

SubOrder_Pacbio 测序验证报告

实验流程

样品经前期准备处理后，使用SMRTbell Express Template Prep Kit 2.0试剂盒构建SMRTbell文库。之后通过Qubit3.1和Qseq100生物分析仪检测文库质量。构建好的SMRTbell文库结合SMRTbell Express Template Prep Kit 2.0试剂盒的上机引物和聚合酶，以自由扩散的方式加入到PacBio平台测序芯片中，进行测序。

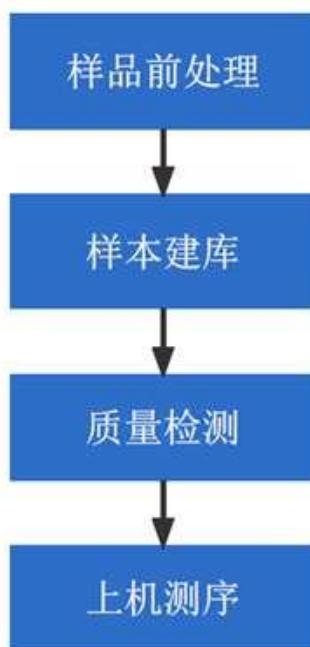


Fig1. 实验流程

生物信息分析流程

- (1) CCS 纠正：使用 ccs (4.2.0)^[1]软件对测序结果中同一个零模波导孔 (ZMW) 产生的 subread 进行矫正得到高质量的 CCS 序列；
- (2) 数据比对：通过 minimap2 (version 2.15-r905)^[2]软件将高质量的 CCS 序列比对到参考序列上；
- (3) 验证分析：比对结果使用 samtools(Version: 1.9)^[3]软件进行质控后，统计比对率、测序数据对参考序列的覆盖情况、变异信息，并进行可视化展示；

