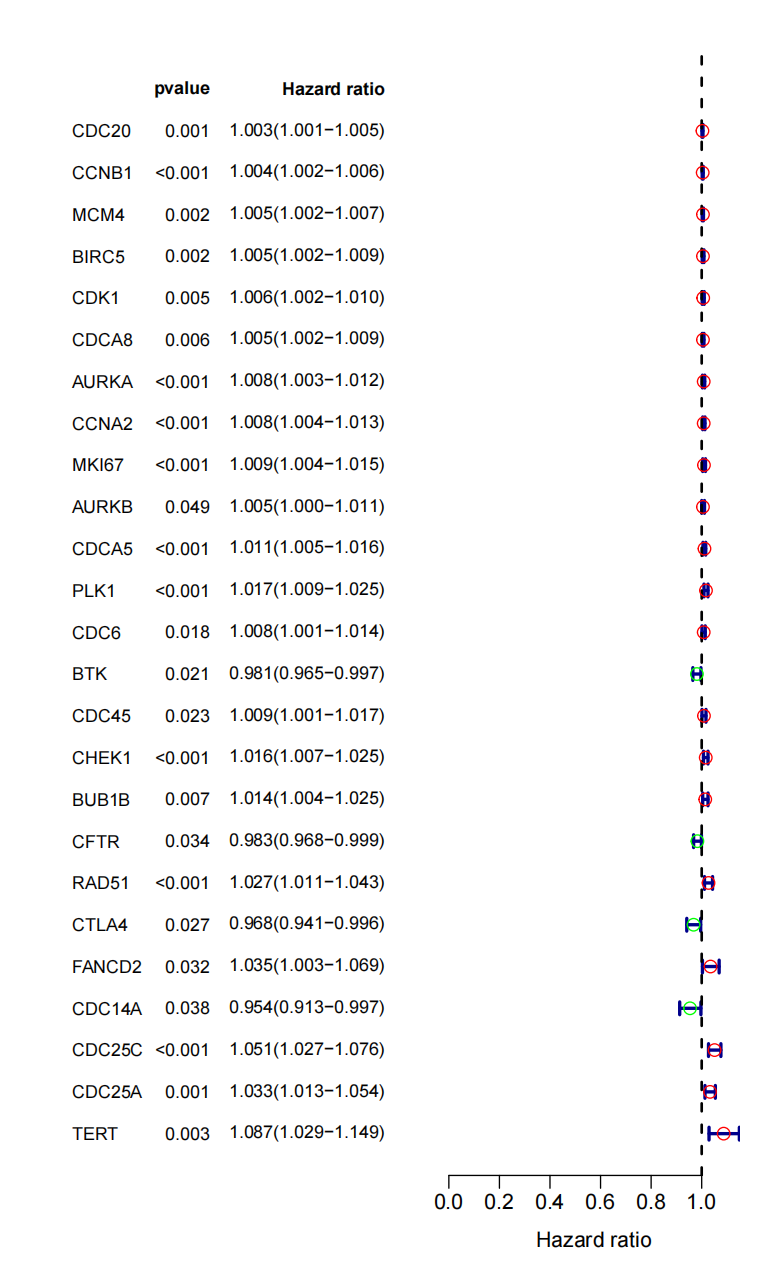


1. 单因素COX回归筛选与LUAD患者生存相关的差异表达基因

Data（UniCoxSurvival.csv）

1. 单因素COX回归分析鉴定训练集中差异细胞周期相关基因与生存之间的关系（p值＜0.05）

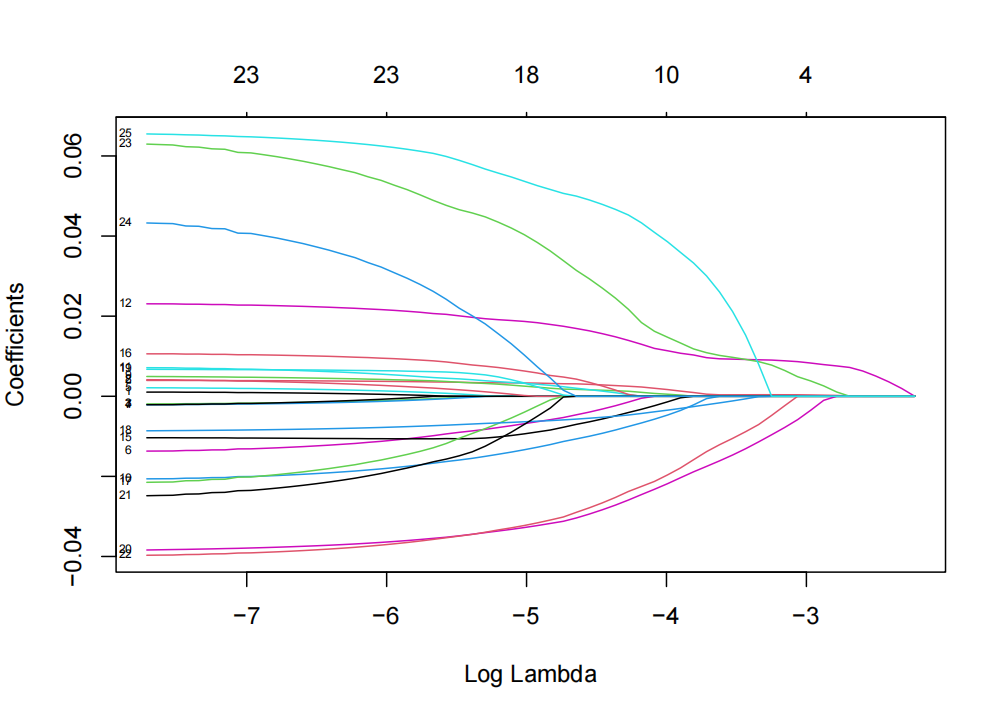
，使用R包survival做单因素COX回归分析，我们找到了25个有趣的基因，其中CDC20也在其中。(picture:UniCoxSurForestPlot.pdf)

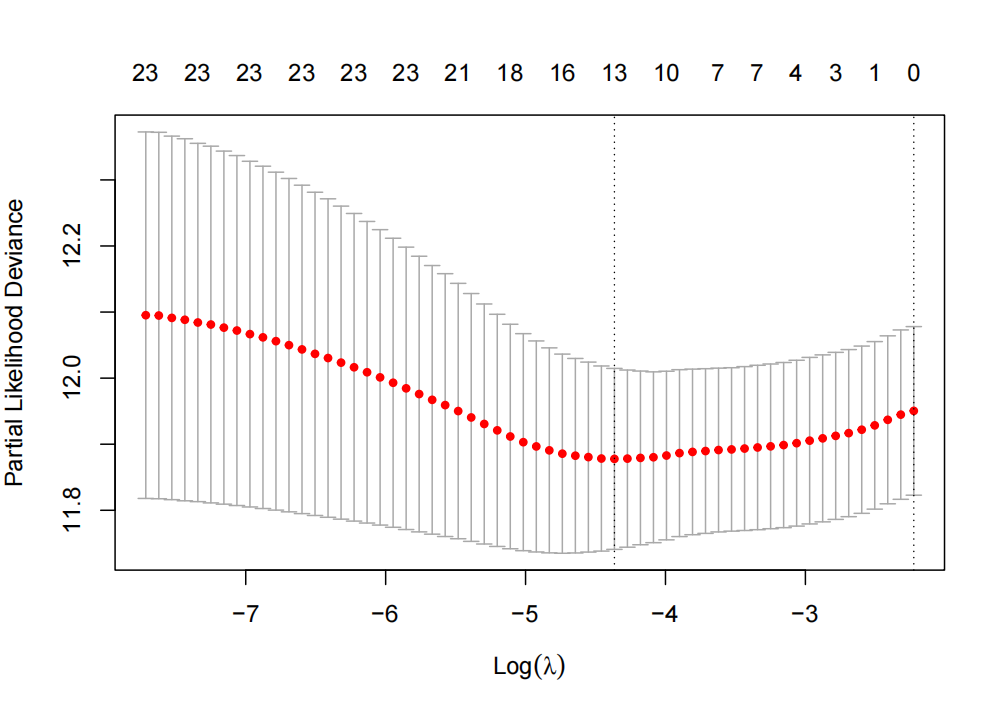


1. LASSO+COX回归建立细胞周期相关基因的预后模型

Data（train.csv、gsea\_df.csv）

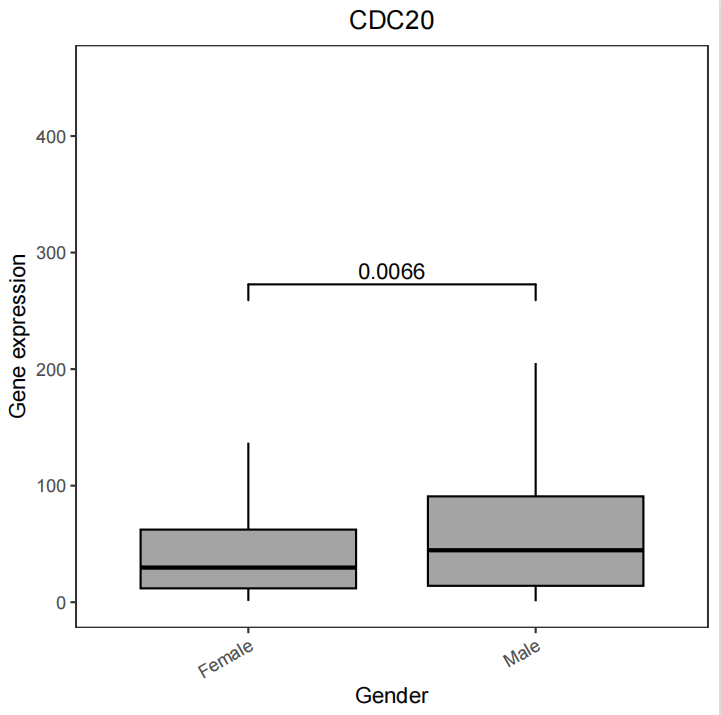
1. 对p值＜0.05的差异细胞周期相关基因做LASSO+COX回归分析，使用R包glmnet对单因素COX回归分析得到的基因，做LASSO+COX回归分析，从结果中我们找到22个有趣的基因，供我们接下来的研究。(picture:lassocox.pdf、lasso\_fit.pdf)

