# Specific Aims:

Alternative splicing of mutually exclusive exons is important but poorly understood. Alternative splicing is a significant factor in gene regulation. However, the regulation of alternative splicing is still incompletely understood. Splicing outcomes can correlate over long distances. A common example of this phenomenon, mutually exclusive exon splicing, occurs when only one exon from a cluster is included in the mature transcript. These long-distance anti-correlations present challenges as a study system because they require measuring multiple exons simultaneously in a single transcript. While most mutually exclusive exon clusters in humans still have no known mechanism, several have been identified, including spliceosome incompatibility, 5’-splice site/branch point steric occlusion, and RNA secondary structures. Further, work on the function of mutually exclusive exons has borne fruit. A common theme is driving switch-like changes in protein function that occur in cell differentiation or through development. These include a Pyruvate Kinase M (PKM) isoform change when cells differentiate that is reverted in cancer causing the Warburg effect. Another example is ketohexokinase, aberrant activation of one isoform causes pathological cardiac hypertrophy but activation of the other is a major driver of hepatocellular carcinoma. Finally, mutations alter an exon switch in the calcium channel CaV1.2 produce timothy syndrome characterized by developmental abnormalities, neurological symptoms, cardiac issues and a life expectancy of only 2.5 years. Understanding the mechanisms and regulation of mutually exclusive splicing offers the potential not only to deepen our knowledge of fundamental processes important for development and differentiation but also to provide new tools we can use to target splice modifying treatments towards a wide variety of human diseases.

Massively parallel splicing assays (MPSAs) provide a method for interrogating the mechanisms of alternative splicing but have technical limitations. There are several different types of MPSA, but fundamentally they all rely on introducing a library of variant mini-genes and using short read sequencing to measure how each variant is spliced. These techniques have great promise for dissecting splicing regulation by enabling quantitative phenotype measurement of large numbers of mutants in parallel. However, because these methods use short-read sequencing they cannot confidently quantify all isoforms of minigenes which are longer than a short read. This is especially problematic for measuring exons in the context of their native introns, for measuring clusters of mutually exclusive exons, and for distinguishing isoforms where cryptic splice sites or intron retention events play a role. These issues can also lead to noisy measurements and limit the dynamic range of the assay by collapsing multiple isoforms which cannot be distinguished using only short-read based methods.

I propose to develop tools for studying complicated splicing regulation. These tools are a long read based MPSA (LR-MPSA) and analysis software for nucleotide resolution MPSA data. While these tools will be broadly useful for understanding a diverse set of splicing decisions, here I will apply them to dissecting the mechanisms of mutually exclusive exon splicing in the specific context of pyruvate kinase M (PKM). I will accomplish these goals in three complementary aims.

**Aim 1: Develop a nucleotide-resolution MPSA using long read sequencing***.* 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 by creating new sample and library preparation pipelines optimized for unbiased isoform enrichment and nanopore sequencing.

**Aim 2: Develop software for analyzing data from nucleotide-resolution MPSAs***.* Calling low abundance RNA isoforms from nanopore reads faces challenges from the low depth and high noise. I will develop analysis software for identifying and quantifying isoforms that optimizes for power to detect low abundance isoforms and minimizes incorrect isoform assignment. I will simulate nanopore reads given a known ground truth isoform distribution for each variant to test and optimize the software at various depths and error profiles. I will make this software publicly available and open-source.

# Aim 3: Define the mechanism of mutually exclusive exon inclusion in PKM. PKM has a pair of mutually exclusive exons and switching between them drives the Warburg effect, this splicing event is a target for cancer therapies. PKM mutually exclusive exon splicing causes a shift from respiration to fermentation that is a driver of cancer. I will use LR-MPSAs to identify motifs that are important for maintaining mutual exclusivity in PKM mutually exclusive exons by identifying mutants that deviate from mutually exclusive exon behavior. I will verify identified motifs with anti-sense oligo based blocking and low throughput experiments, then model their behavior.

# My background positions me well to execute this proposal, yet this proposal offers a training opportunity that will help me found an independent lab. In my doctorate I used massively parallel assays and computational modeling in microbiology, protein engineering, and CRISPR tool development providing me with the foundational skills I will need for this proposal. This proposal will allow me to explore new applications of these skills in a field with opportunities to do fundamental science with applications in human health. While I pursue these scientific goals, I will also be focused on utilizing my position at Cold Spring Harbor lab to gain wider training and mentorship that will help launch my career as an independent researcher. I will attend CSHL meetings and Gordon conferences on RNA processing and quantitative methods, attend lab meetings and journal clubs in the Kinney, Krainer, Koo and McCandlish labs, attend CSHL grant writing and professional development courses, and hone my skill at lecturing through teaching opportunities.

Together, these research and training opportunities will position me to launch an independent research career focusing on applying massively parallel assays and modeling to understanding deep biological questions in RNA processing.