

# 基于MiXCR及VDJtools的免疫组库分析流程

T/B细胞是适应性免疫系统的两大细胞群，细胞表面受体TCR/BCR存在一块区域叫互补决定区（Complementary Determining Region, CDR），包含CDR1、CDR2、CDR3，其中CDR3最高变，在抗原识别中起关键作用。免疫组库测序可通过多重PCR和高通量测序技术，分析编码CDR3区的DNA/RNA序列，获得机体的免疫特征。

## 基于MiXCR及VDJtools的免疫组库分析流程

### 1. MiXCR比对、组装、并输出CDR区域clones文件

- 1.1 Align比对reads
- 1.2 Assemble partial reads拼接部分reads
- 1.3 利用已有V和J基因延长TCR比对结果，基于germline序列补全覆盖度不完全的CDR3s
- 1.4 Assemble拼接clones
- 1.5 Exporting导出clones
- 1.6 在服务器上用单一命令分析

### 2. 使用VDJtools分析clones数据

- 2.1 工具配置
- 2.2 文件格式转换及准备metadata数据
- 2.3 基础信息统计
- 2.4 多样性评估
- 2.5 样本间重叠情况

本文主要介绍在mRNA测序的基础上得到的原始fastq文件如何进行免疫组库分析。

## 1. MiXCR比对、组装、并输出CDR区域clones文件

MiXCR可以用于提取RNA-Seq数据中TCR和BCR的CDR3组库，提取效率取决于样本T/B细胞的丰度和测序长度。推荐2x150bp或者2x100bp的双端测序方法。不过在双端2x50bp的RNA-Seq数据（比如肿瘤样本中），主要clonotypes信息也可以被提取。

### 1.1 Align比对reads

```
mixcr align -s hs -p rna-seq -OallowPartialAlignments=true data_R1.fastq.gz
data_R2.fastq.gz alignments.vdjca
```

所有mixcr align的参数都可以在这里使用（比如-s来指定物种）：

**-OallowPartialAlignments=true**选项保留部分比对结果用于后续的assemblePartial模块

### 1.2 Assemble partial reads拼接部分reads

```
mixcr assemblePartial alignments.vdjca alignmentsRescued.vdjca
```

为了获得包含CDR3全长序列的拼接reads，建议使用迭代mixcr的**assemblePartial**模块多次迭代拼接结果。根据经验，两次迭代后结果最优。

```
mixcr assemblePartial alignments.vdjca alignmentsRescued_1.vdjca
mixcr assemblePartial alignmentsRescued_1.vdjca alignmentsRescued_2.vdjca
```

### 1.3 利用已有V和J基因延长TCR比对结果，基于germline序列补全覆盖度不完全的CDR3s

```
mixcr extendAlignments alignmentsRescued_2.vdjca
alignmentsRescued_2_extended.vdjca
```

### 1.4 Assemble拼接clones

```
mixcr assemble alignmentsRescued_2_extended.vdjca clones.clns
```

所有mixcr assemble的参数都可以在这里使用：

- 对于低质量数据，建议降低输入质量阈值（e.g. -ObadQualityThreshold=15）
- 为了克隆丰度与错误校正算法相结合，增加下面的选项：  
**-OaddReadsCountOnClustering=true**

### 1.5 Exporting导出clones

```
mixcr exportClones -c TRA -o -t clones.clns clones.txt
```

可以指定导出感兴趣的免疫受体链（-c TRA 或者 -c TRB等），也可以去除包含out-of-frame（选项-o）和stop codon的突变体（选项-t）。

**但为方便起见，使用单一analyze shotgun命令完成分析。**

### 1.6 在服务器上用单一命令分析

```
sudo docker run --rm -v /mnt:/mnt milaboratory/mixcr mixcr analyze shotgun --
species hs --starting-material rna --only-productive --report
/mnt/vol2_ws/test/wbzhao/TCR/${i}/${i}.report
/mnt/vol1_root/rendong/NGS_raw_data_tmp/RD206_20210402/R21031599-RD206-20210403-
1-${i}_combined_R2.fastq.gz
/mnt/vol1_root/rendong/NGS_raw_data_tmp/RD206_20210402/R21031599-RD206-20210403-
1-${i}_combined_R1.fastq.gz /mnt/vol2_ws/test/wbzhao/TCR/${i}/${i}
```

