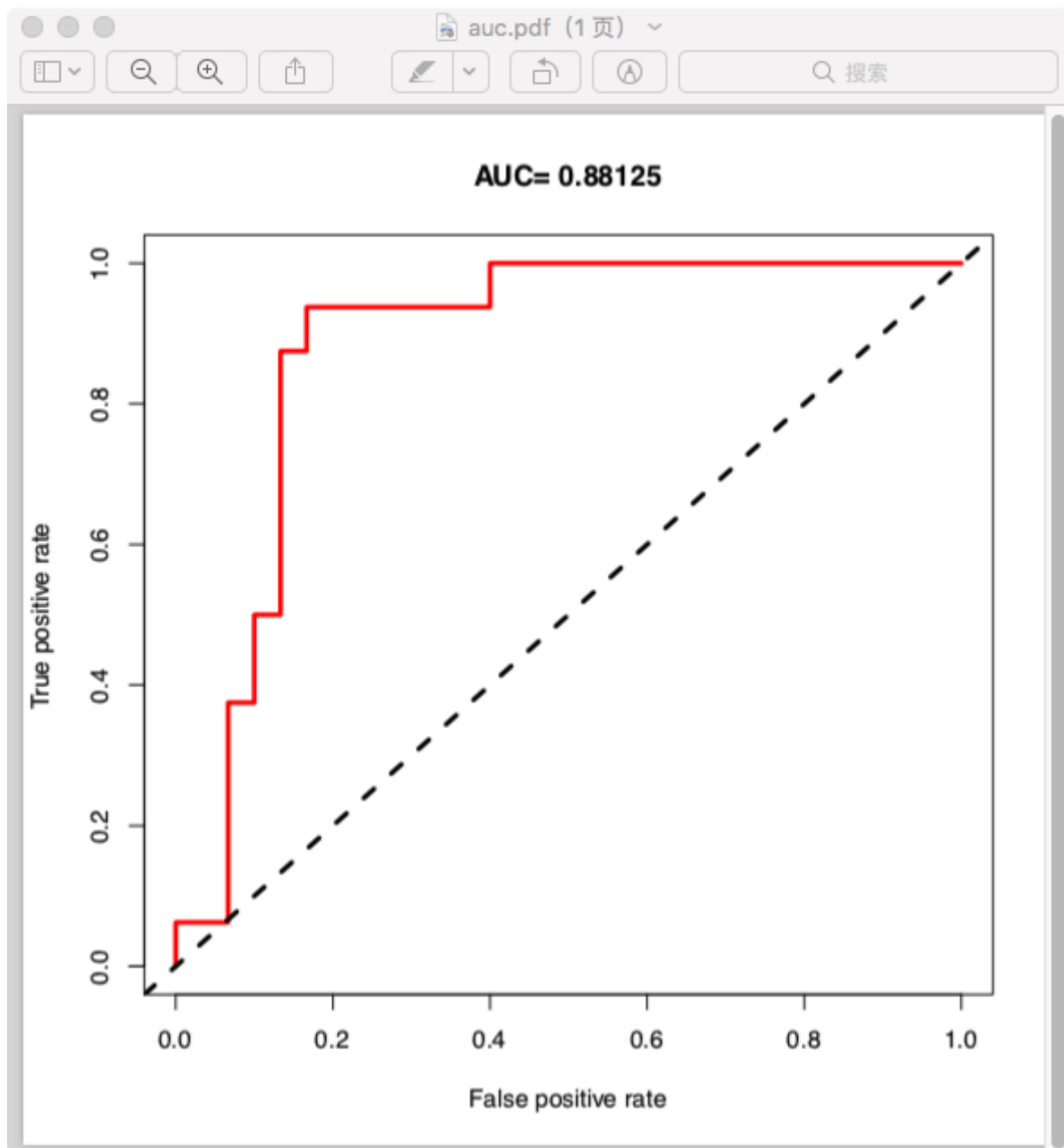
