RNA Seq Workflow for bacterial RNA using Rockhopper

The program, user guide, and FAQ are found at this website:
http://cs.wellesley.edu/~btjaden/Rockhopper/

**Prior to first project, you will need to download the following freeware:*
Zipeg for Mac or 7Zip for PC to unzip files
IGV and Rockhopper
**You will need to have the replicon files for the genome you are using. See addendum at end of this document.*
**Quirks in the program:*
Do not include a degree sign (˚) in filenames used by Rockhopper. It crashes the program.
Do not have the replicon name highlighted when you hit submit. It won't run.
Make sure any open windows from previous runs are closed. It won't run.

1. Download your sequence files from BaseSpace, and unzip them using Zipeg (7Zip if on a PC). Make sure they are organized and in a well named folder.
2. Open up Rockhopper (you may need admin username and password).
3. For replicon name, click on the green DNA strand, hit ok, and browse to folder that you created that contains your replicon file
4. For Experiment Name – write your condition 1 that you want to be the numerator of your differential expression (eg. 37˚C on an upshift)
5. Browse to your first sample in this group and choose it.
6. For further samples in this group, just click on the blue plus sign and navigate to each sample in the group.
7. To add a new condition that will be the denominator of your differential expression, click on the red plus sign to add a new condition and its replicates as performed previously.
8. Finally, go under Options and choose Parameters, and make sure the defaults are appropriate for your experiment. Choose:
    a. IMPORTANT! Click on "verbose output" so you will have the raw data in the spreadsheet files as well as the analyzed. It will help you compare the quality of your various runs and understand your data.
    b. Use the default settings for the other parameters.
    c. IMPORTANT! In the Finder, create a folder where you want your results to be placed. In the Parameters window, use the Browse button to specify the path to the folder.
    d. Make sure to click SAVE before closing the Parameters window.
    e. Note: Clear cache under the options menu allows you to clear the inputs and start fresh.
9. Once all data is loaded and parameters defined, hit submit. The program will run with the time being dependent on genome size and computer power. It can go relatively quickly (5 minutes) on a computer with significant RAM. It's best to use the computer for bioinformatics in the CMB. All results go into the folder you named and designated in the parameters box.

10. In the Rockhopper Results file, you will see three .txt files. To view these, drag and drop them over the Excel icon in the dock. This will directly open them in Excel.
    a. Genome #_transcripts.txt (NC_004431_transcripts.txt) has the raw and normalized along with the statistics on differential expression. This is the file you will use the most.
    b. Genome#_operons.txt (NC_004431_operons.txt)
    c. Summary.txt- this gives an overview of the percentage of how many reads were matched and whether they matched to proteins, RNAs, or unannotated regions.
11. For beginning your analysis:
    i. Rockhopper is set up to use a false discovery rate of 1%. q-values are adjusted p-values and under this model, a q-value of ≤ 0.01 is considered significant. A FDR of 1% is fairly standard, but you can expand this to see a greater gene set by looking at genes with higher q-vlaues.
    ii. The numbers under the corresponding raw columns are the actual sequence counts for each condition.
    iii. Use the Expression Values to figure out your fold-change ratios. Expression values reported by Rockhopper for each transcript in each condition are similar to RPKM (reads per kilobase per million mapped reads) values. However, RPKM values are generally normalized by the total mapped reads, whereas the expression values reported by Rockhopper are normalized by the upper quartile of gene expression, which is a more robust normalizer. These are then normalized These can be used to calculate fold change (one expression value over the other), and then take the log2 of that fold change number to get the actual fold change values we are used to. Genes of interest in the list can be further analyzed in Kegg or other database (trying to learn CummeRbund for further visuals.
12.
13. Once done, try hitting view in IGV. If it fails, you can load things manually as follows:
    i. For visualizing the data, open up IGV
    ii. Under File menu, go to "Load Genome from File" and navigate to your .genome file (for this first EPEC round it is in the Replicon Info folder.
    iii. Under File menu, go to "Load from File" and navigate to and load all files in your Rockhopper_Results Folder, in the "GenomeBrowserFiles" (may want to only load plus or minus strands to start). Hold the Command button down to choose multiple files that are not necessarily in sequential order.
    iv. Slider at the top, right hand corner allows you to zoom in and out on the genome.
    v. Right click (hold down Control button and click if your mouse is not enabled) on the data range numbers in the upper left hand corner of each track for you plus or minus reads, and from the drop down menu change the "Set Data Range" from 0-800.
    vi. You can zoom in on areas of interest just by boxing out a portion of the line that conveys the number of basepairs you are looking at in the window (has a little black arrow on each end)

vii. You can change track height and color, and data range by right clicking on the track name, on the far left) and choosing operation from the drop down menu.

- Enjoy the exploration☺

Addendum: Creating your own replicons

- If you need to create your own replicons either because the replicon listed is not working or your organism is not present, you will need to do the following:
  - You will need a .fna, a .ptt, and a .rnt file placed in a single folder to submit as your replicon.
  - The .fna and .ptt files are the bare essentials. .rnt is optional but is what contains the rRNA and tRNA info.
  - Most of these files can be found at: ftp://ftp.ncbi.nih.gov/genomes/Bacteria/. The folder you download from this source will contain all of these files and many more you won't likely need.
  - If you can't find there try googling your bacteria name and .fna, etc.
  - Once you navigate to the file you need, do "Save page as" and place in your designated replicon folder.