

IGV 使用简介

(RNA-Seq 基因组比对)

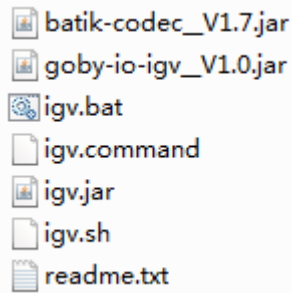
目 录

一、软件准备	2
1. IGV 工具.....	2
2. JAVA 运行库	2
二、启动 IGV	2
三、导入参考数据	3
1. 载入本地 FASTA 格式文件.....	3
2. 从远程服务端载入.....	3
四、导入 BAM 文件	4
五、查看比对信息	4
① 参考数据管理.....	5
② 染色体选择.....	5
③ 具体位点或元件查找.....	5
④ 染色体区域不同尺度的放大或缩小.....	6
⑤ 染色体区域位置.....	6
⑥ 染色体位置刻度尺.....	6
⑦ BAM 文件比对信息显示区.....	6
⑧ 参考数据显示区.....	8
六、其他说明.....	10

一、软件准备

1. IGV 工具

我们已从官网下载了 IGV 工具包，适用于 Windows、Linux 和 Mac OS 三种操作系统。请解压本目录下的 IGV_2.3.32.zip，解压后文件如下：



2. JAVA 运行库

使用 IGV 需要操作系统预先安装 JRE (Java Runtime Environment, JAVA 运行环境库)，如果未安装，请到此网址下载适合系统的 JRE 包：

http://java.com/zh_CN/download/manual.jsp

二、启动 IGV

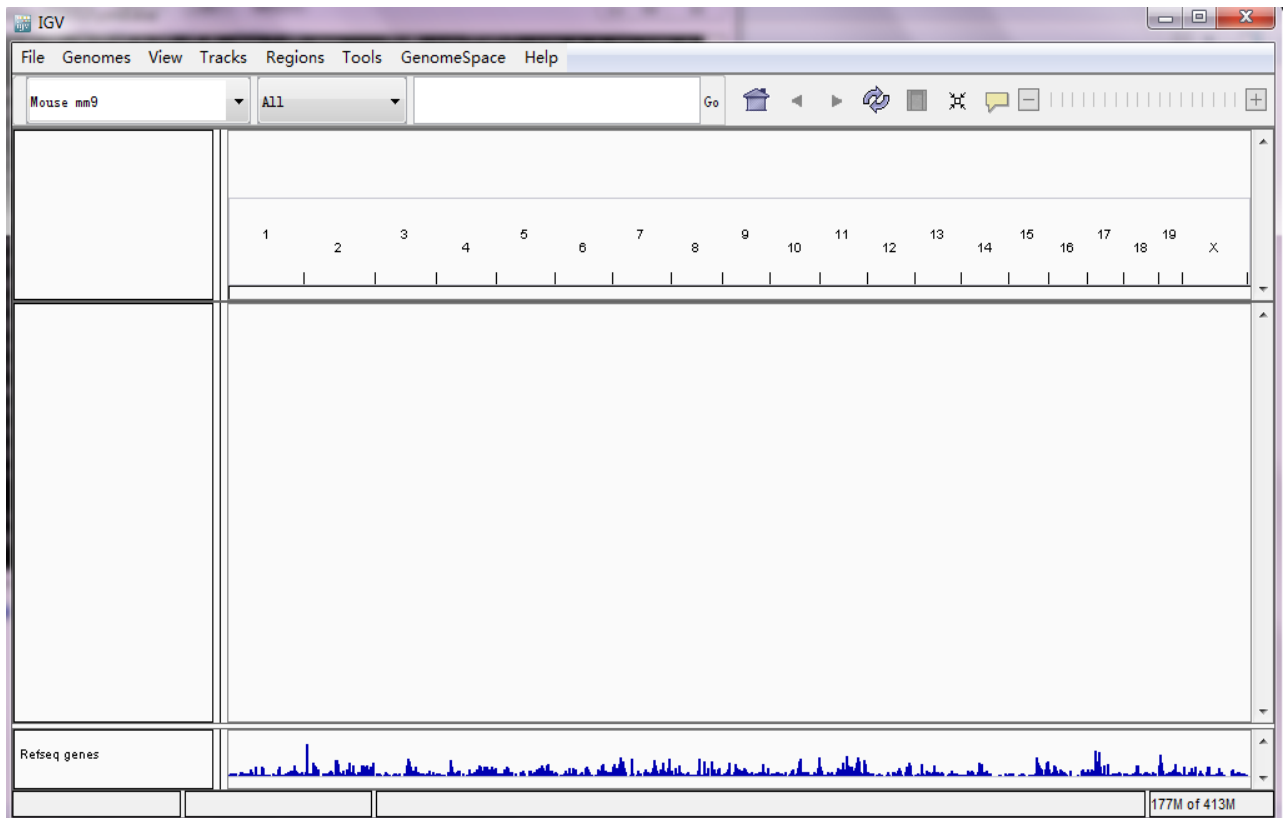
IGV 工具包为不同操作系统的用户，准备了三种类型的脚本来启动 IGV，对应如下：

