Title

Tumor-wide RNA splicing aberrations generate immunogenic public neoantigens

Author

Kwok, Darwin

Publication Date

2024

Tumor-wide RNA splicing aberrations generate immunogenic public neoantigens

by
Darwin Kwok

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Hani Goodarzi
FDD44359FCC6487...

Hani Goodarzi
Chair

DocuSigned by:
Russell Pieper
DocuSigned by:
Aaron Diaz
DocuSigned by:
Hideho Okada
DocuSigned by:
Joseph Costello
917AF3FB58574BE...

Russell Pieper
Aaron Diaz
Hideho Okada
Joseph Costello
Committee Members

Dedication

In memory of Kenneth Kwok (September 7, 1940 – April 10, 2023)

My research, my entire Ph.D., and my passion for science have been driven by my father's decade-long battle with cancer. Diagnosed in 2014 with a vestibular schwannoma, my father underwent an unsuccessful surgery that nearly costed him his life. My mother and I had to work relentlessly to nurse him back to health, and for nearly a decade, our family strived to regain a sense of normalcy. However, in March 2023, my mother returned home to find my father on the ground, having succumbed to a severe bilateral ischemic stroke hours earlier. For the next month, my mother and I stayed by his side in the hospital, hoping each day that he would recognize and return to us. In April 2024, the man who held up my entire world passed away.

Growing up, I rarely saw my father. He worked two shifts as a cab driver and waiter to support my mother as she prepared for her medical board exams. He would leave for work before I woke up and return after I was asleep. When he was diagnosed with cancer in 2014, his physician recommended immediate tumor resection. Despite this, my father asked how long he could continue working without the surgery, determined to support me through college. He made my mother promise not to tell me about his diagnosis until it was absolutely necessary. I only found out later that year.

Although the tumor was successfully resected, complications during the surgery left my father permanently paralyzed and bedridden for months. This experience, along with his recovery journey, inspired me to dedicate my research to understanding cancer and exploring non-invasive treatment methods. I shifted my undergraduate Biomedical Engineering focus to nanoparticles and conducted research at UPMC and Yale University, delving into sonoporation and immunotherapy. Over the years, I developed a deep love for science, sharing every small discovery with my father. This passion led me

to choose science over medicine and pursue a Ph.D. at UCSF in 2018. Here, I have been able to pursue the research I have always dreamed of – discovering a novel non-invasive approach to treating brain tumors.

Thank you, Dad, for shaping me into the person I am today, for inspiring me to explore the unknown and strive for greatness, and for building a network of friends and colleagues who became like family. Thank you for believing in me despite my missteps in high school and beyond. Thank you for being my role model, demonstrating the values of hard work and dedication. And thank you for fighting so hard, supporting Mom and me, and living for us. I hope that everything I've done has made you proud, and I will continue to help others in your name. I did it, Dad. And it's all because of you. I miss you so much and love you beyond measure.

Acknowledgements

None of the work completed here would have been possible without the endless support and guidance from the countless individuals I have had the pleasure to meet throughout my PhD journey. I would like to first thank my two incredible scientific advisors, Dr. Hideho Okada and Dr. Joseph Costello, for being the main two pillars of my past six years. Without them, I would never have had the chance to grow as much as I have, both as a scientist and as a person. I am forever grateful to have had the opportunity to spend a significant portion of my professional career with these talented professors, and I am inspired to one day emulate them and inspire the next generation of learners and curious individuals.

When I first matriculated to UCSF, I wanted to immediately rotate with a lab that studies the disease that my father had battled for many years. Recently coming from a rigorous post-graduate training at Yale University in basic immunology and applied immunotherapy, Hideho's lab seemed to be the best fit. My relationship with Hideho evolved immensely from the first time we met to the day I graduated. From the formalities of all post-doctorates and post-graduates in his lab to the high expectations he holds for all of his trainees, I was initially worried that I would not be able to step up to the plate and deliver. But I could not have been more wrong. Hideho has repeatedly pulled me up when things got tough and pushed me forward when I was unsure. He consistently exercised his dedication to science to my benefit and made nearly all of our big decisions to support my career. Not many would support their student taking a left turn and apply to medical school during their PhD. Instead, Hideho urged me to never lose sight of that goal. When I was at a crossroads, he encouraged me to chase my dream and not worry about the little things, offering to take care of those himself. Hideho ensured that my worries were quelled and that anything I had going moved forward at blistering speeds. Darwin believing he was scooped from a recently published paper became an annual affair, and Hideho's

patience with dealing with these sporadic text messages showed just how much he cared. At the end of 2023, my thesis manuscript was just submitted to *Nature*, and its corresponding pre-print was made publicly available in *bioRxiv*. Not long after, I saw a newly published article titled "Alternative RNA splicing generates shared clonal neoantigens across different types of cancer" appear in my recommended PubMed manuscripts. I frantically texted Hideho and Joe before even reading the publications, and after realizing that the paper was a *Nature Reviews Immunology* Preprint Watch advertising my work on *bioRxiv*, I immediately texted back, "Just kidding. Please ignore it all." Jokes aside, Hideho is the type of person who would respond to me mid-flight with detailed edits of a proposal draft that I just sent to him. He makes me feel less like a student, and more of a member of the greater family that he has built in his lab. For that, I am incredibly lucky and thank Hideho for everything that he has given me all these years.

Joe's lab was where I had my fourth and final rotation. Unlike the rest of my rotations, I was least familiar with the scientific focus of Joe's projects. I chose to rotate with Joe because he was regarded unanimously by my first-year class as the nicest and most fatherly professor across all of BMS. After these last six years, I can confidently say that that was an understatement. I left ever interaction with Joe with a smile on my face. Joe's unwavering positivity for his trainees is unparalleled, and it was his laboratory environment that helped me march through my entire PhD. He places a large emphasis on understanding over doing and made sure that no matter what the topic may be, I would be confident with what I knew before the task begins. He made sure that I felt my opinions were valued and trusted my choices in experimental design. Joe made me feel that everything I did had purpose, and this extended beyond just the science. He would go out of his way to have lengthy conversations with both my mother and partner. He would joke about his time as a graduate student to offset the weight of unsuccessful experimental results. With Joe, I felt that nothing was the end, and that the possibilities were limitless. Thank you, Joe, and I am so grateful to have joined your lab.

I would like to also thank my thesis committee and qualifying exam committee for their infinite wisdom and guidance as my project matured. Drs. Manish Aghi, Mary Barcellos-Hoff, David Raleigh, and Aaron Diaz were stellar mentors and resources for my qualifying exam. They each imbued a substantial degree of confidence into my proposal and gently nudged me in the right direction when I become overly ambitious or short-sighted. Drs. Hani Goodarzi, Aaron Diaz, and Russ Pieper were the members of my thesis committee, and I thank them for their time and efforts in ensuring the success of my project and myself as a scientist. As my thesis committee chair, Hani has dedicated so much of his time towards giving me both scientific and career advice. Hani helped me navigate through the turmoil following my father's passing, and he gave me the greatest insights on career trajectories in academia, industry, and even medicine. He is truly an amazing mentor, and I hope to continue to have opportunities to work with him in the future.

The many collaborators who have hopped aboard my project made everything possible. I want to thank Dr. Chris Klebanoff and Dr. Inaki Etxeberria from Memorial Sloan Kettering Cancer Center for being an integral part of this project. Chris had done more than just collaborated in this work. He had met with me on multiple occasions to give me keen career advice. Inaki took on the huge burden of performing numerous mass spectrometry and *in vitro* sensitization experiments that had been critical in key experiments and revisions. Without Chris and Inaki, our project would not have been at the quality and impact that it is at now. I would also like to thank Dr. Arun Wiita and his lab for collaborating with us and teaching me how to interpret mass spectrometry results in both his course lectures and our one-on-one meetings. In particular, I would like to thank Dr. Emilio Ramos and Dr. Ahbilash Barpanda from his lab for their MaxQuant assistance with our mass spectrometry data.

I would also like to thank my earlier scientific mentors who inspired me to take the leap into science. At Yale University, Dr. Nikhil Joshi took me in as his trainee despite my request for only a one-year

timeframe. I came in with no prior immunology training or education, and within that year, Nik taught me enough to prepare me for a successful PhD in immunotherapy. He had unwavering faith in my ability to code and allowed me to flex my engineering training to design software for his lab. Dr. Martina Damo was my overseeing post-doctoral mentor in Nik's lab and the reason for my meticulousness when it comes towards designing experiments and interpreting data. She and Nik have both inspired me to become incredible mentors like themselves, and I hope to have the opportunity to work together with them in the near future.

The members of the Okada and Costello lab have also been incredible in assisting me with crucial experiments and being involved in my graduate school journey. Takahide was an excellent mentor who taught me the majority of the bioinformatics that I've implemented in the first half of my project. Payal was my initial rotation mentor and has continued advising me years afterward. Hirokazu, Atsuro, Akane, Ryusuke, Marco, and all the post-doctorates in the Okada lab have helped me with experiments, and Kaori, in particular, had dedicated months to assist me with key *in vitro* stimulation experiments. Ryan, Jared, Kira, Tiffany, Abigail, Christian, Su, Landon, and Jeff were many of the underappreciated post-graduates who helped keep the lab running and were the life of the lab. Thank you for the many memories and fun that we've shared together. In the Costello lab, I would like to thank Chibo, Stephanie, Lee, Daniel, Drew, Ben, and Christina for creating such a warm and hospitable environment. I've never looked forward so much to going to school until I started my PhD, and a large part of that was due to the amazing people I spent most of my day with. Being situated in the Costello lab, it was often difficult to get work done because of the many fun conversations happening around me. Many of these fun talks were sparked by none other than Nick. My biggest shoutout has to go to the most talented graduate student I have ever met. Nick would drop everything to help any of his colleagues and friends out of a tough situation. The number of late-night text messages and desperate cries for help sent to him during working hours are in the triple digits, and without him, I

would not have gotten this far. Nick and I shared the same goal throughout our academic careers, and he never stopped running towards it. Yet, when I slipped and fell, he'd run back to pull me up so we could continue running together. Nick deserves his PhD and mine. Thank you for being part of my life.

Beyond the Okada and Costello labs, I am grateful to have met, conversed, and even collaborated with many talented people in the Helen Diller Cancer Center. In the Diaz lab, I've had the pleasure of meeting Dr. Jangham Jung, who goes out of his way to offer his boundless wealth of knowledge despite his multitude of obligations. In the Oldham lab, Daniel Brody was a beacon of excitement who provided delightful talking points at any moment. In the Ho lab, I was fortunate enough to briefly meet Isha and share exciting and hilarious anecdotes from my lab mishaps. I am incredibly lucky to have met Dr. Peter Bruno towards the end of my PhD, through whom I gained a broader insight into what it takes to become a prospective young investigator in modern science.

I mentioned mentorship frequently throughout these acknowledgements, but the greatest role model I look up to for mentorship comes from Dr. Michael Zhang. I would see Mike in the Costello lab office working quietly, and one day, he saw me looking at the scans he was analyzing out of curiosity. He invited me to speak with him, offered to onboard me onto his project, and the rest was history. Long after transitioning to his full-time clinical role, Mike still kept in touch with me on a near weekly basis. Mike had no obligations to do these routine check-ins, but I am entirely grateful for them. It made me feel like someone worth checking in on, and it made our mentor-mentee relationship that much more genuine. This experience with Mike has heavily impacted my take on mentoring my own trainees, and I owe my success and theirs to Mike.

On the note of trainees, I would like to thank all of my many mentees who I have had the pleasure of working with throughout my PhD. Maggie, Jerry, Gary, Aidan, and James have all been spectacular

trainees who have either rotated or interned in the Okada or Costello labs. From being high school students to rotating PhD students, each one of them taught me valuable lessons on how to properly mold my approach to be personalized for every individual. Maggie, in particular, has been one of the most talented students that I've trained, and she has demonstrated in a multitude of instances that the student can become the master. I now defer most of my wet lab transduction experiments to her due to my inability to match her unparalleled transduction efficiencies.

I would also like to express my greatest gratitude to my BMS cohort. Spending time with these immensely talented individuals have made me appreciate new hobbies and activities beyond science. Excursions with Raymond, Adam, and Jackson in the early years of my PhD were some of the most exciting and memorable adventures I've ever experienced. I also must thank the many friends that I've met in San Francisco – both old and new. High school classmates such as Aaron, Shannon, and Jenn pulled me out of my laboratory cave to do fun and wholesome group activities together. New friends such as Abigail, Ben, Khris, and Jonathan took me out on more physically demanding excursions. And I have to give one of the biggest acknowledgements to the many friends scattered across the globe who continue to keep connected with me through long distances. From Gene, Jonathan, Clara, Can, Tianyang, Mursal, and everyone else still at Yale to Janice and Jenn in NYC, thank you so much for always checking in on me for all these years. To my friends from or currently in Hawaii: Robert, Sean, Kal, Shayd, Daniel, Kenton, Evan de Luna, Steve, Drew, Ryan, Alisse, Dillon, Justin, Bolin, Mayumi, Winney, Micah, Ian, and all, thank you for hanging out with me every time I return home or visit where you're all at. No words can explain the effect of having long term friends like you all on my mental health through a turbulent and prolonged PhD. To my League of Loogies crew, you know who you are, I'll probably see you gentlemen on the rift tonight. Finally, thank you Zach for being my best friend and the go-to man for every challenge I faced since I met you. Zach has visited me more times than I can count, and I have often wondered if he stuck around with me this much because I

am a good friend or if I'm more fascinating than his psychiatric patient cohort. Either way, I'm incredibly lucky to have such a caring and reliable individual as my righthand man, and I look forward to following in his footsteps into medicine.

I cannot talk about best friends without talking about Heaven. We met right before the COVID-19 pandemic, and who would have known that the few dates watching the John Wick trilogy would lead to something this magical and life-changing. Heaven supported me through the hundreds of papers that I've read prior to my qualifying exam, shown me countless foodie hotspots in San Francisco, and has been my safe haven in the hardest of times. Thank you for introducing me to your wonderful family and friends, who I love seeing every time I come to Seattle. For four years of my PhD, Heaven had been my lifeline and my joy, and I am the luckiest man to have found such a beautiful and compassionate woman who believes in everything that I am. I am excited to see where the next chapter of our lives take us.

Finally, I want to thank my parents for literally bestowing me the life and experiences that paved way for my opportunity and decision to pursue this PhD. As mentioned in my Dedication, none of this desire to pursue science would have been possible without witnessing my father's challenge with battling cancer and its aftermath. He has motivated my desire to understand the biology of cancer and to make a difference for everyone who is affected by this disease. If my father was the reason, then my mother was the unbridled support. My mother placed all her eggs in one basket and dedicated her entire life towards ensuring that I would have a happy life. Whether it was spending everything on providing me with the best education or getting me out of sticky situations that were entirely my fault, my mother did it all without a second thought. I could only run and jump to these heights because of her support, her belief, and her love. My mother taught me that there is no limit to what I could do as long as I put in the time and work. I hope I made you proud, mom. I will be the happiest man in

the world if I could be half the parent, scientist, physician, and parent that you are and always will be. Thank you, and I love you.

Contributions

Kwok D, Stevers N, Nejo T, Chen L, Etxeberria I, Jung G, Okada K, Colton Cove M, Lakshmanachetty S, Gallus M, Barpanda A, Hong C, Chan G, Wu S, Ramos E, Yamamichi A, Liu J, Watchmaker P, Ogino H, Saijo A, Du A, Grishanina N, Woo J, Diaz A, Chang S, Phillips J, Wiita A, Klebanoff C, Costello J, Okada H. (2024, in revision). Public tumor-wide splice-derived neoantigens drive anti-tumor T-cell responses. ***Nature***. (Pre-print available on bioRxiv: https://pubmed.ncbi.nlm.nih.gov/37904942/)

Author contributions: D.K. conceived the work and designed experimental setup and data analysis with input from H.Okada., J.C., and C.A.K. D.K and T.N. jointly designed and implemented the RNA-seq analysis pipeline and performed RNA-seq analyses, characterization of TCGA-derived neojunctions, and quantitative alternative splicing analysis. Pan-cancer analyses were the result of discussion among D.K., N.S., and J.C. D.K. analyzed publicly available MS data with guidance from A.B., E.R., and A.W. Spatially mapped samples and corresponding sequencing data were obtain through a collaboration between J.C., S.C., and J.P. D.K. and K.O. performed IVS against NEJ-derived neoantigens with input from P.W., H.Ogino., A.S., and C.A.K. N.S. and C.H. provided help with 10x V(D)J scRNA-seq preparation, and L.C. assisted with scRNA-seq analysis and the identification of neoantigen-reactive TCR clones. D.K. performed all ELISA assays with assistance from G.C., A.D., and N.G. J.J. performed in vitro transcription to generate HLA and neojunction-encoding mRNA with experimental design input from I.E. Electroporation transcription and subsequent COS7:TR Jurkat76 co-cultures were performed by D.K. with input from I.E. Tumor killing assays using xCELLigence was conducted by D.K. with guidance from S.C. and A.Y. All flow cytometry data was obtained by D.K. with experimental setup assistance from A.Y and S.L. Flow cytometry analyses and gating was performed by M.G. and D.K. N.S. designed the primers for qPCR

quantification of neojunction reads and performed siRNA KD experiments against splicing factors. D.K. wrote the manuscript, and all authors provided feedback on manuscript drafts.

Usage in this dissertation: Chapters 1 through 4 have been reproduced and modified from this paper.[1,2] The Methods section of this dissertation is reproduced strictly from the contents of the original paper.

Kwok D, Zhang M, Wang C, Colton Cove M, Stevers N, Lakshmanachetty S, Lerman B, Liu J, Borrman T, Pan Z, Yuen B, Peng S, Nguyen D, Martin M, Hong C, Hilz S, Phillips J, Shai A, Oberheim Bush N, Hervey-Jumper S, McDermott M, Mandl S, Okada H, Costello J. (2024, in preparation). Tumor-wide neoantigen-specific T-cells infiltrating mutant IDH1 low-grade gliomas persist in peripheral blood and allow for personalized TCR-based immunotherapies.

Author contributions: D.K. and M.Z. conceived the work. D.K. designed experiments and data analysis with guidance from H.O. and J.C. M.C. performed transfection and transduction experiments, and N.S. designed constructs for the aforementioned transfections. S.L. assisted with flow cytometry. B.L. designed lineage trees. Members of PACT Pharma identified neoantigen candidates and characterized patient-derived CD8+ T-cell clones that were specific against these targets. J.L assisted with tissue culture experiments. Sample acquisition and procurement was performed by S.H., J.P., A.S., N.O., M.M.

Usage in this dissertation: Portions of Chapter 5 have been written based on the findings from this paper.

Tumor-wide RNA splicing aberrations generate immunogenic public neoantigens

Darwin Kwok

Abstract

T-cell-based immunotherapies hold promise in treating cancer by leveraging the immune system's recognition of cancer-specific antigens.[3] However, their efficacy is often limited in tumors with few somatic mutations and significant intratumoral heterogeneity.[4–7] Here, we introduce a previously uncharacterized class of tumor-wide and public neoantigens originating from RNA-splicing aberrations in diverse cancer types. Notably, we identified T-cell receptor clones capable of recognizing and targeting neoantigens derived from aberrant splicing in *GNAS* and *RPL22*. In multi-site-directed biopsies, we detected the tumor-wide expression of the *GNAS* neojunction within glioma, mesothelioma, prostate cancer, and liver cancer. Importantly, these neoantigens are endogenously generated and presented by tumor cells under physiological conditions and are sufficient to trigger cancer cell eradication by neoantigen-specific $CD8^+$ T-cells. Moreover, our study resolves the complex interplay of dysregulated splicing factor expression in specific cancer subtypes, leading to recurrent patterns of neojunction upregulation. These findings establish a molecular basis for T-cell-based immunotherapies addressing the challenges of intratumoral heterogeneity.

Table of Contents

Prologue

Karkinos, Greek for "crab", was the first instance in which the term for cancer was described in literature. Occurring at approximately 400 BC, Hippocrates described the tumor and its prolific vasculature as similar to a crab with legs spread in the shape of a circle. The mention of cancer has ever since been restricted to specific cases across the years. With the emergence of modern medicine, diseases that have once ravaged human history have been conquered and the average lifespan has more than doubled in the last few centuries. However, with the extension of life comes diseases associated with advanced age. Cancer risk increases with age, and as life expectancy has improved, progressively more individuals live long enough to develop the disease.

The traditional approach for treating solid tumors circles around mass removal of the neoplastic lesion, and the current standard-of-care for most cancer types includes bulk resection followed by a combination of chemotherapy and/or radiotherapy. However, cancer is a disease caused by the unrestricted division of a patient's own cells, and missing even a single tumor cell can lead to the repopulation of a more aggressive and therapeutically-resistant tumor. Over the last decade, research involving the manipulation of the immune system to recognize and eradicate tumor cells has surfaced as a promising new approach. Immune cells can mount a cytotoxic response by recognizing non-self-antigens, and in theory, the multitude of mutations generated by uncontrolled tumor cell proliferation would lead to the existence of tumor-specific antigens (TSAs).

Approaches that leveraged the immune system were known as "immunotherapies", and one particular class of immunotherapies involved directing immune cells to target TSAs that were only found in cancer cells. While conceptually sound, the expression of TSAs is hardly consistent across the entire tumor landscape. As a tumor uncontrollably grows, the accumulation and type of mutations often diverge throughout various cellular subpopulations. TSAs have largely been identified through the

analysis of single biopsies from tumors, and as a result, targeting these antigens often results in the selection for neoantigen-null tumor cells and recurrence of an immunotherapy-resistant tumor. As such, this level of intratumoral heterogeneity needs to be considered when designing immunotherapeutic approaches that target specific TSAs.

Chapter 1: Characterizing tumor-wide neojunctions

The identification of TSAs across cancers has largely been focused on characterizing peptides arising from somatic mutations within the coding regions of the genome.[8,9] While nonsynonymous mutations are prevalent in high mutational burden cancers such as melanoma and lung cancer, tumors with less genomic aberrations yield fewer downstream TSA targets.[10,11] To expand the repertoire of potential immunotherapeutic antigens, recent studies have explored aberrant splicing events as a source of TSAs.[12,13] These cancer-specific splicing events are known as neojunctions, and they are shown to be prevalent and capable of generating TSAs that are capable of potentiating CD8+ T-cell responses.[13–20] However, no study has yet examined the spatial and temporal conservation of neojunctions expressed throughout the entire tumor landscape. The question of whether neojunction-derived targets are clonally expressed remains unknown.

Characterization of public, pan-cancer neojunctions

In our study, we aimed to investigate the clonality of neojunctions across various cancer types to identify public tumor-wide neojunction-derived tumor-specific antigens (TSAs). Public targets are those expressed across multiple patients within the same cancer cohort and targeting these would be the optimal approach for developing "off-the-shelf" therapies. We first utilized RNA-sequencing (RNA-seq) data from The Cancer Genome Atlas (TCGA) to identify non-annotated junction reads across various cancer types **(Extended Data Figure 1A)**. We included 12 tumor types with publicly available, multiple spatially-mapped tumor samples to assess intratumoral heterogeneity (ITH) **(Figure 1A)**. Only samples with inferred tumor purities of ≥ 60% were included in our analysis[21,22] **(Figure 1B)** to ensure that the identified protein-coding, non-annotated junctions are predominantly captured from tumor cells **(Extended Data Figure 1B).** The positive sample rate (PSR) of a junction represents the percentage of samples within a cohort expressing the neojunction at a junction read frequency of ≥ 1% compared to the canonical splicing junction.[23] This metric allowed us to define

public neojunctions as those demonstrating elevated PSRs within each TCGA tumor cohort (PSR_{TCGA} ≥ 10%) **(Figure 1C, Extended Data Figure 1C).** Following neojunction nomenclature[13], we selected cancer-specific splicing events as those with a PSR < 1% in normal tissue from the Genotype-Tissue Expression (GTEx) project (n=9166; PSR_{GTEx} < 1%) **(Extended Data Figure 1D).** We identified an average of 94 public neojunctions per TCGA tumor type **(Figure 1D)**, which were expressed at similar frequencies across all cancers **(Figure 1E)**. Further characterization of public neojunctions across tumor types revealed varying distributions of splice types **(Figure 1F)** and consistent proportions of frame shift-generating splicing events **(Figure 1G)**. Some of these neojunctions have been detected in recent splicing studies[19,24] **(Extended Data Figure 1E-1F)**. Unbiased hierarchical clustering revealed that cases from the same tumor type tend to cluster together, indicating similar neojunction expression profiles across patients with the same cancer type. A subset of neojunctions were expressed in multiple tumor types **(Figure 1H)**. Therefore, neojunction expression can be conserved among patients across multiple cancer types, suggesting potential pan-cancer immunotherapy targets produced by aberrant splicing.

