

- 1) Transfer wild type and mutant .fastq files into separate folders in the Galaxy Shared folder.
- 2) Copy these folders into a folder in the Galaxy machine. I chose "Galaxy upload" because the other necessary files for the analysis were in this folder.
- 3) Open RNAmapper in the firefox window.
- 4) Import .fastq files into the RNAmapper history (\*\*run steps 4-7 for wild type first then repeat for mutant)
  - upload each file individually using the "upload file" option under the Get Data menu on the right.
  - select Zv9.69 as the reference genome
  - if they fail to upload after a few minutes, delete them, purge the history and try again
  - if they fail to upload again, close the machine and restart
- 5) Follow Adam's instructions to clean the files and align them to the genome using Tophat in the terminal:
  - a) Run Trimmomatic (In the RNAmapper tools on the left under NGS:QC and manipulation) on all the files - this gets rid of the less ideal reads. Each run will produce 2 files, you need the bigger of the two.
  - b) Save EACH of the bigger of the two - click the file names on the right, check their size, then click the floppy disc icon.
  - c) They will save to your Downloads folder. It is easiest to then go in and rename them similar to the original. For example, use the same name but add \_trim to the end.
  - d) Copy and paste these files into a new folder in the galaxy upload folder. I made new folders named wt\_trim and mutant\_trim.
- 6) Run Tophat
  - a) Bring up the "command line" by going to the left side of the Galaxy desktop, which reveals the programs, and click on "terminal" which is the program that allows you to input commands to the operating system using text commands.
  - b) When it opens you will get a line that should say
 

```
galaxy@galaxy-VirtualBox:~$
```

That line tells you first "who you are" - galaxy@galaxy-VirtualBox  
Then it tells you "where you are" - ~  
The ~ symbol (tilde) means you are "home"  
and then says its ready to accept commands with the "\$"
  - c) type "cd galaxy\_upload" and hit enter  
this brings you into the folder with your trimmed fastq sequences and other pertinent files
  - d) type "cd wt\_trim" (or cd whatever you named your folder with the trimmed fastq files) and hit enter
  - e) type "ls" and hit enter  
This will show you all of the files in this folder which should be your trimmed fastq files.
  - f) Now that you are in the right folder, you will setup some variables for Tophat
  - g) Tophat needs to know where the latest zebrafish gene model file

type "GTF=/home/galaxy/galaxy\_upload/GTFs/Zv9.69.gtf" and hit enter  
this makes "GTF" the path, to call the variable you have to  
type \$GTF

h) to confirm that the file path is correct, type "echo \$GTF" and hit enter. The file path should show up. If you get something like "no such file or directory", you typed something wrong OR the gene model is in a different place.

i) Next, do the same thing with the reference genome  
type "refGenome=/home/galaxy/reference\_genomes/Zv9.69/Zv9.69" and hit enter

j) Run Tophat  
type "tophat -G \$GTF \$refGenome then list all of your trimmed file names separated by commas but no spaces" and hit enter  
For example: mine was "wt\_001.gz,wt\_002.gz,wt\_003.gz,etc" and hit enter

7) This will make a tophat\_out directory where in a day (or more...) you'll have an alignment!

8) Do the same for your mutant sample.

9) Next, index and sort both your wildtype and mutant .bam files using Samtools in the terminal. Do the wildtype first.

a) Go into the tophat\_out folder that has been created in your trimmed sample folder by typing "cd tophat\_out" enter

b) In this folder, you will have several files, including the one that has your aligned reads. It is called "accepted.bam".

c) First, rename this file something useful, like wt.bam

d) To do this, in Galaxy space, navigate to the this file by going through the appropriate folders:

- click on the home folder icon on the left
- double click the galaxy upload folder
- double click on the folder containing your trimmed sequences. In my case, wt\_trim.
- double click on the tophat\_out folder
- right click on the accepted.bam file and rename it. For example: wt.bam

10) Now, go back to your terminal by clicking the terminal icon on the left.

11) You are in your tophat\_out folder and should see your renamed .bam file.