其中\${i}为样本编号，注意修改样本编号、fastq文件路径及结果存储路径；不要在登陆节点使用！请配置进pipeline或在node3启动容器使用。

--species 定义物种 --report 生成报告文件位置及名称

```
for i in `cat id.list`;do sudo docker run --rm -v /mnt:/mnt milaboratory/mixcr
mixcr analyze shotgun --species hs --starting-material rna --only-productive --
report /mnt/vol2_ws/test/wbzhao/TCR/${i}/${i}.report
/mnt/vol1_root/rendong/NGS_raw_data_tmp/RD206_20210402/R21031599-RD206-20210403-
1-${i}_combined_R2.fastq.gz
/mnt/vol1_root/rendong/NGS_raw_data_tmp/RD206_20210402/R21031599-RD206-20210403-
1-${i}_combined_R1.fastq.gz /mnt/vol2_ws/test/wbzhao/TCR/${i}/${i};done
```

id.list为样本编号列表文件，注意修改fastq文件路径及结果存储路径。

## 2.使用VDJtools分析clones数据

### 2.1 工具配置

可以直接通过<https://github.com/mikessh/vdjtools/releases/tag/1.2.1>下载，zip包中包含jar文件可直接调用。VDJtools会调用R去进行绘图，所以需要先安装所需的R包。VDJtools提供了命令可以一键安装相关的R包：`java -jar vdjtools.jar Rinstall`

node3中已拉取VDJtools镜像，直接调用即可。

### 2.2 文件格式转换及准备metadata数据

VDJtools不能直接处理其他软件产生的clones文件，需要一步转换，`-S`参数指明clones文件由什么软件产生：

```
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools convert -S mixcr
/mnt/vol2_ws/test/wbzhao/TCR/${i}/${i}.clonotypes.ALL.txt
/mnt/vol2_ws/test/wbzhao/TCR/
```

其中*i*为样本编号，会在指定的结果路径中生成输入文件同名文件。

```
for i in `cat id.list`;do sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools
vdjtools convert -S mixcr /mnt/vol2_ws/test/wbzhao/TCR/${i}/${i}.clonotypes.ALL.txt
/mnt/vol2_ws/test/wbzhao/TCR/;done
```

批量转换文件格式，id.list为样本编号列表文件。

另外，后续的分析都需要样本的meta信息，需要准备一个Metadata.txt，格式如图：

#file.name	sample.id	group	name	fraction	TPS	CPS	CD8.SP16	PD-L1	CD8.infiltration	TP
/mnt/vol2_ws/test/wbzhao/TCR/BP2101130069RL01S01.clonotypes.ALL.txt	BP2101130069RL01S01	non-relapse	B0461933	85%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2101130070RL01S01.clonotypes.ALL.txt	BP2101130070RL01S01	non-relapse	B0417431	75%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060024RL01S01.clonotypes.ALL.txt	BP2103060024RL01S01	relapse	B0500640	40%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060026RL01S01.clonotypes.ALL.txt	BP2103060026RL01S01	relapse	B0554253	85%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060028RL01S01.clonotypes.ALL.txt	BP2103060028RL01S01	relapse	B0557756	95%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060030RL01S01.clonotypes.ALL.txt	BP2103060030RL01S01	relapse	B0539827	60%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060036RL01S01.clonotypes.ALL.txt	BP2103060036RL01S01	relapse	B0403521	90%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060037RL01S01.clonotypes.ALL.txt	BP2103060037RL01S01	relapse	B558970	30%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060038RL01S01.clonotypes.ALL.txt	BP2103060038RL01S01	relapse	B633932	85%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060039RL01S01.clonotypes.ALL.txt	BP2103060039RL01S01	relapse	B525634	50%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060040RL01S01.clonotypes.ALL.txt	BP2103060040RL01S01	relapse	B525634	50%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060043RL01S01.clonotypes.ALL.txt	BP2103060043RL01S01	non-relapse	B0459023	95%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060044RL01S01.clonotypes.ALL.txt	BP2103060044RL01S01	non-relapse	B0548499	90%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060045RL01S01.clonotypes.ALL.txt	BP2103060045RL01S01	non-relapse	B0460873	35%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060046RL01S01.clonotypes.ALL.txt	BP2103060046RL01S01	non-relapse	B0413540	70%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060048RL01S01.clonotypes.ALL.txt	BP2103060048RL01S01	non-relapse	B0544225	70%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060050RL01S01.clonotypes.ALL.txt	BP2103060050RL01S01	non-relapse	B0468253	20%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060054RL01S01.clonotypes.ALL.txt	BP2103060054RL01S01	non-relapse	B0434618	50%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060057RL01S01.clonotypes.ALL.txt	BP2103060057RL01S01	non-relapse	B0416910	35%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060059RL01S01.clonotypes.ALL.txt	BP2103060059RL01S01	non-relapse	B0525162	40%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060062RL01S01.clonotypes.ALL.txt	BP2103060062RL01S01	non-relapse	B0533998	80%	<1%	<1%	<1%	<1%	<1%	0%
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060065RL01S01.clonotypes.ALL.txt	BP2103060065RL01S01	non-relapse	B0541341	75%	<1%	<1%	<1%	<1%	<1%	0%