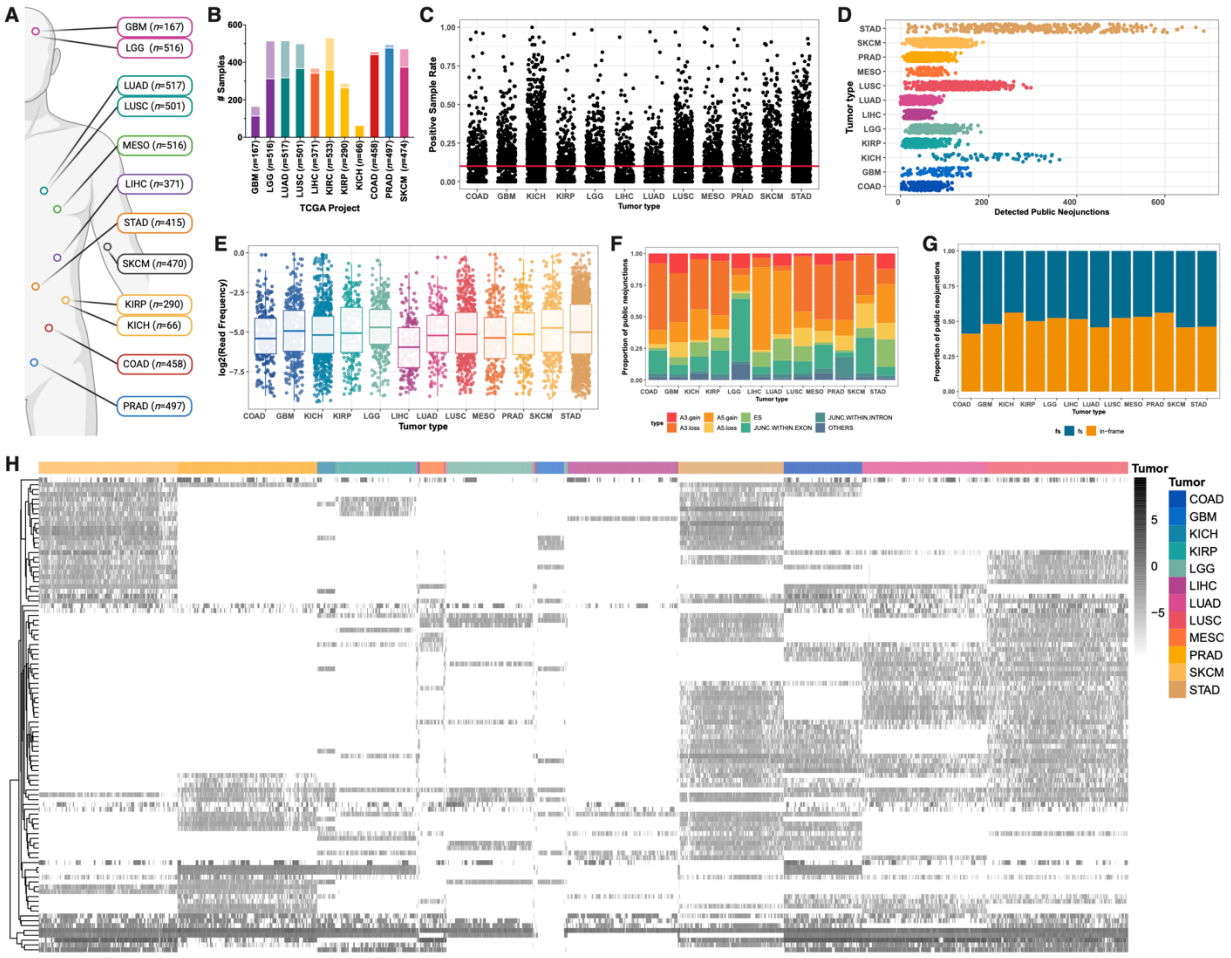
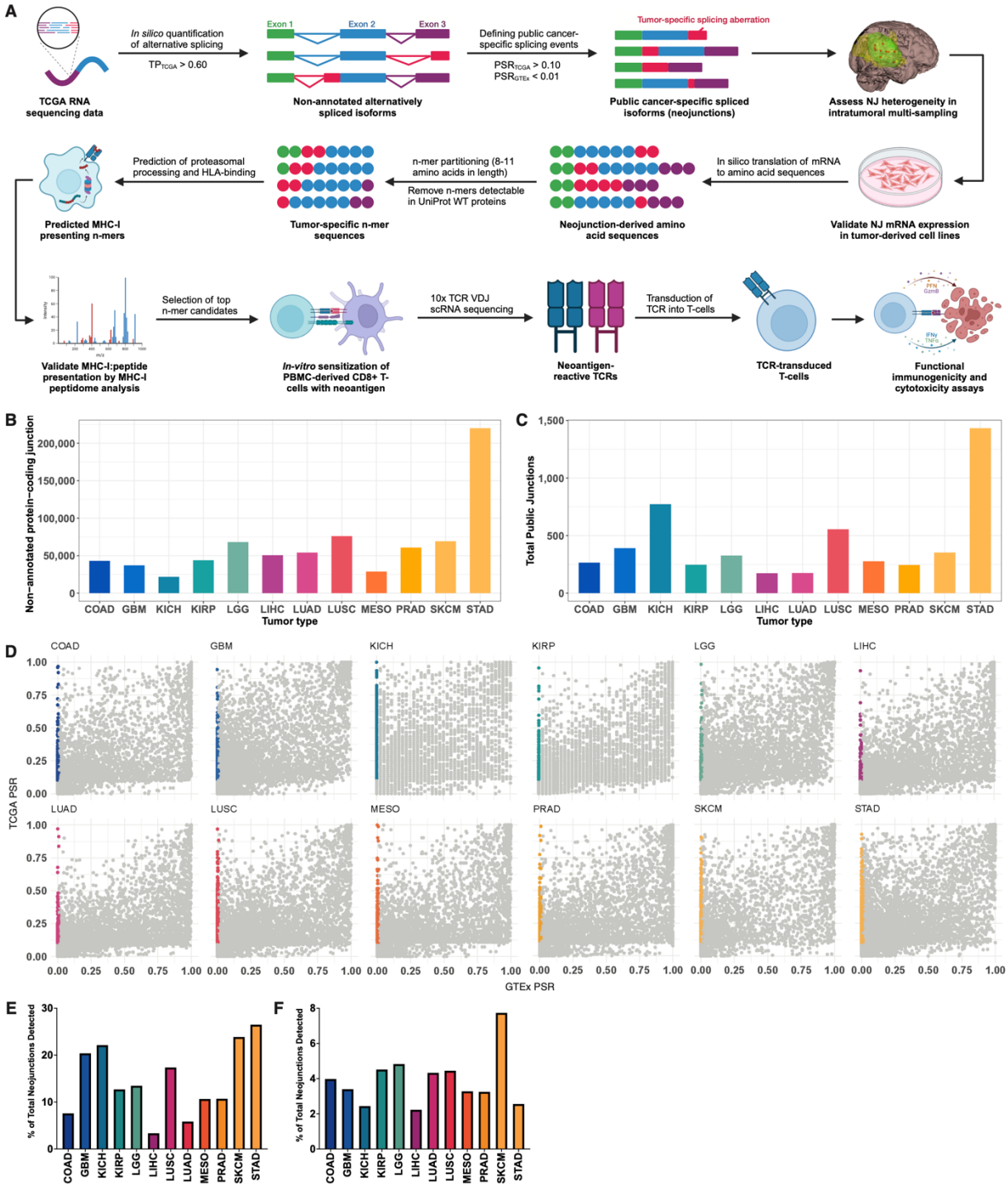


Figure 1: Characterization of public neojunctions across multiple cancer types. A. TCGA tumor sets with corresponding multi-biopsy RNA sequencing data available for analysis. Disease types were selected based on the availability of publicly available data sets that included multi-site sampling within the same tumor, which included glioblastoma (GBM; n=167), low-grade glioma (LGG; n=516), lung adenocarcinoma (LUAD, n=517), lung squamous cell carcinoma (LUSC, n=501), mesothelioma (MESO, n=516), liver hepatocellular carcinoma (LIHC, n=371), stomach adenocarcinoma (STAD, n=415), kidney renal clear cell carcinoma (KIRC; n=533), kidney renal papillary cell carcinoma (KIRP; n=290), kidney chromophobe (KICH, n=66), colon adenocarcinoma (COAD; n=458), and prostate adenocarcinoma (PRAD; n=497). **B**. Tumor purity of TCGA tumor samples. Samples with confirmed tumor purity of 60% and above (solid fraction) were retained for downstream analysis. **C.** Interpatient frequency (positive sample rate; PSR) of putative neojunctions identified in each tumor type. Public neojunctions are defined as those with a PSR ≥ 10% (red line). **D.** Total public neojunctions detected per sample across tumor types. **E.** Log_2 (read frequency) of public neojunctions across tumor types. **F-G.** Distribution of public neojunctions based on splice types: exonic loss due at the 3' or 5' splice site (A3 or A5 loss), intronic gain at the 3' or 5' splice site (A3 or A5 gain), exon skip ES), junction within exon, junction within intron, others. (Figure caption continued on the next page.)

(Figure caption continued from the previous page.) (**F**) and frame-shift status (**G**). **H.** Expression of all pan-cancer-spanning neojunctions (log_2(CPM)) across all studied TCGA tumor types



Extended Data Figure 1: Pan-cancer public neojunctions are characterized from TCGA A. TCGA RNA sequencing data across multiple cancers (*n*=12) were analyzed for non-annotated, protein-coding, and cancer-specific splicing junctions (GTEx positive sample rate < 1%; neojunctions). Interpatiently conserved (TCGA positive (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) sample rate ≥ 10%; public neojunctions) were retained for downstream analysis of intratumoral heterogeneity (ITH). Tumors with sequencing data extracted from multiple intratumoral regions were used to evaluate each public neojunction's ITH. Independent prediction algorithms were used to assess proteasomal processing and MHC-I binding of peptide sequences translated from public, intratumorally conserved neojunctions. The expression of these neojunctions and their peptide derivatives were validated by RNA sequencing and mass spectrometry analysis of patient-derived tumor samples and cell lines. T-cell receptors (TCRs) were cloned and characterized for top predicted candidates through *in vitro* sensitization of PBMC-derived $CD8^+$ T-cells against the corresponding neoantigen-pulsed antigen presenting cells and subsequent 10x V(D)J single-cell sequencing. Transduction of these neoantigen-reactive TCR sequences in TCR-null Jurkat76/CD8 cells and PBMC-derived $CD8^+$ T-cells allowed the demonstration of neoantigen-specific immunogenicity and tumor-specific killing. **C.** Total number of public ($PSR_{TCGA} \geq 10\%$), non-annotated, protein-coding junctions detected pan-cancer. **D.** Dot plots representing the positive sample rate percentage of non-annotated, protein-coding junctions in all studied cancer types (COAD, GBM, KICH, KIRP, LGG, LIHC, LUAD, LUSC, MESO, PRAD, SKCM, STAD). Neojunctions ($PSR_{TCGA} \geq 10\%$ and $PSR_{GTEx} < 1\%$) are denoted by colored dots. **E-F.** Bar plots illustrating the proportion of SNIPP-characterized neojunctions found in the IRIS **(E)** and MAJIQlopedia **(F)** databases.

Characterization of public, pan-cancer neojunctions

ITH presents a significant challenge in cancer treatment, particularly for immunotherapy approaches. Tumors often exhibit diverse genetic and molecular profiles across different regions, complicating the development of effective targeted therapies. This heterogeneity can lead to the outgrowth of cancer cells lacking specific (TSAs), potentially rendering single-target immunotherapies ineffective. For targeted T-cell therapies, this underscores the importance of focusing on multiple neoantigens shared by the entire tumor to prevent immune evasion by antigenic evolution.[3,25] Neojunctions have previously been demonstrated to generate immunogenic antigens.[16–19] To address this challenge, we conducted a comprehensive analysis of public neojunctions across multiple intratumoral samples from various cancer types, including prostate,[26] liver,[27–30] colon,[27,31] stomach,[27] kidney,[27] and lung cancers[32,33] **(Extended Data Figure 2A)**. We filtered for neojunction reads across multiple samples from the same tumor **(Figure 2A)** and investigated whether public neojunctions are present throughout the entirety of the tumor landscape. This analysis revealed public neojunctions expressed in multiple intratumoral samples **(Figure 2B, Extended Data Figure 2B)**. Public spatially-conserved neojunctions were identified across all cases in datasets with multi-site sampling, further highlighting their potential as ideal neojunction targets **(Figure 2C)**.

Among the cancer types analyzed in our study, gliomas exhibit a notably high degree of ITH, complicating immunotherapy efforts.[34–37] Determining the number of samples required to represent a tumor's transcriptomic landscape accurately is crucial.[34,38] Therefore, we increased the number of intratumoral biopsies across the three main glioma subtypes.[39–41] We evaluated approximately ten maximally distanced, spatially mapped samples from each of 51 tumors with exome and RNA-seq data **(Figure 2D, Extended Data Figure 2D-H)**, detecting neojunction expression intratumorally across multiple patients. As the number of samples increased from one to ten, the number of ubiquitously expressed neojunctions inversely correlated with the number of samples **(Extended**

Data Figure 2F-H). These findings underscore the critical need for multiple biopsies per tumor to confidently characterize neojunctions as tumor-wide.

Hierarchical clustering of our extensive intratumoral dataset revealed neojunction subsets associated with either mutant isocitrate dehydrogenase (IDHmut) or wild-type (IDHwt) subtypes **(Figure 2E)**. Notably, IDHmut gliomas expressed significantly more tumor-wide neojunctions than IDHwt gliomas. Although the proportion of tumor-wide neojunctions is significantly lower than sub-clonally expressed neojunctions across patients **(Figure 2F)**, we detected at least one tumor-wide neojunction in 45 (88.2%) of our patients **(Figure 2G)**, with 13 (25.5%) patients expressing over 50 tumor-wide neojunctions **(Extended Data Figure 2D)**. While 774 (98.1%) of the TCGA-characterized LGG and GBM neojunctions were detectable in more than one region of at least one tumor in our cohort, 37 (4.7%) neojunctions were found across all samples in more than 10% of the study cohort **(Extended Data Figure 2E)**. Therefore, the majority of the public neojunctions identified from the TCGA LGG/GBM analysis are expressed in multiple tumor regions but not tumor-wide in our multi-sampled tumor dataset. This suggests that combining neojunctions could effectively cover an entire tumor landscape.

Next, we characterized spatially and temporally conserved neojunctions at metastasis and recurrence. Analysis of public cutaneous melanoma (SKCM) RNA-seq data[42] revealed 13 (9.6%) neojunctions detected across metastatic sites in at least one patient **(Extended Data Figure 2C)**. Examining matched primary/metastasis pairs in TCGA revealed that 43.8% to 72.6% of neojunctions identified in primary tumors persist in metastases across COAD, PRAD, and SKCM cancers **(Extended Data Figure 2I)**. Similarly, an average of 36.4% of neojunctions were conserved at recurrence when investigating matched primary/recurrence pairs across TCGA COAD, GBM, LGG, LIHC, and LUAD cancers **(Extended Data Figure 2J)**. In our glioma dataset, 79.2% and 82.3% of neojunctions

were conserved at recurrence following temozolomide treatment **(Extended Data Figure 2K)**. These findings demonstrate that neojunctions can persist across both space and time.

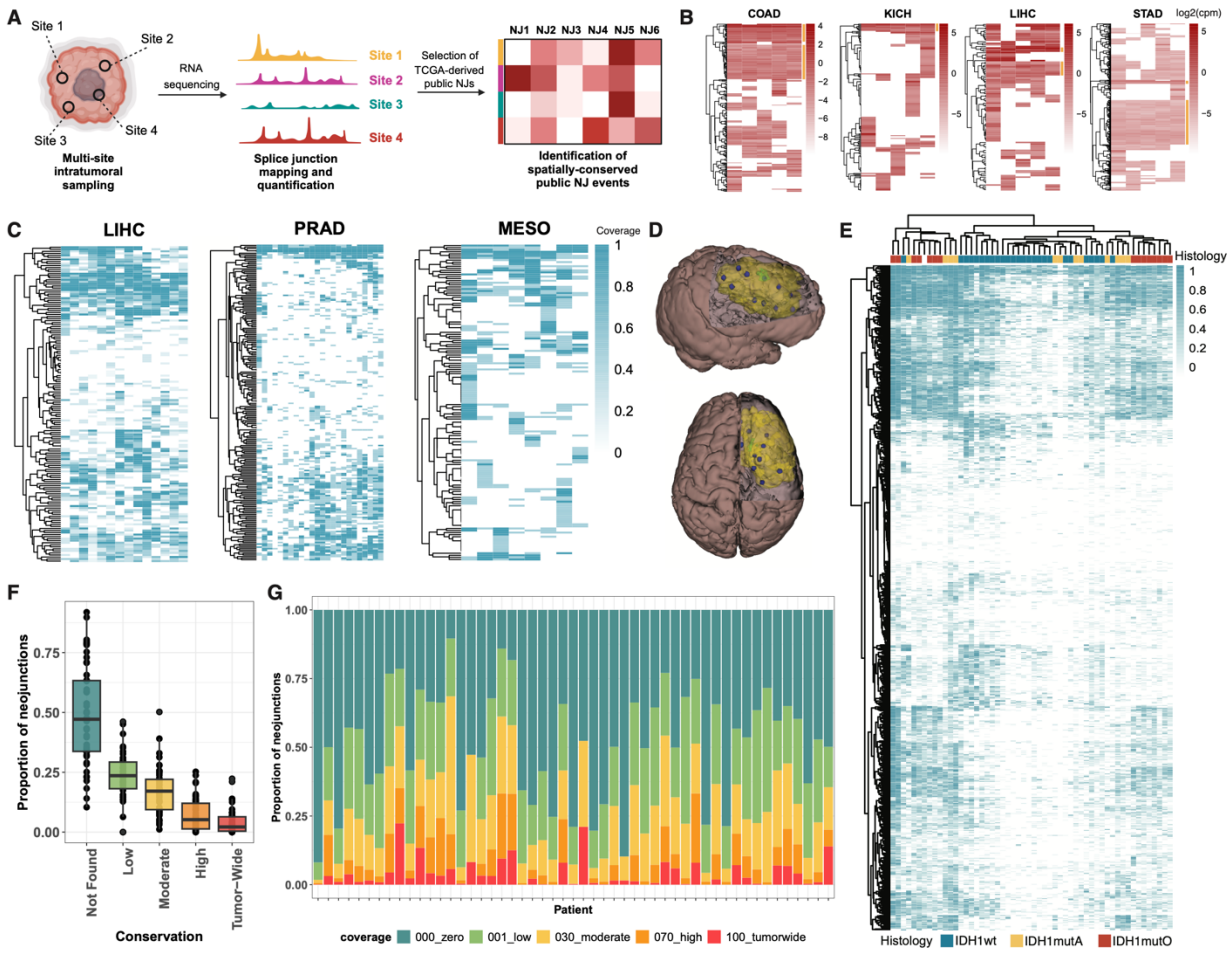
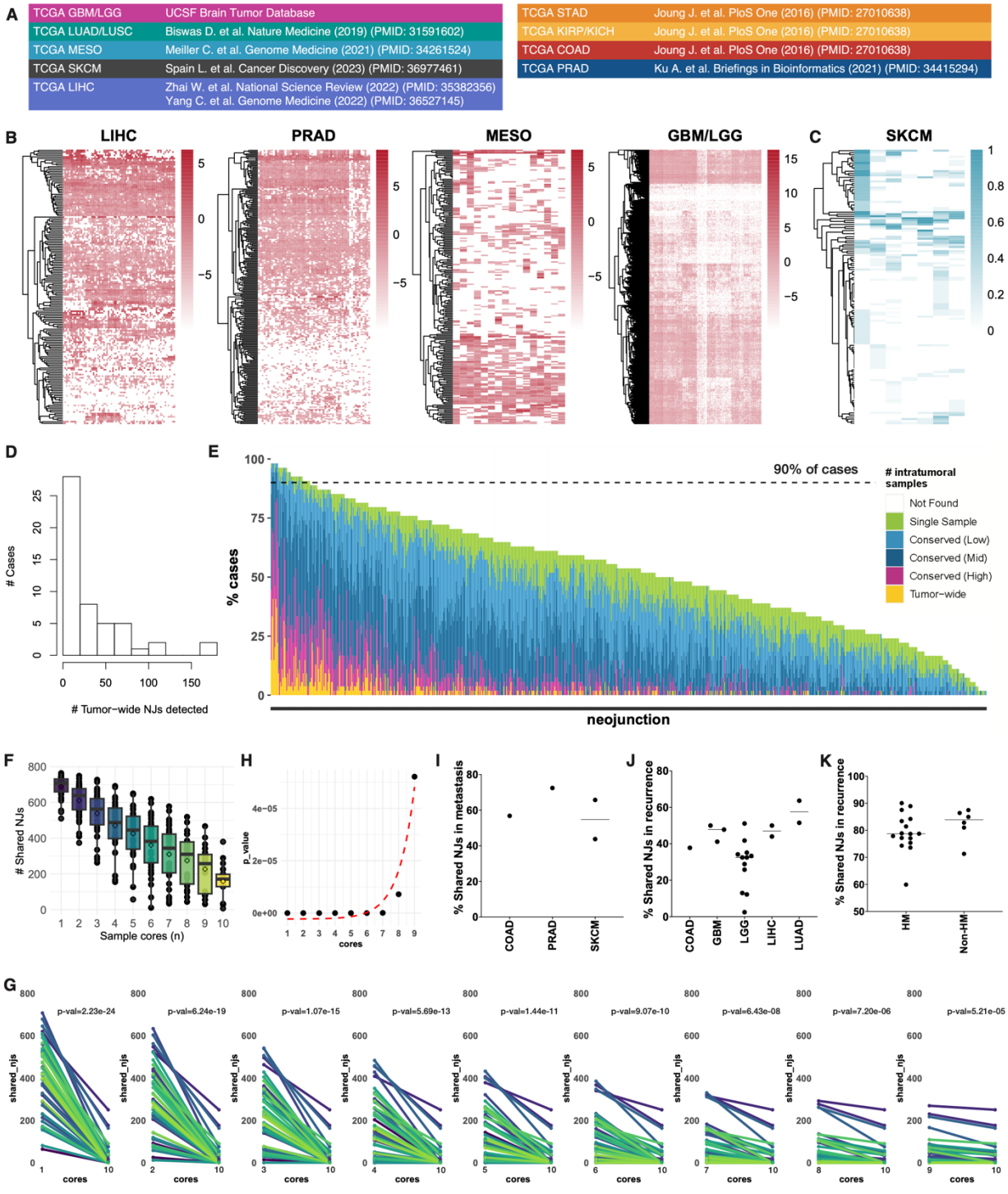


Figure 2: A subset of neojunctions are expressed tumor-wide. A. Overview of the tumor-wide characterization of neojunctions by investigating RNA sequencing of multiple intratumoral regions in various cancer types. **B.** Heatmaps representing $\log_2$(CPM) of neojunctions (rows) across five samples within the same tumor (columns) in colon adenocarcinoma (COAD), kidney chromophobe (KICH), liver cancer (LIHC), and stomach adenocarcinoma (STAD). Neojunctions found across all five intratumoral samples are annotated in yellow. **C.** Heatmap illustrating the number of biopsies within the same tumor sample (columns) that has detectable expression of neojunctions (rows) in LIHC (left), prostate adenocarcinoma (PRAD; center), and mesothelioma (MESO; right). The intensity of each cell indicates the proportion of regions within the same tumor that has putative expression of each neojunction, with intensity of 1 representing a neojunction expressed tumor-wide within a corresponding patient. **D.** 3-D models of the brain and tumor (yellow) derived from patient 470 (P470). Approximately 10 spatially mapped and maximally distanced biopsies (blue) were taken within each tumor (refer to Supplementary Video 1). Whole-exome sequencing, RNA sequencing, and further analyses were conducted on each of these regions. **E.** Heatmap illustrating the number of biopsies within the same tumor sample (columns) that has detectable expression of neojunctions (rows) in across various glioma subtypes, *IDH1*wt (blue), (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) *IDH1*mut-A (yellow), and *IDH1*mut-O (red). The intensity of each cell indicates the percentage of regions within the same tumor that has putative expression of each neojunction. **F-G.** Bar plot **(F)** and parts-of-whole chart **(G)** of proportion of glioma-specific neojunctions (*n*=789) based on the their intratumoral heterogeneity across patients. Neojunctions are considered tumor-wide if they are found across 100% of intratumor samples within the same patient, highly conserved if they are found in > 70% of samples but not 100% of samples, moderately conserved if they are found in > 30% of samples but ≤ 70% of samples, or lowly conserved if found in at least two samples but ≤ 30% of samples.



Extended Data Figure 2: Intratumoral heterogeneity and interpatient characteristics of neojunctions across various cancer types. A. Multi-region RNA-sequencing data of multiple cancer types were collected across various studies. Multi-region sampling is defined in studies in which multiple biopsies were isolated from same tumor (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) for downstream sequencing analyses. **B.** Counts per million (CPM) of non-annotated, protein-coding neojunctions across multi-region samples in kidney cancer**,** prostate cancer, mesothelioma, and glioma cases. **C.** Heatmap illustrating the number of metastases within the a SKCM patient (columns) that has detectable expression of neojunctions (rows). The intensity of each cell indicates the proportion of regions within the same tumor that has putative expression of each neojunction, with intensity of 1 representing a neojunction expressed in all metastases within a corresponding patient. **D.** Histogram of the number of multi-region sampled glioma cases with the corresponding number of tumor-wide neojunctions. **E.** Distribution of glioma-specific neojunctions (n=789, columns) based on the their intratumoral heterogeneity across patients. **F.** Total number of neojunctions found in n cores per patient. **G.** Slope charts demonstrating patient-matched pairs of the number of neojunctions found in n cores compared to 10 cores. Paired t-test analysis was performed on all matched values, and the corresponding p-value is displayed above each slope chart iteration. **H.** Dot-plot with best-fitting curve mapping the p-values of all iterations of paired n core and 10 core comparisons. **I-K.** Percentage of neojunctions identified at primary that were conserved in paired metastases **(I)**, recurrence **(J)**, and recurrence following temozolomide treatment **(K)**.

Conclusion

By leveraging RNA-seq data from The Cancer Genome Atlas (TCGA), we identified non-annotated junction reads across multiple cancer types and revealed an average of 94 public neojunctions per TCGA tumor type, with consistent expression patterns and similar frequencies across different cancers. The study highlighted the significance of public neojunctions, which demonstrated elevated PSRs within each TCGA tumor cohort and were predominantly cancer-specific, with minimal expression in normal tissues. By investigating multiple intratumoral samples from diverse cancer types, we uncovered that public neojunctions were often spatially conserved within tumors, underscoring their potential as ideal immunotherapy targets.

