

=====文件=====

|-- All\_sample\_filter.stat.xls      数据过滤统计表

过滤掉原始序列中含有接头的序列和低质量序列等一系列不符合要求的序列，得到符合要求的高质量序列。

- |-- Samples：样本信息单样品名称；
- |-- BMK-ID：百迈客编号；
- |-- Raw\_reads：测序原始数据；
- |-- Low\_quality：质量值低于 30 的碱基所占比例超过 20%的 reads；
- |-- Containing'N'reads：含至少 10%的未知碱基 N 的 reads；
- |-- Length<18：去掉接头后,长度小于 18 个核苷酸的 reads 数；
- |-- Length>30：去掉接头后,长度小于 18 个核苷酸的 reads 数；
- |-- Clean\_reads：质量值大于或等于 30 的碱基的 Reads 数；
- |-- Q30(%)：碱基质量值大于 30 的比例。

|-- AllSample\_GC\_Q.stat.xls      样本的测序数据质量评估统计表

- |-- SampleID：客户的样本编号；
- |-- BMK-ID：百迈客样品分析编号,为公司内部编号,与客户样品的编号对应；
- |-- ReadSum：Clean Data 中 pair-end Reads 总数；
- |-- BaseSum：Clean Data 中的总碱基数；
- |-- GC(%)：Clean Data 中的 GC 含量,即 Clean Data 中 G 和 C 两种碱基占总碱基的百分比；
- |-- N(%)：Clean Data 中含 N 的碱基占总碱基的百分比；
- |-- Q30(%)：Clean Data 中质量值大于或等于 30 的碱基占总碱基的百分比。

=====文件夹=====

|-- PNG

|-- \*.error.png      样本测序错误率分布图，png 格式

横坐标是碱基的位置,纵坐标是错误率；此图反映每个测序反应的测序质量，通常测序序列 5'端前几个碱基的错误率相对较高。