World Academy of Science, Engineering and Technology
International Journal of Biotechnology and Bioengineering
Vol:14, No:08, 2020

# Systematic Identification of Noncoding Cancer Driver Somatic Mutations

**Authors :** Zohar Manber, Ran Elkon

**Abstract :** Accumulation of somatic mutations (SMs) in the genome is a major driving force of cancer development. Most SMs in the tumor's genome are functionally neutral; however, some cause damage to critical processes and provide the tumor with a selective growth advantage (termed cancer driver mutations). Current research on functional significance of SMs is mainly focused on finding alterations in protein coding sequences. However, the exome comprises only 3% of the human genome, and thus, SMs in the noncoding genome significantly outnumber those that map to protein-coding regions. Although our understanding of noncoding driver SMs is very rudimentary, it is likely that disruption of regulatory elements in the genome is an important, yet largely underexplored mechanism by which somatic mutations contribute to cancer development. The expression of most human genes is controlled by multiple enhancers, and therefore, it is conceivable that regulatory SMs are distributed across different enhancers of the same target gene. Yet, to date, most statistical searches for regulatory SMs have considered each regulatory element individually, which may reduce statistical power. The first challenge in considering the cumulative activity of all the enhancers of a gene as a single unit is to map enhancers to their target promoters. Such mapping defines for each gene its set of regulating enhancers (termed "set of regulatory elements" (SRE)). Considering multiple enhancers of each gene as one unit holds great promise for enhancing the identification of driver regulatory SMs. However, the success of this approach is greatly dependent on the availability of comprehensive and accurate enhancer-promoter (E-P) maps. To date, the discovery of driver regulatory SMs has been hindered by insufficient sample sizes and statistical analyses that often considered each regulatory element separately. In this study, we analyzed more than 2,500 whole-genome sequence (WGS) samples provided by The Cancer Genome Atlas (TCGA) and The International Cancer Genome Consortium (ICGC) in order to identify such driver regulatory SMs. Our analyses took into account the combinatorial aspect of gene regulation by considering all the enhancers that control the same target gene as one unit, based on E-P maps from three genomics resources. The identification of candidate driver noncoding SMs is based on their recurrence. We searched for SREs of genes that are "hotspots" for SMs (that is, they accumulate SMs at a significantly elevated rate). To test the statistical significance of recurrence of SMs within a gene's SRE, we used both global and local background mutation rates. Using this approach, we detected - in seven different cancer types - numerous "hotspots" for SMs. To support the functional significance of these recurrent noncoding SMs, we further examined their association with the expression level of their target gene (using gene expression data provided by the ICGC and TCGA for samples that were also analyzed by WGS).

**Keywords :** cancer genomics, enhancers, noncoding genome, regulatory elements
**Conference Title :** ICSBB 2020 : International Conference on Systems Biology and Biotechnology
**Conference Location :** Paris, France
**Conference Dates :** August 27-28, 2020