- Windows: igv.bat
- Linux: igv.sh
- Mac OS: igv.command

用文本软件打开以上文件，可修改参数配置，主要关注-Xmx1200m 这个默认参数，里面的数字代表运行 IGV 时使用的内存大小（文件大需要的内存也大，可根据实际情况进行修改）。

本文档以 Windows 系统作为示例，进行操作说明。鼠标双击 igv.bat 文件，将会出现 windows 的 cmd 命令窗口，自动调用 JAVA 运行库，过一会儿出现如下 IGV 的初始

界面，就表示启动成功。



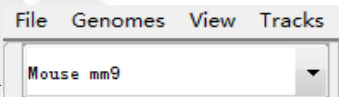
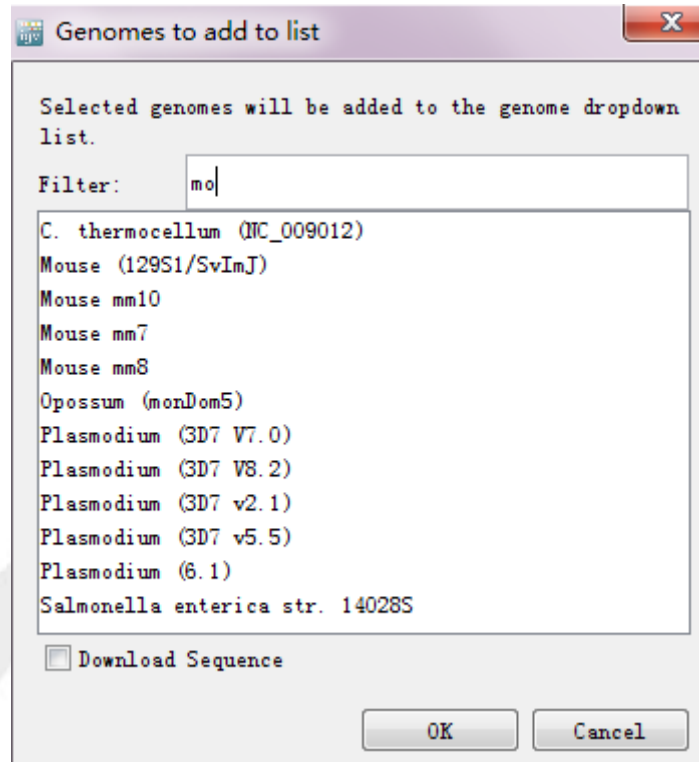
三、导入参考数据

1. 载入本地 FASTA 格式文件

点击菜单栏上的 Genome->Load Genome from File..., 从弹出的系统窗口选择已经准备好的本地 FASTA 文件。

2. 从远程服务端载入

用这种方法需要本机联网，点击菜单栏上的 Genome->Load Genome from server..., 出现如下窗口，在 Filter 栏里输入物种关键字进行选择。也可以勾选“Download Sequence”将远程服务端上的参考数据下到本地，以便下次没联网时也能够使用。



两种方法导入的参考数据，都会在Genomes菜单里面保存，可下拉点击进行切换，也可点击 Genomes->Manage Genome List ...对参考数据进行添加、删除、更改顺序等操作。建议使用第二种方法进行参考数据的导入，因为服务器上的文件较全，会一并导入基因注释文件，有助于在不同结构上查看 reads 的覆盖情况。

