The discovery of novel multiple small deletions within human coding genes associated to known lung cancer pathways

Gabriel Wajnberg[1,2], Raphael Tavares da Silva[1], Nicole de Miranda Scherer[3], Carlos Gil Ferreira[4] e Fabio Passetti[1,2]

1. Laboratory of Functional Genomics and Bioinformatics, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil., 2. Systems and Computational Biology Graduate Program, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil., 3. Laboratório de Bioinformática e Biologia Computacional, Instituto Nacional de Câncer (INCA), Rio de Janeiro, Brazil., 4. Clinical Research Coordination, Instituto Nacional de Câncer (INCA), Rio de Janeiro, Brazil.

Deletions occur naturally in the coding genes of healthy human populations, which, in turn, may affect the protein product sequence. Although the identification of deletions is usually performed using DNA data, the use of transcriptome data is a promising strategy. In addition, the 1000 genomes project can be used as source of data to search for deletions in different healthy human populations. We developed an innovative method to identify small deletions using human transcriptome data. Here, we present the detection and analysis of small deletions in 22 matched tumor and normal lung cancer RNA-Seq data using the human genome GRCh37/hg19 as reference. Using this strategy, we identified 1,778 small deletions using the assembled transcripts in these lung cancer samples. We confirmed 928 (52%) of these on a selected list 66 genome sequences comprising 3 from each of the 22 populations available in the 1000 genomes project. We identified 398 (22%) from the 928 set of small deletions predicted to change the known reading frame in 194 different genes. If considered deletions detected in both matched normal and tumoral samples, 53 (13%) were identified. We identified three altered pathways with significant posterior probability to be overrepresented performing a Gene Set Enrichment Analysis (GSEA): positive regulation of cell proliferation, positive regulation of Notch signaling pathway, and negative regulation of apoptotic process. These findings are supported by other studies. For example, the positive regulation of cell proliferation path (GO:0008284) has been already been previously identified as overrepresented in using lung cancer samples. The *TNFSF13B* is described to influence this pathway and it has been already identified with high expression in lung cancer. We identified a 88 nucleotide small deletion in the *TNFSF13B* in 8 tumor samples from different patients, providing a strong support of our findings. According to our analysis, this deletion affects the main TNF domain encoded by the canonical protein. In conclusion, here we describe the identification of a group of small deletions, which may contribute the reduced control of previously described pathways associated to lung cancer, improving the knowledge of the lung cancer biology. Financial support: INCA/MS, FIOCRUZ, CAPES, Fundação do Câncer, FAPERJ and CNPq.