Gliomas, known for their pronounced ITH, particularly benefitted from our extensive in-house sampling approach. Our findings emphasized the necessity of multiple biopsies per tumor to accurately capture the transcriptomic landscape and identify tumor-wide neojunctions. Hierarchical clustering revealed that certain neojunction expression profiles were associated with either IDHmut or wild-type subtypes, with IDHmut gliomas expressing significantly more tumor-wide neojunctions. Furthermore, our analysis extended to spatially and temporally conserved neojunctions at metastasis and recurrence. We demonstrated that a significant proportion of neojunctions identified in primary tumors persist in metastatic and recurrent sites across various cancer types.

In conclusion, our study underscores the potential of targeting public neojunction-derived TSAs for developing pan-cancer immunotherapies. Tumor-wide or highly conserved neojunctions identified through spatial analysis of multiple regions within the same tumor helps identify targets that are not limited to select subclones or regions of a tumor. By addressing the challenges posed by intratumoral heterogeneity and focusing on conserved neojunctions, we pave the way for more effective and broadly applicable cancer treatments. These findings lay a strong foundation for the downstream

identification of neojunction-derived TSAs and the eventual development of innovative immunotherapeutic strategies aimed at combating a wide range of cancers.

Chapter 2: Cancer-specific splicing dysregulation

Dysregulated splicing in cancer often results from a complex interplay of factors, including mutations in splicing factor genes, alterations in regulatory genes that influence splicing, and cis-acting mutations affecting splice sites or regulatory elements. Splicing factor mutations are common in various cancers, particularly hematologic malignancies.[43] The most frequently mutated splicing factors include *SF3B1*, *SRSF2*, *U2AF1*, and *ZRSR2*, and these mutations can lead to altered splicing patterns in cancer cells that could potentially generate novel protein isoforms.[44–46] In *SF3B1*-mutated uveal melanoma cases, one particular study demonstrated the generation of neojunction-derived neoantigens that could be targeted in cell-based immunotherapy approaches.[16]

However, many cancer types do not carry splicing factor mutations, which raises the question of why they express significantly more neojunctions than healthy tissue[13] and even differ in expression between subcategories of the same cancer. This observation suggests that other mechanisms, such as aberrant expression of splicing factors or alterations in upstream regulatory pathways, may contribute to dysregulated splicing in these cases. Additionally, epigenetic changes, including altered DNA methylation patterns and histone modifications, can influence splicing factor expression and activity, potentially leading to widespread changes in splicing patterns across the transcriptome.

Our study aims to elucidate the underlying mechanisms driving the increased expression of neojunctions in cancer types independent of splicing factor mutations. By investigating the differential expression of canonical splicing-related genes, we seek to uncover novel insights into the complex landscape of splicing dysregulation in cancer. Understanding these mechanisms is crucial for developing targeted therapies and identifying potential biomarkers for cancer diagnosis and prognosis.

Neojunction expression is not associated with splicing factor mutations in gliomas

Noting the subtype-specific expression of neojunctions across glioma subtypes in both TCGA and our spatially-mapped dataset **(Figure 2E)**, we investigated dysregulation in splicing machinery that may contribute to these patterns. Prior investigations report a potential role of the *IDH* mutation in splicing aberrations,[13] but our study revealed much more complexity. Both TCGA and our spatially mapped datasets showed significantly greater numbers of public neojunctions per case in *IDH* mutant (*IDH*mut) gliomas compared to *IDH* wild-type (*IDH*wt) gliomas **(Figures 3A-3B)**. While *IDH*mut-A gliomas demonstrated higher average levels of neojunction expression than *IDH*wt gliomas, *IDH*mut-O neojunction expression was even greater **(Figures 3C-3D)**.

We performed pairwise Pearson correlation analyses to explore whether neojunction expression is associated with somatic mutations in commonly mutated RNA splicing factors[47–50] **(Extended Data Figures 3A-3C)**. A high correlation was found between *FUBP1*, *SF3A1*, or *NIPBL* mutations and the *IDH* mutation. *FUBP1* mutations are prevalent in *IDH*mut-O gliomas,[51,52] yet hierarchical clustering of neojunctions revealed no significant trend of neojunction expression with *FUBP1*, *SF3A1*, or *NIPBL* mutation status **(Extended Data Figures 3D-3I)**. Dysregulation of individual splicing factors can result in aberrant splicing.[48,53] Based on these findings, we investigated whether aberrations in the expression levels of splicing-related genes correlate with neojunction production.

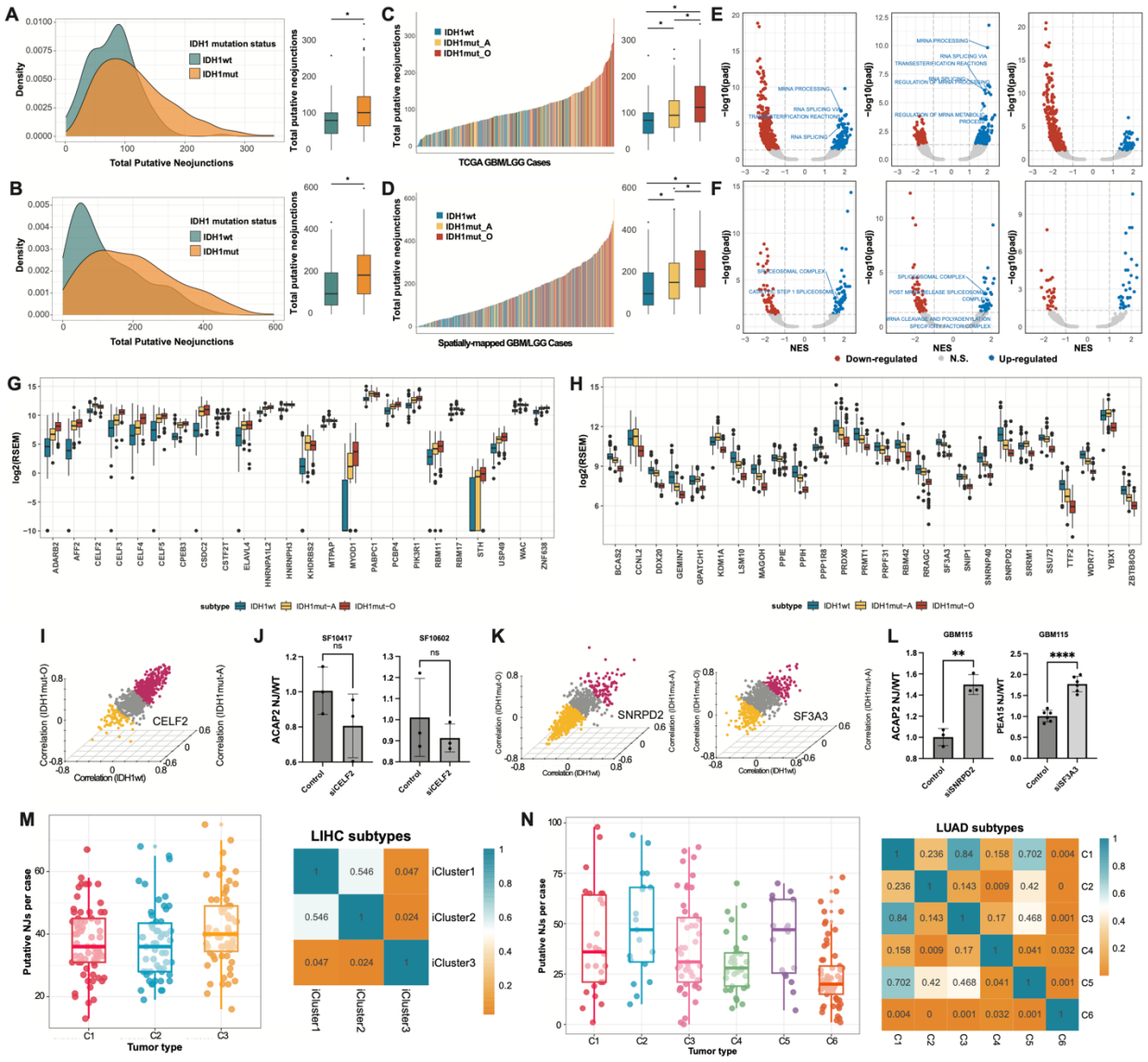


Figure 3: Tumor-specific disease subtypes demonstrate differential levels of neojunction expression. A-B. Density and box-and-whisker plots depict the total number of putative neojunctions expressed in *IDH1*mut cases (orange) and *IDH1*wt cases (green) in TCGA GBM/LGG samples (**A**) and our in-house spatially-mapped GBM/LGG dataset (**B**). **C-D.** Histogram and box-and-whisker plot depict the number total number of putative neojunctions expressed in *IDH1*wt (blue), astrocytoma (yellow), and oligodendroglioma cases (red) in TCGA GBM/LGG samples (**C**) and our in-house spatially mapped GBM/LGG dataset (**D**). **E-F.** Volcano plots illustrating significantly upregulated (blue) and downregulated (red) gene sets comparing *IDH1*mutO cases vs. *IDH1*wt cases (left), *IDH1*mutA cases vs. *IDH1*wt cases (center), and *IDH1*mutO cases vs. *IDH1*mutA cases (right). Gene sets categorized under the Gene Ontology Biological Processes (GOBP, **E**) and Gene Ontology Cellular Component (GOCC, **F**) were investigated in our(Figure captions continued on the next page.)

(Figure captions continued from the previous page.) analyses. Splicing-related gene sets were denoted by points labeled with text. **G.** Box-and-whisker plot depicting the $\log_2$(RSEM) expression level of splicing-related genes detected from GOBP gene set with a significant ($p < 0.05$) $\log_2$fold increase in expression of 1.5 between *IDH1*mutA (yellow) and *IDH1*mutO (red) cases when compared to *IDH1*wt cases (blue). **H.** Box-and-whisker plot depicting the $\log_2$(RSEM) expression level of chromosome 1p or chromosome 19q splicing-related genes detected from GOBP gene sets with a significant ($p < 0.05$) $\log_2$fold decrease in expression of 1.5 between *IDH1*mutO (red) when compared to *IDH1*mutA (yellow) and *IDH1*wt cases (blue). **I.** Pearson correlation analyses of glioma-specific neojunctions against the expression of *CELF2* in *IDH1*mutO (z-axis), *IDH1*mutA (y-axis), and *IDH1*wt (x-axis) cases. Neojunctions with a Pearson correlation greater than or equal to 0.10 with the corresponding gene are denoted with purple dots and those with a Pearson correlation less than or equal to -0.10 with the corresponding gene are denoted with yellow dots. **J.** Expression of NJ_{ACAP2} in LGG cell lines, SF10417 (left) and SF10602 (right), treated with control siRNA or si*CELF2*.**K.** Pearson correlation analyses of glioma-specific neojunctions against the expression of *SNRPD2* (left) and *SF3A3* (right)in *IDH1*mutO (z-axis), *IDH1*mutA (y-axis), and *IDH1*wt (x-axis) cases. **L.** Expression of NJ_{ACAP2} in GBM115 treated with control siRNA or si*SNRPD2* (left) or si*SF3A3* (right). **M-N.** Bar plot (left) showing the total neojunctions expressed per case across all disease subtypes and heatmap (right) displaying the Wilcoxon rank-sum test of neojunction expression between each subtype within TCGA LIHC **(M)** and LUAD **(N)**.

IDH-mut and Chr 1p/19 splicing gene dysregulation modulates neojunction expression

To investigate possible drivers for the glioma-subtype differences in neojunction expression, we evaluated differentially expressed gene sets on the three glioma subtypes from TCGA **(Extended Data Figures 4A-4B)**. Gene set enrichment analysis highlighted significantly upregulated splicing-related gene sets in mutant IDH cases compared to their wild-type counterparts across Gene Ontology Biological Processes (GOBP) **(Figure 3E)** and Gene Ontology Cellular Component (GOCC) databases **(Figure 3F)**. When ordered based on neojunction expression, splicing-related genes highly expressed in both mutant *IDH* tumor subtypes largely clustered together **(Extended Data Figures 4C-E)**.

To further investigate specific splicing-related genes that may lead to increased neojunction expression in *IDH*mut gliomas **(Figures 3G-H)**, we selected GOBP splicing-related genes (n=24) with a statistically significant ($p < 0.05$) 1.5-fold increase in expression in *IDH*mut cases compared to wild-type **(Figure 3G)**. Notably, *CELF2* has previously been reported to generate splice aberrations when overexpressed.[54,55] Our correlation analyses of the expression of *CELF2* against the expression of all 789 public neojunctions identified a greater percentage of neojunctions whose expression generally increased (average Pearson correlation coefficient > 0.10) with the increasing CELF2 expression across all glioma subtypes **(Figure 3I)**. Of the 789 neojunctions, 359 (45.5%) increased in expression with *CELF2* expression, as opposed to 81 (10.3%) neojunctions that tended to decrease. Selecting the highest correlating neojunction associated with this splicing-related gene, we performed siRNA-mediated knockdown of *CELF2* in two patient-derived *IDH* mutant cell lines, SF10417 and SF10602[56] **(Extended Data Figures 5A-5B)**, and observed a trend of decreased expression of the associated neojunction across both lines **(Figure 3J)**. We characterized 244 neojunctions that are significantly upregulated in *IDH* mutant compared with wild-type glioma cases (log2 fold change > 1.5, p-value <

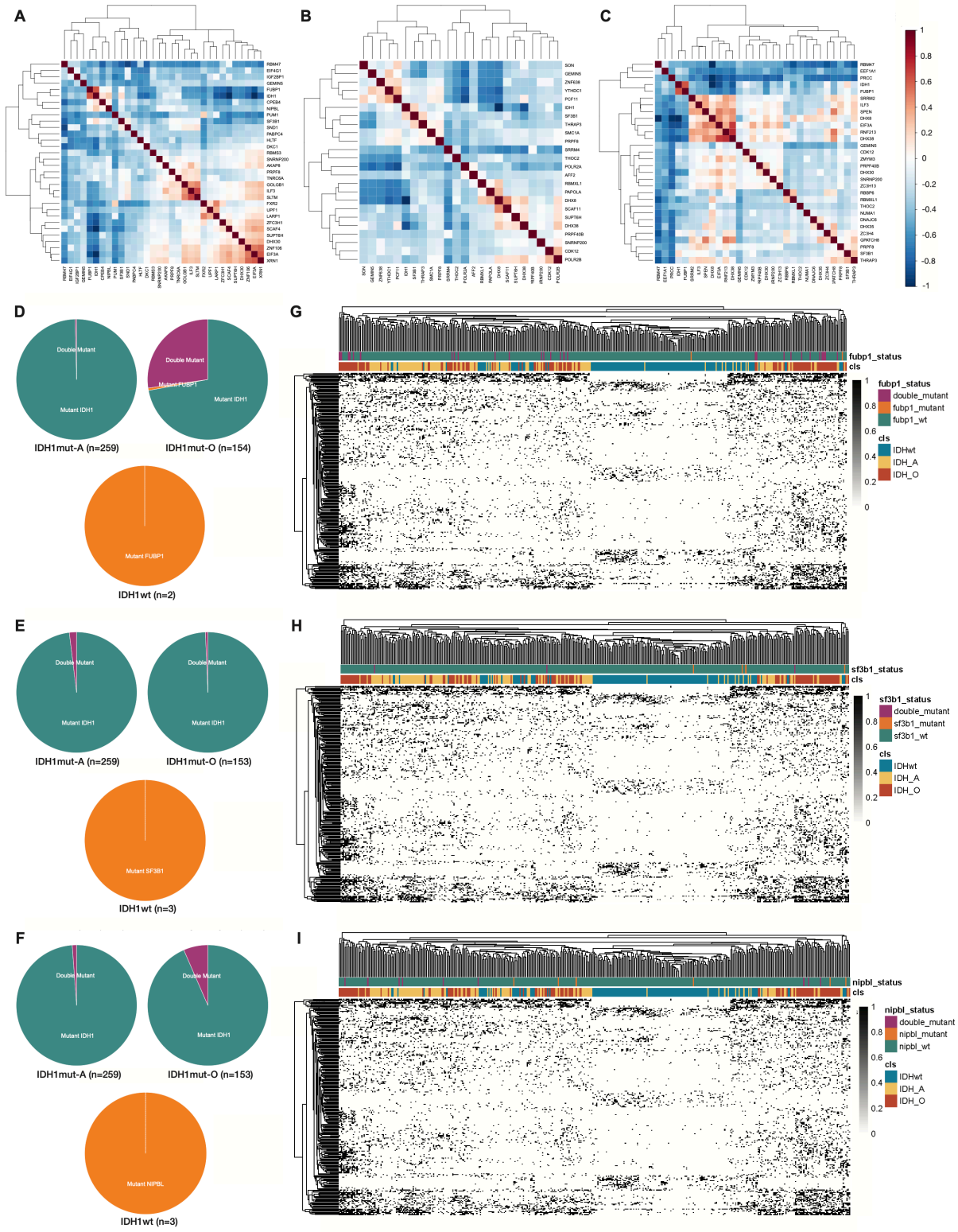
0.05), a subset of which were detected in other TCGA *IDH*mut cancer types **(Extended Data Figure 5K)**. RNA-sequencing analysis demonstrated an associated decrease in the expression of 19 (8.6%) and 28 (12.7%) *IDH*mut-associated neojunctions, respectively, in SF10417 and SF10602 following *CELF2* knockdown compared to non-treated controls **(Extended Data Figure 5C)** and a correlative increase in a candidate IDHmut neojunction expression with increased expression of IDHmut-associated splicing-related genes **(Extended Data Figure 5J)**. These findings suggest that neojunction prevalence is regulated by altered expression of RNA-binding proteins in tumor subtypes and that modulation of these genes can lead to changes in neojunction levels.

When re-examining GOBP splicing-related gene sets **(Extended Data Figures 4C-4E)**, we discovered subclusters of genes significantly downregulated in *IDH*mut-O cases. Most of the genes found within these clusters reside on either chromosome 1p or 19q, co-deletion of which is a distinctive diagnostic feature of *IDH*mut-O gliomas. To evaluate whether the downregulation of these genes could lead to the characteristic increase in putative neojunction expression seen in the *IDH*mut-O subtype, we selected GOBP splicing-related genes (n=26) with a statistically significant ($p < 0.05$) 1.5-fold decrease in expression in *IDH*mut-O cases compared to both *IDH*mut-A and *IDH*wt cases **(Figure 3H)**. Of these splicing genes, disruption of normal *SNRPD2* and *SF3A3* expression was previously reported to lead to splicing aberrations.[57] Correlation analysis of *SNRPD2* and *SF3A3* expression against the expression of the 789 neojunctions across all glioma subtypes supported our hypothesis that decreased *SNRPD2* and *SF3A3* expression might contribute to greater neojunction expression **(Figure 3K)**. Of the 789 neojunctions, 385 (48.8%) increased in expression with decreased *SNRPD2* expression compared with the 93 (11.8%) neojunctions that tended to increase in expression with increasing levels of *CELF2*. Similarly, with increasing levels of *SF3A3* expression, 178 (22.6%) neojunctions tended to increase in expression, and 127 (16.1%) neojunctions tended to decrease in expression. Notably, siRNA knockdown of either *SNRPD2* or *SF3A3* in the GBM115 cell line

(Extended Data Figures 5A-5B) led to a significant increase in the expression levels of their associated neojunctions **(Figure 3L, Extended Data Figure 5E)**. We also characterized 52 *IDH*mut-O-associated neojunctions as significantly upregulated in mutant *IDH*-O glioma cases compared to their mutant *IDH*-A and wild-type counterparts (log2 fold change > 1.5, p-value < 0.05). We found increased expression of 7 (13.5%) and 4 (7.7%) *IDH*mut-O-associated neojunctions in GBM115 cells treated with *SF3A3* or *SNRPD2* siRNA, respectively **(Extended Data Figure 5D)**. While previous studies linked splicing factor mutations to neojunctions in cancers,[58] our results shed light on a previously undescribed mechanism showcasing that decreased wild-type splicing factor expression consistently generates neojunctions. These findings suggest that commonly altered components of the RNA splicing machinery in gliomas are mechanistically linked to increased neojunction expression.

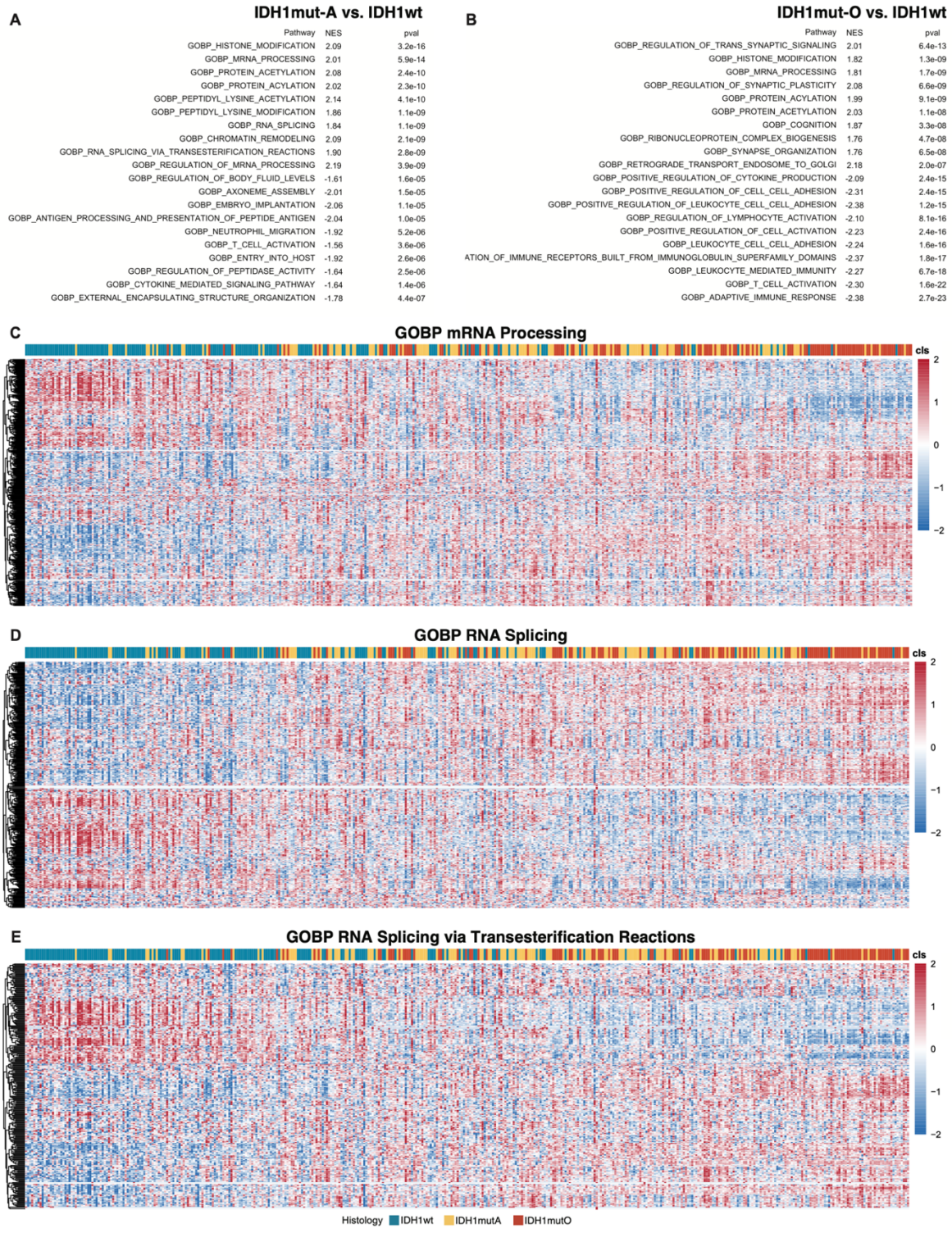
Dysregulation of splicing-related gene expression leads to differential neojunction expression

Finally, we extended our analysis across the remaining TCGA cancer types used in this study to identify tumor subtypes with significantly dysregulated neojunction expression. While neojunction expression remained relatively consistent across SKCM, KIRP, KICH, and PRAD cancers **(Extended Data Figures 5F-5I)**, the iCluster 3 and iCluster 6 subtypes within TCGA LIHC and LUAD, respectively, demonstrated significantly differentiated neojunction expression compared with other iCluster subtypes **(Figures 3M-3N)**. Gene set enrichment analysis of the six LUAD iCluster subtypes revealed decreased expression of splicing-related gene pathways. Notably, 23 of these splicing-related gene sets were consistently downregulated in LUAD iCluster 6 compared with all five other iCluster subtypes. Together, these results indicate that in addition to splicing factor mutations, dysregulated expression of canonical splicing-related genes can lead to the generation of disease-specific neojunctions.



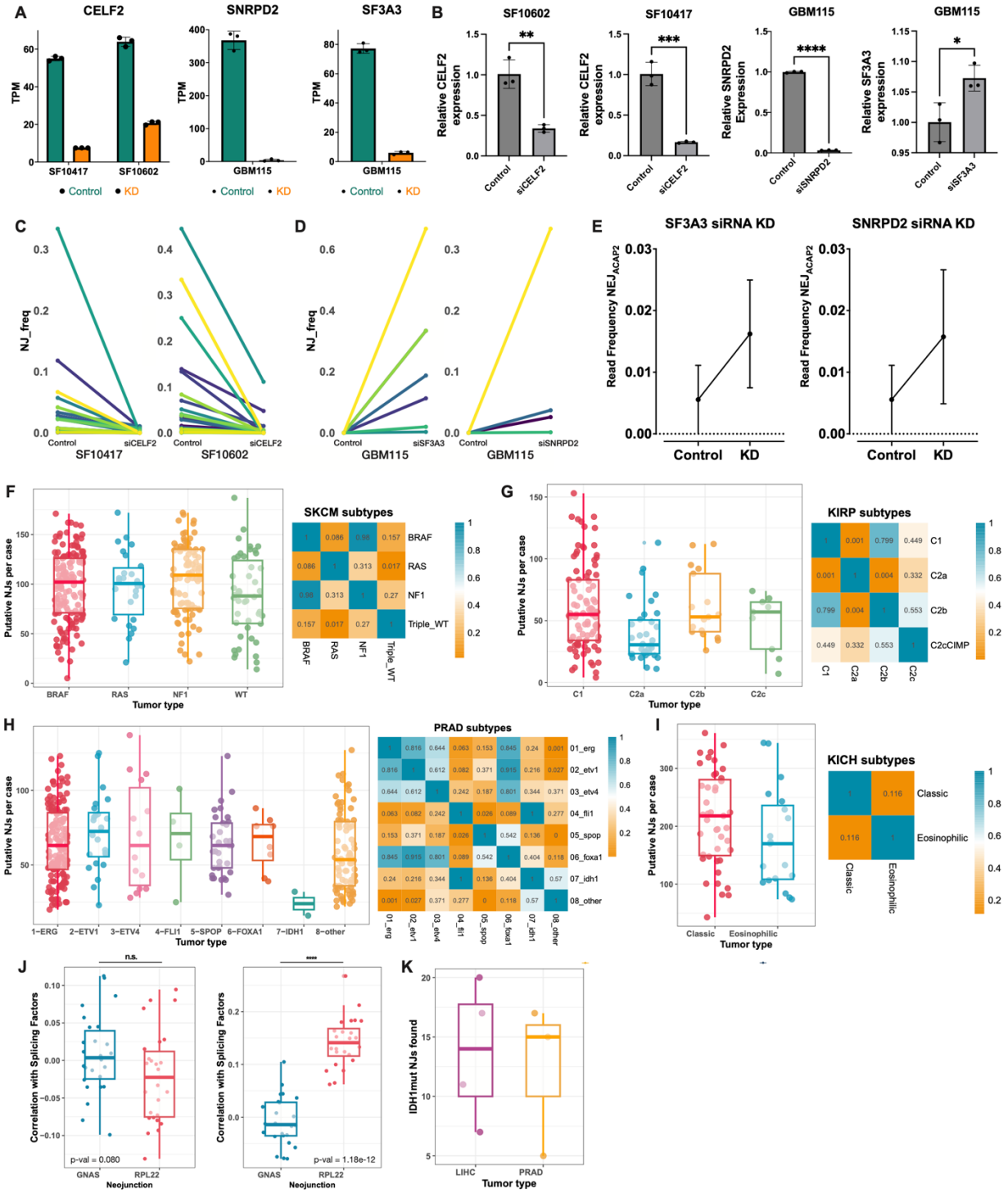
Extended Data Figure 3: Co-occurrence of somatic mutations in splicing-related genes. A-C. Heatmaps showing the pairwise Pearson correlation (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) matrix between gene expression of TCGA GBM/LGG samples computed for each gene pair. Splicing-related gene lists were defined by **A.** Nostrand et al. 2020, **B.** Sveen et al. 2016, and **C.** Seiler et al. 2018. **D-F.** Pie charts illustrating the proportions of IDH1 mutant samples that also contain mutations in **D.** *FUBP1*, **E.** *SF3B1*, and **F.** *NIPBL* in *IDH1*wt GBM samples (bottom), *IDH1*mut astrocytoma samples (top-left), and *IDH1*mut oligodendroglioma samples (top-right). **G-I.** Binary heatmap demonstrating the putative expression of neojunctions in relation to glioma subtypes and mutation status of **G.** *FUBP1*, **H.** *SF3B1*, and **I.** *NIPBL*.



Extended Data Figure 4: Glioma subtype-specific aberrations in splicing factor expression leads to differential levels of neojunction expression. A. Ranked log2 fold change of genes within the top enriched pathways within GOBP when (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) comparing IDH1mutA samples against IDH1wt samples. Green line plots show the running sum with the enrichment score peaking at the red dotted line, indicating leading edge genes that precede the peak. **B.** Ranked log2 fold change of genes within the top enriched pathways within GOBP when comparing IDH1mutO samples against IDH1wt samples. **C-E.** Heatmap demonstrating hierarchal clustering of splicing-related genes (rows) in GOBP mRNA Processing **(C)**, GOBP RNA Splicing **(D)**, and GOBP RNA Splicing via Transesterification Reactions **(E)** in TCGA samples (columns) ordered by the total number of expressed putative neojunctions.



Extended Data Figure 5: Dysregulated expression of canonical splicing-related genes are associated with disease subtype-specific expression of neojunctions. A-B. RNA-sequencing-derived **(A)** and qPCR **(B)** TPM expression of *CELF2*, (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) *SNRPD2*, and *SF3A3* following siRNA knockdown. **C.** Slope plots demonstrating decrease in the frequency of *IDH1*mut-specific NJs in *CELF2* siRNA-treated SF10417 (left) and SF10602 cells (right). **D.** Slope plots demonstrating increase in the frequency of *IDH1*mut-O-specific NJs in *SF3A3* siRNA- (left) and *SNRPD2* siRNA-treated (right) GBM115 cells. **E.** RNA-sequencing-derived read frequency of NJ_{ACAP2} in GBM115 cells treated with control siRNA and si*SF3A3* (left) or control siRNA and si*SNRPD2* (right). **F-I.** Bar plot (left) showing the total neojunctions expressed per case across all disease subtypes and heatmap (right) displaying the Wilcoxon rank-sum test of neojunction expression between each subtype within TCGA SKCM **(F)**, KIRP **(G)**, PRAD **(H)**, KICH **(I)**. **J.** Correlation of *IDH1*mut-specific splicing-related genes (left) and chr1p/19q splicing-related genes (right) with NJ_{GNAS} and NJ_{RPL22}. **K.** Detection of glioma-derived *IDH1*mut neojunctions in *IDH1*mut TCGA LIHC and PRAD samples.

Conclusion

Our comprehensive study of neojunction expression across glioma subtypes and other cancers has revealed several key insights into the mechanisms driving subtype-specific splicing dysregulation. Firstly, *IDH* mutation status significantly influences neojunction expression, with *IDH*mut gliomas exhibiting higher levels of neojunctions compared to *IDH*wt gliomas. This led us to discover that differential expression of splicing-related genes, rather than mutations in splicing factors, can be a major driver of neojunction formation in gliomas. Upregulation of specific splicing-related genes, such as *CELF2*, correlates with increased neojunction expression in *IDH*mut gliomas. Downregulation of splicing-related genes located on chromosomes 1p and 19q, particularly in *IDH*mut-O gliomas, contributes to increased neojunction expression. Modulation of splicing-related gene expression through siRNA knockdown experiments confirmed the causal relationship between these genes and neojunction levels. Analysis of other TCGA cancer types revealed additional tumor subtypes with significantly dysregulated neojunction expression, suggesting that this phenomenon extends beyond gliomas.

These findings have important implications for our understanding of cancer biology and potential therapeutic approaches: The identification of subtype-specific neojunction patterns may serve as valuable biomarkers for cancer diagnosis and prognosis. Targeting dysregulated splicing-related genes or their downstream effects could provide novel therapeutic opportunities for specific cancer subtypes. The complex interplay between genetic alterations, gene expression changes, and splicing dysregulation highlights the need for integrative approaches in cancer research and treatment. Our results underscore the importance of considering tumor heterogeneity and subtype-specific molecular features when developing targeted therapies. In conclusion, this study provides a comprehensive overview of the factors driving neojunction expression in gliomas and other cancers, paving the way for future research into the functional consequences of these aberrant splicing events and their

potential as therapeutic targets. Further investigation into the mechanisms of neojunction formation and their impact on tumor biology may lead to innovative strategies for cancer diagnosis, prognosis, and treatment.

Chapter 3: Neojunction-specific CD8+ T-cell killing

With the characterization of clonal and publicly-expressed neojunctions, the next and final goal is to identify those that would generate an immunogenic target. To accomplish this, we selected a cohort of neojunctions and validate their ability to be endogenously processed and presented by human leukocyte antigen (HLA) molecules. Presentation by HLA is required for the proper recognition of the neoantigen by a specific T-cell receptor (TCR). In this study, we identify neoantigen-specific CD8+ T-cell-derived TCR sequences and evaluate their ability to mount a neojunction-derived neoantigen-specific immune and cytotoxic response.

Detection of Neojunctions in RNA and Protein Levels in Patient-Derived Tumor Samples

We first sought to validate the expression of public neojunctions and their protein products in cell line transcriptomic data and tumor tissue proteomic data. Targeting neojunctions expressed in available cell lines will facilitate their usage in downstream *in vivo* assays, and as such, we looked towards confirming the expression of these neojunctions in publicly available or in-house cell lines. Focusing on gliomas, known for high ITH and poor clinical outcomes, we investigated glioblastoma multiforme (GBM) patient-derived xenografts (PDX) (n=66)[59] and lower-grade glioma (LGG) cell lines (n=2).[56] We measured the expression of 767 (97.2%) and 510 (64.6%) of our characterized public neojunctions in GBM and LGG, respectively **(Figures 4A-4B)**. While short read sequencing allows for higher throughput interpretations of neojunction expression, we wanted to accurately confirm the expression of specific candidate neojunctions in cell lines. Using deep amplicon sequencing with primers spanning a subset of neojunctions and their flanking exons, we confirmed mRNA expression of neojunction-spanning reads in glioma cell lines **(Extended Data Figure 6J)**.

To test whether these neojunctions are translated into proteins, we analyzed mass spectrometry (MS) data from glioma patients (n=447) using publicly available datasets.[60–62] Our analysis detected neopeptides corresponding to 302 (38.3%) unique public neojunctions **(Figure 4C)**. These peptide sequences spanned aberrantly spliced regions, confirmed by sequence-specific searches within the MS data and subsequent analysis of the MS spectra **(Extended Data Figures 6K-L)**. Notably, 41.7% of detected peptides mapped to neojunctions causing frame shifts **(Extended Data Figures 6M)**, indicating that splicing aberrations inducing a frame shift can lead to detectable translated peptides. Our peptidome analysis confirmed that neojunction-encoding transcripts are actively translated into detectable protein products. When considering both RNA-seq and MS confirmation of glioma-specific neojunctions, we validated the presence of 192 (24.3%) public neojunctions expressed across all patient-derived samples **(Figure 4D)**. These findings demonstrate the existence of recurrent public neojunctions that generate tumor-specific polypeptides.

Predicting tumor-wide neojunctions encoding neoantigens presented by HLA

We hypothesized that a subset of translated neojunctions could be presented as targetable neoantigens.[16,17,19] We investigated whether the 789 characterized public neojunctions could generate peptides processed and presented by human leukocyte antigen (HLA) class I following proteasomal processing. All properly translated neojunction-derived sequences from TCGA were translated in silico to generate a neojunction-derived protein dataset. Tumor-specific n-mers (8 to 11 amino acids) **(Figure 4E)** were defined as those absent from a UniProt reference normal human tissue proteome dataset.

Prediction of HLA class I-presented peptides needs to integrate key aspects of antigen-presentation machinery, including peptide processing and HLA binding.[63] Using two independent prediction

algorithms, MHCflurry 2.0 and HLAthena, we identified neoepitope sequences likely to be processed and presented **(Extended Data Figure 6A)**.[64,65] To rank candidate public neopeptides, we assessed the binding potential of n-mer candidates against the most prevalent HLA-A types expressed across a wide range of demographics. Amongst the 36 demographically predominant HLA-A alleles **(Extended Data Figure 4F-4G)**, our analyses investigated the presentation likelihood of neoantigen candidates by HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*11:01, and HLA-A*24:02.[66] Together, these alleles are expressed by the majority of the global population.[67,68] To select high-binding targets, we focused on n-mer candidates that scored in the top 1% with both algorithms **(Figures 4F-4G, Extended Data Figures 6B-6C).** Candidate n-mers that yielded these scores (*n*=832) were predicted to be processed and presented in glioma, and were retained for downstream analysis **(Figure 4H).** Mapping these top candidates to their originating neojunctions, we found that 315 neopeptide-encoding neojunctions (NEJs; 39.9% of the originally characterized public neojunctions) produced cancer-specific peptide sequences with top n-mer candidates.

Mapping these top candidates to their originating neojunctions, we determined that 315 neopeptide-encoding neojunctions (NEJs; 39.9% of the originally characterized public neojunctions) produced cancer-specific peptide sequences containing top n-mer candidates. While a greater number of top-scoring n-mer candidates are generated from frameshifts and alternative exonic 3' splice sites **(Extended Data Figures 6D-6E)**, presentation scores remained relatively consistent across n-mer candidates generated from NEJs with or without frameshift mutations or any specific splice type **(Figures 4I-4J, Extended Data Figures 6H-6I).** To further narrow our NEJ candidate list, we cross-referenced the 315 NEJs with the 192 neojunctions that we previously characterized as being shared across transcriptomic and proteomic platforms **(Figure 4D).** Of the 192 neojunctions, 81 were characterized as NEJs through this analysis **(Figure 4K)**, with many NEJs encoding multiple strongly

predicted candidates. We focused our downstream analyses on 32 candidate NEJs that were predicted to bind strongly to HLA-A*02:01 due to this allele's high prevalence across North American and European populations[66,69] and the ability to benchmark to other neoantigen studies[70–72] **(Figure 4L)**. When ITH of these 32 neojunctions was investigated in the data set from spatially mapped samples, high intratumoral conservation was observed for most of these NEJs, particularly the two nucleotide A3 loss-encoding neojunction located within *GNAS* (NJ_{GNAS}) **(Figure 4M)**. These findings demonstrate that intratumorally conserved public neojunctions may generate HLA-presented neopeptides.

This comprehensive analysis validates the presence of public neojunctions at both the RNA and protein levels in gliomas, highlighting their potential as targets for neoantigen-based immunotherapies. The consistent expression and translation of these neojunctions across diverse glioma samples, coupled with their ability to generate HLA-presented neoepitopes, underscore their therapeutic relevance. These findings pave the way for future research into the therapeutic potential of targeting public neojunction-derived neoantigens in glioma and other cancers with high intratumoral heterogeneity.

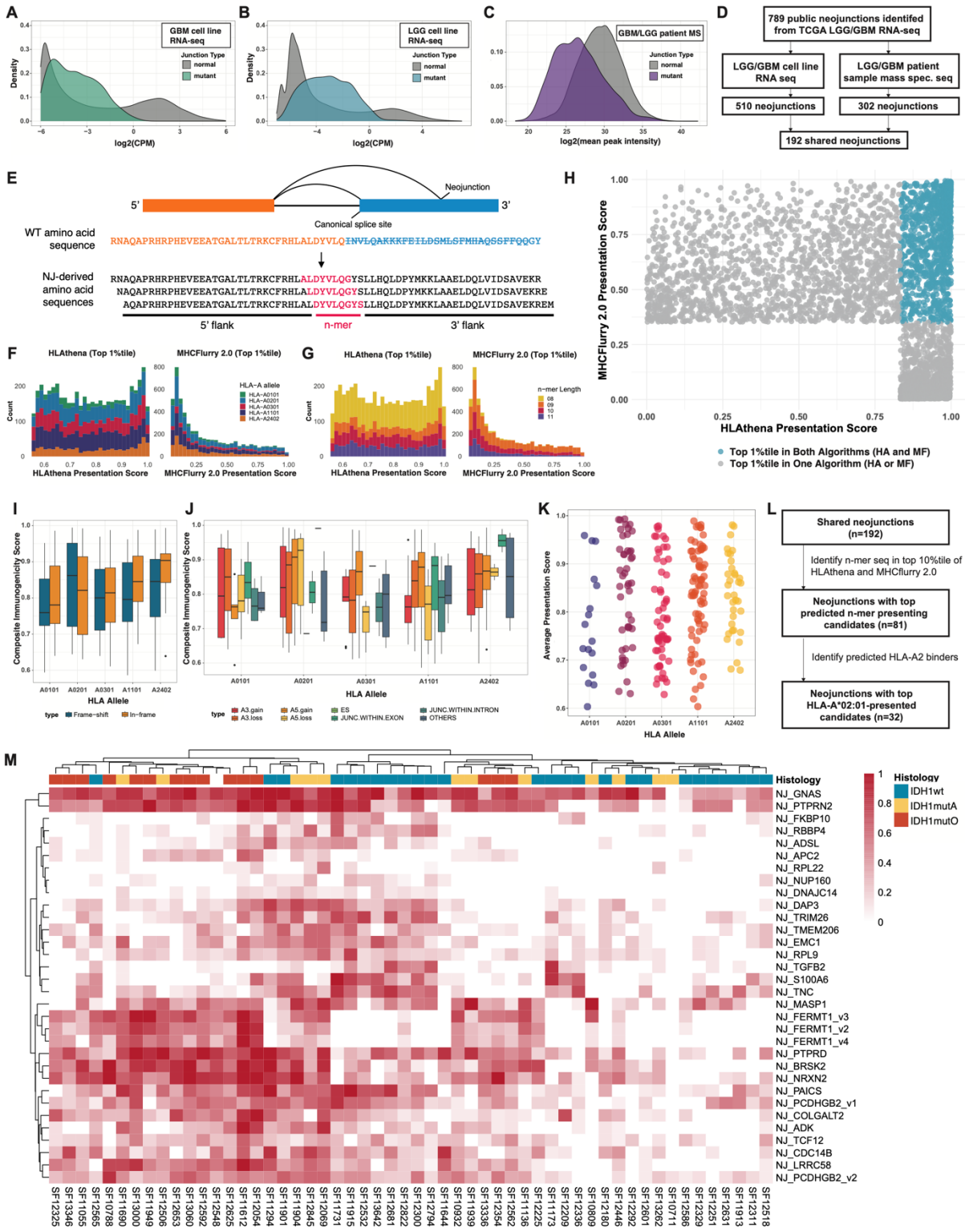
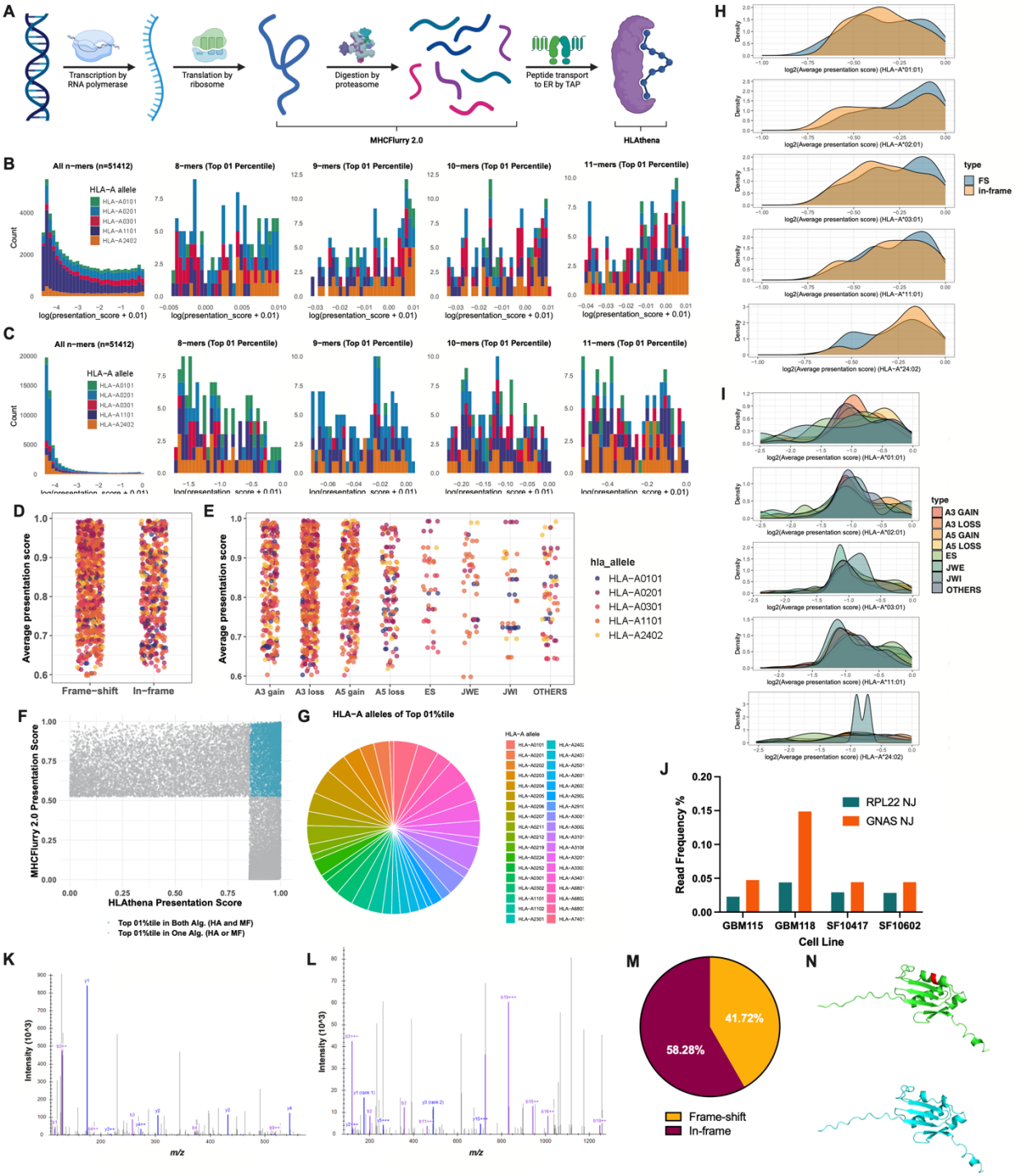


Figure 4: Neojunction-derived neoepitopes are predicted to be processed and presented by HLA. A-B. Density plots depicting $\log_2$(CPM) of junction reads derived from RNA-sequencing from patient-derived GBM (**A**) and LGG (**B**) cell lines that (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) validate detectable levels of neojunction expression (colored) compared to the expression of canonical splicing (gray). **C.** Density plot depicting mass spectrometry analysis of publicly-available LGG and GBM data sets (*n*=447). The log2(peak intensity) of detectable neojunction-derived peptides (purple) indicate comparable intensity compared to other endogenous peptides (gray). **D.** Schematic demonstrating the selection of high-confidence neojunctions for downstream analysis. 192 neojunctions were selected based on expression in all RNA-seq and MS platforms. **E.** Diagram illustrating the mechanism of neoantigen production by the introduction of a neojunction. Multimer partitioning was employed subsequently to generate a peptide bank for prediction analysis. **F-G.** Histogram illustrating the scores pertaining to the likelihood of peptide presentation calculated from each algorithm for the top scoring 10-percentile of n-mers categorized by HLA-allele (**F**) or n-mer length (**G**). **H.** Dot plot showing the overlay of the top scoring 1-percentile of both algorithms. Top-scoring final candidates are indicated in blue as the candidates that scored in the top 1 percentile in both algorithms. **I-J.** Bar graphs illustrating the composite presentation score (average presentation scores from HLAthena and MHCflurry 2.0) of top-scoring candidates binding across HLA-alleles based on frame-shift status (**I**) or alternative splicing category (**J**). **K-L.** Dot plot (**K**) and schematic (**L**) describing the immunogenic potential of the final candidate list based on top-scoring neojunction-derived n-mers derived from neojunctions detectable in patient-derived mass spectrometry and RNA sequencing data. **M.** Heatmap illustrating the intratumoral heterogeneity of final candidate HLA-A*02:01-presented neojunctions across all spatially-mapped glioma samples. Glioma subtypes analyzed in this study include *IDH1*wt (blue), *IDH1*mut astrocytoma (yellow), *IDH1*mut oligodendroglioma (red), and one case (SF12548) that simultaneously exhibited either *IDH1*mut astrocytoma or *IDH1*mut oligodendroglioma across its spatially mapped samples (purple).



Extended Data Figure 6: Neojunction-derived neopeptides are detectable in patient mass spectrometry and predicted to be processed and presented by HLA. A. Schematic depicting the biological steps leading to the generation of HLA class I-presented antigens. Our neoantigen discovery pipeline considers the pre-presentation steps of proteasomal processing and HLA-binding. **B-C.** Histogram depicting the distribution of the top scoring (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) 8-mers, 9-mers, 10-mers, and 11-mers in **B.** HLAthena and **C.** MHCflurry 2.0. Scores pertain to presentation by HLA-A*01:01 (green), HLA-A*02:01 (blue), HLA-A*03:01 (red), HLA-A*11:01 (purple), and HLA-A*24:02 (orange). **D.** Jitter plot corresponding to the average presentation scores of peptides derived from neojunctions generating frame-shifts or in-frame mutations. **E.** Jitter plot corresponding to the average presentation scores of peptides derived from neojunctions derived from various splice types. **F.** Dot plot showing the overlay of the top scoring 1-percentile of HLAthena and MHCflurry 2.0 algorithms against neopeptide candidates presented by all demographically predominant HLA-A haplotypes. Top-scoring final candidates are indicated in blue as the candidates that scored in the top 1 percentile in both algorithms. **G.** Pie chart illustrating the distribution of neopeptide candidates found in the overlapping top 1 percentile based on composite HLA-A haplotype score. **H-I.** Density plots depicting the average presentation scores of neoantigens derived from neojunctions generating **H.** frame shifts or **I.** various splice types presented by HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*11:01, and HLA-A*24:02. **J.** Read frequency of reads spanning neojunctions in *RPL22* and *GNAS* compared to the canonical junction spanning reads in glioma cell lines (n=1). **K-L.** Mass spectra of peptide sequences spanning the aberrantly spliced regions in **K.** RPL22 and **L.** GNAS detected in publicly available GBM and LGG MS data. **M.** Proportion of mass spectrometry peptides that map back to neojunctions that encode for frame-shift or in-frame mutations. **N.** AlphaFold2 renders of the three-dimensional structure of wild-type RPL22 (top) and mutant RPL22 (bottom). The AA loss is indicated in red.

