Quality Control in RNA-Seq Alignment

Nima Hejazi

Division of Biostatistics

University of California, Berkeley

`nh@nimahejazi.org`

1 QC of RNA-Seq Reads (Post-Alignment)

The present report concerns the process and results of performing quality control (QC) post-alignment on RNA-Seq reads from the study described in the "Developmental regulation of human cortex transcription and its clinical relevance at base resolution", Jaffe *et al.*, *Nature Neuroscience*. In order to perform basic quality control on RNA-Seq reads after the alignment procedure, FastQC (v. 0.11.4) was used via the Galaxy computing platform. As the re-analysis performed in this project was focused on only the *paired-end reads*, only these reads were used in the QC process. In all cases, more than 98% of reads aligned successfully to the reference transcriptome. Most reads aligned with exceptionally high quality (this is not usually the case in my past experience), with a (rough) Phred score average of 33 or so. Based on a cursory examination, there does not appear to be a significant trend between read sequence alignment and either classes of samples (fetal versus adult) or sample quality (as indicated by RIN). Based on the QC results, we conclude that the alignment procedure was successful and adequate.

1.1 The Results of Post-Alignment Quality Control

The results of performing quality control on the samples is reported below:

Adult samples:

- SRR1554535: 98.26% of reads aligned concordantly, average quality score ≈ 33, reads with a score above 30 $\approx 77\%$

- SRR1554536: 99.35% of reads aligned concordantly, average quality score ≈ 34, reads with a score above 30 $\approx 84\%$

- SRR1554539: 98.51% of reads aligned concordantly, average quality score ≈ 35, reads with a score above 30 $\approx 92\%$

Fetal samples:

- SRR1554537: 98.42% of reads aligned concordantly, average quality score ≈ 34.1, reads with a score above 30 $\approx 82\%$

- SRR1554538: 98.51% of reads aligned concordantly, average quality score ≈ 34.8, reads with a score above $30 \approx 86\%$

- SRR1554541: 98.59% of reads aligned concordantly, average quality score ≈ 33.6, reads with a score above $30 \approx 79\%$