LARVA/MUSIC

We have worked on statistical methods and software specifically for the analysis of noncoding regulatory regions. We have developed a computational tool called LARVA (Large-scale Analysis of Recurrent Variants in noncoding Annotations), which identifies significant mutation enrichments in noncoding elements by comparing observed mutation counts to the expected counts under a whole genome background mutation model. This model uses a beta-binomial distribution to represent the spread of variants, which better handles the large heterogeneity in mutation rate seen throughout the genome compared to previously used binomial models. LARVA also includes corrections for biases in the mutation rate owing to DNA replication timing. We have also developed MUSIC, which identifies regions enriched in ChIP-seq signal using an advanced statistical framework.

We plan to use LARVA to systematically analyze our cancer whole genome data. We will use a sliding window along the whole genome, and look for variant enrichments in each of these windows. This methodology has the potential to discover known functional elements involved in cancer, and implicate previously unannotated regions in cancer as well. These findings will enable the prioritization of annotations and regions for followup investigation into involvement in cancer driving processes.