Specific Aims:

Background Alternative splicing is of fundamental importance to gene regulation, but the mechanisms that control it are incompletely understood. Massively parallel splicing assays (MPSAs) provide a promising method for understanding these mechanisms, but current MPSAs have major technical limitations that stem from their reliance on short read sequencing. A particularly interesting case of alternative splicing that is poorly understood in the human context is mutually exclusive exons (MXEs). MXEs are difficult to study with modern MPSA techniques because they involve splicing decisions that occur over large distances difficult that cannot be fully characterized with only short reads. Understanding the mechanisms that create MXEs will both improve our understanding of fundamental processes in human biology and guide development of splice modifying therapies for a wide variety of human diseases. Two particularly interesting examples of MXE clusters in the human context are pyruvate kinase M (PKM) and complement decay-accelerating factor (CD55). PKM has a pair of MXEs and the switch between them drives the warburg effect, making this splicing event a potential target for cancer therapeutics. CD55 is involved in immune self tolerance and has been implicated in diseases like CHAPEL syndrome, paroxymal nocturnal hemoglobinuria, viral infection, and malaria. It has a cassette of 5 exons that can be entirely skipped or included as MXEs. The mechanisms by which both of these systems maintain mutual exclusivity have not been shown. Further, there is no known mechanisms allowing clusters of more than 2 MXEs in humans, making the study of CD55's cluster of 5 MXEs particularly interesting.

Rational I am a quantitative biologist with an interest in using massively parallel experimental and computational methods to understand biological problems. In my doctoral research, I applied these lenses to problems in microbiology, protein engineering, and CRISPR tool development. Splicing regulation is a problem that is perfectly suited to this view point, with complex regulation that is amenable to MPSA techniques and quantitative modeling. By applying my skills in quantitative and parallel techniques to splicing, I hope to learn new applications of these skills in a field where there is a lot of opportunity for deep mechanistic exploration. This will help me to gain skills I will need in my pursuit of beginning my own independent quantitative biology lab.

Proposal Here I propose to develop an MPSA which uses long read sequencing (LR-MPSA) to overcome the limitations of previous methods and apply it to dissecting the mechanisms of MXE alternative splicing in PKM and CD55. I will accomplish this in two complementary aims, each with 2 sub-aims.

Aim 1: Develop an MPSA protocol capable of handling the complexity of natural splicing.

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 by creating new sample and library preparation pipelines optimized for unbiased isoform enrichment and nanopore sequencing.

Sub-aim 1.2: Develop analysis software for quantifying isoform abundances from LR-MPSA data. 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 2: Characterize the mechanism for mutual exclusivity in PKM and CD55.

Sub-aim 2.1: Identify mechanisms of mutual exclusivity in PKM splicing. PKM MXE splicing drives a shift from respiration to fermentative metabolism that is a key driver of cancer. I will use the LR-MPSA method to identify motifs that are important for maintaining mutual exclusivity in PKM MXEs. I will focus on high density exon mutant scans and low density intron mutant scans and identify mutations that produce deviations from mutual exclusivity. I will verify identified motifs with anti-sense oligo based blocking and low throughput experiments. I will produce models of MXE behavior in this system.

Sub-aim 2.2: Identify mechanisms of mutual exclusivity in CD55 splicing. CD55 has a large cluster of MXEs that can also be skipped entirely. Dissecting the mechanisms that regulate this behavior will greatly deepen our understanding of splicing and may improve our understanding of CD55 dependent diseases. I will use the LR-MPSA method to identify motifs that are important for maintaining mutual exclusivity in CD55 MXEs. I will produce first coarse mutagenesis libraries and then fine grained libraries to narrow in on important motifs. I will verify identified motifs with anti-sense oligo based blocking and low throughput experiments. I will

produce models of the more complicated MXE behavior in this system.

Training proposal While I perform this research, I also propose to undergo 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.

Career preparation Together, this research and these 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.