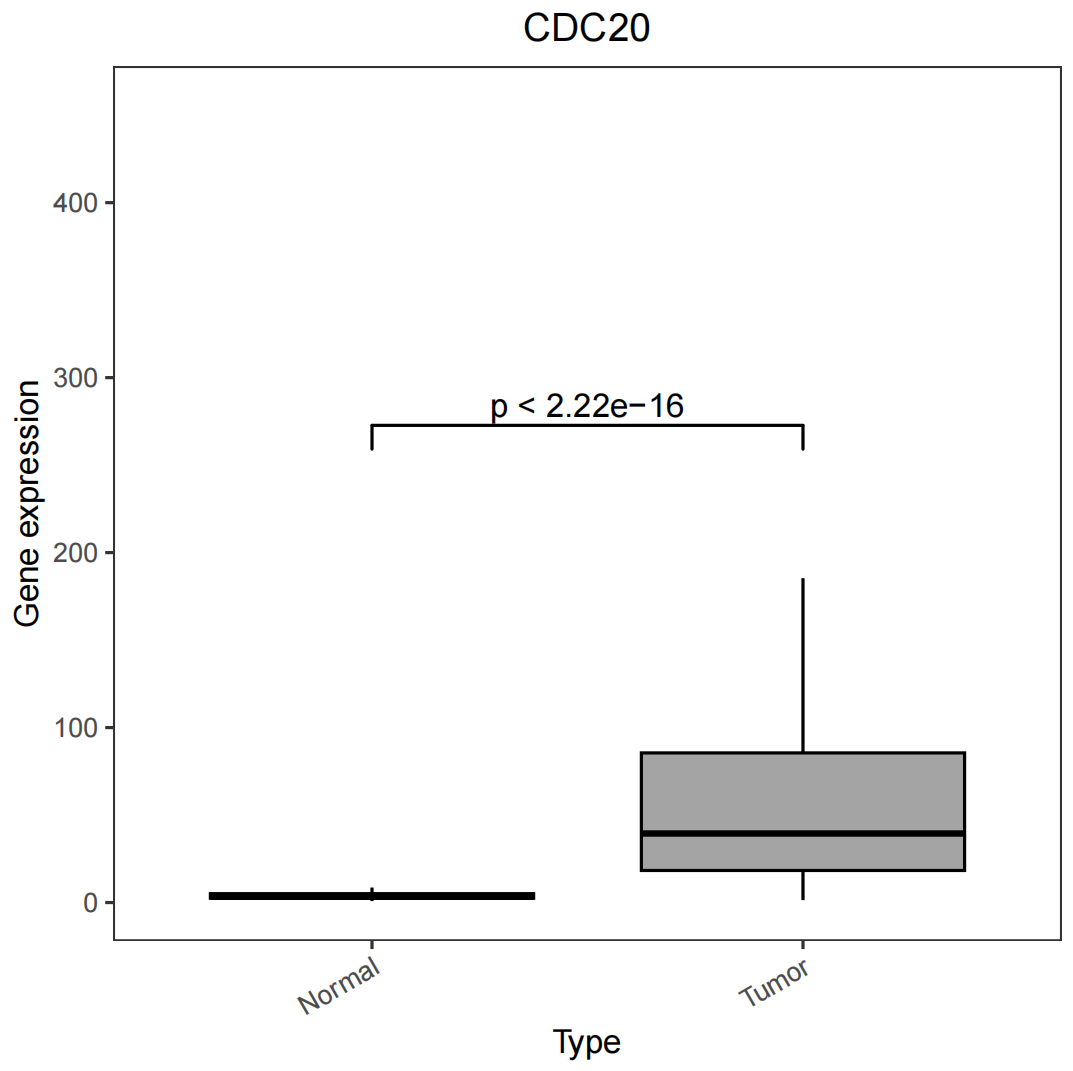


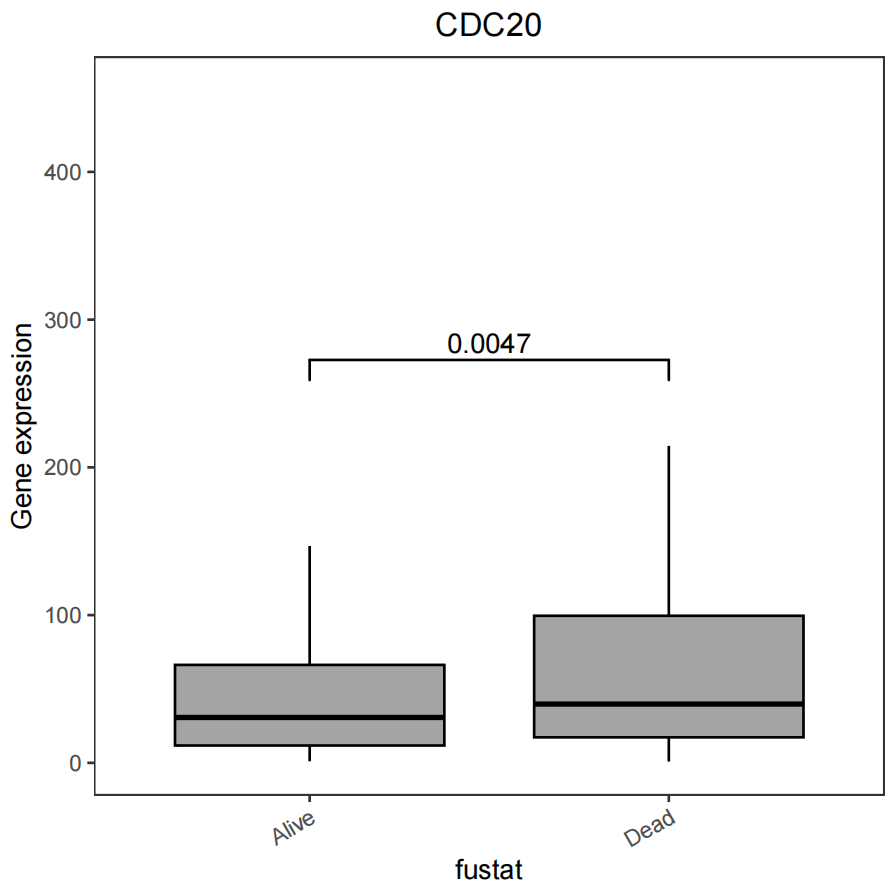


1. 细胞周期相关基因预后特征与临床关系

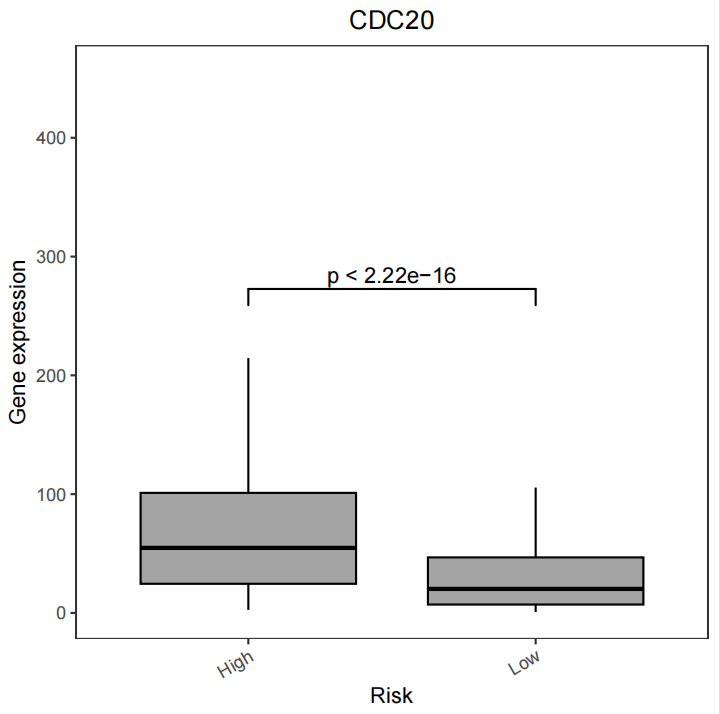
我们通过患者的性别以及患者的高低风险以及患者是否患有肿瘤以及患者是否生存，这四个角度描绘了CDC20与的表达状态，发现CDC20的表达与这些表型有潜在的关系(picture:cdc20\_gender.pdf、cdc20\_normal.pdf、cdc20\_sur.pdf)