12) Run Samtools

a) type "samtools sort wt.bam wt\_sort" and hit enter

- These instructions will call samtools, tell it to sort, tell it the file to sort, and tell it what to call the output.
- This will take a few minutes, and the program doesn't "tell you" its doing anything it just sits there looking stuck but just wait it out and it will produce a file called wt\_sort.bam

13) Run Samtools again, this time to index the files.

a) type "samtools index wt\_sort.bam" and hit enter

b) This will produce a file called wt\_sort.bam.bai which is an indexed version of the sorted file necessary for scanning through the sequencing data.

14) Run both Samtools functions for your mutant files as described for the wildtype. Steps 9-13.

15) Finally, now that you have sorted and indexed .bam files for your wild type and mutant samples, you can run RNAmapper.

16) Open RNAmapper using the firefox icon on the dock on the left.

17) Open a new history using the options tab above the current history on the left.

a) In my experience, RNAmapper works best with as little storage space used as possible, which is shown above the options tab on the right. If possible, delete histories and purge them prior to proceeding with .bam analysis.

18) Using the "Upload File" option under the Get Data menu on the left, upload your wt and mutant sorted .bam files (Select Zv9.69 as your reference genome). Use the .bam and not the .bai files. The .bai files are useful for another step; analyzing the sequences in IGV viewer.

19) Similarly, upload other necessary files including: wt\_SNP\_set.vcf and danio\_rerio\_variation\_dbSNP\_130.vcf (both found in the rnaseq\_mapper\_resources folder in the galaxy upload folder of the downloaded RNAmapper version. Also upload Zv9.69.gtf in the GTFs folder in your galaxy upload folder.

20) Now that you have your sorted .bam files and these other essential files in your history, you can run RNAmapper.

21) To do this, first select the workflow option in the top toolbar in RNAmapper

22) Left click on the RNAmapper\_from\_bam and choose "run"

23) Select the appropriate files from the drop down menu

- a) WT alignment - sorted wildtype .bam file
- b) MUT alignment - sorted mutant .bam file
- c) transcript list Zv9.65 GTF - Zv9.69.gtf
- d) Input: dbSNP.vcf4 - danio\_rerio\_variation\_dbSNP\_130.vcf
- e) Input: WT\_SNP\_list.vcf4 - wt\_SNP\_set.vcf

24) Scroll to the bottom and click "run workflow"

25) Watch the steps on the right change from purple to yellow to green as the RNAmapper works its magic. If any turn red, something has gone terribly wrong.

a) When the workflow completes, you will have several useful things in the history on the right. First, check out your ReportmakerR as described in the RNAmapper instructions by clicking on the top box in your history and then clicking on the eye icon. Make note of the chromosome that contains the highest region of homozygous SNPs as well as a loose chromosomal position. This is found in the top portion of the Reportmaker file. This information will be helpful!

b) If high or moderate impact SNPs are not found in this document, start looking through the other SNPs in the region using a tab separated file found in your history.

- I have found that in 3 out of 3 my mutants, a causative SNP is present in the region that is not shown in the ReportmakerR. I am not sure why

this is but, with all of the RNAseq information, you can typically find your mutation with a little digging. See below.

26) If ReportmakerR does not identify potentially causative SNPS, mine your data!

a) The third box from the top in your history should be titled SnpEff on dataXX. If you click on the title, the box will open and reveal that you have tabular data. There is another similarly titled box above this one but the data is in html format. You want the tabular data!

b) Click on the disk icon to download and save this file. It should save to your downloads folder in your "home" folder. There is a shortcut to this folder in the dock at the top.

c) Open this SNPeff file in excel

- To do this, first open excel and then use the File-open path to open this file. It will not work to simply drag it in!

- This file has all of the homozygous SNPs found in your mutant on all chromosomes.

d) Once open, erase the SNPs found on other chromosomes to make the file easier to use.

e) Now scan through your potential SNPs in order of interest. I chose to sort the excel file by "effect", which is the predicted change caused by the SNP. After sorting, I looked first at the stop\_gained, stop\_lost, and splice acceptor/donor changes. Non-synonomous coding mutations are a bit trickier and, as such, I left them to last.

27) With potential mutations in hand, analyze your sequencing data in IGV viewer, as described in the RNAmapper package, to determine their prevalence in mutant vs wildtype (100% vs 50%) and their effect on the protein. I found that using bioinformatic databases such as ensembl and ncbi were especially helpful during this portion.

28) With your mutation(s) in hand, begin the careful process of verification using rescue and knockdown approaches.

Good luck with your search!