The overall goal of this project is to develop bioinformatics methods and tools for exRNA sequencing (exRNA-seq), construct a reference map for exRNAs in various cancer cells, and identify cancer-related exRNAs that could be potentially used as biomarkers for cancer diagnosis and prognosis.

In our project, we will focus on five of the most common cancer types, including liver, colorectal, lung, stomach, and breast cancer. We will start by using several representative cancer cell lines for each cancer type and perform long and small exRNA-seq for different cellular components, including whole cell, microvesicles, exosomes, and ribonucleoproteins. We will develop visualization tools that employ dimension reduction techniques to display exRNA profiles for intuitive analysis. We will employ both established and new deconvolution techniques to infer sets of genes that occur in conjunction with one another and may correlate with phenotypic effects. We will integrate exRNA-seq data with a wide range of multi-omics data from other large consortia. Finally, for the exRNA biomarkers identified by our computational pipeline, we will design specific PCR primers, and validate the difference in expression levels between normal and cancer samples using quantitative reverse transcription PCR.

We have extensive experience in cancer genomics. We have played key roles in TCGA investigations into prostate [1] and kidney [2] cancers. We participated in TCGA Kidney Chromophobe Rencal Cell Carcinoma (RCC) project [3] and a subsequent pan-subtype kidney analysis. Our team analyzed the whole genome sequencing (WGS) data for the TCGA Kidney Papillary RCC (pRCC) [4]. In recent work, we leveraged our expertise in non-coding regions in the first whole-genome analysis of pRCC samples [2]. We have developed a tool to prioritize non-coding variants in cancer called FunSeq2 [5].

We have extensive experience developing RNA-seq processing pipelines [6] as part of the ENCODE and modENCODE consortia [7, 8]. These pipelines could work on RNA expression analysis, RNA variant calling, and RNA fusion detection techniques. We also have identified a set of disease-associated small exRNAs [9-11], and improved their functional annotations [12].

**References**

1. Cancer Genome Atlas Research N: **The Molecular Taxonomy of Primary Prostate Cancer.** *Cell* 2015, **163:**1011-1025.

2. Li S, Shuch BM, Gerstein MB: **Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations.** *PLoS Genet* 2017, **13:**e1006685.

3. Davis CF, Ricketts CJ, Wang M, Yang L, Cherniack AD, Shen H, Buhay C, Kang H, Kim SC, Fahey CC, et al: **The somatic genomic landscape of chromophobe renal cell carcinoma.** *Cancer Cell* 2014, **26:**319-330.

4. Cancer Genome Atlas Research N, Linehan WM, Spellman PT, Ricketts CJ, Creighton CJ, Fei SS, Davis C, Wheeler DA, Murray BA, Schmidt L, et al: **Comprehensive Molecular Characterization of Papillary Renal-Cell Carcinoma.** *N Engl J Med* 2016, **374:**135-145.

5. Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, Khurana E, Gerstein M: **FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer.** *Genome Biol* 2014, **15:**480.

6. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10:**57-63.

7. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, et al: **Landscape of transcription in human cells.** *Nature* 2012, **489:**101-108.

8. Gerstein MB, Rozowsky J, Yan KK, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, et al: **Comparative analysis of the transcriptome across distant species.** *Nature* 2014, **512:**445-448.

9. Shah R, Tanriverdi K, Levy D, Larson M, Gerstein M, Mick E, Rozowsky J, Kitchen R, Murthy V, Mikalev E, Freedman JE: **Discordant Expression of Circulating microRNA from Cellular and Extracellular Sources.** *PLoS One* 2016, **11:**e0153691.

10. Freedman JE, Gerstein M, Mick E, Rozowsky J, Levy D, Kitchen R, Das S, Shah R, Danielson K, Beaulieu L, et al: **Diverse human extracellular RNAs are widely detected in human plasma.** *Nat Commun* 2016, **7:**11106.

11. Mick E, Shah R, Tanriverdi K, Murthy V, Gerstein M, Rozowsky J, Kitchen R, Larson MG, Levy D, Freedman JE: **Stroke and Circulating Extracellular RNAs.** *Stroke* 2017, **48:**828-834.

12. Cheung KH, Keerthikumar S, Roncaglia P, Subramanian SL, Roth ME, Samuel M, Anand S, Gangoda L, Gould S, Alexander R, et al: **Extending gene ontology in the context of extracellular RNA and vesicle communication.** *J Biomed Semantics* 2016, **7:**19.