Transcriptomic analysis of human endogenous retroviruses in systemic lupus erythematosus

Luis P. Liniguez, Miguel de Mulder Rougvie, Nathaniel Stearrett, Richard B. Jones, Christopher E. Ormsby, Gustavo Reyes-Terán, Keith A. Crandal, Douglas F. Nixon, and Matthew L. Bendall

Endogenous retroviruses (ERVs) are integrated retroviral elements within the human genome. Tokuyama et al. (1) recently published a computational tool, “ERVmap,” to analyze genome-wide, locus-specific expression of human ERVs. The authors found increased expression of 124 ERV loci in patients with systemic lupus erythematosus (SLE), compared to controls, and 0 down-regulated loci. In contrast, our re-analysis of their data using a Bayesian reassignment algorithm, Telescope (2), detected only 23 ERV locations with significant differential expression (DE), including 4 loci with significantly lower expression. We found that the differences between the results could be due to methodological aspects of their analysis, including alignment ambiguity, ERV annotation, and failure to account for sequencing platform as a source of variance (3).

ERVmap does not adequately address the problem of alignment ambiguity, which occurs when sequencing reads align to multiple distinct genomic locations, a major challenge for quantifying expression of repetitive elements (4–7). The ERVmap pipeline aligns reads to the reference genome, allowing for multiple alignments per read, and then applies “very stringent filtering criteria to the mapped reads” (1, p. 12566). These heuristics effectively eliminate suboptimal alignments to the reference genome, but tend to discard data instead of identifying a location for each read. In contrast, Telescope uses a generative model of RNA-seq to reassign ambiguously mapped reads. The choice of annotation is important area for continued research.

In conclusion, ERVmap makes an important attempt to address the challenge of locus-specific ERV quantification but falls short in several ways. However, we agree that the role of ERV expression in SLE is an important area for continued research.

Acknowledgments
Funding, in part, is from NIH Grants AI076059 and CA206488.
Fig. 1. Differential expression analysis. (A) Principal component analysis (PCA) of all genes and ERVs with normalized counts colored based on platform. The number of samples per platform is shown under the PCA plot. (B) Volcano plot depicting differentially expressed ERVs measured by Telescope. Red points represent differentially expressed ERVs. (C) Heatmap of differentially expressed ERVs. Red, higher expression; blue, lower expression. Normalized ERV counts were scaled to Z scores and the Euclidean distances calculated for hierarchical clustering. The complete linkage method was used for clustering. Highlighted loci (†) were also present in ERVmap annotation and are shown in D. (D) Overlapped differentially expressed ERVs between Telescope and ERVmap annotations. Asterisks indicate significantly differentially expressed ERV loci (padj < 0.05, log2 fold-change > 1.0) identified by each analysis. Left shows the normalized Z-score scaled values of Telescope. Right shows the matched annotations in ERVmap. Normalized Z-score scaled values are also illustrated.

1 M. Tokuyama et al., ERVmap analysis reveals genome-wide transcription of human endogenous retroviruses. Proc. Natl. Acad. Sci. U.S.A. 115, 12565–12572 (2018).
2 M. L. Bendall et al., Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. bioRxiv:398172 (2018).
3 A. Conesa et al., A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13 (2016).
4 T. J. Treangen, S. L. Salzberg, Repetitive DNA and next-generation sequencing: Computational challenges and solutions. Nat. Rev. Genet. 13, 36–46 (2011).
5 B. Li, V. Ruotti, R. M. Stewart, J. A. Thomson, C. N. Dewey, RNA-Seq gene expression estimation with read mapping uncertainty. Bioinformatics 26, 493–500 (2010).
6 S. W. Criscione, Y. Zhang, W. Thompson, J. M. Sedivy, N. Neretti, Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC Genomics 15, 583 (2014).
7 Y. Jin, O. H. Tam, E. Paniagua, M. Hammell, TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics 31, 3593–3599 (2015).
8 S. Zhao, B. Zhang, A comprehensive evaluation of ensemble, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. BMC Genomics 16, 97 (2015).
9 M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).