四、导入 BAM 文件

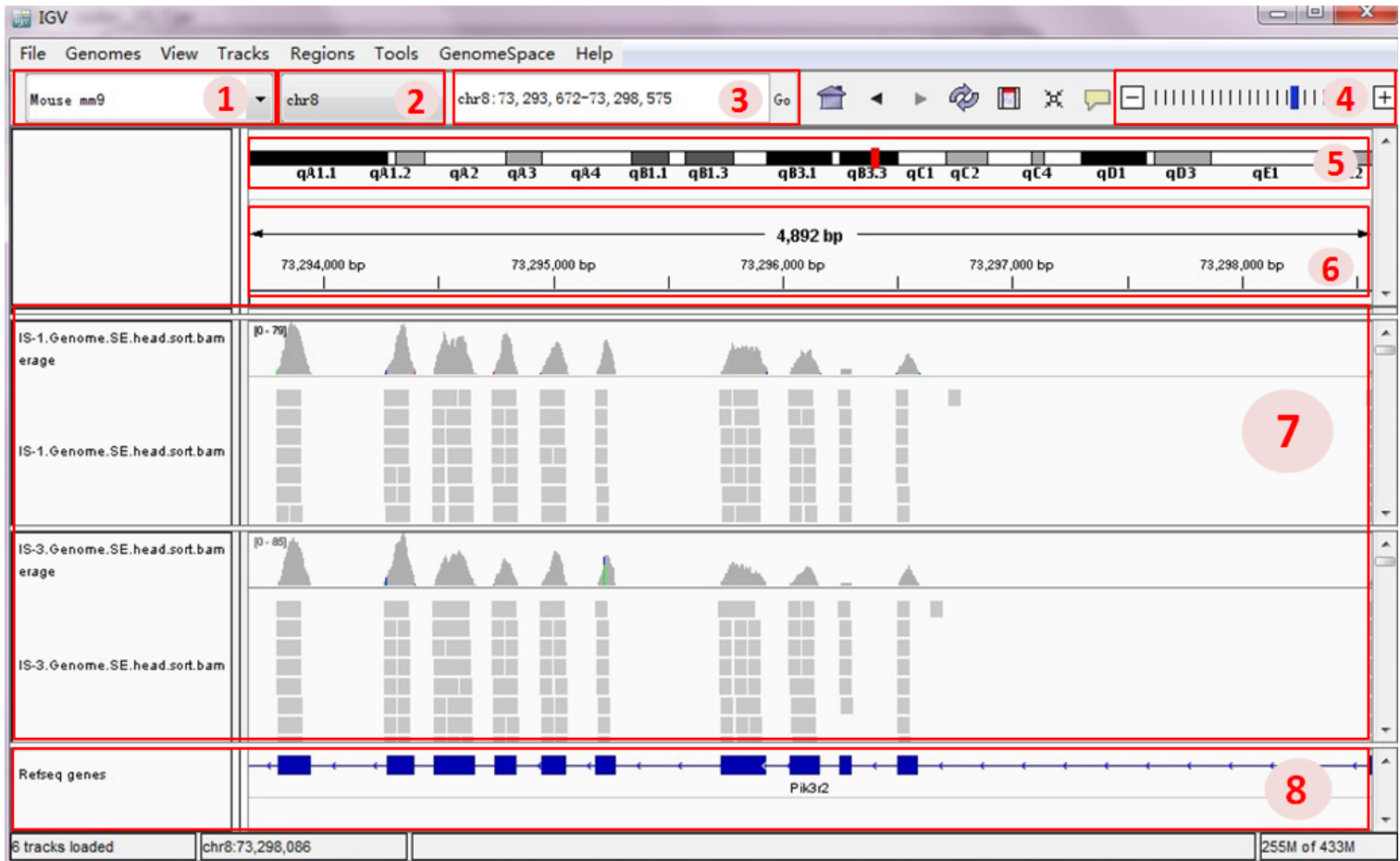
前提条件是样品的 bam 比对文件必须和其对应的 bai 文件在同一目录下，我们结果已提供所有样品的 bam 文件及 bai 。

