Specific Aims:

Background Alternative splicing is of fundamental importance to gene regulation, but the behind it are incompletely understood. Massively parallel splicing assays (MPSAs) provide a promising method for understanding these mechanisms, but have technical limitations because they rely on short read sequencing. An interesting but poorly understood case of alternative splicing human mutually exclusive exons (MXEs). MXEs are difficult to study with current MPSAs because the correlated decisions occur over distances longer than short reads. Understanding the mechanisms that create MXEs will both improve our understanding of fundamental processes and guide development of splice modifying therapies for human diseases. Two interesting MXE clusters are pyruvate kinase M (PKM) and complement decay-accelerating factor (CD55). PKM has a pair of MXEs and switching between them drives the warburg effect, this splicing event is a target for cancer therapies. CD55 is involved in immune self tolerance and implicated in CHAPEL syndrome, paroxysmal nocturnal hemoglobinuria, viral infection, malaria, and cancer. It has a cassette of 5 exons that can be entirely skipped or included as MXEs. CD55's cluster of 5 MXEs is particularly interesting as there is no known mechanisms allowing clusters of more than 2 MXEs in humans.

Rational 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.

Proposal I propose to develop a long read based MPSA (LR-MPSA) and use it to dissect mechanisms of MXE 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 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 MXEs by identifying mutants that deviate from MXE behavior. I will verify identified motifs with anti-sense oligo based blocking and low throughput experiments, then model their behavior.

Sub-aim 2.2: Identify mechanisms of mutual exclusivity in CD55 splicing. CD55 has a large cluster of MXEs that can be skipped entirely. Dissecting the mechanism of this regulation will deepen our understanding of splicing and CD55 dependent diseases. I will use the LR-MPSA method to identify motifs that are important for maintaining mutual exclusivity in CD55 MXEs with 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, then model their behavior.

Training proposal 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.

Career preparation 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.