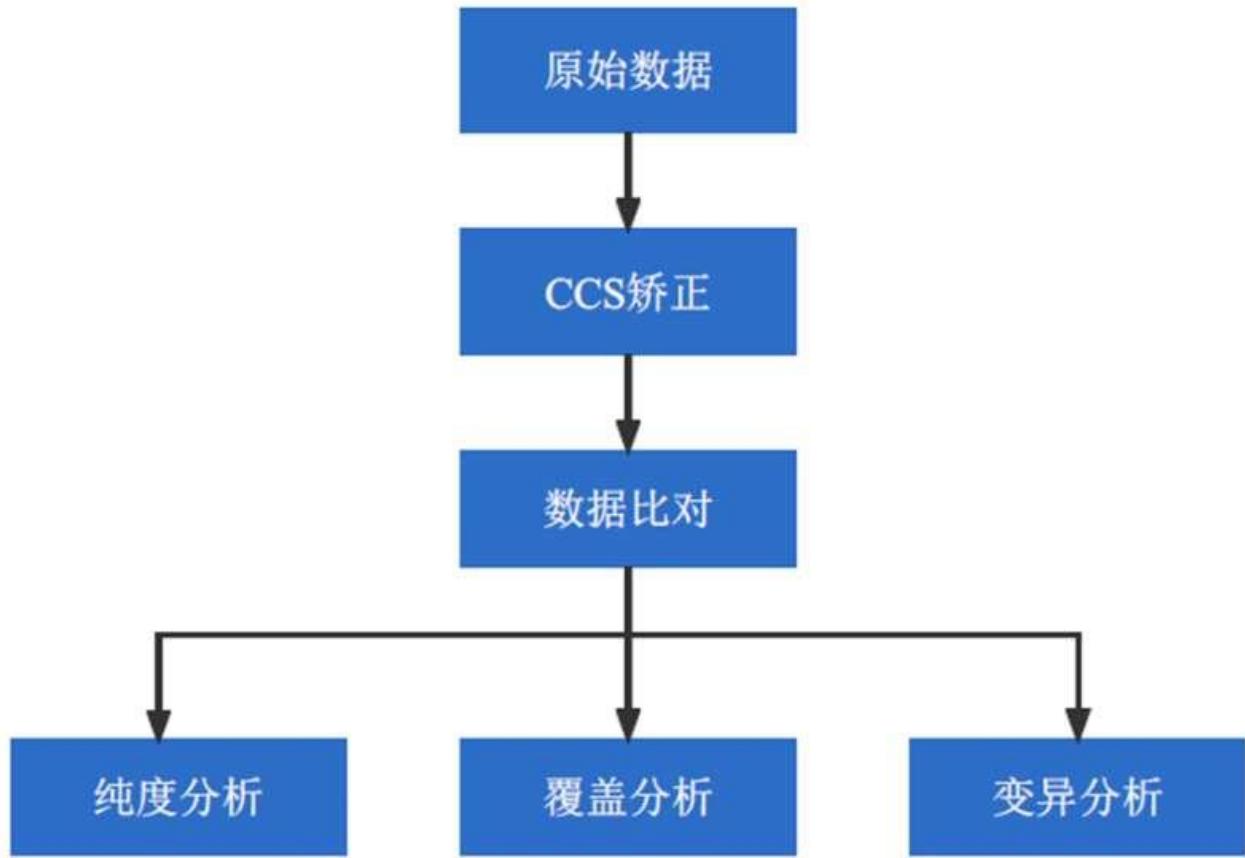


Fig2.生物信息分析流程

分析结果

根据比对结果统计测序数据对参考序列的覆盖情况和变异信息，统计结果见文件“SubOrder.snp.indel.xlsx”。

列名解释：

- (1) Id: 参考序列名称；
- (2) Pos: 参考序列位置；
- (3) RefBase: 参考序列在该位置的碱基；
- (4) Coverage: 测序数据对该位置的覆盖深度；
- (5) A: 支持该位置为 A 碱基的 Reads 数目；
- (6) C: 支持该位置为 C 碱基的 Reads 数目；
- (7) G: 支持该位置为 G 碱基的 Reads 数目；
- (8) T: 支持该位置为 T 碱基的 Reads 数目；
- (9) MutRate(%): 碱基突变频率，该位置支持非参考序列碱基的 Reads 数目/覆盖深度*100；
- (10) Mut: 检出的具体突变碱基和比例；
- (11) InsCov: 支持该位置存在插入碱基的 Reads 数目；
- (12) InsRate(%): 插入频率，该位置支持存在插入碱基的 Reads 数目/覆盖深度*100；
- (13) Ins: 检出的具体插入碱基和比例；
- (14) DelCov: 该位置支持碱基缺失的 Reads 数目；
- (15) DelRate(%): 缺失频率，该位置支持碱基缺失的 Reads 数目/覆盖深度*100；
- (16) Del: 检出的具体缺失碱基和比例；

根据统计结果可视化测序数据对参考序列的覆盖情况及变异信息，结果见 Fig3，横坐标为参考序列的位置信息(单位bp)，纵坐标为对应的统计值。PacBio在多个连续相同碱基处会存在较高误差，此类区域使用最大等位基因频率对应的基因型作为候选结果，当候选结果与参考序列不一致时，建议使用Sanger测序进行验证，并整合结果。将检测到的阳性变异信息标注在图片的左上方，标注包含变异位置、类型和比例，例如：10362 G->A 99.77%，表示参考序列在10,362bp的位置其碱基类型为G，而99.77%的测序数据支持样

品在该位置的碱基类型为A。插入缺失类型的变异仅标注测序数据支持样品在对应位置的碱基类型。

Fig3.测序结果对参考序列的覆盖

选取各个位点最大等位基因频率对应的基因型依次串联组成样品测序结果序列，SEQ格式的结果序列见文件“SubOrder_cons_seq.seq”。

参考文献

- [1] Gorrieri R , Versari C . CCS: A Calculus of Communicating Systems[J]. Springer International Publishing, 2015.
- [2] Heng. Minimap2: pairwise alignment for nucleotide sequences[J]. Bioinformatics, 2018.
- [3] Danecek P , Bonfield J K , Liddle J , et al. Twelve years of SAMtools and BCFtools[J]. GigaScience, 2021, 10(2).