LOHHLA

官方网站：<https://bitbucket.org/mcgranahanlab/lohhla>

**需要另外安装的软件**：

1、Novocraft软件只需要添加到路径中：

export PATH=/data3/pengxb/script\_exon/soft/novocraft:$PATH

2、jellyfish软件安装：

下载<https://github.com/gmarcais/Jellyfish/tree/v2.2.7>

安装<http://blog.sina.com.cn/s/blog_4b91a9e50101off7.html>

按步骤安装完成后，添加到环境变量中：

export PATH=/data3/pengxb/script\_exon/soft/jellyfish/bin:$PATH

3、R包：R版本为3.4.1

export PATH=/data3/pengxb/script\_exon/soft/R-3.4.1/lib64/R/bin:$PATH

seqinr包安装：install.packages("seqinr")

Biostrings包安装：source("https://bioconductor.org/biocLite.R")

biocLite("Biostrings")

beeswarm包安装：install.packages("beeswarm")

zoo包安装：install.packages("zoo")

Rsamtools包安装：source("https://bioconductor.org/biocLite.R")

biocLite("Rsamtools")

optparse包安装：（用于获取参数）

install.packages("optparse")

测试包是否安装完全：library("")

4、python包：

beeswarm包安装：

export PATH=/data3/pengxb/script\_exon/soft/Python-2.7.13/bin:$PATH

pip install beeswarm

5、java版本：1.8.0\_73

export PATH=/data3/dna/soft/jdk1.8.0\_73/bin:$PATH

6、picard的使用：

路径 --gatkDir /data3/xulm/soft/picard-tools-2.2.4/

注意picard软件里的用法：picard.jar SamToFastq，R脚本里用法不同，已改正为现有用法

**测试LOHHLA软件流程：**

由于整个R包是安装在3.4.1版本下的，注意导入R包的路径：

export R\_LIBS="/data3/pengxb/script\_exon/soft/R-3.4.1/lib64/R/library"

jellyfish软件的生成文件自动末尾带“\_0”，后续的文件需要同样带有”\_0”，否则这一步无法运行

1、测试路径：/data3/pengxb/temporary/LOHHLA

2、配置文件：

--hlaPath /data3/pengxb/temporary/LOHHLA/example-file/hlas

hlas：病患HLA的序列名

--HLAfastaLoc /data3/pengxb/temporary/LOHHLA/data/example.patient.hlaFasta.fa

example.patient.hlaFasta.fa：病患HLA的序列

--HLAexonLoc /data3/pengxb/temporary/LOHHLA/data/hla.dat

hla.dat：绘制HLA外显子边界

--CopyNumLoc /data3/pengxb/temporary/LOHHLA/example-file/solutions.txt

solutions.txt：肿瘤样本倍性文件，需要另外软件计算。

其他的参考文件默认

**结果文件说明：**

生成2个excel文件，表格说明如下：

example.10.DNA.HLAlossPrediction\_CI.xls

|  |  |
| --- | --- |
| 表头 | 说明 |
| region | 项目名称或tumor样本名 |
| HLA\_A\_type1 | 等位基因1的类型 |
| HLA\_A\_type2 | 等位基因2的类型 |
| HLAtype1Log2MedianCoverage | 等位基因1的LogR覆盖率中位值 |
| HLAtype2Log2MedianCoverage | 等位基因2的LogR覆盖率中位值 |
| HLAtype1Log2MedianCoverageAtSites | 跨越等位基因1的LogR覆盖率中位值，限于不匹配的位点 |
| HLAtype2Log2MedianCoverageAtSites | 跨越等位基因2的LogR覆盖率中位值，限于不匹配的位点 |
| HLA\_type1copyNum\_withoutBAF | 等位基因1的估计拷贝数，不使用BAF |
| HLA\_type1copyNum\_withoutBAF\_lower | 等位基因1的估计拷贝数低于95%的置信区间 |
| HLA\_type1copyNum\_withoutBAF\_upper | 等位基因1的估计拷贝数高于95%的置信区间 |
| HLA\_type1copyNum\_withBAF | 使用BAF的等位基因1的估计拷贝数，没有分组位点 |
| HLA\_type1copyNum\_withBAF\_lower | 使用BAF估计的等位基因1的拷贝数低于95%的置信区间，没有分组位点 |
| HLA\_type1copyNum\_withBAF\_upper | 使用BAF估计的等位基因1的拷贝数高于95%的置信区间，没有分组位点 |
| HLA\_type2copyNum\_withoutBAF | 未使用BAF的等位基因2的估计拷贝数 |
| HLA\_type2copyNum\_withoutBAF\_lower | 等位基因2的估计拷贝数低于95%的置信区间 |
| HLA\_type2copyNum\_withoutBAF\_upper | 等位基因2的估计拷贝数高于95%的置信区间 |
| HLA\_type2copyNum\_withBAF | 使用BAF的等位基因2的估计拷贝数，没有分组位点 |
| HLA\_type2copyNum\_withBAF\_lower | 使用BAF估计的等位基因2的拷贝数低于95%的置信区间，没有分组位点 |
| HLA\_type2copyNum\_withBAF\_upper | 使用BAF估计的等位基因2的拷贝数高于95%的置信区间，没有分组位点 |
| HLA\_type1copyNum\_withoutBAFBin | 使用分组但没有BAF估计的等位基因1的拷贝数 |
| HLA\_type1copyNum\_withoutBAFBin\_lower | 使用分组但没有BAF的等位基因1的估计拷贝数低于95%的置信区间 |
| HLA\_type1copyNum\_withoutBAFBin\_upper | 使用分组但没有BAF的等位基因1的估计拷贝数高于95%的置信区间 |
| HLA\_type1copyNum\_withBAFBin | 使用分组和BAF估计的等位基因1的拷贝数 |
| HLA\_type1copyNum\_withBAFBin\_lower | 低于95%的置信区间 |
| HLA\_type1copyNum\_withBAFBin\_upper | 高于95%的置信区间 |
| HLA\_type2copyNum\_withoutBAFBin | 使用分组但没有BAF估计的等位基因2的拷贝数 |
| HLA\_type2copyNum\_withoutBAFBin\_lower | 低于95%的置信区间 |
| HLA\_type2copyNum\_withoutBAFBin\_upper | 高于95%的置信区间 |
| HLA\_type2copyNum\_withBAFBin | 使用分组和BAF估计的等位基因2的拷贝数 |
| HLA\_type2copyNum\_withBAFBin\_lower | 低于95%的置信区间 |
| HLA\_type2copyNum\_withBAFBin\_upper | 高于95%的置信区间 |
| PVal | 与等位基因1和等位基因2之间logR差异相关的PVal - p值（配对t检验） |
| UnPairedPval | 与等位基因1和等位基因2之间的logR差异有关的p值（未配对t检验） |
| PVal\_unique | 与等位基因1和等位基因2之间的logR差异有关的p值，确保每个只读贡献一次（配对t检验） |
| UnPairedPval\_unique | 与等位基因1和等位基因2之间的logR差异有关的p值，确保每个只读贡献一次（不成对t检验） |
| LossAllele | 以较低频率存在的HLA等位基因（可能易于丢失） |
| KeptAllele | 以较高频率出现的HLA等位基因（可能不会丢失） |
| numMisMatchSitesCov | 具有足够覆盖范围的不匹配网站的数量 |
| propSupportiveSites | 与错失或等位基因不平衡相一致的错配位点的比例 |