Loss of SMARCA4 leads to Intron Retention and Generation of Tumor-Associated Antigens in Small Cell Carcinoma of the Ovary, Hypercalcemic Type

Authors

Elizabeth A. Raupach[1], Apurva M. Hegde[2,3], Krystine Garcia-Mansfield[2,3], Marice Alcantara[4], David L. Rose[5], Rebecca F. Halperin[1], Krystal A. Orlando[6,7], Jessica D. Lang[1], Ritin Sharma[2,3], Victoria David-Dirgo[2,3], Salvatore J. Facista[1], Rayvon Moore[6,7], Rochelle Kofman[2], Zoe N. Jensen[2], Victoria L. Zismann[1], Anthony N. Karnezis[8], Yemin Wang[9], Lynda B. Bennett[10], Timothy G. Whitsett[11], Marcin Kortylewski[4], William P. D. Hendricks[1], David G. Huntsman[9,12], Lorna Rodriguez-Rodriguez[13], Bernard E. Weissman[6,7], Jeffrey M. Trent[1], Patrick Pirrotte[2,3]

Affiliations

[1] Integrated Cancer Genomics, Translational Genomics Research Institute, Phoenix, Arizona, USA
[2] Cancer & Cell Biology Division, Translational Genomics Research Institute, Phoenix, Arizona, USA
[3] Integrated Mass Spectrometry Shared Resource, City of Hope Comprehensive Cancer Center, Duarte, California, USA
[4]Beckman Research Institute, Immuno-Oncology Department, City of Hope National Medical Center, Duarte, California, USA
[5]Department of Shared Resources, Beckman Research Institute of the City of Hope, Duarte, CA, 91010, USA
[6] Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[7] Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[8] Department of Pathology and Laboratory Medicine, UC Davis, Sacramento, CA USA
[9] Department of Pathology and Laboratory Medicine, University of British Columbia and Department of Molecular Oncology, British Columbia Cancer Research Institute, Vancouver, Canada.
[10]Baylor Scott & White Research Institute, Dallas, Texas, USA
[11]Translational Genomics Research Institute, Phoenix, Arizona, USA
[12]Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, Canada.
[13]Department of Surgery, Division of Gynecologic Oncology, City of Hope National Medical Center, Los Angeles, California, USA

Corresponding author

Patrick Pirrotte
Email: ppirrotte@tgen.org
Phone: +1-602-343-8454
Mail: 445 N. 5th Street, Phoenix, AZ 85004, USA

Running Title: SMARCA4 loss drives intron retention in SCCOHT

Key Words

splicing, intron retention, chromatin remodeling, tumor-associated antigen, tumor microenvironment

Conflicts of Interest

The authors declare they have no competing interest.

Word counts
Significance statement: 32
Abstract: 250
Body: 4,485
Key Words: 5
Figures 6
Table: 1
Supplementary Figures and Tables: Y

Significance Statement:
SCCOHT, a rare ovarian cancer, features splicing dysregulation due to SMARCA4 loss that generates immunostimulatory peptides linked to potential immune responses and therapeutic avenues, challenging traditional views of the role of SMARCA4.

Abstract

Small cell carcinoma of the ovary, hypercalcemic type (SCCOHT) is a rare, deadly form of ovarian cancer that uniformly harbors mutations in SMARCA4, a member of the SWI/SNF chromatin remodeling complex. SWI/SNF impacts RNA splicing, and dysregulation of splicing can generate immunogenic tumor antigens. Here, we explored the relationship between SMARCA4 loss and RNA splicing dysregulation. SCCOHT primary tumors harbored tumor-associated outlier splicing events compared to normal tissues. Many of the tumor events were retained introns encoding novel peptides predicted to bind to MHC-I complexes. Immune cells were observed in primary SCCOHT tumors, suggesting a potentially immune reactive tumor microenvironment. Mutations in several SWI/SNF subunits were associated with higher rates of outlier retained introns across tumor types in TCGA data. Interestingly, RNA sequencing of isogenic SCCOHT cell lines demonstrated a role for SMARCA4 in intron retention. Distinct protein-protein interactions between splicing factors identified in SCCOHT cell lines supported a role for SMARCA4 in splicing regulation. Further, SWI/SNF localized to genes which were differentially spliced. Mass spectrometry analyses confirmed expression of some of these novel peptides and a subset of these are predicted to bind to MHC-I complexes. A pool of these novel peptides derived from retained introns in SCCOHT triggered proliferation and expression of TNFα and INFγ in primary human T cells. Together, these data suggest that SMARCA4 loss in SCCOHT leads to intron retention. Furthermore, T cell activation by novel peptides encoded by these tumor-specific splicing events suggests intron retention could be a source of tumor-associated antigens in SCCOHT.

Introduction

Mutations in members of the switch/sucrose non-fermenting (SWI/SNF) chromatin remodeling complex are observed in 20% of all cancer types (1,2). Mutations often result in loss-of-function of one of its 10-15 SWI/SNF subunits, which act as tumor suppressors (3,4). In contrast to adult cancers where the accumulation of mutations often promotes tumor development, loss of a single SWI/SNF subunit is sufficient to drive aggressive pediatric and young adult cancers such as rhabdoid tumors, epithelial sarcoma, and small cell carcinoma of the ovary, hypercalcemic type (SCCOHT) (3-8). Loss of a single SWI/SNF subunit is also associated with tumor progression and poor patient outcomes in lung cancer, clear cell renal cell carcinoma, breast cancer, cervical cancer, colorectal cancer, and pancreatic cancer (9-14). Therefore, an unmet clinical need exists for treatments that target tumors driven by mutations in the SWI/SNF complex.

Direct targeting of the SWI/SNF complex has been hampered by compensation between paralogous subunits in the complex and the lack of available drugs. In some cases, synthetic lethality can be achieved by genetic inhibition of a subunit paralog such as ARID1B in ARID1A-deficient ovarian clear cell carcinomas (OCCC) (15) or SMARCA2 in SMARCA4-deficient lung adenocarcinomas (16). However, drugs targeting specific subunits of the SWI/SNF complex have had limited success. Strategies targeting either SMARCA2/4 bromodomain or ATPase activity are currently under clinical investigation, though concerns of efficacy and toxicity have been raised (17). Alternatively, targeting critical pathways in SWI/SNF-dependent cancers is an active area of exploration, and include inhibition of the polycomb complex 2 (PRC2) (18-20), receptor tyrosine kinases (21), CDK4/6 (22,23), LSD1 (24), PI3K signaling (25), oxidative phosphorylation (26,27) and DNA repair (28,29). Unfortunately, resistance to some of these targeted therapies has been reported (30).

Response to ICIs is correlated with increased tumor mutation burden (TMB) (31-33), leading to the approval of ICIs in high TMB tumor types. The increased TMB is hypothesized to provide increased tumor-specific antigens for immune cell recognition, sensitizing them to ICIs (31,32,34). However, recent reports have shown promising results deploying immune therapies for the treatment of SWI/SNF-deficient cancers and tumors with quiet genomes (35,36). *SMARCA4* mutation in lung cancer was associated with response to immune checkpoint inhibitors (ICI) (37,38). PBRM1 loss was associated with ICI response in clear cell renal cell carcinoma (39). While studies have shown correlation between SWI/SNF subunit and ICI response, others studies have reported no correlation, suggesting context dependency (40). Thus, understanding the mechanisms that govern how SWI/SNF subunit loss sensitizes tumors to ICI treatment is necessary for patient stratification.

Proposed sources of antigens not encoded by DNA mutations include RNA splicing dysregulation, ribosome slipping during protein translation, protein phosphorylation changes, and peptide splicing by the proteasome (41). Splicing is often dysregulated in cancers (42,43), and is capable of producing putatively antigenic protein sequences (44,45). The SWI/SNF complex regulates transcriptional control and recent evidence suggests a role in splicing regulation. SMARCA2 regulates inclusion or exclusion of variant exons by influencing the state of RNA polymerase II (RNAPII), which may affect transcription elongation rate and co-transcriptional recruitment of splicing factors to pre-mRNAs (46). Further, SWI/SNF associates with many nascent RNAs across the genome (47) and impacts splicing factor association with nascent RNAs (48). Additional evidence from Drosophila indicates that SWI/SNF associates directly with pre-mRNP particles (49) and transcription termination factors (50), suggesting direct roles in splicing and 3' end processing.

Given SWI/SNF involvement in splicing regulation and the potential for splicing dysregulation to generate tumor antigens, we investigated whether splicing dysregulation may provide a source of SCCOHT antigens recognized by immune cells. SCCOHT tumors provide a uniquely advantageous model as they lack expression of both paralogous SWI/SNF ATPase subunits, SMARCA4 and SMARCA2 (5,7,8,51), and harbor remarkably stable genomes with no other recurring mutations identified (52). Using a statistically robust splicing analysis method we developed called Bisbee (53), we identified and characterized outlier splice events in primary SCCOHT tumors encoding novel protein sequences, including a subset predicted to bind to MHC complexes. Supporting a role for SMARCA4 in splicing regulation, we also observed intron retention in isogenic SCCOHT cell lines with and without loss of SMARCA4. Further, we determined that SMARCA4 localized to differentially-spliced genes and interacted with splicing factors, pointing toward a role in splicing regulation. Finally, we found that novel peptides derived from intron retention (IR) events observed in SCCOHT cells could elicit a robust T cell response *in vitro*. Based on these findings, we propose that SMARCA4 loss in SCCOHT leads to intron retention. Because these retained introns are predicted to bind to MHC-I complexes and elicited T cell proliferation in vitro, we hypothesize that IR a potential source of tumor associated-antigens (TAAs) in SCCOHT. These splicing-derived putative TAAs could contribute to therapeutic response to immunotherapies. This finding is of great importance for SCCOHT patients as treatment options are limited and many cases are refractory to available treatments.

Materials and Methods

Cell Culture

BIN67 (BIN-67), SCCOHT1 (SCCOHT-1), and COV434 are SCCOHT cell lines bearing mutations in SMARCA4 and lacking SMARCA2 expression (54-57). SCCOHT1 and BIN67 cells were maintained in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% Fetal Bovine Serum (FBS) and 1% Penicillin/Streptomycin (PS). COV434 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% FBS and 1% PS. HAP1 is a haploid chronic myelogenous leukemia cell line derived from KBM-7 bearing a wild-type SWI/SNF complex (Sequence Read Archive accession SRP044390) (58). HAP1 cells were maintained in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% FBS and 1% PS. All cell lines were cultured at 37°C in a humidified incubator with 5% CO_2. Cell lines were Short Tandem Repeat (STR) profiled for verification and monitored routinely for mycoplasma contamination. For experiments described below, cells were harvested by scraping into cold 1X Phosphate Buffered Saline (PBS), followed by centrifugation, flash freezing, and storage at -80°C. The SCCOHT1, BIN67, and COV434 cell lines were kind gifts from Ralf Hass, Barbara Vanderhyden, and David Huntsman respectively. All cell lines were screened for mycoplasma and STR profiled upon receiving in the lab. Cell lines were routinely screened for mycoplasma using PCR testing monthly and were found to be free of mycoplasma.

RNA sequencing

RNA sequencing (RNA-seq) from BIN67 cells transiently transfected with wild-type SMARCA4 was described in Orlando *et al*. ((59); GEO accession GSE151025), which were sequenced to an average depth of 182.44 +/- 17.29 (standard deviation) million aligned reads. COV434 pIND20 BRG1-2.7 and parental COV434 cell lines were cultured using tetracycline-free FBS and induced with 450 nM doxycycline to re-express SMARCA4 as previously described (21). After 4 days of induction, cells were harvested and RNA was extracted using the RNeasy Mini Kit (Qiagen, Maryland, MD). Three independent replicates were performed for each experimental group. Total RNA (0.25-1 µg) was prepared using the TruSeq mRNA v2 kit (Illumina, San Diego, CA) to produce unstranded RNA libraries following the manufacturer's

protocol. RNA libraries were quantified and quality was evaluated using the Qubit DNA BR Reagent kit (ThermoFisher Scientific, Waltham, MA) and Tapestation D1000 tapes (Agilent, Santa Clara, CA). Libraries were pooled by equimolar ratios and sequenced on a NovaSeq6000 (Illumina) using paired-end sequencing (151x9x151). Average sequencing depth was 250.81 +/- 48.18 (standard deviation) million aligned reads.

SCCOHT tumor RNA-seq was performed on a total of 10 tumors. Four tumors were previously described in Lang *et al.* ((21); dbGap accession phs001528.v1.p1, GEO accession GSE109919), which were sequenced to an average depth of 198.28 +/- 64.12 (standard deviation) million aligned reads. Six additional tumors were collected under IRB protocols: Western IRB protocol #1119451 at TGen (Samples TGEN-E through TGEN-I) and protocol #90-0573 at the University of North Carolina-Chapel Hill (Samples TGEN-J). For the additional 6 tumors, RNA was harvested from formalin-fixed, paraffin-embedded (FFPE) SCCOHT tumors using the RNeasy FFPE Kit (Qiagen). Total RNA (200 ng) was heat fragmented on a GeneAmp PCR System 9700 (Applied Biosystems, Waltham, MA) to a target peak of 150 base pairs and libraries were generated using the SureSelect XT RNA (Agilent) capture following the manufacturer's protocol. RNA libraries were quantified and quality was evaluated using the Qubit DNA BR Reagent kit and Tapestation D1000 tapes. Libraries were pooled by equimolar ratios and sequenced on a HiSeq4000 (Illumina) using paired-end sequencing (77x9x77). Average sequencing depth was 77.05 +/- 26.44 (standard deviation) million aligned reads.