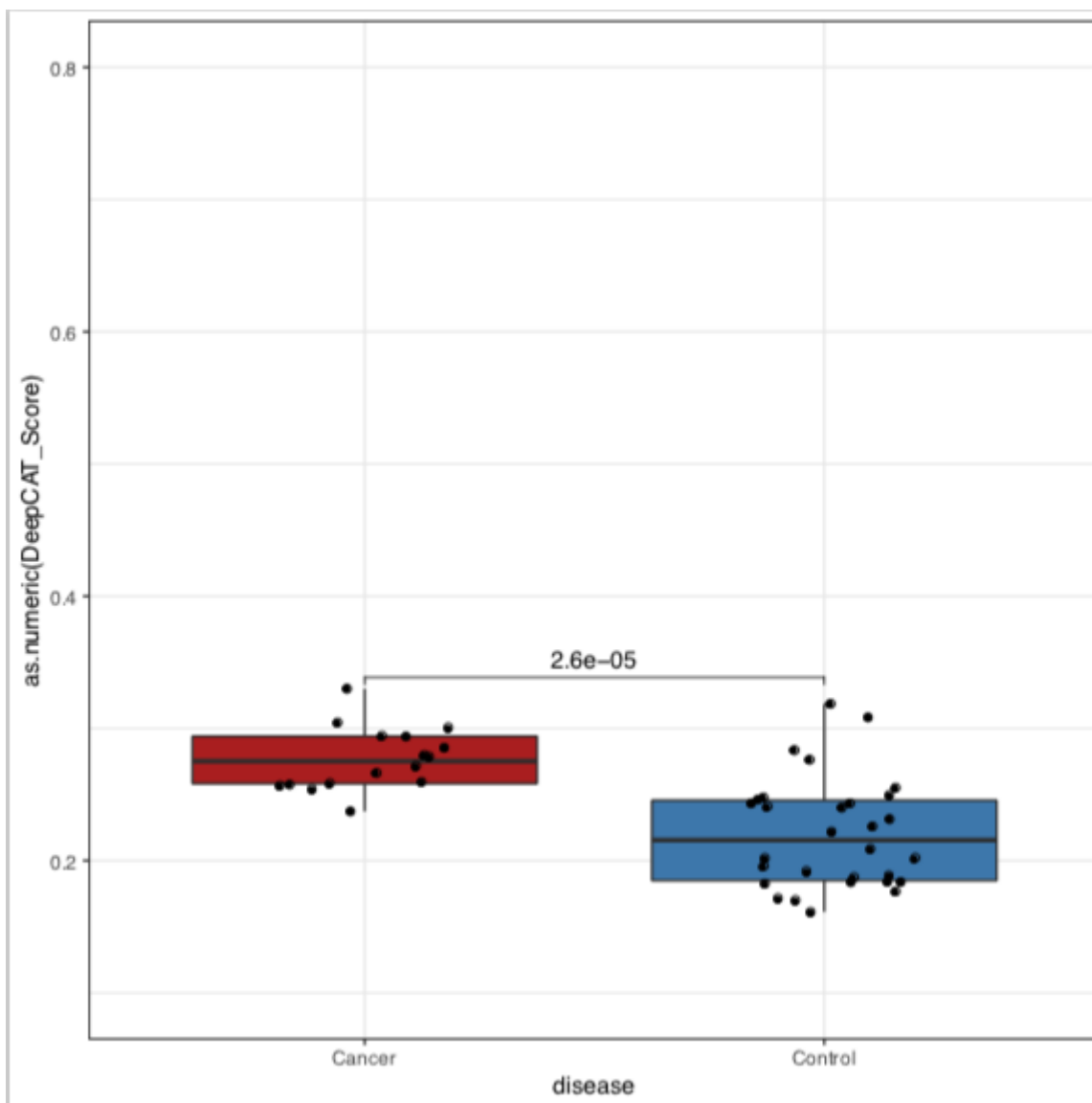