NEJ-reactive T-cell Receptors can be isolated from donor CD8+ T-cells

We next sought to determine whether NEJ-derived neopeptides can drive T-cell immunogenicity. We performed *in vitro* sensitization (IVS) to identify neoantigen-reactive $CD8^+$ T-cell populations from healthy donor (HD)-derived peripheral mononuclear cells (PBMCs)[73,74] **(Figure 5A)**. We focused on a subset (n=4) of our 32 top NEJ candidates predicted to generate high-affinity binders to HLA-A*02:01 **(Figures 4K-4M)**. IVS of naïve CD8+ T-cells was conducted against neopeptide-pulsed autologous monocyte-derived dendritic cells (moDCs) collected from HLA-A*02:01+ HDs (n=5) to retrieve TCR gene sequences that confer specificity against these neoantigens. Subsequent IFNγ ELISA assays on the corresponding APC:CD8+ conditions revealed neoantigen-reactive immunogenicity in two out of four of the public NEJ-derived neoantigens (NeoAs): $NeoA_{RPL22}$ and $NeoA_{GNAS}$ **(Figure 5B)**. $NeoA_{GNAS}$ results in an A3 loss of 2 nucleotides generating a frame-shift and premature stop codon. $NeoA_{RPL22}$ encodes for an in-frame A3 loss of 6 nucleotides, resulting in a loss of two amino acids in an alpha helix **(Extended Data Figure 6N)**. These results indicate that NEJ-reactive CD8+ T-cells can exist within the naturally occurring human T-cell repertoire.

To retrieve TCR gene sequences that confer reactivity to these neoantigens, we repeated the peptide-pulsed-APC:CD8+ T-cell co-culture on $NeoA_{RPL22}$- and $NeoA_{GNAS}$-reactive CD8+ T-cell populations and performed combined single-cell V(D)J and RNA-seq. Neoantigen-reactive TCR clonotypes were associated with significantly elevated *IFNG*, *TNFA*, *GZMB* transcripts in neoantigen peptide-specific manners. Using this method, we identified seven $NeoA_{RPL22}$-reactive TCRs, two from Donor 3 ($TCR_{R3.7}$ and $TCR_{R3.9}$) and five from Donor 4 ($TCR_{R4.5}$, $TCR_{R4.6}$, $TCR_{R4.7}$, $TCR_{R4.9}$, and $TCR_{R4.11}$), and one $NeoA_{GNAS}$-reactive TCR from Donor 4 ($TCR_{G4.1}$) **(Figure 5C)**. Although only one $NeoA_{GNAS}$-reactive TCR clonotype was characterized, this clonotype was the most proliferated TCR clone,

expanding to over 4% of the TCR repertoire in the CD8+ T-cell population **(Figure 5D)**. The expansion of neoantigen-reactive CD8+ T-cell clones suggest a strong immunogenic response to these two neoantigens.

NEJ-reactive TCRs recognize NEJ-derived neoantigens in an HLA-restricted manner

To further test the specificity of the identified $TCR_{R3.9}$ and $TCR_{G4.1}$-reactive T-cell clones, we transduced TCR-null triple-reporter (TR) Jurkat76 cells which express the CD8a/b heterodimer (Jurkat76/CD8) or PBMC-derived $CD8^+$ T-cells with lentiviral vectors encoding the retrieved TCR α- and β-chains.[75] The TR Jurkat76/CD8 cells have response elements for NFAT, NF-κB, and AP-1 which drive expression of eGFP, CFP, and mCherry, respectively **(Figure 5E)**.[76,77] TCR-transduced TR Jurkat76 cells cultured with T2 cells pulsed with varying concentrations of neoantigen peptide demonstrated dose-dependent reactivity **(Figures 5F, Extended Data Figures 7A-7B).** Both TCRs demonstrated nM-level neoantigen-recognition, illustrating a relatively high functional avidity of the corresponding TCRs. The antigen-specificity of these receptors was supported by negligible TCR activation in the presence of supraphysiologic levels of the control peptide (1 μM). TCR-transduced PBMC-derived $CD8^+$ T-cells displayed similar dose-dependent neoantigen-specific behavior **(Figures 5G-5H).** TCR-transduced $CD8^+$ T-cells were stained for surface expression of the T-cell activation and degranulation markers, CD137 and CD107a, to quantify markers of T-cell activation and effector function, respectively. T-cell activation was observed at neoantigen-peptide concentrations as low as 1 pM **(Figure 5G).** Similarly, IFNγ and TNFα expression levels measured by ELISA suggested strong potency of both TCRs as indicated by their half-maximal effective peptide concentrations (EC_{50}s) between 0.01 to 0.1 nM **(Figure 5H, Extended Data Figure 7C).** Treatment of neoantigen-pulsed T2 cells with an HLA-blocking antibody prior to co-culture with the TCR-transduced TR Jurkat76 cells validated that neopeptide T-cell activation is HLA-dependent **(Figure 5I)**.

Next, we performed alanine scanning mutagenesis to determine whether either NEJ-reactive TCR can recognize polypeptides derived from off-target normal human proteins.[78] TCR-transduced triple-reporter Jurkat76/CD8 cells were cultured against residue-substituted neoantigen isoforms, and key residues were defined by those that result in diminished TCR activation **(Extended Data Figure 7E).** Alterations in the recognition of a variant peptide indicate that the substituted residue was critical for TCR recognition.[79] Referencing each TCR's peptide recognition motif to a normal human proteome library (UniProt Proteome ID #UP000005640) demonstrated that no known human proteins share the key residues required for TCR recognition. Together, our results reveal TCRs that exclusively recognize NEJ-derived public neoantigens with robust sensitivity and highlight a potential immunotherapeutic approach utilizing TCR-engineered T-cells to target this novel class of shared neoantigens.

Finally, we determined whether circulating NEJ-reactive $CD8^+$ T-cell populations could be detected in glioma patients by immune monitoring studies on archived PBMC samples obtained from three HLA-A*02:01 glioma patients with known expression of NEJ_{GNAS} **(Figure 4M)**. NEJ_{GNAS} IVS of these PBMC samples led to the detection of an immunogenic response in 1 of 3 glioma patients with no immunogenicity against an irrelevant HLA-A*02-restricted 9-mer neoantigen dextramer control **(Figure 5J)**. These findings further support the immunogenicity and potential clinical application of targeting NEJ-derived neoantigens.

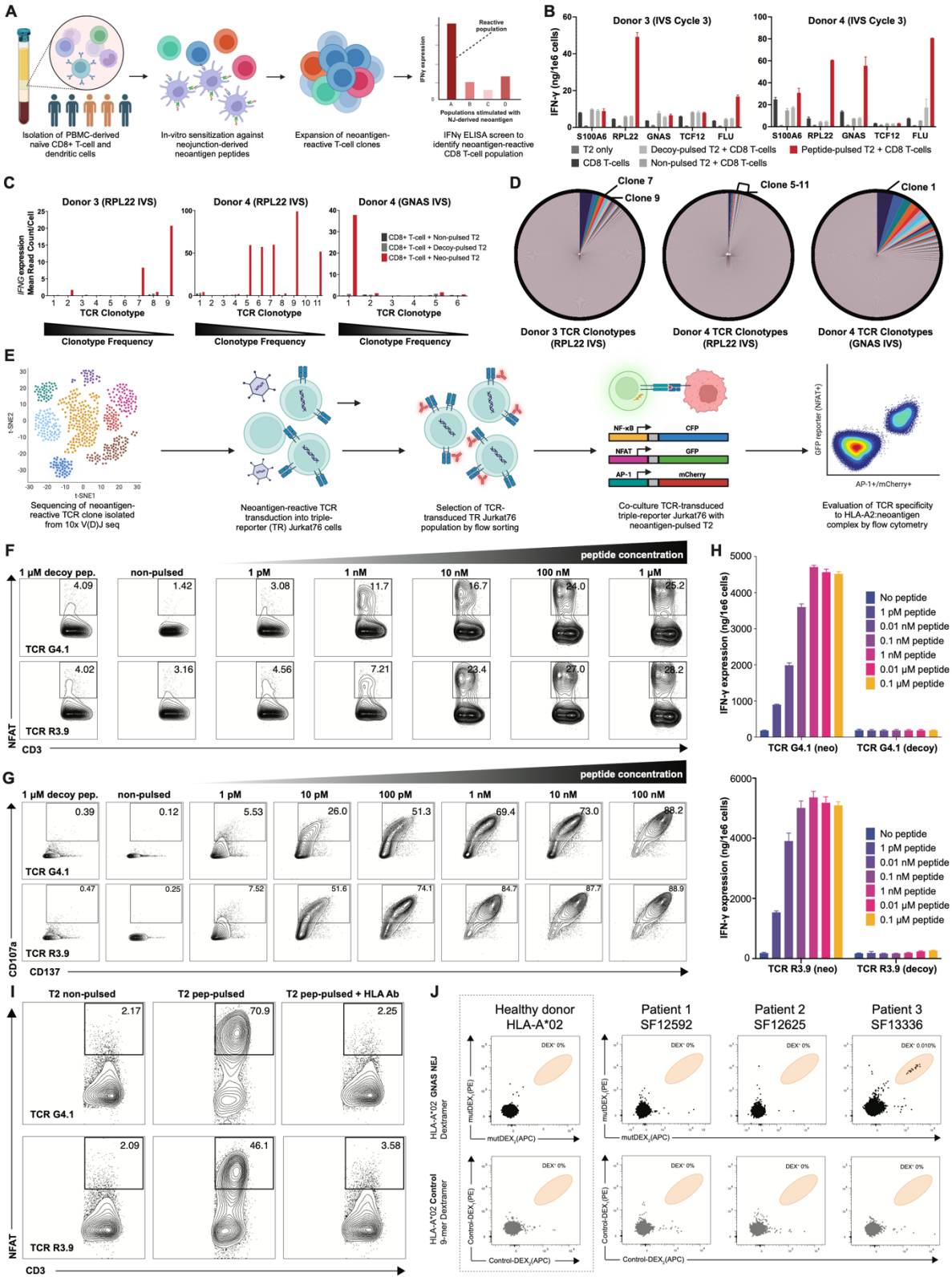
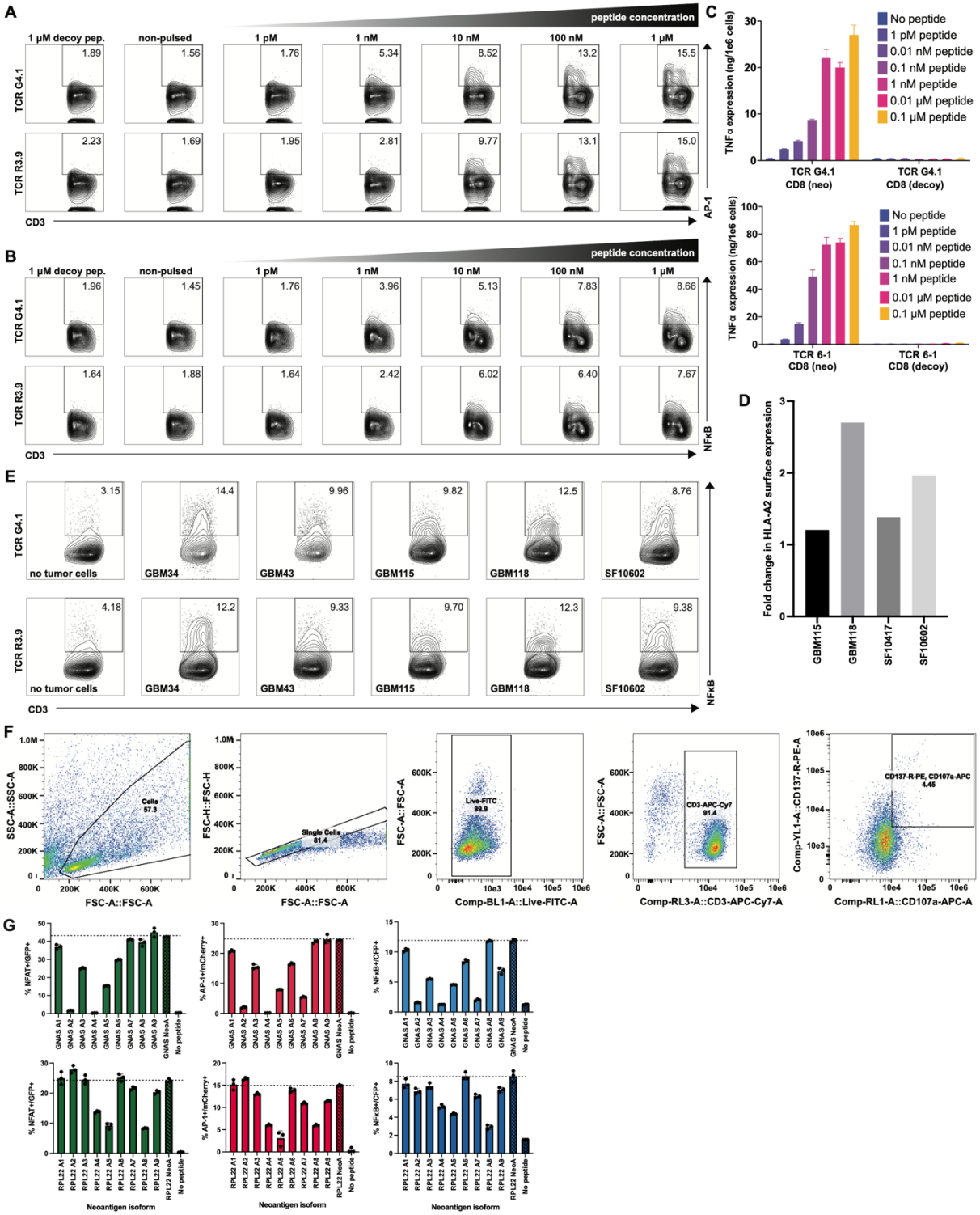


Figure 5: T-cell receptors (TCRs) specifically react to neojunction-derived neoantigens. A. Pipeline overview for identifying neojunction-derived neoantigen-reactive T-cell populations through *in vitro* sensitization (IVS) of healthy-donor (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) PBMC-derived $CD8^+$ T-cells against APC-presented neopeptides. **B.** IFNγ ELISA of reactive $CD8^+$ T-cell populations following IVS with neoantigen. **C.** 10x V(D)J *IFNG* signatures of highly proliferated TCR clonotypes cultured against T2 cells pulsed with the neoantigen (colored), a control peptide (light-gray), or no peptide (dark-gray). Reactive TCR clonotypes were identified in Donor 3 $CD8^+$ T-cells IVS-treated against $NeoA_{RPL22}$ (left), Donor 4 $CD8^+$ T-cells IVS-treated against $NeoA_{RPL22}$ (center), and Donor 4 $CD8^+$ T-cells IVS-treated against $NeoA_{GNAS}$ (right). **D.** Clonotype frequency of all TCR clones identified in Donor 3 $CD8^+$ T-cells IVS-treated against $NeoA_{RPL22}$ (left), Donor 4 $CD8^+$ T-cells IVS-treated against $NeoA_{RPL22}$ (center), and Donor 4 $CD8^+$ T-cells IVS-treated against $NeoA_{GNAS}$ (right). **E.** Pipeline for validating the specificity of neoantigen-reactive TCR clonotypes found in 10x V(D)J single-cell RNA sequencing (scRNA-seq) against neojunction-derived neoantigen candidates utilizing a TCR-transduced triple-reporter Jurkat76/CD8 system followed by flow cytometry analysis. **F-G.** $NeoA_{RPL22}$ (top) and $NeoA_{GNAS}$-specific (bottom) TCR-transduced triple-reporter Jurkat76/CD8 cells (**F**) or PBMC-derived $CD8^+$ T-cells (**G**) were activated against neoantigen-pulsed T2 cells in a dose-dependent manner. TCR-transduced cells were co-cultured with control peptide-pulsed T2 cells at the highest dose concentration (1mM). TCR activation of TCR-transduced triple-reporter Jurkat76/CD8 was measured by flow cytometry analysis of NFAT-GFP. PBMC-derived $CD8^+$ T-cells were stained with CD107a and CD137 antibodies and surface expression of the TCR co-activation markers were analyzed by flow cytometry. **H.** IFNγ ELISA of $NeoA_{RPL22}$ (top) and $NeoA_{GNAS}$-reactive (bottom) TCR-transduced $CD8^+$ T-cells co-cultured with dose-dependent neoantigen-pulsed (left) and control peptide-pulsed T2 cells (right). **I.** $NeoA_{RPL22}$ (top) and $NeoA_{GNAS}$-specific (bottom) TCR-transduced triple-reporter Jurkat76 cells are co-cultured with non-pulsed T2 cells (left), 0.1 μM neoantigen-pulsed T2 cells (center), or 0.1 μM neoantigen-pulsed T2 cells treated with pan-HLA class I blocking antibody (right). Cells were stained with CD3 antibody, and TCR activation was evaluated by NFAT-GFP activity. **J.** $NeoA_{GNAS}$-dextramer staining of bulk $CD8^+$ T-cells derived from an HLA-A*02:01 healthy donor (left) and glioma patients (right) following 2 cycles of NJ_{GNAS} IVS.



Extended Data Figure 7: Neoantigen-reactive T-cell clones are isolated from PBMC and elicit an immune response upon neoantigen recognition. A-B. A. NJ_{RPL22}-derived and NJ_{GNAS}-derived neoantigen-specific TCR-transduced triple-reporter (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) Jurkat76 cells activated against dose-dependent neoantigen-pulsed T2 cells. TCR activation of triple-reporter TCR-transduced triple-reporter Jurkat76 is measured by flow cytometry analysis of NFAT-GFP (top row) and NFκB-CFP (bottom row). **C.** TNFα expression of PBMC-derived $CD8^+$ T-cells (n=3) transduced with TCR 6-1(top) and TCR 1-1 (bottom) against T2 cells pulsed with varying concentrations of corresponding neoantigen or decoy antigen. **D.** Surface HLA-A2 expression of LGG and GBM cell lines (n=1) with or without 48-hour pre-treatment of IFNγ. **E.** Alanine scanning mutagenesis of $NeoA_{RPL22}$ (top) and $NeoA_{GNAS}$-reactive (bottom) TCR-transduced triple-reporter Jurkat76/CD8 cells co-cultured with alanine-substituted neoantigen-pulsed T2 cells, neoantigen-pulsed T2 cells, or non-pulsed T2 cells. Flow analysis was performed to evaluate TCR activity through NFAT-GFP (left), AP-1-mCherry (center), and NFκB-CFP (right) activity. **F.** Gating strategy for flow cytometry experiments.

NEJ-derived public neoantigens are endogenously processed and presented by HLA

Next, we tested whether neojunction-derived transcripts generate peptides that are functionally presented by HLA and recognized by reactive TCRs. We evaluated the presentation of NEJ-derived neoantigens using two approaches: functional TCR-recognition and HLA-immunoprecipitation followed by liquid-chromatography-MS/MS **(Figure 6A).** To determine whether the neojunction transcript expression leads to immune recognition, we co-cultured COS7 cells transfected with the HLA-A2 and full-length mutated transcript together with either TCR-transduced TR Jurkat76 or $CD8^+$ T-cells. $TCR_{R3.9}$ and $TCR_{G4.1}$-transduced TR Jurkat76 and $CD8^+$ T-cells reacted against COS7 cells transfected with their respective neoantigen, demonstrating endogenous processing and presentation of the public NEJs (**Figures 6B-6C**). We then performed affinity-column-based immunopurification of HLA-I ligands on COS7 cells co-transfected with HLA/mutant transcript. The MS analysis identified the same $NeoA_{GNAS}$ peptide as the highly abundant HLA-A2-bound peptide with high-confidence. Likewise, both $NeoA_{RPL22}$ neopeptides were detected with high confidence on COS7 cells co-transfected with HLA-A*02:01 and NJ_{RPL22} with the higher scoring $NeoA_{RPL22}$ 9-mer identified with higher relative abundance **(Figure 6D).** Furthermore, we could detect the HLA-A*02:01-restricted $NeoA_{GNAS}$ peptide in an unmodified GBM cell line (GBM115) **(Figure 6E)**. This finding demonstrates that physiological levels of neojunction expression in tumor cells are sufficient to generate a NEJ-derived neoantigen. Together, these experimental observations confirm our *in silico* predictions for proteasomal processing and HLA-binding **(Figure 4L).**

TCR-transduced $CD8^+$ T-cells mediate cytotoxicity against glioma cells expressing NEJ-derived public neoantigens

Based on the sensitivity of neoantigen-specific TCR **(Figures 5F-5G)** and the endogenous presentation of NEJ-derived neoantigens by tumor cells **(Figure 6E),** we anticipated that

physiological levels of public NEJ expression in tumor cells would elicit a neoantigen-specific cytotoxic T-cell response. We evaluated the cytotoxicity of TCR-transduced $CD8^+$ T-cells against HLA-A*02:01^+ tumor cells endogenously expresssing NJ_{RPL22} and NJ_{GNAS}. As a positive control, we used neoantigen peptide-pulsed tumor cells to define maximum cell killing. At a 1:1 effector:target ratio, $TCR_{R3.9}$ and $TCR_{G4.1}$-transduced $CD8^+$ T-cells mediated TCR-dependent cytotoxicity against GBM115 cells **(Figure 6F).** $TCR_{G4.1}$-transduced $CD8^+$ T-cells mediated comparable levels of tumor killing against a second glioblastoma cell line, GBM102 and two melanoma cell lines, RPMI-7951 and WM-266-4 **(Extended Data Figures 8A).** Adding an HLA-I blocking antibody partially blocked killing compared to an isotype control, verifying tumor cell killing is initiated by TCR recognition of the HLA:peptide complex **(Figure 6G).** Co-culture of $TCR_{G4.1}$-transduced $CD8^+$ T-cell with an HLA-A2-negative, NJ_{GNAS} expressing GBM cell line (Mayo PDX GBM39) revealed HLA-A2 dependent cytotoxicity **(Figure 6H)**. These results illustrate that the recognition and killing of NEJ-expressing tumor cells is mediated by HLA-dependent neoantigen presentation. Relative to non-transduced $CD8^+$ T-cells, increased surface expression of CD107a and CD137 on TCR-transduced $CD8^+$ T-cells co-cultured with tumor cells further confirmed neoantigen-specific T-cell activation **(Figure 6I, Extended Data Figure 8B)**. Together, these data indicate that NEJs are endogenously processed and presented at sufficient levels to enable tumor-specific cytotoxicity by neoantigen-specific $CD8^+$ T-cells.

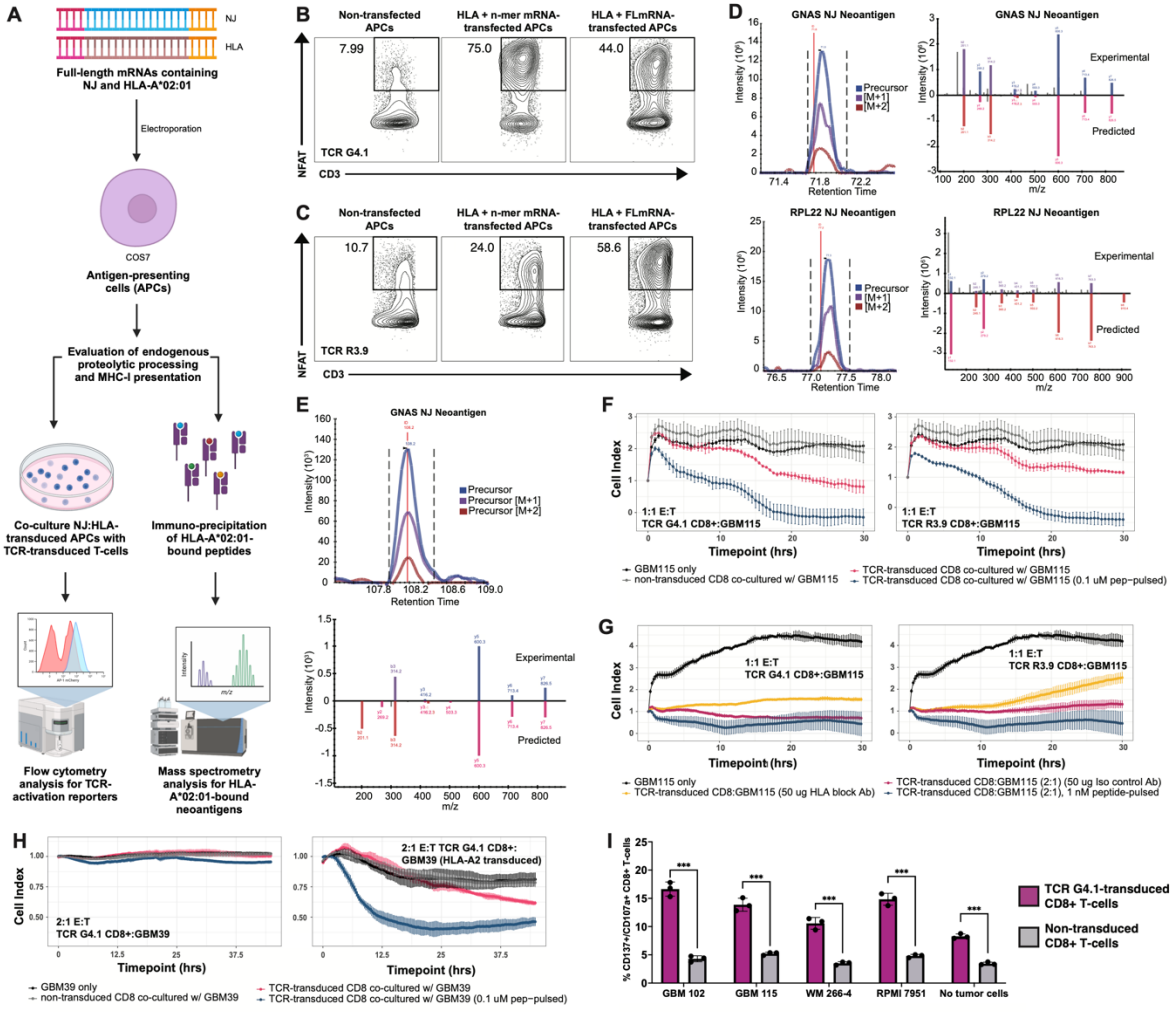
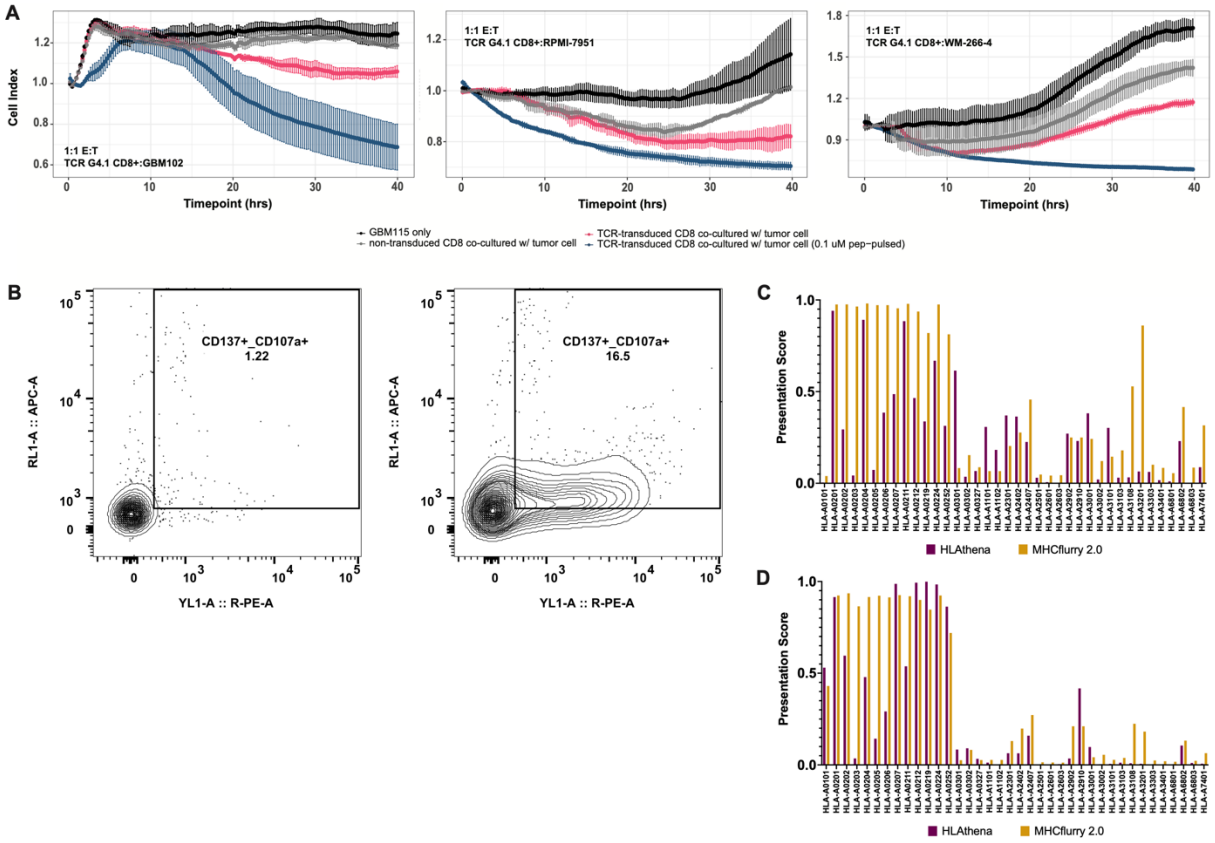