RNA-seq analysis

All RNA-seq was run through TGen's Pegasus pipeline. BCL to FASTQ conversion was performed using Illumina's BCL Converter tool. Alignment of FASTQ files against the GRCh37 human reference genome was performed using TopHat (60,61). Cufflinks/Cuffdiff was used to identify differentially expressed transcripts (61). Gene annotations were performed using Ensembl version 74.

Differential and outlier splice event analysis

SplAdder (v2.3.0) was run on RNAseq bam files aligned to the human reference genome (GRCh37) from 2 cell lines and 10 tumors, using the following run parameters: validation threshold of 2, no extraction, and single file alignment for parallelization. The resulting .counts.hdf5 files were used as inputs into the Bisbee alternative splicing tool ((59), https://github.com/tgen/bisbee) in order to: 1) calculate differentially spliced events between control and SMARCA4 re-expressed cell line models, and 2) identify outlier events in tumors vs. whole body reference tissues. Differential analysis through Bisbee, set the SplAdder version to 2 with a max W of 200 (default). Bisbee outlier analysis used the .counts.hdf5 files of GTEx samples published by Kahles et al. (49) as reference comparison, and was run with the reference SplAdder version set to 1, the test SplAdder version set to 2, and a maximum Beta of 80. Absolute log likelihood ratios (LLR) of 8 and 10 were used as thresholds in the differential splicing and outlier analysis, respectively. The above parameters were chosen based on findings by Halperin, et al (59). Protein sequences were predicted by Bisbee using the hs37d5 human reference genome FASTA containing additional ribosomal contigs and cancer-related viruses used in previous studies (https://github.com/tgen/MMRF_CoMMpass/tree/master/fasta_and_gtf_creation), with a padding of 9 amino acids.

Neoantigen prediction

Human leukocyte antigen (HLA) calls for HLA-A, HLA-B, and HLA-C were determined using seq2HLA 2.2 (62) and phlat (63). Analysis of HLA expression was performed using seq2HLA 2.2. 8 to 11-mer variant peptide sequences from the topmost alternatively spliced transcript as

output by the Bisbee pipeline for each differential/outlier splice event (53), and corresponding wild type peptides were used to predict major histocompatibility complex (MHC) class I binding affinity to the HLA genotypes using NetMHCpan 4.0 (64). Patient harmonic mean best rank (PHBR) scores were calculated to collapse predictions to a single value per splice event per sample (65). Expression of predicted neoantigens was determined using the detection of the splice events in the RNA-sequencing data. Predicted strong-binding neoantigens were defined as a PHBR score <0.5, predicted weak binders as 0.5<PHBR<2 and predicted non-binders were defined as PHBR>2.

Gene Ontology (GO) Analysis

Enriched biological processes among SWI/SNF-interacting proteins in IP-MS data were determined using ClueGO (v2.5.3; (66)) against the Homo sapiens database for GO biological processes (EBI-UniProt-GOA_04.09.2018_00h00) using experimental evidence only (67). GO term fusion was applied and only terms with a p-value ≤0.05, based on a 2-sided hypergeometric test with Benjamini-Hochberg correction, were reported. Kappa score (>0.4) based GO term grouping was utilized to define overarching functional groups across the data. All other parameters were left at default ClueGO settings. Gene ontology enrichment analysis of high confidence interactors from HAP1, BIN67, SCCOHT1, and COV434 was performed using ClueGO and CluePedia (v11.0), and the percent of genes identified were computed from each cell line per term.

Genes expressing outlier splice events shared in 6 or more SCCOHT tumors were analyzed for enriched GO biological processes (EBI-UniProt-GOA-ACAP-ARAP_08.05.2020_00h00) using all codes of evidence in ClueGO v2.5.7. Functionally grouped networks were analyzed before and after GO term fusion as mentioned above and visualized using Cytoscape v3.8.0 (68).

Immunoprecipitation mass spectrometry

Cytoplasmic extracts of BIN67, COV434, SCCOHT1, and HAP1 cells were isolated after thawing cells on ice by incubation in 400 µL Buffer A (20 mM HEPES pH 7.9, 10 mM KCl, 0.2 mM EDTA pH 8, 1X Halt™ protease inhibitor cocktail) on ice for 10 minutes. After addition of 25 µL of 10% NP-40, extracts were vortexed, and nuclei were isolated by centrifugation at 21,000 RCF at 4°C for 30 seconds. Nuclei were resuspended in 400 µL Buffer C (20 mM HEPES pH 7.9, 0.4 M NaCl, 10 mM EDTA pH 8, 1 mM EGTA pH 8, 1X Halt™ protease inhibitor cocktail) and incubated at 4°C for 15 minutes using end-over-end rotation. The insoluble fraction was separated by centrifugation at 21,000 RCF at 4°C for 5 minutes. Soluble nuclear protein extracts were collected and protein concentration was measured using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific). An input of 500 µg of nuclear protein was used for each IP and the total volume was adjusted to 400 µL. Either protein A or G conjugated sepharose beads, used with rabbit and mouse antibodies, respectively, were washed 3 times in Buffer C, and 30 µL of a 50% slurry were added to appropriate IP samples. A titrated amount of each antibody used was added (Supp. Table 1) to a protein/bead mixture and incubated at 4°C overnight using end-over-end rotation. Sepharose beads were pelleted by centrifugation at 6,000 RCF at 4°C for 1 minute and IP samples were washed 5 times with 500 µL Buffer C. Proteins were eluted from dried beads by reconstituting in Laemmli loading buffer containing 2-Mercaptoethanol then incubating for 5 minutes at 95°C. Samples were separated by SDS-PAGE using 4-20% Tris-Glycine gels. Proteins were visualized by Coomassie staining, each sample was divided into 4 equal fractions, and each fraction was cut into 2 mm^3 cubes. Overnight in-gel trypsin digestion was performed following previously published methods (69).

Prior to analysis, peptides were desalted following a modified C18 StageTip protocol (70). Peptides were resuspended in 2% acetonitrile, 0.1% trifluoroacetic acid (TFA) and loaded onto

a self-packed C18 (3M, Maplewood, MN) StageTip. Peptides were washed with 2% acetonitrile, 0.1% TFA, eluted in 60% acetonitrile 0.1% TFA, dried in a vacuum concentrator, and stored at -80°C. The entirety of each sample was reconstituted in 2% acetonitrile, 0.1% formic acid containing 25 fmol of Pierce Peptide Retention Time Calibration (PRTC) mixture. Data were acquired on an Orbitrap Fusion Lumos (ThermoFisher Scientific) interfaced with an Ultimate 3000 UHPLC system (ThermoFisher Scientific) running binary solvent system A (Water, 0.1% formic acid) and B (Acetonitrile, 0.1% formic acid). IP-MS samples (5 µL injection volume) were loaded directly (18.3 minutes loading time) on an analytical column (PepMap RSLC C18, 75 µm ID x 15 cm, 3 µm particle size, 100 Å pore size) kept at 45°C and eluted at a flow rate of 300 nL/minute using the following 60 minutes method: 2% to 19% solvent B in 42 minutes, 19% to 45% B in 6 minutes, 45% to 90% B in 0.5 minute, plateau at 90% B for 1 minute, return to initial conditions in 0.5 minute and re-equilibration for 10 minutes. Data-dependent acquisition was performed in Top Speed mode with a 3-second duty cycle and the following parameters: spray voltage of 1900 V, ion transfer tube temperature of 275°C, survey scan in the Orbitrap at a resolution of 120K at 200 m/z, scan range of 400-1500 m/z, AGC target of 4E5 and maximum ion injection time of 50 milliseconds. Every parent scan was followed by a daughter scan using High Energy Collision (HCD) dissociation of top abundant peaks and detection in the iontrap with the following settings: quadrupole isolation mode enabled, isolation window at 1.6 m/z, AGC target of 5E3 with maximum ion injection time of 35 milliseconds and HCD collision energy of 35%. Dynamic exclusion was set to 60 seconds.

Immunoprecipitation and Protein Simple assay

Immunoprecipitation was performed from nuclear extracts as described above using SMARCC1 and SMARCB1 antibodies compared to a negative control IgG IP. Protein detetion was performed the Jess Simple Western™ platform (ProteinSimple™, Bio-Techne, Minneapolis, MN, USA). The Jess system employs capillary electrophoresis to separate proteins based on their size and charge, antibody probing, and chemiluminescent detection to quantify individual proteins in a sample. Proteins were eluted from the Sepharose A beads using the Protein Simple 5x fluorescent master mix (EZ standard pack 1; Protein Simple), first diluted to 1x with 0.1x sample buffer (ProteinSimple™, Bio-Techne), following the methods outlined in the Peggy Immunoprecipitation Protocol (Bio-Techne n.d.). The eluted proteins were then diluted with 5x fluorescent master mix (EZ standard pack 1; Protein Simple™) to total protein concentrations of 1.2 µg/µL for all SMARCC1 Ips, 1.5 µg/µL for all SNF5 Ips, 0.8 µg/µL for IgG Ips, and .08 µg/µL for nuclear extracts from each cell line. For each well, 3 µL of diluted protein sample was added. The samples were probed with an HNRNPC1/2 antibody (Santa Cruz Biotechnology, sc-32308) at a 1:25 dilution, diluting with antibody diluent 2 (anti-mouse detection module chemiluminescence; ProteinSimple™, Bio-Techne) and loading 10 µL into each well. The secondary antibody (anti-mouse HRP) and chemi-luminescence reagents were added according to the instructions of the kit manufacturer (anti-mouse detection module chemiluminescence; ProteinSimple™, Bio-Techne). The wash buffer, 10x sample buffer, plate, and capillary cartridge used for Jess were taken from the 12-230 kDa separation module (ProteinSimple™, Bio-Techne). Default 12-230 kDa chemiluminescence protocol settings were used for the Jess run.

Deep Expression Proteomics

BIN67 and COV434 cells were lysed in 8M Urea lysis buffer in 50 mM Tris-HCl, pH 8.0 containing 1X protease and phosphatase inhibitor cocktail (Pierce) and sonicated using a tip sonicator. Lysates were clarified by centrifugation at 16,000 g at 4 °C and protein concentration was determined using BCA protein assay. Clarified lysate (1 mg total protein amount) was reduced by dithiothreitol (10 mM final concentration) and alkylated by iodoacetamide (40 mm final concentration). Proteins were digested sequentially by two endoproteases, first with lysC

(Wako) for 4 hours at an enzyme: substrate ratio of 1:100 followed by overnight digestion with trypsin (Promega) at an enzyme: substrate ratio of 1:50. Digestion was stopped by the addition of trifluoroacetic acid (final concentration 1%) and peptides were desalted using C18 Sep-Pak columns (Waters). Tryptic peptides were offline fractionated by high pH Reverse Phase Chromatography into 96 one-minute fractions that were combined to yield the final 24 fractions for LC-MS/MS analysis. Peptides were loaded on Waters XBridge C18 column (4.6 mm ID * 250 mm, 3.5 µm particle size) and separated using a 96-minute gradient formed by Solvent A (Water), Solvent B (Acetonitrile), and Solvent C (50 mM Ammonium Hydroxide in water, pH 10). Deep expression proteomics of BIN67 and COV434 tryptic digests was acquired with minor differences highlighted. Data were acquired on an Orbitrap Eclipse mass spectrometer with FAIMSPro interface (ThermoFisher Scientific) (BIN67) or an Orbitrap Fusion Lumos (COV434), both coupled to Ultimate 3000 RSLCnano LC system (ThermoFisher Scientific). From each of the 24 offline fractions, one µg of peptides was directly loaded on either a 50 cm (BIN67) or a 25 cm (COV434) EasySpray C18 column (ThermoFisher Scientific). Peptides were separated using a 120-minute gradient formed by Solvent A (Water, 0.1% formic acid) and Solvent B (Acetonitrile, 0.1% formic acid) at a flow rate of 300 nL/min. For data acquisition on Eclipse, three FAIMS Compensation Voltages (CV) of -40, -60, and -80 with a cycle time of one second per CV were applied while the data on Lumos was acquired with a duty cycle of three seconds. Using TopSpeed mode, MS1 scans were performed in the Orbitrap (120,000 resolution at 200 m/z) and MS2 scans were performed in the ion trap following fragmentation of most abundant precursors from MS1 by Higher-Energy Collision Dissociation (Normalized Collision Energy 32%). Only precursors with a charge state between 2-7 were selected for MS2 and dynamic exclusion setting of either 45 seconds (BIN67) or 60 seconds (COV434) was applied to prevent resampling of the same precursors.

