
Storming TE-Storm: implications for short RNA-seq feasibility in transposon research exemplified by ALS/FTLD (re)analysis

Natalia Savytska*¹ and Vikas Bansal¹

¹Biomedical Data Science, Bansal AG, DZNE (German Center for Neurodegenerative Diseases),
Tübingen – Otfried-Müller-Str. 23 72076 Tübingen, Germany, Germany

Abstract

Transposon transcriptional activity has been a suspected contributor to ALS/FTLD spectrum disorder since the 2010s (PMID:35415778), leading to the proposal of retrotransposon storm hypothesis of neurodegenerative diseases (PMID:29705598). TDP-43 and C9Orf72 pathologies were hypothesised as associated to TE overexpression (PMID: 22957047, 28637276). A number of studies failed to replicate these findings, while others reported discrepant TE subfamilies as hyperactivated (PMID:35415778). Most recent studies relied on short read RNA-seq and focused on the TE activity at the subfamily level, using inconsistent analysis methods. The latter may have contributed to the discrepancies between the results. Subfamily level analysis remains mainstream, as it theoretically overcomes mapping uncertainty for individual TE copies, however concerns were raised in the community (PMID:32576954 and 36338986).

We aimed at evaluating feasibility of TE analysis at the subfamily level (using synthetic and cell culture derived datasets) in RNA-seq datasets and to systematically analyse previously published and a new in-house human ALS/FTLD datasets to test the "storm" hypothesis and its association to TDP43 or C9Orf72 pathology.

Our simulation and investigation into iPSC transcriptomes demonstrated a high false positive rate for subfamilies analysis, which was driven by both mismapping between related but distinct subfamilies, and exaptation of TEs within longer transcripts, such as LINC00665. Our results suggest TE analysis at the subfamily level is unadvisable and may ultimately lead to the misinterpretation of transcriptomic changes.

Further analysis of ALS/FTLD datasets using a unified pipeline showed no consistent profile changes between different datasets (incl. for the previously suspected elements, e.g. HERVK), and no global transcriptional upregulation of TEs across datasets (equal numbers of up- and down-regulated loci in disease). We find further support for alternative gene isoforms as an explanation that leads to misidentification of TE activity change, illustrated through the STMN2 gene.

Keywords: neurodegeneration, neurodegenerative disease, ALS, FTLD, human disease, human transposons, human, RNA, seq, expression, transposon transcriptomics, transcriptomics

*Speaker