点击 File->Load from File..., 从弹出的系统窗口选择要导入的样品 bam 文件。IGV 支持多个样品的 BAM 导入，重复此步骤即可。

五、查看比对信息

参考数据和 BAM 文件都载入成功后（完整界面如下图），便可对具体的比对信息进

行查看。下面以分区块形式，对界面进行简单的操作说明。



① 参考数据管理

在第三部分的“导入参考数据”已有介绍。

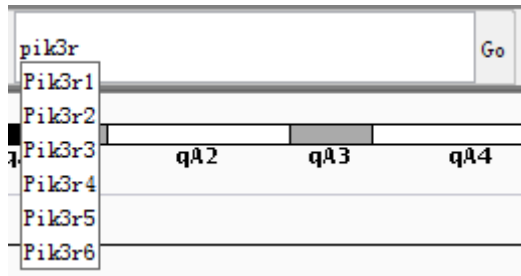
② 染色体选择

此栏会自动将导入的参考数据中的染色体名字，全部显示出来，可下拉点击进行某条染色体的选择。

③ 具体位点或元件查找

此栏是查找输入区，可输入一条染色体名字（如 chr1）、染色体某个具体位点（如 chr1:73298086）、染色体某段区域（如 chr1:73293672-73298575），也可在参考基因导入完整情况下，输入基因元件名，如下图中的 Pik3r2，IGV 软件会自动出现匹配

列表，更方便选择。



将要查看的位置输入后，点击 Go 按钮便行（这时软件运行会有点缓慢，要稍等）。

下方所有相关显示区，都会随此位置改变。

Go 右边的三角按钮 ◀ ▶，是之前查找过的位置的快捷键，点击向左箭头回到前一个查找的位置，点击向右箭头则回到后一个查找的位置。

④ 染色体区域不同尺度的放大或缩小

此功能只能在选定某条染色体的前提下使用（即②区若选了 All，此功能便会失效），减号显示更大范围的区域；加号显示更小范围的区域，最小精度能达到单个碱基的查看。也可以直接点击加减号中间的竖条进行尺度选择，更加快捷。

