

# Human Retrocopies and Genetic Expression in Tumor and Normal Tissues

Helena Beatriz da Conceicao, Gabriela Guardia, PEDRO A F GALANTE

*Instituto de Ensino e Pesquisa - Hospital Sírio Libanês*

## Abstract

Retrocopies are copies of messenger RNAs reverse transcribed into the genome. Since this duplication process occurs from mature messenger RNAs (without introns), retrocopies are characterized by the conservation of only their parental exons. This characteristic has been used to identify retrocopies since 1980. Also, as a consequence of the lack of promoter regions to allow their expression/transcription, the first works describing retrocopies classified them as "dead on arrival" (i.e., never transcribed). However, nowadays there are a few known mechanisms described in the literature that allow retrocopies to gain expression, such as the obtention of regulatory sequences from neighborhood elements, distant CpG islands or even from cis-regulatory regions. In addition to the label of retropseudogenes, an unexpected number of functional retrocopies have been identified due to the increasing practice of complete genome sequencing and multidisciplinary approaches. Thanks to them, we know nowadays that the mechanism of retrocopy generation has a major role in the genesis of genes in humans and other species. However, this role is limited to a few examples and it is rare the studies that systematically analyze the functionality of retrocopies. For example, in humans, it has been identified approximately 8000 retrocopies, of which only a minority (less than 10%) presents evidence of transcription and functionality. Two important examples are PTENP1 (a retrocopy of the tumor suppressor PTEN) and BRAFP1 (a retrocopy of the proto-oncogene BRAF). In this context, previous works have hypothesized that retrotransposition can be a possible mechanism of neoplasia. It is known that genomic instability can promote retrotransposition activity, with reported cases of pseudogene insertions in tumors. Since many functional retrocopies are still poorly characterized, our objective is to investigate retrocopies with expression in humans, in order to give a global vision of retrogenes contribution to normal and cancer cells phenotype. To explore the possible implications of transcribed retrocopies in the human genome, we used a list of retrocopies detected by an in-house pipeline. Based on this list, we performed an analysis of gene expression using RNA-Seq data from healthy (GTEx) and tumoral (TCGA) datasets. In our results, we gathered general characteristics of retrocopies and their parental genes, such as the number of retrocopies per parental gene and how the retrocopies are distributed in the genome. In healthy tissues, we found that 2/3 of retrocopies show distinct expression patterns among tissues, suggesting some kind of expression regulation. For example, testis has the highest number of expressed retrocopies (4067), while whole blood has only 1767 expressed retrocopies, being the tissue with the lowest number. In tumor samples, we found that 75% of retrocopies are transcribed, with some interesting cases such as three aggressive tumors (STAD, GBM, and AB) having the highest number of transcribed retrocopies (5387, 3853, 4775, respectively), suggesting the possibility of using some of these retrocopies as tumor markers. In the end, our results revealed patterns in terms of retrocopies expression in both normal and cancer that will be further explored, including the investigation of features associated with the selection of retrocopies and the incorporation of Ribo-Seq data to our analyses.

Funding: FAPESP