

FindJunctions

- [Introduction](#)
- [FindJunctions algorithm](#)
- [How to use FindJunctions within IGB](#)
- [Using FindJunctions from the command line](#)

Introduction

FindJunctions uses gapped RNA-Seq read to genome alignments to identify exon-exon junctions (i.e., introns). FindJunctions is implemented both as a visual analytics function within IGB and as a stand-alone, command-line program.

Within IGB, FindJunctions produces a new track showing exon-exon junction features labeled with the number of alignments that support that junction. The command-line program produces a BED file in which the score field contains the number of reads supporting a junction.

FindJunctions algorithm

FindJunctions operates on RNA-Seq read alignments, using alignments that contain gaps in the read sequence relative to the genomic sequence. These gaps in the read sequence correspond to introns and typically start and end with the so-called canonical splice site consensus sequences "GT" (5' end) and "AG" (3' end) for genes transcribed from the plus (forward) strand. For genes transcribed from the minus (reverse) strand, the consensus sequences relative to the plus strand are the reverse complement of the consensus splice site sequences (i.e., "CT" on the 5' end of the gap and "AC" on the 3' prime end).

For each alignment containing a gap, FindJunctions inspects the start and end coordinates of the gap and uses the genomic sequence to infer the strand, if available. FindJunctions creates a list of all such gaps, recording the strand and the genomic coordinates of the start and end coordinates. For each unique triplet of start, end, and strand, FindJunctions creates a scored junction feature and increments the score each time a gap supporting that feature is encountered in a dataset. Options are available to limit scoring to read alignments that have a given minimum number of bases flanking a gap and/or which having one unique mapping onto the genomic sequence.

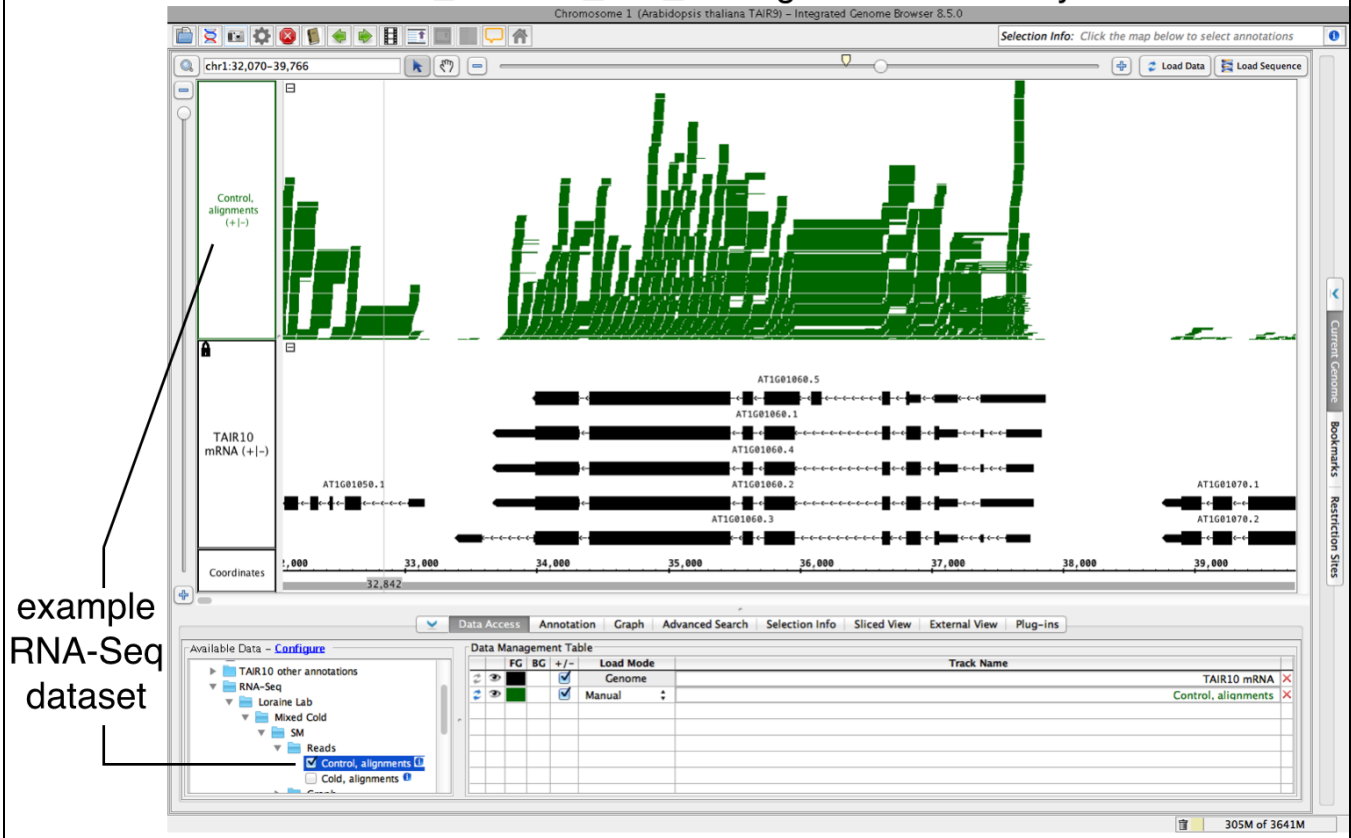
How to use FindJunctions within IGB

To use FindJunctions within IGB, you'll need to open and load your data, then run the FindJunctions track operation.