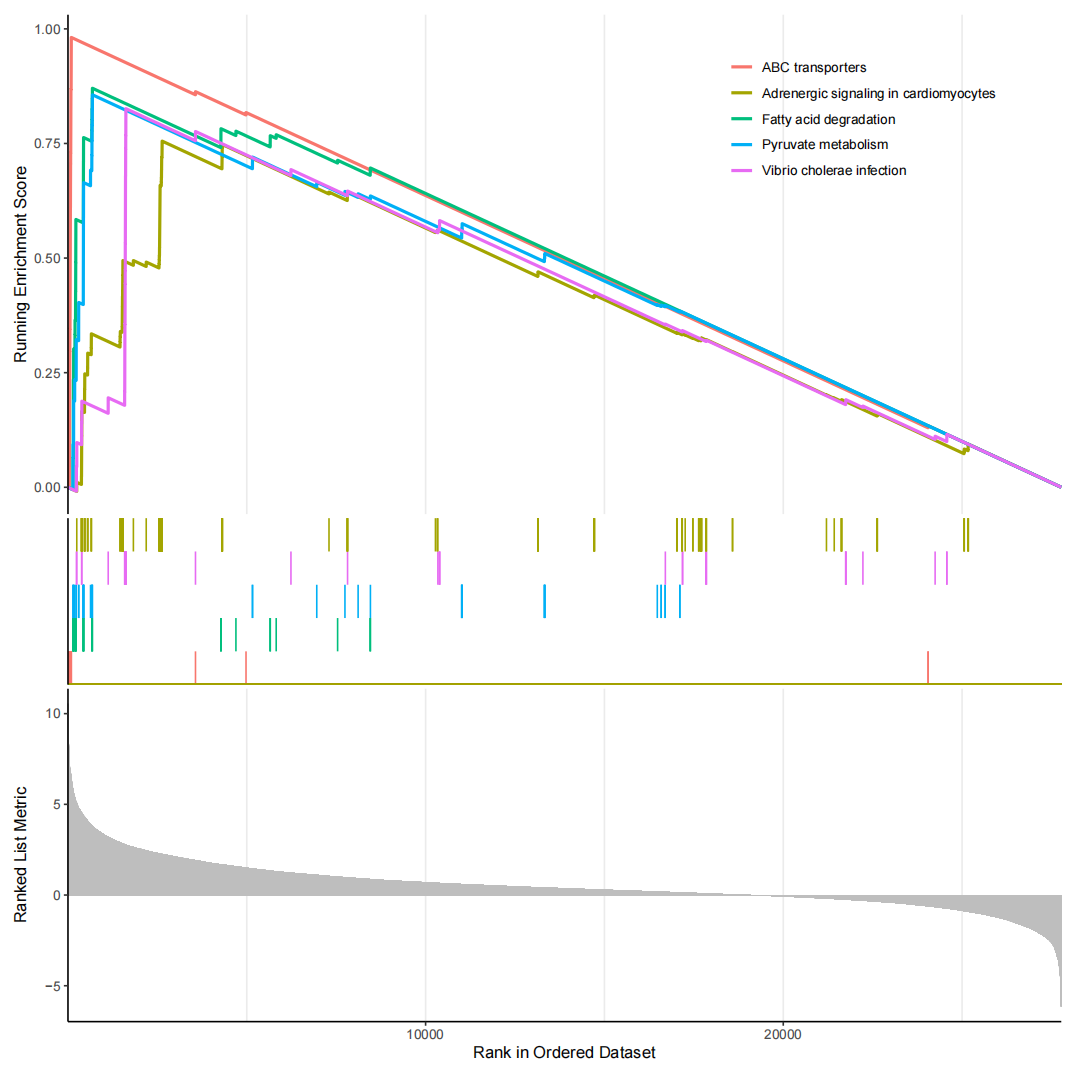


1. 预后特征中细胞周期相关基因的研究（在高危组和底危组中mRNA的表达水平）(picture:cdc20\_risk.pdf)



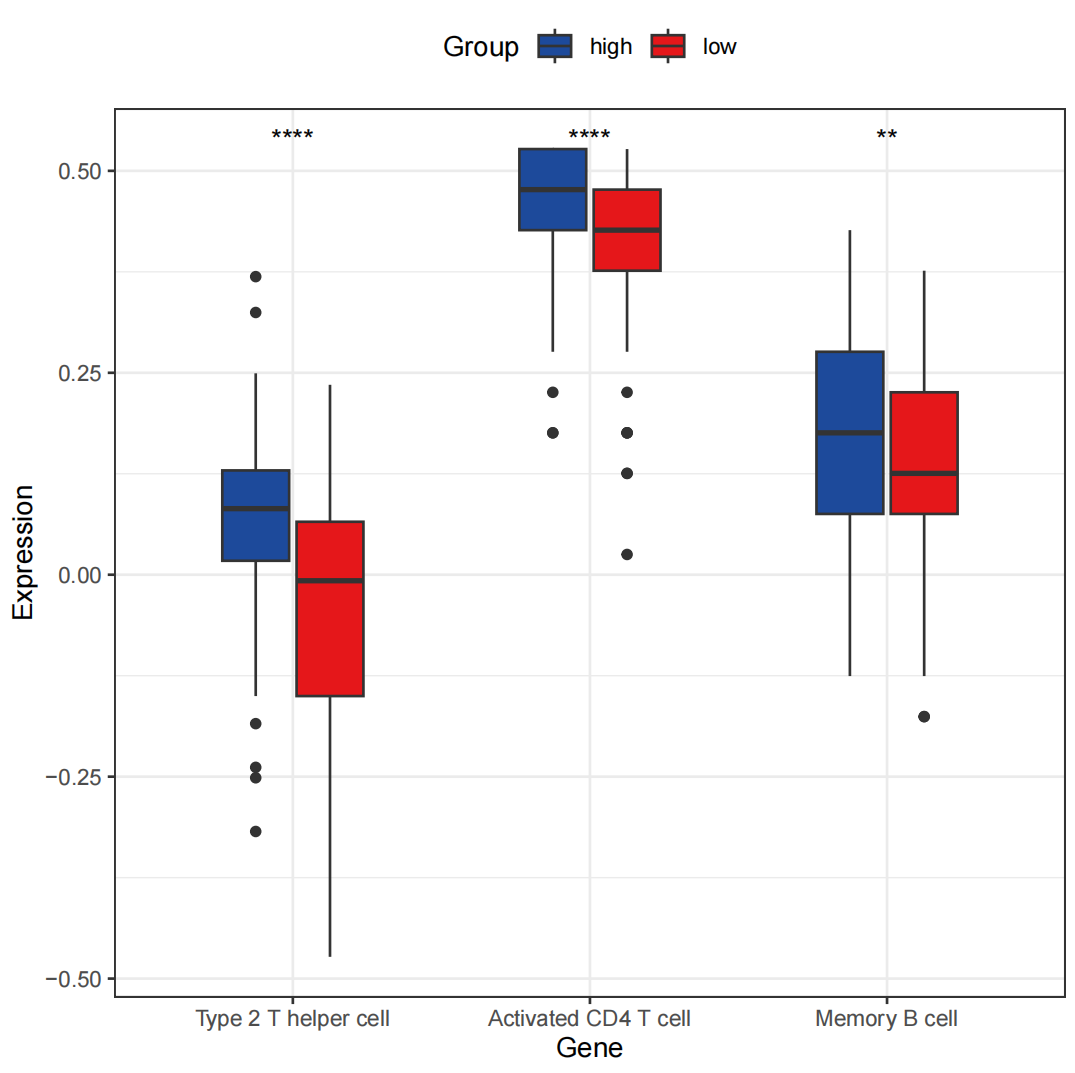
1. GSEA富集分析高危组和底危组参与的信号通路

用差异基因做gsea富集分析，我们以P小于0.05为阈值，找到61调kegg通路，我们画了前5个富集通路。(picture:gsea\_rich.pdf)