Proteogenomics analysis

Mass spectra were searched using Proteome Discoverer 2.4 (ThermoFisher Scientific, Waltham, MA) and Mascot (Matrix Science, London, UK; v2.6.0) against a concatenated FASTA database of amino acid sequences from differential splice events identified by Bisbee as “novel” supplemented with the GRCh37 Ensembl human proteome. Methionine oxidation and carbamidomethyl modifications were included as dynamic and static modifications, respectively. A maximum of two missed cleavages were allowed from full trypsin digestion, with fragment mass tolerance of 0.6 Da and precursor mass tolerance of 10 ppm. Results were filtered using Target Decoy PSM validator at a strict FDR threshold of 0.01. Identified novel splice peptides were further prioritized based on their putative MHC-I binding strength to be used in the T cell stimulation assays.

For immunoprecipitated samples, mass spectra were aligned by antibody within each cell line, followed by peak picking using Progenesis QI for proteomics v4.1.6 with default parameters for automated processing. Pre-processed spectra were then searched against a *Homo sapiens* database (UniprotKB/Swissprot, 2017) in Mascot v2.6.0. Oxidation (M) and acetylation (N-terminus) were specified as variable modifications and carbamidomethyl (C) was defined as a fixed modification. MS/MS tolerance was set to 0.6 Da. Tryptic peptides allowed for up to 2 missed cleavages, peptide charge states ranging between +2 and +4, and a peptide tolerance of 10 ppm. Matched spectra were imported into Progenesis QI for peptide assignment and filtering (Mascot score ≥21). Analyzed gel fractions were recombined and normalized in Progenesis QI. Normalized abundances of identified proteins were used to score *bona fide* protein-protein interactions (SAINTexpress) (71) that were enriched at least 2-fold above IgG controls. As per SAINTexpress recommendation, we used a threshold of an average probability of interaction (AvgP) of ≥0.7. Fold change for SWI/SNF complex members in the SCCOHT cell lines was calculated by first taking the mean normalized SAINTexpress abundance for each

antibody, then the median across antibodies for a single cell line, and dividing by the same value in HAP1.

ChIP-seq analysis

Publicly available ChIP-seq data were used to evaluate SWI/SNF occupancy to differentially spliced genes (GEO: GSE117735) (72). This analysis was performed from downloaded SRA files that were converted to FASTQs. Alignment to GRCh37 human reference genome was performed to generate BAM files using bowtie2 (v0.12.9; (73); arguments: -n 2 mismatches, -l 50 bases as seed, max phred quality -e 70, reportable alignment limit -k 1, suppress alignments with more than -m 1 reportable alignment) and samtools (v1.7) (74). QC was assessed using FastQC (v0.11.8), samtools (v1.9) and deepTools (v3.3.1) (75). BAM files for BRG1 control samples with replicate data were merged. Reads within the ENCODE blacklisted genomic ranges were discarded (https://github.com/Boyle-Lab/Blacklist/blob/master/lists/hg19-blacklist.v2.bed.gz, Encode Project Consortium) (76). Peaks were called using MACS2 (v2.1.2; (77)) (arguments: -f BAM -B -q 0.05 –bw 200). Peaks from the resulting narrowPeak files were annotated using ChipSeeker (v1.18.0) with default tssRegion value (78).

ATAC-seq analysis

Publicly available ATACseq data was used to evaluate chromatin accessibility at differentially spliced genes (GEO: GSE117301) (72). NGMerge (v0.3, (79)) was used to remove adaptor sequences from FASTQs files from both sequencing lanes. QC was assessed using FastQC (v0.11.8). Alignment to GRCh37 human reference genome was performed to generate sorted BAM files using bowtie2 (v2.3.5.1; (73); default bowtie arguments with the addition of --very-sensitive) and samtools (v1.7; (80)). Biological replicate BAM files were merged using samtools merge (v1.7) and mitochondrial reads were eliminated. Duplicate reads were removed using picard (v2.10.3) and only unique reads were retained using samtools view (arguments -b -q 10). Reads within the ENCODE blacklisted genomic ranges were discarded (https://github.com/Boyle-Lab/Blacklist/blob/master/lists/hg19-blacklist.v2.bed.gz, Encode Project Consortium) (76). CSAW (v1.24.3; (81,82)) was run on the resulting bam files with TMM normalization to identify differentially accessible regions using a 150 base-pair window size.

Immunohistochemistry

Immunohistochemistry (IHC) staining was performed on 0.6 mm slides cut from an FFPE fixed tissue microarray (TMA) containing primary tumor samples from 9 SCCOHT patients. IHC was performed on a Leica BOND autostaining instrument following the standard protocol. Presence of immune cells were evaluated using the following antibody clones: CD8, C8-144B 1:100; CD4, 4B12 1:80, CD20, L26 1:200; and CD68, PG-M1 1:50. Heated epitope retrieval was performed for 20 minutes in citrate buffer for CD20 and CD68 or in EDTA buffer for CD4 and CD8. Secondary antibodies were HRP conjugates that were visualized using DAB substrate. The slides were counter-stained with hematoxylin. Slides were imaged on a Leica Aperio CS2 instrument at 40X magnification. CD8 staining intensity was scored as follows: zero positive cells was called negative, 1-2 tumor infiltrating lymphocytes (TILs) was called low, 3-19 TILs was called moderate, and 20 or more TILs was called high (83). CD4 and CD20 staining intensity was scored as follows: zero positive cells was called negative, 1-5 TILs was called low, 6-19 TILs was called moderate, and 20 or more TILs was called high (84). CD68 staining intensity was scored as follows: zero positive cells was called negative, positive cells only located in the stroma was called low, <20 positive cells was called moderate, and 20 or more cells was called high (84).

Splicing-derived peptide pool generation

A total of 18 unique splice events combined from outlier tumor and differentially-spliced events in BIN67 were selected for testing in a T cell stimulation assay (Supp. Table 8A). Tumor outliers

were prioritized based on occurrence in 6 or more cases, higher expression in SCCOHT compared to normal tissues in GTEx, and predicted to bind to the 3 most common MHC-I alleles in the tumor data set (HLA-A*02:01, HLA-B*15:01, HLA-C*03:04). Differentially-spliced events observed in BIN67 were prioritized based on predicted MHC-I binding and filtered to only include events more highly expressed in -SMARCA4 cells. PepSequencer from JPT (Berlin, Germany) was used to generate tiled 15 amino acid peptides from the 18 prioritized peptides, with a 4 amino acid offset, which resulted in 189 total peptides. These peptides were split into two arbitrary SpotMix Plus pools (JPT), named PepMix 1 and PepMix 2, which were resuspended following the manufacturer's protocol.

T cell Stimulation Assays

T cells were enriched from peripheral blood mononuclear cells (PBMCs) obtained from healthy donor peripheral blood using the EasySep Human T cell Isolation Kit (Stemcell technologies catalog #17951). Stimulation assays were carried out in U-bottom 96-well plates with 40,000 enriched T cells per sample treated for 4 days with either PBS, DMSO, Pepmix 1, Pepmix 2, CEFX Ultra Superstim pool (JPT Peptide Technologies, catalog # PM-CEFX-1), or Dynabeads Human T cell activator CD3/CD28 for T cell activation (ThermoFisher Scientific, catalog #11131D) and PMA and Ionomycin with BFA for short term T cell activation (Sigma Aldrich catalog #P1585-1MG and 19657-1MG or BD catalog #555029 respectively). All dosing followed manufacturers' protocols. Flow cytometry was then performed on a BD LSR Fortessa 2 cytometer (BD Biosciences, San Jose, CA, USA) or the Cytek Aurora spectral cytometer (Cytek Biosciences, San Diego, CA) to assess CD3+ T cell proliferation using Cell Trace Violet (CTV) dilution (ThermoFisher, catalog #C34557) and T cell IFN-γ and TNF-α production following stimulation. Antibodies utilized were anti-CD3 (BD Biosciences, catalog #552127), anti-IFN-γ (eBiosciences catalog #E07382-1633), and anti-TNF-α (Biolegend, catalog #502909). Data were analyzed using FlowJo v10.7.1. A one-way ANOVA plus Bonferroni's multiple-comparison test was applied to assess differences between multiple groups. Data were analyzed with Prism v8.30 (GraphPad Software, La Jolla, CA, USA).

Data Availability Statement

RNA-seq from four SCCO tumors were previously published (dbGap accession phs001528.v1.p1, GEO accession GSE109919). RNA-seq from BIN67 is available under GEO accession GSE151025. RNA-seq from additional tumors and COV434 have been uploaded to dbGaP (accession phs001528.v2.p1) and GEO (GSE217699 and GSE216801). All IP-MS data have been uploaded to the MassIVE database hosted by UCSD (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) and are available via the identifier MSV000084780. Previously published ChIP-seq (GEO: GSE117734) and ATAC-seq (GEO: GSE117301) data were also used in this study. RNA-seq data from TCGA and GTex were also analyzed in this study (https://portal.gdc.cancer.gov/analysis_page?app=Downloads and https://www.gtexportal.org/home/downloads/:project). All other raw data generated in this study are available upon request from the corresponding author.

Results

SCCOHT primary tumors exhibit outlier splice events compared to normal tissues

To determine if splicing is dysregulated in SCCOHT, we performed Bisbee outlier splice analysis on 10 SCCOHT patient tumors and compared them to normal tissues in GTEx. Our patient cohort consists of young women (all under 40) with mostly stage II-III diagnoses. The majority of our cases have identified mutations in SMARCA4, while one lacks SMARCA4 expression and one has a mutation in another SWI/SNF subunit, SMARCB1 (Table 1). We previously developed Bisbee as a statistically robust splicing analysis method (53). SplAdder is

first used to detect and quantify annotated and *de novo* splice events based on the portion of RNAseq reads containing each splice isoform termed percent spliced in (PSI) (85). Bisbee performs beta-binomial distribution analysis of PSI values to determine whether a splice event is statistically differential between test comparison groups with replicate data (*in vitro* studies) or a statistical outlier in a given sample compared to a reference database (primary tumor analyses). The binomial model accounts for technical noise while the beta distribution models biological variance. Importantly, Bisbee robustly detected differences in splice isoform abundance across samples with differing levels of expression without the need to normalize to library size or expression level. An advantage of this statistical method is that it can be used for a single sample when compared to a reference database. Bisbee additionally predicts the protein sequences arising from identified splice alterations, enabling validation of splice alterations by detection of protein expression in mass spectrometry data. Bisbee analysis identified between 334-1125 (average 680) outlier splice events in each of 10 SCCOHT tumors (|LLR| >10) when compared to normal tissues in the GTEx database (Supp Figure 1A). These splicing events encode novel protein sequences not categorized as canonical in the Ensembl database (170-628, average 356 per tumor) (Figure 1A, Supp Table 2). Many outlier splice events (1349 total, Figure 1B) were observed in more than 1 tumor with 201 events observed in more than half of the tumors and 17 observed in all 10 tumors examined (Figure 1B, Supp Figure 1B). The majority (>60% in each tumor) of total outlier splice events and novel protein-coding outlier events, including those shared across multiple tumors, arose from intron retention (IR) (Figure 1C-D, Supp Figure 1C-D). Outlier splice events contain an average of 137 (range: 76-227) novel protein sequences predicted to strongly bind to MHC-I per tumor and an average of 91 (range: 31-167) novel protein sequences predicted to weakly bind to MHC1 per tumor (Supp Table 3). These intron retention events may be a rich source of tumor-associated antigens (TAAs) (86).

Recurrent outlier splice events included genes enriched in translational initiation, SRP-dependent co-translational protein targeting to membrane, cristae formation, spliceosomal complex assembly, and positive regulation of histone methylation (Supp Figure 1E, Supp Figure 2). Retained introns bear canonical 5' and 3' splice site sequences at similar frequencies to excised introns, (Supp Figure 3). No conserved sequence motifs were identified as shared among retained introns. The average length of retained introns tends to be shorter (mean 411 bp) than excised introns (mean 2026 bp) even when normalized to overall gene length ($p = 0$ total length, $p = 1.12E-09$ normalized length, Kolmogorov-Smirnov test) (Supp Figure 4A-D). Additionally, the retained introns tend to occur more proximal to the 5' end of the gene (mean 6th intron) compared to excised introns (mean 13th intron), even when normalized to the number of introns per gene ($p = 0$ intron position, $p = 1.24E-10$ normalized position, Kolmogorov-Smirnov test) (Supp Figure 4E-G). No recurrent mutations in known splicing regulators were identified that would explain our observation of intron retention in SCCOHT tumors (Supp Table 4).

SCCOHT tumors display immune infiltration

Because we postulate that the retained introns observed in SCCOHT tumors could be a putative source of TAAs, we evaluated whether the SCCOHT tumor microenvinronment contains infiltrating immune cells. Two previous reports demonstrated immune infiltration in other SCCOHT cases (35,36,87). We evaluated CD8, CD4 (T cell markers), CD20 (B cell marker), and CD68 (macrophage marker) staining in 9 SCCOHT tumors by immunohistochemistry staining. This includes 5 tumors for which splicing analyses were performed and 4 additional tumors. All 9 tumors were positive for CD8 staining with 3 tumors having a high count while the remaining 6 tumors had a moderate count (Figure 1E, Supp Table 5). All 9 tumors had a high level of macrophage infiltration demonstrated by high counts of CD68. CD4+ infiltrating T cells were observed at high levels in 2 tumors, moderate levels in 2 tumors, and low levels in 2

tumors. Infiltrating B cells were observed at high levels in one tumor, moderate levels in one tumor, and low levels in 3 tumors. Consistent with previous reports, we observed T cells, B cells, and macrophages in primary SCCOHT tumors with varying levels of B cells.

