Final Report - Yang Li, PhD, University of Chicago
2018 CRF Young Investigator Award
Using Recurrent Aberrant mRNA isoforms for Early Cancer Detection

We originally proposed to identify recurrent cancer-specific RNA splice isoforms as biomarkers for early cancer detection. This led us to focus on the splice targets of SF3B1 as potential diagnostic targets, as SF3B1 is most commonly recurrently mutated splicing factor in cancers, with particular prevalence in hematological malignancies[1]. However, we reasoned that the sensitivity of RNA splicing measurements to detect SF3B1-mutated tumors is limited by the allelic frequency of mutations, and routine targeted DNA sequencing already has the ability to detect low frequency (<5%) somatic mutations[2] in recurrently mutated genes like SF3B1. We therefore developed an assay to detect SF3B1-mediated mis-spliced isoforms with the intention of aiding in interpretation of SF3B1 mutations which may be detected by routine targeted DNA sequencing but have unknown significance with regards to splicing. Though, the COVID-19 pandemic has limited our capacity to complete the wet-lab aspects of proposal to create RNA-splicing-based assays. Therefore, since the COVID-19 pandemic, we have recently focused on utilizing published datasets to better understanding how different SF3B1 mutations cause unique RNA splicing signatures. Our reasoning is that these unique RNA splicing signatures may have some diagnostic relevance. In addition, this approach will more generally aid in our understanding of SF3B1 biology and the identification of the mis-spliced genes which mediate disease progression in different contexts.

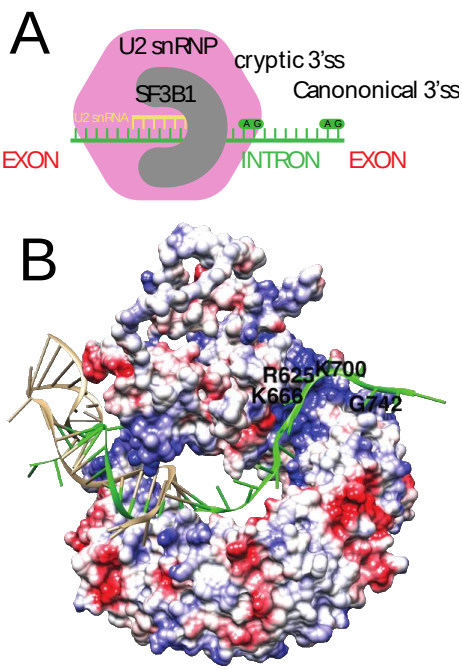

Figure 1 - (A) SF3B1 recognizes the branchsite element of introns. Cancer-associated SF3B1 mutations induce activation of cryptic upstream 3' ss. (B) CryoEM structure of the SF3B1, in complex with pre-mRNA (green) and U2 snRNA (beige). Labelled cancer-associated mutated residues cluster at a positively charged (blue) surface that interacts with pre-mRNA downstream of the branchsite.

SF3B1 is a component of the U2 snRNP complex which recognizes the essential branch site position during spliceosome assembly on the intron. Previous studies have revealed that the cancer-associated recurrent mutations in SF3B1 induce activation of cryptic 3' splice sites (3'ss) 10-30 bases upstream of the canonical 3' splice site, leading to misregulation of hundreds of genes (Fig. 1A)[1,3]. The detailed molecular mechanisms for activation of upstream 3'ss in these mutants remains unknown. Interestingly, while various recurrent SF3B1 mutations physically cluster along the same molecular surface[4] (Fig. 1B) and induce similar molecular splicing defects[1], the relative frequency of these SF3B1 mutations varies in different cancer types (Fig. 2). This suggests some functional differences between these mutations. For example, while the K700E allele is present in many cancers – both solid tumors and hematological malignancies like chronic leukocytic leukemia (CLL), acute myeloid leukemia (AML), and myelodysplastic syndrome (MDS) - the G742D allele seems to occur very specifically in CLL tumors[5] (Fig2). We hypothesize that these cancer-associated SF3B1 alleles are different, leading to mis-regulation of different sets of genes which push cells towards different disease phenotypes. Characterization of these differences may help better understand the molecular mechanisms of SF3B1-induced aberrant splicing.

