颜色代码：步骤，命令，文件/路径名称，非必选步骤，软件名

**分别对肿瘤样本和正常组织进行DNA测序，可用于获得肿瘤样本的体细胞突变图谱。**

**步骤1. 准备样本名文件**

以2个小鼠为例（小鼠1有两个肿瘤灶点，小鼠2有一个肿瘤灶点。此例分别对这个三个肿瘤以及两只小鼠的尾巴进行了DNA测序）。在 ~/DNA-SeqDemo中，有以下测序文件：

Mouse1\_Tumor1\_R1.fq.gz, Mouse1\_Tumour1\_R2.fq.gz （小鼠1的1号肿瘤）

Mouse1\_Tumor2\_R1.fq.gz, Mouse1\_Tumour2\_R2.fq.gz （小鼠2的2号肿瘤）

Mouse1\_Tail\_R1.fq.gz, Mouse1\_Tail.fq.gz （小鼠1的尾巴）

Mouse2\_Tumor1\_R1.fq.gz, Mouse2\_Tumour1\_R2.fq.gz （小鼠2的肿瘤）

Mouse2\_Tail\_R1.fq.gz, Mouse2\_Tail.fq.gz （小鼠2的尾巴）

创建sampleSheet.txt，其中列出文件名为（见同一文件夹下sampleSheet.txt范本）：

Mouse1\_Tumor1

Mouse1\_Tumor2

Mouse1\_Tail

Mouse2\_Tumor1

Mouse2\_Tail

创建sampleSheet2.txt，其中列出肿瘤文件名为（见同一文件夹下sampleSheet.txt范本）：

Mouse1\_Tumor1

Mouse1\_Tumor2

Mouse2\_Tumour1

**步骤3. 基因组拼接**

输入命令：perl mapping.pl

输出例如：~/DNAseqDemo/bam/Mouse\_Tumor1.sam

**步骤4. : 将sam文件转换为bam文件**

输入命令：perl sam2bam.pl

输出例如：~/DNAseqDemo/bam/Mouse1\_Tumor1.bam

**步骤5. GATK Toolkit 4.0格式处理**

按照GATK Toolkit 4.0的要求，将bam文件依次进行加头，排序，去除复制体以及索引，依次执行以下命令：perl addhead.pl; perl sort.pl; perl markdup.pl

输出例如：~/DNAseqDemo/bam/Mouse1\_Tumor1\_addRG\_sorted\_markdup.bam

~/DNAseqDemo/bam/Mouse1\_Tumor1\_addRG\_sorted\_markdup.bai

**步骤6. 碱基质量校正**

通过GATK Toolkit 4.0，对拼接后的碱基质量进行校正，依次执行以下命令：perl recalibrate1.pl; perl recalibrate2.pl

输出例如：~/DNAseqDemo/bam/Mouse1\_Tumor1\_recal.bam

**步骤6. Mutect2计算体突变**

输入命令：：perl mutect.pl

输出例如：~/DNAseqDemo/bam/ Mouse1\_Tumor1.vcf

# -n 10 指调用的集群节点数

# -nct 80 是指线程数

#-I:tumor cancer\_recal.bam 是指输入的肿瘤拼接文件

#- I:normal tissue\_recal.bam 是指输入的正常组织拼接文件

# -o mutation.vcf' 是指将输出