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.[1] Errors in splicing are involved in a wide variety of human diseases including familial dysautonomia, early onset Parkinson disease, and cancer.[2] Recent efforts have produced drugs that modulate splicing outcomes to treat Spinal muscular atrophy, Huntington's Disease, Duchenne muscular dystrophy, and cancer.[3] Improving our understanding of splicing will improve our understanding of development and differentiation and 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.[4–17] However, these efforts have utilized Illumina technologies which restricts the isoforms that can be reliably detected. These issues have been overcome by studying systems shrunk to fit an Illumina read or by ignoring unexpected isoforms. Yet the median human intron is 1.7 kb in length[18] far longer than an Illumina read and perturbations can cause complicated changes in splicing outcomes.[8,19,20] Further, these MPSAs have so far focused on the decision to include or exclude a single exon. However, recent work has begun to show that correlated splicing decisions are far more common than was previously understood.[21–24] 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 of which less than 25% had an identifiable mechanism for maintaining mutual exclusivity (MMX).[23] **In order to overcome these issues, I propose to develop a long read massively parallel splicing assay (LR-MPSA) capable of handling correlated splicing, full size introns, and complicated isoform distributions. I will apply this assay to investigating MXEs with a focus on Pyruvate Kinase M (PKM).**

Aim 1: Develop an MPSA protocol capable of handling more diverse isoforms.
Sub-aim 1.1: Create an MPSA protocol utilizing nanopore sequencing technology. I will develop an LR-MPSA optimized for characterizing the mechanisms of complex splicing decisions in the context of full native introns and all isoform outputs. I will adapt MPSA techniques honed by the Kinney lab[17,25] by creating new sample and library preparation pipelines optimized unbiased isoform enrichment and nanopore sequencing.
Sub-aim 1.2: Develop an analysis pipeline for quantifying isoform abundances from MPSA data. Calling low abundance RNA isoforms with single nucleotide resolution from nanopore reads faces challenges from the low depth and high noise. I will develop an analysis pipeline for identifying and quantifying isoforms optimizing for power to detect low abundance isoforms and correct isoform assignment. I will write a tool to simulate nanopore reads given a known ground truth isoform distribution for each variant to test and optimize the new pipeline.
Aim 2: Characterize the mechanism for mutual exclusivity in Pyruvate Kinase M (PKM) exons 9 and 10. PKM catalyzes the rate limiting step in glycolysis and has 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.[26,27] However, the mechanism for MMX between exons 9 and 10 is unknown.
Sub-aim 2.1: Identify sequence elements necessary for mutual exclusivity in PKM splicing. I will use PKM minigene constructs designed by the Krainer lab[19] to screen for mutants that loose MMX. I will use mutant scans over exons 9 and 10 to find MMX elements because previous work has shown that regulatory sequences are located in these exons[19] broader random mutagenesis will map elements in introns. MMX regions will be mapped at higher resolution with deep mutagenesis. If the nanopore screening technique fails, I will use short-read or low throughput methods for screening.
Sub-aim 2.2: Mechanistically characterize sequence elements enforcing mutual exclusivity. I will follow up identification of MMX elements with mechanistic studies. One hypothesis for the identity of new MMX elements is RNA secondary structural elements. I will use mutant cycle analysis to show dependence on structure for any identified secondary structural elements. Another hypothesis is that MMX elements are motifs for RNA binding proteins (RBPs). I will use pull downs as well as motif mutants coupled with RNAi knockdown and over-expression of RBPs to demonstrate the importance of RBP binding motifs.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] Allison Piovesan et al. "Human protein-coding genes and gene feature statistics in 2019". en. In: *BMC Res. Notes* 12.1 (June 2019), p. 315. ISSN: 1756-0500. DOI: 10.1186/s13104-019-4343-8. URL: http://dx.doi.org/10.1186/s13104-019-4343-8.

[19] 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.

[20] Melina Mathur et al. "Programmable mutually exclusive alternative splicing for generating RNA and protein diversity". en. In: *Nat. Commun.* 10.1 (June 2019), p. 2673. ISSN: 2041-1723. DOI: 10.1038/s41467-019-10403-w. URL: http://dx.doi.org/10.1038/s41467-019-10403-w.

[21] 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.

[22] 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.

[23] 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.

[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] 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.

[26] 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.

[27] 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/.