结核分枝杆菌耐药检测分析流程可靠性测试方案

1. **项目简介**

结核分枝杆菌耐药检测技术是利用二代测序技术对结核菌进行测序，并对结核分枝杆菌样本中多个基因、位点的耐药变异的检测，现需要对流程中结核分枝杆菌鉴定的结果准确性进行评估。

1. **评估方法**

使用模拟的样本，模拟出结核分枝杆菌（mycobacterium tuberculosis，MTB）与非结核分枝杆菌（Nontuberculosis mycobacteria, NTM）的数据，进行多次重复测试，最终得到测试结果。

1. **测试数据**
   1. 在NCBI上搜索并下载MTB与NTM全基因组序列。
   2. 将下载序列进行预处理，将不符合条件的序列进行剔除，筛选条件如下：非全长序列，非DNA序列，非MTB或NTM物种等。
2. **测试方法**
   1. 将预处理后的序列随机打断成fastq格式。需要模拟两种平台的数据：BGISEQ-500数据为PE100，每条reads的错误率随机设置为1-10%，BGISEQ-50平台数据为SE50，每条reads错误率随机设置为 1-10%。每个fastq数据量为1G。
   2. 打断后的样本分为纯MTB样本（100份），纯NTM样本（100份）和MTB与NTM混合样本三大类，其中混合样本又分为1:3:7比例的错误样本：TB：NTM（100份）；1:7:3比例的错误样本：MTB：NTM（100份）；5:5比例的MTB：NTM（100份）。
   3. 得到两个平台下的模拟样本各500份（共1000份），通过流程中比对模块进行数据分析，鉴定是否为MTB样本（比对上的uniq reads数比例与真实比例的误差为±5%则为有效），得到实际鉴定的结果。
3. **评估方法**

评估两个平台下本次鉴定结果的敏感性(Sn)、特异性(Sp)、准确性（Acc）和马修斯相关系数(MCC),统计方式见下方计算公式。