Figure 6: Neojunction-derived neoantigens are endogenously processed and presented by HLA to elicit neoantigen-specific T cells in patients and TCR-dependent tumor-specific killing. A. Pipeline overview for validating endogenous proteolytic cleavage of neoantigen candidates and the subsequent binding and presentation by HLA using two complementary methods. HLA-null antigen-presenting cells (APCs; COS7 or K562) were electroporated with mRNAs encoding the full-length gene with the neojunction mutation and HLA-A*02:01. First, TCR-activation following co-culture with APCs that express neojunction transcripts and HLA-A*02:01 was quantified by FACS using neoantigen-specific TCR-transduced triple-reporter Jurkat76/CD8 cells. Second, HLA-I-bound peptides were validated with IP-MS/MS of transfected APCs followed by identification of HLA-A*0201-bound peptide sequences. **B-C.** NJ_{GNAS}-derived (**B**) and NJ_{RPL22}-derived (**C**) neoantigen-specific TCR-transduced triple-reporter Jurkat76 cells were co-cultured against transfected COS7. Cultured cells were either non-transfected (left), transfected with mRNA encoding the neoantigen n-mer sequence and HLA-A*02:01 (center), or transfected with mRNA encoding the full-length (FL) mutant peptide and HLA-A*02:01 (right). TCR activation of TCR-transduced triple-reporter Jurkat76/CD8 was measured by flow cytometry analysis of NFAT-GFP. **D.** Mass spectrometry spectra of NJ_{GNAS}-derived (bottom) and NJ_{RPL22}-derived (top) neoantigen n-mers detected through IP-MS/MS following HLA-A*02:01 pulldown of HLA-A*02:01 and full-length neojunction-encoding mRNA-transduced COS7 cells. **E.** Mass spectrometry (Figure captions continued on the next page.)

(Figure captions continued from the previous page.) spectra of HLA-A2-presented NJ_{GNAS}-derived neoantigen n-mers detected through IP-MS/MS following HLA-A*02:01 pulldown of GBM115 tumor cells. **F.** NJ_{GNAS}-derived (left; colored), NJ_{RPL22}-derived (right; colored) neoantigen-specific TCR-transduced, or non-transduced (gray) $CD8^+$ T-cells were cultured against GBM115 tumor cells. Cell index indicates tumor cell adherence to the xCELLigence plate platform, with decreased cell index indicating tumor cell death. The assay was performed at an E:T ratio of 1:1. Cytotoxic killing was determined as the reduction of cell index compared to the control group with no $CD8^+$ T-cell introduction (black) at a given timepoint. **G.** xCELLigence live-cytotoxicity assay of $CD8^+$ T-cells co-cultured with GBM115 tumor cells incubated with an anti-HLA-I antibody (yellow), an isotype control antibody (purple), or pulsed with 1 nM of the neoantigen peptide (blue). NJ_{GNAS}-derived (left) and NJ_{RPL22}-derived (right) neoantigen-specific TCR-transduced $CD8^+$ T-cells were cultured against these GBM115 cells. **H.** xCELLigence live-cytotoxicity assay of $CD8^+$ T-cells co-cultured with GBM39 cells (left) or HLA-A*02:01-transduced GBM39 cells (right) co-cultured with non-transduced or NJ_{GNAS}-TCR-transduced $CD8^+$ T-cells. **I.** Bar plot of the surface expression of CD107a and CD137 on NJ_{GNAS}-TCR-transduced (purple) or non-transduced (gray) $CD8^+$ T-cells when cultured with tumor cell lines.



Extended Data Figure 8: NeoA$_{GNAS}$-specific cytotoxicity by NJ$_{GNAS}$-specific TCR-transduced CD8+ T-cells. A. NeoA$_{GNAS}$-specific TCR-transduced (colored), or non-transduced (gray) CD8$^+$ T-cells cultured against GBM102 (left), RPMI-7951 (center), and WM-266-4 (right) on the xCELLigence plate platform. The assay was performed at an E:T ratio of 2:1. Cytotoxic killing was determined as the reduction of cell index compared to the control group with no CD8$^+$ T-cell introduction (black) at a given timepoint. **B.** Representative flow gating of surface CD107a and CD137 expression in co-cultured CD8+ T-cells. **C-D.** Composite HLA-presentation scores of NeoA$_{GNAS}$ **(C)** and NeoA$_{RPL22}$ **(D)** by HLAthena and MHCflurry 2.0 against 36 HLA-A haplotypes.

Conclusion

Our study demonstrates that NEJ-derived neoantigens can elicit a strong T-cell immunogenic response, highlighting their potential as targets for TCR-engineered T-cell immunotherapies. The identification and characterization of NEJ-reactive TCRs with high functional avidity and specificity offer a promising avenue for developing targeted treatments against tumors expressing NEJ-derived public neoantigens. The ability of TCR-transduced CD8+ T-cells to mediate cytotoxicity against glioma and melanoma cells underscores the therapeutic potential of this approach. Further exploration and clinical studies are warranted to harness the full potential of NEJ-derived neoantigens in cancer immunotherapy.

Chapter 4: Discussion

Our analysis of cohorts of multi-site samples indicated the expression of NJ_{RPL22} across multiple samples within the same tumor. Most notably, NJ_{GNAS} was expressed tumor-wide in diverse tumor types, including glioma, mesothelioma, prostate cancer, and hepatocellular carcinoma **(Figure 1H).** The discovery of a targetable tumor-wide neoantigen in GBM provides a novel potential therapeutic approach for this devastating disease. The higher expression level of the canonical *GNAS* allele over *RPL22* may contribute to the prevalence of NJ_{GNAS} detected across all analyses. This supports the observation of greater immunogenicity and tumor-specific killing by $TCR_{G4.1}$ **(Figures 6F-6G)** as there is a greater frequency in generating and presenting $NeoA_{GNAS}$. Indeed, a NJ_{GNAS}-specific $CD8^+$ T-cell population could be detected in the circulation of an HLA-A*02:01$^+$ glioma patient whose tumor exhibited NJ_{GNAS} expression **(Figure 5J)**

We also investigated the dysregulated expression of splicing-related genes associated with the increased expression of neojunctions in *IDH*-mutated gliomas. Mutations in *IDH* and *IDH2* are prevalent in other cancers, including acute myeloid leukemia (AML)[80–82], cholangiocarcinoma[83,84], chondrosarcoma[85], sinonasal undifferentiated carcinoma[86–88], and angioimmunoblastic T cell lymphoma[89,90]. Our study demonstrated dysregulation in splicing factor expression in different disease types and that these aberrations lead to significant changes in neojunction production. In the case of *IDH* mutant oligodendrogliomas, *SNRPD2* expression is decreased due to the characteristic co-deletion of chromosomes 1p and 19q, and targeted knockdown increased neojunction expression. This suggests that components of the RNA splicing machinery are mechanistically linked to the generation of neojunctions. Inhibiting these components in future studies can potentially bolster targetable NEJ-derived neoantigen expression.

Limitations of the Study

Several recent studies have identified HLA class II-restricted neoepitopes that are capable of driving antitumor $CD4^+$ T cell responses in patients with gliomas[91,92] and other solid cancers[93,94]. Due to the limitations of currently available HLA-II binding prediction algorithms, we did not assess whether HLA-II restricted public neoantigens resulting from clonally conserved NEJs are generated. Similarly, our study does not investigate surface-bound neojunction-derived neoantigens as we found they were difficult to characterize.[23] Therefore, validation of neoantigen candidates from our pipeline was performed solely on predicted HLA-A*A02:01 binders to demonstrate proof-of-concept. Future validation studies could include candidates predicted as high-binders for other prevalent HLA class I alleles to further expand the repertoire of targetable neoantigens.

Importantly, the most comprehensive analysis of ITH (average of 10 intratumorally-mapped samples) was conducted using GBM and LGG samples. To fully validate neojunctions and their corresponding neoantigens as tumor-wide across other cancer types, we will need a greater set of intratumoral sites per patient, including a temporal and wide anatomical distribution to maximally represent the evolving tumor. Finally, we did not assess the biological contribution of the studied NEJs to the malignant phenotype.

In conclusion, our study highlights that RNA splicing aberrations are a robust source of intratumorally-conserved and tumor-wide public TSAs that the immune system can recognize. The ability to target tumor-wide neoantigens with engineered T-cells enables a powerful therapeutic approach that could tackle the significant clinical challenge of ITH. Ultimately, the results from our study could allow us to design effective vaccine panels comprising tumor-wide neoantigen targets and to engineer T-cell-based modalities that target tumor-wide splice-derived antigens across a wide range of cancer types.

Chapter 5: Future directions

Intratumoral heterogeneity of personalized neoantigens

While the focus of our study is dedicated to characterizing tumor-wide, public neojunction-derived neoantigens, it would be critical to apply the same approach in investigating the ITH of neoantigens for personalized treatments. Beyond the study presented in Chapters 1 through 4, we additionally investigated the ITH of somatic mutation-derived neoantigens in Grade II astrocytomas.[95] In a collaboration with PACT Pharma, we were able to capture neoantigen-specific T-cell clones that selectively interacted with various sub-clonal and clonal neoantigens across multiple patients. In one patient, we were able to demonstrate neoantigen-specific immunogenicity against one particular sub-clonal neoantigen but not yet against clonal neoantigens. This may suggest a possible mechanism of "immune sculpting" in which neoantigens derived from driver and/or clonal mutations may be less immunogenic due to evolutionary factors influencing a tumor's immunosuppressive or antigen presentation phenotype.

Epigenetic regulation of splicing factors

As discussed in Chapter 2, there is a plethora of mechanisms that may contribute to the generation of neojunctions. In particular, we demonstrated in our studies that mutant *IDH1* gliomas have significantly upregulated expression of neojunctions than their wild-type counterpart **(Figure 3A-3B)**. What remains to be explored is the effect of mutant *IDH1*-mediated global hypermethylation on splicing. Is there a possibility that methylation of splicing-related genes in mutant *IDH1* gliomas leads to downregulated expression which subsequently generates more opportunities for aberrant splicing? Studies have shown that methylation can influence the progression rate of RNA polymerase II,[96,97] and exploring the intricate interplay of *IDH1*-mediated methylation's effect on splicing machinery can help us develop therapies against a multitude of cancer types.

Translation of *in vitro* cytotoxicity to *in vivo* models

Finally, the importance of translating our cytotoxic findings from our *in vitro* studies to an *in vivo* model would solidify the therapeutic potential of targeting splicing-derived neoantigens. At the time of this publication, there has been no confirmation of anti-tumor cytotoxicity in murine models. Evaluating the cytotoxic behavior of our characterized TCRs in xenograft murine models would be the next step in bringing this needed cell-based immunotherapy closer to the clinic.

Methods

Data and code availability

Spatially-mapped biopsy RNA sequencing data will be deposited. HLA-IP and LC-MS/MS data will be deposited and made publicly available at the date of publication. Additionally, single-cell V(D)J sequencing data of identified TCRs will be deposited and made publicly available as of the date of publication. Accession numbers are listed in the key resources table. All original code for the identification of tumor-wide public neojunctions have been deposited at GitHub (https://github.com/dakwok/SSNIP) and are publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contacts upon request.

Human clinical datasets

The intratumoral multi-region sampling cohort for various cancer types utilizes RNA sequencing data from the following studies:

1. This paper, for multi-region sampling of glioblastoma and low-grade glioma
2. Yang et al. (Genome Medicine, 2022), for multi-region sampling of hepatocellular carcinoma.
3. Joung et al. (PLoS One, 2016), for multi-region sampling of hepatocellular carcinoma, stomach adenocarcinoma, renal cell carcinoma, and colon adenocarcinoma.
4. Ku et al. (Briefings in Bioinformatics, 2021), for multi-region sampling of prostate cancer.
5. Meiller et al. (Genome Medicine, 2021), for multi-region sampling of mesothelioma.
6. Bakir et al. (Nature, 2023), for multi-region sampling of non-small cell lung cancer

Analysis of neojunction expression within multi-region samples were conducted immediately with our neojunction prediction pipeline if the FASTQ file is available. If RNA-sequencing data is only available in BAM format, the sequencing file is converted into FASTQ format utilizing the Picard software (version 2.7.7a). Neojunction prediction is detailed in the Method Details section.

Data download

Bulk RNA-sequencing data for glioblastoma (GBM; n=167), low-grade glioma (LGG; n=516), lung adenocarcinoma (LUAD, n=517), lung squamous cell carcinoma (LUSC, n=501), mesothelioma (MESO, n=516), liver hepatocellular carcinoma (LIHC, n=371), stomach adenocarcinoma (STAD, n=415), kidney renal clear cell carcinoma (KIRC; n=533), kidney renal papillary cell carcinoma (KIRP; n=290), kidney chromophobe (KICH, n=66), colon adenocarcinoma (COAD; n=458), and prostate adenocarcinoma (PRAD; n=497) samples were downloaded from TCGA in FASTQ format. Download of intratumoral multi-region sampling sequencing data is detailed in the previous section. Similarly, bulk RNA-sequencing data for 9,166 normal tissue samples in FASTQ format were downloaded from the Genotype-Tissue Expression (GTEx) repository. Bulk-RNA sequencing data for 66 patient-derived GBM cell lines were received from the Mayo Clinic Brain Tumor Patient-Derived Xenograft National Resource.[59] Proteomics data for 100 GBM samples were downloaded from the Clinical Proteomic Tumor Analysis Consortium (CPTAC).[60]

RNA sequencing alignment

All downloaded RNA-sequencing data sets were individually aligned using a STAR aligner-based processing pipeline. Using the STAR software (version 2.7.7a), we constructed a genome index containing non-annotated junctions through the initial alignment pass of the input data. The complete

set of command line parameters: --runThreadN 1 \ --outFilterMultimapScoreRange 1 \ --outFilterMultimapNmax 20 \ --outFilterMismatchNmax 10 \ --alignIntronMax 500000 \ --alignMatesGapMax 1000000 \ --sjdbScore 2 \ --alignSJDBoverhangMin 1 \ --genomeLoad NoSharedMemory \ --limitBAMsortRAM 80000000000 \ --readFilesCommand gunzip -c \ --outFilterMatchNminOverLread 0.33 \ --outFilterScoreMinOverLread 0.33 \ --sjdbOverhang 100 \ --outSAMstrandField intronMotif \ --outSAMattributes NH HI NM MD AS XS \ --limitSjdbInsertNsj 2000000 \ --outSAMunmapped None \ --outSAMtype BAM SortedByCoordinate \ --outSAMheaderHD @HD VN1.4 \ --twopassMode Basic \ --outSAMmultNmax 1 \ and aligned using the GRCH37 STAR index file.

TCGA sample selection and gene expression quantification

TCGA tumor samples with an absolute tumor purity greater than 0.60 were retained downstream *in silico* analysis. (Aran et al. 2015, Ceccarelli et al. 2016) We selected non-mitochondrial, protein-coding transcripts defined by the Ensembl Homo Sapiens GRCH37.87 gene annotation gene transfer format (GTF) file and utilized this curated to select and retain protein-coding transcript isoforms within the TCGA RNA-sequencing data. Transcript-level expression data (log2[RSEM-TPM+0.001]) for all TCGA samples were downloaded from the UCSC Xena Toil-pipeline and transformed into standard TPM values. Protein-coding transcript isoforms with a median TPM ≥ 10 were retained for downstream analysis. In the case of glioma TCGA cases, subsequent expression data in TPM was subset into 6 disease type categories: all cases (n=429), GBM cases (n=115), LGG cases (n=314), *IDH1*-WT cases (n=166), *IDH1*-MUT astrocytoma cases (n=140), *IDH1*-MUT oligodendroglioma (n=123). Protein-coding transcript isoforms with a median TPM ≥ 10 in at least one of the 6 disease types were retained for further analysis.

Characterization of public neojunctions

For public cancer-specific splicing event counting, we designed a custom R script that detected and quantified non-annotated, cancer-specific splicing events found across each corresponding patient cohort. From the output files derived from STAR aligner in the previous step, alternative splicing events were quantified in detected junction counts within the corresponding sj.out.tab file. We removed splicing events detected in the GRCh37.87 GTF sj.out.tab (GENCODE v33) file to define non-annotated splicing junctions. Non-annotated splicing junctions that overlap non-mitochondrial, protein-coding genes identified in the previous step were retained for continued analytical processing. We removed all splicing junctions with less than 10 of its target spliced reads (count) or less than 20 total spliced reads (depth) over the whole cohort. Similarly to previous studies[13], we computed spliced frequency as the sum of the total number of target spliced reads divided by the collective sum of spliced reads from the target and canonical junctions. Splicing junctions with a read frequency greater than 1% were retained for downstream analyses. We defined public splicing junctions as ones that were putatively expressed with the aforementioned criteria of total read count, read depth, and read frequency across at least 10% of the studied patient cohort and retained those for further analysis. To characterize cancer-specific splicing events, otherwise known as neojunctions, we removed all junctions that putatively expressed with the same parameters in more than 1% of GTEx normal samples.

Detection of cancer-specific intron retention events

Intronic splicing events were detected and characterized using IRFinder v1.2.3. RNA sequencing data from TCGA (GBM/LGG) and GTEx (CNS) aligned to GRCh37 (hg19) were imported into the

software for the detection of intron retention events. General linear model (GLM)-based analysis was used for differential intron retention assessment. The intron retention ratio is calculated as (intronic reads/sum(intronic reads, normal spliced reads). Significant intron retention changes are defined as (1) no less than 10% in both directions (2) adjusted p-values less than 0.05. An intron retention event's PSR within TCGA or GTEx is defined as the number of cases that fulfill these criteria divided by the total number of cases within the cohort. Putative cancer-specific intron retention neojunctions are characterized as intron retention events with a TCGA PSR ≥ 0.10 and a GTEx PSR < 0.01.

Transcriptomic validation of expressed neojunctions

Detection of expressed neojunctions in patient-derived GBM/LGG cell lines: RNA sequencing data derived from GBM PDX cell lines were downloaded from the Mayo Clinic Brain Tumor Patient-Derived Xenograft National Resource. Patient-derived LGG cell lines were generated from surgically-resected specimens in University of California, San Francisco's Neurological Surgery Brain Tumor Center.[56] RNA sequencing data from GBM and LGG cell lines were aligned and processed as described above. Public neojunctions with splice junction counts per million (CPM) > 0 are considered detectable in cell line-derived RNA sequencing data. Detection of expressed neojunctions in multi-region cases: In our cohort of spatially-mapped glioma cases, approximately ten or more maximally-distanced anatomical biopsies were collected from each patient, allowing for intratumoral assessment of genetic heterogeneity via bulk RNA-seq and whole-exome sequencing. Multi-region sequencing data of various other cancer types vary in the number of sampled regions per tumor and are detailed in the corresponding references **(Figure S1)**. RNA-sequencing data collected from each multi-region sample was processed and aligned as described above. We searched for putative neojunctions previously characterized from TCGA within each multi-region sampling dataset. Public neojunctions

with CPM > 0 were considered detectable. Public neojunctions with putative expression (≥ 10 spliced reads) in two or more mapped samples within the same case are considered spatial-conserved neojunctions. Neojunctions detected in all multi-region samples within the same tumor are considered tumor-wide neojunctions.

Proteomic validation of expressed neojunction-derived peptides

From the putative neojunctions detected in the above pipeline, we generated a database of all plausible polypeptides derived from all neojunctions. Neojunction-encoding transcripts were generated by mapping the junction coordinates to an hg19 human genome assembly within the Ensembl annotation database (AH13964, EnsDb.Hsapiens.v75). Prediction of neojunction-derived amino acid sequences were subsequently performed, and appropriately translated sequences (methionine starting residue, removal of sequences following first stop codon) were retained for downstream n-mer iteration. To detect neojunction-derived polypeptides within GBM cases, we analyzed .RAW files of GBM and LGG MS dta housed in the Clinical Proteomics Tumor Analysis Consortium (CPTAC, *n*=99), Bader et al. (*n*=99), Lam et al. (*n*=92), and Yanovich-Arad et al. (*n*=84). MaxQuant (v1.6.17.0) was used to identify tryptic sequences from the corresponding MS data sets. Predicted neojunction-derived peptides, decoy sequences, and a human reference proteome (UniProt Proteome ID #UP000005640) were inputted as a FASTA file into MaxQuant, and tryptic sequences derived from the input file were matched against the publicly available MS databases. Cancer-specific peptides spanning neojunction-derived protein sequences were considered MS-confirmed. The relative detection levels of the neojunction-derived peptides and normal tissue-derived peptides were evaluated by their log2(peak intensities). Aside from the default settings, the following commands and parameters were modified

and used for MS analysis in MaxQuant: Digestion mode = Trypsin/P; Max missed = 3; Minimum peptide length = 5; Minimum peptide length for unspecific search = 5.

Peptide processing and HLA binding and presentation predictions

Cancer-specific transcripts with associated neojunctions were translated *in silico* into their corresponding amino acid sequences. A library of all possible peptides of 8 to 11 amino acids in length was then generated, and cancer-specific sequences were selected by removing those detectable in normal tissue peptide isoforms in a reference human proteome dataset (UniProt Proteome ID #UP000005640). All cancer-specific peptides with their upstream and downstream flanking sequences (maximum flanking length of 30 amino acids) were independently analyzed and ranked by MHCFlurry 2.0 and HLAthena MSiC. HLA-I binding affinity was assessed against HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*11:01, and HLA-A*24:02 in both cases. In the HLAthena evaluation of antigen binding and presentation to the corresponding HLA haplotypes, peptides were assigned to alleles by rank with a threshold of 0.1. Context of up to 30 flanking amino acids on both N and C terminus were utilized with aggregation by peptide and no log-transformed expression. Baseline MHCFlurry 2.0 models with both peptide:HLA binding affinity (BA) predictor and antigen processing (AP) predictor was used. Overall, peptide:HLA presentation scores were characterized by mhcflurry_presentation_score and MSiC_HLA scores in MHCFlurry 2.0 and HLAthena, respectively. To select for high-binders, we curated lists of peptide:HLA complexes within the top 10 percentile of scores from both prediction algorithms.

