Proteogenomic Discovery of Neoantigens Facilitates Personalized Multi-antigen Targeted T cell Immunotherapy for Brain Tumors

Samuel Rivero-Hinojosa PhD[1], Melanie Grant PhD[1,2], Aswini Panigrahi PhD[1], Huizhen Zhang MD[1], Veronika Caisova PhD[1], Catherine M. Bollard MD[1,3], Brian R. Rood MD[1].

[1]Center for Cancer and Immunology Research, Children's National Research Institute, Washington DC

[2]Emory University School of Medicine, Department of Pediatrics, Atlanta, GA

[3]George Washington University Cancer Center

Disclosure of Conflicts of Interest

The authors have no conflicts of interest to declare.

ABSTRACT

Neoantigen discovery in pediatric brain tumors is hampered by their low mutational burden and scant tissue availability. We developed a low-input proteogenomic approach combining tumor DNA/RNA sequencing and mass spectrometry proteomics to identify tumor-restricted (neoantigen) peptides arising from multiple genomic aberrations to generate a highly target-specific, autologous, personalized T cell immunotherapy. Our data indicate that novel splice junctions are the primary source of neoantigens in medulloblastoma, a common pediatric brain tumor. Proteogenomically identified tumor-specific peptides are immunogenic and generate MHC II-based T cell responses. Moreover, polyclonal and polyfunctional T cells specific for tumor-specific peptides effectively eliminated tumor cells in vitro. Targeting novel tumor-specific antigens obviates the issue of central immune tolerance while potentially providing a safety margin favoring combination with other immune-activating therapies. These findings demonstrate the proteogenomic discovery of immunogenic tumor-specific peptides and lay the groundwork for personalized targeted T cell therapies for children with brain tumors.

INTRODUCTION

Medulloblastoma (MB) is the most common malignant brain tumor of childhood and carries overall survival rates between 25-75% depending on molecular subgroup, metastatic status, and age at diagnosis[1]. Effective therapy requires intensive chemo- and radiation therapies, leaving survivors with significant long-term burdens including life-altering cognitive deficits. If the tumor recurs after chemo-radiotherapy, there are no standard effective therapies and virtually no long-term survivors. Therefore, there is an urgent need to develop new therapeutics that can augment standard therapies to effectively prevent tumor recurrence without increasing toxicity. A promising strategy is the use of T cells targeting tumor-specific antigens (TSA), which can: (1) actively home to sites of disease, including across the BBB; (2) possess exquisitely sensitive peptide antigen recognition that may differ from their irrelevant counterparts by a single amino acid[2,3]; and (3) mediate continued, life-long protection through generation of immune memory[4].

Strategies to expand T cell populations with specificity for multiple antigens expressed by a range of malignancies have been developed[5,6]. Targeting multiple antigens reduces the possibility of tumor resistance through antigen escape since selection pressure is not applied to a single target[7,8]. In addition, targeting multiple antigens more effectively addresses intra-tumoral antigen heterogeneity. Ex vivo expanded T cells targeting tumor-associated antigens (TAA) derived from differentiation antigens or cancer testis antigens have been evaluated in different cancer types[9-13]; however, these approaches may be limited by central tolerance toward antigens that are not wholly "foreign" as well as by the potential for on-target, off-tumor auto-immune toxicity[14]. When applied to solid tumors, additional challenges arise from the immunosuppressive tumor microenvironment. To enhance efficacy, it will be necessary to combine T cell therapies with immune adjuvants to boost immune activation and subvert the immunosuppressive tumor microenvironment. We propose that developing T cells targeting tumor-specific antigens (TSA), as opposed to tumor-associated antigens (TAA), will potentially increase the potency of tumor antigen-specific T cell products while decreasing the potential for toxicity, especially when administered in combination with immune adjuvants.

To identify sufficient TSA for multi-antigen targeting, it is necessary to expand their sources beyond somatic mutations alone. This is especially true for pediatric cancers which have many fewer mutations than to their adult counterparts[15]. There are two main strategies to identify TSA. The first relies on identification of non-canonical transcriptomic or exome sequencing reads followed by HLA binding prediction and large-scale immunogenicity assays[16,17]. However, this approach results in many false discoveries, relies on error-prone HLA binding prediction algorithms, and ignores the fact that many transcripts are not translated into proteins. The second strategy is the direct immunoprecipitation of MHC-peptide complexes from tumor cells and identification of bound peptide sequences by liquid

chromatography-mass spectrometry (LC-MS/MS) with subsequent matching to exome or transcriptomic reads. This strategy, called ligandomics, is limited by the efficiency of immuno-precipitation, the large amount of tissue required (infeasible for pediatric brain tumors), the need for robust tumor cell MHC expression, the identification of relatively few peptides, and low throughput[18,19] .

