**Significance**

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.

1. ~~Mechanisms of alternative splicing are important~~
   1. ~~Alternative splicing is very prevalent (>90% of human genes)~~
   2. ~~Targeting alternate splicing offers the potential for targeting diseases~~
      1. ~~SMN2, Dystrophin etc~~
   3. ~~Alternative splicing regulation is still poorly understood~~
2. Mutually exclusive splicing is important
   1. MXEs are important in many processes
   2. MXE mutations drive disease
   3. MXEs are enriched for disease causing SNPs
   4. Mutually exclusive splicing was once though uncommon
   5. Recent work has shown it is much more common than previously thought
   6. Mechanisms that are known for MXEs
   7. Most MXEs still have no known mechanism
3. 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
4. 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).