**Significance**

Alternative splicing is important for both understanding fundamental aspects of human biology and treating a wide array of diseases. Splicing is a process where introns are removed from pre-mRNAs and the exons are stitched together. This 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 However, despite these recent advances, mechanisms of splicing regulation are still poorly understood. Improving our understanding of splicing will deepen our knowledge of development and differentiation and our ability to target these processes with therapeutics.

A particularly interesting type of alternative splicing is mutually exclusive exons. Mutually exclusive exons are clusters of exons whose splicing is anticorrelated such that one exon from the cluster is selected for inclusion 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.4 Furthermore, mutations in mutually exclusive exons or changes in their splicing are involved with diseases including Timothy syndrome, cancer, heart disease, and cardiomyopathy.4 In fact, mutually exclusive exons are twice as likely to be associated with a parthenogenic SNP in clinvar.5 These SNPs are most frequently associated with neurological diseases, neuromuscular diseases, cardiomyopathies, or cancer.5 While mutually exclusive exon splicing has long been thought to be uncommon, recent studies have uncovered far more instances that were previously thought.5 Research on mutually exclusive exons has uncovered several mechanisms that drive mutual exclusivity, including Spliceosome incompatibility, steric hindrance, and RNA secondary structure.4 However, despite these advances, more than 75% of human mutually exclusive exon clusters have no known mechanism.5

Pyruvate kinase M (PKM) is a model for mutually exclusive exon splicing with deep disease relevance. This gene has 2 primary isoforms characterized by a switch between inclusion of either exon 9 (PKM1) or 10 (PKM2).6 PKM1 is constitutively active and is the dominant form in energy hungry tissues like muscle and brain, while PKM2 is allosterically regulated by Fructose-1,6-bisphosphate and is dominant in proliferating tissues.7–11 The allosteric regulation of PKM2 causes it to transition from a tetrameric form which is competent as a pyruvate kinase to a dimeric form that is inactive as a pyruvate kinase.8 This dimeric form however, gains activity as a protein kinase and translocates into the nuclease where it phosphorylates stat3 upregulating mek and driving proliferation.12 Cancer is associated with increases in the expression of the PKM2 isoform, and with increased levels and nuclear localization of its dimeric form specifically.7–10,12 The presence of PKM2 instead of PKM1 is necessary for the Warburg effect, and is important for tumorigenesis.9,10,13

1. PKM is particularly important
   1. ~~Differentially regulated across cell types~~
   2. ~~Can translocate to the nucleus and act as a regulator~~
   3. ~~Driver of the Warburg effect~~
   4. Has a role in cardiovascular disease, Alzheimer’s, inflammation, and immune function
2. Current MPSAs have shown promise but are deeply flawed
   1. MPSAs allow rapid characterization of splicing mechanisms
   2. Dissection of MPSA techniques
      1. Known isoforms only
         1. Sort-seq
            1. Nonnative context
            2. Low isoform resolution
            3. Low PSI resolution
         2. PCR-seq
            1. Low isoform resolution
            2. PSI resolution limited by isoform collapse
         3. Gel cut seq
            1. Low isoform resolution
            2. Assumes RNA levels are even
      2. All isoforms
         1. Junction sequencing
            1. Can only differentiate isoforms below read length

Requires non native or small system

Cannot distinguish isoforms with differences beyond the read length

* + - * 1. PCR or RT may cause crossover noise
        2. Interior primer sites may miss isoforms that loose those sites
  1. These techniques are not good fits for large systems or for differentiating isoforms that maybe large
  2. Native context is important
  3. Analysis pipelines have not been made as publicly available software reducing acceptance and spread of the methods

1. How my study fits
   1. Long reads will allow junction sequencing over larger constructs
      1. Native contexts
      2. More complicated situations
         1. MXEs
         2. Cryptic sites
         3. Intron retention
      3. Reduced count noise
   2. Open-source software will democratize the method
      1. Allow robust verification of analysis pipelines on synthetic datasets
      2. Allow experimentalists to go directly from reads to processed data
   3. Applications of these methods to PKM will help understand it’s complicated regulation
      1. Interesting basic biology
      2. Disease relevance
      3. Form a good test ground for the techniques

Much progress has been made using massively parallel splicing assays (MPSAs) to probe the mechanisms of splicing.\supercite{Ke2018-af, Julien2016-wa, Adamson2018-va, Soemedi2017-pz, Cortes-Lopez2022-gy, Schirman2021-ss, Mikl2019-ng, Braun2018-mb, Soucek2019-iq, Baeza-Centurion2020-tn, Cheung2019-ah, Baeza-Centurion2019-hz, Rosenberg2015-zs, Wong2018-vq}

These efforts have utilized Illumina technology which cannot produce reads capable of covering the whole length of most genes, greatly affecting the isoforms that can be detected.

This issue has 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\supercite{Piovesan2019-rp} far longer than an Illumina read and perturbations can cause complicated changes in splicing outcomes.\supercite{Cortes-Lopez2022-gy,Wang2012-dr,Mathur2019-hy}

Further, 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 across exons are far more common than was previously understood.\supercite{Zhu2021-fs, Tilgner2015-sb, Hatje2017-oj,Tilgner2018-jo}

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).\supercite{Hatje2017-oj}

\textbf{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.}

\textbf{I will apply this assay to investigating MXEs with a focus on Pyruvate Kinase M (PKM).}

**Approach**

1. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

2. Scotti, M. M. & Swanson, M. S. RNA mis-splicing in disease. *Nat. Rev. Genet.* **17**, 19–32 (2015).

3. Neil, C. R. *et al.* Reprogramming RNA processing: an emerging therapeutic landscape. *Trends Pharmacol. Sci.* **43**, 437–454 (2022).

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

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

6. Noguchi, T., Inoue, H. & Tanaka, T. The M1- and M2-type isozymes of rat pyruvate kinase are produced from the same gene by alternative RNA splicing. *J. Biol. Chem.* **261**, 13807–13812 (1986).

7. Taniguchi, K. *et al.* Organ-specific PTB1-associated microRNAs determine expression of pyruvate kinase isoforms. *Sci. Rep.* **5**, 1–8 (2015).

8. Mazurek, S., Boschek, C. B., Hugo, F. & Eigenbrodt, E. Pyruvate kinase type M2 and its role in tumor growth and spreading. *Semin. Cancer Biol.* **15**, 300–308 (2005).

9. 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).

10. 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).

11. Jurica, M. S. *et al.* The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. *Structure* **6**, 195–210 (1998).

12. Gao, X., Wang, H., Yang, J. J., Liu, X. & Liu, Z.-R. Pyruvate kinase M2 regulates gene transcription by acting as a protein kinase. *Mol. Cell* **45**, 598–609 (2012).

13. 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.