Steps for processing RNA-Seq data (steps 1-8 was done on Lewis):

1. Remove adaptors:

* Software: Cutadapt
* Base cmd: cutadapt -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -b A{100} -b T{100} -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -B A{100} -B T{100} --minimum-length 50 --trim-n --output RAPiD-Genomics\_HJYM3BBXX\_MIZ\_117501\_P01\_WB12\_i5-\_i7-75\_S622\_L003\_trimmed\_R1\_001.fastq.gz --paired-output RAPiD-Genomics\_HJYM3BBXX\_MIZ\_117501\_P01\_WB12\_i5-\_i7-75\_S622\_L003\_trimmed\_R2\_001.fastq.gz RAPiD-Genomics\_HJYM3BBXX\_MIZ\_117501\_P01\_WB12\_i5-\_i7-75\_S622\_L003\_R1\_001.fastq.gz RAPiD-Genomics\_HJYM3BBXX\_MIZ\_117501\_P01\_WB12\_i5-\_i7-75\_S622\_L003\_R2\_001.fastq.gz
* Ref: <http://cutadapt.readthedocs.io/en/stable/guide.html>

1. Quality Control:

* Software: Fastqc
* Base cmd:
* Ref:

We used the Tuxedo pipeline to analyze the Rna-seq data. Ref: https://www.nature.com/articles/nprot.2016.095

1. Align the RNA-seq reads to the genome:

* Software: Hisat

1. Convert Sam files to Bam:

* Software: Samtools
* Base cmd: samtools sort -@ 8 -o /group/kinglab/Patricka/Learn\_Mem/rna-seq\_data\_aligned/RAPiD-Genomics\_HJYM3BBXX\_MIZ\_117501\_P01\_WB12\_i5-\_i7-75\_S622\_L003\_aligned\_rna-seq.bam /storage/hpc/group/kinglab/Patricka/Learn\_Mem/rna-seq\_data\_aligned/RAPiD-Genomics\_HJYM3BBXX\_MIZ\_117501\_P01\_WB12\_i5-\_i7-75\_S622\_L003\_aligned\_rna-seq.sam

1. Assemble transcripts for each sample:

* Software: Stringtie
* Base cmd: stringtie -p 8 -G /group/kinglab/Patricka/base\_pop/genes/dmel-all-r6.18.gtf -o /group/kinglab/Patricka/Learn\_Mem/rna-seq\_data\_assemble/LL\_1\_assembled.gtf /group/kinglab/Patricka/Learn\_Mem/rna-seq\_data\_bam/LL\_1\_merged.bam

1. Merge transcripts from all samples:
2. Assemble transcripts:
3. Identify transcripts and genes: