





METHOD ARTICLE

Slinker: Visualising novel splicing events in RNA-Seq data

[version 1; peer review: 3 approved]

Breon Schmidt ^{1,2}, Marek Cmero^{1,3,4}, Paul Ekert^{1,3,5-7}, Nadia Davidson^{1,2,4}, Alicia Oshlack ¹⁻³

¹Peter MacCallum Cancer Centre, Melbourne, Victoria, 3000, Australia

²School of BioScience, University of Melbourne, Melbourne, Victoria, 3000, Australia

³Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, 3000, Australia

⁴Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, 3052, Australia

⁵Murdoch Children's Research Institute, Parkville, Victoria, 3052, Australia

⁶Children's Cancer Institute, Lowy Cancer Centre, Kensington, New South Wales, 2033, Australia

⁷School of Women's and Children's Health, University of New South Wales, Sydney, Victoria, 2052, Australia

V1 First published: 07 Dec 2021, **10**:1255
<https://doi.org/10.12688/f1000research.74836.1>
 Latest published: 07 Dec 2021, **10**:1255
<https://doi.org/10.12688/f1000research.74836.1>

Abstract

Visualisation of the transcriptome relative to a reference genome is fraught with sparsity. This is due to RNA sequencing (RNA-Seq) reads being predominantly mapped to exons that account for just under 3% of the human genome. Recently, we have used exon-only references, superTranscripts, to improve visualisation of aligned RNA-Seq data through the omission of supposedly unexpressed regions such as introns. However, variation within these regions can lead to novel splicing events that may drive a pathogenic phenotype. In these cases, the loss of information in only retaining annotated exons presents significant drawbacks. Here we present Slinker, a bioinformatics pipeline written in Python and Bpipe that uses a data-driven approach to assemble sample-specific superTranscripts. At its core, Slinker uses *Stringtie2* to assemble transcripts with any sequence across any gene. This assembly is merged with reference transcripts, converted to a superTranscript, of which rich visualisations are made through *Plotly* with associated annotation and coverage information. Slinker was validated on five novel splicing events of rare disease samples from a cohort of primary muscular disorders. In addition, Slinker was shown to be effective in visualising deletion events within transcriptomes of tumour samples in the important leukemia gene, *IKZF1*. Slinker offers a succinct visualisation of RNA-Seq alignments across typically sparse regions and is freely available on Github.

Keywords




RNA-Seq, Visualisation, Novel Splicing Events, superTranscripts, bioinformatics

Open Peer Review

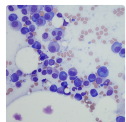
Reviewer Status 

Invited Reviewers

	1	2	3
version 1			
07 Dec 2021	report	report	report

1. **George Wiggins** , University of Otago, Christchurch, New Zealand
2. **Stephen Montgomery** , Stanford University School of Medicine, Stanford, USA
3. **Amy Huei-Yi Lee** , Simon Fraser University, Burnaby, Canada

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the [Cell & Molecular Biology gateway](#).

Corresponding author: Alicia Oshlack (alicia.oshlack@petermac.org)

Author roles: **Schmidt B:** Conceptualization, Formal Analysis, Methodology, Software, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Cmero M:** Writing – Review & Editing; **Ekert P:** Writing – Review & Editing; **Davidson N:** Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Oshlack A:** Conceptualization, Methodology, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the NHMRC under project grant APP1140626.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2021 Schmidt B *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Schmidt B, Cmero M, Ekert P *et al.* **Slinker: Visualising novel splicing events in RNA-Seq data [version 1; peer review: 3 approved]** F1000Research 2021, **10**:1255 <https://doi.org/10.12688/f1000research.74836.1>

First published: 07 Dec 2021, **10**:1255 <https://doi.org/10.12688/f1000research.74836.1>

Introduction

Genomic variants often carry through to the transcriptome. Through these events, gene products can be inhibited, upregulated, or modified, which disrupts typical function. Understanding how a genomic variant has altered transcription can offer insights into causal mechanisms for pathogenesis or potential targets for therapeutic intervention.^{1,2} DNA variants can be identified through genome sequencing and transcriptional disruption can be measured by RNA sequencing (RNA-Seq). In addition, RNA-Seq data can be used to detect an assortment of variants such as single nucleotide polymorphisms (SNPs), deletions or insertions, and altered splicing.²⁻⁶ It is common practice to visualise any novel events relative to control samples with tools such as the [Integrative Genomics Viewer \(IGV\)](#).⁷ This process involves aligning sequencing reads to a reference genome. However, when visualising RNA-Seq data, the size of introns are typically significantly larger than the exons that harbour most of the aligned reads.^{8,9} As a consequence, these visualisations are sparse. The intronic regions with low coverage offer little information, yet significantly reduce the interpretability of the transcribed regions.

In an effort to provide succinct representations of RNA-Seq data, we previously developed the concept of superTranscripts.¹⁰ A superTranscript is the flattened collection of transcripts within a gene and includes only the concatenated sequence of its exons. Subsequently, alignment to a superTranscript reference mostly resolves the sparsity problem through elimination of low-information intronic regions. This concept was further extended in our previous work with the release of [Clinker](#), which uses superTranscripts to visualise fusion genes.¹¹ In the Clinker pipeline, the pre-generated