1) TCR-seq results

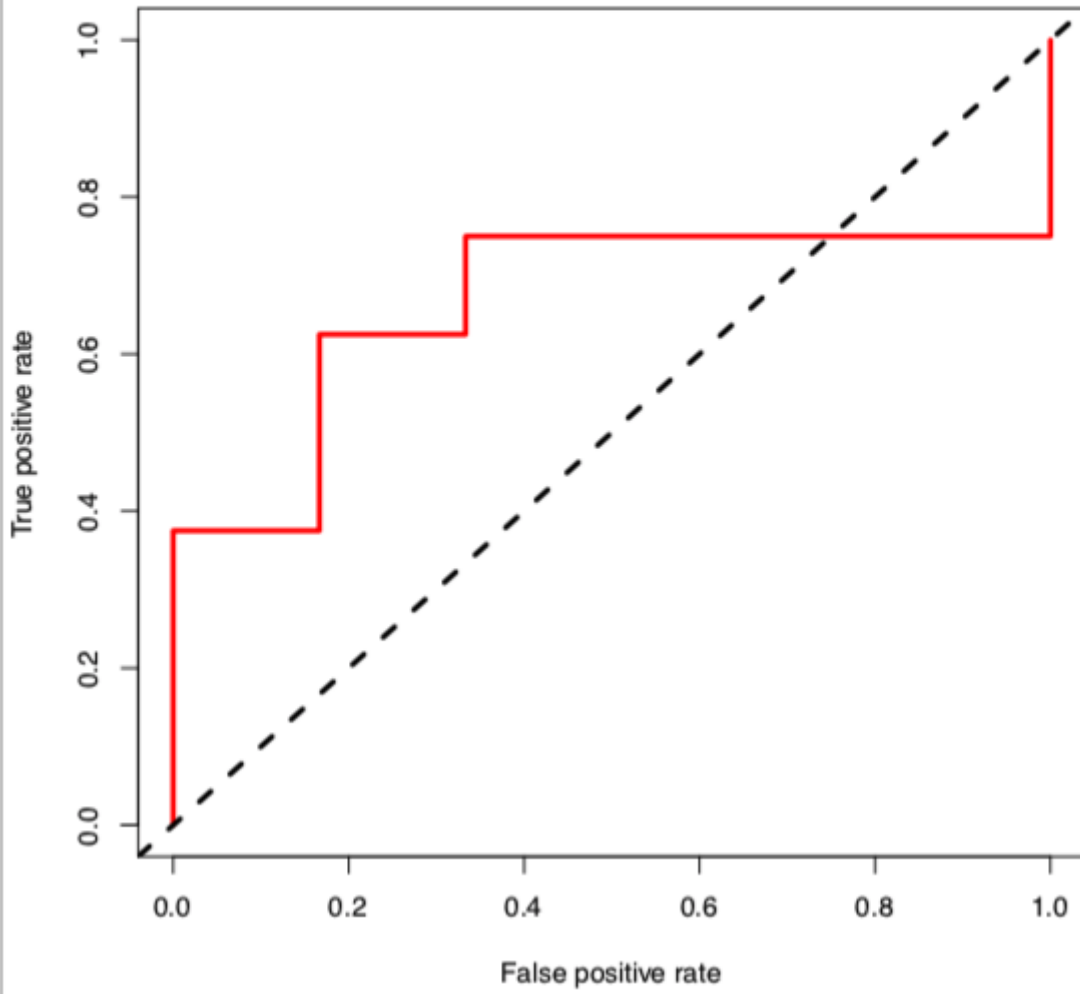


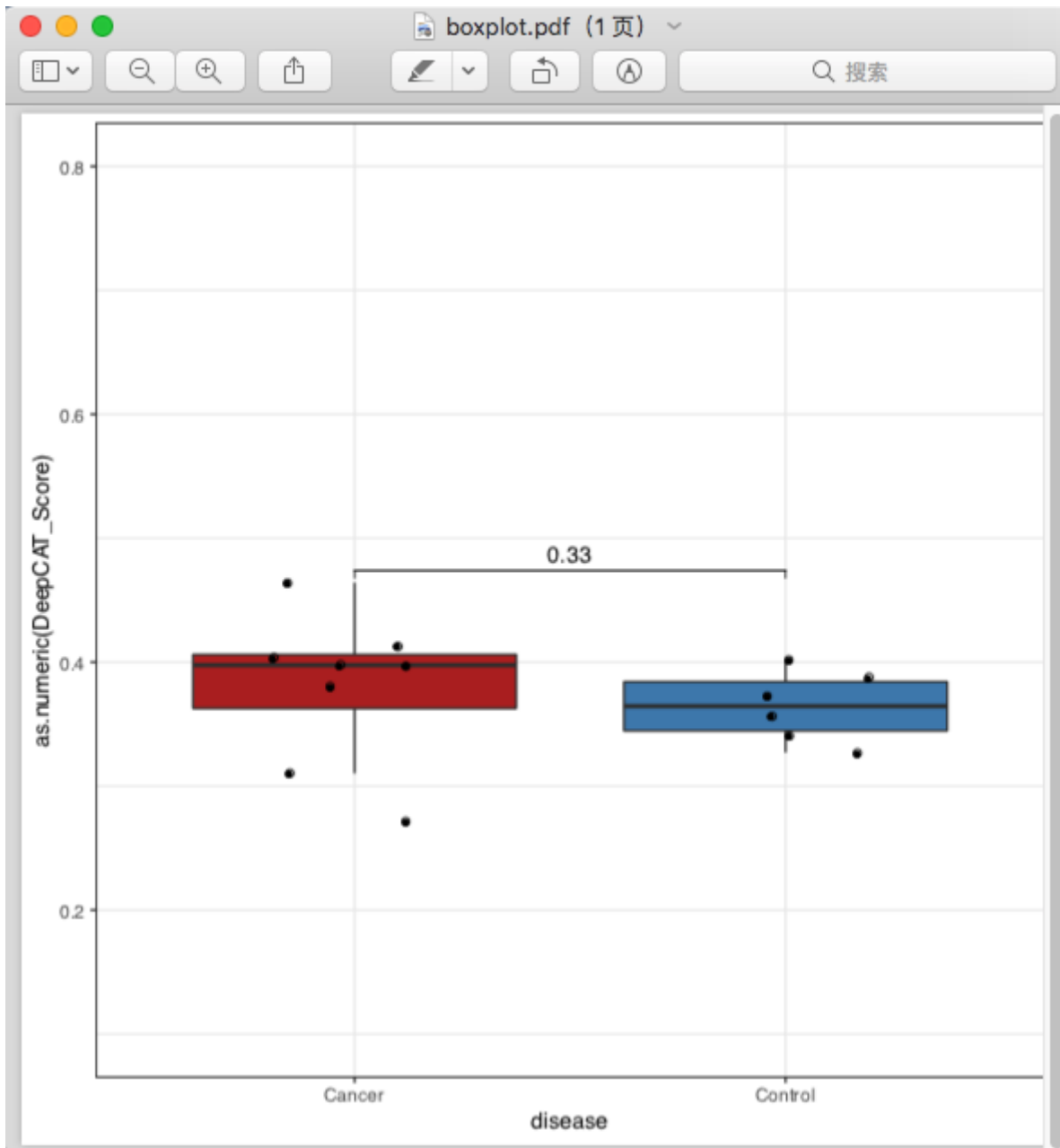


2) RNA-seq results



搜索

AUC= 0.666666666666667



**从TCR-seq与RNA-seq的AUC曲线以及boxplot比较可以看出，TCR-seq的结果明显好于RNA-seq。TCR-seq AUC曲线值更接近于1，disease与control之间的cancer score P-value <0.05 ,说明TCR-seq的方法disease与control之间存在明显差异。而RNA-seq的结果显示，disease与control的cancer score的P-value=0.33 >0.05 ,不能说明二者存在显著差异，综上所述，TCR-seq的测序的方法个在癌症诊断中表现的更好，通过cancer score 能将disease与control之间明显的区分出来。

3) (dis)advantages of both sequencig methods

3.a) (dis)advantages of RNA-seq

- RNA-seq 总文库中存在的TCR文库不是特别的多，因为RNA-seq是非靶向的测序。从刚刚的实际操作中可以看到，RNA-seq数据经过TCR calling 后得到的TCR序列量很少，甚至不足以进行cluster
- RNA-seq 会做RNA片段化处理，TCR最大的特点是VDJ重组，被随机打断后，再进行RNA拼接时，可能会造成将不同的TCR的VDJ拼接起来从而导致数据的不准确。
- RNA-seq样本大多取自肿瘤组织或癌旁组织。组织中的TCR含量本身没有PBMC中高，多样性也不如外周血。

