===========EXAMPLE 1 - arm.Xsubset.fastq =========================

*This dataset contains reads from a mRNA sample after RNAi knockdown of the arm gene, in the drosophila Melanogaster cell line, S2.*

*The arm gene is found on chromosome X of the drosophila Melanogaster genome.*

*Here, to save time, we will restrict our analysis to this chromosome. You should for your own projects, you use all the genome information.*

Task: Align the reads to the X chromosome

To get information on tophat, terminal window and type:

**tophat -h**

\* The raw data is called arm.Xsubset.fastq in your home directory, in folder called course\_data

**arm.Xsubset.fastq**

\* The annotation Data is for chromosome X is:

**drosophilaMelanogaster.X.gtf**

\* The bowtie index is

**bowtieIndex/drosophilaMelanogaster.X**

Note: remember to specify the output directory in the tophat command.

**mkdir -p ~/course\_data\_output/arm**

Note: Specify option **--no-coverage-search** in the tophat command. This will speed things up.

Review questions:

- what does tophat do?

- What kind of aligner is tophat?

- What is the "bowtie-index"?

- How can specifying more mismatches with the -n option change the otuput?

- Why supply a gtf file to tophat?

- What are the output files from the tophat aligner?

- Why assemble transcripts with cufflinks?

- What is the gtf file and why do I need it?

- what data does the gtf file contain?

- what does the last column in the gtf file contain?

===========EXAMPLE 2 - smo.2Lsubset.fastq =========================

*This dataset contains reads from a mRNA sample after RNAi knockdown of the smo gene, in the drosophila Melanogaster cell line, S2.*

*The arm gene is found on chromosome 2L of the drosophila Melanogaster genome.*

*Here, to save time, we will restrict our analysis to this chromosome, 2L. You should for your own projects, you use all the genome information.*

Task: Align the reads to the 2L chromosome

To get information on tophat, terminal window and type:

**tophat -h**

* The raw data is called smo.2Lsubset.fastq

**smo.2Lsubset.fastq**

\* The annotation Data is for chromosome 2L is:

**drosophilaMelanogaster.2L.gtf**

\* The bowtie index is

**bowtieIndex/drosophilaMelanogaster.2L**

Note: Specify option **--no-coverage-search** in the tophat command. This will speed things up.

Review questions:

- What does the -p option in tophat do?

- How does --no-coverage search save time?

- What does the -g option do?