Check for updates

CANCER

Copyright © 2024 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Splicing neoantigen discovery with SNAF reveals shared targets for cancer immunotherapy

Guangyuan Li[1,2]*†, Shweta Mahajan[3]†, Siyuan Ma[3], Erin D. Jeffery[4], Xuan Zhang[3], Anukana Bhattacharjee[1], Meenakshi Venkatasubramanian[1,5], Matthew T. Weirauch[1,6,7,8], Emily R. Miraldi[1,3,8], H. Leighton Grimes[3,8], Gloria M. Sheynkman[4], Tamara Tilburgs[3,8]*, Nathan Salomonis[1,2,8]*

Immunotherapy has emerged as a crucial strategy to combat cancer by "reprogramming" a patient's own immune system. Although immunotherapy is typically reserved for patients with a high mutational burden, neoantigens produced from posttranscriptional regulation may provide an untapped reservoir of common immunogenic targets for new targeted therapies. To comprehensively define tumor-specific and likely immunogenic neoantigens from patient RNA-Seq, we developed Splicing Neo Antigen Finder (SNAF), an easy-to-use and open-source computational workflow to predict splicing-derived immunogenic MHC-bound peptides (T cell antigen) and unannotated transmembrane proteins with altered extracellular epitopes (B cell antigen). This workflow uses a highly accurate deep learning strategy for immunogenicity prediction (DeepImmuno) in conjunction with new algorithms to rank the tumor specificity of neoantigens (BayesTS) and to predict regulators of mis-splicing (RNA-SPRINT). T cell antigens from SNAF were frequently evidenced as HLA-presented peptides from mass spectrometry (MS) and predict response to immunotherapy in melanoma. Splicing neoantigen burden was attributed to coordinated splicing factor dysregulation. Shared splicing neoantigens were found in up to 90% of patients with melanoma, correlated to overall survival in multiple cancer cohorts, induced T cell reactivity, and were characterized by distinct cells of origin and amino acid preferences. In addition to T cell neoantigens, our B cell focused pipeline (SNAF-B) identified a new class of tumor-specific extracellular neoepitopes, which we termed ExNeoEpitopes. ExNeoEpitope full-length mRNA predictions were tumor specific and were validated using long-read isoform sequencing and in vitro transmembrane localization assays. Therefore, our systematic identification of splicing neoantigens revealed potential shared targets for therapy in heterogeneous cancers.

INTRODUCTION

A paramount goal for cancer treatment is standardized and accessible therapeutic strategies for shared targets that will be effective in a large percentage of patients. Tumor heterogeneity has been widely acknowledged as a hallmark of cancer, which poses challenges for developing new targeted therapies (*1*). Such heterogeneity is further responsible for drug resistance that leads to frequent cancer relapse. Because each tumor sample is unique with distinct mutations, the search for tumor-specific neoantigens has been considered the "final common pathway" for our immune system to fight cancer (*1*, *2*). Focused targeting of patients with selective mutations has produced promising results in cancers with a high mutational burden, such as melanoma, non–small cell lung cancer, and microsatellite instability (MSI)–high (MSI-H) colorectal cancer. For example, four out of six melanoma patients vaccinated with precision neoantigen vaccines show no evidence of relapse within 25 months after therapy (NCT01970358) (*3*). Such promising clinical results have been attributed to the long-term persistence of neoantigen-specific memory T cells, illustrating the durability of neoantigen-based therapies (*4*). Other examples include Moderna's mRNA-4157 combination with pembrolizumab, which achieved a 50% response rate in human papillomavirus (HPV)–negative head and neck cancer compared with 14.6% for pembrolizumab monotherapy (NCT03313778) (*5*) and adoptive T cell transfer, in which neoantigen-reactive T cells are cultured and reinfused into the same patient, resulting in a 55% objective response and 24% complete response rate in metastatic melanoma (*6*).

Although immune checkpoint blockade (ICB) has become the frontline clinical treatment in patients with high mutational burden, such therapies are not used in many cancers with low mutation burden, such as glioma and leukemia (*7*, *8*). Although historically attributed to tumor-associated mutations, neoantigens can be produced from diverse posttranscriptional regulatory mechanisms. Alternative splicing is one of the primary mechanisms used to achieve mRNA transcript and proteomic diversity in higher-order eukaryotes (*9*). In cancer, altered mRNA splicing can lead to aberrant protein products that promote oncogenic transformation and metastasis and confer chemotherapy resistance (*10*–*13*). After their initial identification using proteogenomics approaches, splicing neoantigens have become increasingly recognized as a potent source of neopeptides to potentially elicit immune response and induce cancer cell death (*14*).

Depending on the cancer, splicing neoantigens often appear to be the dominant source of tumor-specific peptides (*15*, *16*). Such splicing events include intron retention, which typically results in

[1]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA. [2]Department of Biomedical Informatics, College of Medicine, University of Cincinnati, Cincinnati, OH 45267, USA. [3]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA. [4]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, VA 22903, USA. [5]Department of Computer Science, University of Cincinnati, Cincinnati, OH 45229, USA. [6]Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital, Cincinnati, OH 45229, USA. [7]Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, OH 45229, USA. [8]Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA.
*Corresponding author. Email: li2g2@mail.uc.edu (G.L.); tamara.tilburgs@cchmc.org (T.T.); nathan.salomonis@cchmc.org (N.S.)
†These authors contributed equally to this work.

Downloaded from https://www.science.org on July 31, 2025

nonsense-mediated decay but which produces major histocompatibility complex (MHC)–presented neopeptides that can be detected by mass spectrometry (MS) (*17*, *18*). Such peptides require further experimental validation, because MHC presentation alone does not dictate the ability to mount a robust T cell response (immunogenicity) (*19*). The prediction of such antigens, however, remains nontrivial, because splicing neoantigens must be degraded, bound, and presented by specific cognate human leukocyte antigen (HLA) alleles, and interact with patient-specific T cell receptors on $CD8^+$ T cells to induce an immune response. As such, the precise relationship between splicing neoantigen expression and patient prognosis has remained largely unknown, and it is unclear whether overall splicing neoantigen burden affects response to immunotherapy. Further, a concern for the use of splicing neoantigens as targets for therapy is that the occurrence of a splicing event is often nonbinary (changes in percent exon/intron inclusion), relative to mutations (present or absent), making it difficult to know which splicing events are truly tumor specific. An alternative strategy to target tumor-specific splicing is to focus on events that specifically result in new transmembrane proteins that might expose tumor-specific epitopes, bypassing the need to be presented by HLA. In principle, such peptides could be recognized by new chimeric antigen receptor (CAR)–T therapies, which use B cell receptors to bind epitopes (*20*) or selective monoclonal antibodies to mediate targeted tumor cell death. Although attractive, identifying such neoisoforms requires an accurate prediction or measurement of full-length isoforms that do not undergo nonsense-mediated decay and result in properly folded proteins that conserve overall domain architecture. Given these challenges, no reusable and comprehensive neoantigen prediction workflows exist to unbiasedly and confidently identify splicing neoantigens that can be exploited by current immunotherapy strategies.

Here, we performed a systematic analysis of splicing neoantigens in melanoma using a new computational workflow called Splicing Neo Antigen Finder (SNAF). SNAF is an easy-to-use computational tool to identify, prioritize, and interpret distinct classes of splicing neoantigens. The workflow incorporates advanced deep learning and probabilistic algorithms to discover immunogenic splicing neoantigens (SNAF-T workflow), full-length protein coding transmembrane tumor-specific isoforms (SNAF-B workflow), and regulators of altered splicing (RNA-SPRINT). We demonstrate that splicing neoantigens in melanoma are frequently shared among patients, can predict survival, and can be validated by multiple approaches: immunopeptidomics, targeted MS, MHC stabilization, T cell reactivity, single-cell genomics, long-read isoform sequencing, and neoisoform transmembrane localization. These analyses show that splicing neoantigens represent an untapped reservoir of shared targets for cancer immunotherapy.

RESULTS

Inferring new classes of neoepitopes from RNA-Seq

To identify and characterize new forms of neoepitopes, we created two new computational workflows focused on T cell– and B cell–based therapies. T cell–based therapies include cancer vaccines, which require that target antigens are processed, are presented by MHC, and are immunogenic. B cell–based therapies, such as monoclonal antibodies, require the identification of transmembrane protein encoding neoantigens that will enable targeted approaches to selectively recognize cancer cells. SNAF was developed to recognize and prioritize both classes of neoantigens (SNAF-T and SNAF-B) in individual patient samples while assessing the aggregate importance of each neoantigen at a population scale (Fig. 1A).

This workflow begins with user-supplied BAM files from tumor samples or cancer cell lines, followed by the identification and quantification of diverse classes of posttranscriptional regulation. In particular, the workflow applies a highly accurate approach for local splicing variation (LSV) (MultiPath-PSI) from the AltAnalyze framework (*21*) to detect known and unannotated alternative splicing (cassette exon, 3′/5′ splice site exon, intron retention, alternative terminal exon, trans-splicing) and alternative promoter regulatory events, which produce unique exon-exon or exon-intron junctions for in silico translation (fig. S1). This approach has been benchmarked against diverse LSV approaches (*22*–*24*), with methods to accurately quantify retained introns (fig. S2) (*25*). The produced splice junction/sample count matrix is compared against a MultiPath-PSI preprocessed database of normal human healthy tissues (GTEx and TCGA) to identify those that are tumor specific (fig. S3) (*26*). Tumor-specific splice junctions can be analyzed in parallel with SNAF-T and SNAF-B. SNAF-T consists of (i) HLA type prediction from sample FASTQ files, which are user provided, (ii) in silico translation, (iii) MHC binding prediction (NetMHCpan or embedded calls to MHCflurry) (*27*, *28*), and (iv) HLA allele–specific immunogenicity prediction (DeepImmuno) (*19*). SNAF-B consists of (i) full-length isoform prediction for each tumor-specific splice junction by augmenting existing isoform references, (ii) exclusion of isoforms predicted to induce nonsense-mediated decay (NMD), (iii) transmembrane topology prediction, and (iv) long-read isoform sequence validation. For both workflows, a maximum likelihood estimation and a separate hierarchical Bayesian model (BayesTS) (*29*) are then applied to assess the tumor specificity of each neojunction in SNAF's default and the optional custom healthy tissue reference RNA-sequencing (RNA-Seq) data with custom tissue weighting assigned by the user. Finally, to identify causal regulators of splicing neoantigen production, we developed RNA-SPRINT (RNA-based Splicing PRotein activity INference from multivariate decision Trees) to infer splicing factor activity directly from tumor RNA-Seq splicing profiles.

This workflow is unique in both its design and functionality (*30*–*32*) (Table 1). Unlike prior T cell–based splicing neoantigen prediction approaches, SNAF is fully automated, supports any human genome version, has an embedded diverse database of healthy reference profiles (GTEx and TCGA), performs probabilistic tumor specificity modeling, quantifies splicing factor activities, identifies intron retention–associated antigens, and enables more accurate prediction of immunogenicity (*19*). SNAF-B provides independent evidence of tumor-specific transmembrane proteins (ExNeoEpitopes) that can uncover new extracellular epitopes for antibody recognition. Because the program has a modular design with well-described Python classes, it can be customized to incorporate additional reference datasets for verification, including control RNA-Seq, long-read sequencing, and alternative algorithms such as MHC binding prediction.

Validation of predicted MHC-bound neoantigens

We recently showed that our DeepImmuno workflow can predict immunogenic tumor neoantigens with up to a twofold greater sensitivity than alternative approaches (*19*). To determine whether SNAF-T can identify bona fide MHC-presented splicing neoantigens,

Downloaded from https://www.science.org on July 31, 2025

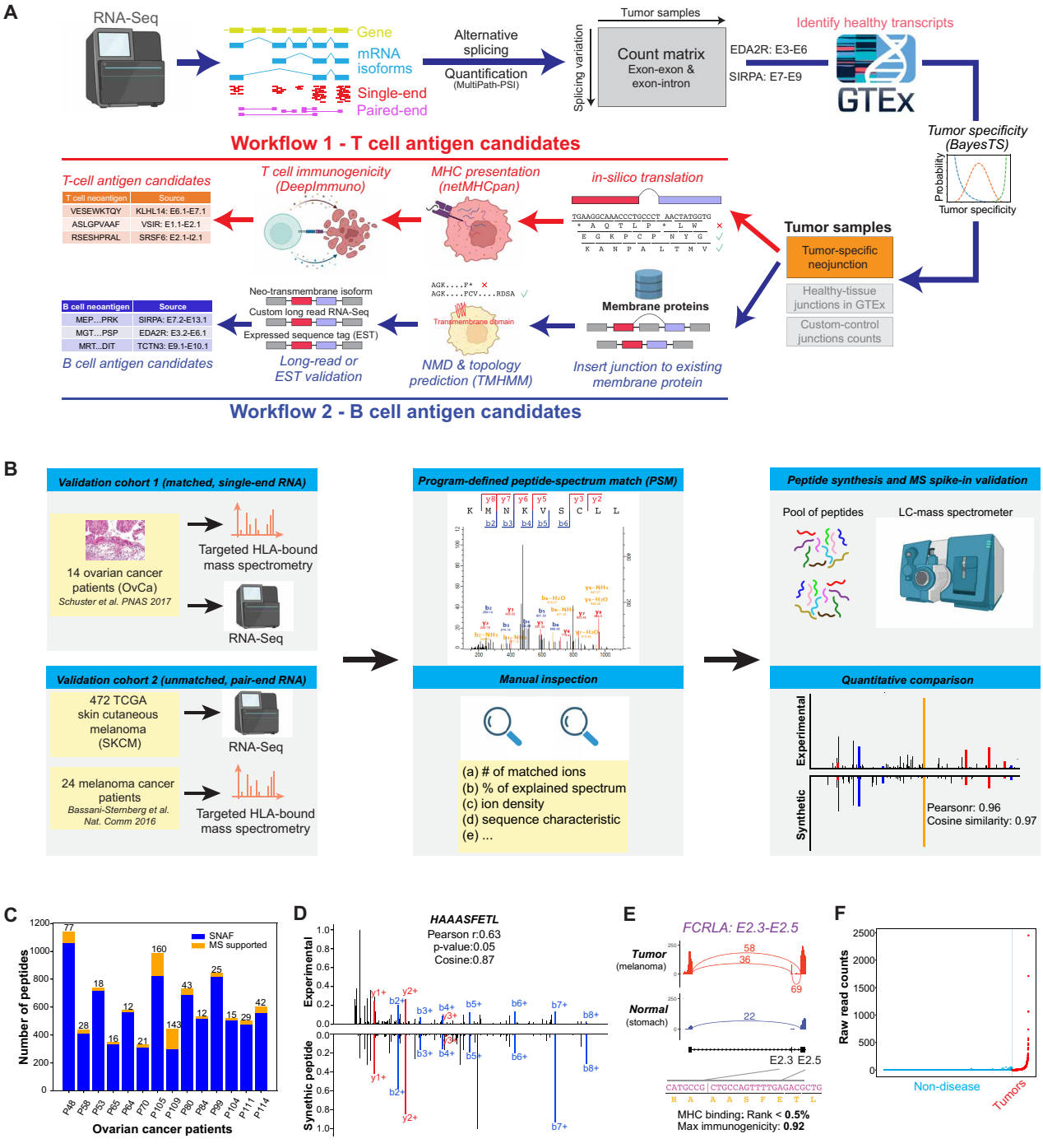

Fig. 1. Automated discovery and confirmation of immunogenic and transmembrane splicing neoantigens with SNAF. (**A**) Outline of the two parallel workflows in the software SNAF to predict splicing neoantigens. SNAF begins with the identification and quantification of alternative splice junctions (exon-exon and exon-intron) from RNA-Seq BAM files and filters these junctions against normal tissue reference RNA-Seq profiles (BayesTS). Retained tumor-specific splice junctions (neojunctions) are evaluated for T cell (SNAF-T) and B cell (SNAF-B) antigen production. SNAF-T performs in silico translation of each junction and MHC binding affinity prediction (netMHCpan or MHCflurry) and identifies high-confidence immunogenic neoantigens through deep learning (DeepImmuno). SNAF-B predicts full-length protein coding isoforms that produce cancer-specific ExNeoEpitopes, considering existing transcript annotations and full-length isoform sequencing. (**B**) Validation workflow for ovarian cancer and melanoma immunopeptidomics with either matched or unmatched RNA-Seq. MaxQuant was applied to find a PSM, followed by quantitative and expert MS2 spectra prioritization. HPLC-MS/MS confirmation was performed on synthesized nominated neoantigens. (**C**) Number of SNAF-T–predicted neoantigens and those evidenced by immunopeptidomics across 14 patients, queried selectively against neoantigen peptides. (**D**) Mirror plot of the immunopeptidomics and spike-in MS spectrum for HAAASFETL. The lines indicate mass-to-charge ratios for distinct types of fragmented ions (red/blue). (**E**) SashimiPlot visualization of HAAASFETL, derived from an exon-exon junction in the gene *FCRLA*, along with the junction/peptide sequence, binding affinity, and immunogenicity prediction. (**F**) Raw read counts of the *FCRLA* neojunction between normal controls (blue) and TCGA melanoma cohort (red).

Downloaded from https://www.science.org on July 31, 2025

Table 1. Comparison of features in SNAF to other published splicing neoantigen workflows. Y, yes; N, no.