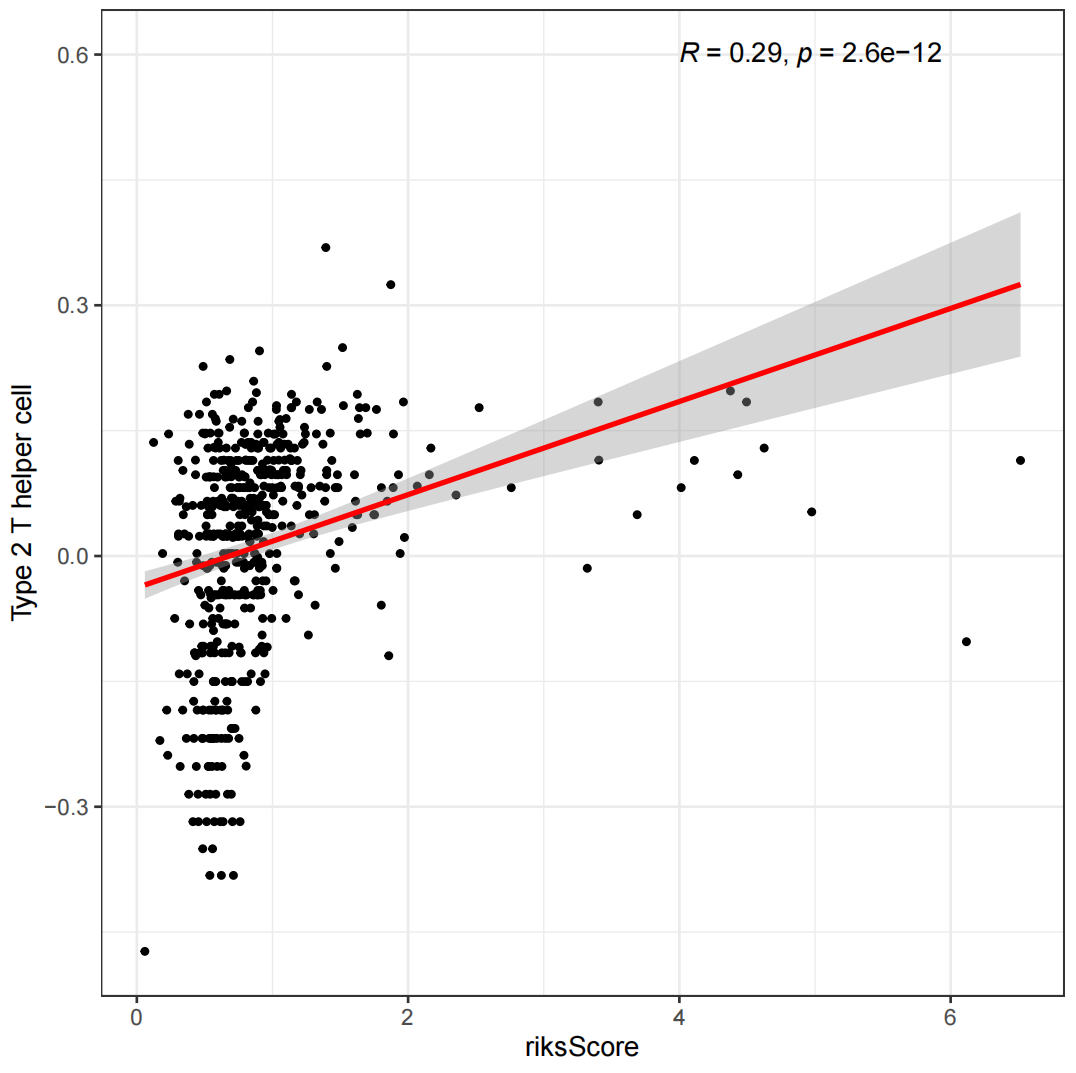
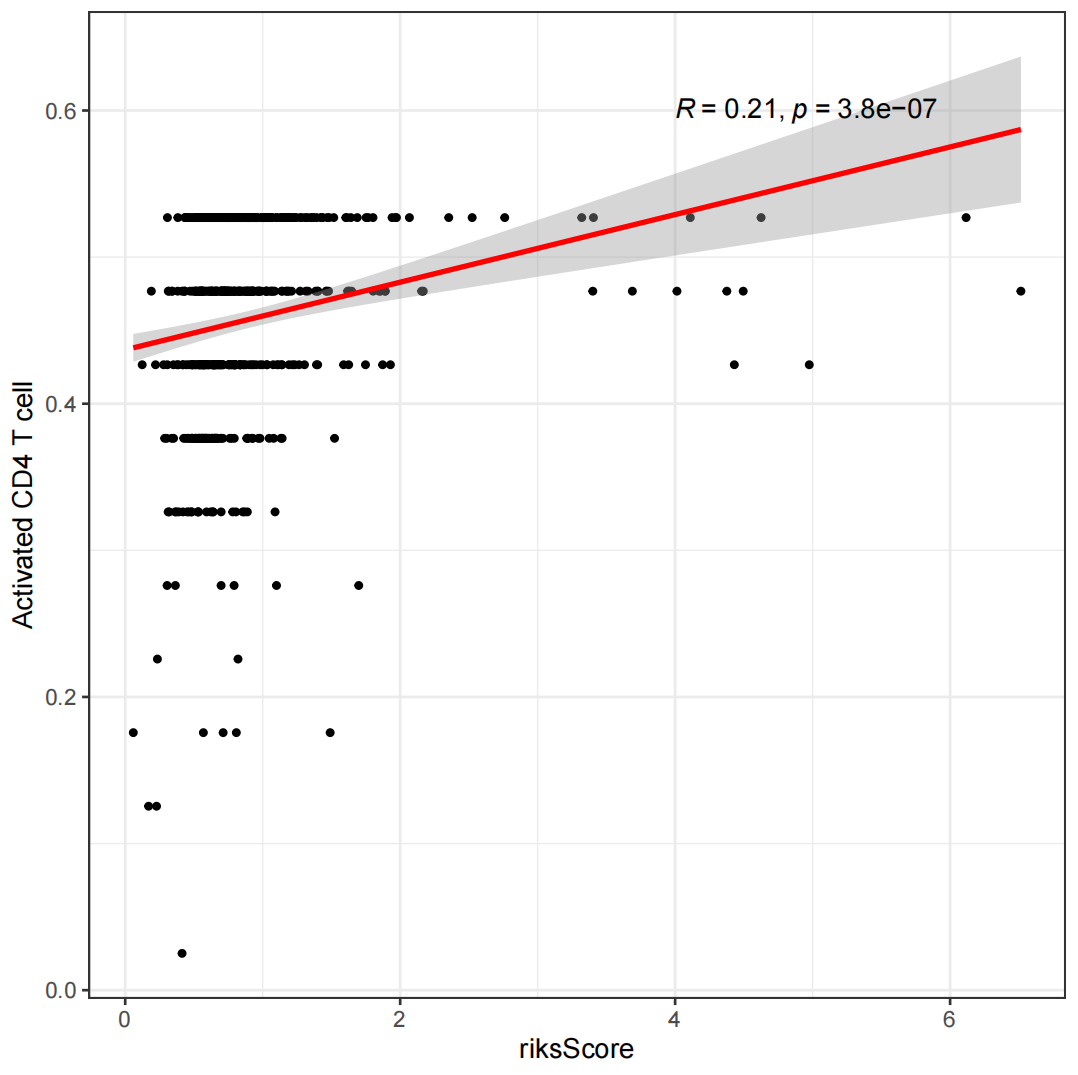
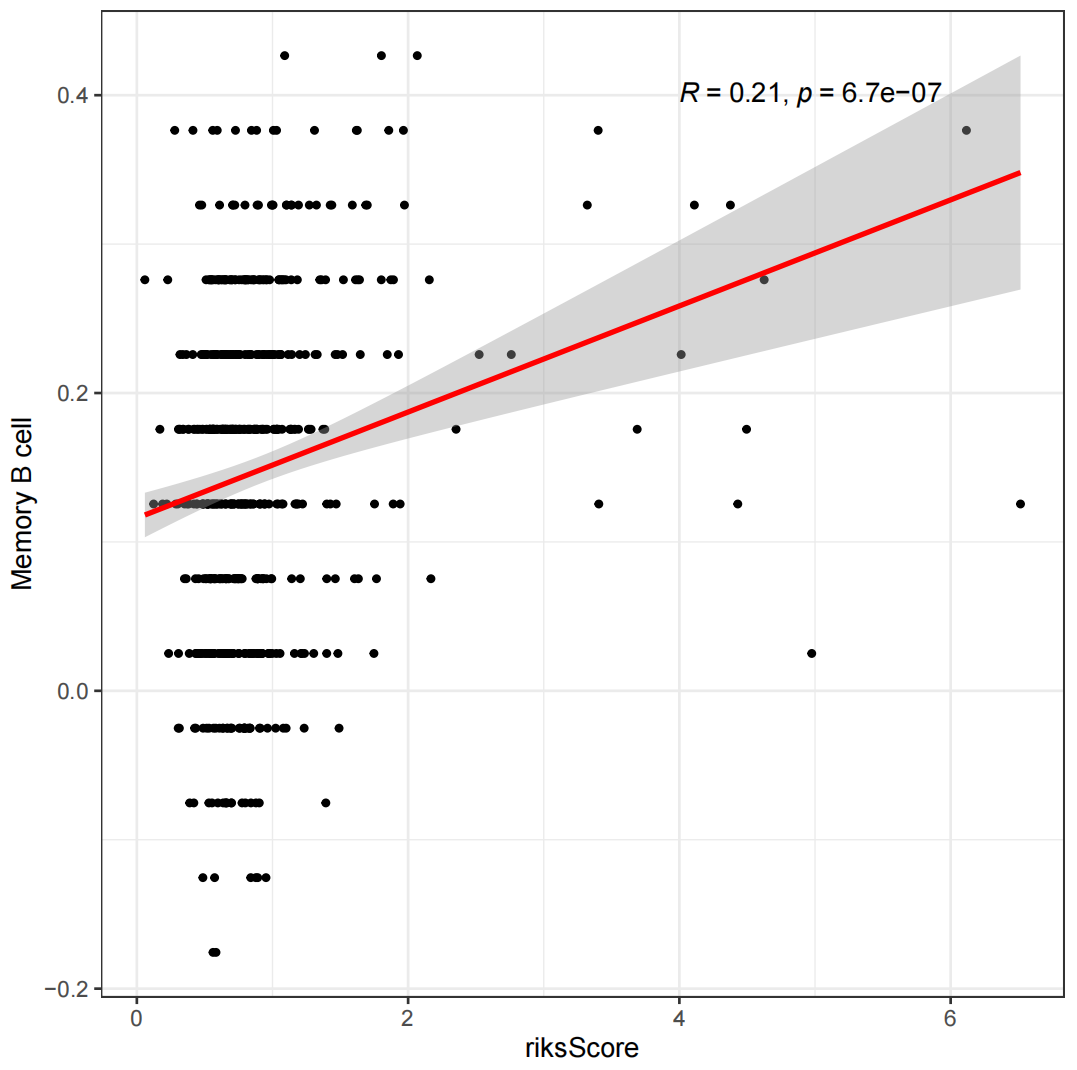


1. 分析预后模型风险评分与免疫浸润水平的相关性

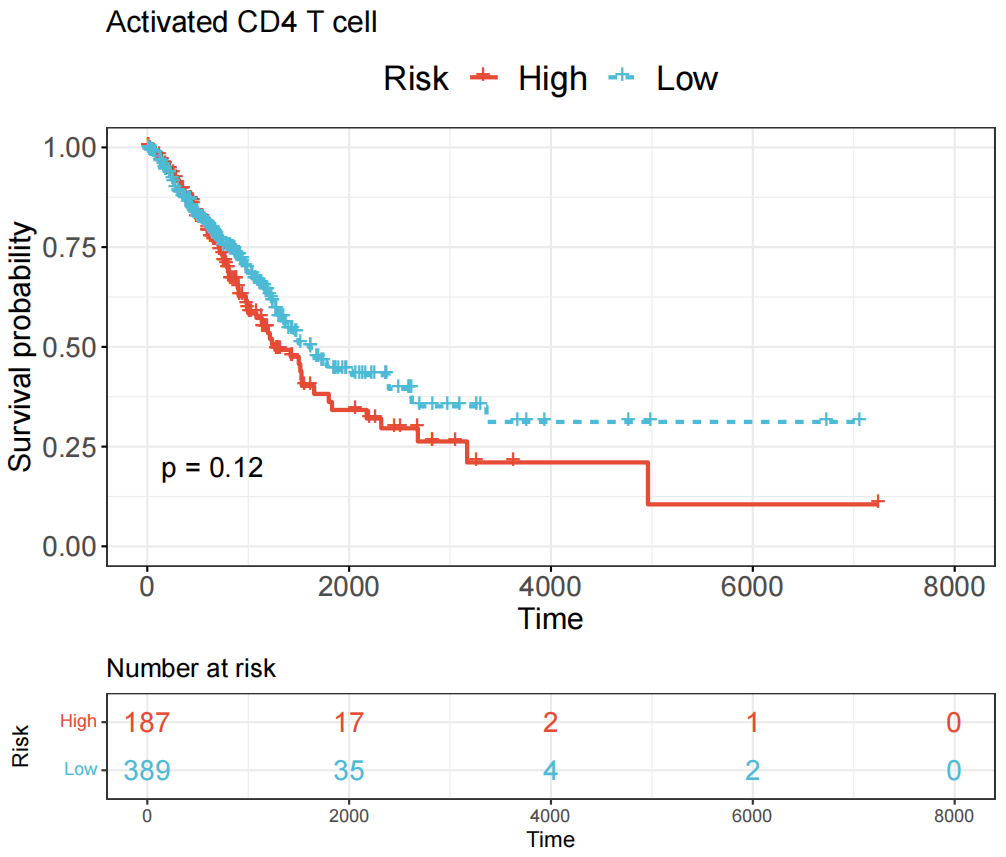
高危组和底危组免疫细胞亚群的差异表达(picture:mianyi\_cell.pdf)

