doi 10.18699/bgrs2024-2.1-25

Retrotransposons expression controls: hubs and networks

Yuditskiy K.I., Samsonova A.A., Kanapin A.A.*
Institute of Translational Biomedicine, Saint Petersburg State University, St. Petersburg, Russia
** akanapin@spbu.ru*

Key words: retrotransposons; transcriptional regulation; gene networks; epigenetics

Motivation and Aim: Over half of the human genome consists of interspersed repeats that result from copy-and-paste events carried out by retrotransposons [1]. These elements are typically kept in check through epigenetic silencing to prevent genome instability. In cancer, changes in the epigenetic landscape can lead to abnormal expression of these genetic elements, which can act as internal mutagens and contribute to clonal evolution and cancer development [2]. Our specific goal is to uncover new regulatory mechanisms, networks, and potentially genomic elements that control the transcriptional activity of retroelements in both cancer and normal cells. By conducting a thorough analysis of high-throughput transcriptional datasets and developing new computational frameworks, we hope to identify potential regulatory mechanisms that control retrotransposon expression in both normal and cancer cells, bridging the gap between seemingly unrelated observations of epigenetic target inhibition and retrotransposon expression.
Methods and Algorithms: The experiments relevant to this study were chosen from a literature search of published RNASeq experiments involving HDAC inhibitors. The publicly available transcriptomic data and corresponding metadata were downloaded from the SRA (Short Read Archive) repository and the Entrez database, respectively. A complete list of data accessions is available in the project database: https://transrt.compbio.ru/db/. The SalmonTE package [3] was used to assess the expression levels of both genes and transposons, which included a total of 688 transposons across 14 different types. The normalized gene and transposon expression data were then used to predict gene regulatory networks using the WGCNA package [4]. Consensus networks were constructed to identify control modules that remained consistent across experiments, and adjacency matrices were created for each experiment using the WGCNA::adjacency method. The obtained minimum and median adjacency matrices were used to calculate the Topological Overlap Matrix (TOM). Subsequently, the dendrogram of co-expression modules was obtained using the TOMs, and the modules were extracted by applying the dynamicTreeCut::cutreeDynamic algorithm to the dendrogram. Additionally, the ARACHNe-AP package [5] was used to identify transcription factors that control retrotransposon expression.
Results: We have successfully constructed distinct gene networks for various experiments and have identified co-expression modules that align with each other. Consensus modules, which remain consistent across cell types, have been determined for two retrotransposon families: LINE1 and HERV. The modules are mainly composed of long non-coding RNAs, making it difficult to perform standard gene set enrichment analyses and functional annotation.
Using ARACHNe-AP algorithms, we conducted a systematic search for transcription factors that control transposon activity. This allowed us to identify specific transcription

factor-transposon pairs and generate lists of factors that exclusively regulate certain classes of transposon elements, as well as general regulators. Based on the results from WGCNA, which revealed two distinct regulatory modules for LINE and ERV transposons, we carried out an enrichment analysis of Reactome and KEGG signaling pathways associated with the transcription factors that exclusively regulate these transposon classes. Notably, we observed a significant difference in the enriched transcription factors for LINE and HERV transposons.

Conclusions: Our findings offer fresh insights into the intricate regulatory networks governing the expression of retrotransposons in both normal and cancer cells. It was observed that transposons from at least two classes, namely LINE and HERV are subject to regulation through distinct mechanisms. Long non-coding RNAs were found to play a significant role in the regulation of both types of transposons. Furthermore, the distinct sets of transcriptional regulators controlling the expression of different mobile elements suggest differing mechanisms governing retrotransposon functions in normal and pathological cells.

Funding: The study is supported by Russian Science Foundation (Grant No. 23-14-00134).

References

1. de Koning A.P., Gu W., Casto T.A., Batze M A., Polloc D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 2011;7(12):e1002384. doi 10.1371/journal.pgen.1002384
2. Roulois D., Loo Yau H., Singhania R., Wang Y., Danesh A., Shen S.Y., Han H., Liang G., Jones P.A., Pugh T.J., O'Brien C., De Carvalho D.D. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell.* 2015;162(5):961-973. doi 10.1016/j.cell.2015.07.056
3. Jeong H.H., Yalamanchili H.K., Guo C., Shulman J.M., Liu Z. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pac Symp Biocomput.* 2018;23:168-179
4. Langfelder P., Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559. doi 10.1186/1471-2105-9-559
5. Lachmann A., Giorg, F.M., Lopez G., Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics*. 2016;32(14):2233-2235. doi 10.1093/bioinformatics/btw216