To open and load your data:

1. Open a BAM file.
2. Zoom in on the gene or region of interest.
3. Click **Load Data**.

A_thaliana_Jun_2009 genome assembly



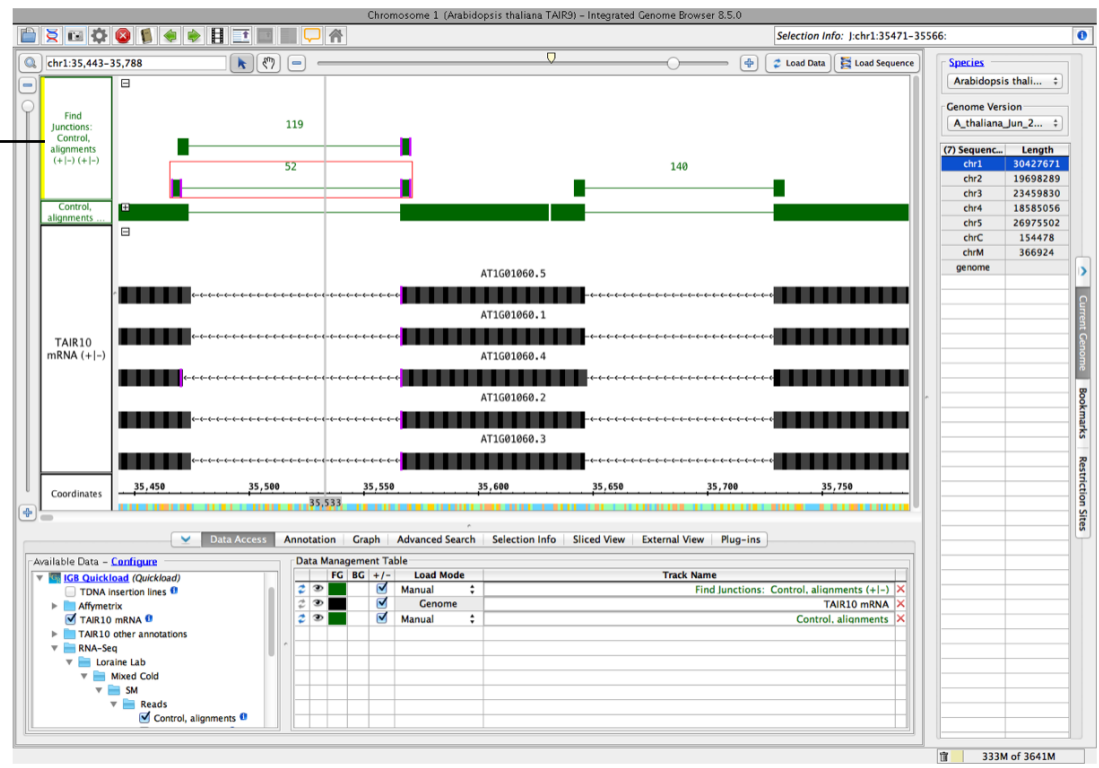
To run FindJunctions:

1. Right-click your data's track label.
2. Select **Track Operations > FindJunctions**.
3. Enter a value or use the default. At least this many bases must align across a putative intron for a read to be counted as support for a junction.
4. Select **OK** to run FindJunctions.

A new track will then appear containing junction features bracketing introns (see below). Labels report the number of spliced alignments that supported the junction.

A_thaliana_Jun_2009 genome assembly

FindJunctions
track



Using FindJunctions from the command line

To run FindJunctions as a stand-alone program, visit <https://bitbucket.org/lorainelab/findjunctions>. Follow the instructions there to compile FindJunctions and create a "jar" file using Apache Maven.

FindJunctions takes one argument, the name of a BAM file, and multiple options:

- -u (unique) Ensures that only uniquely mapping spliced reads (with NH tag equal to 1) will be used to construct junctions. Default is to use all reads regardless of mapping quality or number of mappings obtained.
- -n Number of bases that must map to either side of a putative intron for a spliced read to be used to create or support a junction feature. Default is 5.
- -b Absolute path to a .2bit format genomic sequence file that will be used to identify junction strand. **Required.**
- -o (output) Name of the junctions file that will be written. Default is print to stdout.

Output is tab-delimited BED12 format. The name field contains a name constructed from the location of the junction and the score field contains the number of spliced alignments supporting each junction.

Example:

```
java -Xmx1g -jar FindJunction_exe.jar -u -n 5 -b Genome.2bit -o FJ.bed sample1.bam,sample2.bam
```

Here, the -Xmx1g option specifies that the program can run with up to 1 Gb of computer memory (RAM) using the code in jar file (-jar) FindJunction_exe.jar.