**Significance**

Background

One of the major outstanding questions in the field of splicing is how are the sections of a pre-mRNA that will be included in the final transcript determined. **Massively parallel splicing assays (MPSAs)** have revolutionized our ability to understand the regulatory mechanisms of splicing. These assays use high-throughput sequencing to measure quantitative splicing phenotypes for 10s of thousands of variants in parallel.1–15 They work by linking the sequence of the pre-mRNA to a barcode and then either compare barcode abundances between two conditions as a proxy for splicing1,7,10–12,14–16 or directly sequence the splice junction of interest and the barcode to observe the isoform distribution of each variant.3,5–7,9,13 These assays have allowed identification of human sequence variants that cause splicing changes,11 in depth study of how drugs interreact with splicing regulation,15 and detailed mechanistic studies on splicing regulations.1,6,7,10,12 However, MPSAs have been held back by three barriers to progress.

Barrier to progress 1: Current MPSAs cannot report the full isoform diversity of alternative splicing. Many MPSAs rely on counting barcodes across two conditions as a proxy for splicing outcomes, but this forcibly binarizes the isoform data that is collected. It is common for there to be multiple alternative isoforms that can arise from mutations.5,9,17,18 Missing this diversity by binarizing the data with a proxy assay causes two problems. The first is that unexpected isoforms will be binned into one of the binary conditions causing inaccuracy and noise in the results. The second is that loosing data on the true isoform distributions of each variant misses out on information that is highly mechanistically informative. One solution to overcome this issue is to sequence the splice junction of interest. However, current methods use Illumina short-read sequencing which limits them to isoforms that are small enough to be covered by a short read.3,5–7,9,13 Since the median human intron is 1.7 kb,19 this technique cannot distinguish isoforms that contain some or all of the intron for most human splice junctions unless the introns are cut down or replaced. However, intronic sequences can be highly important for determining splicing outcomes,17,18,20–22 so shrinking or replacing them can reduce generalizability of the data to native contexts. Creating MPSA methods that are able to measure the full diversity of isoforms produced by alternative splicing will allow both higher quality data in simple splicing contexts and allow extension of MPSA methods to the study of more complicated splicing events.

Barrier to progress 2: Current MPSAs cannot handle mutually exclusive splicing. Recent work has revealed that correlated splicing is far more common than previously understood.23–26 The simplest and best characterized form of correlated splicing is known as mutually exclusive exon splicing. Mutually exclusive exons are clusters of exons where exactly one exon from the cluster is included in the final transcript. Mutually exclusive exons are important for regulating a wide variety of processes including cell fate determination, neural development, and muscle development.20 Furthermore, mutations in mutually exclusive exons or changes in their splicing are involved in diseases including Timothy syndrome, cancer, heart disease, and cardiomyopathy.20 Despite their importance to biological processes, more than 75% of human mutually exclusive exon clusters have no known mechanism of maintaining mutual exclusivity.23 Correlated splicing and mutually exclusive exons represent an important class of splicing events that has been largely inaccessible to MPSA techniques. This is because determining the mechanisms maintaining mutual exclusivity between exons relies on tracking the fate of multiple splice junctions that are at a remove for each other. This carries with it all of the difficulties with tracking the diverse isoforms that might result from an MPSA of one splice junction (see barrier to progress 1) with the added complication that several splice junctions need to be tracked. To make things worse, in some cases these junctions might not even be adjacent, and can be at quite a distance and have other exons in between.23–26 This sort of splicing phenomenon where the fate of more than one exon is important can also easily generate isoforms with no included introns that are too large to be covered by an illumina read even in the simplest cases of two short exons close together.17 Creating new MPSA methods that are capable of handling these more complicated systems will open up large numbers of mutually exclusive exon clusters and other correlated splicing events to rapid mechanistic characterization, and help us discover the missing mechanisms of mutual exclusivity.

Barrier to progress 3: There is no community MPSA analysis software. Current MPSA techniques have been independently developed and few experiments have the same experimental design or analysis methods. These factors make it hard to compare findings across multiple experiments and raise barriers to entry because the methods and analysis pipelines need to be redeveloped by each group. Further, analysis pipelines have not been tested against datasets with known ground truth. Stronger vetting would increase confidence in the robustness and accuracy of these methods. Finally, MPSA data holds information useful for modeling the mechanisms of splicing but there is no tool set that integrates MPSA data analysis and modeling. Development of a user-friendly, publicly available, robustly verified, and open-source MPSA toolkit will greatly ease wider use of the method and help unify different techniques.

