rnaseq-methods

April 6, 2015

Methods

All samples are processed using RNA-seq pipeline implemented in bcbio-nextgen project. Raw reads will be examined for quality issues using FastQC to ensure library generation and sequencing are suitable for further analysis. Adapter sequences, other contaminant sequences such as polyA tails and low quality sequences with PHRED quality scores less than five will be trimmed from reads using cutadapt (Martin (2011)). Trimmed reads will be aligned to build hg19 of the Hsapiens genome, augmented with transcript information from Ensembl release GRCh37.75 using STAR (Dobin et al. (2013)).

Alignments will be checked for evenness of coverage, rRNA content, genomic context of alignments (for example, alignments in known transcripts and introns), complexity and other quality checks using a combination of FastQC (Andrews), RNA-SeQC (DeLuca et al. (2012)) and custom tools. Counts of reads aligning to known genes are detected by featureCounts (Liao, Smyth, and Shi (2014)).

Normalilzation and differential expression at the gene level are called with DESeq2 (Love, Huber, and Anders (2014)), which has been shown to be a robust, conservative differential expression calller.

Bibliography

Andrews, S. "FastQC A Quality Control Tool for High Throughput Sequence Data." *http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

DeLuca, David S, Joshua Z Levin, Andrey Sivachenko, Timothy Fennell, Marc-Danie Nazaire, Chris Williams, Michael Reich, Wendy Winckler, and Gad Getz. 2012. "RNA-SeQC: RNA-Seq Metrics for Quality Control and Process Optimization." *Bioinformatics (Oxford, England)* 28 (11): 1530–32. doi:10.1093/bioinformatics/bts196. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22539670&retmode=ref&cmd=prlinks.

Dobin, Alexander, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. 2013. "STAR: ultrafast Universal RNA-Seq Aligner." *Bioinformatics (Oxford, England)* 29 (1). Oxford University Press: 15–21. doi:10.1093/bioinformatics/bts635. http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/bts635.

Liao, Yang, Gordon K Smyth, and Wei Shi. 2014. "featureCounts: an Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." *Bioinformatics (Oxford, England)* 30 (7): 923–30. doi:10.1093/bioinformatics/btt656. http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btt656.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12): 550. doi:10.1186/PREACCEPT-8897612761307401. http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=25516281&retmode=ref&cmd=prlinks.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." *EMBnet.Journal* 17 (1): pp. 10–12. http://journal.embnet.org/index.php/embnetjournal/article/view/200/479.