To counter these limitations, we present a personalized, low-input (10-15 mg of tumor tissue) proteogenomic approach to identify TSAs resulting from an individual tumor's genomic aberrations and their use to manufacture T cells specific for multiple TSAs. We effectively identify TSAs arising from four types of genomic events: small insertions/deletions, single nucleotide variations (SNVs), fusions and novel splice junctions. To verify that our findings are not simply unannotated normal proteins, we also developed a multi-step strategy to ensure that the novel peptides are not present in normal tissues. This approach succeeds in identifying a mean of tens of neoantigens per tumor, making multi-antigen targeting possible. As a proof-of-principle, we demonstrate that T cells selected and expanded in response to these peptides contain both CD4 and CD8 populations and are immunogenic, as demonstrated by cytokine profiling and robust cytotoxicity *in vitro.* We posit that this tool can be used to identify personalized TSA peptides for the creation of a T cell therapy using autologous TSA peptide-loaded dendritic cells to select and expand autologous T cells.

RESULTS

Low-input proteogenomic workflow identifies multiple neoantigen peptides from individual medulloblastoma tumors

Pediatric brain tumors present a difficult challenge for immunotherapy development given their low mutational burden, location behind the BBB in an immunosuppressive tumor microenvironment, intra-tumoral heterogeneity and the frequently small amount of tumor tissue available for multiple diagnostic demands. We developed a low-input personalized proteogenomic approach for the identification and curation of tumor-specific neoantigens, which can be used to generate T cells, for autologous use, that are specific for multiple neoantigens in pediatric brain tumors (Figure 1).

In order to identify novel tumor-specific genomic events, we obtained 46 freshly frozen tumor tissues and high coverage whole genome sequencing (WGS) and RNA-seq data from the Children's Brain Tumor Network (CBTN) (Supplementary dataset 1). Four different types of tumor-specific genomic events were identified from genomic data: gene fusions, novel splice junctions, small insertions/deletions and single nucleotide variants (SNVs). Gene fusions were called using three different algorithms: STAR-Fusion[20], ericscript[21] and Defuse,[22] with different numbers of intra- and inter-chromosomal fusions detected per tumor, a mean of 1,044 inter- and intra-chromosomal gene fusions per tumor were detected (Supplementary Fig. 1a). Coding SNVs were called on WGS using the GATK pipeline[23], detecting a mean of 0.8 coding mutations per Mb (Supplementary Fig. 1b). As expected, the number of SNVs indicates a low mutational burden as previously described for medulloblastoma[6]. Novel junctions were considered if the junction was supported by more than 5 reads and was not present in the human Ensembl version 84 transcript annotation, detecting a mean of 12 novel junctions per Mb of the human genome (Supplementary Fig. 2). We detected a mean of 39,000 novel genomic events with the majority (37,000) belonging to the novel splice junction category (Fig. 2a). The novel genomic events for each tumor were translated into 1, 3 or 6 frames depending on the event and included in individualized databases (one for each tumor) together with the normal human proteome from UniProt (UP000005640, See Methods section for details).

MHC class I molecules typically present 8-10 amino acid peptides while MHC class II presents 11-30 amino acid peptides[24]. Our translational workflow employs autologous dendritic cells (DCs) to process tumor-specific peptides to the proper length and sequence for efficient MHC binding in the appropriate autologous HLA context. To test whether the lengths of the identified peptides were suitable to generate both MHC class I and II T cell responses after DC processing, we compared Lys-C and trypsin enzymatic digestion of the 7316-3778 tumor lysate. We also induced missed cleavages by reducing the digestion time to obtain longer average peptides, with an intent to yield more potential tumor-specific peptides suitable for MHC class I and II presentation after DCs processing. For this tumor, the Lys-C enzyme, which cleaves

proteins only at lysine, resulted in peptides with a median of 17 amino acids in contrast to a median of 15 for trypsin digestion, which cleaves proteins at lysine and arginine (Supplementary Fig. 3a). These results indicate that Lys-C yielded longer peptides suitable to elicit both MHC Class I and II T cell responses after DCs processing.

We next performed high-resolution LC-MS/MS on the same panel of freshly frozen tumors for which we previously created individualized protein databases (Fig. 1, Supplementary dataset 1) using total protein lysates partially digested with LysC. In addition to the tumors, we also performed LC-MS/MS on five healthy childhood cerebellums for comparison. LS-MS/MS spectra from MB tumors and healthy cerebellums were searched together against the individualized tumor databases described above. Full and partial LysC digest (i.e., cleavage at a minimum of one C-terminal Lys) searches were performed. A total of 241,224 unique peptides were identified across the sample set with an average of 33,576 peptides per tumor with a 1% FDR (Supplementary Fig. 3b, Supplementary dataset 2). In order to reduce the potential for false discovery in our proteogenomic findings, we developed a filtration strategy to remove peptides from annotated and unannotated normal proteins while preserving novel tumor-specific peptides from undiscovered protein-coding loci (Supplementary Fig. 4). We first removed all peptides matched to the human UniProt proteome and all novel peptides found in healthy childhood cerebellum. Additionally, peptides with a length < 8-mer and an Xcorr < 1 were culled. Xcorr (cross correlation) is a measure of the "goodness of fit" of experimental peptide fragments to theoretical spectra created from the predicted b and y ions. Xcorr > 1 was selected based on our experience identifying confident matches between experimental and theoretical spectra. Next, the novel identified peptides were processed through BLASTP[25] to remove exact matches to known proteins from the Human NCBI(GRCh38), RefSeq, UniProt Isoforms proteome (UP000005640), neXtProt, and Ensembl (version 84) protein annotations. Overall, 481 novel peptides were identified, and 17 (3.53 %) peptides were removed as they were also present in normal cerebellar tissues (Figs. 2b and 2c). Finally, we removed peptides that originate from novel genomic events also present in normal healthy tissues from The Genotype-Tissue Expression (GTEx) project (Supplementary dataset 1); 102 (21%) peptides were removed as the same genomic event that originates that peptide were detected in the GTEX RNA-seq collection. After these filtration steps, we identified a total of 362 novel unique peptides across all 46 MBs (Fig. 2b and 2c, Supplementary dataset 3). A total of 230 novel peptides originated from novel junctions, 85 from SNVs, and 49 from gene fusions across all MB tumors analyzed. Thus, novel junctions are the main source of neoantigens identified in medulloblastoma tumors (Fig. 2b and 2c). On a per tumor basis, a mean of 9 novel peptides were identified (range 1-43) with an average of 5.9 originating from novel junctions, 1.1 from gene fusions and 1.8 from SNVs (Fig. 2b and 2c). The vast majority of these tumor specific peptides were non-overlapping between the tumors. These results indicate

that this approach identifies a significant number of high confidence tumor specific peptides from minimal input tissue and that novel splice junctions are the main source of neoantigens in MB tumors[26].

Tumor-specific peptides were validated by retention time and spectral match

We employed three methods to evaluate the validity of the peptides discovered by our approach. First, we correlated the mean retention time (RT) of the identified peptides to their predicted hydrophobicity index (HI). The HI is a relative value that corresponds to the organic solvent concentration at which the peptide elutes from the HPLC system and is proportional to the retention time. The predicted HI for the UniProt matched peptides, calculated with the SSRCalc tool[27], shows a significant positive correlation with the experimental RT as shown in Supplementary Fig. 5 with a distribution along the regression line. Similarly, the predicted HI of the tumor-specific peptides have significant positive correlations with the retention times supporting their accurate amino acid sequence (Supplementary Fig. 5a). Second, we used AutoRT[28] , a retention time prediction algorithm based in deep learning. A learning model was created using all peptides identified in our cohort (MB model) and applied to a subset of random identified uniport peptides and to all specific peptides (Supplementary Fig. 5b). We identified significant positive correlations (p-values shown in the Supplementary Fig. 5b) between predicted retention times and experimental retention times using the MB model for both UniProt and tumor-specific peptides, (Supplementary Fig. 5b, left panel). We next used a completely independent learning model created using the PXD006109 dataset[29], to evaluate retention times of our peptides. Using this independent dataset, we found significant positive correlations (p-values shown in the Supplementary Fig. 5b) between predicted retention times and experimental retention for both UniProt and tumor specific peptides (Supplementary Fig. 5b, right panel). Third, synthetic versions of 7 tumor specific peptides from patient 7316-3778 were analyzed and, as shown in Supplementary Fig. 6, their MS/MS spectra and retention times were identical to those detected endogenously (Supplementary Fig. 6). These results indicate that our “proof of principle” proteomic approach properly identifies novel MB tumor-specific peptides.

