

### Specific Aims:

**Rationale:** Splicing is a process where pre-mRNAs are processed to remove introns and stitch together exons. This process offers the potential to introduce tremendous diversity into the proteome by changing which sequences from the pre-mRNA are included as exons. Accordingly, more than 90% of human genes are alternatively spliced in different conditions, such as at different developmental stages or across cell types.[25] Errors in this process have been implicated in a wide variety of human diseases including familial dysautonomia, early onset Parkinson disease, and cancer.[20] Further recent efforts have produced drugs that modulate splicing outcomes to produce improvements in Spinal muscular Atrophy, Huntington’s Disease, Duchenne muscular dystrophy, and cancer.[17] Work to understand the mechanisms of splicing stands to benefit not only our understandings of development and differentiation but also our ability to target these processes with therapeutics. Much progress has been made using massively parallel splicing assays (MPSAs) to probe the mechanisms of splicing.[1–5, 7, 12, 14, 16, 18, 19, 21, 22, 28] These efforts have largely focused on understanding how the sequence characteristics of the pre-RNA dictate the inclusion or exclusion of a potential exon and have made great strides. However, while most alternative splicing events are believed to be local affecting an individual intron or exon, recent work has begun to show that correlated splicing decisions are far more common than was previously understood.[9, 23, 29] A particularly interesting case of correlated splicing is mutually exclusive exons (MXEs). MXEs are clusters of exons that are spliced such that every isoform includes exactly one exon from the cluster. A recent analysis of human RNA-seq data identified 629 MXE clusters, of which less than 25% had an identifiable mechanism for enforcing mutual exclusivity.[9] **I propose to develop a long read massively parallel splicing assay well suited for characterizing correlation in splicing. I will apply it to investigating MXEs with a focus on Pyruvate Kinase M (PKM).**

**Aim 1: Develop an MPSA protocol capable of handling more diverse isoforms.** Current MPSAs utilize Illumina technologies to read out the population of the various isoforms they detect.[1–5, 7, 12, 14, 16, 18, 19, 21, 22, 28] However, this decision puts strict length limits on reads that can be obtained and therefore restricts the isoforms that can be reliably detected. As mutations in MXE clusters can cause the production of isoforms that include part or all of the introns in addition to just inappropriate inclusion or exclusion of exons,[26] longer reads are needed to fully characterize mutated MXE clusters with their native introns. Further, a long read MPSA opens the door to studies of more complicated correlated splicing events like large MXE clusters[9] or long distance splicing correlations.[23, 24, 29]

*Sub-aim 1.1: Create an MPSA protocol utilizing nanopore sequencing technology.* I will redesign our library production protocols to maximize the throughput with which we can generate test libraries and the length and complexity of the libraries we can generate. I will trial overlap extension PCR, gibson assembly, and golden gate assemblies for single step assembly and barcoding of multi-component libraries. I will map barcodes to mutant sequences with long read technologies to avoid intermediate cloning steps that could introduce post mapping biases and complicate library production while enabling detection of spontaneous mutations in unexpected regions of the minigene or vector. I will adapt MPSA techniques honed by the Kinney lab [10, 28] to produce samples well suited to nanopore sequencing. I will develop and optimize protocols to prepare single-transcript very deep sequencing libraries for nanopore that minimize bias while maximizing read depth on our transcript of interest. I will actively work to detect and minimize sources of error or bias in both mutant library and sequencing library preparation to maximize the dynamic range of the assay, the number of variants that can be reliably measured, and the diversity of isoforms that can be reliably quantified.

*Sub-aim 1.2: Develop an analysis pipeline for quantifying isoform abundances from MPSA data.* Reliably calling low abundance RNA isoforms with single nucleotide resolution from nanopore reads faces challenges. These challenges stem from the lower depth and higher noise inherent in nanopore reads compared to short read technologies like Illumina. I will develop an analysis pipeline for robustly identifying and quantifying isoforms, including low abundance minor isoforms from nanopore data. Since our goal is to provide quantitative data on the isoform distribution of each variant in our study, including low abundance minor isoforms, it is important that this method has the power to detect low abundance isoforms and does not falsely identify isoforms. In order to test and optimize the new pipeline, we will write a tool to simulate nanopore reads given a known ground truth isoform distribution for each variant.

