



华大基因纯测序项目报告

2022/7/18



@2022 BGI All Rights Reserved

目 录

1 项目信息	3
2 数据统计	3
3 数据质控	3
4 帮助文档	3

1 项目信息

项目编号: F22FTSEUHT0562_MUScwfR
样本数量: 1

2 数据统计

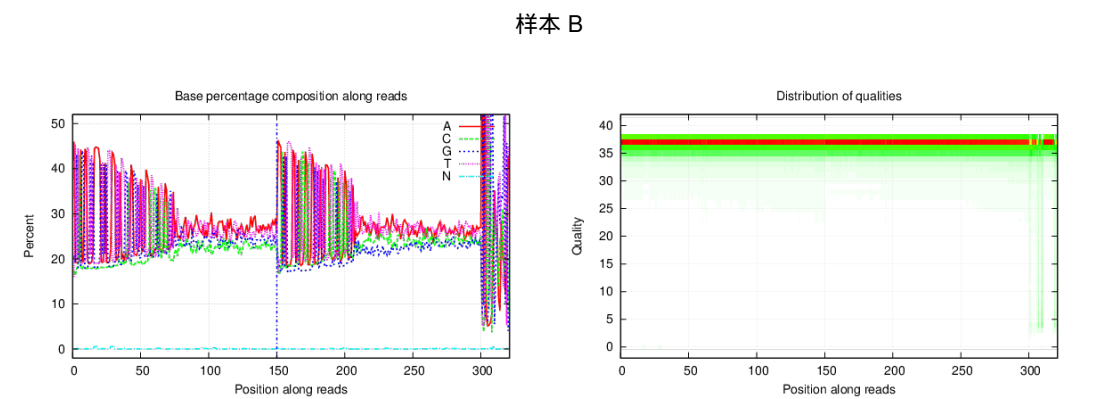
从测序仪产生的原始读数包含接头、未知或低质量的碱基。原始数据的统计数据如下所示。

样本	读长	Q20(%)	Q30(%)	GC含量(%)	序列数量	碱基数量
B	150;170	97.91;96.81	94.22;92.83	46.14;46.57	923,794,956	147,807,192,960

- 格式说明:
- 1. 样本: 样本名
 - 2. 读长: 测序序列长度
 - 3. Q20 (%): 质量值大于20的碱基在序列中的占比
 - 4. Q30 (%): 质量值大于30的碱基在序列中的占比
 - 4. GC含量(%): C和G碱基在序列中的占比
 - 5. 序列数量: 原始序列数量
 - 6. 碱基数量: 原始碱基数量

3 数据质控

数据过滤中碱基比例和质量值分布如下图所示（如果一个样本有多个lane，则只显示其中一个）。左图是样本碱基百分比分布，右图是样本序列的质量分布。



4 帮助文档

原始图像数据转换为序列数据，定义为原始数据并保存为 FASTQ 文件。FASTQ 文件中的每个条目由 4 行组成：

- 1. 包含有关测序运行和集群信息的序列标识符。此行的确切内容因使用的 BCL 到 FASTQ 转换软件而异；
- 2. 测序序列(包含碱基: A, C, T, G and N)；
- 3. 分隔符；
- 4. 测序质量值，通常是 Phred +33 编码,使用 ASCII 字符表示质量分数。

以下是 FASTQ 文件中单条序列的示例：

```
@V300029029L1C001R0010000210/1
GCGACCCCAGGTCAGTCGGGACTACCCGCTGAAGTCGGAGGCCAAGCGGT
+
FFFFFFFFFFFDFEFFFFEF0FFFFEFFFFEFFFFEFCGFFFF
```

DNBseq测序仪测序错误率与测序质量值的关系如下式所示。具体来说，如果测序错误率记为“E”，DNBseq测序碱基质量值记为“sQ”，关系如下：

$sQ = -10\log_{10} E$

测序错误率	测序碱基质量值	Phred +33 编码字符
5%	13	.
1%	20	5
0.1%	30	?