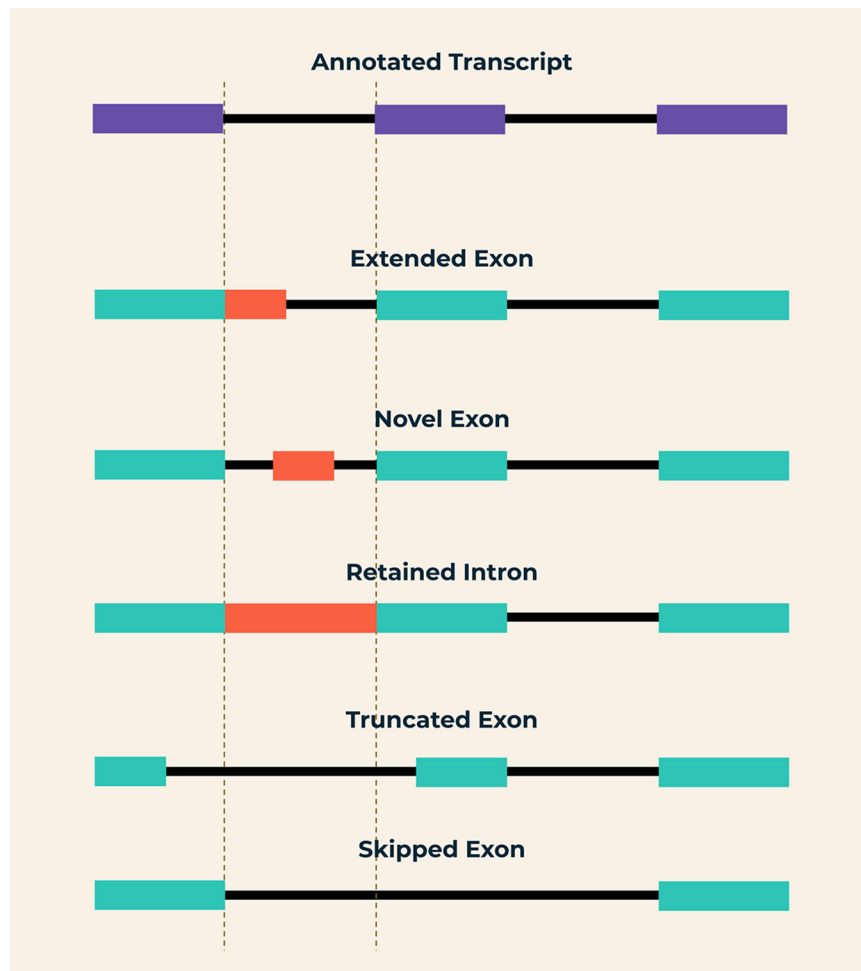


Figure 1. Depiction of each of the novel splicing events that Slinker is designed to visualise. Compared to the reference annotation: retained introns are defined by the inclusion of regions between exons, up to and including the exon boundary. Novel exons are defined by the inclusion of regions between exons, not including the exon boundaries. Skipped exons are defined by the existence of a splice junction in the case that starts and ends at known exon boundaries, but does not exist within the controls. Truncated exons occur when an exon in the case does not extend to the annotated exon boundary. Conversely, an extended exon is where an assembled exon within the case exceeds an annotated exon boundary.

superTranscripts of genes identified as belonging to a fusion are combined to form a new reference. The sequencing reads are then mapped to this with the splice-aware alignment tool, STAR.¹² Any splice junctions spanning the two genes in the superTranscripts are indicative of a fusion breakpoint. Although this method offers succinct representations of fusion events, it is not applicable to other types of events that can produce novel structures in transcripts. Specifically, in both the rare disease and cancer context, it is important to visualise any events that might occur within the excised intronic regions, or those that add sequences to the reference transcripts upstream or downstream of annotated exons.

To address these requirements in the visualisation of novel transcripts, we present Slinker, a superTranscript generation and visualisation method. Slinker builds on the concept of visualisation using superTranscripts, but uses genome-guided assembly rather than predefined annotation to incorporate the novel transcribed regions of interest into a reference. Through this, Slinker retains novel transcribed regions outside of annotated exons, that can be described as: novel exons, retained introns, skipped exons, truncated exons, and extended exons (Figure 1).

These events can be an important consequence of genomic variants and may be associated with a disease phenotype. To demonstrate the utility of Slinker, we applied it to rare disease variants, discovered with the aid of RNA-Seq in genes associated with primary muscle disorders in skeletal muscle² and events detected in childhood leukemia.^{6,13} We determined that Slinker offers a succinct and complementary method to visualise and explore RNA-Seq data. Slinker is a Bpipe pipeline publicly available via [Github](#) (Extended data).

Methods

Slinker is a pipeline that is used to visualise novel gene events using RNA-Seq. The main steps in the pipeline are outlined in Figure 2. Slinker requires multiple inputs: a gene name, aligned RNA-Seq reads (for the case and each control), a

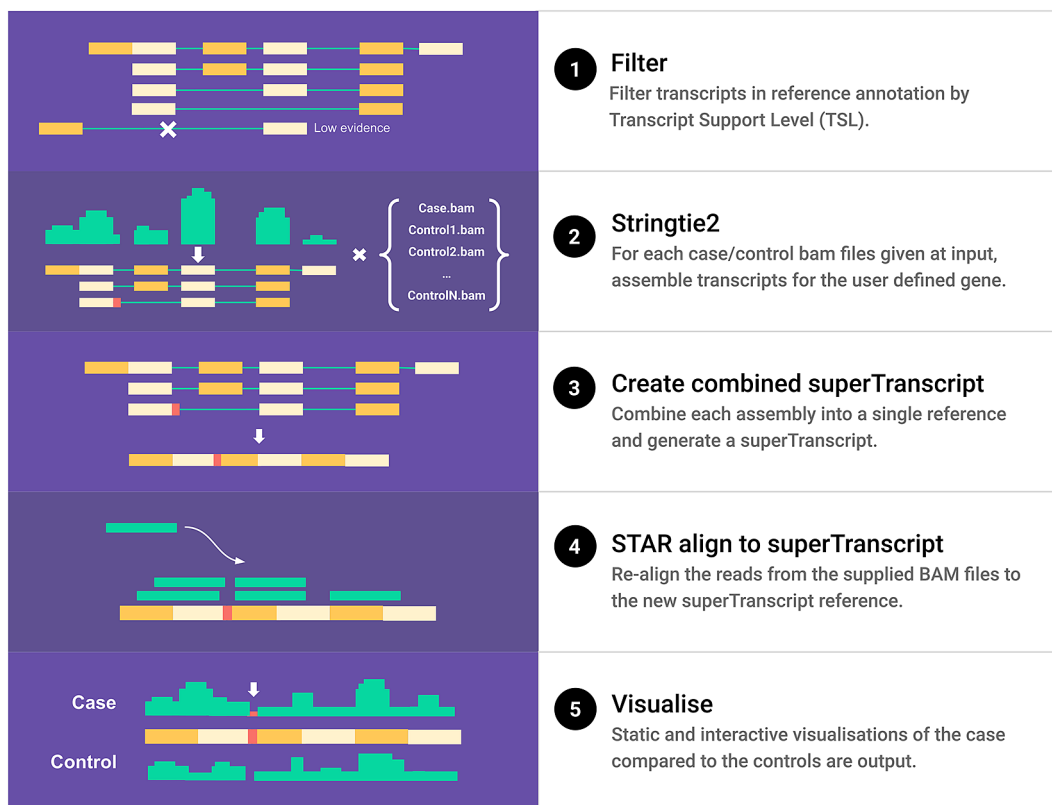


Figure 2. A schematic of the Slinker pipeline for a single user-input gene. Reference transcripts are first filtered by Transcript Support Level (TSL) as indicated in the reference annotation. The TSL is a tag included within Ensembl annotation that informs on the size and quality of evidence supporting the existence of the associated transcript. These transcripts, along with the input BAM files, are used to perform genome-guided assembly and output the predicted transcripts for each control and the case sample. The merged result of these assemblies is then flattened, exons concatenated, and oriented in the transcriptionally forwards direction to create the genome-guided superTranscripts. The reads previously mapped to the user-supplied gene are extracted from each input BAM and then re-aligned to this new reference. Finally, the coverage from each alignment, splice junctions, and transcript annotation are combined into both interactive and static visualisations upon completion of the pipeline.

genomic sequence reference, and a transcriptome reference annotation. A custom annotation can also be supplied. These must be provided in GTF format with the naming conventions for the attributes reflecting those used within **Ensembl** annotations.¹⁴ Upon completion, a PNG file containing a static image of the visualisation and an interactive HTML version are generated and saved to a user-defined destination.

