Dr. Gerstein's team has made contributions to several human genome projects, including the Encyclopedia of DNA Elements (ENCODE) {PMID: 32728249}, model organism ENCODE (modENCODE) {PMID: 21177976}, the GENCODE gene annotation project {PMID: 30357393}, the EN-TEx {PMID: 37001506}, the Developmental Genotype-Tissue Expression(dGTEX), and Impact of Genomic Variation on Function (IGVF) Consortium. As a member of the IGVF consortium's Data Analysis Coordination Center, Dr.Gerstein’s team produced a comprehensive catalog of human genetic variation for the biomedical research community. Dr.Gerstein’s team has long-standing expertise in prioritizing protein-coding variants, non-coding variants, and structural variants at multiple levels.

**Dr.Gerstein’s team has much experience in studying the protein-coding variants**.

We were among the earliest to use statistical models to reconstruct PPI networks systematically by integrating high-throughput data and using networks to infer the direct and indirect effects of variants. We also developed an efficient computational framework that incorporates protein structure and dynamics to predict allosteric residues on surfaces and interiors. This enables high-throughput analysis of large protein datasets. We predict interior allosteric residues laying crucial communication pathways between protein subregions. These predicted allosteric residues tend to be evolutionarily conserved, as demonstrated by ExAC, HGMD, and ClinVar data. Notably, many disease-associated variants at predicted allosteric sites were previously poorly understood. Rare variants present challenges in evaluating potential deleteriousness using simple phenotype-genotype associations. We have expertise in measuring the impacts of rare and common variants in protein structures {PMID:27915290} leveraging metrics of localized structural frustration perturbations. This approach reveals biologically relevant insights when applied to the PDB. Specifically, disease-associated variants cause greater changes in localized frustration compared to non-disease variants. Additionally, rare SNVs more severely disrupt local interactions relative to common variants.

**Dr.Gerstein’s team has much experience in interpreting the non-coding variants.**

We have experience in estimating the effects of noncoding variants using data integration based on inter- and intra-species conservation, loss- and gain-of-function events for transcription-factor binding, enhancer-gene linkages and network centrality, and per-element recurrence across samples. Using data from multiple sources, we have developed a weighted scoring system to prioritize variants based on their predicted phenotypic effect. We also developed RADAR which extends FunSeq2 by combining RNA regulome and tissue-specific information. We have also built the GRAM pipeline, which is a generalized framework to predict the cellular expression-modulating effect of a noncoding variant by incorporating TF disruptive information, histone modification, and cell-specific expression and regulatory network information. GRAM can be extensively applied to fine-map causal variants within an LD-associated region.

Allele-specific (AS) variants are implicated in several diseases. Recently, we created the EN-TEx resource including the largest catalog of non-coding AS variants. We leveraged this catalog to build generalized models of variant impact. We developed a deep-learning transformer model that can predict the allele-specific activity based only on local nucleotide-sequence context, highlighting the importance of transcription-factor-binding motifs particularly sensitive to variants*.*

Previous work related to Genomic Privacy

We have developed a formalism for eQTL leakage from gene expression values and SV leakage from signal profiles of functional genomics data {\cite 26828419}. We showed that through various linking attack scenarios, we were able to connect individuals to databases of personal information. We also developed tools to deal with the leakage from signal tracks {\cite 29934598}. The most effective way to protect against a linking attack scenario is to ensure that deletion genotypes cannot be inferred from signal profiles. Deletions are a major source of leakage of genetic information from functional genomics signal profiles. We proposed solutions to the signal profiles where we can mask the sensitive information leakage while providing high utility. Our proposed solution systematically removes the dips in signal profiles to anonymize the profiles against the prediction of deletions. To remove these dips systematically, we used median filtering-based signal processing to locally smooth the signal profile around the deletion. This signal processing technique has been used to remove shot noise in two-dimensional imaging data and one-dimensional audio signals.