# Specific Aims:

Mutually exclusive alternative splicing of exons is fundamentally and therapeutically important but poorly understood. Alternative splicing is highly important in gene regulation. However, while there has been a lot of research on the regulation of splicing, these processes are still incompletely understood. In some cases, splicing decisions can correlate over long distances. A particularly common version of this phenomenon known as mutually exclusive exon splicing occurs when exon inclusion is anticorrelated within a cluster of exons such that only one exon from the cluster is ever included in a mature isoform. Mutually exclusive exons can be particularly challenging to study because in order to observe them, you need a sequencing strategy which is capable of measuring the inclusion or exclusion of multiple exons simultaneously. However, mutually exclusive exons have been identified that are driven by spliceosome incompatibility, 5’-splice site/branch point steric incompatibility, or RNA secondary structures however most mutually exclusive exon clusters in humans have no known mechanism. Recent work has found far more mutually exclusive exon clusters in the human genome than were previously suspected. Mutually exclusive exons are important for driving switch-like changes in protein function that occur in cell differentiation or through development. These include the exon switch in Pyruvate Kinase M (PKM) that occurs when cells differentiate and is reverted in cancer to cause aerobic fermentation and the Warburg effect. Another example is an exon switch in ketohexokinase that regulates fructose metabolism, activating fructose metabolism through this mechanism in the heart is important for pathological cardiac hypertrophy but deactivating it in liver is a major driver of hepatocellular carcinoma. Finally, mutations that cause a delay in a developmental exon switch in the calcium channel CaV1.2 produce the autosomal dominant genetic disease 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 enabeling 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 a long read based MPSA (LR-MPSA), to create LR-MPSA analysis software, and apply these to dissecting the mechanisms of mutually exclusive exon splicing in the specific context of pyruvate kinase M (PKM). I will accomplish this 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.

I am interested in massively parallel experimental and computational methods to ask biological questions. In my doctorate, I applied these lenses to microbiology, protein engineering, and CRISPR tool development. Splicing is perfectly suited to this approach, with complex regulation amenable to MPSAs and quantitative modeling. By applying my skills to splicing, I will learn new applications of these techniques in a field with deep fundamental and applied questions. This will provide skills I need to begin my own independent quantitative biology lab. Concurrent with this research, I propose a directed effort to gain training and mentorship for my future career as an independent scientist. 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.