

genome survey SOP

- [1. 流程简介](#)
- [2. 预先准备](#)
- [3. 运行准备](#)
- [4. 运行流程](#)

1. 流程简介

使用一个高深度测序文库进行基因组调查。

2. 预先准备

如果第一次使用该流程，请git clone

```
git clone git@git.majorbio.com:long.huang/dna.git
```

如果已有repo，请使用git pull。务必每次运行前运行git pull，以确保代码为最新代码。

nextflow.config文件根据172环境开发，如果搬迁到236服务器，请自行调整对应参数。如果不会调整，后续询问小汤。

3. 运行准备

1. 将双端reads下载至工作目录
2. project.info文件，与云平台的project.txt格式相同：制表符分隔6列（合同号，分析号，老师姓名，空，空，样品数）

4. 运行流程

运行代码示例：

```
nextflow run -bg -resume GenomeSurvey/genome_survey.nf --r1  
6539c6ad0a78ca7e3a5431d5--L1EHG1901449_TKS_10.R1.raw.fastq.gz --r2  
6539c6ad0a78ca7e3a5431d5--L1EHG1901449_TKS_10.R2.raw.fastq.gz --name TKS_10 --  
project project.info --report
```

命令参数说明：

```
Usage:  
The typical command for running the pipeline is as follows:  
nextflow run genome_survey.nf --outdir gs_result --name sample --r1 r1.fq.gz  
--r2 r2.fq.gz --report  
  
--reportdb <file> 报告模板文件夹
```

```

--r1      <file>  r1.fq.gz, required
--r2      <file>  r2.fq.gz, required
--outdir  <dir>   输出文件夹
--name     <str>   样品名, required
--a1      <str1>  r1接头, 华大测序时可以为""
--a2      <str>  r2接头, 华大测序时可以为""
--kmer     <int>  kmer长度, 奇数, 默认21, 基因组很大时可以酌情调大kmer
--readlen  <int>  测序读长
--report   <bool> 是否生成报告
--project  <file>  project.info, required
--help     <bool> 打开帮助

```

最终结果一般位于输出文件夹下的publish文件夹，输出文件夹组成如下：

```

disease_result/
├─ qc          # 质控结果文件
├─ publish     # 生成报告和可以交付的结果目录
└─ gs         # genome survey结果文件

```