*This proposal will overcome these barriers by developing new long read MPSA (LR-MPSA) methods and analysis software. These methods will be evaluated by testing on simulated data sets, a well characterized SMN2 5’-splice site library, and new PKM mutually exclusive exon libraries. The analysis software will be made freely available to the research community as software package.*

The LR-MPSA will enable performing splice junction sequencing in constructs with full size introns without ignoring intron inclusion isoforms. The LR-MPSA will also allow analysis of more complicated splicing situations that are expected to produce multiple minor isoforms and isoforms too large to fit in a short read. This includes the study of systems with intron retention or cryptic splice sites as well as mutually exclusive exons and other forms of correlated splicing outcomes. The key advance of this technique will be to use oxford nanopore sequencing that is capable of sequencing the full length of any isoform or even un-spliced transcript.

Our analysis software will provide tools for end-to-end analysis of LR-MPSA data. This will include parsing Nanopore reads, clustering and aligning reads to identify isoforms, quantifying isoforms by library variant, providing visualization functions, and integration with modeling tools like MAVE-NN. The software will be rigorously vetted on simulated data and well ununderstood experimental systems. The key advances of this technique will be to rigorously verify performance, to integrate seamlessly with quantitative modeling packages, and to provide an easy-to-use, widely-accessible software package.

These new techniques will be demonstrated through investigating the mechanism of producing mutual exclusivity between exons 9 and 10 of pyruvate kinase M (PKM). With the inclusion of exon 9, PKM becomes a constitutively active pyruvate kinase. With the inclusion of exon 10, PKM becomes sensitive to allosteric regulation that switches it from acting as a pyruvate kinase to acting as a protein kinase phosphorylating Stat3 and a transcriptional co-activator of HIF-1 and β-catenin. Increases in the level of the exon 10 included transcript have been associated with cancer,27–31 Alzheimer’s,32 Crohn’s,33,34 and cardiovascular disease.35 However, there has been no systematic analysis of the regulatory mechanisms that produce the mutually exclusive splicing behavior of PKM. This makes PKM a perfect test case for the new LR-MPSA and analysis software.

We propose pursuing these goals through 3 complementary aims. Aim 1 will develop the LR-MPSA experimental techniques. Aim 2 will develop the MPSA analysis software package. Finally, aim 3 will apply these new tools to understanding the mechanisms of mutually exclusive splicing in PKM.

1. Ke, S. *et al.* Saturation mutagenesis reveals manifold determinants of exon definition. *Genome Res.* **28**, 11–24 (2018).

2. Julien, P., Miñana, B., Baeza-Centurion, P., Valcárcel, J. & Lehner, B. The complete local genotype–phenotype landscape for the alternative splicing of a human exon. *Nat. Commun.* **7**, 1–8 (2016).

3. Adamson, S. I., Zhan, L. & Graveley, B. R. Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency. *Genome Biol.* **19**, 71 (2018).

4. Soemedi, R. *et al.* Pathogenic variants that alter protein code often disrupt splicing. *Nat. Genet.* **49**, 848–855 (2017).

5. Cortés-López, M. *et al.* High-throughput mutagenesis identifies mutations and RNA-binding proteins controlling CD19 splicing and CART-19 therapy resistance. *Nat. Commun.* **13**, 1–17 (2022).

6. Schirman, D., Yakhini, Z., Pilpel, Y. & Dahan, O. A broad analysis of splicing regulation in yeast using a large library of synthetic introns. *PLoS Genet.* **17**, e1009805 (2021).

7. Mikl, M., Hamburg, A., Pilpel, Y. & Segal, E. Dissecting splicing decisions and cell-to-cell variability with designed sequence libraries. *Nat. Commun.* **10**, 1–14 (2019).

8. Braun, S. *et al.* Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis. *Nat. Commun.* **9**, 1–18 (2018).

9. Souček, P. *et al.* High-throughput analysis revealed mutations’ diverging effects on SMN1 exon 7 splicing. *RNA Biol.* **16**, 1364–1376 (2019).

10. Baeza-Centurion, P., Miñana, B., Valcárcel, J. & Lehner, B. Mutations primarily alter the inclusion of alternatively spliced exons. *Elife* **9**, (2020).