The first step in the Slinker pipeline is to filter the gene transcripts to be visualised by the level of support in the Ensembl annotation. Specifically, all transcripts that either have a Transcript Support Level of 1 or are lacking this information are retained, the rest are filtered out. This ensures that only the best-supported transcripts are included in the assembly. The next step uses **Samtools** (v1.13) to extract all the reads aligned to the genomic region that the user-supplied gene defines in each input BAM. From each of these reduced BAM files, **Stringtie2** (v2.1.7) performs genome-guided assembly using either the supplied annotation or the hg19 or hg38 references packaged with Slinker.¹⁵ The output of **Stringtie2** is a GTF file for each BAM that contains all potential reference and novel transcripts existing within that sample. The transcripts are further filtered by the **Stringtie2** coverage estimate, removing any novel transcripts that do not meet the defined threshold (default, $c = 1$). These transcripts are then merged into a single GTF file. The reason for generating transcripts for both the case and controls is so that deletion events in the case are exposed. Next, we flatten these transcripts into a single linear representation, with concatenated exons, and then visually orient the strand left-to-right, creating a data-driven superTranscript.

A reference is then created from the generated superTranscript through **gffread** (v0.9.9) and STAR's (v2.7.3a) genomeGenerate mode.^{12,16} The reads extracted from each input BAM that overlap the gene of interest are then aligned to this new reference using STAR in alignReads mode, specifically with all splice junction score penalties set to 0. Assembled transcripts are then annotated according to the new superTranscript coordinates. Finally, a custom plotting package using the Python library, **Plotly** (v4.0 and above), is used to generate both a static image and interactive HTML output. These plots contain coverage and splice junction tracks for the test sample and controls with the superTranscript annotations included for context. Slinker automatically detects and highlights novel splicing events in the gene of interest. These events are discovered through comparisons between the reference and novel transcripts and highlighted according to the novel event's proximity to known exon boundaries, or whether an exon has been skipped. For instance, retained introns are defined by the inclusion of the region between exons in a gene, including the exon boundary. Novel exons are defined by the inclusion of regions between exons, not including the exon boundaries. Skipped exons are defined by the existence of a splice junction in the studied case that starts and ends at known exon boundaries, but does not exist within the controls. Truncated exons occur when an exon does not extend to the annotated exon boundary in the case. Conversely, an extended exon is where an assembled exon within the case exceeds an annotated exon boundary.

Table 1. A list of the cases and controls used within this study to validate Slinker. The GTEx samples were used as controls for the COL6A1, POMGNT1, RYR1, and NEB genes. The B-ALL20_7 and B-ALL9_4 samples were validated through Multiplex-ligation dependent probe amplification (MLPA) to not contain IKZF1 deletions and were used as controls against samples that were validated to have DEL4-7, DEL2-7, and DEL4-8.

Sample ID	Gene	Variant	Source	Verified
Cases				
SRR5020918	COL6A1	Novel Exon	(Cummings <i>et al.</i> 2017)	MINTIE
SRR5034830	POMGNT1	Skipped Exon	(Cummings <i>et al.</i> 2017)	MINTIE
		Retained Intron	(Cummings <i>et al.</i> 2017)	MINTIE
SRR5033001	RYR1	Truncated Exon	(Cummings <i>et al.</i> 2017)	MINTIE
SRR5038795	NEB	Extended Exon	(Cummings <i>et al.</i> 2017)	MINTIE
B-ALL18_4	IKZF1	DEL4-7	(Brown <i>et al.</i> 2020)	MLPA
B-ALL19_11	IKZF1	DEL2-7	(Brown <i>et al.</i> 2020)	MLPA
B-ALL7_8	IKZF1	DEL4-8	(Brown <i>et al.</i> 2020)	MLPA
Controls				
SRR811771	Multiple	N/A	(Carithers <i>et al.</i> 2015)	N/A
SRR810249	Multiple	N/A	(Carithers <i>et al.</i> 2015)	N/A
SRR809595	Multiple	N/A	(Carithers <i>et al.</i> 2015)	N/A
B-ALL20_7	IKZF1	N/A	(Brown <i>et al.</i> 2020)	MLPA
B-ALL9_4	IKZF1	N/A	(Brown <i>et al.</i> 2020)	MLPA

Slinker can be executed through a single command line option, where the inputs and outputs of each particular stage of the pipeline are handled by Bpipe.¹⁷ However, users who do not wish to use Bpipe may also run each step manually. The core dependencies of Slinker are the STAR aligner, Stringtie2, Plotly (v4 and above), gffread, and Samtools.^{12,15,16,18,19} Runtime for a single gene is under 10 minutes with 20gb of memory and four cores allocated, with a control and case sample containing approximately 125 and 275 million reads, respectively.

Results

In order to demonstrate the utility of Slinker for visualising transcriptional variants that are associated with rare disease, we applied it to muscle biopsy samples obtained from Cummings *et al.*² Of these, four were selected for visualisation