Cell culture

GBM PDX cell culture: Patient-derived xenograft (PDX) glioblastoma cell lines, GBM34, GBM43, GBM108, GBM115, GBM118, GBM102, GBM137, GBM148, GBM164, and GBM195, were obtained from the Mayo Clinic Brain Tumor PDX national resource. Xenograft lines were cultured as by recommended conditions in previous literature[98,99] and passaged a maximum of 20 times before restoration to earlier passages. Cells were cultured in Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and 1% penicillin and streptomycin (P/S). Cell culture plates were treated overnight at 4 °C with DPBS (with calcium and magnesium) and 10% laminin (Gibco™ Cat. #23017015) prior to use. Primary patient-derived GBM/LGG cell culture: Primary patient-derived wildtype *IDH1* GBM (SF7996), mutant *IDH1* astrocytoma (SF10602), and mutant *IDH1* oligodendroglioma (SF10417) cell lines were previously internally generated from dissociated glioma biopsies and cultured as previously described.[56] Cells were cultured in serum-free, glioma neural stem (GNS) cell medium, which comprises of Neurocult NS-A (STEMCELL Technologies Cat. #05751) supplemented with N-2 supplement (Invitrogen Cat. #17502048), B-27 supplement minus vitamin A (Invitrogen Cat. #12587010), 1% P/S, 1% glutamine, and 1% sodium pyruvate. Prior to immediate use in culture, GNS media is supplemented with 20 ng/mL EGF (Peprotech Cat. #AF-100-15), bFGF (Peprotech Cat. #AF-100-18B), and PDGF-AA (Peprotech Cat. #AF-100-13A). Similarly with the GBM PDX cell lines, cell culture plates were incubated overnight at 4 °C with DPBS (with calcium and magnesium) and 10% laminin (Gibco™ Cat. #23017015) prior to use. Jurkat76 cell culture: Jurkat76 cells were used as the TCR α- and β-negative human T-cell derivative that allowed for noncompeting introduction of exogenous TCRs.[100] CD8+ Jurkat76 cells were cultured in RPMI supplemented with 10% fetal bovine serum and 1% P/S. T2 cell culture: T2 cells were used in the study to monitor immune cell response to the exogenous antigen of interest in a non-competitive environment. T2 cells are deficient in a peptide transporter involved in antigen processing (TAP), and

as such, induction of these cells with exogenously administered peptides allows for their association and presentation by HLA molecules, HLA-A*0102 in particular.[75,101] We cultured T2 cells in IMDM medium supplemented with 20% FBS. COS7 and K562 cell culture: We opted to use COS7 (ATCC Cat. #CRL-1651) and K562 (ATCC Cat. #CCL-243) cell lines as our respective primate and human artificial antigen presentation cell (aAPC) models.[74,102,103] These cell lines do not express HLA molecules, which allows for the introduction of the HLA allele of interest. COS7 cells were cultured in DMEM medium supplemented with 10% FBS and 1% P/S. K562 cells were cultured in IMDM medium supplemented with 10% FBS and 1% P/S. THP-1 cell culture: THP-1 cells (ATCC Cat. #TIB-202) were used to investigate immune reactivity against neoantigen presentation by dendritic cells. THP-1 cells were cultured in RPMI-1640 supplemented with 10% FBS.

siRNA-mediated knockdowns of splicing related genes

Cells were seeded in 2 mL of antibiotic free media in a 6-well plate at the following densities: GBM115 – 45,000 cells/well, SF10417 – 100,000 cells/well, and SF10602 – 100,000 cells/well. Twenty-four hours post seeding cells were transfected by adding 400 uL reaction containing serum free media, 2.0 uL DharmaFECT 1 reagent (Horizon, #T-2001-02), and their respective siRNA pools (4 siRNA equimolar mix) at a final concentration of 30 nM. Twenty-four hours post transfection, media was changed to complete media. At seventy-two hours post transfection, RNAs were isolated and purified via the Zymo Quick-RNA microprep kit (Zymo Research, #R1058).

Reverse Transcription Quantitative PCR (RT-qPCR)

1000 ng of DNAse-treated RNA was converted to cDNA using the iScript cDNA synthesis kit (BioRad, #1708891). This cDNA was then diluted 1:3 using ultrapure, nuclease-free water, and 2 uL

was used per qPCR reaction. Reverse transcription quantitative polymerase chain reaction (RT-qPCR) was performed using the Applied Biosystems POWER SYBR Green Master Mix (Applied Biosystems, #4367659). All samples were run in biological triplicates, with technical triplicates for each biological triplicate using the Quantstudio 5 (Thermo Scientific) and all gene expression data were normalized to the housekeeping gene *GUSB*. The cycling protocol is as follows: 2 minutes at 50 °C, 10 minutes at 95 °C, followed by 40 cycles of 95 °C for 15 s, and 60 °C for 60 s. Dissociation curves were performed to confirm specific product amplification. Primer sequences corresponding to each gene for the mRNA expression analysis were designed using NCBI Primer.

Amplicon-sequencing for validation of neojunction expression

RNAs from respective cell lines were isolated and purified via the Zymo Quick-RNA microprep kit (Zymo Research, #R1058). 1000 ng of DNAse-treated RNA was converted to cDNA using the iScript cDNA synthesis kit (BioRad, #1708891). This cDNA was then diluted 1:3 using ultrapure, nuclease-free water, and 2 uL was used per PCR reaction. 16 reactions were carried out per amplicon per cell line using Q5 High-Fidelity 2x master mix (NEB, #M0492L) with primers containing partial Illumina adapters. Reaction mixtures were set up according to manufacturer guidelines. These products were then purified by separation on a 1.0% agarose gel at 100 volts constant for 1 hour and were then purified via the Monarch DNA gel extraction kit (NEB, #T1020L). Purified products were quantified with qubit high sensitivity dsDNA kit (Invitrogen, #Q32851) and prepared and submitted according to Azenta life sciences (Genewiz) guidelines for amplicon-sequencing.

In-vitro sensitization of healthy donor PBMCs

HLA-A*02:01:01-positive PBMCs were purchased from StemExpress in either fresh or cryopreserved format. Fresh PBMCs (StemExpress Cat. #LE001F) of approximately $1x10^9$ cells were immediately proportioned into aliquots of 3 x 10^8 cells and cryopreserved in liquid nitrogen, with one aliquot actively used for downstream IVS. Cryopreserved PBMCs (StemExpress Cat. #PBMNC300C) totaling approximately $3x10^8$ cells per cryovial were used in one vial per IVS procedure. PBMCs were thawed with 1:1000 Benzonase:RPMI (Sigma Aldrich Cat. #E8263). The CD14+ population was isolated from the PBMCs using CD14+. Miltenyi microbeads (Miltenyi Biotec Cat. #130-050-201) per manufacturer's instructions. The CD14- flowthrough was cryopreserved for 6 days prior to naïve CD8+ T-cell isolation. Isolated CD14+ cells were cultured in CellGenix GMP DC medium (CellGenix Cat #20801-0500) supplemented with 1% human serum (Sigma Aldrich Cat #H6914), 1% P/S, 1000 U/mL recombinant human IL-4 (Peprotech Cat. #200-04), and GM-CSF (Peprotech Cat. #300-03) in non-treated 24-well plates at a seeding density of $5x10^5$ cells per well. On Day 3, recombinant human IL-4 and GM-CSF (1000 U/mL each) is added to the DC culture. On Day 5, DC culture is matured with 250 ng/mL LPS (Sigma Aldrich Cat. #L6529) in addition to supplementation of recombinant human IL-4 and GM-CSF (1000 U/mL each). Naïve CD8+ T-cells were isolated from the thawed CD14- population on Day 6 using the EasySep Human Naïve CD8+ T Cell Isolation Kit (STEMCELL Technologies Cat. #19258) as per manufacturer's instructions. Isolated naïve CD8+ T-cells were cultured in X-Vivo 15 medium (Lonza Cat. #04-418Q) supplemented with 5% human serum, 1% P/S, and 10 ng/mL of recombinant human IL-7 (Peprotech Cat. #200-07) in 48-well plates at a seeding density of $5x10^5$ cells per well. On Day 8, adherent matured DCs were harvested from the plate using cold PBS. The collected DCs ($1x10^6$ cells/mL) were exogenously pulsed with either 1 μM of the neoantigen peptide, influenza peptide, or no peptide for 1 hour at 37 ^oC. The peptide-pulsed or non-pulsed DCs were then co-cultured with naïve CD8+ T-cells at an optimal DC:T-cell ratio of 1:4

in 48-well plates. The co-culture was maintained with X-Vivo 15-medium supplemented with 10 ng/mL of recombinant human IL-7, 10 ng/mL recombinant human IL-15 (Peprotech Cat. #200-15), and 60 ng/mL of recombinant human IL-21 (Peprotech Cat. #200-21) for 10 days with IL-7 and IL-15 restimulation every 2 days. Cells were reseeded into subsequent 24-well, 12-well, and 6-well plates based on confluency. This concludes the first cycle of IVS of the neoantigens and influenza peptides. On Days 19 and 29, sensitized-CD8+ T-cells are reintroduced to a second and third round of stimulation with newly-pulsed DCs, and the co-culture is maintained for 10 additional days until the end of the second and third cycle of IVS. Immunogenic cytokine assays were performed at the end of the second and third cycles of IVS to determine whether a peptide-reactive T-cell population has expanded.

Mutation-specific ELISA screen

Aliquots containing $CD8^+$ T-cells from individual parent IVS wells were harvested and split equally into 96-well plate daughter wells containing $1x10^5$ cells per well. Daughter wells in triplicate were stimulated with T2 cells pulsed with either the neoantigen peptide of interest, control peptide, no peptide, or no T2 cells at all for 16 hours at an effector-to-target (E:T) ratio of 1:1. T2 cells were pulsed with 1 pM to 1 μM of the neoantigen peptide of interest, control peptide, or no peptides for 1 hour at 37 °C. Influenza-reactive T-cells were co-cultured against influenza peptide-pulsed T2 cells as a positive control. Co-culture supernatant was collected and diluted for use in IFNγ (BD Biosciences Cat. #555142) and TNFα (BD Biosciences Cat. #555212) ELISAs as per manufacturer's instructions. ELISA readouts were performed on the Epoch Microplate Spectrophotometer (BioTek Instruments) using the BioTek Gen5 Data Analysis software (version 1.11). Wells with significantly increased

expression levels of IFNγ and TNFα were selected for downstream single-cell immune profiling using single-cell RNA and V(D)J sequencing.

Single-cell immune profiling

Once an expanded neoantigen-reactive $CD8^+$ T-cell population from IVS is identified, single-cell RNA and V(D)J sequencing were performed using the 10x Genomics platform. Prior to sequencing, $CD8^+$ T-cells from the expanded neoantigen-reactive (ELISA screen-positive) wells were harvested and co-cultured with T2 cells pulsed with 1 μM of the neoantigen peptide of interest, a control peptide, or no peptides at an E:T ratio of 1:1. One co-culture replicate was performed for 3 hours for single-cell RNA sequencing analysis, and another was performed for 16 hours for IFNγ and TNFα ELISA confirmation. The final cell concentration was adjusted to approximately $1x10^4$ cells/μL with an initial cell viability of at least 90% to maximize the likelihood of achieving the desired cell recovery target. Independent $CD8^+$ T-cell and non-pulsed T2 single cultures were sequenced alongside the co-culture conditions for differentiating cell types in the downstream single-cell sequencing analysis. The Chromium Next GEM Single Cell 5' Reagent Kit v2 (Dual Index) (10xGenomics, Cat. #CG000331) was used for preparation for single-cell sequencing analysis. Gel bead in emulsions (GEMs) were generated by combining the single cell 5' gel beads, partitioning oil, and the master mix containing the cells onto the Chromium Next GEM Chip K. Cell lysis and barcoded reverse transcription of RNAs in all single cells were finished inside their corresponding GEM. Barcoded cDNA product was recovered through post-GEM-RT cleanup and PCR amplification. cDNA quality control and quantification were performed on the Fragment Analyzer System (Agilent Technologies). 50 ng of cDNA was used for the construction of the 5' gene expression library, and each sample was indexed by a Chromium i7 Sample Index Kit. This process was performed on an Illumina NovaSeq 6000

sequencer at the UCSF Institute of Human Genetics (IHG) with a minimum of 20,000 read pairs per cell for the 5' Gene Expression library. The enriched product was measured by the Fragment Analyzer System. 50 ng of enrichment TCR product was used for library construction. Single-cell V(D)J enriched libraries were subsequently sequenced on the Illumina NovaSeq 6000 with a minimum of 5,000 read pairs per cell for the V(D)J library. Cell Ranger 7.0.0 (10x Genomics Cloud Analysis) was used to pre-process raw single-cell RNA sequencing and identifying V(D)J clonotypes. The annotation files 'vdj_GRCh38_alts_ensembl-3.1.0-3.1.0' and 'GRCh38-3.0.0' were used for demultiplexing cellular barcodes, performing read alignments, and generating feature-barcode matrices. Only cells for which clonotype information was available were retained for downstream analysis. Single-cell gene expression and corresponding V(D)J sequences of candidate T-cell clonotypes were analyzed on the Loupe V(D)J browser. Single cells with detectable *CD8A* expression were specifically isolated and characterized as the $CD8^+$ T-cell population and subsequently grouped according to their TCR clonotypes. To identify T-cell clonotypes associated with a neoantigen-specific response, we selected expanded TCR clonotypes with significantly increased levels of *IFNG*, *TNF*, and *GZMB* expressions in the T-cell:neoantigen-pulsed T2 condition compared to the T-cell:control-pulsed T2 and T-cell:non-pulsed T2 conditions.

HLA typing

OptiType 1.3.1 was used for genotyping HLA alleles from available WES data available for glioma cell lines with default parameters.

Plasmids and peptides

HLA-A*02:01 and neojunction-derived gene sequences were all synthesized and cloned into the pTwist Lenti SFFV Puro WPRE vector (Twist Biosciences). Constructs encoding full-length and truncated multi-mer versions of the wildtype and mutant *GNAS* and *RPL22* sequences were generated. TCR α/β was synthesized and cloned into the pTwist Lenti SFFV vector (Twist Biosciences). HPLC grade neojunction-derived neoantigen peptide multi-mers (>95%) were manufactured by TC Laboratories.

Lentiviral transduction

HEK293T cells were plated in 6-well culture plates at a density of $1x10^6$ cells per well with 2 mL DMEM medium supplemented with 10% FBS without antibiotics. After approximately 18 to 24 hours or at 90% confluency, HEK293T cells were transfected with the expression construct, see above, and lentiviral packaging plasmids, pMD2.G (Addgene, #12259) and psPAX2 (Addgene, #12260). TCR α/β transduction: 1.0 μg TCR α/β transfer plasmid, 0.75 μg psPAX2, and 0.25 μg pMD2.G were combined with 200 μL Opti-MEM (Thermo Fischer Scientific Cat. #31985062). 6 μL of Xtremegene HP was added to this mixture and complex formation was allowed to occur for 15 minutes at room temperature at which point this reaction mixture was added to corresponding HEK293T cells. Transfection medium was replaced with fresh DMEM media after 24 hours. Viral supernatant was collected after 48 hours, and a functional virus titer was performed on 6-well plates seeded with Jurkat76/CD8 cells or PBMC-derived $CD8^+$ T-cells at 60-70% confluency. Viral transduction was performed with 3-fold serial dilutions of the virus stock supplemented with polybrene at a final concentration of 4 μg/mL. Media was changed 24 hours following viral transduction. Cells were assessed for transduction efficiency after 3-4 days by measuring surface expression of TCR α/β and CD3 by fluorescence-activated cell sorting (FACS) analysis. Cells demonstrating high double-positive

expression of TCR α/β and CD3 were flow-sorted and maintained for downstream co-culture and immunogenicity assays. HLA and neoantigen transduction: Constructs expressing HLA-A*02:01 were linearized and restricted with BamHI and XhoI (New England Biolabs) and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research Cat. #D4007). The HLA-A*0201 sequence was then ligated into a lentiviral construct downstream of an EF1A-core promoter and upstream of an IRES followed by a Blasticidin resistance gene. 1.0 μg of either HLA-A*02:01 or neoantigen transfer plasmid, 0.75 μg psPAX2, and 0.25 μg pMD2.G were combined with 200 μL Opti-MEM (Thermo Fischer Scientific Cat. #31985062). 6 μL of Xtremegene HP was added to this mixture, and complex formation was allowed to occur for 15 minutes at room temperature at which point this reaction mixture was added to corresponding HEK293T cells. As stated above, neoantigen constructs encode either the full-length or truncated version of the neojunction-derived peptide. The transfection medium was replaced with fresh DMEM media after 24 hours. HLA-A*02:01 lentiviral transduction and screening was performed first prior to neoantigen lentiviral transduction and screening for streamlined drug selection. Viral supernatant was collected after a subsequent 48 hours, and a functional virus titer was performed on 6-well plates seeded with COS7 or K562 cells at 60-70% confluency. Viral transduction was performed with 3-fold serial dilutions of the virus stock supplemented with 4 μg/mL polybrene. Media was changed 24 hours following viral transduction and replaced with complete media supplemented with blasticidin. Cells were assessed for transduction efficiency after 3-4 days by drug screening. HLA-A*02:01-transduced APCs were cultured in medium treated with 10 μg/mL Blasticidin for approximately 7 days before assessing for cell viability across titers. Neoantigen-lentiviral transduction is subsequently performed, and APCs transduced with both HLA-A*02:01 and neoantigen-expressing constructs are then cultured in medium treated with 3 μg/mL puromycin for approximately 7 days. Cell viability was assessed afterwards across all titer

conditions. Cells were assessed for transduction efficiency after 3-4 days by measuring surface expression of HLA-A2 fluorescence-activated cell sorting (FACS) analysis.

Dose-dependent assessment of TCR reactivity against neoantigen

Specificity of neoantigen-reactive $CD8^+$ T-cells and TCR-transduced T-cells was assessed by human IFNγ (BD Biosciences Cat. #555142), IL-2 (BD Biosciences Cat. #555190), and TNFα ELISA (BD Biosciences Cat. #555212). Assessment of TCR recognition against exogenously introduced neoantigen peptides presented by HLA molecules was conducted by co-culturing T-cells with peptide-pulsed T2 cell conditions. T2 cells are either pulsed with neoantigen peptide of interest at a concentration between 1 pM through 1 μM, decoy peptide, or no peptides for 1 hour at 37°C. Influenza-reactive T-cells are co-cultured against influenza peptide-pulsed T2 cells as a positive control. T-cells and T2 cells were co-cultured in a 96-well round-bottom plate at a concentration of 1×10^5 of each cell type in 200 μL of medium for 16 hours. Supernatant was collected and diluted for cytokine release assays per manufacturer's instructions. ELISA assay readouts were performed on Epoch Microplate Spectrophotometer (input wavelength 450 nm and output wavelength 570 nm) using the BioTek Gen5 Data Analysis software. To characterize the dose-dependent activation of the TCRs in transduced triple-reporter Jurkat76/CD8 cells, we performed flow analysis to assess the level of expression of NFAT-GFP, NFκB-CFP, AP-1-mCherry following 16 hours of co-culture. Similarly, reactivity of TCR-transduced PBMC-derived $CD8^+$ T-cells was evaluated by flow analysis following anti-CD107a (BioLegend, Cat #328620) and anti-CD137 antibody (4-1BB; Biolegend Cat #309804) staining.

In vitro transcription (IVT) synthesis of mRNA

All constructs were subcloned into pcDNA3.1 (Invitrogen, 2520855) and linearized by XhoI restriction enzyme with the plasmid DNA template transcribed downstream from the bacteriophage T7 promoter sequence. For long (> 0.5 kb) and short (< 0.5 kB) transcripts, 1 μg and 0.5 μg of template were used, respectively. Reactions were assembled at room temperature using the mMESSAGE mMACHINE T7 Transcription Kit as per manufacturer's instructions (Invitrogen, 2582905) and incubated at 37°C for 1 hour for long transcripts and 16 hours for short transcripts. Following DNase treatment, a Poly(A) tailing reaction was performed for 1 hour according to the HiScribe T7 ARCA manual (NEB, E2060S). Subsequently, the synthesized mRNA was purified by LiCl precipitation using 70% DEPC-based ethanol. Synthesized mRNA was heat-shocked (70℃, 5mins) with the formaldehyde loading dye to verify quality via gel electrophoresis.

mRNA transfection of HLA-A*02:01, truncated neoantigen, and full-length neojunction-encoding mRNA

Transfection of IVT-synthesized mRNA into COS7 and K562 cells was performed with electroporation using the Neon Transfection System 100 μL Kit (Invitrogen, MPK10096) per manufacturer's instructions. 1 x 10^6 COS7 and K562 cells were washed and resuspended with 100 μL of Neon™ Resuspension Buffer. 5 μg of HLA-A2 and 5 μg of candidate (either the truncated neoantigen sequence or the full-length neojunction sequence) mRNA were added into the cell solution. Electroporation was performed on the Neon NxT Electroporation System (Invitrogen, NEON1). Electroporation of COS7 cells was performed with the following optimized conditions: pulse voltage of 1200 V, width of 30 ms, and 2 pulses. Electroporation of K562 was performed with the following optimized conditions: pulse voltage of 1450 V, width of 10 ms, and 3 pulses. Transfected cells were immediately transferred into warm RPMI with no antibiotics. Aliquots of transfected cells were

retained for validation of HLA-A2 expression by staining with HLA-A2 monoclonal antibody (BB7.2, Thermo Scientific, 17-9876-42) and subsequent flow cytometry analysis.

Evaluation of TCR specificity against endogenously processed and HLA-presented neoantigen

Characterization of neoantigens that are endogenously processed and presented by surface HLA is conducted by co-culturing HLA-A*02:01/neoantigen-transfected COS7 or K562 cells with TCR-transduced T-cells. Similarly, T-cells and COS7/K562 cells were co-cultured in a 96-well flat-bottom plate at a concentration of 1 x 10^5 of each cell type in 200 µL of medium for 16 hours. Supernatant was collected and diluted for cytokine release assays per manufacturer's instructions, and cytokine release levels were assessed with the Epoch Microplate Spectrophotometer and BioTek Gen5 Data Analysis software. In all cytokine release assay experiments, maximum cellular cytokine release per well was determined by the addition of 0.2 µL Cell Activation Cocktail (without Brefeldin A) (BioLegend Cat. #423302) per 100 µL cell solution. Evaluation of endogenously processed and presented neoantigens in glioma cell lines was performed by co-culturing TCR-transduced triple-reporter Jurkat76 cells with glioma cells at a 1:1 E:T ratio (1 x 10^5 per well in a 96-well plate). Flow analysis was performed to assess the level of expression of NFAT-GFP, NFκB-CFP, AP-1-mCherry following 16 hours of co-culture.

HLA-IP and LC–MS/MS

COS-7 cells were co-electroporated with 10 µg of each mRNA encoding HLA-A*02:01 allele and the full-length coding sequence of the mutated GNAS or RPL22 using the Neon Transfection system (100-µl tip, setting: 1,050 V/10 ms/2 pulses). 20×10^6 cells were electroporated per condition and

plated in six-well non-TC plates overnight. For the GMB115 cell line sample, approximately 100×10^6 cells were used. Cells were harvested by incubating with 1 mM EDTA (Millipore Sigma) for 10 minutes at 37 °C. For the immunoprecipitation experiments, cells were lysed in 8 ml of 1% CHAPS (Millipore Sigma) for 1 hour at 4°C, lysates were then spun down for 1 hour at 20,000g and 4°C, and supernatant collected. For the affinity-column based immunopurification of HLA-I ligands, 40 mg of cyanogen bromide–activated–Sepharose 4B (MilliporeSigma) was activated with 1 mM hydrochloric acid (MilliporeSigma) for 30 minutes. Subsequently, 1 mg of W6/32 antibody (Bio X Cell) were coupled to Sepharose in the presence of binding buffer (150 mM sodium chloride, 50 mM sodium bicarbonate, pH 8.3; sodium chloride) for 2 hours at room temperature. Sepharose was blocked for 1 hour with glycine and washed 3 times with PBS. Supernatants of cell lysates were run over the affinity of column through peristaltic pumps at 6 mL/min flow rate overnight at 4°C. HLA complexes and binding peptides were eluted from the column five times using 1% TFA. Peptides and HLA-I complexes were separated using C18 columns (Sep-Pak C18 1 cc Vac Cartridge, 50 mg of sorbent per cartridge, 37–55-μm particle size, Waters). C18 columns were pre-conditioned with 80% ACN (Millipore Sigma) in 0.1% TFA and equilibrated with two washes of 0.1% TFA. Samples were loaded, washed twice with 0.1% TFA and eluted in 300 μl of 30%, 40% and 50% acetonitrile in 0.1% TFA. All three fractions were pooled, dried down using vacuum centrifugation and stored at −80 °C until further processing. HLA-I ligands were isolated by solid-phase extractions using in-house C18 mini-columns. Samples were analyzed by high-resolution/high-accuracy LC–MS/MS (Lumos Fusion, Thermo Fisher Scientific). COS-7 samples we run at DDA mode while GMB115 samples at DIA. MS and MS/MS were operated at resolutions of 60,000 and 30,000, respectively. Only charge states 1, 2 and 3 were allowed. The isolation window was chosen as 1.6 Thomson, and collision energy was set at 30%. For MS/MS, maximum injection time was 100 ms with an automatic gain control of 50,000. MS data were processed using FragPipe. Protein FDR was set at 1%. Oxidization of methionine, phosphorylation

of serine, threonine and tyrosine, as well as N-terminal acetylation were set as variable modifications for all samples. Samples were searched against a database comprising UniProt Cercopithecus aethiops or Uniprot Human reviewed proteins supplemented with human HLA-A*02:01 allele sequence, mutRPL22 and mutGNAS, as well as common contaminants.

Characterization of $CD8^+$ T-cell-mediated anti-tumor reactivity

To determine whether TCR-transduced T-cells were capable of mounting an anti-tumor response, TCR-transduced Jurkat76/CD8 or PBMC-derived $CD8^+$ T-cells were co-cultured with patient-derived GBM or LGG cell lines. CD8+ T-cells were isolated from healthy donor-derived PBMCs using the EasySep™ Human $CD8^+$ T Cell Isolation Kit (STEMCELL Technologies, Cat. # 17953). $CD8^+$ T-cells were then activated with Dynabeads™ Human T-Activator CD3/CD28 for T Cell Expansion and Activation (Thermo Scientific, Cat. #11161D) at a concentration of 25 μL/1 x 10^6 cells. $CD8^+$ T-cells were cultured for 7 days with IL-7 (30 μL/1 x 10^6 cells) supplemented every 2 days. $CD8^+$ T-cells were then LV transduced with neoantigen-specific TCRs with a hybridized murine TCR constant region using the above transduction procedure. This additional step removes the likelihood of TCR α-chain and β-chain mispairing and allows us to evaluate TCR-transduction efficiency by staining with anti-murine TCR constant region antibody (Clone H57-597; BioLegend Cat. #109208). Flow sorting was performed to isolate highly-transduced $CD8^+$ T-cells by selecting for cells stained strongly with anti-CD3 and anti-murine TCR constant region antibody. Sorted transduced $CD8^+$ T-cells were expanded for 7 days before use in co-culture assays. Killing assays were performed using an xCELLigence RTCA S16 Real-Time Cell Analyzer. Tumor cells were cultured in media pre-treated with 100 ng/mL IFNγ (Peprotech, Cat. #300-02) for 48 hours and washed twice with PBS prior to seeding. 1 x 10^4 tumor cells were plated per well in a 96-well E-plate (Agilent), and impedance is read

for 16 hours during incubation. TCR-transduced $CD8^+$ T-cells were introduced to each well at E:T ratio of either 1:1 or 2:1, and tumor-specific killing is measured by changes in cell index over 24-48 hours.

