PA11004

Summary

v3bc22 has 8,265,420 reads, v3bc23 has 17,027,622 reads. 8,016 genes were detected with at least 5 reads in v3bc22 and 11,373 in v3bc23 - these are reasonably consistent with other RNA-seq data we’ve seen with similar numbers of reads.

A brief summary of the mappings shows a high degree of rRNA contamination which could probably be corrected by a poly-A or ribosomal subtraction step on the front end, but even so the % mapping stats are very low. “filter” below would include mapping to adaptor, tRNA, and ribosomal RNA sequences; rRNA and adaptor sequence matches are then enumerated. Details are below.

Overall exon coverage is good - more uniform than the directional RNA cloning approach we usually take, as shown in the following two panels (Ex14_1 was a sample generated by the ABI WT RNA-seq kit). Double-stranded cDNA library preps are typically more uniform than strand-specific preps. Note you can see a SNP in v2bc22.

I haven’t aggregated the data enough to know whether this is real or not, but there might be a slight bias toward extra coverage at the 5’ end of the gene, for example here (this is the gene with the most reads in v3bc22):
Note that all the reads in the ABI kit example are shaded red, indicating they originate in one orientation (in the mRNA sense strand), while the reads from Sabine's prep are fairly random as they should be.

What categories did the reads map to?

```bash
$ cat v3bc22/mappingstats.txt
Total number of f3 reads = 8265420
f3 Reads mapping to filter = 3101853
f3 reads mapping to adaptor filter sequences = 2553
f3 reads mapping to rRNA = 2186015
f3 Reads mapping to genome = 2670250

Total number of f5 reads = 8265420
f5 Reads mapping to filter = 2271195
f5 Reads mapping to genome = 2675251
```

```bash
$ cat v3bc23/mappingstats.txt
Total number of f3 reads = 17027622
f3 Reads mapping to filter = 6509508
f3 reads mapping to adaptor filter sequences = 88938
f3 reads mapping to rRNA = 2542554
f3 Reads mapping to genome = 5332390

Total number of f5 reads = 17027622
f5 Reads mapping to filter = 2864844
f5 Reads mapping to genome = 5538265
```

Where to find the processed data

The final data files consist of .bam files with the maps, and "countagresult.txt" files which contain RPKM-normalized values for each gene. There are several other outputs, such as exon junction reports, which I can point you to if you'd like to look at them.
These results are available on Fourierseq in:
/home/solid/SA11024/iyer
in two subdirectories, one per barcode.