Methods and Results: DNA Workflow and Sequencing with UMI Adaptors

Methods: RNA-Seq Workflow

- Universal Human Reference RNA (Agilent) with ERCC RNA Spike-in-Mix (Thermo Scientific) was used for library input. mRNA was isolated using the NEBNext Poly(A) mRNA Magnetic Isolation Module and libraries were prepared using the Ultra II Directional RNA Library Prep Kit (NEB #ET766).
- Libraries were sequenced on a NextSeq™ 500 (2x70) and downsampled to a minimum of 1M reads.
- Sequencing reads were aligned to GRCh38, duplication rates were computed using Picard MarkDuplicates or bgsub AlignateBamWithUmis

Results: RNA-Seq with UMI Adaptors

96 NEBNext Unique Dual Index UMI Adapters perform consistently

A) Normalized Yields

B) Clustering Efficiency

Removing duplicate reads detected by UMI’s significantly shifts transcript abundance

A significant number of transcripts have a 2x2 change in transcript counts when duplicates are removed. (A) The average number of transcripts with a 2x2 shift in abundance is shown when duplicate reads are removed based on UMI analysis versus no removal of duplicates. An average of 4 technical replicates at three inputs (1,000 ng, 100 ng, and 10 ng) is shown before and after removal of duplicate reads. Each sample was downsampled to 10 million reads. Libraries prepared with a 10 ng input showed the greatest number of transcripts affected by PCR amplification. (B) MDM4 is an example of a gene with a high proportion of mapped reads determined to be PCR duplicates (red bars) based on UMI analysis. Utilizing this information, it is possible to remove duplicate reads introduced by PCR amplification for downstream analysis.

Comparison of unique dual index UMI adapters used for library preparation.

Comparison of library yields and duplication rates with various unique dual index UMI adapters. (A,B) During adaptor ligation the NEBNext Unique Dual Index UMI adapters (UMI length = 11 bases) or IDT iSeq Dual Index UMI Adapters (UMI length = 9 bases) were used. (A) The average library yield of triplicates is shown for three starting total RNA inputs: 10, 100, and 1,000 ng. Final library yields were quantified using the Agilent TapeStation 4200. (B) Libraries were sequenced on the Illumina NextSeq 500 and downsampled to 5 million reads. Duplication rate was determined utilizing the UMI sequence and mapping location. NEBNext Unique Dual Index UMI Adapter libraries produced libraries with a lower percentage of read duplicates. (C) Duplication rate for libraries produced with NEBNext UMI adapter libraries analyzed by two computational methods: utilization of UMI’s (right orange) or read mapping position alone (dark orange).

Conclusions

Incorporating our NEBNext Unique Dual Index UMI adapters into DNA sequencing results in:
- PCR-free library prep
- Improved detection of low variant frequencies
- Error correction through consensus sequence building
- More accurate removal of duplicate reads

Incorporating our NEBNext Unique Dual Index UMI adapters into RNA-sequencing results in:
- Robust library yields and high-quality sequencing metrics
- No introduction of sequencing bias with UMIs
- More accurate assessment of duplicate reads increases the number of reads that can be used for downstream analysis
- Improved quantification of transcript abundance

Acknowledgements:
Thank you to the NEB Sequencing Core (Laura Mazzaola, Danielle Fuchs, Kirsten Augukewicz) for all their sequencing support.