⑤ 染色体区域位置

此栏显示染色体位置信息，也是在选定某条染色体的情况才会出现。红色竖框显示当前染色体位置，可拖动红框改变显示的位置。

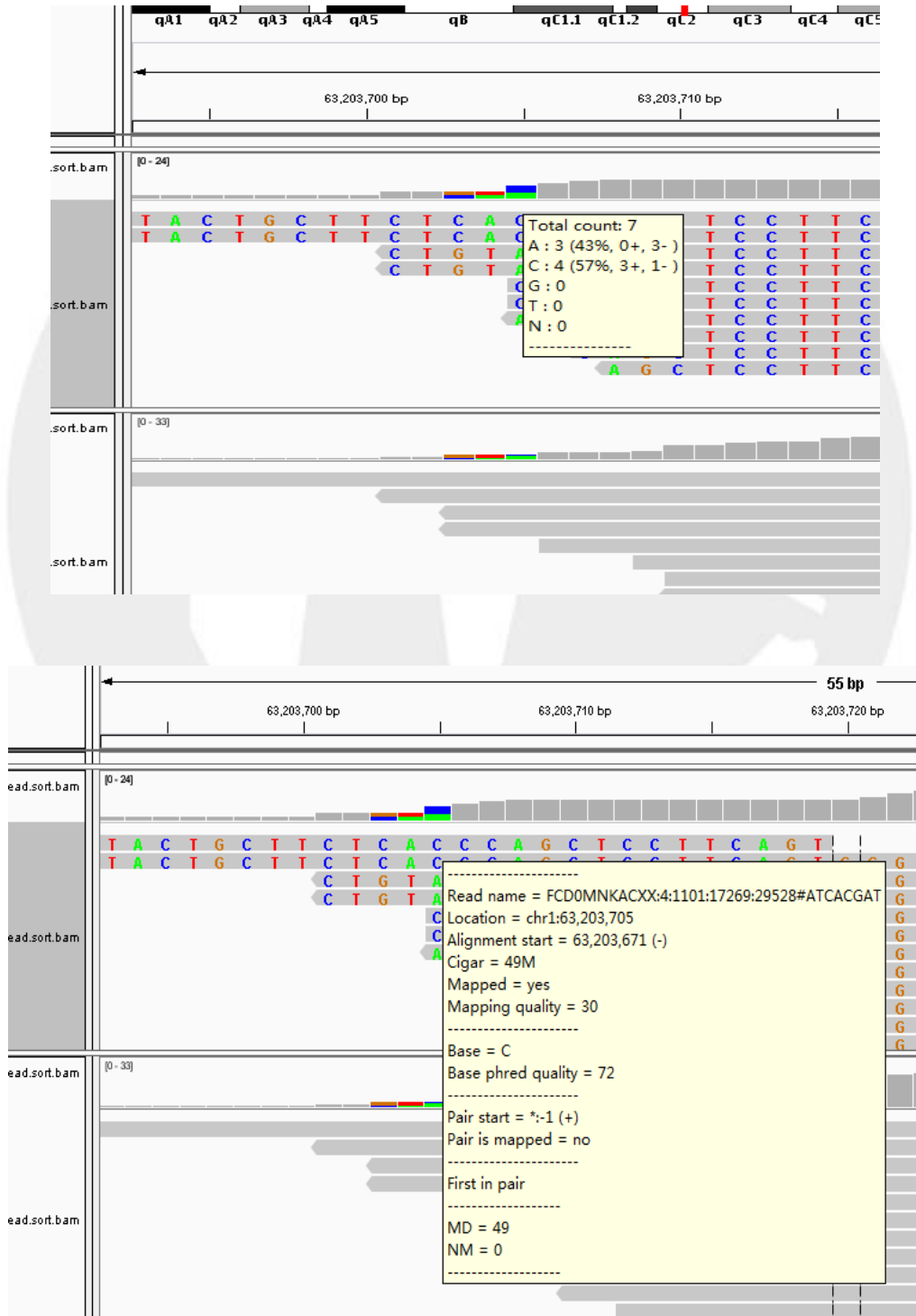
⑥ 染色体位置刻度尺

此项同样要在选定某条染色体的情况下用，两箭头之间显示当前区域总长度，下方的刻度尺显示在染色体的位置，单位有 bp、kb、mb，根据区域大小自动调整显示。可用鼠标选定刻度尺寸上的某段区域，进行放大。

⑦ BAM 文件比对信息显示区

此区是 BAM 比对信息的关键显示区，可导入多个样品，进行相同维度的比较。

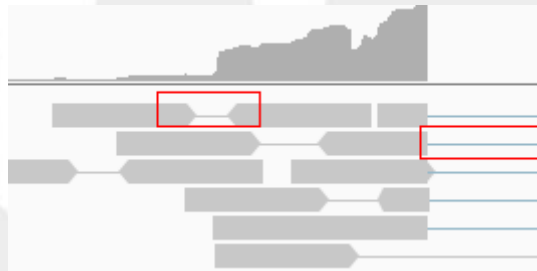
每导入一个样品,便会自动出现对应的两条 track,一个显示 coverage 覆盖深度情况,另一个是具体比对情况。每个样品的比对信息显示区窗口,都有垂直方向的滚动条,当 reads 较多时,可上下拖动进行查看,也可直接用键盘上的上下键操作。对于水平方向,可用键盘上左右箭进行拖动查看,但由于所有样品共用一个刻度尺,所以水平方向的操作会对所有样品适用。



上面两张图是一个例子，显示两个样品在某一小段区域内的各自比对情况。四种碱基用不同颜色代表（若出现某个碱基质量值较差，颜色会相应变浅），若某个位点只有一种碱基比对上，coverage 区便是灰色的；若有不同类型碱基比对上，coverage 区则会根据个数比例显示对应的碱基颜色。将鼠标放在某位点的 coverage 区上，会显示比对情况的统计信息，如第一张图的黄框所示（每种碱基比对上的个数、所占比例、正链几个、负链几个）；将鼠标放在某条 read 的碱基上，也会显示具体信息，如第二张图的黄框所示（各项内容便不详述）。

