**Characterization of splicing-derived neoantigens using splicemutr shows their independence from TMB and potential as a biomarker for immunotherapy response**

Theron Palmer, Elana Fertig

February 09, 2022

Alternative-splicing dysregulation has been shown to play a role in cancer oncogenesis, conferring survival benefits to tumors. Yet, similar to mutation-caused tumor immunity, alternative-splicing dysregulation in cancer has also been shown to create neoepitopes capable of causing an immune response. In this work, we develop a new bioinformatics pipeline, splicemutr, to characterize alternative splicing as a missing source of antigens in cancer from a splice-junction perspective. The splicemutr pipeline identifies antigenic, splice variants by kmerizing differentially spliced transcripts and performing mhc-binding affinity prediction using mhcnuggets. To benchmark this algorithm on established splicing alterations, we first apply our approach to HPV-positive head and neck tumors and find that splicemutr shows enrichment of immunogenic trans-splicing events and evidence for selection against immunogen-causing splice-junctions when compared to normal. Further expanding this to pan-cancer analysis in The Cancer Genome Analysis (TCGA) enables us to evaluate the extent to which alternative splicing relates to common immunogenic biomarkers, including notably tumor mutation burden. Using splicemutr, we show that within 14 cancer types, splicing antigenicity does not correlate with the tumor mutation burden, showing splicing antigenicity to be independent from a common biomarker of tumor response to treatment. To further relate splicing antigens to the efficacy of immunotherapy, we apply our pipeline to compare splicing-based antigenicity in PD1 resistant 4T1MSI implanted Balb/cJ mice sensitized to nivolumab immunotherapy through combination therapy with a BET inhibitor (BETi) and in post-treament samples from an advanced melanoma clinical trial cohort treated with combination nivolumab and ipimulimab. In the mouse experiment, the nivolumab-BETi combination arm showed marked tumor shrinkage compared to the control and nivolumab alone arms. Using splicemutr, we show that the splicing antigenicity increases from combination therapy to nivolumab alone to the untreated arm, suggesting depletion of splicing-based antigens in those samples that respond well to treatment. In the advanced melanoma clinical trial cohort treated with combination nivolumab and ipimulimab, we show that the splicing-based antigenicity is higher in the non-responders than responders. Based upon our mouse data, we hypothezies that the post-treament human data has decreased splicing-based antigens as the tumor clones containing them may be targeted by the immune system and depleted during successful immunotherapy response. Altogether, these data suggest the potential role of splicing as candidate antigens in cancer and use of our splicemutr pipeline for biomarker discovery based upon splicing antigenicity.

# Introduction

# Methods

# Results

## The Splicemutr Pipeline

|To fully use splicemutr, the pipeline requires a set of splice-junctions identified per sample, target junctions to focus on, a gene expression matrix per sample, and genotype information for the samples being analyzed. To use splicemutr in its most basic form requires only a a set of target junctions. Splicemutr first performs transcript modification using the set of target junctions identified. Per target junction, the flanking exons associated with the target junction are identified. The pair of flanking exons are used to identify the flanking genes that the pair of exons can exist in. Using the pair of genes, of which there can either be two unique or a single gene, those transcripts from the flanking genes that contain the flanking exons are identified. The flanking transcripts are then joined together based on the flanking exon locations within each transcript. The end coordinate of the leading exon is modified to reflect the start coordinate of the splice-junction, while the start coordinate of the trailing exon is modified to reflect the end coordinate of the splice-junction. After both transcripts are joined, the five-prime UTR of the leading transcript is removed and the three-prime UTR of the trailing transcript is removed. UTR removal is strand dependent.

|After UTR removal the splice-junction modified transcript is translated. As unnanotated splice-junctions can cause the annotated orf to be overwritten, splicemutr find the modified orf and documents the changes that have occured. The modified orf is found by scanning for the first ATG start codon in the modified transcript. Once the first start codon is found, the transcript is scanned for the next in-frame TAG, TAA, or TGA stop codon. From start codon, to stop codon, the transcript is translated and stored in a fasta file along with the junction metadata for MHC binding affinity prediction. If no orf can be formed due to the junction modification, the transcript, but not the protein is stored. The junction metadata contains all modifications the junction makes to the transcripts that it modifies.

|The set of junction-modified proteins are then processed for MHC binding affinity predictions. The translated proteins are kmerized, then splicemutr extracts out the unique set of kmers found in the set of translated proteins. Splicemutr then uses mhcnuggets to predict the raw binding affinity of each kmer using the HLA alleles specific to the samples being analyzed. Splicemutr then extracts out those kmers with a percentile rank less than or equal to two percent per HLA allele. This percentile rank is unique to the specific HLA allele and has been shown to reduce the number of false positive hits from MHC binding affinity predictors. Using the sample genotype, the number of immunogenic kmers per junction-modified transcript and per sample are determined. Splicemutr then uses this information to calculate a per-gene splicing antigenicity metric.

|The gene splicing antigenicity metric is calculated as follows: where is the variance stabilized read count for the outlier splice-junction j, is the number of immunogenic kmers predicted from the transcript associated with the junction j, is the variance-stabilized read count for the specific gene G, and J is the set of junctions associated with the gene G. This metric serves as a weighted sum for the splicing-based immunogenic impact a gene has.

## Coding potential prediction using the LGG method

| | is the probability of the of ORF in the coding sequence | is the probability of ORF in non-coding sequence | is the probability of finding a stop codon in coding sequence | is the probability of finding a stop codon in non-coding sequence L > 0 indicates it is a protein-coding RNA and L < 0 indicates that it is a non-coding RNA