PostDoc Journal
Vol. 1, No. 4, April 2013

Journal of Postdoctoral Research
www.postdocjournal.com

It's all in the Prep

Joshua Hill, Ph.D.
Texas A&M AgriLife Genomics and Bioinformatics Service, 2123 TAMU, College Station, TX 77843, USA

Email: joshuaehill@gmail.com, joshua.hill@ag.tamu.edu

Wery's et al. recent paper "Zinc-mediated RNA Fragmentation allows robust transcript reassembly upon whole genome transcriptome RNA-Seq" sheds light on a problem that affects many Next Generation Sequencing (NGS) experiments; the ability to accurately represent the RNA present in a given sample. Bias for RNA-seq can be simply defined as a false increase or decrease of a given RNA transcript. If that definition is expanded over a whole transcriptome then it is easy to see how false conclusions can be made and why bias must be considered.

There are several different methods for preparing libraries for sequencing and they all have some inherent potential for bias. DNA or RNA samples to be sequenced go through several processes, which include some type of mRNA enrichment (poly A selection and/or ribosome depletion), shearing to a proper size, DNA/RNA repair, ligations and PCR amplifications. Additionally, any type of DNA contamination will have negative consequences for RNA-Seq. The initial cleaving or shearing step is a key part of short read NGS technology. If the final library size is too large or too small the machine itself will not run correctly and the data compromised. Shearing can be accomplished mechanically through ultrasonication or nebulization, chemically or enzymatically.

The paper highlighted here compares two methods for shearing RNA, 1) enzymatic-based using RNase III and 2) chemical-based Zinc-mediated. RNase III cleaves at specific structures or motifs which may result in an over/under representation of transcripts sequenced as compared to Zinc-mediated that cleaves RNA at the 2' hydroxyl group of ribose irrespective of motif or structure. Wery et al. show that by using Zinc-mediated cleaving, overall coverage (the number of times a single base is sequenced, often referred to as X coverage) is increased and reassembly is more robust as there was an increase in overlapping reads. With more overlapping reads (identical short sequences that are found on different pieces in the library) reassembly is easier because the computer algorithms can then logically piece together longer sequences with a higher confidence. When using 5' rapid amplification of cDNA ends (5' RACE) and 3' serial analysis of gene expression (3' SAGE), both of which allow the identification of known sequences using small regions of sequenced material, Zinc-mediated cleaving had a tighter dispersion around the mean as compared to RNase III cleaving. This allows a more precise

definition of the 5' and 3' ends to be calculated.

There are many chances in an NGS library protocol that have the opportunity to introduce bias if not thought out carefully. While there are other steps that can be taken to help eliminate bias as presented in Wary et al. NGS is still new and improvements are being made everyday in kit construction and data analysis. Careful consideration of problems that might be encountered and how they affect down stream processes can help mitigate bias in NGS.

Original paper:
"Zinc-mediated RNA fragmentation allows robust transcript reassembly upon whole transcriptome RNA-Seq Methods". Maxime Wery, Marc Descrimes, Claude Thermes, Daniel Gautheret, Antonin Morillon. Available online 21 March 2013

http://www.sciencedirect.com/science/article/pii/S1046202313000789

http://dx.doi.org/10.1016/j.ymeth.2013.03.009