Degradome re-annotation of single-cell RNAseq datasets identifies signatures of miRNA biogenesis and apoptosis

Jonathan Preall (jpreall@cshl.edu), Cold Spring Harbor Laboratory, **Jonathan Preall**, Cold Spring Harbor Laboratory

Typically, 3'-end directed single cell sequencing data is compressed into a simple matrix quantifying digital expression counts of annotated genes. Intriguingly, we have found that a common side-product of the 10X Genomics chemistry, namely the spurious incorporation of adapter template-switch oligos (TSO) into sequenced products, can report on a number of biologically relevant processes. cDNA molecules containing this adapter were short enough to avoid fragmentation during Illumina library preparation, and are enriched for RNA molecules that were truncated at the time of cell lysis due to RNA degradation. After applying informatic filters for other likely sources of TSO-containing reads, we can annotate reads as likely endonucleolytic cleavage products and quantify this signal as a per-cell degradation score. By comparing with other common quality-control measures such as mitochondrial read fraction and transcriptome complexity, we observe that our degradation score is a unique measure of cell viability, and correlates strongly with gene expression signatures related to apoptosis and cell death in cells that otherwise appear healthy by virtue of their transcriptome alone. Exploring this new modality in published scRNA-seq datasets could reveal cell dynamics in complex tissues such as the immune-infiltrated tumor microenvironment or developing stem cell niche.

Separately, we have discovered that many of the TSO-containing reads re-annotated by our pipeline align precisely to the predicted 3'-fragment of pre-miRNA precursors that is discarded by the enzyme Drosha during miRNA biogenesis. Currently, there exists no robust means to simultaneously quantify mature miRNA abundance and global gene expression in droplet-based scRNA-seq platforms such as 10X Genomics. We provide functional validation that this signature is attributable to Drosha processing, thus providing a new method for indirect detection of mature miRNAs in high throughput scRNA-seq data. These tools have been developed into an open-source Python package called Clippings that can be retroactively applied to any raw sequencing data produced by the 10X Genomics platform for functional re-annotation of TSO-containing reads.