1. Obtain reference genome (.fa) , annotation file (.gff / .gtf) and your reads (.fastq probably). It will be helpful to make a working directory to put your files into and organize outputs.
2. Build Bowtie indexes:

bowtie2-build <reference\_genome> <index\_base\_name>

ie. bowtie2-build Ss\_SK36.fa Ss\_SK36

1. Run tophat to map each set of reads to the genome. Note that:

\* name in the seq\_id column (1st column) in the .gff has to exactly match the header name in the reference FASTA file

\* the reference FASTA file should be in the same directory as the indexes generated by Bowtie

\* basename of Bowtie indexes should match the basename of the reference FASTA (otherwise tophat will just have to reconstruct the reference genome)

tophat -p 8 -G <annotation\_file> -o <output\_folder> <base\_name> <reads\_file>

\* Add “--no-novel-juncs” option to exclude junction discovery

tophat -p 8 -G <annotation\_file> --no-novel-juncs <output\_folder> ...

ie. tophat –p 8 –G Ss\_SK36.gff --no-novel-juncs Ss\_S1\_thout Ss\_SK36 S1.fastq.gz

1. Run cufflinks to assemble transcripts (using the .bam file outputted by cufflinks):

cufflinks -p 8 –o <output\_folder> <.bam file>

ie. cufflinks -p 8 –o Ss\_S1\_clout Ss\_S1\_thout/Ss\_S1\_thout.bam

1. Create an “assemblies.txt” file that lists the assembly file for each sample to be merged(.gtf files, outputted by cufflinks)

./ Ss\_S1\_clout / Ss\_S1\_clout.gtf

./ Ss\_S2\_clout / Ss\_S2\_clout.gtf

./ Ss\_S3\_clout / Ss\_S3\_clout.gtf

…

1. Run cuffmerge on all the assembled transcripts (.gtf) outputted from cufflinks (for each organism)

cuffmerge -g <annotation\_file> -s <reference\_genome> -p 8 assemblies.txt

ie. cuffmerge -g Ss\_SK36.gff -s Ss\_SK36.fa -p 8 assemblies.txt

1. Run cuffdiff for each comparison you wish to make:

cuffdiff -o <output\_folder> -b <reference\_genome> -p 8 -L C1,C2 –u <merged\_assembly> \

<C1\_replicate\_1.bam>,<C1\_replicate\_2.bam>,<C1\_replicate\_3.bam> \

<C2\_replicate\_1.bam>,<C2\_replicate\_2.bam>,<C2\_replicate\_3.bam>

ie. cuffdiff -o Ss\_\_Sk\_Pg--CiaR\_Pg\_\_diffout -b Streptococcus\_sanguinis\_SK36.fa -p 8 -L Sk\_Pg,CiaR\_Pg -u Ss\_merged\_asm/merged.gtf \

S13\_Ss\_wnj\_thout/accepted\_hits.bam, S14\_Ss\_wnj\_thout/accepted\_hits.bam, S15\_Ss\_wnj\_thout/accepted\_hits.bam \

S16\_Ss\_wnj\_thout/accepted\_hits.bam, S17\_Ss\_wnj\_thout/accepted\_hits.bam, S18\_Ss\_wnj\_thout/accepted\_hits.bam

\* Note the “-L C1,C2” option – it defines the names of your conditions.

\* You can define more than two conditions – cuffdiff will then compare each possible pair

\* When you list the alignment files (.bam) that correspond to each condition, they need to be in the same order of the conditions you specified in the “-L C1,C2” option

\* Syntax needs to be perfect here – the replicate .bam files for each condition must be grouped together by commas (**no spaces in between**). Each group must be separated by a space or new-line.

1. Run cummeRbund analysis. You can read in the output folder from cuffdiff (see the cummeRbund commands document)