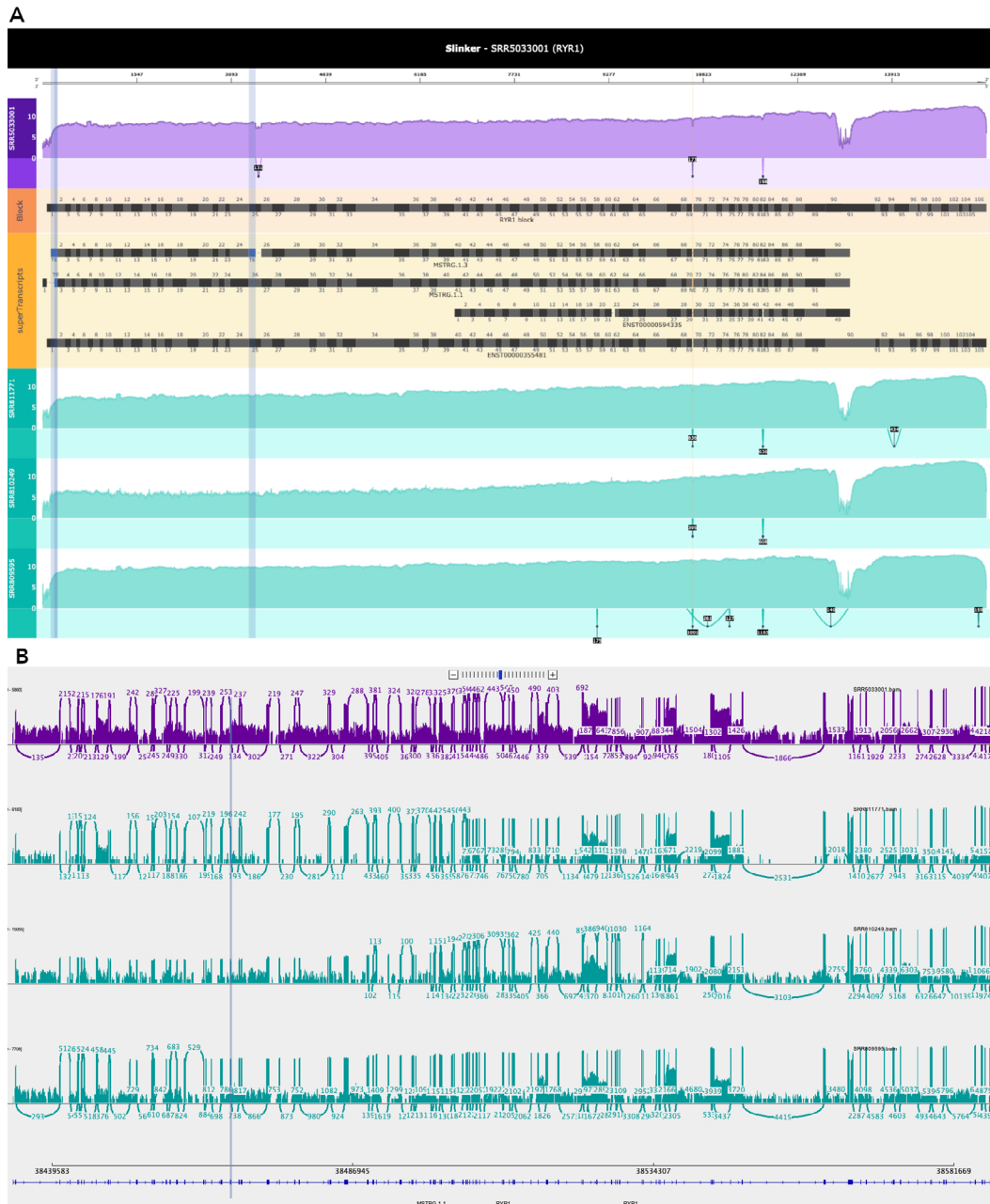


Figure 3. Integrated Genomics Viewer (IGV) and Slinker visualisations of the RYR1 gene. A. Slinker visualisations can show coverage (log scale) across the entire gene including events such as truncated exons (purple shading with adjacent drop in coverage), in a single, compact visualisation. Such a visualisation is desirable as users can consider the event in the greater context of the gene's expression. B. Integrative Genomics Viewer (IGV) Sashimi plot including the same case, controls, and transcripts as in A but not in the superTranscript format. Case is shown in purple and the controls in teal. The truncated exon is highlighted with the purple vertical line.

based on the existence of relevant novel splicing events in disease related genes (Table 1). These events were found to have consensus between Cummings *et al.*² and the variant caller, MINTIE.⁶ Three skeletal muscle controls were selected from the Genotype-Tissue Expression (GTEx) consortium that were also used as controls within Cummings *et al.*^{2,20} The selected samples were then aligned to hg38 with the STAR aligner in 2-pass mode, of which the resulting BAM files were inputted into the Slinker pipeline.¹²

Figure 3 demonstrates Slinker’s superTranscript visualisation for the RYR1 gene, which has 105 exons across 154kb of the genome. This was compared to the visualisation of the same region using a refined IGV sashimi plot. Log coverage settings were enabled in both Slinker and IGV. Slinker clearly revealed the novel event of a truncated exon in exon 25, which was far less obvious in the genomic view that included the unnecessary splice junctions and alignment sparsity.

Next, we demonstrated Slinker’s ability to visualise a range of novel events involving both added and deleted sequences in transcripts, each confirmed through MINTIE. Figure 4A depicts two novel splicing events in POMGNT1 found within a single case sample, and is highlighted through a comparison with the three GTEx controls (Table 1). In this case, we could see a retained intron in the genome-guided assembly (highlighted in yellow) and an exon-skipping event (highlighted in purple) in exon 7. Applying Slinker to the COL6A1 gene in another sample clearly revealed the novel exon as reported by Cummings *et al.*² (Figure 4B). Furthermore, a truncated exon is shown in the RYR1 gene (Figure 3A) and an extended exon is shown within the NEB gene (Figure 5).

Finally, we applied Slinker to three childhood B-Cell Acute Lymphoblastic Leukemia samples obtained from the Royal Children’s Hospital (RCH), that harboured various IKZF1 deletions; these were validated in DNA using a