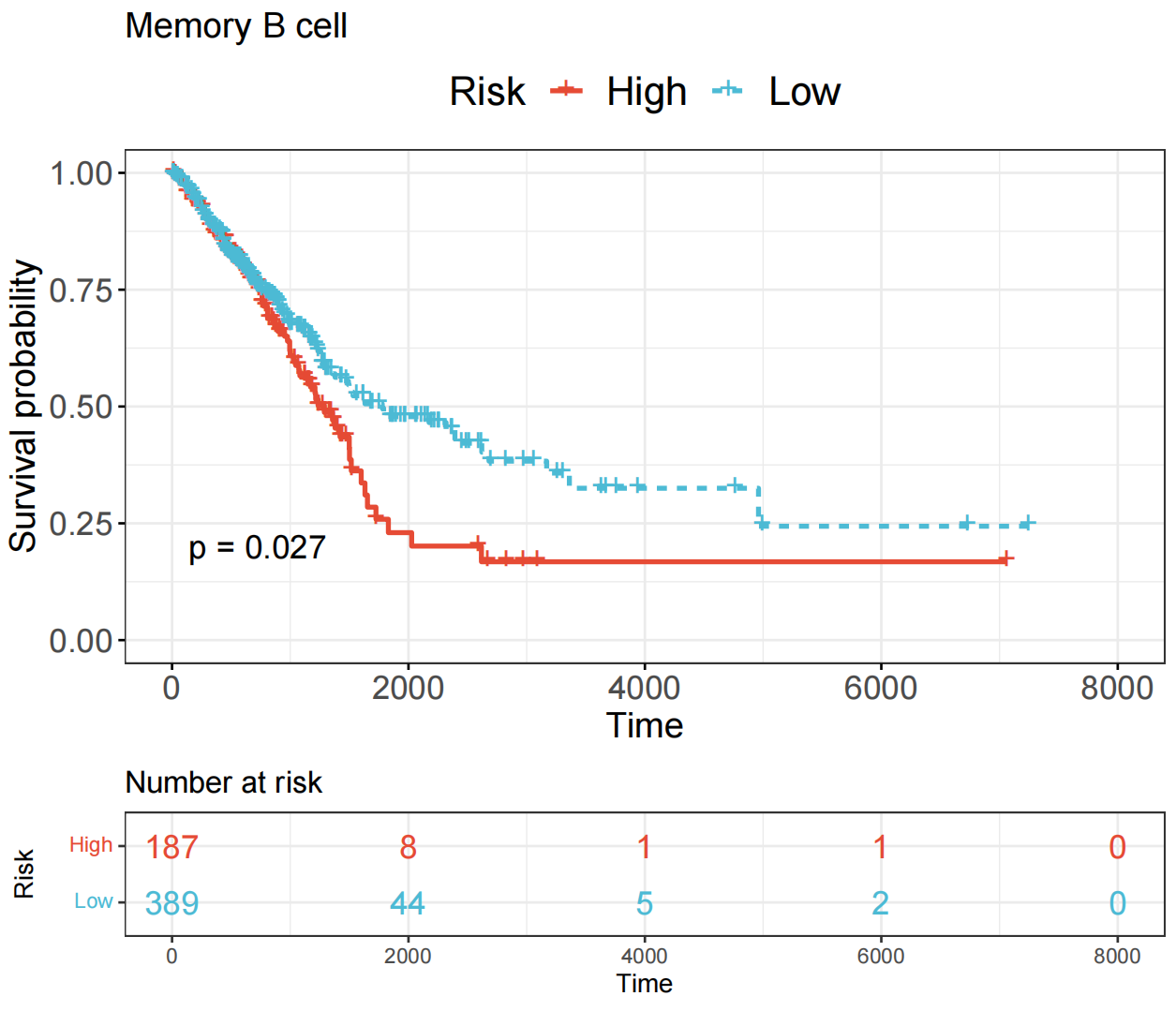


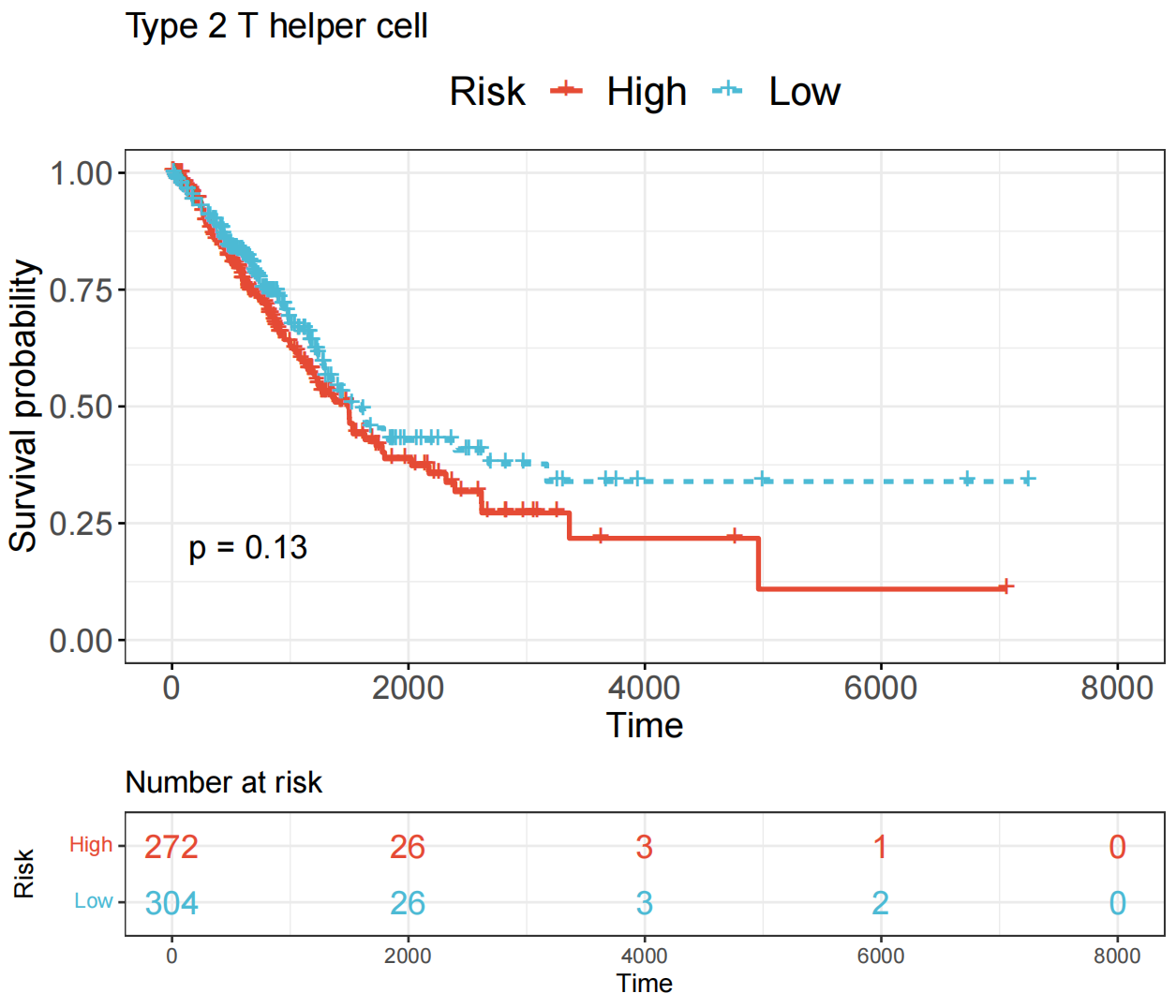
在训练集中，用Spearman相关性分析风险评分和差异免疫细胞浸润水平之间的相关性。我们富集到三类免疫细胞，与风险评分做相关性分析，发现风险评分与免疫细胞Type 2 T helper cell相关性最高，达到0.29。(picture:mianyin\_cell1.pdf、mianyin\_cell2.pdf、mianyin\_cell3.pdf)



