|  |  |  |
| --- | --- | --- |
| 文件编号：SOP-XX-110 | **爱汝康illumina平台**  **数据分析操作规范** | 拟制人： |
| 生效日期： 2016.X.X | 审核人： |
| 版 本：第一版 | 批准人： |

1. 目的：

根据实验数据，在集群下分析，生成分析结果，结果用于生成报告。

1. 范围：

测序下机数据，用于检益康产品的数据分析和结果生成。

1. 职责：

信息部。

1. 操作步骤：
   1. 公共密钥

如果已经做过设置，则忽略公共密钥设置教程。

* + 1. 注意事项：
       1. 修改密码

首次登录需在10.20.4.11上进行，如果修改时有交互，回车即可。

* + - 1. 投放任务

登录到节点上，目前可以在此节点上跑程序。

* + - 1. 保存数据

申请开通工作目录后，今后的工作在工作目录进行，/home下的个人目录中仅允许放小的程序，不允许放大数据，空间有限。

* + 1. 秘钥生成
       1. 生成密钥 (可忽略)

## ssh-keygen -t rsa

输入命令后一直按回车

* + - 1. 切换到家目录的.ssh目录

## cd ~/.ssh

## cat id\_rsa.pub > authorized\_keys

具体操作如下图内容：

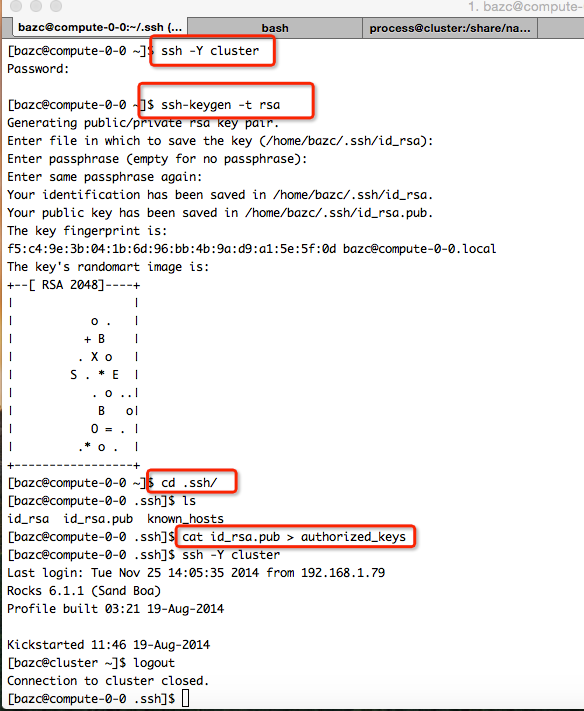


图1

公共密钥配置成功才能在集群上做下面的操作。

* 1. 操作方法：
     1. 准备软件包

需要软件包Reseq ，Reseq的内容如下图1

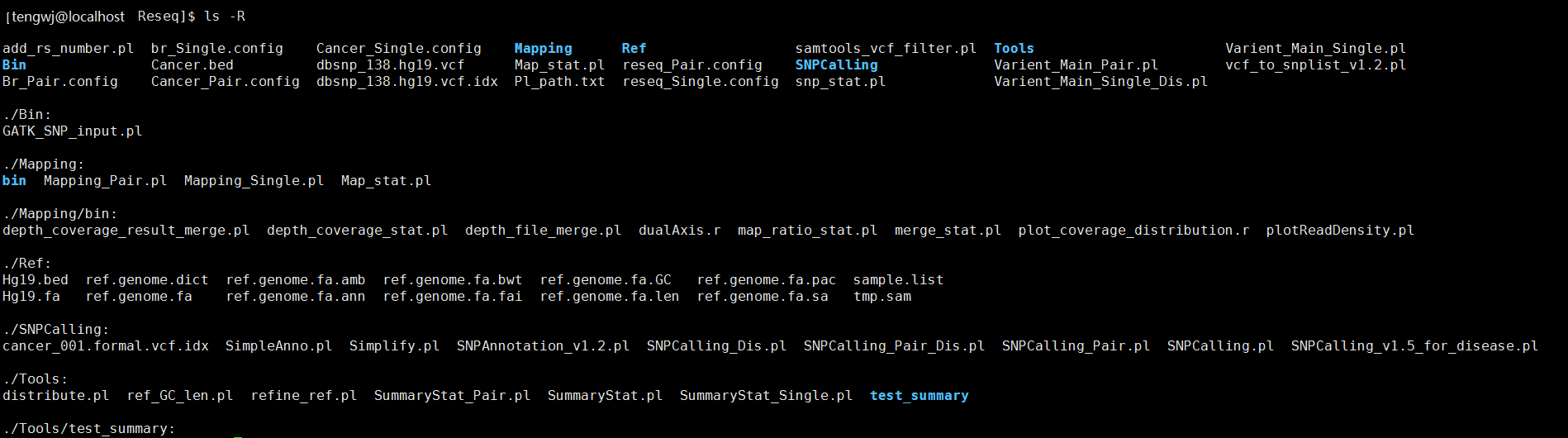
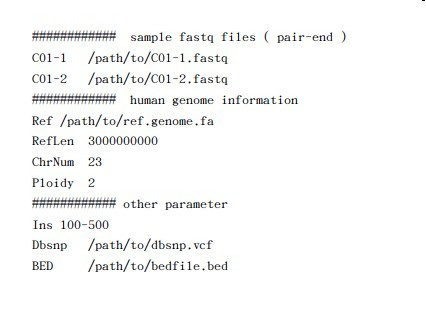