Tumor-specificpeptide frequency

Although most of these tumor specific peptides were non-overlapping between the tumors, 18 of 362 (4.97%) were identified in more than one tumor; 12 of them (3.3%) were found in 2 tumors and 6 of them (1.6%) were identified in 5 or more tumors (Supplementary Fig. 7). Only one peptide, NSSVSGIFTFQK, was identified in more than 20% (12 of 46) of the samples. All of these shared peptides result from novel splice junctions. The peptide NSSVSGIFTFQK can arise from a number of different novel junction events in the DDX31 gene. We detected novel alternative splicing between the exon 14 and

exons 17, 18 and 19. In addition, we detected a novel junction between intron 13 and exon 14. This novel junction changes the frame of exon 14, originating the peptide NSSVSGIFTFQK. Interestedly, The splicing between exon 14 and exon 17, 18 and 19 returns to the canonical frame for the DDX31 protein. The annotated splicing between the Exon 14 and 15 will introduce stop codons as consequences of this change in the frame of exon 14.

DDX31 is a DEAD-box RNA helicase conserved across eukaryotes, but it has not been extensively studied. DDX31 was found to be mutated in several group 4 medulloblastomas in a previous sequencing study[30]. Complex rearrangement and focal deletions of the DDX31 gene have also been observed in several Group 4 medulloblastomas; these deletions occur concurrently with amplification of the OTX2 locus, a known medulloblastoma oncogene[31,32]. This finding suggests that DDX31 mutation (either by deletion or truncation) may cooperate with the oncogenic role of OTX2. Linking our findings to previously published work, such a deletion or rearrangement could originate the NSSVSGIFTFQK peptide by modifying the splicing partner of DDX31. In addition, novel peptides arising from novel junctions in the genes CARF, EEA1, LMNB1, LIZIC and VANGL2 have been discovered with lower frequency in 5 out of 46 tumors (Supplementary Fig. 7).

TSA-specific T cells respond to medulloblastoma tumor-specific neoantigens

As a further "proof of principle" to determine the clinical feasibility of using this neoantigen identification approach to create an autologous T cell immunotherapy product from heavily pre-treated patients, we identified a subject (Patient ID: 7316-3778) from whom tumor tissue and blood were available. Employing our proteogenomic pipeline, we identified 25 novel unique peptides. Among these 25 peptides, 3 peptides were also identified in the healthy cerebellum proteome. In addition, we discovered that the genomic events giving rise to 6 novel peptide sequences were present in normal tissues from GTEX. After removal of those normal unannotated peptides, we were left with 17 novel unique peptides derived from this subject's tumor sample. One peptide resulted from a fusion, 2 from SNVs and 14 from novel junctions (Fig. 3a).

To create an autologous Tumor Specific Antigen T cell product (TSA-T) specific for this patient's newly identified tumor-specific peptides, we stimulated their peripheral blood mononuclear cells (PBMC) with antigen presenting cells (dendritic cells; DC) pulsed with 15 out of 17 peptides derived from patient 7316-3778's tumor (two peptides could not be synthesized; Supplementary dataset 4). Peptide-pulsed DC stimulation was repeated on days 7 (2^{nd} stimulation) and day 14 (3^{rd} stimulation). Seven days after the 3rd stimulation, the polyclonality and polyfunctionality of the resultant T cells were evaluated using anti-interferon gamma (IFN-γ) ELISpot assay and intracellular flow cytometric staining for IFN-γ and tumor necrosis factor alpha (TNF-α). TSA-T cell products stimulated with DCs loaded with 13/15 peptides elicited

a statistically significant IFN-γ response (Fig. 3b). This response was reproducible even when using cryopreserved, thawed and rested TSA-T cells as opposed to fresh product.