- 目前RNA-seq产生的数据非常的多，通过TCR calling的工具，可以从中挖掘出许多的TCR sequence。且RNA-seq技术相较于TCR-seq技术更加成熟。

3.b) (dis)advantages of TCR-seq

- TCR-seq 根据不同的测序策略有DNA-based methods和RNA-based methods 两组基于不同材料的方式。

1. DNA-based methods 更适用于单TCR克隆的定量，但是很贵
2. RNA-based methods 对于检测TCR的丰富度以及基因表达的检测更加敏感，并且通过UMIs能够减少PCR偏差，并且增强对variants和rare mutation 的准确识别

- 根据TCR-seq的三种不同的测序策略，multiplex PCR、Rapid Amplification of cDNA Ends PCR(5'RACE-PCR)、Target enrichment。各有优劣

1. Multiplex PCR 由于引物设计参考了特定V等位基因，导致不能检测新的V等位基因，此外，多重PCR方法也存在扩增偏好性，影响TCR产物的相对丰度。兼容gDNA与RNA两种材料，使用最为普遍。
2. Target enrichment 定制TCR- $\alpha\beta$ 链目标区域序列互补的探针，与目标gDNA/cDNA杂交后，对目标进行捕获，使用较少的PCR过程，PCR偏好小，但是TCR区域本身高度多样性，这种方法并不主流。
3. 5'RACE+Switch Oligo+nested PCR 必须使用RNA为材料，逐渐成为群体TCR分析的金标准。cDNA 5'末端快速扩增技术。能够合成完整的TCR 5'cDNA，通常覆盖到完整的V基因，保留完整VDJ区域。由于扩增产物一致，也规避了扩增的偏好性。由于这种方法是RNA为起始材料，相对于其他技术对操作要求相对较高，并且全部流程繁杂，重复性可能会受到影响。

- 相比较于RNA-seq,TCR-seq对TCR sequence具有更强的靶向性，对于TCR序列进行富集，能较为完整的保留整个VDJ区域，准确性高，且能保证TCR sequence的多样性与可变性。此外TCR-seq技术可用于单克隆的TCR-seq

4) RNA-seq with raw fastq input

data from /Bioll/lulab_b/wangsiqi/exRNA/exRNA-panel/pico/02.rawdata_PBMC
or [here](#)

Running Steps

进入TCR目录，激活TCR的conda 环境

1. cp TCR_input data,divide the data into control and disease

```
# enter the TCR dir
cd TCR
# conda activate TCR env
conda activate TCR
# cp TCR_input data,divide the data into control and disease
cp -r /Bioll/lulab_b/wangsiqi/exRNA/exRNA-panel/pico/02.rawdata_PBMC ./
```

```

mv CRC-241* disease/
mv NC_PKU-* control/

# extract RN-seq ID of disease and control
cd /home/zhaoyizi/TCR/02.rawdata_PBMC/control
ls |cut -c1-19|uniq > /home/zhaoyizi/TCR/02.rawdata_PBMC/controlID.txt

cd /home/zhaoyizi/TCR/02.rawdata_PBMC/disease
ls |cut -c1-19|uniq > /home/zhaoyizi/TCR/02.rawdata_PBMC/diseaseID.txt

```

2. make working directory tree

```

# 确认路径在TCR目录下
cd /home/zhaoyizi/TCR
mkdir -p ./rawdata_PBMC_RNA-seq/{01_TCRcalling_output,02_filter_output,03_deepcat_output}
mkdir -p ./rawdata_PBMC_RNA-seq/{01_TCRcalling_output,02_filter_output}/{disease,control}

```

3. TCR calling

```

# 从双端测序PE的原始文件作为输入文件，从fastq(.gz)原始文件中得到TCR序列信息，使用 for loop循环
for idx in `cat 02.rawdata_PBMC/controlID.txt`;
do
./packages/TRUST4/run-trust4 -1 ./02.rawdata_PBMC/control/${idx}_1.fq.gz
-2 ./02.rawdata_PBMC/control/${idx}_2.fq.gz
-f ./packages/TRUST4/hg38_bcrtrcr.fa
--ref ./packages/TRUST4/human_IMGT+C.fa
-t 4

-o ./rawdata_PBMC_RNA-seq/01_TCRcalling_output/${idx};
done

```