第一个样品显示了所有碱基，可通过在其相应的显示窗口点击右键，选择 Show all bases，第二个样品没有显示就是把 Show all bases 关掉了。若要移除某个样品，右键选择 Remove Track 即可。

对于 PE 比对，还可以右键选择 View as pairs...进行成对 reads 的查看。此时鼠标放在 read 上便会出现两个黄色框，同时展示第一条、第二条 reads 的信息。下图左边红框里的横线，便是连接成对 reads；但注意右边红框所标的横线，是表示一条 read 的 junction 比对，另一边未截到图的是比对到其他外显子上了（有的 read 会 junction 比对上好几个外显子）。



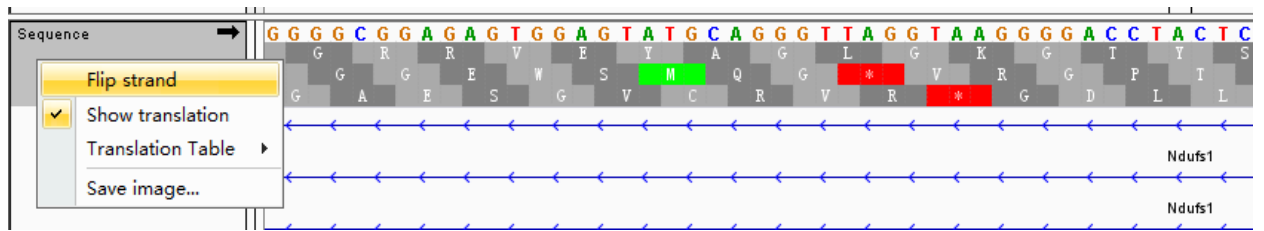
更多比对查看方法，可点击右键，在出现的菜单栏进行操作。

⑧ 参考数据显示区

有两条 track，一条是参考基因组的碱基序列区（sequence），另一条是基因转录本结构区（Refseq genes）。

1. Sequence 区要放大到一定程度后才会出现碱基和向右的黑色箭头，右击选择 Flip

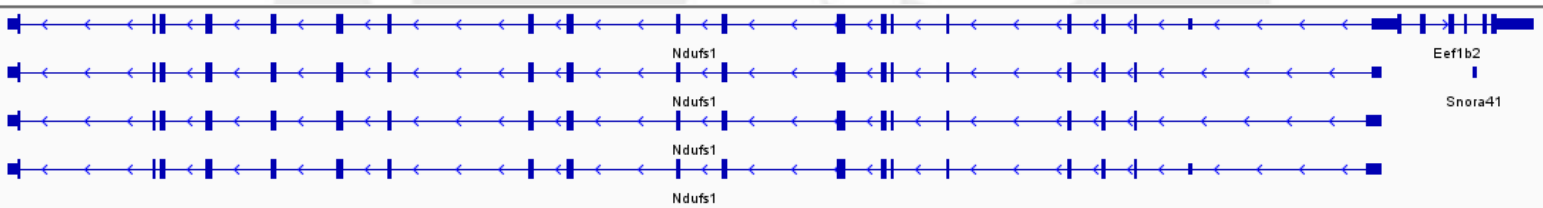
strand 可显示互补链，选择 Show translation 可显示三种情况的开放阅读框翻译情况，如下图。



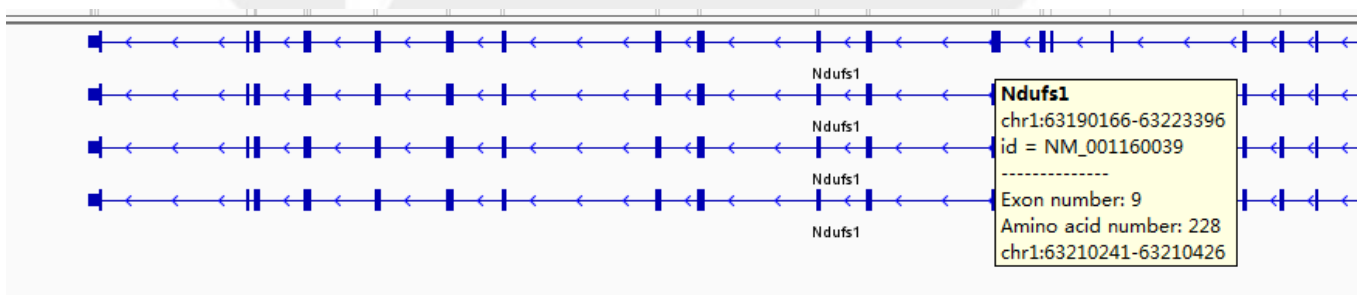
2. Refseq genes 区为参考基因对应的转录本结构情况，显示方式共有如下三种：