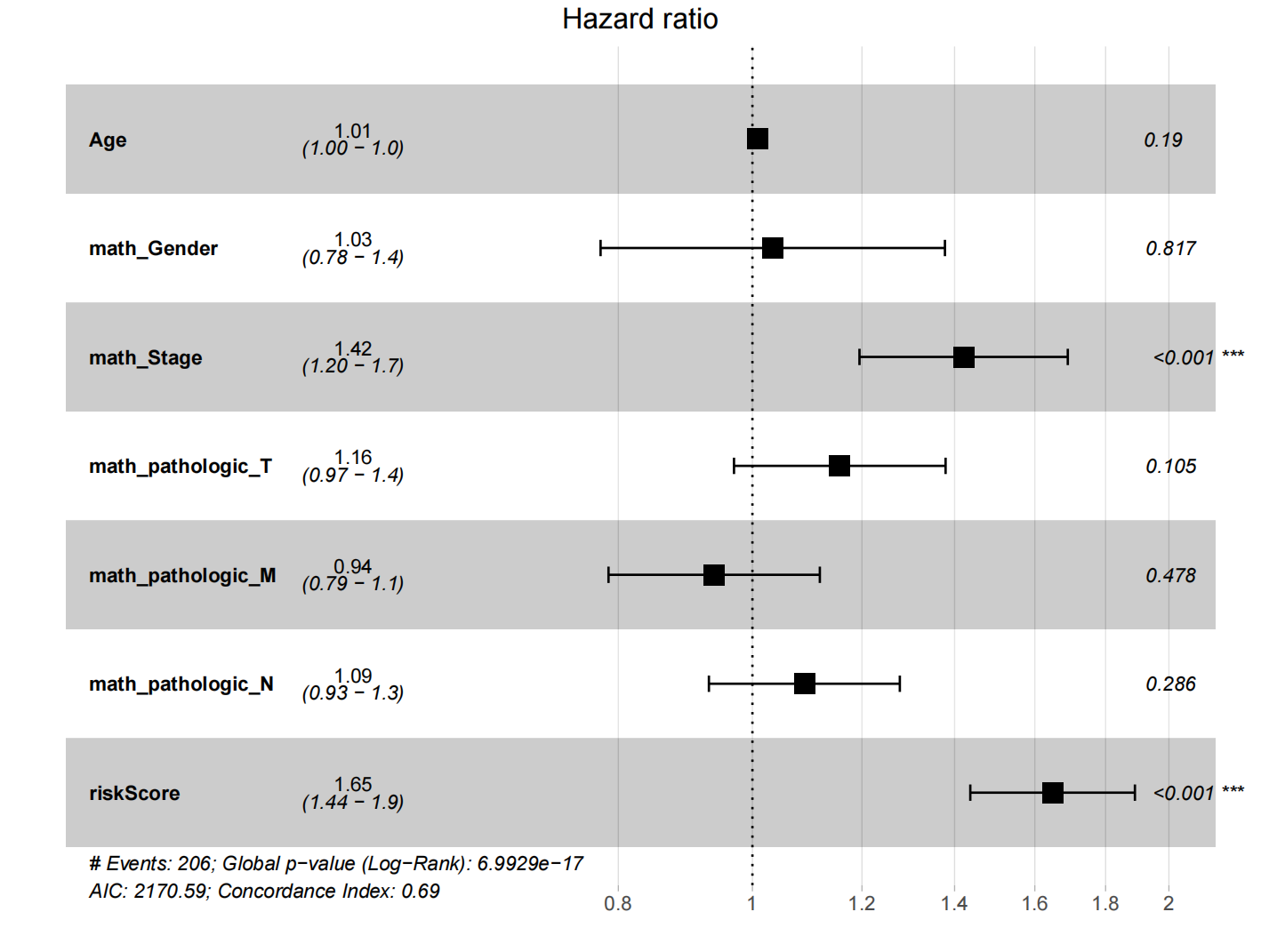
Kaplan-Meier生存分析高危和低危组之间相关性高的差异免疫细胞与预后的相关性，但是P值只有Memory B cell满足条件(picture:mianyi\_KM1.pdf、mianyi\_KM2.pdf、mianyi\_KM3.pdf)







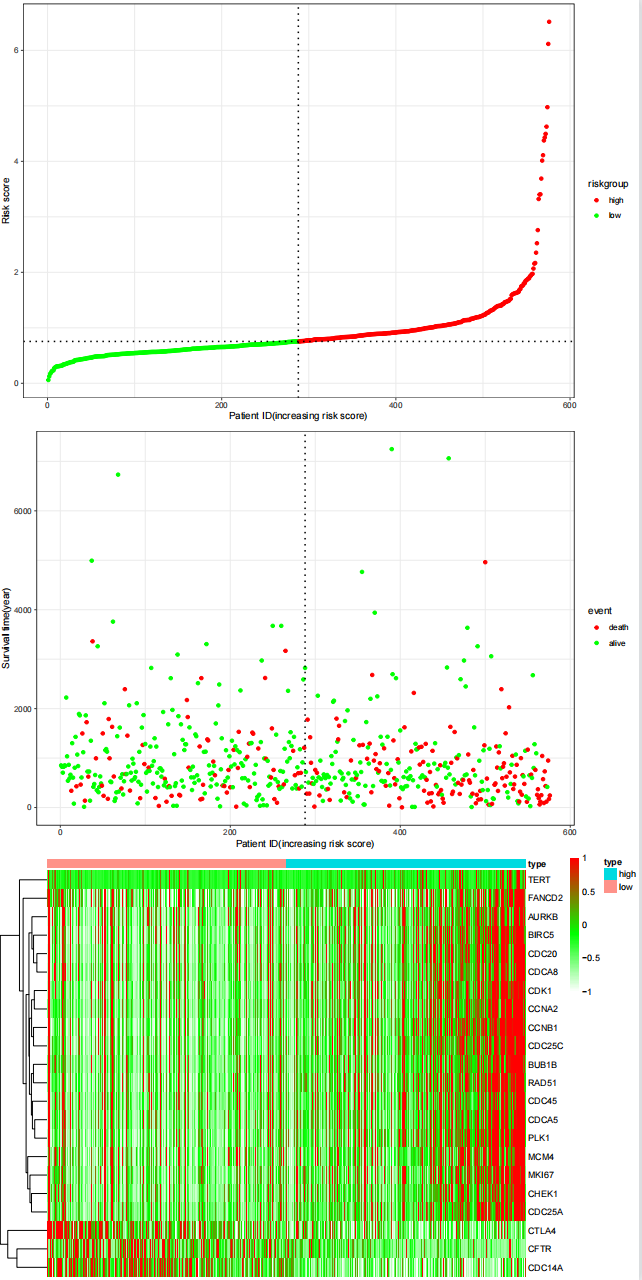
6、通过多因素Cox回归分析，评估预后模型与年龄、性别、组织学分、病理分期和TNM分期等常见的临床因素(picture:snm\_mutil\_cox.pdf)