11. Cheung, R. *et al.* A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. *Mol. Cell* **73**, 183-194.e8 (2019).

12. Baeza-Centurion, P., Miñana, B., Schmiedel, J. M., Valcárcel, J. & Lehner, B. Combinatorial Genetics Reveals a Scaling Law for the Effects of Mutations on Splicing. *Cell* **176**, 549-563.e23 (2019).

13. Rosenberg, A. B., Patwardhan, R. P., Shendure, J. & Seelig, G. Learning the sequence determinants of alternative splicing from millions of random sequences. *Cell* **163**, 698–711 (2015).

14. Wong, M. S., Kinney, J. B. & Krainer, A. R. Quantitative Activity Profile and Context Dependence of All Human 5’ Splice Sites. *Mol. Cell* **71**, 1012-1026.e3 (2018).

15. Ishigami, Y. *et al.* Specificity, synergy, and mechanisms of splice-modifying drugs. *bioRxiv* 2022.12.30.522303 (2022) doi:10.1101/2022.12.30.522303.

16. North, K. *et al.* Synthetic introns enable splicing factor mutation-dependent targeting of cancer cells. *Nat. Biotechnol.* **40**, 1103–1113 (2022).

17. Wang, Z. *et al.* Exon-centric regulation of pyruvate kinase M alternative splicing via mutually exclusive exons. *Journal of molecular* (2012).

18. Wang, Z., Jeon, H. Y., Rigo, F., Bennett, C. F. & Krainer, A. R. Manipulation of PK-M mutually exclusive alternative splicing by antisense oligonucleotides. *Open Biol.* **2**, 120133 (2012).

19. Piovesan, A. *et al.* Human protein-coding genes and gene feature statistics in 2019. *BMC Res. Notes* **12**, 315 (2019).

20. Jin, Y., Dong, H., Shi, Y. & Bian, L. Mutually exclusive alternative splicing of pre‐mRNAs. *Wiley Interdiscip. Rev. RNA* **9**, e1468 (2018).

21. Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008).

22. Kalinina, M. *et al.* Multiple competing RNA structures dynamically control alternative splicing in the human ATE1 gene. *Nucleic Acids Res.* **49**, 479–490 (2021).

23. Hatje, K. *et al.* The landscape of human mutually exclusive splicing. *Mol. Syst. Biol.* **13**, 959 (2017).

24. Tilgner, H. *et al.* Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events. *Nat. Biotechnol.* **33**, 736–742 (2015).

25. Tilgner, H. *et al.* Microfluidic isoform sequencing shows widespread splicing coordination in the human transcriptome. *Genome Res.* **28**, 231–242 (2018).

26. Zhu, C. *et al.* Single-molecule, full-length transcript isoform sequencing reveals disease-associated RNA isoforms in cardiomyocytes. *Nat. Commun.* **12**, 4203 (2021).

27. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–233 (2008).

28. Ma, W. K. *et al.* ASO-based PKM Splice-switching Therapy Inhibits Hepatocellular Carcinoma Cell Growth. *Cancer Res.* 2020.09.01.278580 (2022) doi:10.1158/0008-5472.CAN-20-0948.

29. Clower, C. V. *et al.* The alternative splicing repressors hnRNP A1/A2 and PTB influence pyruvate kinase isoform expression and cell metabolism. *Proceedings of the National Academy of Sciences* **107**, 1894–1899 (2010).

30. Luo, W. *et al.* Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1. *Cell* **145**, 732–744 (2011).

31. Yang, W. *et al.* Nuclear PKM2 regulates β-catenin transactivation upon EGFR activation. *Nature* **480**, 118–122 (2011).

32. Traxler, L. *et al.* Warburg-like metabolic transformation underlies neuronal degeneration in sporadic Alzheimer’s disease. *Cell Metab.* **34**, 1248-1263.e6 (2022).

33. Liu, C., Liu, C. & Fu, R. Research progress on the role of PKM2 in the immune response. *Front. Immunol.* **13**, 936967 (2022).

34. Tang, Q. *et al.* Pyruvate kinase M2 regulates apoptosis of intestinal epithelial cells in Crohn’s disease. *Dig. Dis. Sci.* **60**, 393–404 (2015).

35. Rihan, M. & Sharma, S. S. Role of Pyruvate Kinase M2 (PKM2) in Cardiovascular Diseases. *J. Cardiovasc. Transl. Res.* (2022) doi:10.1007/s12265-022-10321-1.