**新基因预测**

基于所选参考基因组序列，使用StringTie软件对Mapped Reads进行拼接，并与原有的基因组注释信息进行比较，寻找原来未被注释的转录区，发掘该物种的新转录本和新基因，从而补充和完善原有的基因组注释信息。过滤掉编码的肽链过短（少于50个氨基酸残基）或只包含单个外显子的序列。

**=====================================文件=====================================**

**|-- Human.newGene\_final.filtered.gff 新基因预测gff文件**

|-- #Seq\_ID：染色体号

|-- Source：注释信息的来源,Cufflinks 软件

|-- Type：注释特征(Feature)类 型

|-- Start/End：特征序列的起止位置

|-- Score：得分,数字,注释信息可能性的说明,“.”表示缺失值

|-- Strand：特征序列所在的正负 链

|-- Phase：仅对注释类型为CDS有效，表示起始编码的位置,有效值为0、1、2，“.”表示缺失值；

|-- Attributes：以多个键值对组成的注释信息描述。

**|-- Human.newGene.longest\_transcript.fa 新基因最长转录本fa文件**

**====================================文件夹=====================================**

**|-- BMK\_1\_NewGene\_Anno 新基因注释结果**