Molecular Pathology Selected Abstracts, 2/16

Editors: Donna E. Hansel, MD, PhD, chief, Division of Anatomic Pathology, and professor, Department of Pathology, University of California, San Diego; John A. Thorson, MD, PhD, associate professor of pathology, director of the Clinical Genomics Laboratory, Center for Advanced Laboratory Medicine, UCSD; Sarah S. Murray, PhD, professor, Department of Pathology, and director of genomic technologies, Center for Advanced Laboratory Medicine, UCSD; and James Solomon, MD, PhD, resident, Department of Pathology, UCSD.

Identifying significantly mutated regions across cancer types

The clinical application of tumor genome profiling to aid in determining prognosis and in selecting the most effective therapies is dependent on fundamental knowledge of the identity and nature of significant driver mutations. In most cases, the current approach is to focus only on changes in the DNA sequence that subsequently lead to a significant change in the encoded protein—that is, an altered amino acid or alternatively spliced messenger RNA. Changes in regulatory regions or untranslated regions of DNA are largely ignored. The authors conducted a study that provides evidence that this approach likely leads to a significant underestimation of the number and type of mutations underlying many cancers. They examined approximately 3 million previously identified somatic, single-nucleotide variants (SNVs) from 4,735 tumors of 21 cancer types. They noted that 79 percent of the somatic mutations were found to neither alter protein coding sequences nor lead to splicing changes and, therefore, were not previously considered for further evaluation as potentially important driver mutations. The authors re-evaluated this same data for significant variants using an annotation-independent, density-based clustering algorithm. They were able to identify regions of the genome (coding as well as exon-proximal noncoding) that display a high density of mutations, referred to as significantly mutated regions (SMRs). A subset of 872 of these SMRs was present in more than two percent of patients, representing 20 cancer types, and were, therefore, selected for further study. The validity of this approach for identifying functionally important variants is demonstrated by the fact that SMRs with higher density clustering scores correlate with enrichment for known somatic-mutation driven cancer-associated genes, as found in the Cancer Genome Census database. In addition, although many SMRs identified within protein-coding regions of the genome were consistently observed in a variety of cancer types, the authors also uncovered examples of coding region SMRs that are very specific to a particular type of cancer, exemplified by differences observed between breast and ovarian cancer in PIK3CA variants. Interestingly, a significant portion (31 percent) of identified SMRs are not predicted to alter encoded proteins, suggesting that a large number of unrecognized pathologic variants may reside in noncoding regions of the genome, including promoter regions and 5’ untranslated regions, as well as splice sites and transcription factor binding sites. Overall, SMR analysis is shown to not only correlate well with current knowledge of significant variants but also highlight additional areas and types of variation likely to be of importance for understanding and interpreting genomic sequencing data in the context of cancer. Although identification of recurrently mutated sites is only a first step in understanding the varied mechanisms of cancer initiation and progression, this study presents compelling evidence that the current approach to interpreting genomic profiling data is not fully assessing significant driver mutations. Expanding these types of studies to encompass the entire genome will likely provide evidence that additional significant cancer-related variants exist.

Araya CL, Cenik C, Reuter JA, et al. Identification of significantly mutated regions across cancer types highlights a rich landscape of functional molecular alterations [published online ahead of print December 21, 2015]. *Nat Genet.* doi:10.1038/ng.3471.

Correspondence: Carlos L. Araya at claraya@stanford.edu