	SNAF	IRIS	ASNEO	NeoSplice
Automated program features	Y	Y	Y	Y
In silico translation	Y	Y	Y	Y
MHC binding predictions	Y	Y	Y	Y
Tumor-specific expression	Y	Y	Y	Y
Immunogenicity predictions	Y	Y	Y	Y
Surface antigen predictions	Y	Y	N	N
Parallelization	Y	Y	N	N
Interactive web app	Y	Y	N	N
Interface to proteomics analysis	Y	Y	N	N
Principal tumor specificity score	Y	N	N	N
Custom normal tissue reference	Y	N	N	N
Long-read validation	Y	N	N	N
Differential gene expression	Y	N	N	N
Gene set enrichment	Y	N	N	N
Splicing factor activity prediction	Y	N	N	N
Advanced visualization	Y	N	N	N
Integrated survival analysis	Y	N	N	N
Stand-alone python module	Y	N	N	N

we selected two prior produced cancer immunopeptidome datasets to validate its predictions. First, we evaluated bulk RNA-Seq (single-end) and matched immunopeptidome profiling data (HLA-bound peptides) from 14 patients with ovarian cancer (*33*). To expand these predictions, we applied SNAF-T to skin cutaneous melanoma (SKCM) biopsy RNA-Seq from the Cancer Genome Atlas (TCGA) initiative ($n = 472$) and unmatched immunopeptidome data from 24 patients with melanoma, herein referred to as the Bassani Sternberg cohort (Fig. 1B and data S1 to S4) (*34*). For these analyses, we performed mapping of the MS spectra to both a neoantigen-only and combined neoantigen plus normal proteome database (Supplementary Materials). In ovarian cancer, searching for MS spectra that map directly to neoantigens, SNAF identified 46 splicing neoantigens on average, ranging from 12 to 160 antigens per patient (Fig. 1C). We expect this number to be an underestimate, because HLA-bound peptides arise from nonspecific protease cleavage, which alters the resultant MS spectra, and hence, traditional tryptic-based search engines cannot confidently recover all neoantigens (*35*, *36*). Here, we observe that the absolute number of MS-supported splicing neoantigens exceeds what was previously reported (an average of two peptides per sample) using untargeted proteome data (CPTAC) (*14*), indicating increased sensitivity of targeted immunopeptidome data for identifying both common and rare neoantigens. Because MS-based predictions are subject to inherent type 1 errors (*37*), we validated select ovarian and melanoma SNAF-T neoantigens using targeted MS on synthesized neopeptides. Specifically, we selected 14 ovarian and 22 melanoma peptides with high-confidence spectra (*38*). Of the 36 tested, 27 were sufficiently detected by MS (data S5). Comparison of the synthetic to the original mass spectra found 11 matches, of which 7 were high-scoring based on all match criteria, with varying levels of confidence (Fig. 1D and fig. S4).

These seven neoantigens were derived from multiple mechanisms, including known alternative exons, alternative 3′ and 5′ splice sites, intron retention, and undocumented cassette exons (AltAnalyze defined). For example, the shared melanoma splicing neoantigen HAAASFETL in the gene *FCRLA* occurs because of a known cassette exon-exon junction in an isoform that is weakly detected in blood and spleen (average read count = 0.51, BayesTS: 0.03) (Fig. 1E). Expression of *FCRLA*, which is a member of the Fc receptor–like immunoglobulins, is correlated with good prognosis in melanoma (*39*). *FCRLA* gene expression is only weakly tumor specific (BayesTS: 0.13). However, this junction was detected in >34% of TCGA patients with melanoma (162 of 472, average read count = 52.91) (Fig. 1F). The resultant mass spectra of this antigen had a high-confidence match (Andromeda score: 149.06, $P < 0.0001$), with a synthetic spectrum Pearsonr similarity of 0.63 ($P = 0.05$) and cosine similarity of 0.87. This peptide was only mapped to this melanoma-specific FCRLA isoform in the original mass spectrum search, using the extended human isoform proteome database (Fig. 1D). The other mass spectrum confirmed neoantigens were derived from diverse protein families, including ubiquitin protein ligase complex (*FBXO7*), asparagine amidase (*NGLY1*), cytoskeletal motor protein (*DYNLT5*), negative regulator of RAS signaling (*RASA3*), and currently uncharacterized protein coding genes (*C20orf204*, *C6orf52*) (fig. S4). We observed lower spectrum similarity scores in the ovarian cohort (Pearsonr: 0.55) compared with the melanoma cohort (Pearsonr: 0.84), which we attributed to

Downloaded from https://www.science.org on July 31, 2025

differences in the MS technology applied in the synthetic peptide MS data acquisition (ion trap MS versus Fourier transform MS) and fragmentation methods (collision-induced dissociation versus higher-energy collisional dissociation).

Splicing neoantigen burden predicts overall survival and response to immunotherapy

To broadly assess the clinical relevance of splicing neoantigens in a large cancer cohort, we applied SNAF to more than 500 melanoma patient biopsies with (Van Allen cohort) (*40*) and without immunotherapy (TCGA-SKCM) (*41*). For these analyses, we separately considered the number of predicted neojunctions, MHC-bound peptides, immunogenic neoantigens, and overall neoantigen burden, considering RNA-Seq–determined HLA alleles for each patient. In the TCGA cohort of 472 samples, we found 528 tumor-specific splice junctions (neojunctions) per patient on average, ranging from 28 to 1549. From these neojunctions, we predicted an average of 1090 MHC-bound peptides, ranging from 75 to 2981 peptides per patient. DeepImmuno predictions reduced the number of neoantigens to 915 on average (ranging from 74 to 2486 per patient), filtering out 16% of potentially nonimmunogenic bound peptides, which is expected because all current immunogenicity approaches suffer from low precision (data S6) (*19*).

To investigate the relationship between splicing neoantigen burden and clinical outcome, we performed survival analyses on both the TCGA and Van Allen cohorts (data S6). This analysis found that patients in TCGA with high MHC-bound neoantigen burden trended toward poor overall survival (log-rank $P < 0.05$) (Fig. 2A). Conversely, we found that patients with a high neoantigen burden that received ICB (Van Allen cohort) had improved overall survival (log-rank $P = 0.18$). These trends were also reflected in neojunction and immunogenic peptide burden (Fig. 2B). Because the bulk RNA-Seq data profiled in these cohorts were obtained before therapy, one plausible explanation is that patients with high neoantigen burden exhibited frequent tumor immune escape, which was overcome by CTLA-4 inhibition. Examination of differential gene expression in TCGA patients with high splicing neoantigen burden versus low found 597 up- and 227 down-regulated genes [fold > 1.5 and eBayes *t* test $P < 0.05$, false discovery rate (FDR)] (Fig. 2C and data S7). The top most differentially up-regulated genes were those previously implicated in immune evasion such as *ADAM10* (*42*), *PTPN11* (*43*), *TGFBR1* (*44*), *TNPO1* (*45*), *ANKRD52* (*46*), *MIB1* (*47*), and *KIF3B* (*48*). In contrast, down-regulated genes were markers of active tumor immune cell infiltration, in particular plasma cells and T cells (Fig. 2C), along with genes involved in antigen binding and complement activation (fig. S5A). Gene set enrichment of high burden–induced transcripts identified significant up-regulation of p53 signaling ($P = 0.001$, FDR), mitotic cell cycle ($P = 1.24 \times 10^{-5}$, FDR), cell motility ($P = 0.02$, FDR), DNA damage response ($P = 0.002$, FDR), fatty acyl–CoA (coenzyme A) metabolism ($P = 0.007$, FDR), lipid biosynthesis ($P = 0.03$, FDR), and epithelial-mesenchymal transition genes ($P = 0.02$, FDR), among others (Fig. 2D and data S8). Many of these genes are directly associated with chemotherapy or radiotherapy resistance, in particular those that mediate DNA repair, including *ATM*, *RB1*, and *RAD21*, and acyl-CoA synthesis/ferroptosis, such as *ACSL3*, *ACSL4*, and *FASN* (*49*–*55*). Combination radiotherapy and immunotherapy have been proposed as a strategy to overcome single-therapy drug resistance, which aligns with our observation of elevated immune evasion genes in high-burden groups (*56*). Thus, patients with melanoma (TCGA) who exhibit a high burden of splicing neoantigens may have tumors that are resistant to chemotherapy and radiotherapy with up-regulated immune evasion capacity. As a result, such patients represent possible candidates for combination therapy. In contrast, low neoantigen burden patients with melanoma show evidence of immune reactivity, which may partially explain their favorable prognosis.

CAMMK2 mutations correlate with increased neoantigen burden in melanoma

To determine whether splicing neoantigen burden in patients with melanoma is associated with specific mutations, we compared mutations in high- versus low-burden neoantigen patients. Although no individual mutations achieved an FDR-corrected $P < 0.05$, multiple mutations trended toward significance, with the most enriched mutations in the gene *CAMKK2* versus wild type (Mann-Whitney test $P = 0.0004$, nonadjusted) (Fig. 2E and data S9). Out of 19 patients with reported *CAMKK2* mutations, 13 were found to occur in the high burden group, out of a total of 222 patients. Mutation or inhibition of *CAMKK2* is known to lead to increased anti-PD1 immunotherapy efficacy (*57*). This effect is believed to occur as a result of *CAMKK2*'s ability to negatively regulate ferroptosis, a mechanism of cell death that is induced by iron-dependent lipid peroxidation, through the AMPK-NRF2 pathway (*57*). Our results suggest that *CAMKK2* mutations may contribute to splicing neoantigen burden and immunotherapy outcome, through either direct or indirect splicing regulatory networks, in a subset of patients with melanoma.

Individual splicing neoantigens can predict response to immunotherapy

Splicing neoantigens that predict prognosis in the presence or absence of immune checkpoint blockade (ICB) represent potential biomarkers to guide treatment decisions. To evaluate this hypothesis, we compared the individual splicing profiles of SNAF immunogenic neoantigens with patient survival. This analysis identified 2970 parental junctions associated with poor overall survival in TCGA melanoma [likelihood ratio test (LRT) $P < 0.05$ and $z \geq 1$] (Fig. 2F and data S1). Among these neoantigens, we noted that a subset ($n = 108$ junctions) was present in over 15% of patients (shared neoantigens), suggesting that these represented new potential survival biomarkers. Although fewer neoantigens were associated with survival in the Van Allen cohort because of the limited sample size, we found 1755 poor and 227 good prognosis-associated neoantigen junctions (LRT $P < 0.05$ and $z \geq 1$ or $z \leq 1$, respectively). However, we found that seven unique neoantigen junctions in TCGA and Van Allen were associated with opposite survival associations (poor in TCGA and good in Van Allen) and present in >10 TCGA patients (Fig. 2F). For example, a splicing neoantigen in the melanin synthesis–associated glucose transporter *SLC45A2* (TEFQTRRAM) was detected in 212 of 472 patients from the TCGA SKCM cohort and was correlated with poor overall survival in TCGA (Wald $P < 0.05$, z score > 2) and positive overall survival in the Van Allen cohort (Wald $P < 0.05$, z score < -2), suggesting that it may represent a relatively common target for therapy and prognosis (Fig. 2F). Other overlapping neoantigens produced from this same junction were present in over >64% of patients in both melanoma cohorts (FQTRRAMTL), with differences in their abundances due to varying MHC-I allele binding and immunogenicity preferences. Dozens of other shared neoantigens were associated

Downloaded from https://www.science.org on July 31, 2025

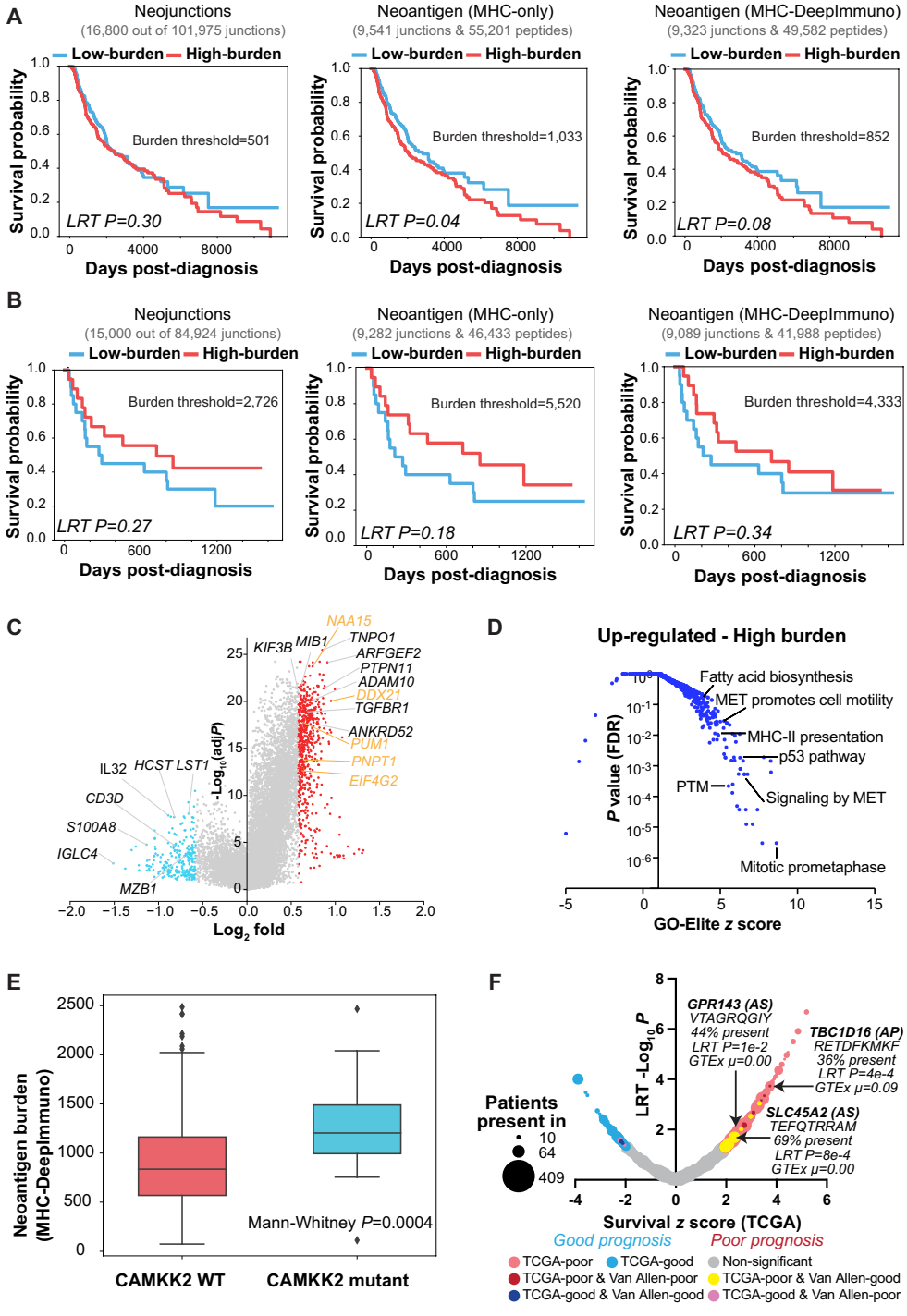

Fig. 2. Splicing neoantigen burden predicts response to therapy in melanoma. (**A** and **B**) Kaplan-Meier (KM) survival plots of melanoma patient samples stratified into low and high neoantigen burden, considering overall survival for (A) TCGA (n = 472) and (B) Van Allen (n = 39). Each column indicates a sequential stage of SNAF neoantigen filtering: tumor-specific neojunctions (left column), MHC-bound neoantigens (middle column), and immunogenic neoantigens (right column). Van Allen cohort RNA-Seq samples were subjected to immune checkpoint inhibitors, whereas TCGA samples were not. The number of neojunctions or neoantigen peptides is shown at the top of each plot. (**C**) Volcano plot of genes differentially expressed in patients with high versus low immunogenic splicing neoantigen burden in TCGA-SKCM, with fold > 1.5 and eBayes t test $P < 0.05$ (FDR-corrected). Red, genes are up-regulated in the high burden group; blue, down-regulated genes in the high burden; orange, representative RNA binding proteins. (**D**) Gene set enrichment with GO-Elite of ToppFun pathway gene sets for genes up-regulated in patients with high splicing versus low neoantigen burden (C). (**E**) Immunogenic splicing neoantigen burden between patients in the TCGA melanoma cohort with or without mutations in *CAMKK2*. Mann-Whitney two-sided test. (**F**) Bubble plot of survival-associated splicing neoantigens from SNAF in TCGA melanoma. Dot size corresponds to the number of patients with melanoma that the splicing neoantigen is detected in (10–470) and are colored according to their survival significance in the TCGA melanoma and Van Allen cohorts (LRT $P < 0.05$ and $z \geq 1$). AS, alternative splicing.

Downloaded from https://www.science.org on July 31, 2025

with poor prognosis in TCGA and trending toward good prognosis in the Van Allen cohorts (data S1). The occurrence and tumor specificity of these and similar neojunctions were confirmed by Integrated Genome Viewer (IGV) visualization (fig. S5B). These data reveal that individual shared splicing neoantigens can predict outcomes in patients before and after ICB.

