The assessment of the impact of small deletions within human protein domains using transcriptome data: a preliminary analysis.

Fernanda Oliveira, Gabriel Wajnberg, Fabio Passetti

Laboratory of Functional Genomics and Bioinformatics, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil.

Deletions is an example of sequence polymorphism, which can alter the encoded protein sequence. These alterations within the amino acid sequence can be associated to many diseases in human, such as cancer. High throughput sequencing data can be used to identify novel small polymorphisms, such as deletions. This work aims the analysis of the impact of deletions within protein coding domains associated with tumors, with the use of transcriptome data generated by RNA-Seq. Our preliminary analysis identified coding protein domains affected by small deletions, up to 100 nucleotides in length, in RNA-Seq data available in the public database SRA from matched normal and tumor samples in six lung cancer patients. We identified 734 unique affected protein domains: 123 only in normal samples, 71 exclusively to tumor samples and 540 found affected in both normal and tumor samples. Deletions were detected in protein domains with high probability of being associated with cancer biology, such as deletions occurring in protein tyrosine kinase domains in FLT4 and HTR3A genes in both normal and tumor samples. This type of protein domain is frequently affected by mutations in cancer, including lung cancer. We also identified the zinc finger protein domain C2H2 affected in five different genes only in tumor samples, such as: H1NFP, RBAK, ZNF468, ZNF234 and ZNF571. These genes have different functions, but all of them are associated in the regulation of transcription interacting with transcription factors and other components such as E2F1 transcription factor (RBAK gene function). In conclusion, our preliminary results shows that these small deletions detected using transcriptome data of 6 lung cancer patients may be disrupting some important protein domains associated with cancer biology and we can contribute to identify novel cancer genetic markers with this data.

Financial Support: CAPES, FIOCRUZ, FAPERJ and CNPq.