SWI/SNF mutated cancers have higher outlier IR rates in TCGA tumors

We expanded our analysis to look at splicing burden across 31 tumor types using TCGA data in comparison to GTex. While TCGA does not include SCCOHT cancers, we did observe that high-grade ovarian carcinoma had more outlier splice events than every other cancer, with a median of 25 outlier events per tumor (Supp Figure 5). We observed an average of 680 outlier events per SCCOHT tumor (Supp Figure 1A), which is substantially higher than any tumor type in TGCA. It should be noted that this stark difference might be attributed to the unique genetic context of SCCOHT which bears dual loss of both SMARCA4 and SMARCA2. Ovarian cancer showed higher alternate 3' (A3), alternate 5' (A5), and mutually exclusive exon (MUT) outlier splice events than any other cancer type and additionally had high incidence of IR and exon skipping (ES) outlier events. We also compared splice burden from intron retention events in tumors from TCGA with and without mutations in individual SWI/SNF components. Distribution of outlier IR events showed that mutations in some SWI/SNF members corresponded with higher rates of intron retention. Since the proportion of tumors in TCGA with wild-type SWI/SNF members was much greater than those with mutations, we performed permutation analysis, which revealed that ARID1B, BRD7, SMARCD2, and SMARCE1 had significantly higher number of IR events in >90% of trials and had significantly higher outlier IR events regardless of cancer type or sample size (Figure 2). While we did not observe a significant contribution of SMARCA4 mutation to splice burden compared to tumors with wild-type SMARCA4, it should be noted that the majority of these tumors likely still have a functional SMARCA2 and/or one wild-type copy of SMARCA4 which may be sufficient to carry out normal splicing regulation. SCCOHT is a unique model because it completely lacks both copies of SMARCA4 in a context where SMARCA2 is epigenetically silenced. Thus, comparing SCCOHT tumors to SMARCA4 mutated tumors in TCGA does not reflect the same genetic context for SWI/SNF functions and it may not be surprising that this comparison did not reveal a role for SMARCA4.

SMARCA4 deficiency leads to intron retention in SCCOHT cell lines

We hypothesized that SMARCA4 loss may lead to novel tumor antigens through splicing dysregulation. We analyzed previously published RNA sequencing data from isogenic SCCOHT cell lines where SMARCA4 had been exogenously re-expressed to near endogenous levels(21,59). Bisbee differential splice analysis of RNAseq data from the SCCOHT cell lines BIN67 and COV434 cells +/- exogenous SMARCA4 expression revealed SMARCA4-dependent differential splice events (|LLR|>8) in both isogenic cell lines (Figure 3A-B). Both cell lines showed increased IR when SMARCA4 was absent (Figure 3A-B, Supp Table 6A-B). IR was a predominant feature across primary tumors and cell lines, although specific IR events were not shared across sample types. IR events showed an inverse correlation between PSI values and RNA expression levels of that gene (Supp Figure 6), many of which were statistically significant from a Student's t-test ($p<0.05$). We observed 61 and 24 total differential splice events in BIN67 and COV434, respectively, upon SMARCA4 re-expression. Of those, 42 and 17 events were in protein-coding regions with 24 and 14 events which generated novel protein sequences. We observed 13 and 8 novel protein-coding events which were predicted to bind to MHC-I complexes, 6 and 3 of which were detected at low abundance in LC-MS/MS proteomics data providing evidence of their protein expression (Figure 3C, Supp Table 7). Transcript abundance tended to be lower when retained introns were observed (Figure 3D). These events provide candidate splicing-derived tumor-associated antigen sequences. Thus, SMARCA4 loss drives IR in SCCOHT cells, which may provide a source of splicing-derived antigens.

The SWI/SNF complex interacts with splicing factors and localizes to differentially-spliced genes

To explore the molecular consequences of SMARCA4 loss in SCCOHT, we performed immunoprecipitation and mass spectrometry (IP-MS) comparing three SCCOHT cell lines (BIN67, COV434, and SCCOHT1) to a cell line that harbors a wild-type SWI/SNF complex (HAP1). IP-MS experiments were performed using antibodies to five SWI/SNF subunits (SMARCA4, SMARCC1, SMARCE1, SMARCB1, and ARID2) in triplicate compared to IgG controls to allow statistical calculations for probabilities of interaction using SAINTexpress. These results allowed us to examine how SWI/SNF protein-protein interactions (PPIs) are changed in the context of ATPase subunit loss with high confidence. Although previous reports showed IP-MS evaluation of SWI/SNF complex integrity in the context of SMARCA4 loss (72,88), our data are the first report of a SWI/SNF protein-protein interactome in the context of SMARCA4 loss. Further, our data are the first description of a SWI/SNF protein-protein interactome using statistical probability thresholds for interaction. We observed 10,901 PPIs total prior to thresholding and 1,021 high-confidence interactors with a probability of interaction (AvgP) greater than 0.7, providing the most robust and detailed SWI/SNF interactome ever described across any model system (Supp Table 8, 9A-E). We combined the data from our 3 most reproducible IP antibodies to core SWI/SNF subunits (SMARCB1, SMARCC1, and SMARCE1) to determine a consensus model of protein-protein interactions. Consistent with a previous reports (72,88), we found that a number of ATPase module and PBAF-specific SWI/SNF subunits were exclusively maintained in the wild-type SWI/SNF including SMARCA4, SMARCA2, BCL7A/C, SS18, ACTL6A/B, DPF3, PHF10, BRD7, and PBRM1 (Figure 4A-B), while core SWI/SNF subunits including SMARCC1/2, SMARCD1/2/3, SMARCB1, and SMARCE1, BAF-specific SWI/SNF subunits ARID1A/B and DPF1/2, and ncBAF subunits BRD9 and BICRA/L remain intact across the three SCCOHT cell lines (Figure 4C, Supp Figure 7A-F). Additionally, interactions with KMT2C (MLL3) and NuRD complex subunits (GATAD2B and RBBP7) were only observed in the wild-type cell line (Figure 4A,B). The absence of PPIs with MML3 may partially explain the differences in chromatin localization of MLL3 and its dependent histone PTM, H3K4me1 in SCCOHT cells (72).

The residual SWI/SNF complex showed PPIs with RNA processing and metabolic proteins that were not shared with the wild-type complex (Figure 4D). Additionally, the residual SWI/SNF complexes showed more PPIs with RNA processing proteins than the wild-type complex (Figure 4E, Supp Figure 8A). We validated the interaction between SWI/SNF and the splicing factor HNRNPC1/2 using IP followed by a ProteinSimple assay. We observed HNRNPC1/2 bands in the SMARCC1 and SMARCB1 IP samples in COV434 and SCCOHT, but not in IgG IP samples or in the HAP1 cell line (Supp Figure 8B). Because IP-MS requires a high cell input, it was not practical to perform these experiments in SCCOHT cells where SMARCA4 expression was exogenously re-expressed. Thus, we validated these results using an isogenic comparison between HAP1 parental and HAP1 SMARCA4 knock-out lines. Similar to our observations with SCCOHT lines, a number of chromatin regulators and SMARCA4 were exclusive to HAP1 parental, the SWI/SNF core subunits were shared across the parental and SMARCA4 KO lines, and a number of RNA splicing and processing factors were exclusive the SMARCA4 KO line (Supp Figure 9A-D). Additionally, a substantial increase in splicing-associated SWI/SNF-interacting proteins were exclusively observed in the HAP1 SMARCA4 KO line. This includes many of the splice factors that were in SCCOHT cell lines and many additional splice factors that were not observed in SCCOHT cell lines (Supp Figure 9E). ChIPseq analysis showed that SMARCA4 localizes to nearly all (40 of 42) of the differentially-spliced genes in BIN67 cells upon re-expression (Figure 5A). SMARCA4 and other SWI/SNF subunit localization to the differentially-spliced genes and protein-protein interactions with splicing factors could support a role for SWI/SNF in splicing regulation. At the vast majority of differentially-spliced genes, no

significant changes in chromatin accessibility were observed in BIN67 cells. A small subset of differentially-spliced genes showed increased accessibility upon SMARCA4 re-expression. However, differential accessibility is not sufficient to explain the splicing differences observed. This may suggest that SWI/SNF may regulate splicing through non-catalytic functions (Figure 5B-C).

Splicing-derived peptides are sufficient to stimulate T cells

To determine if the novel protein-coding splice events we observed in SCCOHT tumors and cell lines are capable of eliciting an immune response, we measured whether the observed splicing-derived peptides stimulate T cells *in vitro*. Using outlier events shared in 6+ tumors combined with SMARCA4-dependent differential splice events observed in cell lines, we selected splice events encoding novel and tumor-associated protein sequences predicted to bind to MHC-I complexes *in silico*. Events detected in cell lines were additionally prioritized based on their protein expression in mass spectrometry data. This yielded a tiled peptide library of 189 peptides as candidate TAAs (Supp Table 10A-B). These high-confidence peptides were pooled into two arbitrary test groups. Naive T cells were exposed to vehicle controls, splicing-derived peptide pools, a positive control stimulatory peptide pool (SuperStim), or CD3/CD28 antibody-coated magnetic beads as a positive control. Peptide-induced CD3+ T cell proliferation was measured using flow cytometry following the dilution of a cell division tracking dye. Following a four-day exposure, we observed enhanced T cell expansion in response to one of the two tested splicing-derived peptide pools compared to negative controls in each of the donors (Figure 6A). Additionally, PepMix 1 and PepMix 2 resulted in a higher portion of CD3+ T cells which express TNFα or INFγ (Figure 6B). Increased proliferation and cytokine expression was observed with PepMix 1 compared to PepMix 2 pool, which suggests stronger immunogenic capacity in PepMix 1. This may indicate that PepMix 1 contains stronger antigens than PepMix 2. Taken together, these data support that SMARCA4-dependent, splicing-derived antigens have immune-stimulatory potential.

Discussion

Loss of the SWI/SNF complex member SMARCA4 is the key driver event in SCCOHT (5,7,8,89). We identified outlier splicing events in primary human SCCOHT tumors, the majority of which arise from intron retention (Figure 1). Given the rarity of SCCOHT (<500 cases in literature worldwide), our cohort of 10 primary tumors represents a substantial portion of available SCCOHT samples (89,90). Our results demonstrated a surprisingly high level of outlier splice events per tumor (average of 680 outliers/patient). For a comparison, the level of outlier splice events from a high mutation rate tumor (e.g. uveal melanoma tumors bearing mutations in the splicing factor SF3B1) demonstrated an average 180 outliers/patient (53). We additionally observed infiltration of immune cells into each of 9 primary SCCOHT tumors (Figure 1E). This finding supports a potentially “hot” immune reactive tumor microenvironment in SCCOHT. When compared across tumor types in TCGA, high grade serous ovarian carcinoma showed a higher rate of outlier splice events compared to other tumor types. This is consistent with previous reports (44). Additionally, we found an association between several SWI/SNF subunits and higher rates of IR across TCGA (Figure 2). The TCGA data did not yield an association between SMARAC4 mutation and IR. However, most SMARCA4 mutated tumors in TCGA retain a functional copy of SMARAC4 and/or SMARCA2 which may be sufficient to promote normal splicing. This surprising finding may point to an under-explored role of the SWI/SNF complex in splicing regulation. Further, using isogenic SCCOHT cell lines with and without re-expression of wild-type *SMARCA4*, we demonstrated a role for SMARCA4 in regulating outlier splice events observed in SCCOHT cells (Figure 3). SCCOHT cell lines displayed markedly fewer IR events than individual tumor samples. It may not be surprising to observe different frequencies between

primary tumors and cell lines, but we postulate that normal splicing may not have been completely restored by re-expression of exogenous SMARCA4. Nonetheless, a role for SWI/SNF in splicing was supported by our observation of protein-protein interactions between splicing factors and the SWI/SNF complex that were exclusive to SCCOHT cells and SMARCA4 localization to differentially-spliced genes (Figures 4, 5). We validated the interaction between SWI/SNF and splicing factors in an isogenic HAP1 cell line with SMARCA4 KO, where splicing factors similarly interacted with the SWI/SNF complex. Importantly, we also validated the protein expression of novel protein sequences derived from retained introns by mass spectrometry (Figure 3C-D). Additionally, most of the IR events were not expressed in normal tissues, and several of the novel, splicing-derived protein sequences were predicted to bind MHC complexes (Figure 3C, Supp Table 3). Finally, we demonstrated that the novel peptides derived from IR events could elicit T cell responses *in vitro* demonstrated by T cell expansion and expression of TNFα and INFγ (Figure 6). Together, these data support that SMARCA4 loss in SCCOHT tumors impacts RNA splicing in a manner that leads to intron retention. We additionally provide preliminary evidence that splicing-derived putative TAAs may have a stimulatory effect on the immune system. Due to the rarity of SCCOHT, there are not currently immunocompetent models available to evaluate how those TAAs contribute to therapeutic responses. However, we postulate that retained introns could be a source of TAAs in SCCOHT patients with therapeutic potential.

