**Gerstein lab experience in cancer genomics**

The Gerstein lab has extensive experience with analyzing cancer genomes through our involvement in a number of cancer-focused consortia, including TCGA and PCAWG. In TCGA, we were involved in studies of prostate (Cancer Genome Atlas Research Network, 2015 ; Cancer Genome Atlas Research Network, 2016 ; Augspach et al, 2021) and kidney (Li et al, 2017) cancers, and some of this work involved detailed analyses into minor splicing in cancer contexts. The Gerstein has built software tools and developed computational methods for finding sites splice sites throughout the human genome. We have also co-lead a PCAWG sub-group to investigate the impact of non-coding mutations in cancer. We have continued to expand upon our widely-used Function-based Prioritization of Sequence Variants (FunSeq) tool to study somatic cancer variants (Khurana et al, 2013). We also developed FusionSeq (Sboner et al, 2010), which is a computational framework to identify fusion transcripts from paired-end RNA-sequence data. FusionSeq includes filters to remove spurious candidate fusions with artifacts such as misalignments or random pairings of transcript fragments, and it provides rankings for identified candidates. We also explored the properties and consequences of recurrent repeat expansions (rREs) spanning 29 cancer types (Erwin et al, 2022). We emphasize that many of these previous studies were carried out as part of joint efforts with the other groups, so our experience in these efforts is scientific and collaborative in nature.

**Gerstein lab experience in minor splicing analysis**

 The Gerstein lab has worked closely with the research groups led by Drs. Mark Rubin and Rahul Kanadia to study systematic differences in gene expression associated with minor-intron-containing genes (MIGs). This work was published in Molecular Cell. In the context of this study, we used silhouette scores to rigorously demonstrate that MIGs exhibit stronger differential gene expression across cancer types and stages of cancer development, relative to genes that do not contain minor introns (that is, relative to non-MIG genes).

**References**

Augspach, Anke, et al. "Minor intron splicing is critical for survival of lethal prostate cancer." Molecular cell 83.12 (2023): 1983-2002.

Balasubramanian, S. et al. Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. Nature communications 8, 382, doi:10.1038/s41467-017-00443-5 (2017).

Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163(4):1011-25. Epub 2015/11/07. doi: 10.1016/j.cell.2015.10.025. PubMed PMID: 26544944; PMCID: PMC4695400.

Cancer Genome Atlas Research Network. "Comprehensive molecular characterization of papillary renal-cell carcinoma." New England Journal of Medicine 374, no. 2 (2016): 135-145.

Chen, J. et al. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. Nature communications 7, 11101, doi:10.1038/ncomms11101 (2016).

Erwin, Graham S., Gamze Gürsoy, Rashid Al-Abri, Ashwini Suriyaprakash, Egor Dolzhenko, Kevin Zhu, Christian R. Hoerner et al. "Recurrent repeat expansions in human cancer genomes." Nature (2022): 1-7.

Khurana, Ekta, Yao Fu, Vincenza Colonna, Xinmeng Jasmine Mu, Hyun Min Kang, Tuuli Lappalainen, Andrea Sboner et al. "Integrative annotation of variants from 1092 humans: application to cancer genomics." Science 342, no. 6154 (2013): 1235587.

Li S, Shuch BM, Gerstein MB. Whole-genome analysis of papillary kidney cancer finds significant noncoding alterations. PLoS genetics. 2017;13(3):e1006685. Epub 2017/03/31. doi: 10.1371/journal.pgen.1006685. PubMed PMID: 28358873; PMCID: PMC5391127.

Rozowsky, J. et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. Molecular systems biology 7, 522, doi:10.1038/msb.2011.54 (2011).