**Gerstein lab experience in prioritizing genomic variants in cancer contexts**:

The Gerstein lab has developed software tools to prioritize cancer-associated genomic alterations, which can be combined to provide multiple lines of evidence for elucidating variant impacts in PCa. These tools include ALoFT, which predicts disease-causing potential of loss-of-function (LoF) events (Balasubramanian et al, 2017). ALoFT discriminates between LoF mutations that are deleterious in heterozygous states from those that may cause disease in the homozygous state. ALoFT enabled us to find that deleterious somatic variants that are enriched in canonical cancer driver genes. In another approach (Kumar et al, 2016), localized perturbations were examined to show how changes in molecular frustration can elucidate differential effects of variants in oncogenes and tumor suppressor genes (TSGs). We also developed SVFX, which uses a machine learning to quantify of SV pathogenicity (Kumar et al, 2020).

In addition to coding variants, our FunSeq tool (Khurana et al, 2013) prioritizes non-coding variants based on network connectivity and their disruptiveness by identifying deleterious variants in non-coding functional elements. FunSeq was used to integrate large-scale data from various resources, including cancer genomics data. By comparing patterns of inherited polymorphisms from 1,092 humans with somatic variants, FunSeq identified candidate non-coding cancer driver mutations. A further method to analyze non-coding regulatory regions, LARVA identifies significant mutation enrichment in non-coding elements (Lochovsky et al, 2015). In a pan-cancer analysis of variants in 760 cancer whole genomes, LARVA replicated established coding and non-coding cancer drivers.

We have played key roles in TCGA investigations into PCa (The Cancer Genome Atlas, 2015) as well as kidney cancer (Lee, 2016). We are part of the driver discovery subgroup in PCAWG to generate a catalog of driver elements in many cancer cohorts. Furthermore, this team led the PCAWG group to investigate the impact of non-coding mutations on cancer using the FunSeq pipeline on each variant in PCAWG. This approach also identified many high-impact mutations that can potentially influence cancer progression. Mining the variant dataset from the ICGC/TCGA PCAWG project identified genes and variants that operate as medium-impact putative passengers (Kumar et al, 2020b).

**References**:

Balasubramanian S, Fu Y, Pawashe M, McGillivray P, Jin M, Liu J, Karczewski KJ, MacArthur DG, Gerstein M. Using ALoFT to determine the impact of putative loss-of-function variants in protein-coding genes. Nat Commun. 2017;8:382. doi: 10.1038/s41467-017-00443-5

Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A, et al. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science. 2013;342:1235587. doi: 10.1126/science.1235587

Kumar S, Clarke D, Gerstein M. Localized structural frustration for evaluating the impact of sequence variants. Nucleic Acids Res. 2016;44:10062-10073. doi: 10.1093/nar/gkw927

Kumar S, Harmanci A, Vytheeswaran J, Gerstein MB. SVFX: a machine learning framework to quantify the pathogenicity of structural variants. Genome Biol. 2020;21:274. doi: 10.1186/s13059-020-02178-x

Kumar S, Warrell J, Li S, McGillivray PD, Meyerson W, Salichos L, Harmanci A, Martinez-Fundichely A, Chan CWY, Nielsen MM, et al. Passenger Mutations in More Than 2,500 Cancer Genomes: Overall Molecular Functional Impact and Consequences. Cell. 2020;180:915-927 e916. doi: 10.1016/j.cell.2020.01.032

Lee BH. Commentary on: "Comprehensive molecular characterization of papillary renal-cell carcinoma." Cancer Genome Atlas Research Network.: N Engl J Med. 2016 Jan 14;374(2):135-45. Urol Oncol. 2017;35:578-579. doi: 10.1016/j.urolonc.2017.07.022

Lochovsky L, Zhang J, Fu Y, Khurana E, Gerstein M. LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations. Nucleic Acids Res. 2015;43:8123-8134. doi: 10.1093/nar/gkv803

The Cancer Genome Atlas. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163:1011-1025. doi: 10.1016/j.cell.2015.10.025