We have developed a number of computational tools to perform *in silico* functional studies of the validated TE candidates and their impact on cancers. We developed Variant Annotation Tool (VAT) to annotate protein sequence changes of mutations. VAT provides transcript-specific annotations and annotates mutations as synonymous, missense, nonsense or splice-site disrupting changes33. We used VAT to systematically survey loss-of-function (LoF) variants in a cohort of 185 healthy people as part of the Pilot Phase of the 1000GP34. We also participated in the 1000GP Phase 3 studies on LoF variants and the functional impact of SVs, and found that a typical genome contains ~150 LoF variants. Our ncVar pipeline further analyzes genetic variants across biotypes and subregions of ncRNAs, e.g. showing that miRNAs with more predicted targets show higher sensitivity to mutation in the human population35. We have extensive experience performing annotation of non-coding regulatory regions and expertise in developing tools to analyze ChIP-Seq data to identify genomic elements and interpret their regulatory potential. For ChIP-Seq, we developed two tools—PeakSeq and MUSIC—that identify regions bound by transcription factors and chemically modified histones36,37. PeakSeq has been widely used in consortium projects such as ENCODE36,22. MUSIC is a newly developed tool that uses multiscale decomposition to help identify enriched regions in cases where strict peaks are not apparent.

A powerful way to integrate diverse genomic data is through networks representations. We have great experience studying regulatory network and relating variants to networks. In particular, we have integrated multiple biological networks to investigated gene functions. We found that functionally significant and highly conserved genes tend to be more central in various networks38. Furthermore, we have extensive experience with allelic activity analysis. We recently applied this pipeline on a population scale to RNA-Seq data from the 1000GP, and used this analysis to create AlleleDB, a database of genomic regions with high allelic activity39.