Disrupted splicing factor activity underlies high splicing neoantigen burden

The coordinated regulation of alternative splicing and alternative promoters is orchestrated by the expression or activity of conserved cis- and trans-regulatory interactions, such as splicing factor binding to RNA recognition elements. Splicing neoantigens are likely produced by modulation of such interactions, driven by direct mutations or RNA editing that results in unique tumor-specific mRNAs. The most likely immediate mediator of differential splicing neoantigen burden is a change in the expression or activity of one or more splicing factors. Examination of differential gene expression in the high versus low splicing neoantigen group found many up-regulated splicing regulators (Fig. 2C and data S3). These included 117 mRNA splicing regulators, up-regulated with fold > 1.2 and eBayes *t* test $P < 0.05$, FDR (data S7). Patients with mutations in SF3B1 were further enriched in the high-burden group (Mann-Whitney $P = 0.02$). Given this finding, we asked whether splicing factor activity was also increased with neoantigen burden. To infer splicing factor activity, we turned to existing methods for transcription regulatory network inference, which rely on a defined set of target genes or regulons (*58*, *59*). The up- and down-regulation of these regulatory targets can indirectly provide evidence of transcription factor (TF) activity. To establish a reliable link between splicing factors and downstream splice junctions in an analogous manner, we used a large dataset of RNA binding protein (RBP) knockdowns ($n = 191$) in K562 cells from the ENCODE project. By observing the changes in splicing events upon the knockdown of a specific RBP and accounting for batch effects, we derived initial splicing regulatory targets. To refine these targets, we further incorporated evidence of direct RBP target regulation using enhanced cross linking and immunoprecipitation (eCLIP) sequencing data for 120 RBPs in the K562 cell line. The resulting data were used to construct a prior network to infer splicing factor activity (Fig. 3A).

We found that a multivariate decision tree (MDT)–based method produced the most accurate RBP activity predictions, compared

Downloaded from https://www.science.org on July 31, 2025

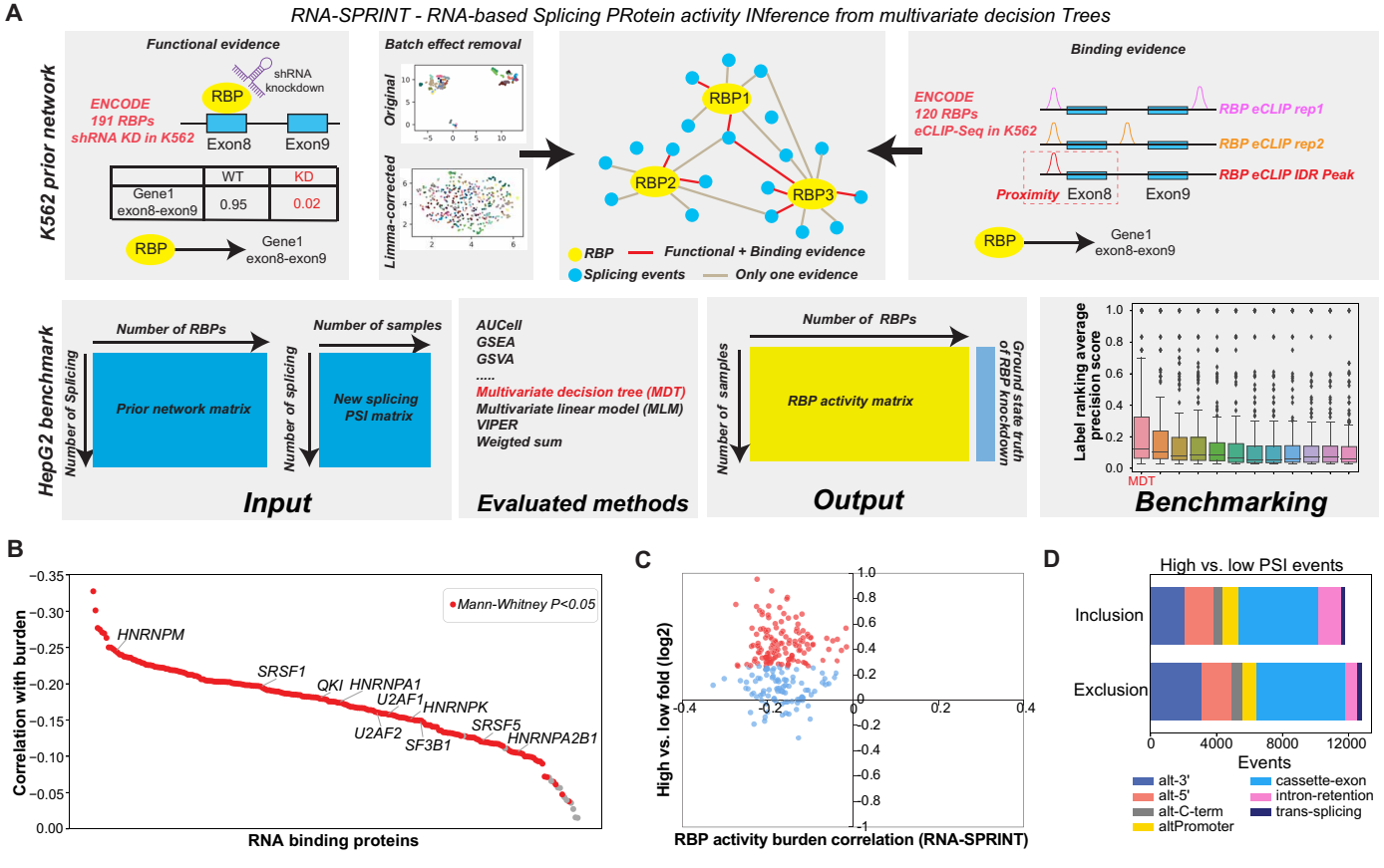

Fig. 3. Regulatory networks mediating splicing neoantigen burden in melanoma. (**A**) Schematic overview of the software RNA-SPRINT and associated benchmarking steps. The workflow involves construction of an RBP prior network to predict splicing regulatory interactions. Evaluation of the method is overviewed, consisting of RNA-SPRINT benchmarking relative to 12 TF activity methods in HepG2 cell line RBP knockdown RNA-Seq datasets. (**B**) Correlation of inferred RBP activity with splicing neoantigen burden for all TCGA patients with melanoma. (**C**) Comparison of RBP activity–burden correlations with RBP differential gene expression for high versus low burden (TCGA SKCM). Red, up-regulated genes (fold > 1.2 and eBayes *t* test $P < 0.05$, FDR-corrected) in high burden. (**D**) Type of splicing events observed with exon/intron inclusion or exclusion comparing high versus low burden.

with 12 other commonly used TF activity inference methods (*60*), when considering 116 RBP knockdown splicing profiles from HepG2 cells (ENCODE) (fig. S6, A to C, and data S10). Application of this approach, which we call RNA-SPRINT, to all TCGA melanoma cohort splicing profiles revealed that 209 of 221 RBPs had reduced activity in the high versus low splicing neoantigen group (Mann-Whitney $P < 0.05$) (Fig. 3, B and C, and data S10). Although contrary to the observed up-regulation of these RBPs, these data suggest that there exists a coordinated failure to properly splice transcripts, as previously described (*61*, *62*). This result is further supported by an observed increase in intron retention for patients in the high-burden group (failure to excise introns) (Fig. 3D). It is also possible that broad up-regulation of splicing factors is a compensatory mechanism to counteract global splicing defects or simply a by-product of cell type heterogeneity within the tumor such as immune infiltration.

Shared splicing neoantigens are presented by MHC and display amino acid compositional bias

One observation from our SNAF-T analyses was the co-occurrence of splicing neoantigens among many patients. We found 940 T cell neoantigens predicted in >15% of patients in TCGA cohort (shared) (Fig. 4A). Although a much smaller cohort, we found 439 of 8422 shared neoantigens in the Van Allen cohort, overlapping with the 940 in TCGA. TCGA shared neoantigens were found to more frequently result from cassette exons as compared with unique neoantigens that are only observed in a single patient (Fig. 4B). Genes with shared neoantigens were significantly enriched in gene sets for melanocyte biology, such as melanocyte differentiation ($P = 9.3 \times 10^{-4}$, FDR), melanin biosynthetic process ($P = 3.4 \times 10^{-3}$, FDR), and cell division ($P = 1.0 \times 10^{-4}$, FDR) (Fig. 4C and data S11).

Because shared splicing neoantigens are rare relative to all unique predicted neoantigens, we would expect such peptides to be more frequently detected using immunopeptidome profiling in an independent cohort. Hence, we again used a previously published melanoma dataset of HLA-I–bound peptides detected by MS in 24 independent patients (Bassani Sternberg cohort) (*34*). Considering both shared ($n = 613$) and unique ($n = 16{,}753$) peptides in our reference peptide database (combined unique reference database, excluding peptides in UniProt), we found that shared splicing neoantigens could be found at a higher rate than unique neoantigens (paired t test, $P = 0.006$) (Fig. 4D). The higher recovery rate was still observed even with the inclusion of the human normal proteome (data S2). Additionally, of the 613 shared neoantigens examined, we found immunopeptidome evidence for 34% ($n = 210$), with 98 shared neoantigens evidenced in at least 15% of patients by immunopeptidome analysis in the Bassani-Sternberg cohort, as illustrated using kernel density estimates or estimator of the empirical cumulative distribution function (eCDF) (Fig. 4E and fig. S7A). Inspection of neojunctions produced from these MHC-presented neoantigens confirmed that they were tumor specific (fig. S7B).

Whereas HLA genes are highly polymorphic and have different binding preferences for neoantigens, the existence of shared splicing neoantigens suggests that diverse HLA genotypes may bind in a more promiscuous manner to these peptides versus those present only in one or few individuals. To evaluate the potential recurrence of amino acid sequences, we first redefined the shared and unique neoantigens by normalizing the frequency of their parental splice junctions (Fig. 4F). Because 9-mer neoantigens had a balanced number of shared and unique neoantigens, we focused specifically on these peptides. To derive a physicochemical profile of the amino acids associated with each, we encoded each 9-mer as a numerical vector based on all amino acid physicochemical parameters in the AAIndex database (566 parameters) and projected each neoantigen vector as a point in Uniform Manifold Approximation and Projection (UMAP) space (Fig. 4G). Whereas most of the shared and unique splicing neoantigens have similar physicochemical characteristics, a few empirically observed clusters were enriched in shared versus unique neoantigens (circled regions, Fig. 4G). To find preferentially detected amino acids present in the shared and unique neoantigens, we performed motif analysis with MEME (*63*) in these two sets. MEME finds that shared neoantigens frequently end in lysine and phenylalanine ($e = 9.8 \times 10^{-14}$), whereas the most dominant motif found in unique neoantigens disproportionately ends in arginines ($e = 0.14$). The less significant E value also suggested more diverse binding modes for unique neoantigens (Fig. 4H). We observed near-identical motif enrichments for shared versus unique splicing neoantigens in the Van Allen cohort as compared to TCGA (fig. S7C). These data suggest that amino acid sequence bias could be used to find shared neoantigens that are bound by most MHC alleles.

Splicing neoantigens stabilize MHC and elicit T cell responses

To demonstrate that such splicing neoantigens represent viable targets for broadly applicable immunotherapies, we experimentally verified their ability to directly bind MHC and induce T cell reactivity. For validation, we selected five shared splicing neoantigens derived from three neojunctions in *PMEL*, *SLC45A2*, and *CDH19*. Although PMEL is an existing target for immunotherapy in melanoma (NCT00509496), the predicted splicing neoantigen has not been previously described (unannotated exon-exon junction). These neoantigens were selected based on availability of HLA-matched donor cells, high neoantigen frequency in melanoma, and tumor specificity. An MHC stabilization assay using transporter associated with antigen processing (TAP)–deficient HLA-A*02–containing T2 cells confirmed the binding of two neoantigens to HLA-A*02 (Fig. 5, A and B). To assess immunogenicity, we leveraged HLA-typed healthy blood peripheral blood mononuclear cells (PBMCs) to prime T cells and tested interferon-γ (IFN-γ) responses upon pMHC stimulation (fig. S8). Frozen PBMCs were primed three times with autologous peptide-loaded dendritic cells (DCs) and thereafter tested for their IFN-γ response against single HLA-expressing 721.221 (221) cells to ensure that immunogenicity was specific for a single HLA genotype. The five neoantigen peptides loaded on 221 cells expressing their respective allele generated similar IFN-γ responses compared to known immunogenic flu virus (FLU) and/or human cytomegalovirus (HCMV) peptide antigens in three donors (Fig. 5, C and D). Only weak IFN-γ responses were detected in unstimulated T cells and in response to 221 without peptide, showing peptide specificity. Thus, these analyses provide strong evidence of T cell immunogenicity by all five shared splicing neoantigens tested. These results are in contrast to previously described neoantigen prediction workflows that reported far lower validation accuracies (*64*, *65*).

Downloaded from https://www.science.org on July 31, 2025

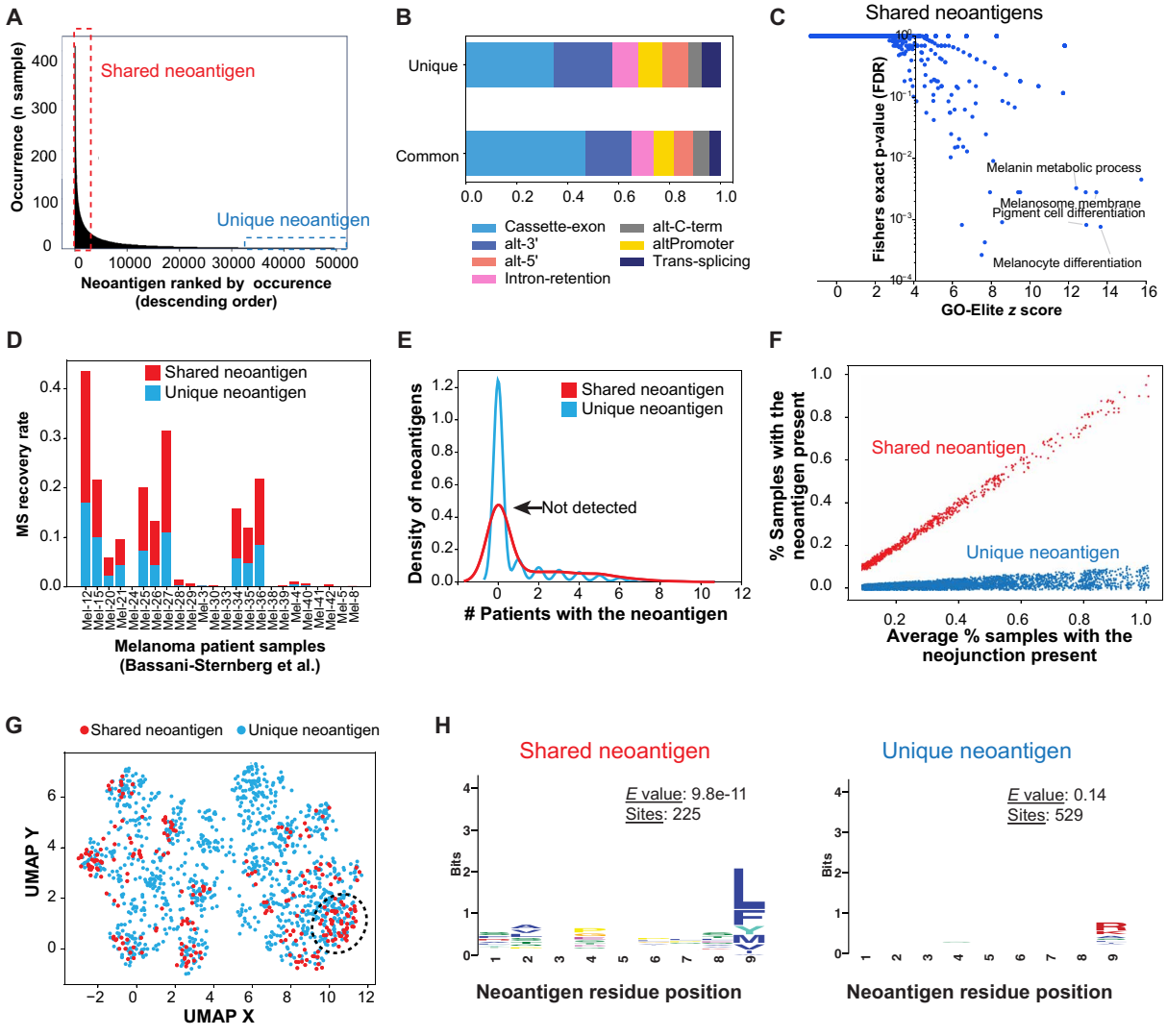

Fig. 4. Shared splicing neoantigens are frequently evidenced by MS and are defined by their sequence composition in melanoma. (**A**) Identification of common (shared in >15% of patients) and unique immunogenic splicing neoantigens in the TCGA melanoma cohort, based on their frequency of occurrence among patients. (**B**) Frequency of splicing event types for shared and unique splicing neoantigen junctions in TCGA. (**C**) Gene Ontology and pathway gene set enrichment of shared neoantigens. (**D**) Comparison of MS recovery rate in an independent melanoma immunopeptidome cohort (*34*) between shared and unique neoantigens from the TCGA melanoma cohort, relative to the neoantigen-only MS database. (**E**) Kernel density estimate plot comparing the observed occurrence of all TCGA shared versus unique splicing neoantigens in the Bassani-Sternberg immunopeptidome dataset. (**F**) Redefined shared and unique neoantigens in TCGA by normalizing the occurrence of their parental splice junction. (**G**) UMAP of splicing neoantigens based on their amino acid physiochemical properties (AAindex) in TCGA melanoma, highlighting neoantigens that cluster based on shared amino acid physicochemical features. (**H**) Distinct enriched amino acid motifs (MEME), comparing shared versus unique neoantigens.

