Gene expression

**ggtranscript**: an R package for the visualization and interpretation of transcript isoforms using *ggplot2*

Emil K. Gustavsson 1,2,†, David Zhang 1,2,*,†, Regina H. Reynolds 1,2, Sonia Garcia-Ruiz 1,2,3 and Mina Ryten 1,2,3

1Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London WC1N 1EH, UK, 2Aligning Science Across Parkinson’s (ASAP) Collaborative Research Network, Chevy Chase, MD 20815, USA and 3NIHR Great Ormond Street Hospital Biomedical Research Centre, University College London, London WC1N 1EH, UK

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

*To whom correspondence should be addressed.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: Anthony Mathelier

Received on March 31, 2022; revised on June 6, 2022; editorial decision on June 7, 2022; accepted on June 22, 2022

Abstract

**Motivation**: The advent of long-read sequencing technologies has increased demand for the visualization and interpretation of transcripts. However, tools that perform such visualizations remain inflexible and lack the ability to easily identify differences between transcript structures. Here, we introduce *ggtranscript*, an R package that provides a fast and flexible method to visualize and compare transcripts. As a *ggplot2* extension, *ggtranscript* inherits the functionality and familiarity of *ggplot2* making it easy to use.

**Availability and implementation**: *ggtranscript* is an R package available at https://github.com/dzhang32/ggtranscript (DOI: https://doi.org/10.5281/zenodo.6374061) via an open-source MIT licence. Further documentation is available at https://dzhang32.github.io/ggtranscript/.

**Contact**: david.zhang.12@ucl.ac.uk

1 Introduction

Alternative splicing is a crucial post-transcriptional step through which introns are excised from messenger RNA (mRNA) precursors, and exons are spliced together to form mature mRNA isoforms. In fact, ~95% of human genes undergo alternative splicing resulting in various forms of mature mRNA (Wang et al., 2008). This process is often regulated in a tissue-specific, disease-specific or developmental manner, resulting in multiple different transcripts being generated from the same gene.

It is well-recognized that it is challenging to identify full-length transcript structures from standard transcriptomic assays relying on short-read RNA-sequencing, as short-reads rarely span multiple splice junctions and therefore make it difficult to infer transcript structures (Conesa et al., 2016). However, long-read sequencing platforms such as PacBio and Oxford Nanopore have transformed the field and enabled the discovery of new transcript isoforms that could not have been recognized by the assembly of short-reads. In addition, long reads facilitate better transcript quantifications and improve mapping of highly homologous sequences.

Current tools to visualize transcript structures are often inflexible, allowing users very limited control over the outputted plot aesthetics or lack the ability to compare transcript structures. For example, UCSC genome browser (Kent et al., 2002), IGV Browser (Robinson et al., 2011) and Geiz (Hahne and Ivanek, 2016) are genome-based tracks that allow for visualization of transcripts, but are not accessible programmatically. IsoformSwitchAnalyzeR (Vitting-Seerup et al., 2019), wiggleplotr and ggashimi (Li et al., 2018) offers limited customization of plot aesthetics and comparisons of transcript structures. SWAN (Reese and Mortazavi, 2021) does offer some customizable transcript visualization functions, but has limited functionality to highlight differences and is within the python framework.

Here, we introduce the R package *ggtranscript* which makes it easy to both visualize and compare transcript structures using ggplot2 (Wickham, 2016), a popular R-based framework for data visualization based upon an intuitive grammar system that permits flexibility via combination of independent components. As a ggplot2 extension, *ggtranscript* inherits a vast amount of flexibility when determining the plot aesthetics, as well as interoperability with existing ggplot2 geoms and ggplot2 extensions. Furthermore, the input data for *ggtranscript* matches widely used formats in genetic and transcriptomic analyses.

2 Implementation

*ggtranscript* is an R package released that extends the incredibly popular tool ggplot2 (RRID: SCR_014601 version: 3.3.5, https://
For longer, more complex transcripts, small differences between exons of interest can be more difficult to visualize (Fig. 1D). For this reason, ggtranscript includes a helper function `shorten_gaps()` which shortens regions that do not overlap an exon to a fixed, user-inputted width. Plotting of the rescaled exons and introns enables easier comparison between transcript structures when genes are long (Fig. 1E). In addition, the function `to_diff()` facilitates this by highlighting differences in comparison to a reference transcript (Fig. 1F).

### 3 Conclusion

ggtranscript enables a fast and simplified way to visualize, explore and interpret transcript isoforms. It allows users to combine data from both long-read and short-read RNA-sequencing technologies, making systematic assessment of transcript support easier. Finally, by being a ggplot2 extension it is highly flexible and can easily generate high-quality and publication-ready plots.

### Acknowledgement

The authors thank Siddharth Sethi and Geo Pertea for their valuable feedback and suggestions.

### Funding

This research was funded in whole or in part by the Aligning Science Across Parkinson’s [ASAP-000478 and ASAP-000509] through the Michael J. Fox Foundation for Parkinson’s Research (MJFF). For the purpose of open access, the author has applied a CC BY public copyright licence to all Author Accepted Manuscripts arising from this submission.
E.K.G. was also supported by the Postdoctoral Fellowship Program in Alzheimer’s Disease Research from the BrightFocus Foundation [A2021009F]. M.R. was supported through the award of a Tenure Track Clinician Scientist Fellowship [MR/N008324/1].

Conflict of Interest: none declared.

References
Conesa, A. et al. (2016) A survey of best practices for RNA-seq data analysis. Genome Biol., 17, 13.
Hahne, F. and Ivanek, R. (2016) Visualizing genomic data using gviz and bioconductor. Methods Mol. Biol., 1418, 335–351.
Kent, W. J. et al. (2002) The human genome browser at UCSC. Genome Res., 12, 996–1006.
Li, Y. I. et al. (2018) Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet., 50, 151–158.
Reese, F. and Mortazavi, A. (2021) Swan: a library for the analysis and visualization of long-read transcriptomes. Bioinformatics, 37, 1322–1323.
Robinson, J. T. et al. (2011) Integrative genomics viewer. Nat. Biotechnol., 29, 24–26.
Vitting-Seerup, K. et al. (2019) IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. Bioinformatics, 35, 4469–4471.
Wang, E. T. et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature, 456, 470–476.
Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.