

# The assessment of the impact of small deletions within human protein domains using transcriptome data: a pilot study in lung cancer

Fernanda Cristina Medeiros de Oliveira<sup>1</sup>, Gabriel Wajnberg<sup>2</sup>, Fabio Passetti<sup>3</sup>,

*1 FIOCRUZ-IOC*

*2 Fiocruz-IOC*

*3 FIOCRUZ - IOC*

## Abstract

Deletions are examples of polymorphisms that can alter the protein sequence encoded by genes. These changes within the amino acid sequence can be connected with numerous human pathologies, such as cancer. Lung cancer has the highest worldwide incidence of all tumors, with an increase of 2% per year. For this reason, there is an interest to find new methods for diagnosis and treatments. High-throughput sequencing generates a large amount of genome or transcriptome data in a short time when compared to other sequencing methodologies. High-throughput sequencing can be used to identify new polymorphisms, such as deletions in the human genome. In this study, we used RNA-Seq data from six lung adenocarcinoma patients available at the Sequence Read Archive database (study ID SRP012656). We identified 2,388 protein domains affected: 1,137 in control tissue adjacent to the tumor, 729 in tumor samples and 522 in both control and tumor samples. We identified deletions in protein domains with high probability to be associated with cancer biology, such as deletions found in the genes SASH1, GRINA, TP53BP2, RAVR1 and NCOR2, which share the same protein domain termed large tegument protein UL36. Changes in this domain may be decisive to the development of lung cancer, because SASH1 plays an important role as a tumor suppressor in lung cancer. We also identified different altered domains in the TP53BP2 gene, such as ankyrin repeats Ank and Ank 2. The p53BP2 protein binds to the p53 tumor suppressor by these ankyrin repeats to increase its DNA binding activity. We identified, through this work, changes in amino acid sequences caused by small genomic deletions up to 100 nucleotides in length that affect protein domains of proteins previously associated with lung cancer. These new findings may be useful for new studies related to the identification of new biomarkers for diagnosis and new therapeutic targets.

Funding: CAPES, FIOCRUZ, FAPERJ, PIBITI/CNPq