Shared splicing neoantigens derive from tumor cells rather than the tumor microenvironment

Although bulk tumor RNA-Seq enables the detection of splicing neoantigens, it cannot clarify their precise cellular source. Emerging data suggest that neoantigens can also be derived from the tumor microenvironment, including immune cells (*66*). To determine which cell types splicing neoantigens derive from, we reanalyzed single-cell RNA-Seq (scRNA-Seq) from tumor biopsies from 17 patients with melanoma (4454 cells) (*67*). Because these data were profiled using SmartSeq2 chemistry, we were able to detect over 340,000 exon-exon and exon-intron junctions present throughout the gene body for seven prior annotated cell types (tumor, endothelial, cancer-associated fibroblasts, B cell, T cell, natural killer cells, and macrophages). This analysis found that only a small proportion of scRNA-Seq–detected junctions were enriched in tumor cells >2-fold ($n = 30{,}523$) versus those enriched in immune cells ($n = 236{,}954$) (Fig. 6A). Considering all TCGA-predicted melanoma neojunctions in this dataset, we found roughly equal immune and tumor-enriched neojunctions (1289 and 979, respectively). However, restricting this analysis to shared splicing neoantigens in >15% of patients, we found that these neojunctions were almost entirely derived from tumor cells ($n = 195$) versus immune cells ($n = 10$) (Fig. 6B and data S12). Tumor-specific neojunctions included experimentally confirmed splicing neoantigens (*SLC45A2*, *CDH19*, and *FCRLA*) (Fig. 6C). Thus, the tumor microenvironment appears to be a contributor to splicing neoantigen burden but not a primary source for shared neoantigens.

Downloaded from https://www.science.org on July 31, 2025

To ensure that such neoantigens are not simply a by-product of a normal proliferative program, which could result in unintended off-target effects, we examined their expression in multiple induced proliferative epidermal RNA-Seq datasets. We found that proliferation-associated junctions only accounted for a small percentage (~6%) of all TCGA melanoma-identified neojunctions (fig. S9, A to C). Neojunctions from shared splicing neoantigens were only frequently observed in 1 of 170 proliferative samples, specifically from 16-week fetal fibroblasts, but not from other fetal fibroblasts profiled (data S13) (*68*). Furthermore, these overlapping melanoma neojunctions were only weakly detected (1 to 50 reads) in cultured epidermal samples, in genes associated with cell cycle regulation (fig. S9D and data S14). Thus, nonmalignant melanocyte proliferation is not a dominant source of splicing neoantigen production.

SNAF accurately predicts full-length mRNAs and stable proteoforms

We observed several SNAF-T splicing neoantigens in transmembrane proteins, such as GPR143 and SLC45A2, which occur due to undocumented in-frame alternative splice sites (5′ or 3′). In principle, such splicing events could result in undescribed cell surface–expressed proteins. As an alternative source of neoantigens that do not require degradation, MHC presentation, and T cell receptor recognition, we applied SNAF-B to the same TCGA melanoma cohort. The SNAF-B workflow can be used to identify ExNeoEpitopes that have retained transmembrane domains, but altered N-terminal or other extracellular sequences that could serve as new epitopes for the design of new CAR-T therapies. To predict full-length isoforms, unannotated exon-exon or exon-intron junctions (not in the Ensembl or UCSC mRNA database) are inserted into the best-matching

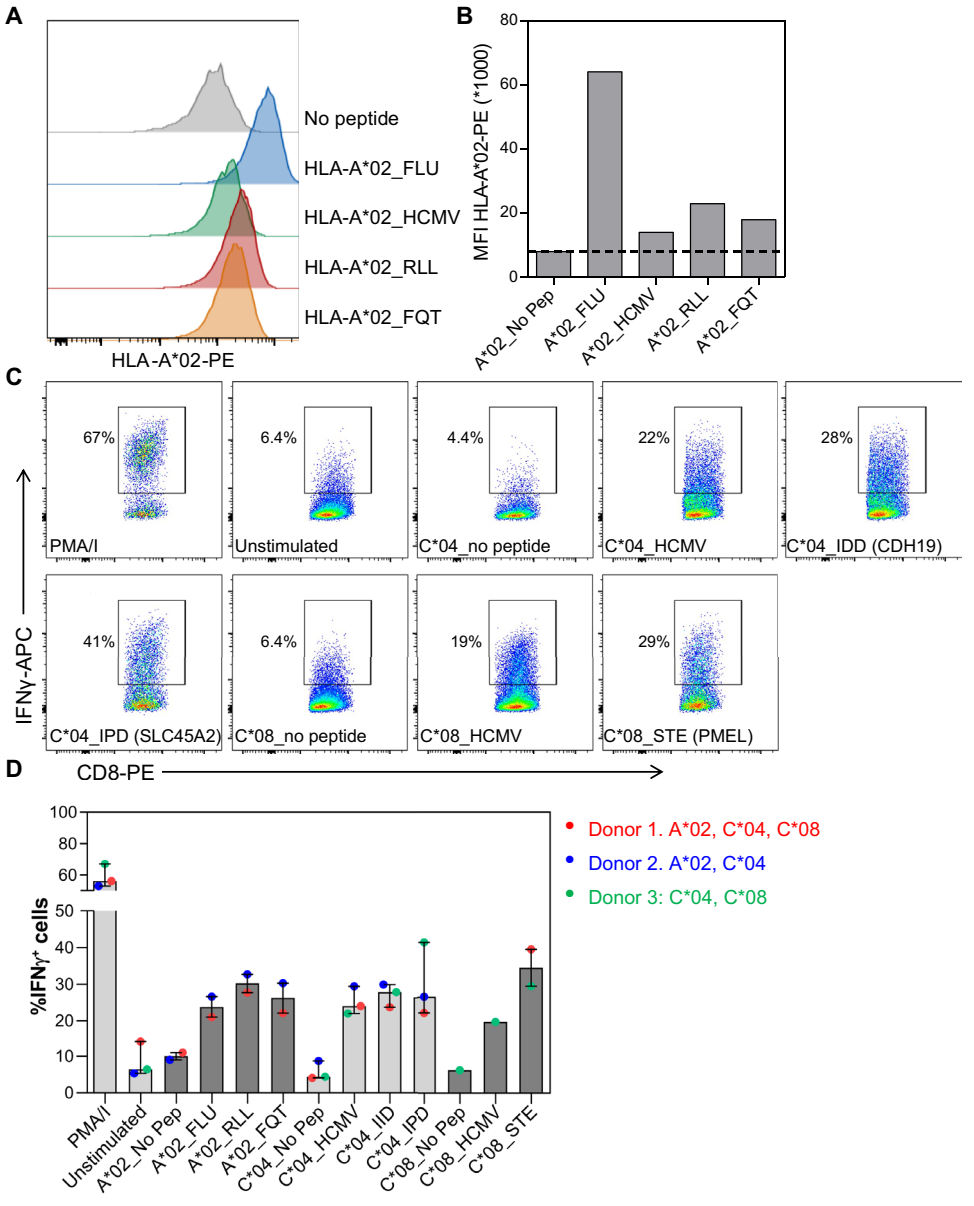

Fig. 5. Shared splicing neoantigens bind HLA and induce T cell reactivity. (**A**) Histograms and (**B**) graph show HLA-A*02-PE staining on HLA-A*02–containing TAP-deficient T2 cells without peptide (No Pep), loaded with FLU and HCMV control peptides and RLLGTEFQT (RLL) and FQTRRAMTL (FQT) peptide neoantigens. MFI, median fluorescence intensity; PE, phycoerythrin-conjugated antibodies. (**C**) Dot plots and (**D**) graph show the percentage of IFN-γ^+ $CD8^+$ T cells in response to five melanoma shared splicing antigens compared with negative (unstimulated, No Pep) or positive (PMA/I, FLU, HMCV) controls. $CD8^+$ T cells were primed using peptide-loaded monocyte-derived DCs and thereafter tested against 721.221 cells selectively expressing the indicated HLA allele with and without peptide loading. Bars indicate median of two to three donors, and lines indicate interquartile range.

Downloaded from https://www.science.org on July 31, 2025

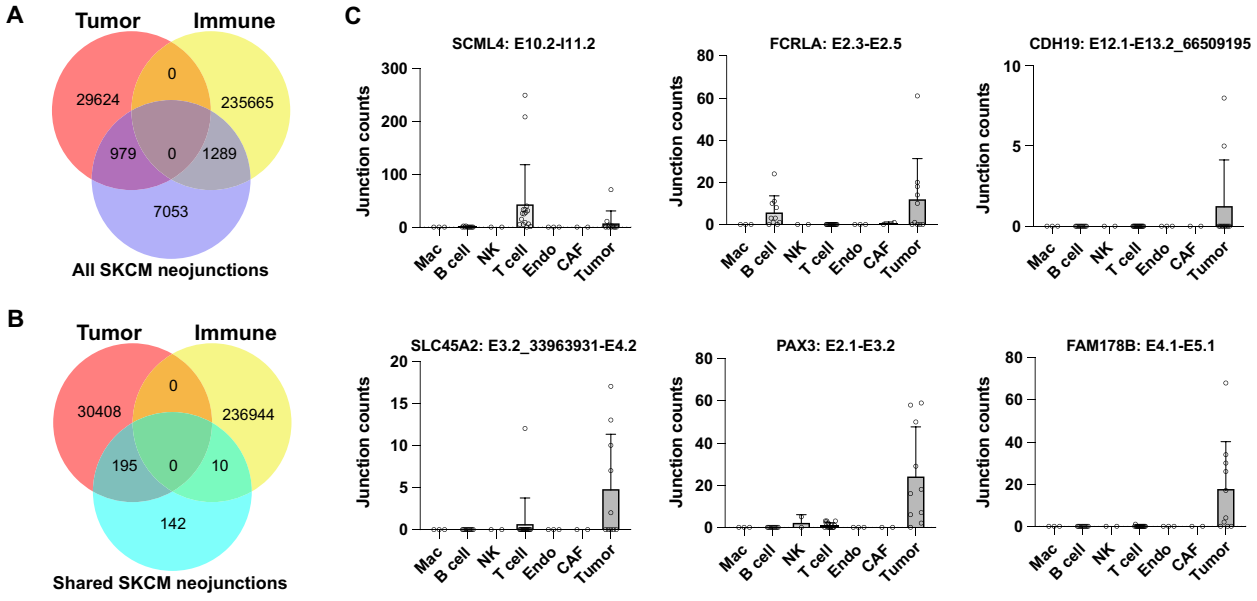

Fig. 6. Splicing neoantigen cell of origin is dependent on its mechanism of regulation. (**A** and **B**) Venn diagrams comparing the number of parental neojunctions for TCGA SKCM splicing neoantigens unique to a single patient (A) or shared in >15% of patients (B) with the specific cell types they derive from in independent melanoma tumor biopsies by scRNA-Seq analysis. Neojunctions are defined as tumor or immune if they are >2-fold enriched in either cell population (absolute number of reads in all patients and cells for each lineage). (**C**) Neojunction expression in individual cell populations for select shared splicing neoantigens. Each dot denotes the combined neojunction read counts in a single patient ($n = 19$) with melanoma, separately per annotated cell population.

isoform models based on exon composition, followed by in silico translation (Fig. 7A). This workflow can optionally exclude predicted mRNAs expected to result in NMD and select those that have high-confidence transmembrane domains based on a prior published hidden Markov model–based topology prediction approach (TMHMM) (*69*). Putative ExNeoEpitopes that result in deleted or new extracellular polypeptides can be further interrogated using the SNAF-B interactive viewer.

This analysis found 378 initial ExNeoEpitopes in the TCGA melanoma cohort, using prior well-annotated transcripts as reference models (Ensembl, UCSC mRNAs) (data S15). To initially assess the validity of such ExNeoEpitopes, we performed long-read RNA isoform sequencing ("Iso-Seq") in four commonly used melanoma cell lines using the PacBio Sequel II platform. This Iso-Seq analysis found 17 of our predicted ExNeoEpitope isoforms that perfectly matched our in silico predictions and an additional 20 that partially matched (overlapping neojunction) (data S15). One example was an undocumented alternative 3′ site in signal regulatory protein α (SIRPA), which resulted in an in-frame deletion of the 5′ end of exon 13, missing 21AA. The predicted full-length isoform was directly evidenced by Iso-Seq (Fig. 7B).

In addition to predicting full-length isoforms, SNAF-B can match short-read identified junctions to other supplied isoform models including long-read or expressed sequence tags (EST). Matching exon-exon and exon-intron junctions from the TCGA melanoma cohort to a publicly available PacBio Iso-Seq dataset of 10 commonly used cancer cell lines (Universal Human Reference), we found 1207 additional full-length neoisoforms associated with junctions not detected in GTEx (data S15). This included the same SLC45A2 neojunction, predicted by SNAF-T to be associated with poor survival, shared in 69% of patients with melanoma, found to be MHC-presented (MS) and only expressed in tumors (fig. S10, A and B). This alternative 5′ donor site resulted in the deletion of 80 amino acids, which disrupted the sixth transmembrane domain. The deleted region (amino acids 215 to 295) appeared to be composed of a transmembrane segment (amino acids 215 to 237) and a cytoplasmic segment (amino acids 237 to 295) in the reference protein (fig. S10, C and D). As a result of the deletion, a region of polypeptide sequence normally positioned at the cytoplasmic face of the membrane was now predicted to reside in the extracellular domain, representing a potential new neoepitope.

Considering all SNAF-B short- and long-read predictions from 10 pan-cancer and 5 melanoma cell lines together, we ran SNAF-B to identify a total of 514 unique ExNeoEpitope proteins (data S15). We filtered these to 187 predictions in which the neojunction overlapped with an extracellular domain (UniProt) and was not contained within any other Ensembl or UCSC protein isoforms. In addition to proteoforms with missing polypeptide sequences, we identified 12 initial candidate long-read supported isoforms with high GTEx evidence of tumor specificity that result in the inclusion of undocumented alternative first or cassette exons. These were initially identified through manual inspection of BLAT sequence matches to the human genome and mRNA transcript databases (UCSC and Ensembl) and biased to ExNeoEpitopes detected in >15% of patients with melanoma. Visualization in the SNAF-B viewer found that 5 of these 12 ExNeoEpitopes result in new inserted polypeptide sequences that affect a cytoplasmic region of the protein (OCA2, SLC2A10, TMEM9, IL13RA1, and ATP13A1), one occurring within a transmembrane domain (ANO10), and 5 predicted to alter the extracellular region of the protein and result in stable transmembrane predictions (DCBLD2, NALCN, MET, SEMA6A, and IGSF11). Although orthogonal long-read RNA