前两列指定文件路径（#file.name）、对应样本名（sample.id），后面的列可以标注样本的分组信息等，在后续的分析中会用到。

## 2.3 基础信息统计

```
#计算统计样本的基础信息，如：read counts，平均克隆型大小，非功能性克隆型的数量等
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools CalcBasicStats -m
/mnt/vol2_ws/test/wbzhao/TCR/metadata.txt /mnt/vol2_ws/test/wbzhao/TCR/result/0
# 计算样本中V/J段中的每个片段的频率
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools CalcSegmentUsage
-m /mnt/vol2_ws/test/wbzhao/TCR/metadata.txt -p -f group
/mnt/vol2_ws/test/wbzhao/TCR/result/2
# 计算spectratype，即CDR3核苷酸长度的读计数直方图
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools CalcSpectratype -m
/mnt/vol2_ws/test/wbzhao/TCR/metadata.txt /mnt/vol2_ws/test/wbzhao/TCR/result/1
# 绘制样本频谱类型，及给定样本中前N个克隆类型的CDR3长度
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools
PlotFancySpectratype
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060062RL01S01.clonotypes.ALL.txt
/mnt/vol2_ws/test/wbzhao/TCR/result/BP2103060062RL01S01
# 绘制圈式的V-J图，显示各种V-J连接的频率
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools PlotFancyVJUsage
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060062RL01S01.clonotypes.ALL.txt
/mnt/vol2_ws/test/wbzhao/TCR/result/BP2103060062RL01S01
# 绘制V区域在CDR3序列长度上的分布
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools PlotSpectratypev
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060062RL01S01.clonotypes.ALL.txt
/mnt/vol2_ws/test/wbzhao/TCR/result/BP2103060062RL01S01
```

-p 打开画图，-f 指定分组factor（metadata中的列），如果是数值型参数，增加-n

## 2.4 多样性评估

```
#画一个三层的甜甜圈图，形象化展示克隆类型层级关系
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools PlotQuantileStats
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060062RL01S01.clonotypes.ALL.txt
/mnt/vol2_ws/test/wbzhao/TCR/result/BP2103060062RL01S01
#绘制样本的稀疏曲线，即样本多样性与样本大小的相关性，并按照-f指定列着色
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools RarefactionPlot -m
/mnt/vol2_ws/test/wbzhao/TCR/metadata.txt -f group -l name
/mnt/vol2_ws/test/wbzhao/TCR/result/8
#计算样本多样性
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools CalcDiversityStats
-m /mnt/vol2_ws/test/wbzhao/TCR/metadata.txt
/mnt/vol2_ws/test/wbzhao/TCR/result/7
```

-l 样本标签，指定metadata中的一列

## 2.5 样本间重叠情况

#对一对样本进行克隆型共享的综合分析

```
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools overlapPair -p  
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060062RL01S01.clonotypes.ALL.txt  
/mnt/vol2_ws/test/wbzhao/TCR/BP2103060038RL01S01.clonotypes.ALL.txt  
/mnt/vol2_ws/test/wbzhao/TCR/result/9
```

#对样本列表执行overlap分析，并计算相似度。至少要提供3个样品

```
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools  
CalcPairwiseDistances -m /mnt/vol2_ws/test/wbzhao/TCR/metadata.txt  
/mnt/vol2_ws/test/wbzhao/TCR/result/10
```

#进行样本聚类

```
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools clusterSamples -p  
-f group -l name /mnt/vol2_ws/test/wbzhao/TCR/result/10  
/mnt/vol2_ws/test/wbzhao/TCR/result/10.group
```

#进行克隆类型跟踪

```
sudo docker run --rm -v /mnt:/mnt yyasumizu/vdjtools vdjtools TrackClonotypes -m  
/mnt/vol2_ws/test/wbzhao/TCR/metadata.txt -f group -p  
/mnt/vol2_ws/test/wbzhao/TCR/result/11
```

---

除以上功能外，MiXCR和VDJtools还有其他更多丰富的功能，请参阅官方文档及以下链接。

参考链接：

1. <https://www.jianshu.com/p/5ee941d726f6>
2. <https://www.bilibili.com/read/cv9461265/>
3. <https://github.com/mikessh/vdjtools-examples>
4. <https://vdjtools-doc.readthedocs.io/en/master/overlap.html#trackclonotypes>