



诺禾致源数据结果说明文档

--参考序列比对分析结果

有效测序数据通过 BWA(version 0.7.8-r455)比对到人类参考基因组 (UCSC hg19), 得到 BAM 格式的最初的比对结果。

在最初的比对结果 (BAM 文件) 的基础上做如下处理:

- (1) 用 SAMtools(Li H et al.)对比对结果进行排序;
- (2) 用 Picard 标记重复 reads (mark duplicate reads)。经过以上处理, 得到 BAM 格式的最终比对结果。后续变异检测基于这个比对结果。

如果一个或一对 read(s) 在基因上可以有多个比对位置, BWA 的处理策略是从中选择一个最好的, 如果有两个或以上最好的比对位置, 则从中随机选择一个。这种多重比对 (multiple hit) 的处理对 SNP、indel 以及 CNV 等的检测有重要影响。通常检测 SNP 或 INDEL 的时候要使用高质量的比对 (alignment), 即比对质量值大于 0 或更高。

本文件夹下包含每个样本(样本名)的比对结果数据结果文件:

- 1.后缀 bam 文件 比对结果文件
 - 2.后缀.bam.bai 文件 对 bam 文件构建索引, 用于对 bam 文件的快速处理
- =====

结果文件说明

- 1.bam 文件是 sam 文件的压缩格式
解读请参考附件 SAMv1.pdf
- 2.后缀.bam.bai 文件是由命令 samtools index *.bam 得到