Downloaded from https://www.science.org on July 31, 2025

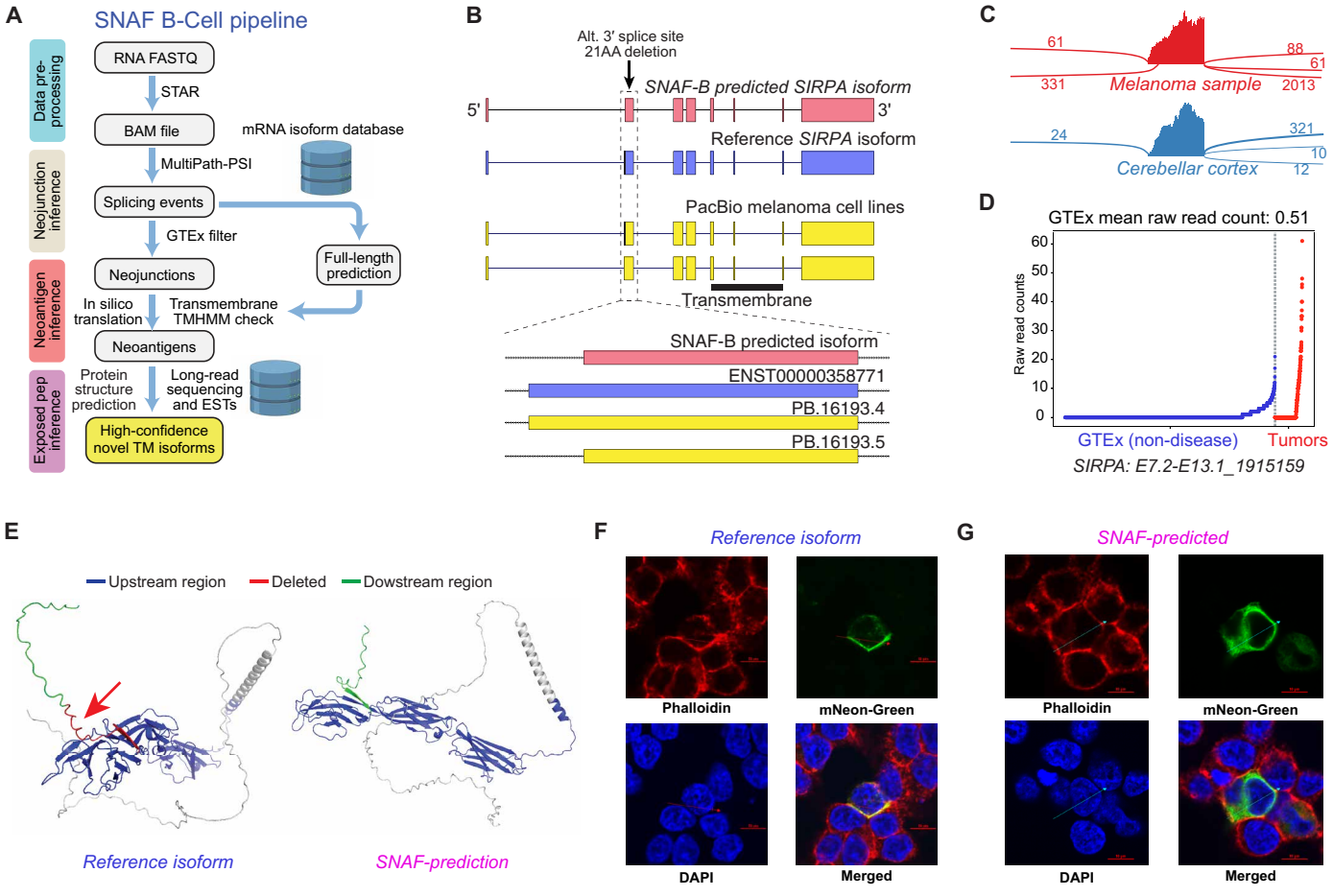

Fig. 7. SNAF-B finds full-length mRNAs and stable proteoforms for targeted therapies. (**A**) Overview of the SNAF-B prediction workflow to define ExNeoEpitopes. The workflow begins with bulk RNA-Seq datasets and optional long-read sequencing data integration to produce results with multiple levels of in silico evidence. (**B**) Comparison of a SNAF-B–predicted full-length isoform in the transmembrane protein SIRPA to documented mRNA isoforms and those predicted from PacBio long-read Iso-Seq of melanoma cell lines. (**C**) SashimiPlot of alternative 3′ splice site selection in melanoma and brain RNA-Seq for SIRPA. (**D**) Specificity of the indicated SIRPA ExNeoEpitope for TCGA melanoma samples versus an integrated healthy control tissue database (GTEx + TCGA). (**E**) Alphafold2 3D modeling of the reference isoform and the long-read verified ExNeoEpitope. Arrow denotes the deleted region in the alternative isoform. (**F** and **G**) Colocalization of the SIRPA reference (F) or melanoma-specific (G) splice isoform by confocal microscopy with a cell surface stain (phalloidin). The arrow indicates the cross section used to quantify fluorophore spatial coincidence.

sequencing indicated the validity of these transcripts from neojunction inference, it is possible that such isoforms are not properly folded or inserted into the cell membrane. First, to initially show the ability of these isoforms to produce functional transmembrane proteins, we predicted two-dimensional (2D) and 3D protein structures for the tumor-specific and reference mRNA isoforms using Protter (*70*) and Alphafold2 (*71*), respectively. In each of the examples, we observed high-confidence 2D and 3D structures of both the tumor and reference isoforms, suggesting that the deleted or inserted polypeptide sequence selectively affected the extracellular portion of these proteins (Fig. 7 and figs. S10 to S12). Finally, we tested the ability of ExNeoEpitopes with distinct in silico evidence to traffic to the plasma membrane. Specifically, we synthesized and transfected C-terminal fluorescent-tagged cDNAs for four ExNeoEpitopes that gained new peptide sequences (SEMA6A, ANO10, IGSF11, and DCBLD2) and three isoforms with deletions (SIRPA, MET, and SLC45A2), along with the reference isoform for each. The synthesized reference isoforms for five of the seven ExNeoEpitopes (SEMA6A, SIRPA, MET, SLC45A2, and DCBLD2) were expressed and found to colocalize to the transmembrane using a selective cell surface stain in human embryonic kidney (HEK)–293 cells. Confirming their expression and trafficking, we could also observe partial transmembrane localization in four of seven of the evaluated ExNeoEpitopes predicted by SNAF-B (SEMA6A, SIRPA, MET, and SLC45A2) (Fig. 7, F and G, and fig. S12, B to F). These data illustrate the potential of tumor-specific splice isoforms as higher precision candidates for CAR-T or monoclonal antibodies over existing targets, which include conformational epitopes that affect protein structure (*72*, *73*).

Interactive neoantigen web explorer

To facilitate the exploration and prioritization of the predicted neoantigens from SNAF, we developed two interactive web applications to visualize both T cell and B cell neoantigens (fig. S13). The SNAF-T viewer allows users to explore different global neoantigen features, including amino acid length and frequency within a cohort, in a projected 2D UMAP space. Here, each neoantigen is embedded based on its physicochemical properties. Users can manually select

Downloaded from https://www.science.org on July 31, 2025

clusters to identify enriched web logo amino acid motifs and explore individual neojunctions and neoantigens (movie S1). The SNAF-B viewer can be used to interactively visualize GTEx and tumor counts for neojunctions, generate protein sequence alignments between a putative ExNeoEpitope and selected reference proteins, compare protein feature composition, secondary structure, and solvability prediction (*74*), and perform topology modeling (*69*). Hence, these tools can be used to select optimal targets for experimental validation.

DISCUSSION

The identification and prioritization of shared neoantigens within and across cancers provides the potential to lead to new targeted immunotherapies. However, the development of existing targeted neoantigen immunotherapies is time consuming and costly, as they must exploit specific MHC-presented mutations that are evidenced by precision proteomics and immunogenicity assays (*75*). Our study provides evidence that aberrant splicing in melanoma frequently results in shared MHC-presented neoantigens, which can be confirmed in different patient cohorts and used to predict survival and response to immunotherapy. Patients with high splice neoantigen burden skew toward poor outcomes and associate with genes important to block immune tumor recruitment. Further, although prior computational strategies for splicing neoantigen discovery have been proposed, SNAF is unique in its inclusion of probabilistic modeling to quantify immunogenicity and tumor specificity, interactive exploratory methods, quantification of splicing factor activities, and interfaces for long-read and immunopeptidomics analysis. We attribute these new methods to our high validation rate. These analyses establish the broad existence of highly shared splicing neoantigens in melanoma and nominate coordinated splicing failure as a broad mediator of mis-splicing. Shared versus patient-specific splicing neoantigens were found to have distinct physicochemical characteristics and cells of origins, suggesting distinct mechanisms of regulation. From these analyses, one prediction, SLC45A2, stands out due to its (i) prognostic indication for immunotherapy outcomes, (ii) specificity for tumor versus immune cells, (iii) MHC-I stabilization, (iv) high immunogenicity, (v) support from long-read sequencing, and (vi) localization to the cell surface, making it an immensely promising target for future therapies.

A key challenge faced by SNAF and other tools is the identification of optimal targets for experimental validation. Improving such predictions in the future will depend on well-designed prospective studies, experimentally validating a range of predicted immunogenic and nonimmunogenic peptides (MHC presentation, immunogenicity) from mutations and splicing that are shared or unique and derived from different tools and statistical cutoffs. Similarly, the validation of ExNeoEpitopes will rely on new proteogenomics approaches that leverage targeted long isoform sequencing and proteomics along with antibodies that target specific conformational epitopes. Such antibodies could represent powerful new molecular reagents for CAR-T or monoclonal antibody strategies for shared and patient-specific neoantigens. Although our current pipeline enables the identification of likely ExNeoEpitopes and deep visual interrogation of the impact and position of residues in undocumented cancer protein isoforms, ultimately improved automated bioinformatics methods are needed to determine which introduced or removed residues will result specifically in new extracellular or transmembrane sequences that retain the conformational integrity of the new protein isoforms. Moreover, transposable elements have been reported to contribute to a subset of alternative splicing events (*76*) and give rise to neoantigens (*77*–*80*), including endogenous retrovirus (*81*). Although we do not observe such overlapping elements in our validated shared splicing neoantigens (UCSC genome browser), these warrant more systemic analysis to assess the potential convergence between these two types of neoantigen and their contribution, respectively.

A final important consideration is the tumor specificity of such neoantigens. As antigen assays do not provide information on tumor specificity, there is a need for more comprehensive normal tissue references. Most conventional RNA-Seq studies are on a limited set of adult human tissues, without considering rare cell types or fetal/pediatric developmental isoforms. It is likely that an expanded atlas of normal tissues, with extremely high sequencing depths (>100 million reads) and longer reads (>100 nucleotides), is needed, which may be aided by newly reported cheaper and longer sequencing approaches (*82*) and ideally single-cell resolution for hundreds of cell types.

Finally, it is important to note that our study has several limitations and outstanding questions. It has been suggested that cancer cells maintain a delicate balance between mutations in oncogenes and suppression of their presentation by MHC (*83*). Whether a similar mechanism exists for splicing neoantigens could further inform the type of therapy administered. Second, although current neoantigen predictions focus on HLA-I presentation and CD8 T cell function, HLA-II and CD4 T cells have also been reported to play an essential role in enhancing antitumor activities, together with other major immune cell types such as neutrophils and DCs (*84*). As current bioinformatics pipelines do not consider the activities of these other HLA mechanisms and T cell subsets, future methods may need to incorporate additional T cell and antigen presentation mechanisms. Further, to confirm SNAF observations, we leveraged existing immunopeptidomics and synthetic peptide MS data. Although it represents an important high-throughput validation, we note that MS-based neoantigen validations suffer from both false positives and false negatives due to the nontryptic nature of immunopeptides and the complex peptide search space considering combinatorial posttranslational modifications. The incomplete nature of the MS2 spectrum necessitates orthogonal assays and manual annotation to confirm individual splicing neoantigen predictions, which accounts for low prior reported MS/MS identification rates (5%) compared to normal proteome (50%) (*35*, *36*, *85*–*87*).

Given its flexibility, SNAF can be easily extended to new datasets, sequencing technologies, and neoantigen prediction libraries, which can be deployed in a modular manner in custom bioinformatics pipelines. Applied broadly to new cancers and distinct forms of malignancy, we believe that SNAF could be used to identify splicing neoantigens that are unique and shared across malignancies and discover new sequence motif preferences that expand the repertoire of targets for precision cancer therapy.

MATERIALS AND METHODS

Study design

The objective of this study was to comprehensively define tumor-specific and potentially immunogenic neoantigens produced from posttranscriptional regulation, particularly through alternative

Downloaded from https://www.science.org on July 31, 2025

splicing. A systematic pipeline for the identification of splicing neoantigens in heterogeneous cancers is presented as a strategy to reveal new shared targets for therapy. To broadly assess the presence and specificity of splicing neoantigens shared among patients with melanoma and ovarian cancer, our described bioinformatics workflow was applied to existing well-curated cancer molecular omics datasets. These cancers and datasets were selected as they have matched or unmatched multiomic measurements (immunopeptidome, RNA-Seq), clinical outcomes, and diverse therapy regimens. Bulk long-read RNA-Seq was applied in melanoma cell lines (one library replicate per cell line) to capture a sufficient diversity of full-length mRNA isoforms not necessarily present in other included cancer cell line long-read datasets. Because most of the analyses in the study are retrospective, sample size for these bulk RNA-Seq, immunoproteomics, and scRNA-Seq datasets is dependent on the original study design. For in vitro functional validation, neoantigen-MHC binding was confirmed using the TAP-deficient T2 cell line's capacity to stabilize HLA-A*02 upon the binding of candidate peptides. The immunogenicity and T cell reactivity of neoantigens were evaluated using peripheral blood collected from a minimum of three healthy donors. The selection of donors was contingent on the availability of MHC-I matches predicted by DeepImmuno and NetMHCpan SNAF. The experiment was conducted three times to ensure the reliability of the results, and each replication followed the same protocol and conditions to minimize variability and enhance the robustness of the findings. The analyses were blinded to the study participants by the clinical study coordinators.

SNAF architecture

SNAF was designed as a modular python package to automate splicing neoantigen identification using a series of embedded workflows. This workflow consists of distinct steps (Supplementary Materials and Methods) that are divided into functionally distinct modules, which can be called on a single SNAF python object or independently produced data files. These functions can be mixed and matched to identify T cell or B cell neoantigens or perform orthogonal, such as survival, MS proteomics, and long-read, analysis. Additional documentation and tutorials are provided from the GitHub repository (https://github.com/frankligy/SNAF). Specific SNAF algorithm details are provided in Supplementary Materials and Methods.

Bulk and single-cell melanoma splicing evaluation datasets

To evaluate shared and unique splicing neoantigens identified by SNAF, we reanalyzed prior reported bulk and scRNA-Seq datasets through SNAF using the same genome alignment and splicing quantification workflows applied to TCGA samples. To assess the association of melanoma splicing neoantigens with noncancerous proliferative skin cell splicing events, we reanalyzed five proliferative melanocyte RNA-Seq datasets in the Gene Expression Omnibus (GEO) database (GSE102983, GSE111786, GSE149189, GSE197471, and GSE202700) (fig. S9). To determine the cell of origin for melanoma splicing antigens, we obtained access to the controlled access raw sequencing data (DUOS-000002). This study obtained transcriptome measurements for 4645 single cells coming from 19 patients with melanoma (GSE72056). Only 3877 with cell annotations were retained for further analysis. These individual cell-level FASTQ files were re-analyzed in STAR and AltAnalyze to produce aggregate junction read counts for each patient and author-annotated cell populations. These junction read counts were summed per cell population to identify tumor versus immune neojunction enrichments (fold > 2 enriched). These analyses are biased toward immune cells, because twice as many immune cells (n = 2605) versus tumor (n = 1174) were present.

Peptide synthesis and MS spike-in validation

Thirty-six peptide candidate splicing neoantigens were synthesized (GenScript, Piscataway, NJ) at minimum of 70% purity with an average yield of 0.2 to 0.5 mg. Peptides were reconstituted with water to a final stock concentration of 1 pmol/μl. Peptides were pooled (except for LELLVKGTV and STLEFGLRV, which did not solubilize sufficiently) at a concentration of 1 pmol/μl and then diluted 1:10 for a working solution of 100 fmol/μl. Liquid chromatography–MS analysis was performed on a 50-fmol injection of pooled peptides using an Ultimate 3000 nanoflow high-performance liquid chromatography (HPLC) system (Dionex) and an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific) as described below. Injections were loaded onto an Acclaim PepMap 100 trap column (300 μm by 5 mm by 5 μm C18) and gradient-eluted from an Acclaim PepMap 100 analytical column (75 μm by 25 cm, 3 μm C18) equilibrated in 96% solvent A (0.1% formic acid in water) and 4% solvent B (80% acetonitrile in 0.1% formic acid). The peptides were eluted at 300 nl/min using the following gradient: 4% B from 0 to 5 min, 4 to 10% B from 5 to 10 min, 10 to 35% B from 10 to 60 min, 35 to 55% B from 60 to 70 min, 55 to 90% B from 70 to 71 min, 90% B from 71 to 73 min, 90 to 4%B from 73 to 74 min, and 4% B from 74 to 90 min. The Orbitrap Eclipse was operated in positive ion mode with 2.0 kV at the spray source, radio frequency lens at 30%, and data-dependent MS/MS acquisition with XCalibur version 4.3.73.11. Positive ion full MS scans were acquired in the Orbitrap from 375 to 1500 mass/charge ratio (m/z) with 120,000 resolution. Data-dependent selection of precursor ions was performed in cycle time mode, with 3 s in between master scans, using an intensity threshold of 2×10^4 ion counts and applying dynamic exclusion (n = 1 scan within 30 s for an exclusion duration of 60 s and ±10 ppm mass tolerance). Monoisotopic peak determination was applied, and charge states 2 to 6 were included for higher-energy collisional dissociation (HCD) MS2 scans (quadrupole isolation mode; 1.6 m/z isolation window; normalized collision energy at 30%). The resulting fragments were detected in the Orbitrap at 15,000 resolution with standard automatic gain control (AGC) target and dynamic maximum injection time mode.

Validation of peptide-MHC binding by MHC stabilization assay

To test MHC-I binding of synthesized neoantigen peptides, we used TAP-deficient T2 cells that are defective in transporters required for endogenous peptide loading (*88*, *89*). T2 cells were obtained from the American Type Culture Collection (ATCC) and grown at 37°C, 5% CO_2 in Iscove's modified Dulbecco's medium supplemented with 20% fetal bovine serum (FBS) and penicillin/streptomycin. T2 cells (1×10^5) were used without or loaded with peptides (100 μg/ml) and thereafter incubated overnight. All T2 cells were harvested and stained with HLA-A2–PE (phycoerythrin) antibody (clone BB7.2; BioLegend) for 30 min on ice. Cells were washed once with cell culture medium and acquired on a Fortessa II flow cytometer. Median fluorescence intensity was determined using FlowJo software.