- Collapsed，将同源转录本合并显示
- Expanded，将同源转录本分开显示
- Squished，同 Expand 类似，但不显示转录本名字

可以右击鼠标选择其中一种，下图便是选择 Expanded 的一个例子，可直观看到基因 Ndufs1 有 4 种转录本，只是个别外显子组成不同。点击其中一个转录本，会跳转到 NCBI 相关页面来查看更多信息。



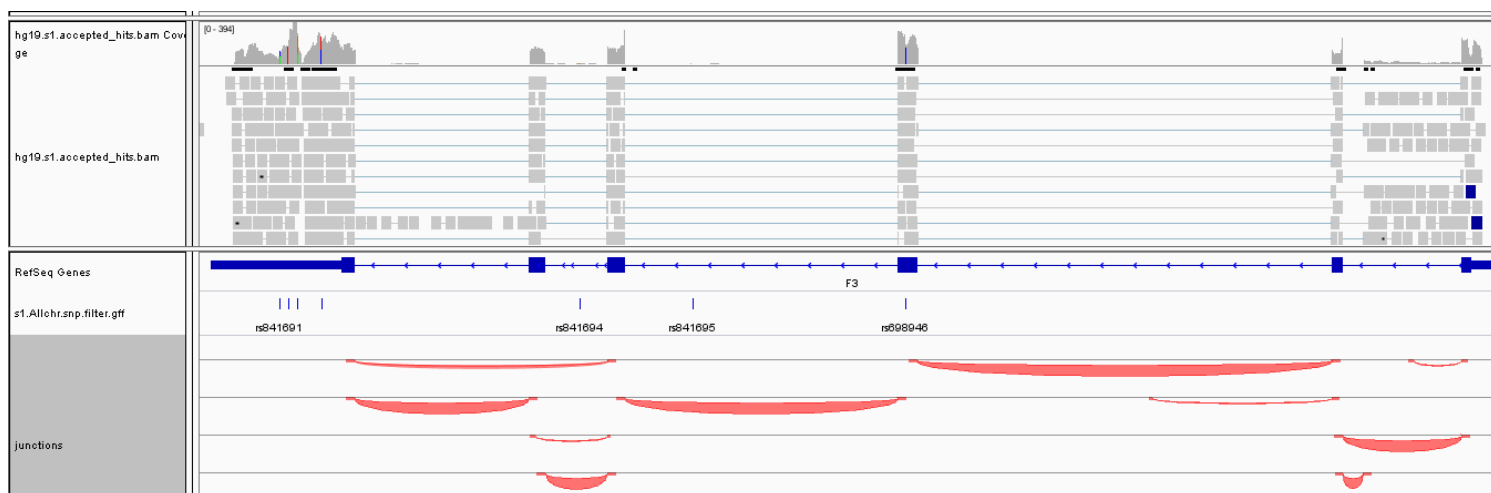
将鼠标置于转录本中的某个外显子，也会显示如下具体信息（基因名、基因在染色体位置、此转录本 ID、外显子编号、氨基酸编号、此外显子在染色体位置）：



快捷键 ctrl+b 可快速跳转到前一个基因的位置，ctrl+f 则跳转到下一个。

六、其他说明

IGV 可在不同尺度水平上，对比对数据（或基因变异结构）进行可视化浏览，界面友好、操作性强、直观、易用。如下图可清楚地在一整个基因水平上，看到样品 reads 的分布情况。



本文档主要针对 RNA 数据的基因组比对，进行 IGV 可视化操作说明。更多的 IGV 使用方法请浏览官网上的详细使用手册：

<http://www.broadinstitute.org/software/igv/UserGuide>