This will provide a tool for stress testing our analysis pipeline and identifying issues and biases so they can be resolved.

**Aim 2: Characterize the mechanism for mutual exclusivity in Pyruvate Kinase M (PKM) exons 9 and 10.** PKM is a central metabolic gene that controls the rate limiting step in glycolysis with two primary isoforms defined by the inclusion of either exon 9 or exon 10. The conversion between these two isoforms is a primary driver of the Warburg effect and is important in a wide variety of cancers.[6, 15] However, while much is known about the regulation of PKM splicing, the mechanism for enforcing exclusivity between exons 9 and 10 is still unknown.

*Sub-aim 2.1: Identify sequence elements necessary for mutual exclusivity in PKM splicing.* I will utilize PKM minigene constructs designed by the Krainer lab for studying PKM splicing [26] as a platform for screening for mutants that break the anti-correlation of MXE splicing in PKM. Any deviations from anti-correlation between MXEs will indicate that that mutant disrupted an element important for the maintenance of mutual exclusivity(MMX). Previous work has shown that MMX elements for PKM are located internal to exons 9 and 10,[26] so we will utilize saturation mutagenesis of these exons to identify these elements. We will couple these experiments with broader random mutagenesis strategies to map MMX elements that may also reside in the introns. In order measure how these effects depend on the exon 9 to exon 10 ratio, I will modulate splice ratios with ASOs[15, 27] or by engineering in risdiplam sensitive elements and using drug titration MPSA techniques[10]. Identified regions will be mapped at higher resolution with deep mutagenesis. I will perform the screen using the techniques developed in aim 1, however, if these techniques prove impractical I can use short read sequencing to gather the data instead without losing isoform resolution in the most important parts of the gene or pursue the same approach using lower throughput methods and rt-qPCR or rt-PCR band analysis approaches.

*Sub-aim 2.2: Mechanistically characterize sequence elements enforcing mutual exclusivity.* We will follow up identification of MMX elements in sub-aim 2.1 with mechanistic studies of the sequence elements that enforce mutual exclusivity. In many arthropods and the human ATE1 gene [8, 11, 13] MMXs are RNA secondary structural elements. We will use mutant cycle analysis to show dependence on structure for any identified secondary structural elements. We will also use RNA based pull downs as well as motif mutants coupled with RNAi knockdown and over-expression of RNA binding proteins (RBPs) to demonstrate the importance identified RBP binding motifs. The Krainer lab is adept at such techniques.[26] This data will be used to inform a mechanistic model of PKM MXE splicing.

## References

- [1] Scott I Adamson, Lijun Zhan, and Brenton R Graveley. “Vex-seq: high-throughput identification of the impact of genetic variation on pre-mRNA splicing efficiency”. en. In: *Genome Biol.* 19.1 (June 2018), p. 71. ISSN: 1465-6906. DOI: 10.1186/s13059-018-1437-x. URL: <http://dx.doi.org/10.1186/s13059-018-1437-x>.
- [2] Pablo Baeza-Centurion et al. “Combinatorial Genetics Reveals a Scaling Law for the Effects of Mutations on Splicing”. en. In: *Cell* 176.3 (Jan. 2019), 549–563.e23. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2018.12.010. URL: <http://dx.doi.org/10.1016/j.cell.2018.12.010>.
- [3] Pablo Baeza-Centurion et al. “Mutations primarily alter the inclusion of alternatively spliced exons”. en. In: *Elife* 9 (Oct. 2020). ISSN: 2050-084X. DOI: 10.7554/eLife.59959. URL: <http://dx.doi.org/10.7554/eLife.59959>.
- [4] Simon Braun et al. “Decoding a cancer-relevant splicing decision in the RON proto-oncogene using high-throughput mutagenesis”. en. In: *Nat. Commun.* 9.1 (Aug. 2018), pp. 1–18. ISSN: 2041-1723, 2041-1723. DOI: 10.1038/s41467-018-05748-7. URL: <https://www.nature.com/articles/s41467-018-05748-7>.
- [5] Rocky Cheung et al. “A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions”. en. In: *Mol. Cell* 73.1 (Jan. 2019), 183–194.e8. ISSN: 1097-2765, 1097-4164. DOI: 10.1016/j.molcel.2018.10.037. URL: <http://dx.doi.org/10.1016/j.molcel.2018.10.037>.
- [6] Heather R Christofk et al. “The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth”. en. In: *Nature* 452.7184 (Mar. 2008), pp. 230–233. ISSN: 0028-0836. DOI: 10.1038/nature06734. URL: <https://www.nature.com/articles/nature06734>.