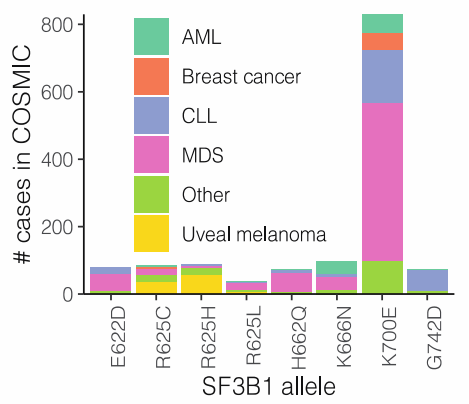

Figure 2 - Prevalence of different recurrent SF3B1 mutations, stratified by cancer type in Catalogue of Somatic Mutations in Cancer (COMSIC)

To address this question, we utilized published patient-derived transcriptomic data (RNA-seq) on the K700E alleles and G742D alleles from bone marrow extracts from CLL patients[3,6]. We specifically chose these alleles to focus because of the availability of cell-type matched transcriptomic data to compare these alleles without confounding factors. Through a standardized analysis pipeline, we compared the genome-wide splicing patterns of both alleles, comparing them to CLL samples without SF3B1 mutations to identify mis-spliced introns. We identified hundreds of significantly mis-spliced introns for each allele, including many introns that are specifically detected as mis-spliced in K700E or G742D but not both (Fig. 3A). We are currently analyzing the features which make introns uniquely susceptible to one allele or the other. For example, we first asked if both alleles share the same profile of cryptic 3'ss activation, relative to canonical 3'ss. Consistent with previous findings, we find that both alleles have generally similar splicing phenotypes, including strong enrichment for activation of 3'ss 20-30bp upstream of the canonical 3'ss (Fig3B). Though the profile of cryptic 3'ss positions is not statistically different between the two alleles, we are continuing to explore other features that may highlight differences between the two alleles, such as branchpoint positioning or sequence motifs.

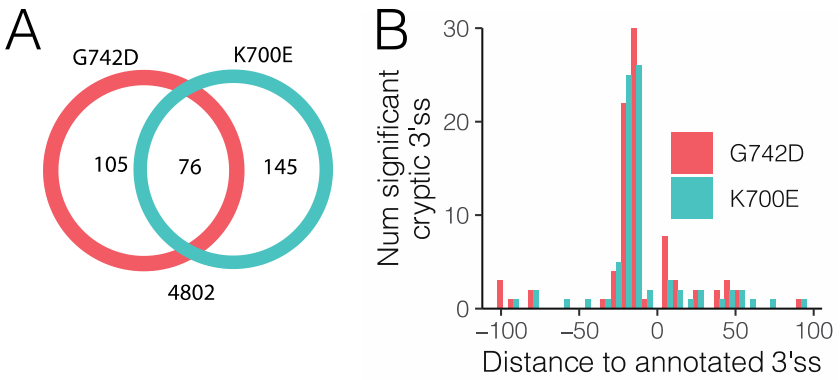

Figure 3 - (A) overlap of significantly mis-spliced intron regions (FDR<0.05) when comparing each mutant SF3B1 allele to matched wildtype. The high degree of overlap implies similar splicing phenotypes (OddsRatio=23.2, P<2x10^{-16}, hypergeometric test), though many allele-specific effects exist. (B) The positional distribution of 3'ss of significantly upregulated cryptic 3'ss, relative to annotated canonical 3'ss in each SF3B1 mutant allele. These differences in these distributions are not statistically meaningful.

Furthermore, our transcriptomic analysis may provide a prioritized list of candidate genes that mediate the effects of SF3B1 specifically in CLL. A similar transcriptomic analysis approach identified the gene BRD9 as a key mis-spliced gene in SF3B1 mutants across multiple cancer types, leading to the development of BRD9 targeted anti-sense oligo therapies to correct BRD9 mis-splicing[7]. Though we find that BRD9 is similarly mis-spliced and similarly mis-expressed in both of these SF3B1 alleles, we believe that our identification of allele-specific mis-spliced and mis-expressed genes may help prioritize which genes mediate the effects of SF3B1-mutants on CLL disease progression, as the unique effects of the G742D allele appears to be CLL-specific. Over the next year we will continue our work on contrasting these two SF3B1 alleles.

1. Agrawal, A. A., Yu, L., Smith, P. G. & Buonamici, S. Targeting splicing abnormalities in cancer. *Current Opinion in Genetics & Development* **48**, 67–74 (2018).
2. Surrey, L. F. *et al.* Clinical utility of custom-designed NGS panel testing in pediatric tumors. *Genome Med* **11**, 32 (2019).
3. Darman, R. B. *et al.* Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3′ Splice Site Selection through Use of a Different Branch Point. *Cell Reports* **13**, 1033–1045 (2015).
4. Zhang, X. *et al.* Structure of the human activated spliceosome in three conformational states. *Cell Res* **28**, 307–322 (2018).
5. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research* **47**, D941–D947 (2019).
6. Agrawal, A. A. *et al.* Novel SF3B1 in-frame deletions result in aberrant RNA splicing in CLL patients. *Blood Advances* **1**, 995–1000 (2017).
7. Inoue, D. *et al.* Spliceosomal disruption of the non-canonical BAF complex in cancer. *Nature* **574**, 432–436 (2019).