**A large scale integrative resource from ENCODE for cancer research**

Cancer is caused by mutations in the DNA which disrupt the normal physiology of cells. While mutations on coding genes have been well characterized, the preponderance of mutations in tumors occur in non-coding regions and are still poorly understood. The new release of ENCODE data provides an opportunity to bridge these knowledge gaps. For a variety of cancer-derived cell lines, as well as non-cancerous cell lines derived from relevant tissues, ENCODE provides diverse genome-wide assays, such as Repli-seq, ChIP-seq, DNase-seq, STARR-seq, Hi-C, and ChIA-PET. The resulting data and functional maps of the human genome provide a framework to assess the potential for non-coding mutations to dysregulate genes.

In this paper, we first integrated diverse assays from ENCODE to define high-confidence regulatory elements and their gene linkage to define the extended gene neighborhood. We also developed a regression based method for background mutation rate calibration. It removes confounding effects from chromatin and replication timing and search for genes with higher than expected mutation frequency in the extended gene neighborhood. This approach successfully identified novel highly mutated genes, such as *BCL6* in leukemia, that are associated with patient prognosis.

Besides, we also integrated extensive binding profiles from ENCODE to build up tissue specific regulatory networks for both transcription factors (TFs) and RNA-binding proteins (RBPs). Intriguingly, through networks hierarchy analysis we found that TFs with higher mutation burden tend to be located at the bottom of the hierarchy (e.g., EZH2 and NR2C2), whereas those with dysregulated expression tend to reside at the top. Furthermore, by comparing tumor and normal network, we identified highly “rewired” TFs with changed targets and prognostic value, such as IKZF1 and MYC. We then extended tissue specific network to build up generalized networks across cancers. After combining with expression profiles from other cohorts, we pinpointed MYC and SUB1 as key regulators that significantly drive tumor to normal differential expression and then validated their effects through knockdown experiments.

Finally, we proposed a prioritization scheme for key mutations in cancer. We identified active enhancers and seven high impact mutations therein in breast cancer and validated their functional effects through luciferase assays.