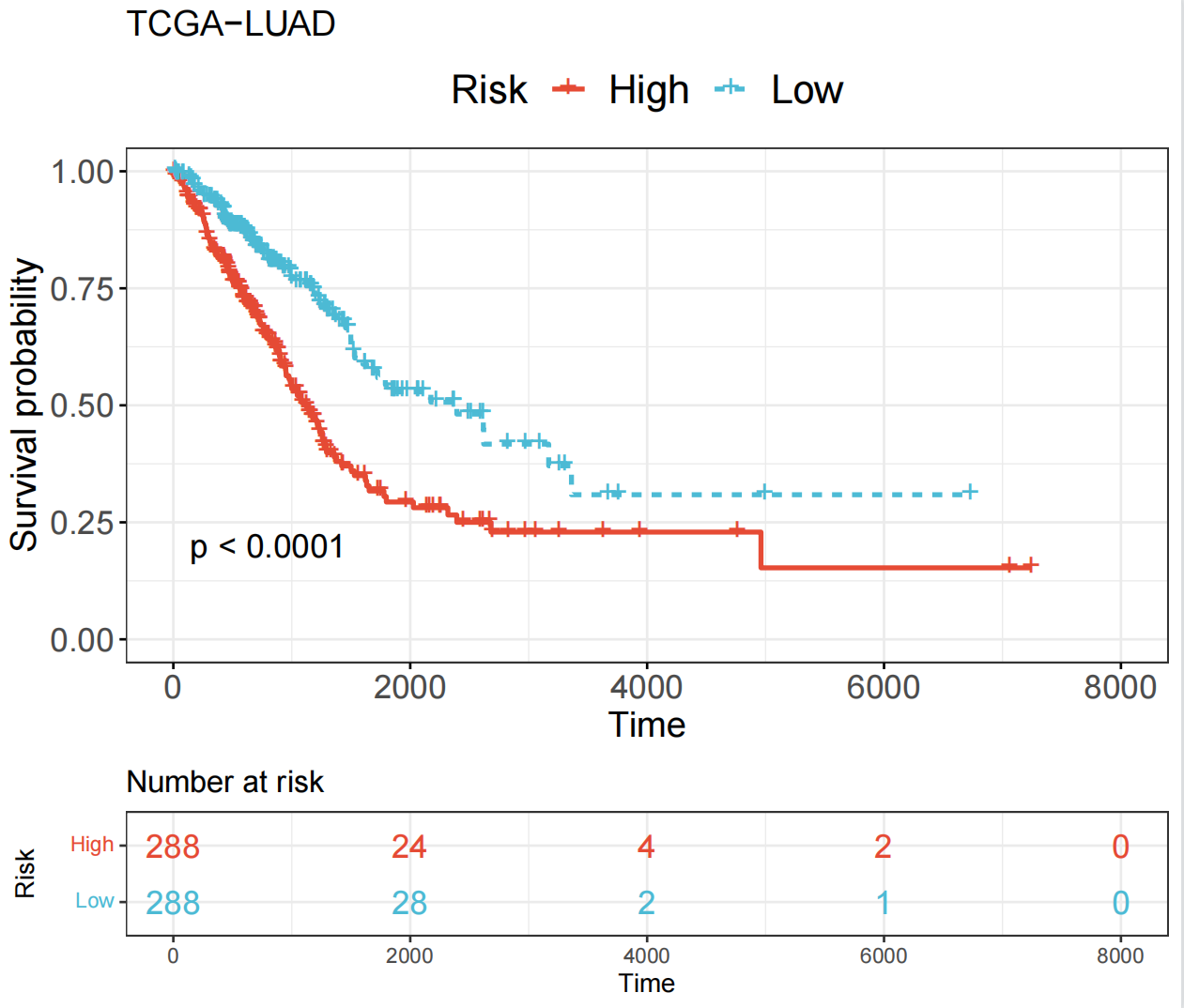
1. 基于预后模型进行风险评分，将患者分为底危组和高危组。

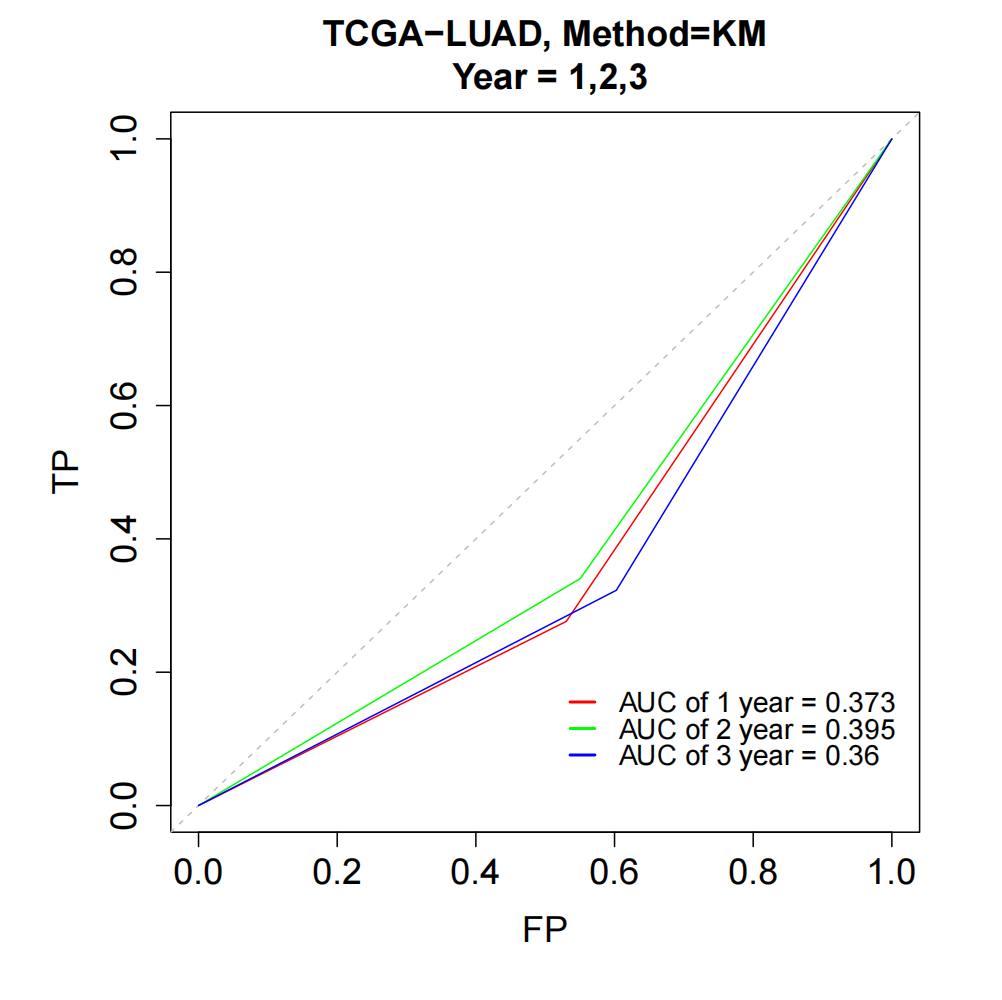
基于预后模型进行风险评分，将患者分为高危组和底危组,我们根据LASSO+COX找到的22个有趣的基因做了生存分析，发现高风险患者CDC20的表达要高于低风险患者。低风险患者生存状况更好。但是AUC曲线分析没有得到好的结果。

Data（train.csv）(picture:risk\_all.pdf)

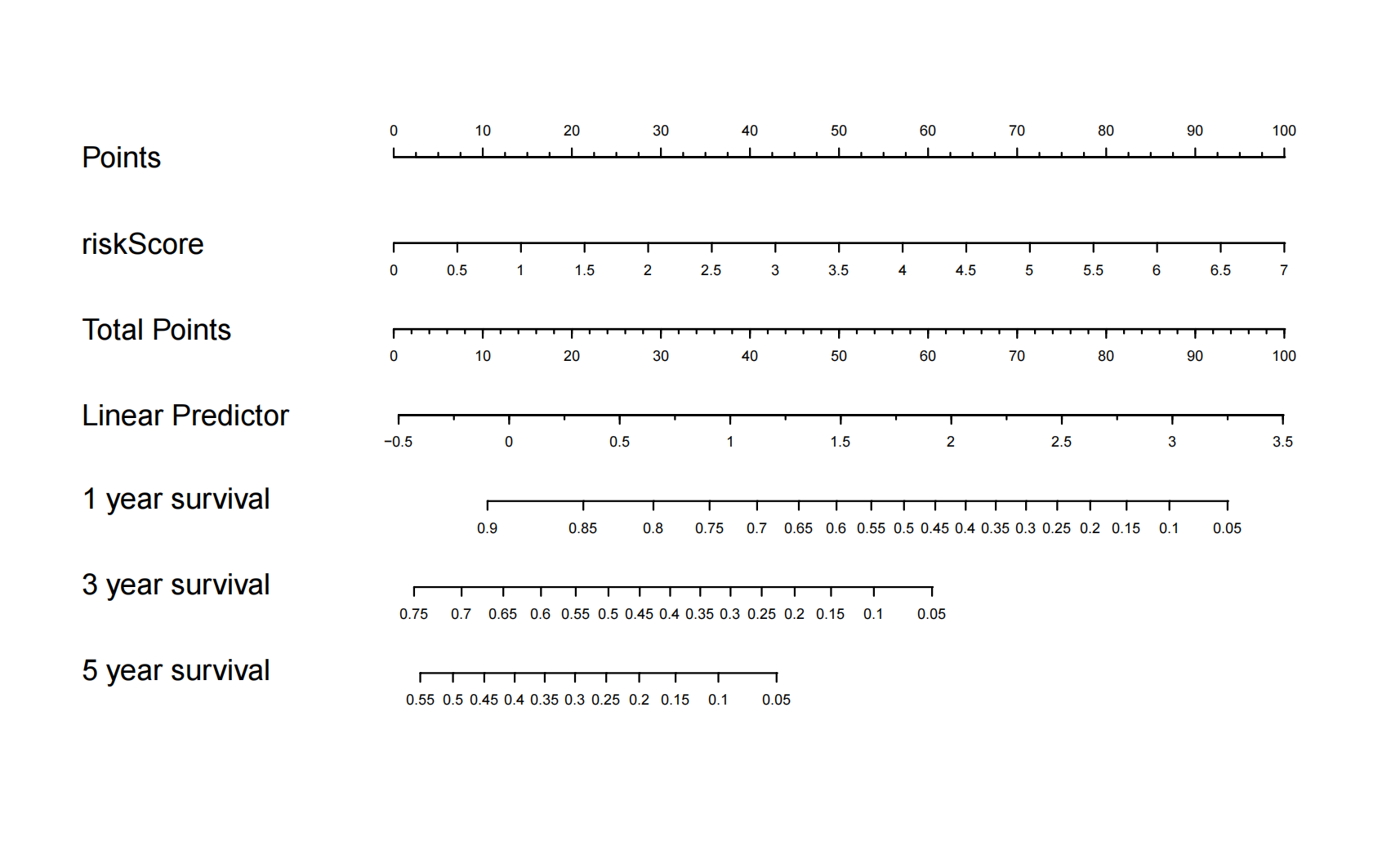


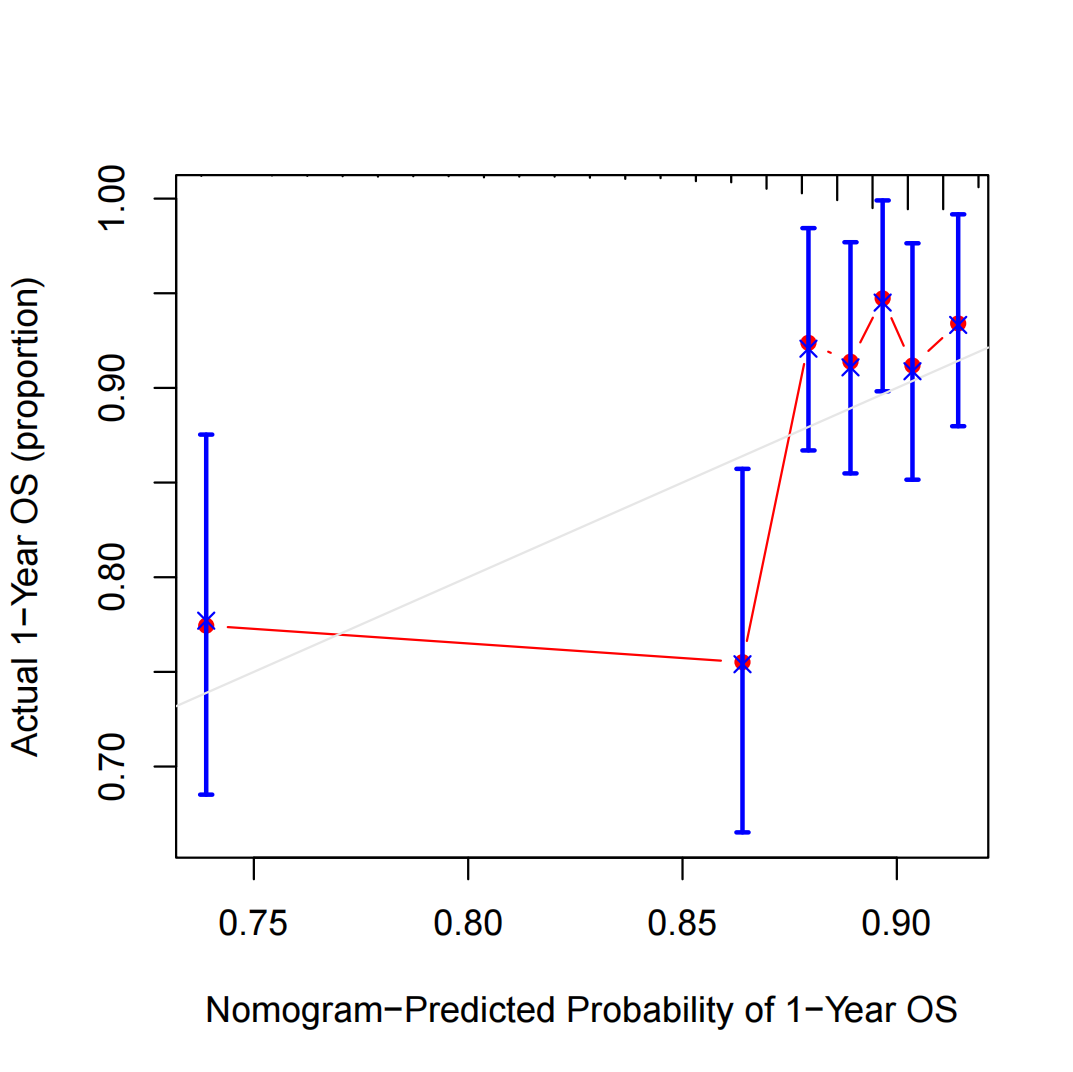
六、KM和ROC曲线分析两组的评估预后特征(picture:train\_KM.pdf、train\_ROC.pdf)