- [7] Mariela Cortés-López et al. “High-throughput mutagenesis identifies mutations and RNA-binding proteins controlling CD19 splicing and CART-19 therapy resistance”. en. In: *Nat. Commun.* 13.1 (Sept. 2022), pp. 1–17. ISSN: 2041-1723, 2041-1723. DOI: 10.1038/s41467-022-31818-y. URL: <https://www.nature.com/articles/s41467-022-31818-y>.
- [8] Brenton R Graveley. “Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures”. en. In: *Cell* 123.1 (Oct. 2005), pp. 65–73. ISSN: 0092-8674. DOI: 10.1016/j.cell.2005.07.028. URL: <http://dx.doi.org/10.1016/j.cell.2005.07.028>.
- [9] Klas Hatje et al. “The landscape of human mutually exclusive splicing”. en. In: *Mol. Syst. Biol.* 13.12 (Dec. 2017), p. 959. ISSN: 1744-4292. DOI: 10.15252/msb.20177728. URL: <http://dx.doi.org/10.15252/msb.20177728>.
- [10] Yuma Ishigami et al. “Specificity, synergy, and mechanisms of splice-modifying drugs”. en. Dec. 2022. DOI: 10.1101/2022.12.30.522303. URL: <https://www.biorxiv.org/content/10.1101/2022.12.30.522303v1>.
- [11] Yongfeng Jin et al. “Mutually exclusive alternative splicing of pre-mRNAs”. en. In: *Wiley Interdiscip. Rev. RNA* 9.3 (May 2018), e1468. ISSN: 1757-7004, 1757-7012. DOI: 10.1002/wrna.1468. URL: <https://onlinelibrary.wiley.com/doi/10.1002/wrna.1468>.
- [12] Philippe Julien et al. “The complete local genotype–phenotype landscape for the alternative splicing of a human exon”. en. In: *Nat. Commun.* 7.1 (May 2016), pp. 1–8. ISSN: 2041-1723, 2041-1723. DOI: 10.1038/ncomms11558. URL: <https://www.nature.com/articles/ncomms11558>.
- [13] Marina Kalinina et al. “Multiple competing RNA structures dynamically control alternative splicing in the human ATE1 gene”. en. In: *Nucleic Acids Res.* 49.1 (Jan. 2021), pp. 479–490. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkaa1208. URL: <http://dx.doi.org/10.1093/nar/gkaa1208>.
- [14] Shengdong Ke et al. “Saturation mutagenesis reveals manifold determinants of exon definition”. en. In: *Genome Res.* 28.1 (Jan. 2018), pp. 11–24. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.219683.116. URL: <http://dx.doi.org/10.1101/gr.219683.116>.
- [15] Wai Kit Ma et al. “ASO-based PKM Splice-switching Therapy Inhibits Hepatocellular Carcinoma Cell Growth”. en. Mar. 2022. DOI: 10.1158/0008-5472.CAN-20-0948. URL: <https://repository.cshl.edu/id/eprint/40463/>.
- [16] Martin Mikl et al. “Dissecting splicing decisions and cell-to-cell variability with designed sequence libraries”. en. In: *Nat. Commun.* 10.1 (Oct. 2019), pp. 1–14. ISSN: 2041-1723, 2041-1723. DOI: 10.1038/s41467-019-12642-3. URL: <https://www.nature.com/articles/s41467-019-12642-3>.
- [17] Christopher R Neil et al. “Reprogramming RNA processing: an emerging therapeutic landscape”. en. In: *Trends Pharmacol. Sci.* 43.5 (May 2022), pp. 437–454. ISSN: 0165-6147. DOI: 10.1016/j.tips.2022.02.011. URL: [https://www.cell.com/trends/pharmacological-sciences/fulltext/S0165-6147\(22\)00046-3](https://www.cell.com/trends/pharmacological-sciences/fulltext/S0165-6147(22)00046-3).
- [18] Alexander B Rosenberg et al. “Learning the sequence determinants of alternative splicing from millions of random sequences”. en. In: *Cell* 163.3 (Oct. 2015), pp. 698–711. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2015.09.054. URL: <http://dx.doi.org/10.1016/j.cell.2015.09.054>.
- [19] Dvir Schirman et al. “A broad analysis of splicing regulation in yeast using a large library of synthetic introns”. en. In: *PLoS Genet.* 17.9 (Sept. 2021), e1009805. ISSN: 1553-7390, 1553-7404. DOI: 10.1371/journal.pgen.1009805. URL: <http://dx.doi.org/10.1371/journal.pgen.1009805>.
- [20] Marina M Scotti and Maurice S Swanson. “RNA mis-splicing in disease”. en. In: *Nat. Rev. Genet.* 17.1 (Nov. 2015), pp. 19–32. ISSN: 1471-0056. DOI: 10.1038/nrg.2015.3. URL: <https://www.nature.com/articles/nrg.2015.3>.
- [21] Rachel Soemedi et al. “Pathogenic variants that alter protein code often disrupt splicing”. en. In: *Nat. Genet.* 49.6 (June 2017), pp. 848–855. ISSN: 1061-4036, 1546-1718. DOI: 10.1038/ng.3837. URL: <http://dx.doi.org/10.1038/ng.3837>.