Downloaded from https://www.science.org on July 31, 2025

Immunogenicity assay

Immunogenicity of predicted splicing neoantigen peptides was determined as described previously (*16*, *79*). In short, HLA-typed PBMCs from leukocyte reduction system (LRS) (Cincinnati Hoxworth Blood Center) chambers were isolated using Ficoll Hypaque density gradient centrifugation, aliquoted in 20×10^6 cells per vial, and frozen in liquid nitrogen until use. At day 0, PBMCs were thawed and used to set up monocyte-derived DCs by plating 4×10^6 PBMCs in a 24-well plate. Cells were incubated at 37°C, 5% CO_2 in DC medium [RPMI 1640 supplemented with 10% FBS, 1% L-glutamine (200 mM) + IL-4 (1000 U/ml), and granulocyte-macrophage colony-stimulating factor (GM-CSF) (800 U/ml)]. After 4 hours, the nonadherent fraction was removed by rinsing the wells twice with phosphate-buffered saline (PBS). Adherent cells were cultured for 7 days in DC medium. On day 7, DCs were loaded with peptides (10 μg/ml) dissolved in DC medium and incubated at 37°C, 5% CO_2. After 4 hours, 1.5 ml of DC maturation medium was added [RPMI 1640 supplemented with 10% FBS, 1% L-glutamine, IL-4 (1000 U/ml), GM-CSF (800 U/ml), IL-1β (10 ng/ml), IL-6 (10 ng/ml), tumor necrosis factor–α (TNF-α) (10 ng/ml), and lipopolysaccharide (LPS) (30 ng/ml)]. After 16 hours of DC maturation, peptide-loaded DCs were used to stimulate autologous PBMCs by adding 1×10^6 PBMCs of the same donor to the DC cultures. DC and PBMC cocultures were grown in T cell medium [60% RPMI 1640, 40% Click's medium supplemented with 10% FBS, 1% L-glutamine, IL-6 (100 ng/ml), IL-7 (10 ng/ml), IL-12 (10 ng/ml), and IL-15 (5 ng/ml)]. The medium was changed on days 3 and 6 based on medium color change. On days 14 and 21, T cells were harvested and stimulated with new autologous peptide-loaded DCs. After three rounds of T cell priming, at day 28, T cells were harvested and tested for their IFN-γ response to peptide-loaded 721.221 (221) single HLA antigen target cells. The target cells 221.A*02:01, 221.C*04:01, and 221.C*08:01 were loaded with relevant peptides at 10 μg/ml in separate wells. 221 target cells and peptides were incubated for 4 hours at 37°C, 5% CO_2 in RPMI 1640 supplemented with 10% FBS. Peptide-loaded 221 target cells with and without peptides were cocultured with primed T cells at a 3:1 ratio in the presence of phorbol 12-myristate 13-acetate (PMA) (50 ng/ml) for 6 hours at 37°C, 5% CO_2 in RPMI 1640 supplemented with 10% FBS and monensin. PMA and ionomycin stimulation (each at 1 μg/ml) was used as positive control. Thereafter, the cells were harvested and stained for extracellular CD45-BV786 (BioLegend), CD14-PerCP (peridinin chlorophyll protein) (BioLegend), and CD8-PE (BioLegend) for 30 min on ice. Cells were fixed and permeabilized using the CytoFix/CytoPerm kit (BD) according to the manufacturer's instructions and thereafter stained for intracellular IFN-γ–APC (allophycocyanin) expression (clone 4S.B3; BioLegend) for 20 min on ice and directly analyzed on a BD Fortessa flow cytometer. Analysis of $CD45^+CD14^-CD8^+$ IFN-γ^+ cells was determined using FlowJo software.

Validation of ExNeoEpitope localization

To confirm the expression and cell membrane localization of SNAF-B neoisoforms, we synthesized the long-read sequencing evidenced alternative isoforms and their annotated reference isoforms as C-terminal–tagged cDNAs with either mNeon-Green or enhanced green fluorescent protein (eGFP) on a plasmid vector (VectorBuilder, USA). Streak LB agar plates with ampicillin (100 μg/ml) were made for each isoform. A single colony was picked from each plate and expanded in 1 ml of LB broth for 8 hours at 37°C. The pre-expansion broth (20 μl) was then taken and pipetted into an Elenmyer flask with 50 ml of LB broth. The competent cells were expanded overnight at 37°C. Medi-preps were performed for each construct with the ZymoPure II plasmid midiprep kit. On an Ibidi four-well chamber μ-slide, human embryonic kidney (HEK)–293T cells were seeded in a concentration of 0.15 M/ml. When HEK-293T cells reached 60% confluency, these constructs (CMV promoter) were transfected into the cells separately with 1 μg of plasmid (99% pUC19-negative control plasmid + 1% engineered plasmid) with TransIT-LT1 (Mirus) following the manufacturer's protocol.

The cells were fixed and permeabilized with 4% paraformaldehyde 24 hours after transfection. After two rounds of washing, the cells were treated with a warm 1× citrate buffer (diluted from 10× stock; Sigma-Aldrich) to break the protein cross-links. The cells were then washed once with PBS again, and a membrane actin stain was performed with 1× phalloidin 647 in PBS with 1% bovine serum albumin (BSA) (Abcam) at room temperature for an hour. To stain the nuclei of the fixed cells, the cells were washed with PBS and stained with DAPI (4′,6-diamidino-2-phenylindole) (Thermo Fisher Scientific) (1:4000 diluted in PBS with 1% BSA) at room temperature for 5 min. The stained cells were immediately washed with PBS, and the PBS was removed by vacuum. ProLong Gold mounting medium (Thermo Fisher Scientific) was evenly applied to the fixed cell surface. Twenty-four hours after transfection, the expression and colocalization of the reference and alternative isoforms were assessed with a confocal spinning disk microscope (Yokogawa SoRa W1 dual camera system) using the Nikon Elements software. The confocal raw imaging files with all z-stacks are provided in Synapse (https://synapse.org/#!Synapse:syn52063953).

Melanoma cell line long-read isoform sequencing

For long-read mRNA isoform sequencing, cells were grown and isolated from five independent melanoma cell lines: A375, SKMEL, MeWo, UACC62, and UACC257 (ATCC). Total RNA was isolated (TRIzol) and analyzed on a Thermo Nanodrop UV-Vis and an Agilent Bioanalyzer to confirm the nominal concentration and to ensure RNA integrity. From the RNA, cDNA was synthesized using the Clontech SMARTr cDNA Synthesis Kit, in which a barcode was added to the oligo-dT at the 3′ end. Each melanoma cell line cDNA was pooled and then converted into an SMRTbell library using the Iso-Seq Express Kit SMRT Bell Express Template prep kit 2.0 (Pacific Biosciences). We sequenced each library on an SMRT cell on the Sequel II system using a 30-hour movie collection time. The "ccs" command from the PacBio SMRTLink suite (SMRTLink version 9) was used to convert raw reads into circular consensus sequence (CCS) reads. The resulting data were analyzed in SQUANTI to assign reads to full-length collapsed reference or neoisoforms. The isoform GTF files, barcode sequences, and raw data are available in Synapse (https://synapse.org/#!Synapse:syn32785802) (table S1).

Melanoma RNA-Seq analyses

RNA-Seq paired-end BAM files from 472 patients with melanoma collected by TCGA (SKCM) were obtained from the GDC portal after dbGAP approval (phs000178.v10.p8). A second collection of 40 RNA-Seq FASTQ files from patients who underwent immunotherapy (archival formalin-fixed, paraffin-embedded) in Van Allen cohort (*40*) was obtained from the dbGAP database (phs000452.v2.p1). These FASTQ files were aligned to the reference human

Downloaded from https://www.science.org on July 31, 2025

genome (hg38) and transcriptome (Ensembl 91) using STAR. Among these 40 Van Allen RNA-Seq samples, patient 41 was excluded (only partial sequencing data available), as previously reported (*90*). The HLA genotype of each patient sample was determined from the RNA-Seq FASTQ files using the software Optitype 1.3.3 (*91*). We chose Optitype based on its superior performance in calling HLA-I alleles (over 99% accuracy) from RNA-Seq data based on a recent large-scale benchmarking study, evaluated on "gold standard" HLA genotyping data (*92*, *93*). AltAnalyze v2.1.4 was used to quantify splicing independently in these two cohorts using the Ensembl version 91 database. MultiPath-PSI–identified splicing events were used as inputs for SNAF. The TCGA survival and mutation data were downloaded from Xena Browser (*94*). Survival analysis was performed using the snaf.survival_analysis function with stratification argument $n = 2$ (high burden is equivalent to greater than the median burden, low burden is equivalent to less than the median burden, with the outliers excluded). Mutation analysis was conducted using snaf.mutation_analysis function. To identify individual neojunctions or neoantigens associated with survival, a univariate Cox regression analysis was used to identify events/antigens with a parental percent spliced in (PSI) value that was positively or negatively associated with patient outcome (Wald test P value and z score). The neojunction and its parental PSI value were ignored if the neoantigen was not predicted to be presented in that sample, resulting in different survival associations for different neoantigens produced from the same neojunction. For analysis of melanoma RNA-Seq TCGA samples in the SNAF-B workflow, long-read Iso-Seq cDNA sequences were obtained from pan-cancer cell line sequencing using the PacBio-provided isoform GTF file (https://downloads.pacbcloud.com/public/dataset/UHRRisoseq2021/Final-MappedTranscripts/).

Statistics

Replicate sample genomic comparative analyses used a two-sided empirical Bayes moderated t test ($P < 0.05$) for all bulk RNA-Seq gene expression and alternative splicing analyses. All analyses in which greater than 30 measurements were obtained were subjected to false discovery testing procedure (Benjamini-Hochberg). Associations for individual neojunctions or neoantigens with patient survival were derived using a univariate Cox regression analysis, with positive and negative association results reported for all results and significant results reported for a Coxph < 0.05.

Supplementary Materials

This PDF file includes:
Materials and Methods
Figs. S1 to S13
Table S1
References (*88*–*199*)

Other Supplementary Material for this manuscript includes the following:
Data S1 to S15
Movie S1
MDAR Reproducibility Checklist

REFERENCES AND NOTES

1. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).
2. E. Tran, P. F. Robbins, S. A. Rosenberg, 'Final common pathway' of human cancer immunotherapy: Targeting random somatic mutations. *Nat. Immunol.* **18**, 255–262 (2017).
3. P. A. Ott, Z. Hu, D. B. Keskin, S. A. Shukla, J. Sun, D. J. Bozym, W. Zhang, A. Luoma, A. Giobbie-Hurder, L. Peter, C. Chen, O. Olive, T. A. Carter, S. Li, D. J. Lieb, T. Eisenhaure, E. Gjini, J. Stevens, W. J. Lane, I. Javeri, K. Nellaiappan, A. M. Salazar, H. Daley, M. Seaman, E. I. Buchbinder, C. H. Yoon, M. Harden, N. Lennon, S. Gabriel, S. J. Rodig, D. H. Barouch, J. C. Aster, G. Getz, K. Wucherpfennig, D. Neuberg, J. Ritz, E. S. Lander, E. F. Fritsch, N. Hacohen, C. J. Wu, An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).
4. Z. Hu, D. E. Leet, R. L. Allesøe, G. Oliveira, S. Li, A. M. Luoma, J. Liu, J. Forman, T. Huang, J. B. Iorgulescu, R. Holden, S. Sarkizova, S. H. Gohil, R. A. Redd, J. Sun, L. Elagina, A. Giobbie-Hurder, W. Zhang, L. Peter, Z. Ciantra, S. Rodig, O. Olive, K. Shetty, J. Pyrdol, M. Uduman, P. C. Lee, P. Bachireddy, E. I. Buchbinder, C. H. Yoon, D. Neuberg, B. L. Pentelute, N. Hacohen, K. J. Livak, S. A. Shukla, L. R. Olsen, D. H. Barouch, K. W. Wucherpfennig, E. F. Fritsch, D. B. Keskin, C. J. Wu, P. A. Ott, Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. *Nat. Med.* **27**, 515–525 (2021).
5. J. Bauman, H. Burris, J. Clarke, M. Patel, D. Cho, M. Gutierrez, R. Julian, A. Scott, P. Cohen, J. Frederick, C. Robert-Tissot, H. Zhou, K. Mody, K. Keating, R. Meehan, J. Gainor, 798 Safety, tolerability, and immunogenicity of mRNA-4157 in combination with pembrolizumab in subjects with unresectable solid tumors (KEYNOTE-603): An update. *J. Immunother. Cancer* **8**, 0798 (2020).
6. S. L. Goff, M. E. Dudley, D. E. Citrin, R. P. Somerville, J. R. Wunderlich, D. N. Danforth, D. A. Zlott, J. C. Yang, R. M. Sherry, U. S. Kammula, C. A. Klebanoff, M. S. Hughes, N. P. Restifo, M. M. Langhan, T. E. Shelton, L. Lu, M. L. M. Kwong, S. Ilyas, N. D. Klemen, E. C. Payabyab, K. E. Morton, M. A. Toomey, S. M. Steinberg, D. E. White, S. A. Rosenberg, Randomized, prospective evaluation comparing intensity of lymphodepletion before adoptive transfer of tumor-infiltrating lymphocytes for patients with metastatic melanoma. *J. Clin. Oncol.* **34**, 2389–2397 (2016).
7. M. S. Lawrence, P. Stojanov, P. Polak, G. V. Kryukov, K. Cibulskis, A. Sivachenko, S. L. Carter, C. Stewart, C. H. Mermel, S. A. Roberts, A. Kiezun, P. S. Hammerman, A. McKenna, Y. Drier, L. Zou, A. H. Ramos, T. J. Pugh, N. Stransky, E. Helman, J. Kim, C. Sougnez, L. Ambrogio, E. Nickerson, E. Shefler, M. L. Cortés, D. Auclair, G. Saksena, D. Voet, M. Noble, D. DiCara, P. Lin, L. Lichtenstein, D. I. Heiman, T. Fennell, M. Imielinski, B. Hernandez, E. Hodis, S. Baca, A. M. Dulak, J. Lohr, D.-A. Landau, C. J. Wu, J. Melendez-Zajgla, A. Hidalgo-Miranda, A. Koren, S. A. McCarroll, J. Mora, B. Crompton, R. Onofrio, M. Parkin, W. Winckler, K. Ardlie, S. B. Gabriel, C. W. M. Roberts, J. A. Biegel, K. Stegmaier, A. J. Bass, L. A. Garraway, M. Meyerson, T. R. Golub, D. A. Gordenin, S. Sunyaev, E. S. Lander, G. Getz, Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
8. M. Yarchoan, A. Hopkins, E. M. Jaffee, Tumor mutational burden and response rate to PD-1 Inhibition. *N. Engl. J. Med.* **377**, 2500–2501 (2017).
9. T. W. Nilsen, B. R. Graveley, Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**, 457–463 (2010).
10. S. M. Park, J. Ou, L. Chamberlain, T. M. Simone, H. Yang, C.-M. Virbasius, A. M. Ali, L. J. Zhu, S. Mukherjee, A. Raza, M. R. Green, U2AF35(S34F) promotes transformation by directing aberrant ATG7 pre-mRNA 3′ end formation. *Mol. Cell* **62**, 479–490 (2016).
11. S. Calabretta, P. Bielli, I. Passacantilli, E. Pilozzi, V. Fendrich, G. Capurso, G. D. Fave, C. Sette, Modulation of PKM alternative splicing by PTBP1 promotes gemcitabine resistance in pancreatic cancer cells. *Oncogene* **35**, 2031–2039 (2016).
12. S.-J. Zhang, R. Rampal, T. Manshouri, J. Patel, N. Mensah, A. Kayserian, T. Hricik, A. Heguy, C. Hedvat, M. Gönen, H. Kantarjian, R. L. Levine, O. Abdel-Wahab, S. Verstovsek, Genetic analysis of patients with leukemic transformation of myeloproliferative neoplasms shows recurrent *SRSF2* mutations that are associated with adverse outcome. *Blood* **119**, 4480–4485 (2012).
13. M. J. Oudin, S. K. Hughes, N. Rohani, M. N. Moufarrej, J. G. Jones, J. S. Condeelis, D. A. Lauffenburger, F. B. Gertler, Characterization of the expression of the pro-metastatic Mena^{INV} isoform during breast tumor progression. *Clin. Exp. Metastasis* **33**, 249–261 (2016).
14. A. Kahles, K. V. Lehmann, N. C. Toussaint, M. Hüser, S. G. Stark, T. Sachsenberg, O. Stegle, O. Kohlbacher, C. Sander, G. Rätsch, Comprehensive analysis of alternative splicing across tumors from 8,705 patients. *Cancer Cell* **34**, 211–224.e6 (2018).
15. O. D. Rivera, M. J. Mallory, M. Quesnel-Vallières, R. Chatrikhi, D. C. Schultz, M. Carroll, Y. Barash, S. Cherry, K. W. Lynch, Alternative splicing redefines landscape of commonly mutated genes in acute myeloid leukemia. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2014967118 (2021).
16. S. Rivero-Hinojosa, M. Grant, A. Panigrahi, H. Zhang, V. Caisova, C. M. Bollard, B. R. Rood, Proteogenomic discovery of neoantigens facilitates personalized multi-antigen targeted T cell immunotherapy for brain tumors. *Nat. Commun.* **12**, 6689 (2021).
17. A. C. Smart, C. A. Margolis, H. Pimentel, M. X. He, D. Miao, D. Adeegbe, T. Fugmann, K.-K. Wong, E. M. Van Allen, Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* **36**, 1056–1058 (2018).
18. G. Ehx, J.-D. Larouche, C. Durette, J.-P. Laverdure, L. Hesnard, K. Vincent, M.-P. Hardy, C. Thériault, C. Rulleau, J. Lanoix, E. Bonneil, A. Feghaly, A. Apavaloaei, N. Noronha,