Immunotherapeutic agents, especially ICIs, rely on the patient's immune response and have shown great promise in treatment across a wide range of tumor types. Initial successes in melanoma and lung cancer, tumor types with mutagenic etiologies, implicated tumor mutation burden as a factor in clinical response to ICIs (31-33). The abundance of novel, mutation-encoded proteins provide a source of neoantigens thought to be necessary for the observed immune responses. This led to the prevailing theory that high TMB tumors are immunologically "hot" and amenable to ICI treatment while low TMB tumors were expected to be "cold" and not respond to ICI. In the clinic, TMB has not proved to be predictive of responses to ICIs in all settings. Some high TMB tumors fail to respond, and some 'cold' tumors respond to ICI treatment (35,36,91). SCCOHT is one example of a low TMB tumor type with demonstrated responses to ICI treatments (35,36). The low mutation burden (92), absence of recurrent mutations in other genes-encoded antigens (Supp Table 4), and our observation that SMARCA4 loss leads to IR in SCCOHT cells suggest that SMARCA4-dependent retained introns may be a source of antigens in SCCOHT.

This is in line with a growing body of evidence that altered splicing is a source of tumor-associated antigens. Recent work examining RNA splicing in more than 8,700 TCGA tumors demonstrated pervasive splicing dysregulation across tumor types compared to normal tissues (44). The widespread splicing dysregulation generates novel peptides predicted to bind MHC-I, suggesting that splicing-derived, tumor-associated antigens may be common across tumor types. However, the mechanisms underlying splicing dysregulation in cancer have not been fully elucidated. Mutation of splicing factors has been shown to alter splicing in cancer cells, such as *SF3B1* mutations in melanoma leading to alternative 3' splice site selection (93). Further evidence for mutations in genes which are not splice factors will provide additional biomarkers for splicing dysregulation.

While we have investigated the role of SMARCA4 loss in SCCOHT splicing regulation, its paralog SMARCA2, has previously been shown to impact splicing through RNAPII elongation rate and phosphorylation status, which are necessary for proper recruitment of splicing factors (46). Prior studies in model organisms and our protein-protein interaction data have further confirmed that SWI/SNF complexes interact with splicing and RNA processing factors (Figure 4)

(49,50,94). The molecular mechanisms by which SMARCA4 regulates splicing have not been fully elucidated. A recent report showed that SMARCA4 and SMARCA2 influenced the association of splicing factors to nascent RNAs (48). We found no evidence for changes in chromatin accessibility or RNAPII phosphorylation state in mediating SMARCA4-dependent splicing outcomes in SCCOHT cells. Nonetheless, our data implicate SMARCA4 loss in intron retention as a potential source of antigens in the context of SCCOHT. Thus, specific driver mutations and/or specific tumor contexts may be highly susceptible to splicing dysregulation and consequently produce targetable splicing-derived, tumor-associated antigens.

The SWI/SNF protein-protein interaction networks we describe provide a valuable resource for exploring how SWI/SNF complexes may function differently in the context of SMARCA4 loss in tumor cells. SCCOHT cells provide an ideal model for this type of discovery as they do not express SMARCA2 and cannot compensate for the loss of SMARCA4. In addition to RNA processing and splicing factors, we found that the SWI/SNF complexes lacking ATPase module members showed differential protein-protein interactions with epigenetic and metabolic regulators. The SCCOHT SWI/SNF complex lacked interactions with the enhancer-associated histone methyltransferase MLL3/4/COMPASS complexes, which may play a functional role in recently described enhancer dysregulation in SCCOHT (72). The differential association of SWI/SNF with metabolic regulators poses interesting questions for future research. Because the IPs were performed using nuclear extracts, the observed interactions represent nuclear-localized, metabolic proteins. Recent research has illuminated a clear interdependence of histone modification states and metabolic states (95-97). It will be interesting to learn how altered metabolic states impact the chromatin landscape of SWI/SNF-deficient cancers in future studies.

This work adds to a growing literature showing that tumor-associated splicing events may provide a rich source of tumor-associated antigens. We posit that the appearance of novel splicing events or splicing events detected repeatedly in particular tumor types could be exploited to activate an immune response. For example, the 94 novel, tumor-associated, protein-coding outlier splice events observed in more than half (>6) primary SCCOHT tumors (Figure 1D) may serve as candidate immunogens for targeted immune therapies such as CAR-T. Future studies should evaluate the efficacy of the SMARCA4-dependent tumor-associated antigens identified here in the context of an active immune environment. Our results demonstrate that thorough investigations of altered splicing events, driven by specific genomic alterations such as SMARCA4 or others, may identify unique biomarkers of tumor-associated immunogenic targets.

Acknowledgments
We thank the SCCOHT patients and their families who contributed to this work. We thank the Collaborative Sequencing Center at TGen for their help in our sequencing studies. The graphical abstract was created using BioRender.com.

Author contributions
EAR, VDD, LBB, WPDH, BEW, and PP designed the study.
EAR, MA, DLR, KAO, RS, VDD, SJF, RK, ZNJ, and RM performed experiments.
AMH, KGM, RFH, EAR, JDL, MA, and RS analyzed data.
ANK provided secondary pathological review and IHC data analysis.
EAR, AMH, KGM, RFH, JDL, YW, VLZ, TGW, MK, WPDH, DGH, LRR, BEW, JMT, and PP provided scientific data interpretation.
EAR, AMH, KGM, MA, RFH, JDL, RS, VDD, TGW, and PP prepared the manuscript.
EAR, LBB, WPDH, DGH, LRR, BEW, JMT, and PP secured funding for this study.

Funding
This work was supported by NCI R01 CA195670 to PP, BEW, DGH, and JMT, the Baylor Scott & White Research Institute to PP, EAR, WPDH, JDL, the Markel Friedman Accelerator Fund to PP, JDL, EAR, LRR, and the Ovarian Cancer Research Alliance [Mentored Investigator Grant, 65001] to EAR. Research reported in this publication included work performed in the Integrated Mass Spectrometry Shared Resource and the Analytical Cytometry Core supported by the National Cancer Institute of the National Institutes of Health under grant number P30CA033572. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Table 1. Clinical Parameters of SCCOHT cohort (n=10).

Age of onset (years)	**n**	
0-10	1	
11-20	1	
21-30	4	
31-40	4	
Age at tumor resection (years)	**n**	
0-10	1	
11-20	0	
21-30	4	
31-40	5	
Specimen collection site	**n**	
Primary	2	
Metastasis	3	
Unknown	5	
FIGO stage	**n**	
I	1	
II	3	
III	4	
IV	0	
Unknown	2	
Variants (SWI/SNF genes)	**SMARCA4**	**SMARCB1**
TGEN-A	p.Q413*	
TGEN-B	no known mutation; IHC negative	
TGEN-C	p.E667fs p.L1161fs	
TGEN-D	p.R1189*	
TGEN-E	p.V204fs p.W764fs	
TGEN-F	p.G836*	

TGEN-G	no known mutation; IHC positive	p.N34fs p.F412fs
TGEN-H	p.G990fs	
TGEN-I	c.2123+1G>A	
TGEN-J	p.A945fs	

References

1. Kadoch C, Hargreaves DC, Hodges C, Elias L, Ho L, Ranish J*, et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. Nat Genet **2013**;45:592-601
2. Shain AH, Pollack JR. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. PLoS One **2013**;8:e55119
3. Versteege I, Sevenet N, Lange J, Rousseau-Merck MF, Ambros P, Handgretinger R*, et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. Nature **1998**;394:203-6
4. Biegel JA, Zhou JY, Rorke LB, Stenstrom C, Wainwright LM, Fogelgren B. Germ-line and acquired mutations of INI1 in atypical teratoid and rhabdoid tumors. Cancer Res **1999**;59:74-9
5. Witkowski L, Carrot-Zhang J, Albrecht S, Fahiminiya S, Hamel N, Tomiak E*, et al.* Germline and somatic SMARCA4 mutations characterize small cell carcinoma of the ovary, hypercalcemic type. Nat Genet **2014**;46:438-43
6. Sullivan LM, Folpe AL, Pawel BR, Judkins AR, Biegel JA. Epithelioid sarcoma is associated with a high percentage of SMARCB1 deletions. Mod Pathol **2013**;26:385-92
7. Ramos P, Karnezis AN, Craig DW, Sekulic A, Russell ML, Hendricks WP*, et al.* Small cell carcinoma of the ovary, hypercalcemic type, displays frequent inactivating germline and somatic mutations in SMARCA4. Nat Genet **2014**;46:427-9
8. Jelinic P, Mueller JJ, Olvera N, Dao F, Scott SN, Shah R*, et al.* Recurrent SMARCA4 mutations in small cell carcinoma of the ovary. Nat Genet **2014**;46:424-6
9. Wei XL, Wang DS, Xi SY, Wu WJ, Chen DL, Zeng ZL*, et al.* Clinicopathologic and prognostic relevance of ARID1A protein loss in colorectal cancer. World J Gastroenterol **2014**;20:18404-12
10. Reisman DN, Sciarrotta J, Wang W, Funkhouser WK, Weissman BE. Loss of BRG1/BRM in human lung cancer cell lines and primary lung cancers: correlation with poor prognosis. Cancer Res **2003**;63:560-6
11. Numata M, Morinaga S, Watanabe T, Tamagawa H, Yamamoto N, Shiozawa M*, et al.* The clinical significance of SWI/SNF complex in pancreatic cancer. Int J Oncol **2013**;42:403-10
12. Hakimi AA, Chen YB, Wren J, Gonen M, Abdel-Wahab O, Heguy A*, et al.* Clinical and pathologic impact of select chromatin-modulating tumor suppressors in clear cell renal cell carcinoma. Eur Urol **2013**;63:848-54
13. Cho H, Kim JS, Chung H, Perry C, Lee H, Kim JH. Loss of ARID1A/BAF250a expression is linked to tumor progression and adverse prognosis in cervical cancer. Hum Pathol **2013**;44:1365-74
14. Bai J, Mei P, Zhang C, Chen F, Li C, Pan Z*, et al.* BRG1 is a prognostic marker and potential therapeutic target in human breast cancer. PLoS One **2013**;8:e59772
15. Helming KC, Wang X, Wilson BG, Vazquez F, Haswell JR, Manchester HE*, et al.* ARID1B is a specific vulnerability in ARID1A-mutant cancers. Nat Med **2014**;20:251-4
16. Hoffman GR, Rahal R, Buxton F, Xiang K, McAllister G, Frias E*, et al.* Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. Proc Natl Acad Sci U S A **2014**;111:3128-33
17. Wang ZQ, Zhang ZC, Wu YY, Pi YN, Lou SH, Liu TB*, et al.* Bromodomain and extraterminal (BET) proteins: biological functions, diseases, and targeted therapy. Signal Transduct Target Ther **2023**;8:420
18. Wilson BG, Wang X, Shen X, McKenna ES, Lemieux ME, Cho YJ*, et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. Cancer Cell **2010**;18:316-28

19. Wang Y, Chen SY, Karnezis AN, Colborne S, Santos ND, Lang JD*, et al.* The histone methyltransferase EZH2 is a therapeutic target in small cell carcinoma of the ovary, hypercalcaemic type. J Pathol **2017**;242:371-83
20. Knutson SK, Warholic NM, Wigle TJ, Klaus CR, Allain CJ, Raimondi A*, et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. Proc Natl Acad Sci U S A **2013**;110:7922-7
21. Lang JD, Hendricks WPD, Orlando KA, Yin H, Kiefer J, Ramos P*, et al.* Ponatinib Shows Potent Antitumor Activity in Small Cell Carcinoma of the Ovary Hypercalcemic Type (SCCOHT) through Multikinase Inhibition. Clin Cancer Res **2018**;24:1932-43
22. Xue Y, Meehan B, Macdonald E, Venneti S, Wang XQD, Witkowski L*, et al.* CDK4/6 inhibitors target SMARCA4-determined cyclin D1 deficiency in hypercalcemic small cell carcinoma of the ovary. Nat Commun **2019**;10:558
23. Xue Y, Meehan B, Fu Z, Wang XQD, Fiset PO, Rieker R*, et al.* SMARCA4 loss is synthetic lethal with CDK4/6 inhibition in non-small cell lung cancer. Nat Commun **2019**;10:557
24. Soldi R, Ghosh Halder T, Weston A, Thode T, Drenner K, Lewis R*, et al.* The novel reversible LSD1 inhibitor SP-2577 promotes anti-tumor immunity in SWItch/Sucrose-NonFermentable (SWI/SNF) complex mutated ovarian cancer. PLoS One **2020**;15:e0235705
25. Samartzis EP, Gutsche K, Dedes KJ, Fink D, Stucki M, Imesch P. Loss of ARID1A expression sensitizes cancer cells to PI3K- and AKT-inhibition. Oncotarget **2014**;5:5295-303
26. Molina JR, Sun Y, Protopopova M, Gera S, Bandi M, Bristow C*, et al.* An inhibitor of oxidative phosphorylation exploits cancer vulnerability. Nat Med **2018**;24:1036-46
27. Lissanu Deribe Y, Sun Y, Terranova C, Khan F, Martinez-Ledesma J, Gay J*, et al.* Mutations in the SWI/SNF complex induce a targetable dependence on oxidative phosphorylation in lung cancer. Nat Med **2018**;24:1047-57
28. Shen J, Peng Y, Wei L, Zhang W, Yang L, Lan L*, et al.* ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. Cancer Discov **2015**;5:752-67
29. Park Y, Chui MH, Suryo Rahmanto Y, Yu ZC, Shamanna RA, Bellani MA*, et al.* Loss of ARID1A in Tumor Cells Renders Selective Vulnerability to Combined Ionizing Radiation and PARP Inhibitor Therapy. Clin Cancer Res **2019**;25:5584-94
30. Wu S, Fatkhutdinov N, Fukumoto T, Bitler BG, Park PH, Kossenkov AV*, et al.* SWI/SNF catalytic subunits' switch drives resistance to EZH2 inhibitors in ARID1A-mutated cells. Nat Commun **2018**;9:4116
31. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. Science **2015**;348:69-74
32. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer **2012**;12:252-64
33. Kostron H, Grunert V. [Photodynamic therapy of malignant brain tumors]. Wien Klin Wochenschr **1987**;99:389-92
34. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science **2018**;359:1350-5
35. Li G, Jiang Y. Case Report: A Durable Response to Camrelizumab and Apatinib Combination Therapy in a Heavily Treated Small Cell Carcinoma of the Ovary, Hypercalcemic Type. Front Oncol **2022**;12:916790
36. Jelinic P, Ricca J, Van Oudenhove E, Olvera N, Merghoub T, Levine DA*, et al.* Immune-Active Microenvironment in Small Cell Carcinoma of the Ovary, Hypercalcemic Type: Rationale for Immune Checkpoint Blockade. J Natl Cancer Inst **2018**;110:787-90
37. Abou Alaiwi S, Nassar AH, Xie W, Bakouny Z, Berchuck JE, Braun DA*, et al.* Mammalian SWI/SNF Complex Genomic Alterations and Immune Checkpoint Blockade in Solid Tumors. Cancer Immunol Res **2020**;8:1075-84