- [22] Přemysl Souček et al. “High-throughput analysis revealed mutations’ diverging effects on SMN1 exon 7 splicing”. en. In: *RNA Biol.* 16.10 (Oct. 2019), pp. 1364–1376. ISSN: 1547-6286, 1555-8584. DOI: 10.1080/15476286.2019.1630796. URL: <http://dx.doi.org/10.1080/15476286.2019.1630796>.
- [23] Hagen Tilgner et al. “Comprehensive transcriptome analysis using synthetic long-read sequencing reveals molecular co-association of distant splicing events”. en. In: *Nat. Biotechnol.* 33.7 (May 2015), pp. 736–742. ISSN: 1087-0156. DOI: 10.1038/nbt.3242. URL: <https://www.nature.com/articles/nbt.3242>.
- [24] Hagen Tilgner et al. “Microfluidic isoform sequencing shows widespread splicing coordination in the human transcriptome”. en. In: *Genome Res.* 28.2 (Feb. 2018), pp. 231–242. ISSN: 1088-9051, 1549-5469. DOI: 10.1101/gr.230516.117. URL: <http://dx.doi.org/10.1101/gr.230516.117>.
- [25] Eric T Wang et al. “Alternative isoform regulation in human tissue transcriptomes”. en. In: *Nature* 456.7221 (Nov. 2008), pp. 470–476. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature07509. URL: <http://dx.doi.org/10.1038/nature07509>.
- [26] Z Wang et al. “Exon-centric regulation of pyruvate kinase M alternative splicing via mutually exclusive exons”. In: *Journal of molecular* (2012). URL: <https://academic.oup.com/jmcb/article-abstract/4/2/79/908162>.
- [27] Zhenxun Wang et al. “Manipulation of PK-M mutually exclusive alternative splicing by antisense oligonucleotides”. en. In: *Open Biol.* 2.10 (Oct. 2012), p. 120133. ISSN: 2046-2441. DOI: 10.1098/rsob.120133. URL: <http://dx.doi.org/10.1098/rsob.120133>.
- [28] Mandy S Wong, Justin B Kinney, and Adrian R Krainer. “Quantitative Activity Profile and Context Dependence of All Human 5’ Splice Sites”. en. In: *Mol. Cell* 71.6 (Sept. 2018), 1012–1026.e3. ISSN: 1097-2765, 1097-4164. DOI: 10.1016/j.molcel.2018.07.033. URL: <http://dx.doi.org/10.1016/j.molcel.2018.07.033>.
- [29] Chenchen Zhu et al. “Single-molecule, full-length transcript isoform sequencing reveals disease-associated RNA isoforms in cardiomyocytes”. en. In: *Nat. Commun.* 12.1 (July 2021), p. 4203. ISSN: 2041-1723. DOI: 10.1038/s41467-021-24484-z. URL: <http://dx.doi.org/10.1038/s41467-021-24484-z>.