Downloaded from https://www.science.org on July 31, 2025

C. M. Laumont, J.-S. Delisle, L. Vago, J. Hébert, G. Sauvageau, S. Lemieux, P. Thibault, C. Perreault, Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. *Immunity* **54**, 737–752.e10 (2021).
19. G. Li, B. Iyer, V. B. S. Prasath, Y. Ni, N. Salomonis, DeepImmuno: Deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity. *Brief. Bioinform.* **22**, bbab160 (2021).
20. V. Narayan, J. S. Barber-Rotenberg, I. Y. Jung, S. F. Lacey, A. J. Rech, M. M. Davis, W. T. Hwang, P. Lal, E. L. Carpenter, S. L. Maude, G. Plesa, N. Vapiwala, A. Chew, M. Moniak, R. A. Sebro, A. M. Farwell, J. Gilmore, L. Lledo, K. Dengel, S. E. Church, T. D. Hether, J. Xu, M. Gohil, T. H. Buckingham, S. S. Yee, V. E. Gonzalez, I. Kulikovskaya, F. Chen, L. Tian, K. Tien, W. Gladney, C. L. Nobles, H. E. Raymond, E. O. Hexner, D. L. Siegel, F. D. Bushman, C. H. June, J. A. Fraietta, N. B. Haas, PSMA-targeting TGFβ-insensitive armored CAR T cells in metastatic castration-resistant prostate cancer: A phase 1 trial. *Nat. Med.* **28**, 724–734 (2022).
21. A. M. Fenix, Y. Miyaoka, A. Bertero, S. M. Blue, M. J. Spindler, K. K. B. Tan, J. A. Perez-Bermejo, A. H. Chan, S. J. Mayerl, T. D. Nguyen, C. R. Russell, P. P. Lizarraga, A. Truong, P.-L. So, A. Kulkarni, K. Chetal, S. Sathe, N. J. Sniadecki, G. W. Yeo, C. E. Murry, B. R. Conklin, N. Salomonis, Gain-of-function cardiomyopathic mutations in RBM20 rewire splicing regulation and re-distribute ribonucleoprotein granules within processing bodies. *Nat. Commun.* **12**, 6324 (2021).
22. S. Shen, J. W. Park, Z.-X. Lu, L. Lin, M. D. Henry, Y. N. Wu, Q. Zhou, Y. Xing, rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E5593–E5601 (2014).
23. Y. I. Li, D. A. Knowles, J. Humphrey, A. N. Barbeira, S. P. Dickinson, H. K. Im, J. K. Pritchard, Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2017).
24. J. Vaquero-Garcia, A. Barrera, M. R. Gazzara, J. González-Vallinas, N. F. Lahens, J. B. Hogenesch, K. W. Lynch, Y. Barash, A new view of transcriptome complexity and regulation through the lens of local splicing variations. *eLife* **5**, e11752 (2016).
25. S. S. Itskovich, A. Gurunathan, J. Clark, M. Burwinkel, M. Wunderlich, M. R. Berger, A. Kulkarni, K. Chetal, M. Venkatasubramanian, N. Salomonis, A. R. Kumar, L. H. Lee, MBNL1 regulates essential alternative RNA splicing patterns in MLL-rearranged leukemia. *Nat. Commun.* **11**, 2369 (2020).
26. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
27. B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
28. T. J. O'Donnell, A. Rubinsteyn, U. Laserson, MHCflurry 2.0: Improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst.* **11**, 42–48.e7 (2020).
29. G. Li, A. Bhattacharjee, N. Salomonis, Quantifying tumor specificity using Bayesian probabilistic modeling for drug target discovery and prioritization. bioRxiv 2023.03.03.530994 [Preprint] (2023); https://doi.org/10.1101/2023.03.03.530994.
30. Y. Pan, J. W. Phillips, B. D. Zhang, M. Noguchi, E. Kutschera, J. McLaughlin, P. A. Nesterenko, Z. Mao, N. J. Bangayan, R. Wang, W. Tran, H. T. Yang, Y. Wang, Y. Xu, M. B. Obusan, D. Cheng, A. H. Lee, K. E. Kadash-Edmondson, A. Champhekar, C. Puig-Saus, A. Ribas, R. M. Prins, C. S. Seet, G. M. Crooks, O. N. Witte, Y. Xing, IRIS: Discovery of cancer immunotherapy targets arising from pre-mRNA alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2221116120 (2023)
31. S. Chai, C. C. Smith, T. K. Kochar, S. A. Hunsucker, W. Beck, K. S. Olsen, S. Vensko, G. L. Glish, P. M. Armistead, J. F. Prins, B. G. Vincent, NeoSplice: a bioinformatics method for prediction of splice variant neoantigens. *Bioinform. Adv.* **2**, vbac032 (2022).
32. Z. Zhang, C. Zhou, L. Tang, Y. Gong, Z. Wei, G. Zhang, F. Wang, Q. Liu, J. Yu, ASNEO: Identification of personalized alternative splicing based neoantigens with RNA-seq. *Aging (Albany, NY)* **12**,14633 (2020).
33. H. Schuster, J. K. Peper, H.-C. Bösmüller, K. Röhle, L. Backert, T. Bilich, B. Ney, M. W. Löffler, D. J. Kowalewski, N. Trautwein, A. Rabsteyn, T. Engler, S. Braun, S. P. Haen, J. S. Walz, B. Schmid-Horch, S. Y. Brucker, D. Wallwiener, O. Kohlbacher, F. Fend, H.-G. Rammensee, S. Stevanović, A. Staebler, P. Wagner, The immunopeptidomic landscape of ovarian carcinomas. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E9942–E9951 (2017).
34. M. Bassani-Sternberg, E. Bräunlein, R. Klar, T. Engleitner, P. Sinitcyn, S. Audehm, M. Straub, J. Weber, J. Slotta-Huspenina, K. Specht, M. E. Martignoni, A. Werner, R. Hein, D. H. Busch, C. Peschel, R. Rad, J. Cox, M. Mann, A. M. Krackhardt, Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **7**, 1–16 (2016).
35. M. Wilhelm, D. P. Zolg, M. Graber, S. Gessulat, T. Schmidt, K. Schnatbaum, C. Schwencke-Westphal, P. Seifert, N. de Andrade Krätzig, J. Zerweck, T. Knaute, E. Bräunlein, P. Samaras, L. Lautenbacher, S. Klaeger, H. Wenschuh, R. Rad, B. Delanghe, A. Huhmer, S. A. Carr, K. R. Clauser, A. M. Krackhardt, U. Reimer, B. Kuster, Deep learning boosts sensitivity of mass spectrometry-based immunopeptidomics. *Nat. Commun.* **12**, 3346 (2021).
36. A. Declercq, R. Bouwmeester, A. Hirschler, C. Carapito, S. Degroeve, L. Martens, R. Gabriels, MS^{2}Rescore: Data-driven rescoring dramatically boosts immunopeptide identification rates. *Mol. Cell. Proteomics* **21**, 100266 (2022).
37. C. Shen, Q. Sheng, J. Dai, Y. Li, R. Zeng, H. Tang, On the estimation of false positives in peptide identifications using decoy search strategy. *Proteomics* **9**, 194–204 (2009).
38. R. M. Miller, B. T. Jordan, M. M. Mehlferber, E. D. Jeffery, C. Chatzipantsiou, S. Kaur, R. J. Millikin, Y. Dai, S. Tiberi, P. J. Castaldi, M. R. Shortreed, C. J. Luckey, A. Conesa, L. M. Smith, A. Deslattes Mays, G. M. Sheynkman, Enhanced protein isoform characterization through long-read proteogenomics. *Genome Biol.* **23**, 69 (2022).
39. Y. Liu, Y. Chen, X. Hu, J. Meng, X. Li, Development and validation of the B cell-associated Fc receptor-like molecule-based prognostic signature in skin cutaneous melanoma. *Biomed. Res. Int.* **2020**, 8509805 (2020).
40. E. M. Van Allen, D. Miao, B. Schilling, S. A. Shukla, C. Blank, L. Zimmer, A. Sucker, U. Hillen, M. H. G. Foppen, S. M. Goldinger, J. Utikal, J. C. Hassel, B. Weide, K. C. Kaehler, C. Loquai, P. Mohr, R. Gutzmer, R. Dummer, S. Gabriel, C. J. Wu, D. Schadendorf, L. A. Garraway, Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
41. Cancer Genome Atlas Network, Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
42. T. M. Smith Jr., A. Tharakan, R. K. Martin, Targeting ADAM10 in cancer and autoimmunity. *Front. Immunol.* **11**, 499 (2020).
43. Y. Wang, M. Mohseni, A. Grauel, J. E. Diez, W. Guan, S. Liang, J. E. Choi, M. Pu, D. Chen, T. Laszewski, S. Schwartz, J. Gu, L. Mansur, T. Burks, L. Brodeur, R. Velazquez, S. Kovats, B. Pant, G. Buruzula, E. Deng, J. T. Chen, F. Sari-Sarraf, C. Dornelas, M. Varadarajan, H. Yu, C. Liu, J. Lim, H.-X. Hao, X. Jiang, A. Malamas, M. J. LaMarche, F. C. Geyer, M. McLaughlin, C. Costa, J. Wagner, D. Ruddy, P. Jayaraman, N. D. Kirkpatrick, P. Zhang, O. Iartchouk, K. Aardalen, V. Cremasco, G. Dranoff, J. A. Engelman, S. Silver, H. Wang, W. D. Hastings, S. Goldoni, SHP2 blockade enhances anti-tumor immunity via tumor cell intrinsic and extrinsic mechanisms. *Sci. Rep.* **11**, 1399 (2021).
44. E. Batlle, J. Massagué, Transforming growth factor-β signaling in immunity and cancer. *Immunity* **50**, 924–940 (2019).
45. B. Yang, J. Chen, Y. Teng, TNPO1-mediated nuclear import of FUBP1 contributes to tumor immune evasion by increasing NRP1 expression in cervical cancer. *J. Immunol. Res.* **2021**, 9994004 (2021).
46. T.-Y. Song, M. Long, H.-X. Zhao, M.-W. Zou, H.-J. Fan, Y. Liu, C.-L. Geng, M.-F. Song, Y.-F. Liu, J.-Y. Chen, Y.-L. Yang, W.-R. Zhou, D.-W. Huang, B. Peng, Z.-G. Peng, Y. Cang, Tumor evolution selectively inactivates the core microRNA machinery for immune evasion. *Nat. Commun.* **12**, 7003 (2021).
47. T. Vlaykova, L. Talve, M. Hahka-Kemppinen, M. Hernberg, T. Muhonen, K. Franssila, Y. Collan, S. Pyrhönen, MIB-1 immunoreactivity correlates with blood vessel density and survival in disseminated malignant melanoma. *Oncology* **57**, 242–252 (1999).
48. Y. Gao, H. Zheng, L. Li, C. Zhou, X. Chen, X. Zhou, Y. Cao, KIF3C promotes proliferation, migration, and invasion of glioma cells by activating the PI3K/AKT pathway and inducing EMT. *Biomed. Res. Int.* **2020**, 6349312 (2020).
49. L. Tang, F. Wei, Y. Wu, Y. He, L. Shi, F. Xiong, Z. Gong, C. Guo, X. Li, H. Deng, K. Cao, M. Zhou, B. Xiang, X. Li, Y. Li, G. Li, W. Xiong, Z. Zeng, Role of metabolism in cancer cell radioresistance and radiosensitization methods. *J. Exp. Clin. Cancer Res.* **37**, 87 (2018).
50. E. A. Zaal, C. R. Berkers, The influence of metabolism on drug response in cancer. *Front. Oncol.* **8**, 500 (2018).
51. F. A. Dick, D. W. Goodrich, J. Sage, N. J. Dyson, Non-canonical functions of the RB protein in cancer. *Nat. Rev. Cancer* **18**, 442–451 (2018).
52. M. Shen, Z. Xu, W. Xu, K. Jiang, F. Zhang, Q. Ding, Z. Xu, Y. Chen, Inhibition of ATM reverses EMT and decreases metastatic potential of cisplatin-resistant lung cancer cells through JAK/STAT3/PD-L1 pathway. *J. Exp. Clin. Cancer Res.* **38**, 149 (2019).
53. P. Deng, Z. Wang, J. Chen, S. Liu, X. Yao, S. Liu, L. Liu, Z. Yu, Y. Huang, Z. Xiong, R. Xiao, J. Gao, W. Liang, J. Chen, H. Liu, J. H. Hong, J. Y. Chan, P. Guan, J. Chen, Y. Wang, J. Yin, J. Li, M. Zheng, C. Zhang, P. Zhou, T. Kang, B. T. Teh, Q. Yu, Z. Zuo, Q. Jiang, J. Liu, Y. Xiong, X. Xia, J. Tan, RAD21 amplification epigenetically suppresses interferon signaling to promote immune evasion in ovarian cancer. *J. Clin. Invest.* **132**, e159628 (2022).
54. M. Rossi Sebastiano, G. Konstantinidou, Targeting long chain acyl-CoA synthetases for cancer therapy. *Int. J. Mol. Sci.* **20**, 3624 (2019).
55. N. Germain, M. Dhayer, M. Boileau, Q. Fovez, J. Kluza, P. Marchetti, Lipid metabolism and resistance to anticancer treatment. *Biology* **9**, 474 (2020).
56. G. Li, Y. Jiang, G. Li, Q. Qiao, Comprehensive analysis of radiosensitivity in head and neck squamous cell carcinoma. *Radiother. Oncol.* **159**, 126–135 (2021).
57. S. Wang, X. Yi, Z. Wu, S. Guo, W. Dai, H. Wang, Q. Shi, K. Zeng, W. Guo, C. Li, CAMKK2 defines ferroptosis sensitivity of melanoma cells by regulating AMPK–NRF2 pathway. *J. Invest. Dermatol.* **142**, 189–200.e8 (2022).
58. E. R. Miraldi, M. Pokrovskii, A. Watters, D. M. Castro, N. De Veaux, J. A. Hall, J. Y. Lee, M. Ciofani, A. Madar, N. Carriero, D. R. Littman, R. Bonneau, Leveraging chromatin accessibility for transcriptional regulatory network inference in T Helper 17 Cells. Genome Res. **29**, 449–463 (2019).