图2

* + 1. 准备文件

需要准备配置文件（Project .config）和相关的一些参考文件（如SusGD\_P2\_capture\_targets.bed）。其中（.bed）文件是为区域捕获所设计的一个文件，爱汝康应用的是区域捕获测序及分析。



说明：

C01-1: 双端测序fastq reads1文件

C01-2: 双端测序fastq reads2文件

Ref： 参考基因组fa序列文件路径

RefLen： 参考基因组碱基数

ChrNum： 物种的染色体数量

Ploidy： 样本的倍体数

Ins： 插入片段大小

Dbsnp： Dbsnp数据库文件

BED： 基因区间文件

图3

需要准备报告配置文件（addconf2.txt），需要填写姓名（Name），年龄（Age），性别（Gender），样本类型（SampleType），样本编号（SampleNo），收样日期（Date1），报告日期（Date2）；如下图所示。

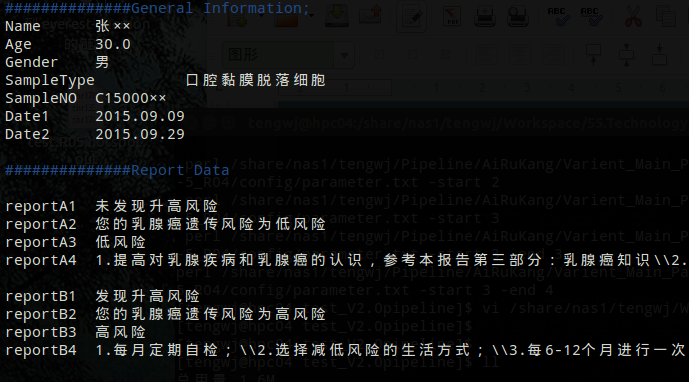


图4

* + 1. 运行和投放任务

perl /share/nas1/tengwj/Pipeline/AiRuKang/Varient\_Main\_Pair\_V2.0.pl -c Project.config -p test -od ./ -addconf2 addconf2.txt

图5

-c 配置文件

-p 样本前缀

-od 输出目录

-addconf2 报告配置文件

* 1. 数据解读判定标准
     1. 样本质控标准：

|  |  |  |
| --- | --- | --- |
| **参数** | **阈值** | **图示** |
| **Mean Depth** | >300X |  |
| **Uniformity** | >80% |  |
| **Target base coverage at 20x** | >90% |  |

表1

* + 1. 位点质控标准：

|  |  |  |
| --- | --- | --- |
| **参数成ccd** | **阈值** | **图示** |
| **MaqQ** | >50 |  |
| **Coverage** | >=30 |  |
| **Frequcency** | >=30 |  |

表2

* + 1. 标准解读

如果符合上面的标准，判定为此分析结果可用；如果不可用，主观查找错误原因并改正后重新分析。

* 1. 分析过程中建议标准（表二）

|  |  |
| --- | --- |
| **建议操作** | **符合条件** |
| **重新采集材料或放弃检测** | 1.三次以上达不到样本质控标准的 |
| **重建库** | 1.两次或两次以上达不到样本质控标准的 |
| **重上机** | 1. 关键位点达不到位点质控标准的 |

表3

1. 相关文件：

**无。**

1. 相关记录：

无。