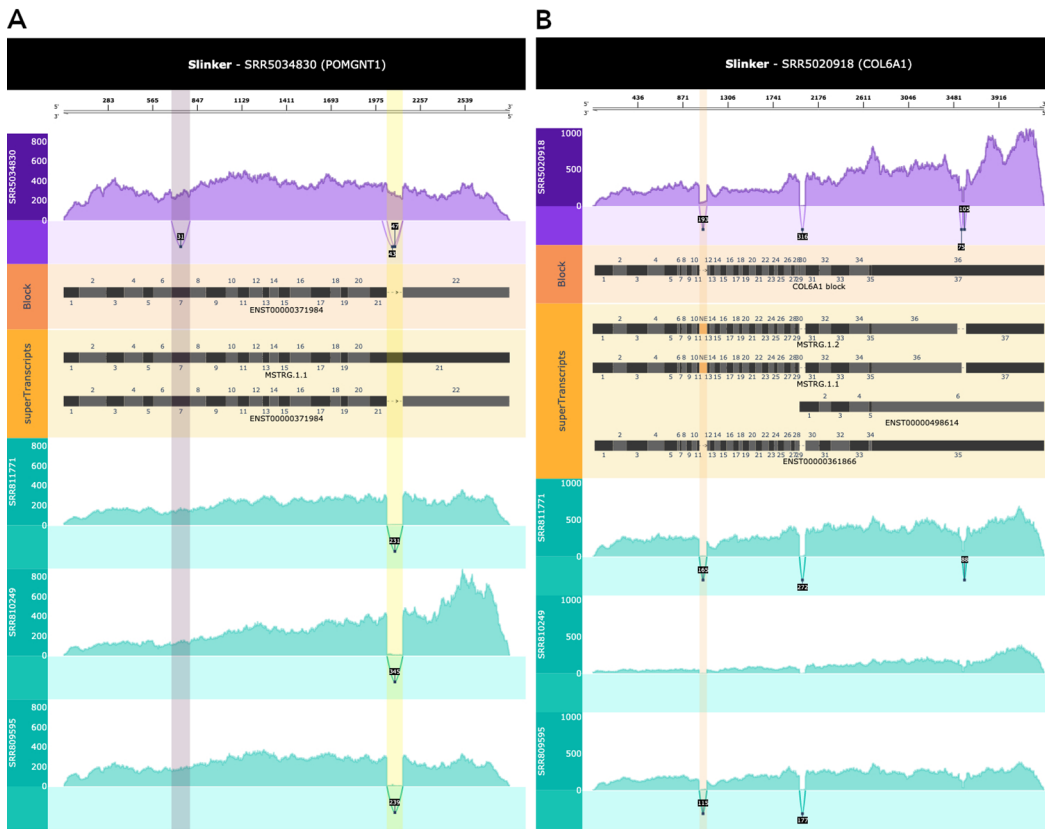


Figure 4. A. Slinker visualisation of the POMGNT1 gene in a sample containing two variants which resulted in both removed and added sequence. A skipped exon (purple vertical highlight) and a retained intron (yellow vertical highlight) are highlighted. From the top: Title bar, Axis in superTranscript coordinates, case’s coverage in raw counts, case’s splice junctions, superTranscript of only the reference transcripts, novel and reference superTranscripts as defined by Stringtie2, and a coverage and junctions track for each control. Slinker highlights relevant novel splicing events (retained introns, novel exons, extended exons, truncated exons, and skipped exons) in a distinct colour to enable users to quickly identify what variation exists between a case and controls. B. Slinker visualisation of COL6A1 in a sample with a novel exon event (orange vertical highlight) between exons 11-12.

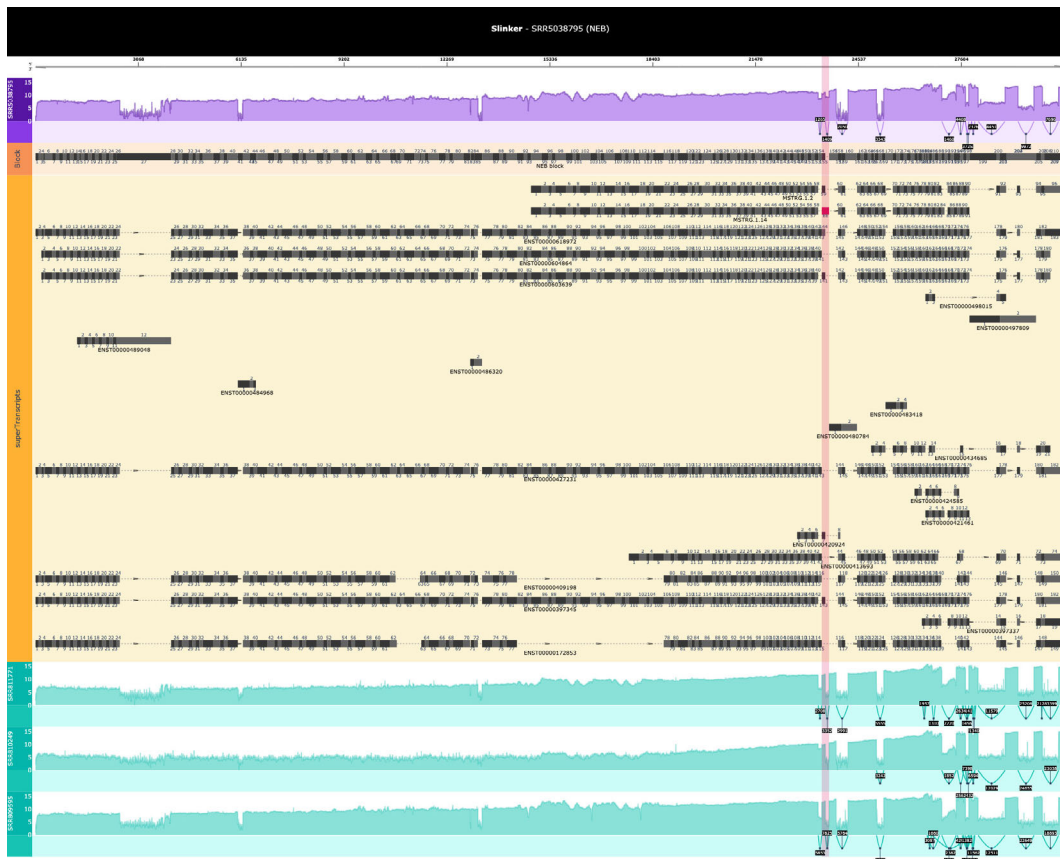


Figure 5. Slinker visualisation in a sample with an extended exon event (red vertical highlight) between exons 143-144 in the NEB gene. In this case, the TSL filter was not applied and a reference with all transcripts was passed through to Slinker. This was due to the large number of transcripts of NEB, which were assembled by Stringtie and masked the interesting event. This demonstrates the utility of being able to pass through custom annotations which may be more suitable to the visualisation at hand.

multiplex-ligation-dependent probe amplification (MLPA) assay (Table 1). Figure 6A depicts an exon 4-7 deletion within one such case compared to two other leukemia samples in the RCH cohort. These two samples were chosen due to their high IKZF1 expression relative to other samples within the cohort, but did not contain a deletion either. In addition to the automatic highlighting denoting a SE event, a clear drop in expression across exons 4 and 7 can be seen in the case relative to the control samples. The 2-7 deletion event (Figure 6B) was also clearly highlighted as a SE event. However, this was not true for the 4-8 deletion sample as no splice junctions existed and the deletion included the final exon (Figure 6C). However, the drop in coverage between the case and controls in this succinct form demonstrates the utility of Slinker for providing a visual comparison between any samples. An interesting aspect of the 4-7 IKZF1 deletion was the assembly of multiple transcripts containing the deletion event (Figure 6B). This demonstrates that Slinker may help provide more information for deletion events over exclusive alignment to predefined superTranscripts.