Downloaded from https://www.science.org on July 31, 2025

59. V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts, Inferring regulatory networks from expression data using tree-based methods. *PLOS ONE* **5**, e12776 (2010).
60. P. Badia-I-Mompel, S. J. Vélez, J. Braunger, C. Geiss, D. Dimitrov, S. Müller-Dott, P. Taus, A. Dugourd, C. H. Holland, R. F. Ro, J. Saez-Rodriguez, decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Adv.* **2**, vbac016(2022).
61. J. D. Thomas, S. X. Lu, E. De Neef, E. Sabio, B. Rousseau, M. Gigoux, D. A. Knorr, B. Greenbaum, Y. Elhanati, S. J. Hogg, A. Chow, A. Ghosh, A. Xie, D. Zamarin, D. Cui, C. Erickson, M. Singer, H. Cho, E. Wang, B. Lu, B. H. Durham, H. Shah, D. Chowell, A. M. Gabel, Y. Shen, J. Liu, J. Jin, M. C. Rhodes, R. E. Taylor, H. Molina, J. D. Wolchok, T. Merghoub, L. A. D. Jr, O. Abdel-Wahab, R. K. Bradley, Abstract 5742: Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cancer Res.* **83**, 5742 (2023).
62. L. Wan, K. T. Lin, M. A. Rahman, Y. Ishigami, Z. Wang, M. A. Jensen, J. E. Wilkinson, Y. Park, D. A. Tuveson, A. R. Krainer, Splicing factor SRSF1 promotes pancreatitis and $KRAS^{G12D}$-mediated pancreatic cancer. *Cancer Discov.* **13**, 1678–1695 (2023).
63. T. L. Bailey, Fitting a mixture model by expectation maximization to discover motifs in bipolymers (UCSD Technical Report CS94-351, 1994).
64. W. Yang, K.-W. Lee, R. M. Srivastava, F. Kuo, C. Krishna, D. Chowell, V. Makarov, D. Hoen, M. G. Dalin, L. Wexler, R. Ghossein, N. Katabi, Z. Nadeem, M. A. Cohen, S. K. Tian, N. Robine, K. Arora, H. Geiger, P. Agius, N. Bouvier, K. Huberman, K. Vanness, J. J. Havel, J. S. Sims, R. M. Samstein, R. Mandal, J. Tepe, I. Ganly, A. L. Ho, N. Riaz, R. J. Wong, N. Shukla, T. A. Chan, L. G. T. Morris, Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat. Med.* **25**, 767–775 (2019).
65. S. X. Lu, E. De Neef, J. D. Thomas, E. Sabio, B. Rousseau, M. Gigoux, D. A. Knorr, B. Greenbaum, Y. Elhanati, S. J. Hogg, A. Chow, A. Ghosh, A. Xie, D. Zamarin, D. Cui, C. Erickson, M. Singer, H. Cho, E. Wang, B. Lu, B. H. Durham, H. Shah, D. Chowell, A. M. Gabel, Y. Shen, J. Liu, J. Jin, M. C. Rhodes, R. E. Taylor, H. Molina, J. D. Wolchok, T. Merghoub, L. A. Diaz Jr., O. Abdel-Wahab, R. K. Bradley, Pharmacologic modulation of RNA splicing enhances anti-tumor immunity. *Cell* **184**, 4032–4047.e31 (2021).
66. M. A. Savanur, H. Weinstein-Marom, G. Gross, Implementing logic gates for safer immunotherapy of cancer. *Front. Immunol.* **12**, 780399 (2021).
67. I. Tirosh, B. Izar, S. M. Prakadan, M. H. Wadsworth II, D. Treacy, J. J. Trombetta, A. Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J.-R. Lin, O. Cohen, P. Shah, D. Lu, A. S. Genshaft, T. K. Hughes, C. G. K. Ziegler, S. W. Kazer, A. Gaillard, K. E. Kolb, A.-C. Villani, C. M. Johannessen, A. Y. Andreev, E. M. Van Allen, M. Bertagnolli, P. K. Sorger, R. J. Sullivan, K. T. Flaherty, D. T. Frederick, J. Jané-Valbuena, C. H. Yoon, O. Rozenblatt-Rosen, A. K. Shalek, A. Regev, L. A. Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
68. J. L. Endicott, P. A. Nolte, H. Shen, P. W. Laird, Cell division drives DNA methylation loss in late-replicating domains in primary human cells. *Nat. Commun.* **13**, 6659 (2022).
69. A. Krogh, B. Larsson, G. von Heijne, E. L. Sonnhammer, Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
70. U. Omasits, C. H. Ahrens, S. Müller, B. Wollscheid, Protter: Interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics* **30**, 884–886 (2014).
71. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
72. C.-Y. Chang, I.-C. Cheng, Y.-C. Chang, P.-S. Tsai, S.-Y. Lai, Y.-L. Huang, C.-R. Jeng, V. F. Pang, H.-W. Chang, Identification of neutralizing monoclonal antibodies targeting novel conformational epitopes of the porcine epidemic diarrhoea virus spike protein. *Sci. Rep.* **9**, 2529 (2019).
73. A. M. Dreyer, J. Beauchamp, H. Matile, G. Pluschke, An efficient system to generate monoclonal antibodies against membrane-associated proteins by immunisation with antigen-expressing mammalian cells. *BMC Biotechnol.* **10**, 87 (2010).
74. R. Adamczak, A. Porollo, J. Meller, Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins* **59**, 467–475 (2005).
75. T. N. Schumacher, R. D. Schreiber, Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
76. E. A. Clayton, L. Rishishwar, T.-C. Huang, S. Gulati, D. Ban, J. F. McDonald, I. K. Jordan, An atlas of transposable element-derived alternative splicing in cancer. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190342 (2020).
77. N. M. Shah, H. J. Jang, Y. Liang, J. H. Maeng, S.-C. Tzeng, A. Wu, N. L. Basri, X. Qu, C. Fan, A. Li, B. Katz, D. Li, X. Xing, B. S. Evans, T. Wang, Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. *Nat. Genet.* **55**, 631–639 (2023).
78. M. Burbage, A. Rocañín-Arjó, B. Baudon, Y. A. Arribas, A. Merlotti, D. C. Rookhuizen, S. Heurtebise-Chrétien, M. Ye, A. Houy, N. Burgdorf, G. Suarez, M. Gros, B. Sadacca, M. Carrascal, A. Garmilla, M. Bohec, S. Baulande, B. Lombard, D. Loew, J. J. Waterfall, M.-H. Stern, C. Goudot, S. Amigorena, Epigenetically controlled tumor antigens derived from splice junctions between exons and transposable elements. *Sci. Immunol.* **8**, eabm6360 (2023).
79. A. Merlotti, B. Sadacca, Y. A. Arribas, M. Ngoma, M. Burbage, C. Goudot, A. Houy, A. Rocañín-Arjó, A. Lalanne, A. Seguin-Givelet, M. Lefevre, S. Heurtebise-Chrétien, B. Baudon, G. Oliveira, D. Loew, M. Carrascal, C. J. Wu, O. Lantz, M.-H. Stern, N. Girard, J. J. Waterfall, S. Amigorena, Noncanonical splicing junctions between exons and transposable elements represent a source of immunogenic recurrent neo-antigens in patients with lung cancer. *Sci. Immunol.* **8**, eabm6359 (2023).
80. C. M. Laumont, K. Vincent, L. Hesnard, É. Audemard, É. Bonneil, J.-P. Laverdure, P. Gendron, M. Courcelles, M.-P. Hardy, C. Côté, C. Durette, C. St-Pierre, M. Benhammadi, J. Lanoix, S. Vobecky, E. Haddad, S. Lemieux, P. Thibault, C. Perreault, Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* **10**, eaau5516 (2018).
81. K. W. Ng, J. Boumelha, K. S. S. Enfield, J. Almagro, H. Cha, O. Pich, T. Karasaki, D. A. Moore, R. Salgado, M. Sivakumar, G. Young, M. Molina-Arcas, S. de Carné Trécesson, P. Anastasiou, A. Fendler, L. Au, S. T. C. Shepherd, C. Martínez-Ruiz, C. Puttick, J. R. M. Black, T. B. K. Watkins, H. Kim, S. Shim, N. Faulkner, J. Attig, S. Veeriah, N. Magno, S. Ward, A. M. Frankell, M. Al Bakir, E. L. Lim, M. S. Hill, G. A. Wilson, D. E. Cook, N. J. Birkbak, A. Behrens, N. Yousaf, S. Popat, A. Hackshaw, C. T. Hiley, K. Litchfield, N. McGranahan, M. Jamal-Hanjani, J. Larkin, S.-H. Lee, S. Turajlic, C. Swanton, J. Downward, G. Kassiotis, Antibodies against endogenous retroviruses promote lung cancer immunotherapy. *Nature* **616**, 563–573 (2023).
82. G. Almogy, M. Pratt, F. Oberstrass, L. Lee, D. Mazur, N. Beckett, O. Barad, I. Soifer, E. Perelman, Y. Etzioni, M. Sosa, A. Jung, T. Clark, G. Lithwick-Yanai, S. Pollock, G. Hornung, M. Levy, M. Coole, T. Howd, M. Shand, Y. Farjoun, J. Emery, G. Hall, S. Lee, T. Sato, R. Magner, S. Low, A. Bernier, B. Gandi, J. Stohlman, C. Nolet, S. Donovan, B. Blumenstiel, M. Cipicchio, S. Dodge, E. Banks, N. Lennon, S. Gabriel, D. Lipson, Cost-efficient whole genome-sequencing using novel mostly natural sequencing-by-synthesis chemistry and open fluidics. bioRxiv. 2022.05.29.493900 (2022).
83. D. Hoyos, R. Zappasodi, I. Schulze, Z. Sethna, K. C. de Andrade, D. F. Bajorin, C. Bandlamudi, M. K. Callahan, S. A. Funt, S. R. Hadrup, J. S. Holm, J. E. Rosenberg, S. P. Shah, I. Vázquez-García, B. Weigelt, M. Wu, D. Zamarin, L. F. Campitelli, E. J. Osborne, M. Klinger, H. S. Robins, P. P. Khincha, S. A. Savage, V. P. Balachandran, J. D. Wolchok, M. D. Hellmann, T. Merghoub, A. J. Levine, M. Łuksza, B. D. Greenbaum, Fundamental immune-oncogenicity trade-offs define driver mutation fitness. *Nature* **606**, 172–179 (2022).
84. S. E. Brightman, M. S. Naradikian, A. M. Miller, S. P. Schoenberger, Harnessing neoantigen specific CD4 T cells for cancer immunotherapy. *J. Leukoc. Biol.* **107**, 625–633 (2020).
85. J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, M. Mann, Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
86. A. W. Purcell, Is the immunopeptidome getting darker?: A commentary on the discussion around Mishto et al., 2019. *Front. Immunol.* **12**, 720811 (2021).
87. T. Ouspenskaia, T. Law, K. R. Clauser, S. Klaeger, S. Sarkizova, F. Aguet, B. Li, E. Christian, B. A. Knisbacher, P. M. Le, C. R. Hartigan, H. Keshishian, A. Apffel, G. Oliveira, W. Zhang, S. Chen, Y. T. Chow, Z. Ji, I. Jungreis, S. A. Shukla, S. Justesen, P. Bachireddy, M. Kellis, G. Getz, N. Hacohen, D. B. Keskin, S. A. Carr, C. J. Wu, A. Regev, Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* **40**, 209–217 (2022).
88. G. Stuber, G. H. Leder, W. T. Storkus, M. T. Lotze, S. Modrow, L. Székely, H. Wolf, E. Klein, K. Kärre, G. Klein, Identification of wild-type and mutant p53 peptides binding to HLA-A2 assessed by a peptide loading-deficient cell line assay and a novel major histocompatibility complex class I peptide binding assay. *Eur. J. Immunol.* **24**, 765–768 (1994).
89. M. Grommé, J. Neefjes, Antigen degradation or presentation by MHC class I molecules via classical and non-classical pathways. *Mol. Immunol.* **39**, 181–202 (2002).
90. Z. Zhang, C. Zhou, L. Tang, Y. Gong, Z. Wei, G. Zhang, F. Wang, Q. Liu, J. Yu, *ASNEO*: Identification of personalized alternative splicing based neoantigens with RNA-seq. *Aging* **12**, 14633–14648 (2020).
91. A. Szolek, B. Schubert, C. Mohr, M. Sturm, M. Feldhahn, O. Kohlbacher, OptiType: Precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
92. R. Orenbuch, I. Filip, D. Comito, J. Shaman, I. Pe'er, R. Rabadan, arcasHLA: High-resolution HLA typing from RNAseq. *Bioinformatics* **36**, 33–40 (2020).
93. M. L. Buchkovich, C. C. Brown, K. Robasky, S. Chai, S. Westfall, B. G. Vincent, E. T. Weimer, J. G. Powers, HLAProfiler utilizes k-mer profiles to improve HLA calling accuracy for rare and common alleles in RNA-seq data. *Genome Med.* **9**, 86 (2017).
94. M. J. Goldman, B. Craft, M. Hastie, K. Repečka, F. McDade, A. Kamath, A. Banerjee, Y. Luo, D. Rogers, A. N. Brooks, J. Zhu, D. Haussler, Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).
95. D. Emig, N. Salomonis, J. Baumbach, T. Lengauer, B. R. Conklin, M. Albrecht, AltAnalyze and DomainGraph: Analyzing and visualizing exon expression data. *Nucleic Acids Res.* **38**, W755–W762 (2010).

Downloaded from https://www.science.org on July 31, 2025

96. Y. Dong, X. Liu, B. Jiang, S. Wei, B. Xiang, R. Liao, Q. Wang, X. He, A genome-wide investigation of effects of aberrant DNA methylation on the usage of alternative promoters in hepatocellular carcinoma. *Front. Oncol.* **11**, 780266 (2021).
97. D. K. Wells, M. M. van Buuren, K. K. Dang, V. M. Hubbard-Lucey, K. C. F. Sheehan, K. M. Campbell, A. Lamb, J. P. Ward, J. Sidney, A. B. Blazquez, A. J. Rech, J. M. Zaretsky, B. Comin-Anduix, A. H. C. Ng, W. Chour, T. V. Yu, H. Rizvi, J. M. Chen, P. Manning, G. M. Steiner, X. C. Doan, T. Merghoub, J. Guinney, A. Kolom, C. Selinsky, A. Ribas, N. H. Hellmann, A. Sette, J. R. Heath, N. Bhardwaj, F. Ramsdell, R. D. Schreiber, T. N. Schumacher, P. Kvistborg, N. A. Defranoux, Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* **183**, 818–834.e13 (2020).
98. S. Paul, N. P. Croft, A. W. Purcell, D. C. Tscharke, A. Sette, M. Nielsen, B. Peters, Benchmarking predictions of MHC class I restricted T cell epitopes in a comprehensively studied model system. *PLOS Comput. Biol.* **16**, e1007757 (2020).
99. G. Nibeyro, R. Girotti, L. Prato, G. Moron, H. D. Luján, E. A. Fernandez, MHC-I binding affinity derived metrics fail to predict tumor specific neoantigen immunogenicity. bioRxiv 2022.03.14.484285 [Preprint] (2022). https://doi.org/10.1101/2022.03.14.484285.
100. N. Salomonis, B. Nelson, K. Vranizan, A. R. Pico, K. Hanspers, A. Kuchinsky, L. Ta, M. Mercola, B. R. Conklin, Alternative splicing in the differentiation of human embryonic stem cells into cardiac precursors. *PLOS Comput. Biol.* **5**, e1000553 (2009).
101. N. Salomonis, C. R. Schlieve, L. Pereira, C. Wahlquist, A. Colas, A. C. Zambon, K. Vranizan, M. J. Spindler, A. R. Pico, M. S. Cline, T. A. Clark, A. Williams, J. E. Blume, E. Samal, M. Mercola, B. J. Merrill, B. R. Conklin, Alternative splicing regulates mouse embryonic stem cell pluripotency and differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10514–10519 (2010).
102. S. K. Behura, D. W. Severson, Codon usage bias: Causative factors, quantification methods and genome-wide patterns: With emphasis on insect genomes. *Biol. Rev. Camb. Philos. Soc.* **88**, 49–61 (2013).
103. M. J. Bowman, J. A. Pulman, T. L. Liu, K. L. Childs, A modified GC-specific MAKER gene annotation method reveals improved and novel gene predictions of high and low GC content in Oryza *sativa*. *Bioinformatics* **18**, 522 (2017).
104. S. Brogna, J. Wen, Nonsense-mediated mRNA decay (NMD) mechanisms. *Nat. Struct. Mol. Biol.* **16**, 107–113 (2009).
105. C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, L. Pachter, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
106. A. C. Frazee, A. E. Jaffe, B. Langmead, J. T. Leek, Polyester: Simulating RNA-seq datasets with differential transcript expression. *Bioinformatics* **31**, 2778–2784 (2015).
107. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
108. E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. Szerlip, P. Horsfall, N. D. Goodman, Pyro: Deep universal probabilistic programming. arXiv:1810.09538 [cs.LG] (2018).
109. B. Slaff, C. M. Radens, P. Jewell, A. Jha, N. F. Lahens, G. R. Grant, A. Thomas-Tikhonenko, K. W. Lynch, Y. Barash, MOCCASIN: A method for correcting for known and unknown confounders in RNA splicing analysis. *Nat. Commun.* **12**, 3353 (2021).
110. J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shoresh, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X.-O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
111. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
112. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.-C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
113. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
114. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: Gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 7 (2013).
115. M. J. Alvarez, Y. Shen, F. M. Giorgi, A. Lachmann, B. Belinda Ding, B. Hilda Ye, A. Califano, Functional characterization of somatic mutations in cancer using Network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
116. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
117. Y. Chen, S. W. Kwon, S. C. Kim, Y. Zhao, Integrated approach for manual evaluation of peptides identified by searching protein sequence databases with tandem mass spectra. *J. Proteome Res.* **4**, 998–1005 (2005).
118. S. Kawashima, M. Kanehisa, AAindex: Amino acid index database. *Nucleic Acids Res.* **28**, 374 (2000).
119. G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, WebLogo: A sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

Acknowledgments: We thank D. J. Schnell for his valuable feedback on statistical analysis recommendations and J. M. Kofron and the Confocal Imaging Core for assistance with the imaging analyses. **Funding:** This work was partially supported by the Cincinnati Children's Hospital Research Foundation, the University of Cincinnati Cancer Center, the Melanoma Research Foundation (Career Development Award to G.M.S.), an American Cancer Society Institutional Research Grant to G.M.S., and the National Institutes of Health (R01CA226802 to N.S., RC2DK122376 to H.L.G. and N.S., and R01HG013328 to M.T.W.) **Author contributions:** G.L. and N.S. conceptualized and developed SNAF, BayesTS, RNA-SPRINT, and associated new workflows. G.L., N.S., A.B., and M.V. performed bioinformatics analyses. M.T.W. and E.R.M. advised on the design and evaluation of the RNA binding regulatory analysis workflow (RNA-SPRINT). S. Mahajan, E.D.J., S. Ma, X.Z., T.T., H.L.G., and G.M.S. conceptualized the experimental design. E.D.J. and G.M.S. generated the long-read and spike-in mass spectrometry experimental data. S. Mahajan and T.T. performed peptide-MHC stabilization and peptide immunogenicity experimental analyses. S. Ma and X.Z. performed ExNeoEpitope cDNA cloning and transmembrane localization experiments. G.L., S. Mahajan, G.M.S., T.T., and N.S. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The SNAF application is available as a Python3 package (https://pypi.org/project/SNAF/). The source code is available at (https://github.com/frankligy/SNAF) and has been deposited to Zenodo (https://zenodo.org/records/10252900). The scripts and data for reproducing the results are available at https://github.com/frankligy/SNAF/tree/main/reproduce along with the raw and processed SNAF results (https://synapse.org/#!Synapse:syn32057176/files/). The raw peptide spectrum match (PSM) reported by MaxQuant using different search strategies is available at https://synapse.org/#!Synapse:syn53038566.

Submitted 5 August 2022
Accepted 13 December 2023
Published 17 January 2024
10.1126/scitranslmed.ade2886

Downloaded from https://www.science.org on July 31, 2025