38. Schoenfeld AJ, Bandlamudi C, Lavery JA, Montecalvo J, Namakydoust A, Rizvi H*, et al.* The Genomic Landscape of SMARCA4 Alterations and Associations with Outcomes in Patients with Lung Cancer. Clin Cancer Res **2020**;26:5701-8
39. Miao D, Margolis CA, Gao W, Voss MH, Li W, Martini DJ*, et al.* Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science **2018**;359:801-6
40. Alessi JV, Ricciuti B, Spurr LF, Gupta H, Li YY, Glass C*, et al.* SMARCA4 and Other SWItch/Sucrose NonFermentable Family Genomic Alterations in NSCLC: Clinicopathologic Characteristics and Outcomes to Immune Checkpoint Inhibition. J Thorac Oncol **2021**;16:1176-87
41. Berkers CR, de Jong A, Schuurman KG, Linnemann C, Geenevasen JA, Schumacher TN*, et al.* Peptide Splicing in the Proteasome Creates a Novel Type of Antigen with an Isopeptide Linkage. J Immunol **2015**;195:4075-84
42. Sebestyen E, Singh B, Minana B, Pages A, Mateo F, Pujana MA*, et al.* Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. Genome Res **2016**;26:732-44
43. Jayasinghe RG, Cao S, Gao Q, Wendl MC, Vo NS, Reynolds SM*, et al.* Systematic Analysis of Splice-Site-Creating Mutations in Cancer. Cell Rep **2018**;23:270-81 e3
44. Kahles A, Lehmann KV, Toussaint NC, Huser M, Stark SG, Sachsenberg T*, et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell **2018**;34:211-24 e6
45. Frankiw L, Baltimore D, Li G. Alternative mRNA splicing in cancer immunotherapy. Nat Rev Immunol **2019**;19:675-87
46. Batsche E, Yaniv M, Muchardt C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. Nat Struct Mol Biol **2006**;13:22-9
47. Raab JR, Smith KN, Spear CC, Manner CJ, Calabrese JM, Magnuson T. SWI/SNF remains localized to chromatin in the presence of SCHLAP1. Nat Genet **2019**;51:26-9
48. Ganez-Zapater A, Mackowiak SD, Guo Y, Tarbier M, Jordan-Pla A, Friedlander MR*, et al.* The SWI/SNF subunit BRG1 affects alternative splicing by changing RNA binding factor interactions with nascent RNA. Mol Genet Genomics **2022**;297:463-84
49. Tyagi A, Ryme J, Brodin D, Ostlund Farrants AK, Visa N. SWI/SNF associates with nascent pre-mRNPs and regulates alternative pre-mRNA processing. PLoS Genet **2009**;5:e1000470
50. Yu S, Jordan-Pla A, Ganez-Zapater A, Jain S, Rolicka A, Ostlund Farrants AK*, et al.* SWI/SNF interacts with cleavage and polyadenylation factors and facilitates pre-mRNA 3' end processing. Nucleic Acids Res **2018**;46:8557-73
51. Ramos P, Karnezis AN, Hendricks WP, Wang Y, Tembe W, Zismann VL*, et al.* Loss of the tumor suppressor SMARCA4 in small cell carcinoma of the ovary, hypercalcemic type (SCCOHT). Rare Dis **2014**;2:e967148
52. Karnezis AN, Wang Y, Ramos P, Hendricks WP, Oliva E, D'Angelo E*, et al.* Dual loss of the SWI/SNF complex ATPases SMARCA4/BRG1 and SMARCA2/BRM is highly sensitive and specific for small cell carcinoma of the ovary, hypercalcaemic type. J Pathol **2016**;238:389-400
53. Halperin RF, Hegde A, Lang JD, Raupach EA, Group CRR, Legendre C*, et al.* Improved methods for RNAseq-based alternative splicing analysis. Sci Rep **2021**;11:10740
54. Otte A, Gohring G, Steinemann D, Schlegelberger B, Groos S, Langer F*, et al.* A tumor-derived population (SCCOHT-1) as cellular model for a small cell ovarian carcinoma of the hypercalcemic type. Int J Oncol **2012**;41:765-75
55. Upchurch KS, Parker LM, Scully RE, Krane SM. Differential cyclic AMP responses to calcitonin among human ovarian carcinoma cell lines: a calcitonin-responsive line derived from a rare tumor type. J Bone Miner Res **1986**;1:299-304

56. Wang Y, Chen SY, Colborne S, Lambert G, Shin CY, Santos ND*, et al.* Histone Deacetylase Inhibitors Synergize with Catalytic Inhibitors of EZH2 to Exhibit Antitumor Activity in Small Cell Carcinoma of the Ovary, Hypercalcemic Type. Mol Cancer Ther **2018**;17:2767-79
57. Karnezis AN, Chen SY, Chow C, Yang W, Hendricks WPD, Ramos P*, et al.* Re-assigning the histologic identities of COV434 and TOV-112D ovarian cancer cell lines. Gynecol Oncol **2021**;160:568-78
58. Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, Mulherkar N*, et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. Nature **2011**;477:340-3
59. Orlando KA, Douglas AK, Abudu A, Wang Y, Tessier-Cloutier B, Su W*, et al.* Re-expression of SMARCA4/BRG1 in small cell carcinoma of ovary, hypercalcemic type (SCCOHT) promotes an epithelial-like gene signature through an AP-1-dependent mechanism. Elife **2020**;9
60. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **2009**;25:1105-11
61. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR*, et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc **2012**;7:562-78
62. Boegel S, Lower M, Schafer M, Bukur T, de Graaf J, Boisguerin V*, et al.* HLA typing from RNA-Seq sequence reads. Genome Med **2012**;4:102
63. Bai Y, Wang D, Fury W. PHLAT: Inference of High-Resolution HLA Types from RNA and Whole Exome Sequencing. Methods Mol Biol **2018**;1802:193-201
64. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. J Immunol **2017**;199:3360-8
65. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB*, et al.* MHC-I Genotype Restricts the Oncogenic Mutational Landscape. Cell **2017**;171:1272-83 e15
66. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A*, et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics **2009**;25:1091-3
67. Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ*, et al.* The GOA database: gene Ontology annotation updates for 2015. Nucleic Acids Res **2015**;43:D1057-63
68. **!!! INVALID CITATION !!!**
69. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc **2006**;1:2856-60
70. Dayal B, Salen G. Stereospecific synthesis and two-dimensional 1H-NMR investigation of isoursocholic acid. J Lipid Res **1991**;32:1381-7
71. Teo G, Liu G, Zhang J, Nesvizhskii AI, Gingras AC, Choi H. SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. J Proteomics **2014**;100:37-43
72. Pan J, McKenzie ZM, D'Avino AR, Mashtalir N, Lareau CA, St Pierre R*, et al.* The ATPase module of mammalian SWI/SNF family complexes mediates subcomplex identity and catalytic activity-independent genomic targeting. Nat Genet **2019**;51:618-26
73. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods **2012**;9:357-9
74. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO*, et al.* Twelve years of SAMtools and BCFtools. Gigascience **2021**;10
75. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS*, et al.* deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res **2016**;44:W160-5

76. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature **2012**;489:57-74
77. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE*, et al.* Model-based analysis of ChIP-Seq (MACS). Genome Biol **2008**;9:R137
78. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics **2015**;31:2382-3
79. Gaspar JM. NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. BMC Bioinformatics **2018**;19:536
80. Higgins JR, Halpin DM, Midgley AK. Tension pneumoperitoneum: a surgical emergency. Br J Hosp Med **1988**;39:160-1
81. Lun AT, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Res **2016**;44:e45
82. Lun AT, Smyth GK. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. Nucleic Acids Res **2014**;42:e95
83. Ovarian Tumor Tissue Analysis C, Goode EL, Block MS, Kalli KR, Vierkant RA, Chen W*, et al.* Dose-Response Association of CD8+ Tumor-Infiltrating Lymphocytes and Survival Time in High-Grade Serous Ovarian Cancer. JAMA Oncol **2017**;3:e173290
84. Milne K, Kobel M, Kalloger SE, Barnes RO, Gao D, Gilks CB*, et al.* Systematic analysis of immune infiltrates in high-grade serous ovarian cancer reveals CD20, FoxP3 and TIA-1 as positive prognostic factors. PLoS One **2009**;4:e6412
85. Kahles A, Ong CS, Zhong Y, Ratsch G. SplAdder: identification, quantification and testing of alternative splicing events from RNA-Seq data. Bioinformatics **2016**;32:1840-7
86. Smart AC, Margolis CA, Pimentel H, He MX, Miao D, Adeegbe D*, et al.* Intron retention is a source of neoepitopes in cancer. Nat Biotechnol **2018**;36:1056-8
87. Altmann J, Schmitt W, Bashian N, Sehouli J. A dramatic response to checkpoint inhibitor in a woman with small cell carcinoma of the hypercalcemic type of the ovary. Gynecol Oncol **2024**;181:99-101
88. Schick S, Rendeiro AF, Runggatscher K, Ringler A, Boidol B, Hinkel M*, et al.* Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalities in human cancers. Nat Genet **2019**;51:1399-410
89. Tischkowitz M, Huang S, Banerjee S, Hague J, Hendricks WPD, Huntsman DG*, et al.* Small-Cell Carcinoma of the Ovary, Hypercalcemic Type-Genetics, New Treatment Targets, and Current Management Guidelines. Clin Cancer Res **2020**;26:3908-17
90. Witkowski L, Goudie C, Foulkes WD, McCluggage WG. Small-Cell Carcinoma of the Ovary of Hypercalcemic Type (Malignant Rhabdoid Tumor of the Ovary): A Review with Recent Developments on Pathogenesis. Surg Pathol Clin **2016**;9:215-26
91. McGrail DJ, Pilie PG, Rashid NU, Voorwerk L, Slagter M, Kok M*, et al.* High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. Ann Oncol **2021**;32:661-72
92. Auguste A, Blanc-Durand F, Deloger M, Le Formal A, Bareja R, Wilkes DC*, et al.* Small Cell Carcinoma of the Ovary, Hypercalcemic Type (SCCOHT) beyond SMARCA4 Mutations: A Comprehensive Genomic Analysis. Cells **2020**;9
93. Zhou Z, Gong Q, Wang Y, Li M, Wang L, Ding H*, et al.* The biological function and clinical significance of SF3B1 mutations in cancer. Biomark Res **2020**;8:38
94. Patrick KL, Ryan CJ, Xu J, Lipp JJ, Nissen KE, Roguev A*, et al.* Genetic interaction mapping reveals a role for the SWI/SNF nucleosome remodeler in spliceosome activation in fission yeast. PLoS Genet **2015**;11:e1005074
95. Reid MA, Dai Z, Locasale JW. The impact of cellular metabolism on chromatin dynamics and epigenetics. Nat Cell Biol **2017**;19:1298-306
96. Sabari BR, Zhang D, Allis CD, Zhao Y. Metabolic regulation of gene expression through histone acylations. Nat Rev Mol Cell Biol **2017**;18:90-101

97. Zhang T, Gong Y, Meng H, Li C, Xue L. Symphony of epigenetic and metabolic regulation-interaction between the histone methyltransferase EZH2 and metabolism of tumor. Clin Epigenetics **2020**;12:72

Figure Legends:

Figure 1. Identification of novel protein-coding outlier intron retention events in primary SCCOHT tumors. A. Percentage of novel protein-coding splice event types that were outliers compared to normal tissues in GTEx using Bisbee outlier analysis in each of 10 primary SCCOHT tumors. The total number of novel protein-coding outlier splice events for each tumor is indicated above its column. **B.** Number of primary SCCOHT tumors sharing a particular novel protein-coding outlier splice event (x-axis) as in (A). **C.** Distribution of the total number of outlier novel protein-coding splice event types observed across 10 primary SCCOHT tumors as in (A-B). A3 = alternative 3' splice site, A5 = alternative 5' splice site, MUT = mutually exclusive exons, ES = exon skipping, IR = intron retention. **D.** Distribution of the total number of outlier novel protein-coding splice event types shared in 6+ tumors. **E.** IHC staining images of CD8, CD4, CD20, and CD68 in SCCOHT primary tumors TGEN-A and TGEN-B.

