Investigating the Role of Transcribed Pseudogenes in Breast Cancer

Joshua Welch[1], Jan Prins[1]

[1]Department of Computer Science, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract

Pseudogenes are genomic sequences closely resembling genes but possessing sequence differences that prevent them from encoding functional proteins. Although the human genome contains thousands of pseudogenes, these sequences are generally disregarded in functional genomic studies and are widely viewed as non-functional. However, there is increasing evidence that some pseudogenes are actually transcribed into RNA and can contribute to cancer when dysregulated. In particular, pseudogene transcripts can sequester miRNAs that would otherwise target mRNAs. In this role pseudogenes function as competing endogenous RNA (ceRNA).

To investigate the hypothesis that transcribed pseudogenes play a role in cancer, we developed a bioinformatics method for studying pseudogene transcription using RNA-seq and applied this method to 820 breast cancer samples from The Cancer Genome Atlas project. We incorporated sample-paired gene and miRNA expression data and miRNA target prediction to assess the potential ceRNA function of transcribed pseudogenes. We also performed a clustering analysis using the pseudogene expression data, determining how variation in pseudogene expression relates to known breast cancer subtypes.

Our results indicate that many pseudogenes are transcribed in breast cancer. A subset of these exhibit significant differential expression between tumor and normal samples. The expression levels of the differentially expressed pseudogenes correlate with a number of known cancer-related genes. Furthermore, our analysis incorporating miRNA target prediction and miRNA expression data suggests that a number of transcribed pseudogenes are strong candidates for ceRNA function. Taken together, these results indicate that pseudogene transcription in cancer plays a larger role than previously appreciated.