The Landscape of Long Non-Coding RNAs in Metastatic Prostate Cancer

Yashar S. Niknafs[1,2], Matthew K. Iyer[1], Arul M. Chinnaiyan[1,2,3,7]

[1]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, Michigan USA.
[2]Department of Pathology, University of Michigan, Ann Arbor, Michigan USA.
[3]Comprehensive Cancer Center, University of Michigan, Ann Arbor, Michigan USA.
[4]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, Michigan USA.

Introduction & Background: RNA sequencing data in cancers are being generated on a large scale at a rapid rate. *Ab initio* assembly of RNA-sequencing reads into transcript predictions provides an unbiased modality for predicting transcripts in unannotated regions of the genome, namely long non-coding RNAs (lncRNAs). lncRNAs have emerged as crucial aspects of human biology including cancer. Despite efforts to catalog lncRNAs in the human transcriptome, our current knowledge of cancer-associated lncRNAs remains infantile. These previous efforts were performed on a small number of cell lines or normal tissue samples, but because cancers possess highly heterogeneous gene expression patterns, detecting recurrent expression of subtype-specific lncRNAs will likely require analysis of large tumor cohorts. We hypothesized that large-scale transcriptome reconstruction from thousands of RNA-Seq datasets would reveal uncharacterized prostate cancer- and prostate tissue-associated lncRNAs and allow the scientific community to begin researching their biological relevance.

Methods & Design: We have downloaded, curated, and processed 23,623 RNA-seq samples largely from the TCGA, ICGC, GTEx, and CCLE, comprising 37 tissue types and over 35 cancer types. RNA-seq data processing was performed using STAR, Cufflinks, Kallisto, and TACO. The web tool for visualization and access to these data and analyses was built using a JavaScript-based front- and back-end server infrastructure (Noje.js) and a relational PostgreSQL database.

Results: The SSEA tool was used to nominate lncRNAs that are lineage specific and those that are differentially expressed in cancers versus normals. Hundreds of novel and previously annotated tissue-specific and cancer-specific lncRNAs were identified and nominated to be potentially implicated in conferring oncogenic phenotype in each of the 12 tissue cohorts for which we have matched normal samples. A number of novel lncRNAs were identified to be exquisitely specific to prostate tissue and prostate cancer, serving as prime potential biomarkers for aggressive prostate cancer.

Conclusions & Future Directions: Discovery of novel prostate tissue and cancer associated lncRNAs provides new and exciting targets that can be used as biomarkers and to further understand the disease mechanism for multiple cancer types. Further studies characterizing the biological function of the lncRNAs we have discovered will be necessary.

Conflicts of Interest: None to report.