Figure 2. A subset of SWI/SNF members are associated with increased rates of intron retention. Proportion of IR events among total outlier splicing events comparing WT versus mutant for each of the SWI/SNF members above.

Figure 3. Effect of SMARCA4 expression on splicing in SCCOHT cells. A. The number of novel protein-coding splice events differ between SCCOHT cells lacking SMARCA4 compared to cells where wild-type SMARCA4 has been exogenously re-expressed in the BIN67 and COV434 cell lines (n=3) using Bisbee differential analysis. **B.** Predicted MHC-I binding strength of SMARCA4-dependent differentially-spliced events observed in (A). **C.** Classification of SMARCA4-dependent novel protein-coding differentially-spliced genes with MHC-I binding predictions and protein expression detection by mass spectrometry as in (A-B). **D.** Log2TPM values of genes with novel sequences found in -SMARC4 conditions compared to +SMARCA4 conditions.

Figure 4. SCCOHT SWI/SNF complex interacts with more splicing factors than the wild-type complex. A. UpSet plot comparing cumulative overlap of SWI/SNF PPIs (AvgP≥ 0.7) from IP-MS data across three antibodies recognizing core SWI/SNF subunits (SMARCC1, SMARCB1, SMARCE1) comparing HAP1, BIN67, SCCOHT1, and COV434. **B.** Interaction network of SWI/SNF interacting proteins identified exclusively in HAP1 bearing a wild-type SWI/SNF (blue). **C.** Interaction network of proteins shared between wild-type and SCCOHT SWI/SNF complexes identified in all cell lines (orange). **D.** Interaction network of SWI/SNF-interacting proteins exclusive to SCCOHT cells. Proteins interacting with SWI/SNF in 2 or 3 SCCOHT cell lines are colored green or purple, respectively. For clarity, proteins showing no first-degree interactions in StringDB are not shown (B-D). **E.** Number of SWI/SNF-interacting proteins associated with splicing regulation by Gene Ontology detected in immunoprecipitation and mass spectrometry using the union of proteins detected with SMARCC1, SMARCB1, and SMARCE1 antibodies comparing SCCOHT cells (BIN67, SCCOHT1, and COV434) to cells bearing a wild-type SWI/SNF complex (HAP1) (n=3, probability of interaction ≥ 0.7) (detailed in Supp Figure 7).

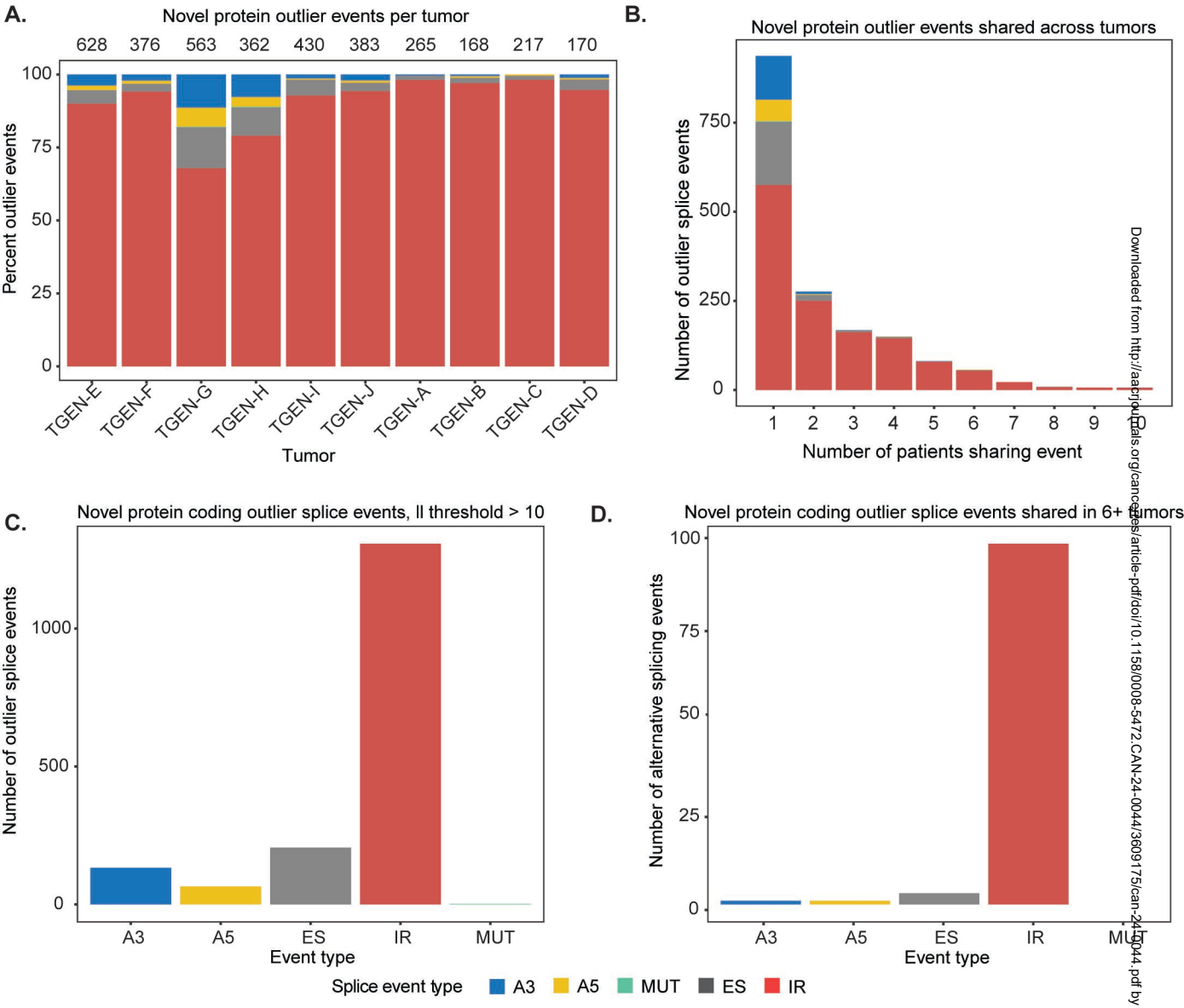
Figure 5. SMARCA4 localization to differentially-spliced genes does not coincide with changes in chromatin accessibility. A. SMARCA4, SMARCC1, DPF2, and ARID2 occupancy at differentially-spliced genes identified in BIN67 cells measured by chromatin immunoprecipitation followed by sequencing comparing BIN67 cells lacking SMARCA4 expression to BIN67 cells in which wild-type SMARCA4 has been exogenously re-expressed (n=1, Pan, et al. (2019)). **B.** Differential chromatin accessibility at differentially-spliced genes in BIN67 as measured by ATACseq comparing BIN67 cells lacking SMARCA4 expression to

BIN67 cell in which SMARCA4 has been exogenously re-expressed (n=2, Pan, et al. (2019)). **C.** Differential chromatin accessibility at ATACseq peaks within BIN67 differentially-spliced genes where -SMARCA4 indicates peaks that were exclusive to -SMARCA4 cells, +SMARCA4 indicates peaks that were exclusive to +SMARCA4 cells, and STATIC indicates peaks that were shared between conditions as in (B)

Figure 6. Peptides derived from SMARCA4-dependent differentially-spliced genes treatment induce T cell proliferation. T cells enriched from healthy donor PBMCs (n = 3-5) were stimulated for 96 hours using custom tiled peptide libraries of peptides resulting from alternative splicing events identified specifically in SCCOHT tumors and cell lines (PepMix 1 & 2). The UltraStim library (JPT) containing common antigenic peptides against all major HLA subtypes and Dynabeads coated with CD3 and CD28 antibodies serve as positive controls for stimulation. **A.** T cell proliferation measured by dilution of the CellTrace Violet dye monitored by flow cytometry**.** Plots show the gating strategy used to assess T cell proliferation, representative histograms of proliferation measured by CellTrace Violet, and quantification of replicate T cell proliferation measurements. **B.** Intracellular staining of TNFa and INFg expression in CD3+ T cells measured by flow cytometry. Plots show the gating strategy used to asses TNFa and INFGg expression, representative TNFa and INFg measurements, and quantification of TNFa and INFg across multiple healthy donors. Statistical significance determined by one-way ANOVA with Bonferroni multiple comparisons test.

Graphic abstract. Graphical representation of SMARCA4 loss leading to intron retention, antigen expression, and T cell stimulation. Created in BioRender. Raupach, E. (2025) https://BioRender.com/tljp61c.

FIGURE 1



E.