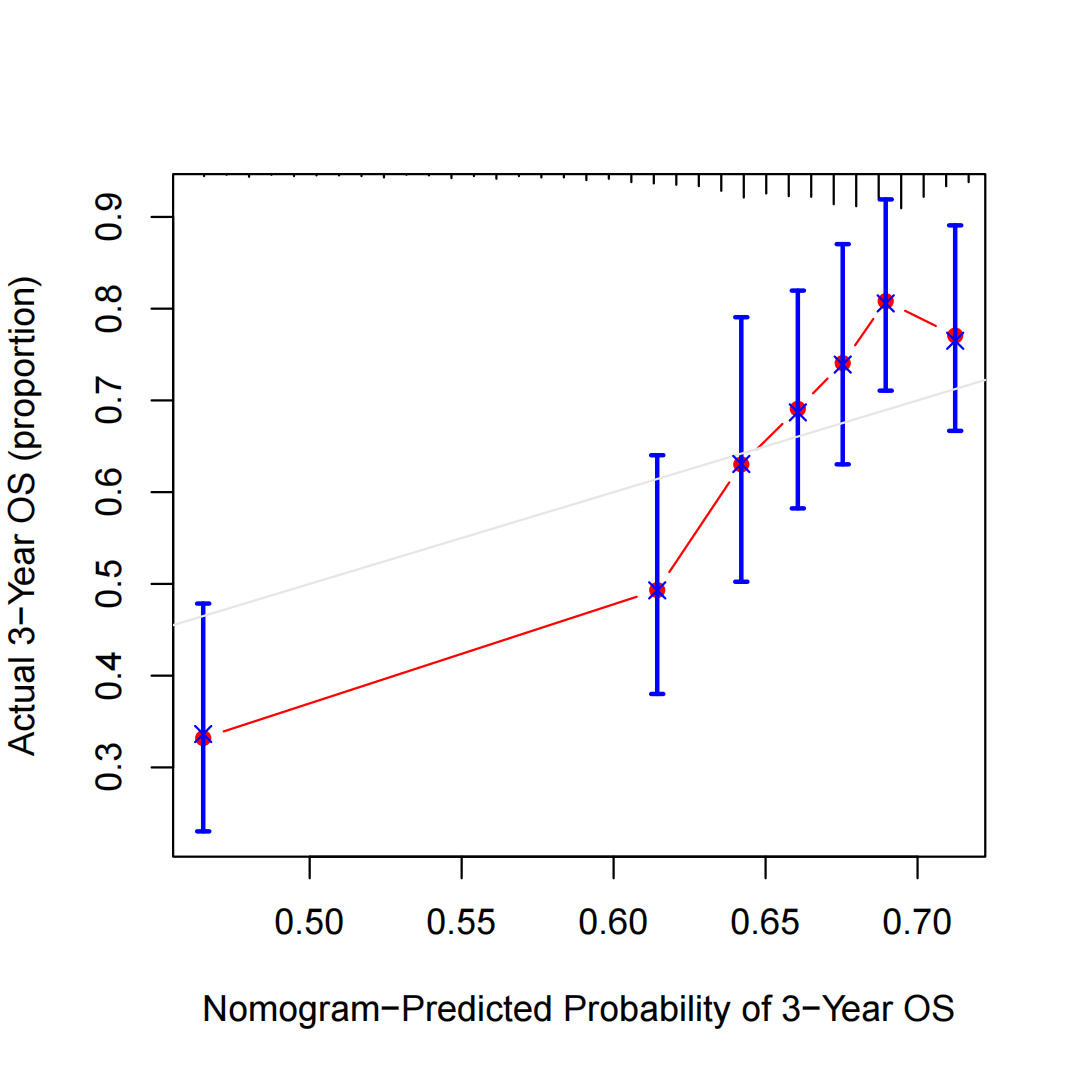


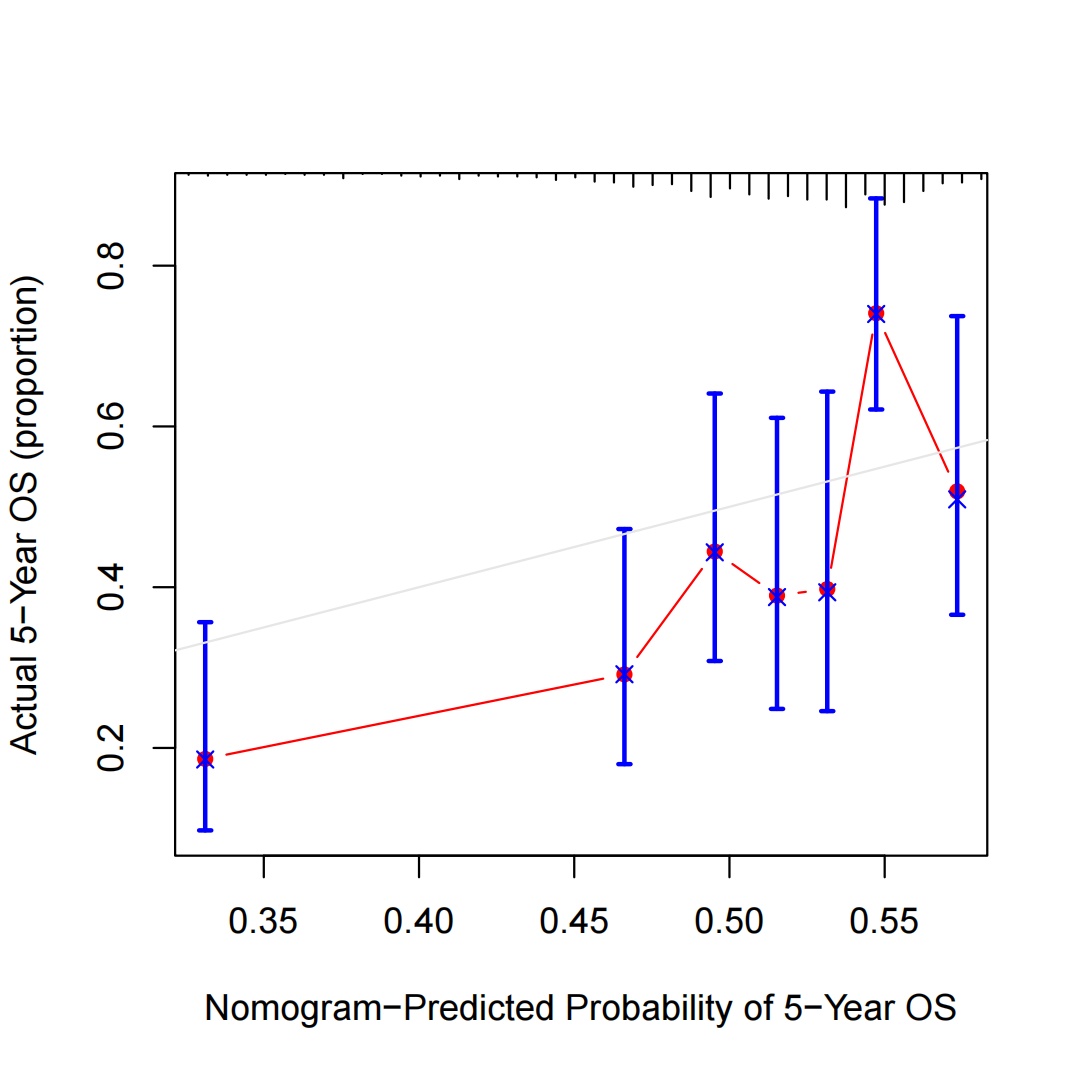


七、构建列线图评估患者1-5年的生存率(picture:liexiantu.pdf、jiaozhun1.pdf、jiaozhun2.pdf、jiaozhun3.pdf)









1. 验证集进行验证

Data（07.expr\_GSE68465.txt、08.survival\_GSE68465.txt）(picture:test\_risk\_all.pdf、test\_KM.pdf、test\_ROC.pdf)