Discussion

DNA mutations can be causal drivers of disease.^{2,6} While genomic sequencing is becoming more commonly used to diagnose genetic diseases, there are still many cases where the variants cannot be directly identified as disease-causing, and are therefore called variants of unknown significance. Mutations with the potential to impact splicing are of particular interest. One way to assess the effect of a variant is to investigate the resulting transcript using RNA sequencing and compare these transcripts to a set of controls. While there are several approaches proposed for this purpose, the visualisation of this data can be improved.^{2,21,22} Here we address this issue with Slinker, which was built on the superTranscript visualisation framework. One of the fundamental advantages of the superTranscript method is the removal of the visualisation sparsity due to uninformative intronic information. However, novel splicing events can retain an intronic sequence, and it is therefore necessary to first determine which sequences are transcribed in the sample. Slinker utilizes the Stringtie2 data-driven method for assembling the transcripts in the sample which is then combined with reference transcripts to capture all potentially expressed sequences.

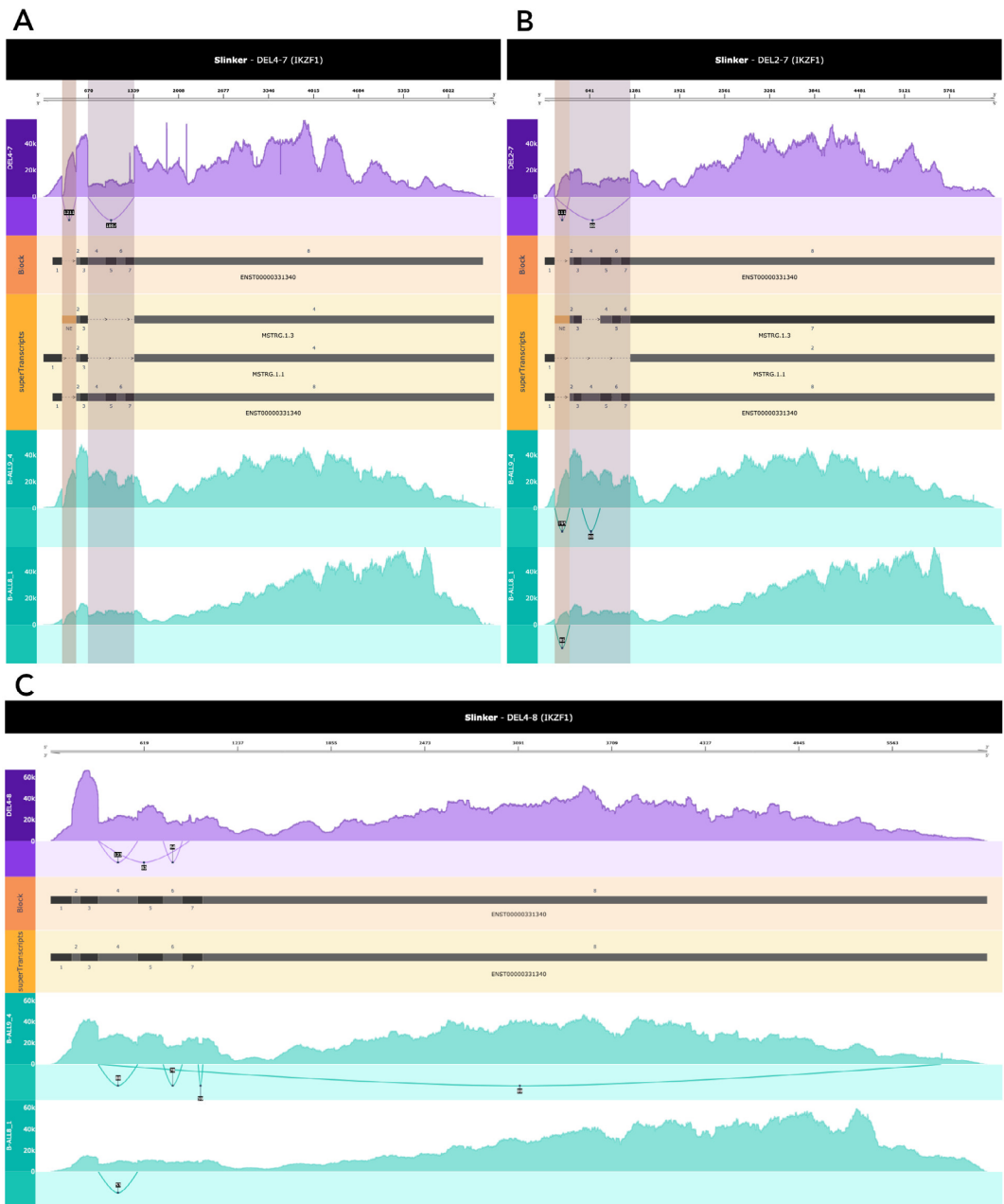


Figure 6. Deletions within the IKZF1 gene. A. An exon 4-7 deletion in IKZF1. A clear dip in coverage in the case compared to the controls can be seen within the purple highlighted region. Slinker was run using only a single transcript (ENST00000331349), which was representative of the canonical. The highlighted novel exon was annotated in a transcript that is within the greater reference. Transcripts were filtered using the minimum coverage parameter to include all assemblies with the known deletion event. Coverage was normalised by total reads in IKZF1. B. A 2-7 deletion in IKZF1. A clear dip in coverage in the case in comparison to the controls can be seen within the purple highlighted region. In this figure, Slinker was run using only a single transcript (ENST00000331349) which was representative of the canonical. Coverage was normalised by total reads in IKZF1. C. A 4-8 deletion in IKZF1. This comparison demonstrates a slight change in coverage post exon 3 to the end of the transcript in the case. This is representative of the deletion event known to have occurred. Coverage was normalised by total reads in IKZF1.

Due to the large insertion of sequences relative to the superTranscript, retained introns and novel exons were expected to be the simplest events to visualise with Slinker. This is reflected in Figure 4A and 4B. The same is true for exon skipping, where the existence of a splice junction at annotated exon boundaries in the case versus lack thereof in the control was clear (Figure 4A). However, truncated and extended exons appeared to be less obvious in Slinker given the relatively

small amount of sequence which were removed/added (Figures 3A and 5). Nevertheless, the bespoke superTranscripts, in conjunction with the highlighting and comparisons with controls could reveal these types of events in the static visualisation even in the largest of genes (Figure 3A). In addition, Slinker generates an interactive plot so that the user can simply zoom in on these regions to better understand the event whilst also zooming out to appreciate the greater context if required.