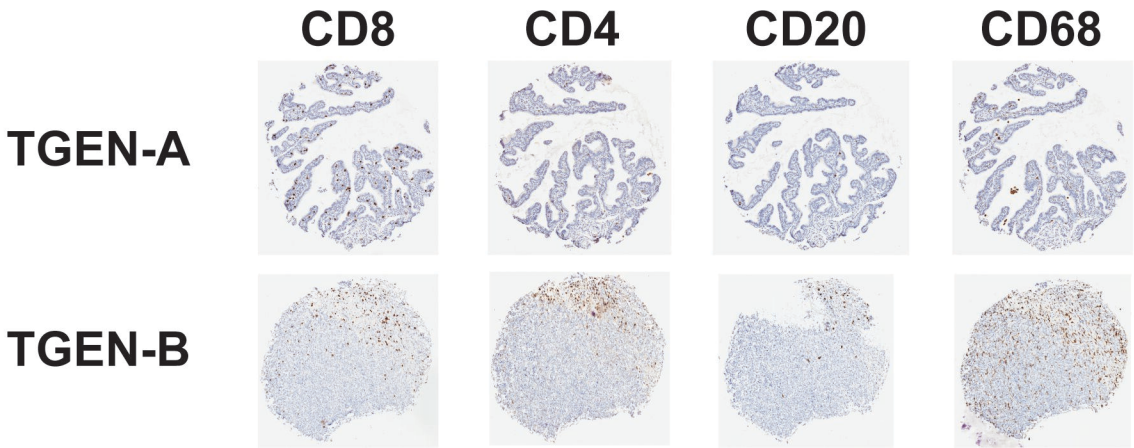


FIGURE 2

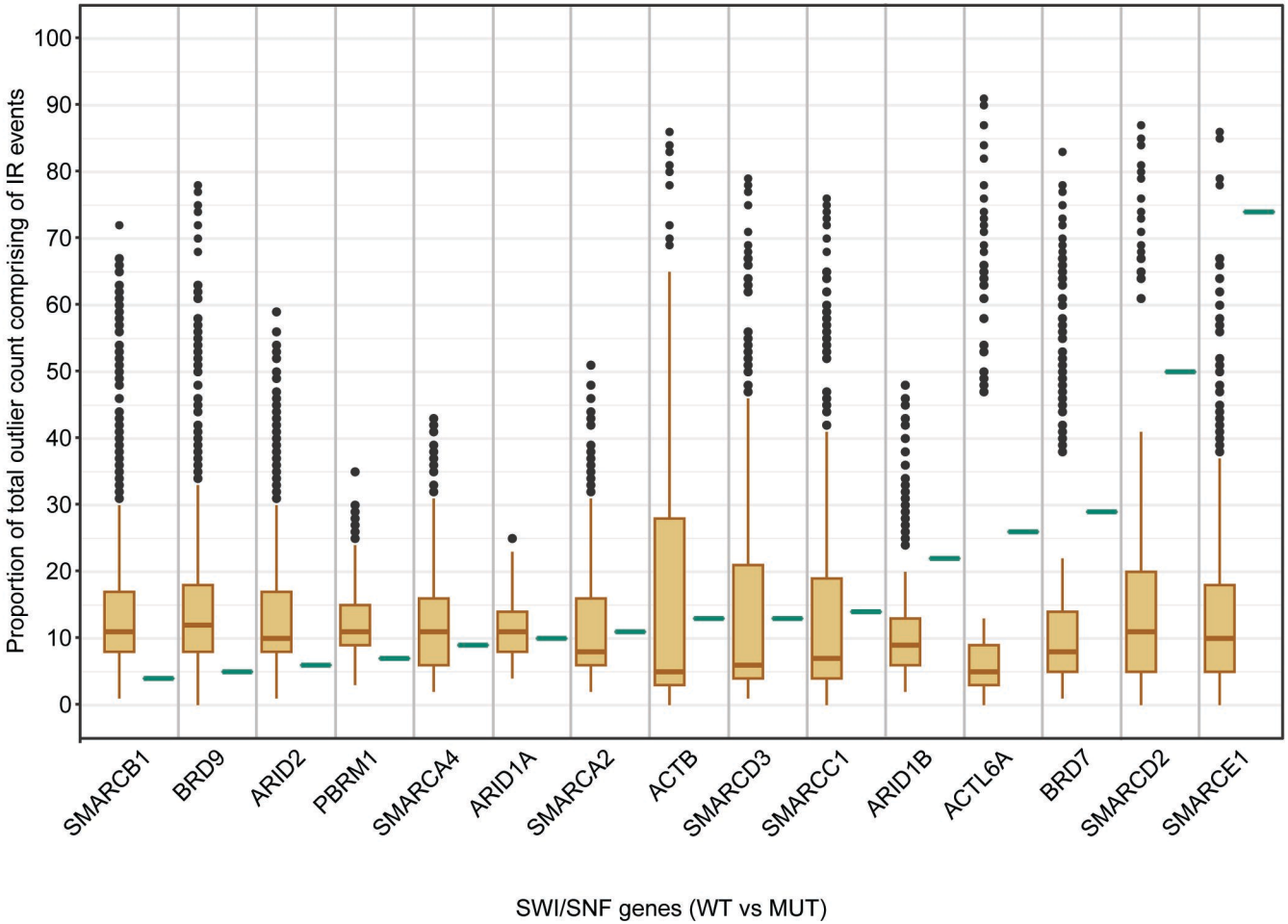


FIGURE 3

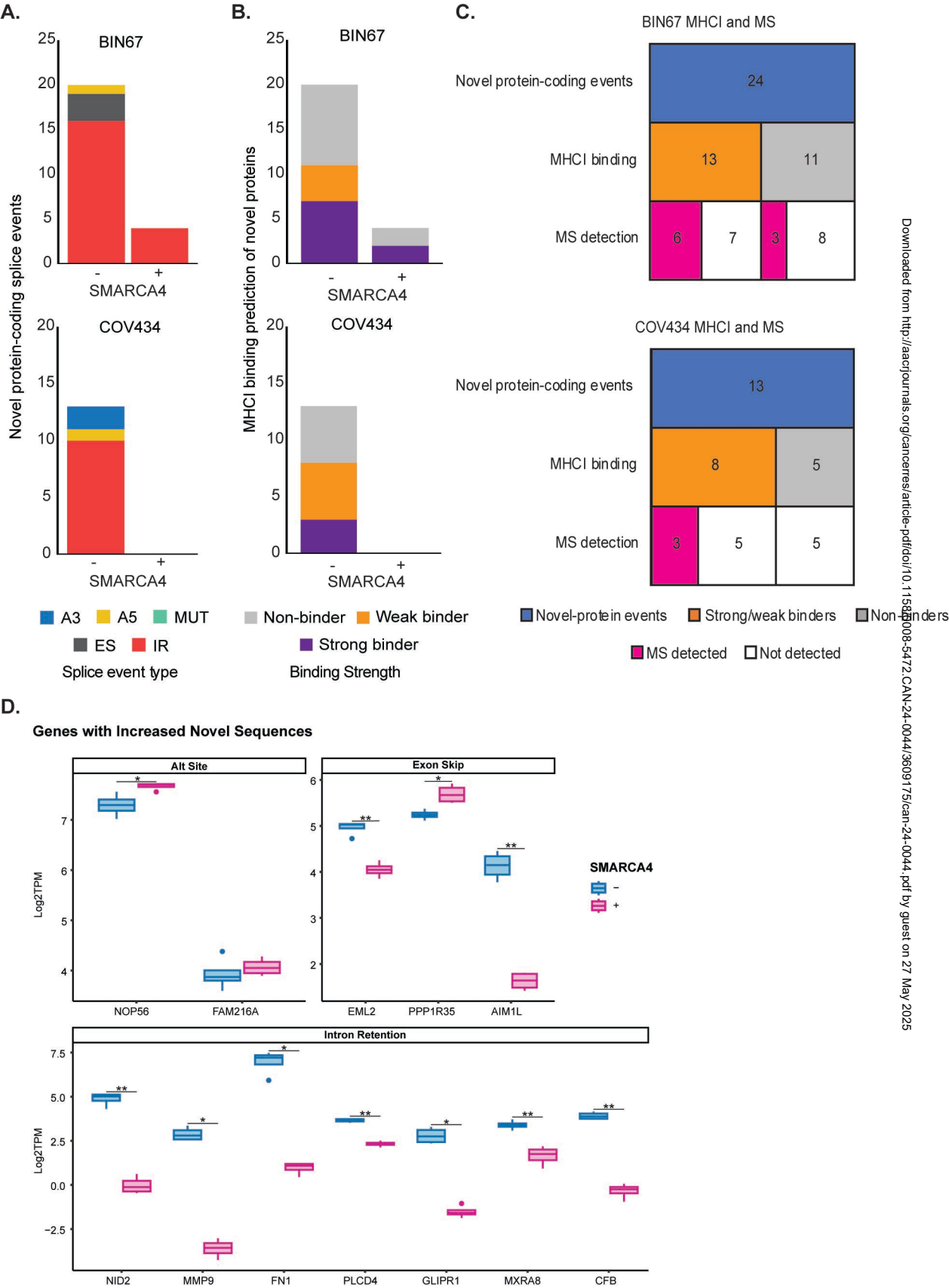


FIGURE 4

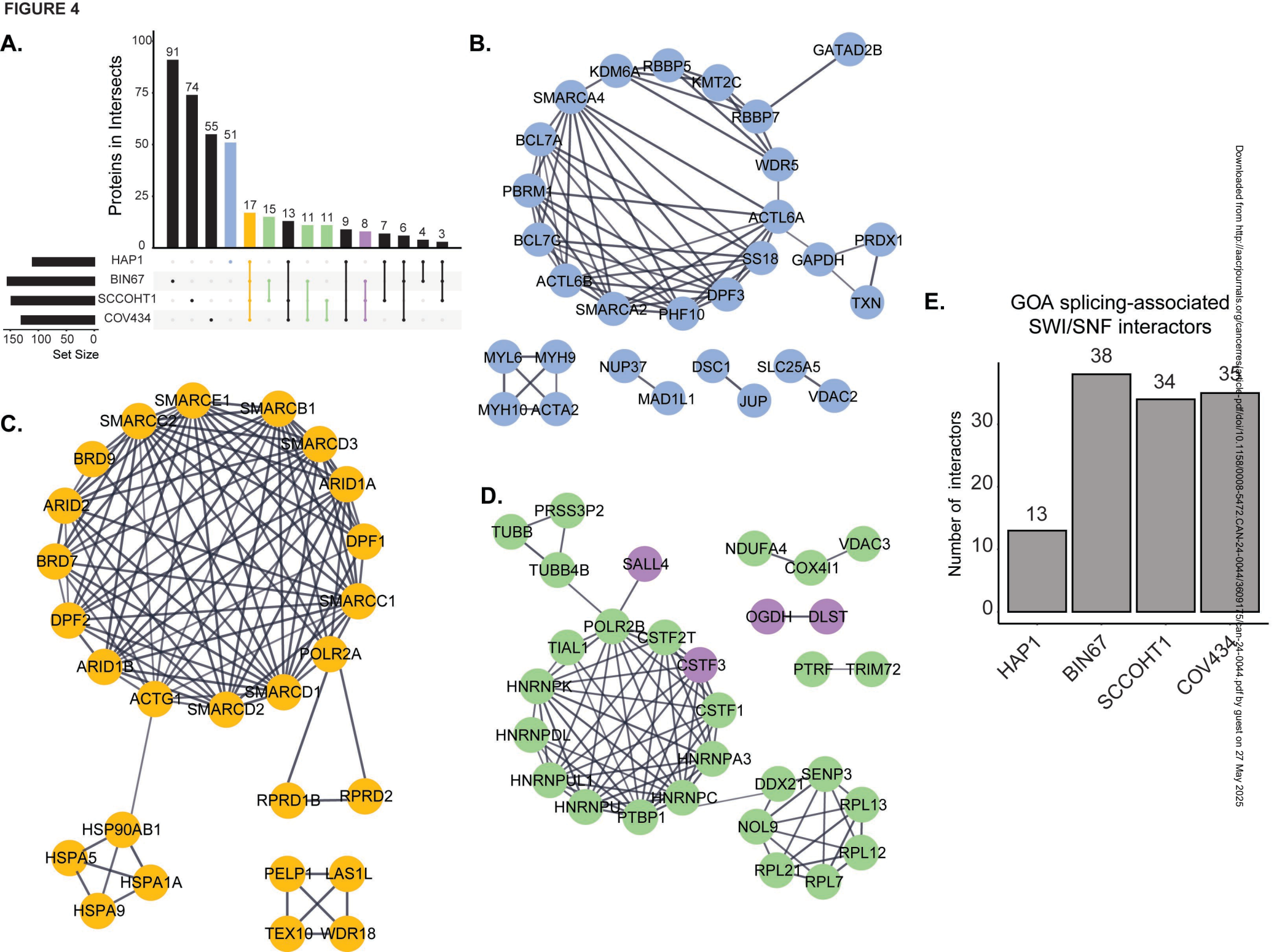


FIGURE 5

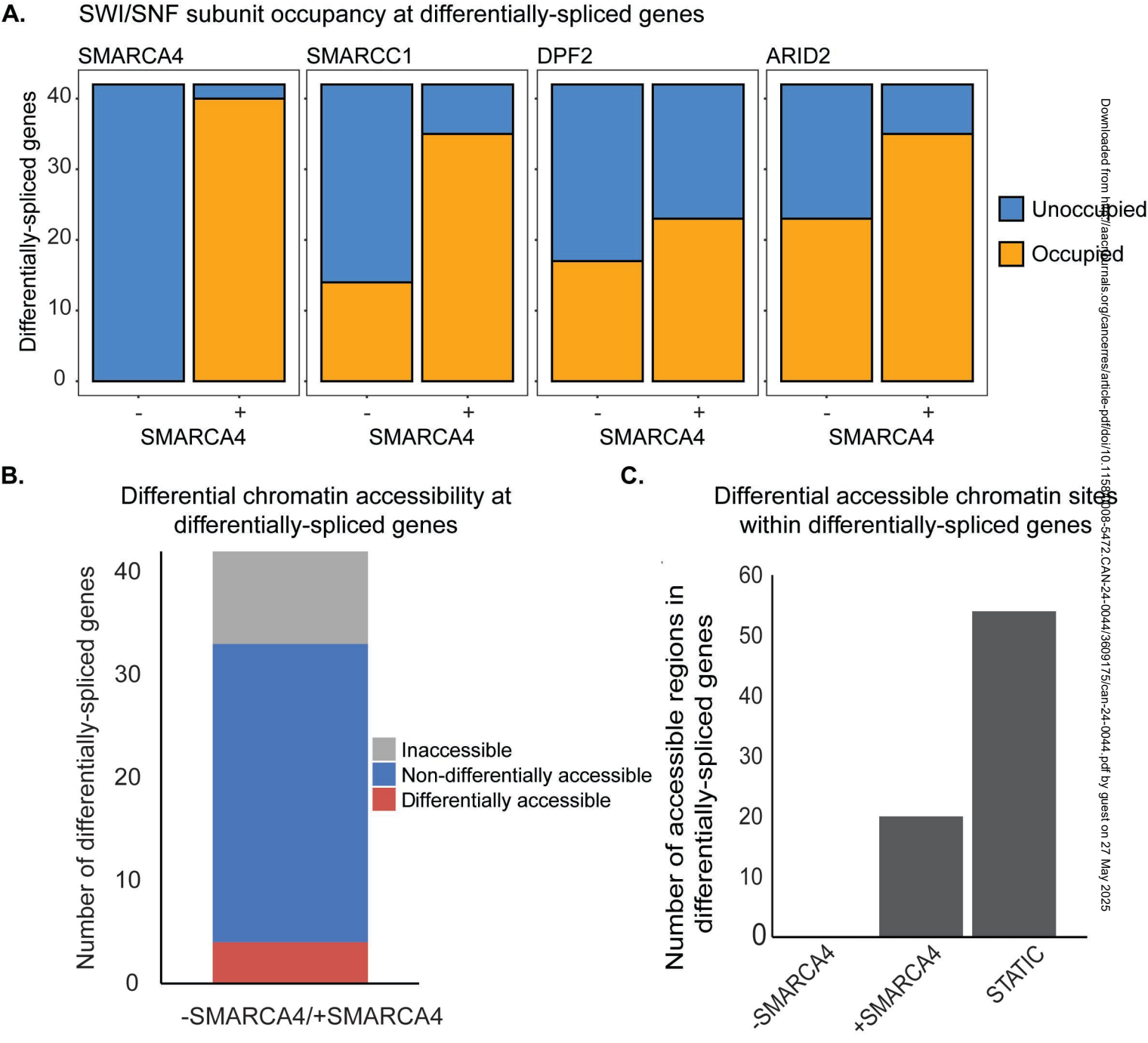


FIGURE 6