Learning Goals

1. Run Tophat/Bowtie alignments of reads to see what are expressed regions
2. Run EST/Transcripts to genome alignments to find genes
3. Run protein to genome alignments to find genes
4. Visualize these results in Browser (IGV)

Running RNASeq Alignments

1. Download sequence for genome, proteins, and RNAs

   ```
   wget http://stajichlab.github.io/GenomeAnnotation/data/locus.tar.gz
   ```

2. Uncompress the file

   ```
   tar zxf locus.tar.gz  # uncompress the small dataset
   ```

3. Align the raw sequence reads against the genome locus with Bowtie/TopHat

   ```
   bowtie2-build locus.fa locus # index the database
   tophat locus RNASeq_locusonly.3H.fq # run the search
   samtools index tophat_out/accepted_hits.bam
   ```

- Let's investigate that alignment file.
- Open IGV.
- Load locus.fa from the Genomes menu
- File - Load the tophat_out/accepted_hits.bam
- File - Load locus.fungidb.gff

Aligning ESTs to the genome

1. Align ESTs to genome with exonerate

```
exonerate -m e2g ESTs.fa locus.fa --showtargetgff > EST.aln.gff
```

- Now load this GFF into IGV to visualize

5. Align proteins to genome with BLASTX

```
makeblastdb -in mory_proteins.fa -dbtype prot # format the db for BLAST
blastx -query locus.fa -db mory_proteins.fa -outfmt 6  # run BLASTX to find homologs
python blast2gff.py mory.BLASTX.tab BLASTX LGV_locus test > mory_proteins.BLASTX.gff
```

- Now load this GFF into IGV to visualize

6. Align proteins to genome with exonerate

```
exonerate  -m p2g  mory_proteins.fa locus.fa --showtargetgff > mory_proteins.aln.gff
```

- Now load this GFF into IGV to visualize