Though Slinker was developed to highlight novel splicing events in rare diseases, it can also be used to visualise cancer transcripts. Slinker was applied to three B-Cell Acute Lymphoblastic Leukemia (B-ALL) cases harbouring 4-7, 4-8, and 2-7 IKZF1 deletions. Deletions within this gene are a known risk factor in aggressive forms of this cancer and are used to monitor disease progression.¹³ In each of these examples, a clear deletion event could be seen through either the presence of skipped exons or the relative drop in coverage when compared to the controls (Figure 6). However, generally, cancer is a genetically complex disease and consequently a genetic visualisation is also challenging.²³ Nevertheless, these results demonstrate that Slinker can be an appropriate choice for providing a succinct visualisation of this complexity and may be applicable to a broader set of diseases.

Further work may involve improving this aspect of the software, including highlighting other novel events, such as inversions and duplications. However, this is limited by the assembler's capacity to retrieve an accurate assembly, including the variant, and for a superTranscript to be an appropriate visual representation.⁶

Conclusion

Slinker is a bespoke superTranscript generation and visualisation tool with a demonstrated ability to succinctly present novel splicing events in RNA-Seq data. Its key advantage consists in removing redundant information without manual intervention, leaving room for more useful information, such as the expression across the entire gene.

We have produced a tool that has been validated on a number of known novel splice variants and is publicly available from Github (<https://github.com/Oshlack/Slinker>).

Data availability

Underlying data

European Genome-Phenome Archive: The application of RNA sequencing for the diagnosis and genomic classification of pediatric acute lymphoblastic leukemia, <https://identifiers.org/ega.study:EGAS00001004212>²⁰ This study contains the paediatric B-ALL samples (B-ALL18_4, B-ALL19_11, B-ALL7_8, B-ALL8_1, and B-ALL9_4).

NCBI dbGaP: Genetics of Inherited Muscle Disease, <https://identifiers.org/dbgap:phs000655.v3.p1>

NCBI dbGaP: Common Fund (CF) Genotype-Tissue Expression Project (GTEx), <https://identifiers.org/dbgap:phs000424.v6.p1>

Extended data

Analysis code available from: <https://github.com/Oshlack/Slinker>

Archived code as at time of publication: <https://doi.org/10.5281/zenodo.5719747>

License: MIT

Competing interests

No competing interests were disclosed.

Acknowledgements

- Tumour samples and coded data were supplied by the Children's Cancer Centre Tissue Bank at the [Murdoch Children's Research Institute](#) and [The Royal Children's Hospital](#). Establishment and running of the Children's Cancer Centre Tissue Bank is made possible through generous support by [Cancer In Kids @ RCH](#), The Royal Children's Hospital Foundation and the Murdoch Children's Research Institute.

References

1. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, *et al.*: **Translating RNA sequencing into clinical diagnostics: opportunities and challenges.** *Nat. Rev. Genet.* 2016 May; **17**(5): 257–271.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Cummings BB, Marshall JL, Tukiainen T, *et al.*: **Improving genetic diagnosis in Mendelian disease with transcriptome sequencing.** *Sci. Transl. Med.* 2017 Apr 19; **9**(386)
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Kim D, Pertea G, Trapnell C, *et al.*: **TopHat2: accurate alignment of transcripts in the presence of insertions, deletions and gene fusions.** *Genome Biol.* 2013 Apr 25; **14**(4): R36.
[Publisher Full Text](#)
4. Van der Auwera GA, O'Connor BD: *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra.* O'Reilly Media, Inc.; 2020; 496.
5. Mortazavi A, Williams BA, McCue K, *et al.*: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat. Methods.* 2008 Jul; **5**(7): 621–628.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Cmero M, Schmidt B, Majewski JJ, *et al.*: **MINTIE: Identifying novel structural and splice variants in transcriptomes using RNA-seq data.** Cold Spring Harbor Laboratory; 2020 [cited 2020 Nov 30]; 2020.06.03.131532.
[Publisher Full Text](#)
7. Robinson JT, Thorvaldsdóttir H, Winckler W, *et al.*: **Integrative genomics viewer.** *Nat. Biotechnol.* 2011 Jan 1; **29**(1): 24–26.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Venter JC, Adams MD, Myers EW, *et al.*: **The sequence of the human genome.** *Science.* 2001 Feb 16; **291**(5507): 1304–1351.
[Publisher Full Text](#)
9. ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome.** *Nature.* 2012 Sep 6; **489**(7414): 57–74.
[PubMed Abstract](#) | [Publisher Full Text](#)
10. Davidson NM, Hawkins ADK, Oshlack A: **SuperTranscripts: a data driven reference for analysis and visualisation of transcriptomes.** *Genome Biol.* 2017 Aug 4; **18**(1): 148.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Schmidt BM, Davidson NM, Hawkins ADK, *et al.*: **Clinker: visualizing fusion genes detected in RNA-seq data.** *Gigascience.* 2018 Jul 1; **7**(7).
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Dobin A, Davis CA, Schlesinger F, *et al.*: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics.* 2013 Jan 1; **29**(1): 15–21.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Brown LM, Lonsdale A, Zhu A, *et al.*: **The application of RNA sequencing for the diagnosis and genomic classification of pediatric acute lymphoblastic leukemia.** *Blood Adv.* 2020 Mar 10; **4**(5): 930–942.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Howe KL, Achuthan P, Allen J, *et al.*: **Ensembl 2021.** *Nucleic Acids Res.* 2021 Jan 8; **49**(D1): D884–D891.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. Kovaka S, Zimin AV, Pertea GM, *et al.*: **Transcriptome assembly from long-read RNA-seq alignments with StringTie2.** *Genome Biol.* 2019 Dec 16; **20**(1): 278.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Pertea G, Pertea M: **GFF Utilities: GffRead and GffCompare.** *F1000Res.* 2020; **9**: 304.
[Publisher Full Text](#)
17. Sadedin SP, Pope B, Oshlack A: **Bpipe: a tool for running and managing bioinformatics pipelines.** *Bioinformatics.* 2012 Jun 1; **28**(11): 1525–1526.
[PubMed Abstract](#) | [Publisher Full Text](#)
18. *Collaborative data science.* Montréal, QC: Plotly Technologies Inc; 2015.
19. Li H, Handsaker B, Wysoker A, *et al.*: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics.* 2009 Aug 15; **25**(16): 2078–2079.
[PubMed Abstract](#) | [Publisher Full Text](#)
20. Carithers LJ, Ardlie K, Barcus M, *et al.*: **A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project.** *Biopreserv Biobank.* 2015 Oct; **13**(5): 311–319.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Yépez VA, Mertes C, Müller MF, *et al.*: **Detection of aberrant gene expression events in RNA sequencing data.** *Nat. Protoc.* 2021 Feb; **16**(2): 1276–1296.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Ullirsch JC, Verboon JM, Kazerounian S, *et al.*: **The Genetic Landscape of Diamond-Blackfan Anemia.** *Am. J. Hum. Genet.* 2019 Feb 7; **104**(2): 356.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Greaves M, Maley CC: **Clonal evolution in cancer.** *Nature.* 2012 Jan 18; **481**(7381): 306–313.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 1