Identification of HLA restricted $CD8^+$ T-cell mediated reactivity against neoantigens

Evaluation of HLA restricted reactivity was performed by perturbing TCR and HLA:peptide interaction with the introduction of anti-HLA antibodies. In dose-dependent immunogenicity assays, T2 cells at a concentration of 1 x 10^5 tumor cells per well of a 96-well plate were washed twice with PBS and incubated for 30 min with blocking anti-HLA antibody (50 μg/well; clone W6/32, Bio X Cell, Cat. #BE0079) or isotype control (50 μg/well; Bio X Cell, Cat. #BE0085) at a total volume of 100 μL. Without any additional washes, T-cells were added in to achieve a final volume of 200 μL. In tumor-killing assays, tumor cells were added to each well of a 96-well E-plate in a total volume of 50 μL for initial seeding. Anti-HLA antibody or isotype control (50 μg/well) is added to each well 30 minutes prior to the addition of T-cells to reach a total volume of 100 μL. T-cells were added in each well to achieve a final volume of 200 μL, and impedance was measured for the following 24-48 hours.

Immune-monitoring of cancer patients expressing mutGNAS-NJ

The presence of mutGNAS specific CD8+ T cells in the circulation of GNAS NJ expressing HLA-A*02 cancer patients was tested by FACS using dual-color HLA-A*02 dextramers loaded with the MS-identified mutGNAS peptide. Patient CD8+ cells were stimulated in vitro with NJ expressing HLA-A*02 matched moDCs for 2 weeks prior the FACS staining. To generate moDCs, HLA-A*02 healthy donor PBMCs were plated in tissue culture flasks at 1 × 106 cells per cm2 in complete media

without cytokines for 2 hours at 37 °C to separate the adherent (monocyte-containing) and non-adherent (T-cell-containing) fractions. The adherent fraction was washed with PBS, and fresh human A/B serum-containing media supplemented with recombinant human IL-4 and GM-CSF (400 IU ml−1) was provided every 3 days. On day 6, moDCs were matured with LPS (Invitrogen) and IFN-γ (Miltenyi Biotec) for 24 hours before transfection. moDCs were electroporated with 100 μg ml−1 of mRNA encoding full-length mutGNAS using the Neon Transfection system (10-μl tip, setting: 1,325 V/10 ms/3 pulses). Patient CD8+ cells were enriched from PBMCs by negative selection (STEMCELL Technologies) and co-cultured with mutGNAS-NJ expressing HLA-A*02 matched moDCs at a 2:1 ratio in non-tissue-treated 24-well plates (FALCON) in the presence of 300 IU ml-1 of IL-2 and 50ng ml-1 of IL-7, IL-15, and IL-21. Cytokines were replenished every 3 days. As a control, similarly isolated and co-cultured HLA-A*02:01 matched CD8+ cells from a healthy donor were used. For dextramer labeling, HLA-A*02 multimers bound to mutGNAS and conjugated to PE or APC were purchased from Immudex. As a specificity control, HLA-A*02 multimers bound to the 9-mer peptide from P53 R175H (HMTEVVRHC) were used. Cells were labeled with dual fluorophore-conjugated dextramers for 15 minutes at room temperature, followed by surface antibodies against CD3-BV785, CD4-BV421, CD8-BV650 (Biolegend) for an additional 15 minutes at 4 °C. Cells were washed twice, stained with the viability dye 7-AAD (biolegend) and acquired on a BD Fortessa X20 flow cytometer.

FACS analysis and antibodies

TCR-transduced cell lines were stained with anti-human TCR α/β (Clone IP26, BioLegend Cat. #306717) and anti-human CD3 antibody (Clone HIT3a, BioLegend Cat. #300307) to assess the surface-level expression of the transduced TCR. $CD8^+$ T-cells were stained with anti-CD107a

(BioLegend, Cat #328620) and anti-CD137 antibody (4-1BB; Biolegend Cat #309804) to assess $CD8^+$ T-cell degranulation and TCR activation, respectively. Viability of cells were assessed with the Zombie Green™ Fixable Viability Kit (BioLegend, Cat. #423111) APCs and patient-derived glioma cell lines were stained with HLA-A2 monoclonal antibody (Clone BB7.2, Thermo Fisher Scientific Cat. #17-9876-42). Approximately $1x10^6$ cells per 100 µL FACS buffer (PBS supplemented with 1% BSA (Sigma Aldrich Cat. #L6529)) is incubated with one test volume of antibody for 20 minutes as indicated by the manufacturer. Stained cells were washed once with FACS buffer before resuspension to a concentration of $4x10^5$ cells per 100 µL FACS buffer. Cells were then analyzed with the Attune NxT flow cytometer (Thermo Fischer Scientific). Please have Inaki add patient immune-monitoring methods here.

Gene set enrichment analysis

Differential gene expression of TCGA, GTEx, and UCSF GBM/LGG RNA-sequencing was performed and quantified using DESeq2.[104] Only genes with an absolute fold change > 1.5 and a Benjamini-Hochberg adjusted *P* value <0.05 called by DESeq2 were considered to be differentially expressed.[105] Pre-ranked gene set enrichment analysis (GSEA)[106] was carried out by ranking genes with the product of their fold-change sign and the -log10(adjusted *P* value). Disease subtype-specific differential gene analysis: GSEA comparison was performed between *IDH1* mutation subtypes (wildtype *IDH1* and mutant *IDH1*) as well as glioma disease subtypes (wildtype *IDH1* glioblastoma, mutant *IDH1* astrocytoma, and mutant *IDH1* oligodendroglioma). Splicing-related gene sets were selected based on keyword search, and gene sets with an adjusted *P* value <0.05 when comparing two groups are considered differentially enriched. Unbiased hierarchical clustering of differentially enriched gene sets allows the characterization of subgroup-specific upregulated genes. Neojunction

load-specific differential gene analysis: TCGA LGG and GBM samples were ranked according to the total putative neojunctions expressed per sample. High (NJ_{HI}) and low neojunction load (NJ_{LO}) samples within each disease subtype were characterized as the upper and lower 0.10 percentile of ranked samples, respectively. GSEA is carried out between the NJ_{HI} and NJ_{LO} samples of each disease subgroup. Gene sets with a unidirectional fold-change and adjusted *P* value <0.05 were considered to be enriched gene sets associated with neojunction load. Splicing-related gene sets were selected based on keyword searches. Leading edge genes shared across all disease subgroups within the same gene set are defined as enriched genes associated with neojunction load.

Neojunction and splicing-related gene correlation analysis

Selection of mutant *IDH1* upregulated genes was determined by splicing-related genes expressed with a significant ($p < 0.05$) log2fold increase of 1.5 in mutant *IDH1* cases when compared to their wild-type counterpart. Selection of splicing-genes affected by oligodendroglioma-specific loss of chromosomes 1p/19q was determined by chromosome 1p/19q splicing-related genes expressed with a significant ($p < 0.05$) log_2fold decrease of 1.5 in IDH-O cases compared to both IDH-A and IDH-wt cases. Splicing-related genes that were selected for *in vitro* validation was chosen based on previously reported confirmation of aberrant splicing due to their dysregulated expression.[54,55,57,107,108] To determine correlation factors between each of the identified public neojunctions with each splicing-gene of interest, we performed a Pearson correlation analysis against each neojunction and splicing-related gene pair. Neojunctions with the highest positive correlation score against the select mutant *IDH1* upregulated genes (*CELF2*, *ELAVL4*) averaged across all three glioma subtypes were tested in downstream qPCR assays. Similarly, neojunctions with the most negative correlation score against

select chromosome 1p or 19q splicing-related genes downregulated in *IDH1*mut-O cases (*SNRPD2*, *SF3A3*) averaged across all three glioma subtypes were also tested in downstream qPCR assays.

AlphaFold2 structure predictions

AlphaFold v2.3.2 and its reference databases were installed. AlphaFold was run in multimer mode with default options and the highest rank resulting pdb file was visualized using Pymol. Image was exported with settings "ray 5000,5000" and "png image,dpi=2400".

References

1. Kwok, D. W. *et al.* Tumor-wide RNA splicing aberrations generate immunogenic public neoantigens. *bioRxiv* 2023.10.19.563178 (2023) doi:10.1101/2023.10.19.563178.
2. Brown, M. & Vabret, N. Alternative RNA splicing generates shared clonal neoantigens across different types of cancer. *Nat. Rev. Immunol.* (2024) doi:10.1038/s41577-023-00986-3.
3. Weber, E. W., Maus, M. V. & Mackall, C. L. The Emerging Landscape of Immune Cell Therapies. *Cell* **181**, 46–62 (2020).
4. Touat, M. *et al.* Mechanisms and therapeutic implications of hypermutation in gliomas. *Nature* **580**, 517–523 (2020).
5. Russo, M. *et al.* Adaptive mutability of colorectal cancers in response to targeted therapies. *Science* **366**, 1473–1480 (2019).
6. Huang, Y., Jiang, X., Liang, X. & Jiang, G. Molecular and cellular mechanisms of castration resistant prostate cancer. *Oncol. Lett.* **15**, 6063–6076 (2018).
7. Kim, C. *et al.* Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell* **173**, 879-893.e13 (2018).
8. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).
9. Strønen, E. *et al.* Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science* **352**, 1337–1341 (2016).
10. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
11. Klebanoff, C. A., Rosenberg, S. A. & Restifo, N. P. Prospects for gene-engineered T cell immunotherapy for solid cancers. *Nat. Med.* **22**, 26–36 (2016).

12. Tretter, C. *et al.* Proteogenomic analysis reveals RNA as a source for tumor-agnostic neoantigen identification. *Nat. Commun.* **14**, 4632 (2023).

13. Kahles, A. *et al.* Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **34**, 211-224.e6 (2018).

14. Dvinge, H. & Bradley, R. K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **7**, 45 (2015).

15. Smart, A. C. *et al.* Intron retention is a source of neoepitopes in cancer. *Nat. Biotechnol.* **36**, 1056–1058 (2018).

16. Bigot, J. *et al.* Splicing Patterns in SF3B1-Mutated Uveal Melanoma Generate Shared Immunogenic Tumor-Specific Neoepitopes. *Cancer Discov.* **11**, 1938–1951 (2021).

17. Merlotti, A. *et al.* 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).

18. Burbage, M. *et al.* Epigenetically controlled tumor antigens derived from splice junctions between exons and transposable elements. *Sci Immunol* **8**, eabm6360 (2023).

19. Pan, Y. *et al.* IRIS: Discovery of cancer immunotherapy targets arising from pre-mRNA alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2221116120 (2023).

20. Kim, G. B. *et al.* Quantitative immunopeptidomics reveals a tumor stroma-specific target for T cell therapy. *Sci. Transl. Med.* **14**, eabo6135 (2022).

21. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).

22. Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550–563 (2016).

23. Nejo, T. *et al.* Challenges in the discovery of tumor-specific alternative splicing-derived cell-surface antigens in glioma. *Sci. Rep.* **14**, 6362 (2024).

24. Quesnel-Vallières, M., Jewell, S., Lynch, K. W., Thomas-Tikhonenko, A. & Barash, Y. MAJIQlopedia: an encyclopedia of RNA splicing variations in human tissues and cancer. *Nucleic Acids Res.* **52**, D213–D221 (2024).

25. Yamamoto, T. N., Kishton, R. J. & Restifo, N. P. Developing neoantigen-targeted T cell-based treatments for solid tumors. *Nat. Med.* **25**, 1488–1499 (2019).

26. Ku, A. T., Wilkinson, S. & Sowalsky, A. G. Comparison of approaches to transcriptomic analysis in multi-sampled tumors. *Brief. Bioinform.* **22**, bbab337 (2021).

27. Joung, J.-G. *et al.* Genomic Characterization and Comparison of Multi-Regional and Pooled Tumor Biopsy Specimens. *PLoS One* **11**, e0152574 (2016).

28. Yan, T. *et al.* Multi-region sequencing unveils novel actionable targets and spatial heterogeneity in esophageal squamous cell carcinoma. *Nat. Commun.* **10**, 1670 (2019).

29. Jeon, A.-J. *et al.* Multi-region sampling with paired sample sequencing analyses reveals sub-groups of patients with novel patient-specific dysregulation in Hepatocellular Carcinoma. *BMC Cancer* **23**, 118 (2023).

30. Zhai, W. *et al.* Dynamic phenotypic heterogeneity and the evolution of multiple RNA subtypes in hepatocellular carcinoma: the PLANET study. *Natl Sci Rev* **9**, nwab192 (2021).

31. Árnadóttir, S. S. *et al.* Characterization of genetic intratumor heterogeneity in colorectal cancer and matching patient-derived spheroid cultures. *Mol. Oncol.* **12**, 132–147 (2018).

32. Biswas, D. *et al.* A clonal expression biomarker associates with lung cancer mortality. *Nat. Med.* **25**, 1540–1548 (2019).

33. Meiller, C. *et al.* Multi-site tumor sampling highlights molecular intra-tumor heterogeneity in malignant pleural mesothelioma. *Genome Med.* **13**, 113 (2021).

34. Tirosh, I. *et al.* Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature* **539**, 309–313 (2016).

35. Neftel, C. *et al.* An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).

36. Bhaduri, A. *et al.* Outer Radial Glia-like Cancer Stem Cells Contribute to Heterogeneity of Glioblastoma. *Cell Stem Cell* **26**, 48-63.e6 (2020).

37. Venteicher, A. S. *et al.* Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, (2017).

38. Brennan, C. W. *et al.* The somatic genomic landscape of glioblastoma. *Cell* **155**, 462–477 (2013).

39. McKinney, A. M. *et al.* GABP couples oncogene signaling to telomere regulation in TERT promoter mutant cancer. *Cell Rep.* **40**, 111344 (2022).

40. Appin, C. L. *et al.* Whole tumor analysis reveals early origin of the TERT promoter mutation and intercellular heterogeneity in TERT expression. *Neuro. Oncol.* **26**, 640–652 (2024).

41. Mathur, R. *et al.* Glioblastoma evolution and heterogeneity from a 3D whole-tumor perspective. *Cell* **187**, 446-463.e16 (2024).

42. Spain, L. *et al.* Late-Stage Metastatic Melanoma Emerges through a Diversity of Evolutionary Pathways. *Cancer Discov.* **13**, 1364–1385 (2023).

43. Saez, B., Walter, M. J. & Graubert, T. A. Splicing factor gene mutations in hematologic malignancies. *Blood* **129**, 1260–1269 (2017).

44. Zhang, Y., Qian, J., Gu, C. & Yang, Y. Alternative splicing and cancer: a systematic review. *Signal Transduct Target Ther* **6**, 78 (2021).

45. Anczuków, O. & Krainer, A. R. Splicing-factor alterations in cancers. *RNA* **22**, 1285–1301 (2016).

46. Bashari, A., Siegfried, Z. & Karni, R. Targeting splicing factors for cancer therapy. *RNA* **29**, 506–515 (2023).

47. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).

48. Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A. & Skotheim, R. I. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **35**, 2413–2427 (2016).

49. Seiler, M. *et al.* Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. *Cell Rep.* **23**, 282-296.e4 (2018).

50. PCAWG Transcriptome Core Group *et al.* Genomic basis for RNA alterations in cancer. *Nature* **578**, 129–136 (2020).

51. Bettegowda, C. *et al.* Mutations in CIC and FUBP1 contribute to human oligodendroglioma. *Science* **333**, 1453–1455 (2011).

52. Chan, A. K.-Y. *et al.* Loss of CIC and FUBP1 expressions are potential markers of shorter time to recurrence in oligodendroglial tumors. *Mod. Pathol.* **27**, 332–342 (2014).

53. Ajiro, M. *et al.* Therapeutic manipulation of IKBKAP mis-splicing with a small molecule to cure familial dysautonomia. *Nat. Commun.* **12**, 4507 (2021).

54. Piqué, L. *et al.* Epigenetic inactivation of the splicing RNA-binding protein CELF2 in human breast cancer. *Oncogene* **38**, 7106–7112 (2019).

55. Yanaizu, M., Washizu, C., Nukina, N., Satoh, J.-I. & Kino, Y. CELF2 regulates the species-specific alternative splicing of TREM2. *Sci. Rep.* **10**, 17995 (2020).

56. Jones, L. E. *et al.* Patient-derived cells from recurrent tumors that model the evolution of IDH-mutant glioma. *Neurooncol Adv* **2**, vdaa088 (2020).

57. Wyatt, C. D. R. *et al.* A developmentally programmed splicing failure contributes to DNA damage response attenuation during mammalian zygotic genome activation. *Sci Adv* **8**, eabn4935 (2022).

58. Furney, S. J. *et al.* SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* **3**, 1122–1129 (2013).

59. Vaubel, R. A. *et al.* Genomic and Phenotypic Characterization of a Broad Panel of Patient-Derived Xenografts Reflects the Diversity of Glioblastoma. *Clin. Cancer Res.* **26**, 1094–1104 (2020).

60. Edwards, N. J. *et al.* The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *J. Proteome Res.* **14**, 2707–2713 (2015).

61. Bader, J. M. *et al.* Proteomics separates adult-type diffuse high-grade gliomas in metabolic subgroups independent of 1p/19q codeletion and across IDH mutational status. *Cell Rep Med* **4**, 100877 (2023).

62. Wong, D., Lee, T. H., Lum, A., Tao, V. L. & Yip, S. Integrated proteomic analysis of low-grade gliomas reveals contributions of 1p-19q co-deletion to oligodendroglioma. *Acta Neuropathol Commun* **10**, 70 (2022).

63. Blass, E. & Ott, P. A. Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nat. Rev. Clin. Oncol.* **18**, 215–229 (2021).

64. O'Donnell, T. J., Rubinsteyn, A. & Laserson, U. MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. *Cell Systems* **11**, 42-48.e7 (2020).

65. Sarkizova, S. *et al.* A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).

66. Gonzalez-Galarza, F. F. *et al.* Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* **48**, D783–D788 (2020).

67. Arrieta-Bolaños, E., Hernández-Zaragoza, D. I. & Barquera, R. An HLA map of the world: A comparison of HLA frequencies in 200 worldwide populations reveals diverse patterns for class I and class II. *Front. Genet.* **14**, 866407 (2023).

68. Olivier, T., Haslam, A., Tuia, J. & Prasad, V. Eligibility for Human Leukocyte Antigen-Based Therapeutics by Race and Ethnicity. *JAMA Netw Open* **6**, e2338612 (2023).

69. Lo, W. *et al.* Immunologic Recognition of a Shared p53 Mutated Neoantigen in a Patient with Metastatic Colorectal Cancer. *Cancer Immunol Res* **7**, 534–543 (2019).

70. Hsiue, E. H.-C. *et al.* Targeting a neoantigen derived from a common *TP53* mutation. *Science* **371**, (2021).

71. Carreno, B. M. *et al.* Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803–808 (2015).

72. Chen, F. *et al.* Neoantigen identification strategies enable personalized immunotherapy in refractory solid tumors. *J. Clin. Invest.* **129**, 2056–2070 (2019).

73. Ali, M. *et al.* Induction of neoantigen-reactive T cells from healthy donors. *Nat. Protoc.* **14**, 1926–1943 (2019).

74. Chandran, S. S. *et al.* Immunogenicity and therapeutic targeting of a public neoantigen derived from mutated PIK3CA. *Nat. Med.* **28**, 946–957 (2022).

75. Bossi, G. *et al.* Examining the presentation of tumor-associated antigens on peptide-pulsed T2 cells. *Oncoimmunology* **2**, e26840 (2013).

76. Jutz, S. *et al.* Assessment of costimulation and coinhibition in a triple parameter T cell reporter line: Simultaneous measurement of NF-κB, NFAT and AP-1. *J. Immunol. Methods* **430**, 10–20 (2016).

77. Rosskopf, S. *et al.* A Jurkat 76 based triple parameter reporter system to evaluate TCR functions and adoptive T cell strategies. *Oncotarget* **9**, 17608–17619 (2018).

78. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8950–8954 (2000).

79. Sanderson, J. P. *et al.* Preclinical evaluation of an affinity-enhanced MAGE-A4-specific T-cell receptor for adoptive T-cell therapy. *Oncoimmunology* **9**, 1682381 (2020).

80. Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).

81. Cancer Genome Atlas Research Network *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).

82. Marcucci, G. *et al.* IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *J. Clin. Oncol.* **28**, 2348–2355 (2010).

83. Borger, D. R. *et al.* Frequent mutation of isocitrate dehydrogenase (IDH)1 and IDH2 in cholangiocarcinoma identified through broad-based tumor genotyping. *Oncologist* **17**, 72–79 (2012).

84. Boscoe, A. N., Rolland, C. & Kelley, R. K. Frequency and prognostic significance of isocitrate dehydrogenase 1 mutations in cholangiocarcinoma: a systematic literature review. *J. Gastrointest. Oncol.* **10**, 751–765 (2019).

85. Amary, M. F. *et al.* IDH1 and IDH2 mutations are frequent events in central chondrosarcoma and central and periosteal chondromas but not in other mesenchymal tumours. *J. Pathol.* **224**, 334–343 (2011).

86. Glöss, S. *et al.* IDH2 R172 Mutations Across Poorly Differentiated Sinonasal Tract Malignancies: Forty Molecularly Homogenous and Histologically Variable Cases With Favorable Outcome. *Am. J. Surg. Pathol.* **45**, 1190–1204 (2021).

87. Guilmette, J. & Sadow, P. M. High-Grade Sinonasal Carcinoma: Classification Through Molecular Profiling. *Arch. Pathol. Lab. Med.* **143**, 1416–1419 (2019).

88. Mito, J. K. *et al.* Immunohistochemical Detection and Molecular Characterization of IDH-mutant Sinonasal Undifferentiated Carcinomas. *Am. J. Surg. Pathol.* **42**, 1067–1075 (2018).

89. Cairns, R. A. *et al.* IDH2 mutations are frequent in angioimmunoblastic T-cell lymphoma. *Blood* **119**, 1901–1903 (2012).

90. Wang, C. *et al.* IDH2R172 mutations define a unique subgroup of patients with angioimmunoblastic T-cell lymphoma. *Blood* **126**, 1741–1752 (2015).

91. Schumacher, T. *et al.* A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature* **512**, 324–327 (2014).

92. Grassl, N. *et al.* A H3K27M-targeted vaccine in adults with diffuse midline glioma. *Nat. Med.* (2023) doi:10.1038/s41591-023-02555-6.

93. Tran, E. *et al.* Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* **344**, 641–645 (2014).

94. Linnemann, C. *et al.* High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat. Med.* **21**, 81–85 (2015).

95. Kwok, D. W. *et al.* Abstract 895: Tumor-wide neoantigen-specific T-cells infiltrating mutant IDH1 low-grade gliomas and persisting in peripheral blood allow for personalized TCR-based immunotherapies. *Cancer Res.* **83**, 895–895 (2023).

96. Muniz, L., Nicolas, E. & Trouche, D. RNA polymerase II speed: a key player in controlling and adapting transcriptome composition. *EMBO J.* **40**, e105740 (2021).

97. Saldi, T., Cortazar, M. A., Sheridan, R. M. & Bentley, D. L. Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. *J. Mol. Biol.* **428**, 2623–2635 (2016).

98. Stathias, V. *et al.* Drug and disease signature integration identifies synergistic combinations in glioblastoma. *Nat. Commun.* **9**, 5315 (2018).

99. Jermakowicz, A. M. *et al.* The novel BET inhibitor UM-002 reduces glioblastoma cell proliferation and invasion. *Sci. Rep.* **11**, 23370 (2021).

100. Zhang, Y. *et al.* Transduction of human T cells with a novel T-cell receptor confers anti-HCV reactivity. *PLoS Pathog.* **6**, e1001018 (2010).

101. Hosken, N. A. & Bevan, M. J. Defective presentation of endogenous antigen by a cell line expressing class I molecules. *Science* **248**, 367–370 (1990).

102. Britten, C. M. *et al.* The use of HLA-A*0201-transfected K562 as standard antigen-presenting cells for CD8(+) T lymphocytes in IFN-gamma ELISPOT assays. *J. Immunol. Methods* **259**, 95–110 (2002).

103. Suhoski, M. M. *et al.* Engineering artificial antigen-presenting cells to express a diverse array of co-stimulatory molecules. *Mol. Ther.* **15**, 981–988 (2007).

104. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

105. Matsushima, S. *et al.* Chemical induction of splice-neoantigens attenuates tumor growth in a preclinical model of colorectal cancer. *Sci. Transl. Med.* **14**, eabn6056 (2022).

106. Subramanian, A. *et al.* 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).

107. Bowles, K. R. *et al.* ELAVL4, splicing, and glutamatergic dysfunction precede neuron loss in MAPT mutation cerebral organoids. *Cell* **184**, 4547-4563.e17 (2021).

108. Cieśla, M. *et al.* Oncogenic translation directs spliceosome dynamics revealing an integral role for SF3A3 in breast cancer. *Mol. Cell* **81**, 1453-1468.e12 (2021).

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Darwin Kwok
48699E845F40493...
Author Signature

7/16/2024
Date