Reviewer Report 03 February 2022

<https://doi.org/10.5256/f1000research.78631.r118438>

© 2022 Hwei-Yi Lee A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Amy Hwei-Yi Lee 

Simon Fraser University, Burnaby, BC, Canada

Visualizing RNA-Seq data using the reference genome via a tool like Integrative Genomics Viewer is challenged by the sparsity of transcriptome data, as RNA-Seq reads predominately map to exons. This impedes our abilities to detect SNPs, deletions/insertions and altered splicing events using RNA-Seq reads, and the associative impact on expression levels of the impacted gene. Schmidt *et al.* extend the concept of superTranscripts, which is a concatenated sequence of all the exons with a gene, with genome-guided assembly to allow visualization of novel exons, retained introns, skipped exons, truncated exons and extended exons. The manuscript is very well written with clear, detailed figures (especially the beautiful schematic of the workflow as shown in Figure 2). This is definitely a tool that I am looking forward to using in my own work.

I only have two minor comments:

1. The colors used in Fig 1 need to be explained in the legends – and the green exons/red introns should probably be different colors to make it colorblind friendly.
2. The comparison of case and control requires coverage/sequencing depth to help with detection of some of the novel splice events (as shown in Fig 3 – 6). What happens when the sequencing depth between case and control are drastically different?

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Systems immunology, multi-omics integration, neonatal sepsis, host-pathogen interactions

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 07 January 2022

<https://doi.org/10.5256/f1000research.78631.r102214>

© 2022 Montgomery S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Stephen Montgomery 

Department of Pathology and Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

Schmidt et al provide a pipeline for visualizing novel splicing events in RNA-seq data. The key limitations with existing approaches that they have identified are that sparsity of mapped RNA-seq reads to the genome making visualization of novel events harder to detect and further, that existing approaches are mainly guided by existing/known gene annotations. The combined approach appears useful however there are a number of minor weaknesses.

1. In practice, novel products/events are first found statistically and then visualized. It appears that the authors are suggesting that this is mostly done first by manual inspection and Slinker makes it easier to find them. From the examples provided with Slinker or with IGV it would be very difficult to eyeball the differences without considerable experience. In fact, even the authors in the figures have to highlight the critical regions so the readers know what to look at. I am not entirely sure the problem they are describing is that large or they have sufficiently solved it.
2. Overall, I think the clearly described pipeline in combination with the visualizations provides more utility and impact. I would be interested in knowing how dependent the pipeline is on the specific tools or is it compatible with specific conventional file formats and the user can change the underlying components as desired. Can the genome build be easily changed?

3. I think Figure 1 would benefit from an example of the corresponding superTranscripts. It wasn't immediately clear to me how a Skipped Exon would be conveniently visualized compared to the retained intron in Figure 2. Overall, some better documentation on how to visualize specific events as the practical examples were hard to read and very small.
4. The examples (Fig 3, 4, 5) have abundant text that is nearly unreadable. I am a bit concerned that the output of Slinker is hard to convert into publication quality figures. Does it output as vector graphics for Illustrator/Inkscape if someone was making a multi-panel figure?
5. In some of the figures there appear to be other events but they are not highlighted. It would be interesting for the authors to comment on that. For instance in Figure 3 exon 69-70.
6. The number of controls compared to is pretty small. In practice, it may be useful to compare to 100s or 1000s of controls to find an aberrant event. Is this possible? Figure 4 focused on 3 GTEx control samples only.
7. With the splice junction penalty score set to 0, are there many more false positives than true events? How many events would visually look interesting in other randomly selected comparisons of a couple samples between genes? Does setting this score influence the recovery of the known disease event examples?
8. In the discussion, 1st para, the word mutation and variant is used interchangeably. I would recommend being more precise here.
9. You don't have to cite it. But you might find our splicePlot tool interesting too. It was designed for sQTL but one can summarize data across lots of samples. However, not with the flexibility you have here since it only focused on junctions that one previously prioritized/found interesting. <https://doi.org/10.1093/bioinformatics/btt733>

Overall, nice work. The paper was well-written. I hope the tool provides some help for people looking to explore novel junctions.

References

1. Wu E, Nance T, Montgomery SB: SplicePlot: a utility for visualizing splicing quantitative trait loci. *Bioinformatics*. 2014; **30** (7): 1025-6 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computational biology, transcriptomics, human genetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 05 January 2022

<https://doi.org/10.5256/f1000research.78631.r118437>

© 2022 Wiggins G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



George Wiggins 

Mackenzie Cancer Research Group, Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand

The authors have developed a novel method (Slinker) to visualize RNA isoforms that improves interpretation. The method requires condensing the genomic region to view only the sequence that is supported by RNA-sequencing reads.

In my opinion Slinker does address the need in visualizing RNA isoforms that are assembled from RNA-sequencing data. A limitation that should be discussed is the reliance on StringTie2 for transcript assemble. RNA isoform discovery is a growing field, in which Slinker would offer great use, if adaptable to the development of new tools (e.g. FLAIR, bambu).

Minor comments are listed below:

- I have assumed long and short-read sequencing can be visualized equally? If not, this needs to be highlighted as a limitation in the discussion.
- In the discussion, the authors highlight how RNA species can help determine the pathogenicity of a given variants. For interpretation of RNA splicing events, it is important to determine whether splicing shift the coding frame. Is this possible to implement? This is not a requirement for indexing but a suggested improvement to Slinker.

All figures are presented well but could be improved with increased font sizes. Additionally, is it possible to provide an HTML version of the one plot in the supplementary information or in the

github repo?

- Figure3A has purple shading at the 5` end of the gene that I assume is not relevant, this should be removed for clarity.
- Please provide definitions for acronyms used in plots, e.g. superTranscript panels have TE, NE, TF etc.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cancer genetics, RNA-sequencing, bioinformatics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research