The TSA-T cell product derived from patient 7316-3778 was polyclonal comprising 93% CD3+, 31% CD8+, 29% CD4+, 9.42% NKT, and 2.11% NK cells. The differentiation and memory status included primarily TEM (CD4+: 44%, CD8+: 21%) and TCM (CD4+: 16%, CD8+: 44%) with minimal TEFF and TSCM populations (Fig, 3c). TSA-T cells primed and expanded with autologous tumor-specific peptides comprised approximately equal proportions of CD8+ (31%) and CD4+ T cells (29%). TSA-T cells showed polyfunctionality producing IFN-γ, TNF-α or both in response to peptide loaded DCs. Polyfunctional responses were observed in both CD8+ (IFN-γ: 3.1%; TNF-α: 2.71%; IFN-γ+ TNF-α+: 1.55%) and CD4+ cells (IFN-γ: 3.25%; TNF-α: 5.70%; IFN-γ+/TNF-α+: 6.66%). As expected, CD8+ and CD4+ T cells did not produce TNF-α and/or IFN-γ in the presence of peptides that were not presented by DCs (Fig. 3d and Supplementary Fig. 8). Thus, antigen presenting cells pulsed with TSA peptides can prime and expand autologous T cells comprising a polyfunctional polyclonal population of CD4 and CD8 cells with robust TEM and TCM fractions (Fig. 3d and Supplementary Fig. 8).

Conventional assays such as anti-IFN-γ ELISpot and intracellular flow cytometry for detecting tumor-specific T cell responses in cancer patients can underestimate the breadth of antigen-specific T cell responses, and do not assess antigen-specific T-cell repertoires[33]. Therefore, we also performed T cell receptor Vβ sequencing to evaluate the expansion of autologous antigen-specific clones. Dominant clones were identified with the top TCR clone accounting for 18% of all unique rearrangements and the top 10 clones making up 43.77% of all TCRs (Fig. 3e and Supplementary dataset 5). Thus, in combination with the ELISpot data, TCR sequencing further supported that the presence of multiple specifically expanded clones within this patient's TSA-T cell product.

TSA-specific T cells can be generated from medulloblastoma cell line-derived neoantigens

The preceding findings demonstrate a proteogenomic approach to identifying novel tumor-specific peptides and their ability to select and expand a mixed lineage, multi-functional, multi-antigen specific autologous TSA-T cell product. The last remaining measure of activity to be demonstrated is the ability of TSA-T cell products to recognize and lyse tumor cells. Because there was no autologous tumor cell line was generated from this subject's tumor tissue, it was necessary to replicate the previous steps using *in vitro* MB cell lines. Using four MB cell lines (D556, MB002, MB004, and D283), we generated high coverage RNA-seq data from poly-A RNA and three distinct types of tumor-specific genomic events were identified from genomic data: gene fusions, novel splice junctions and small insertions/deletions (SNVs). We detected

a mean of 67,000 novel events with the majority (59,000) of them belonging to the novel splice junction category (Fig. 4a). We detected a mean of 6,000 inter- and intra-chromosomal gene fusions per cell line, 0.7 coding mutations per Mb, and 20 novel junctions per Mb of the human genome (Supplementary Fig. 9a, 9b and 9c). These novel genomic events were translated into 1, 3 or 6 frames depending on the event and included in cell line-specific databases together with the normal human proteome from UniProt (See Methods section for details).

Similar to the approach used for identifying TSAs from primary MB tumor tissues, we performed global high-resolution peptide LC-MS/MS on the MB cell lines using total protein lysates after protease digestion. Applying the same proteogenomic cohort described for MB tumors a total of 100,771 unique peptides were identified across datasets with an average of 53,649 peptides per cell line with a 1% FDR (Supplementary Fig. 9d and Supplementary dataset 6). The removal of normal peptides (from UniProt and our BLASTP approach) and unannotated peptides found in the healthy cerebellum proteome reduced the number to 53, 75, 74, and 132 novel peptides remained for the D556, MB002, MB004 and D283 cell lines, respectively (Fig. 4b and 4c). Finally, we removed novel peptides that originate from novel genomic events also present in normal healthy tissues from The Genotype-Tissue Expression (GTEx) project (Supplementary dataset 1 and Fig. 4c). We identified a total of 269 unique novel peptides across the four MB cell lines - 47, 62, 68, and 100 novel peptides in the D556, MB002, MB004 and D283 lines, respectively (Supplementary dataset 7). The vast majority of these were non-overlapping between the cell lines. Moreover, no peptides were found to be in common between cell lines and primary tumor tissues. In regard to the types of genomic events giving rise to novel peptides, 157 originated from novel junctions, 37 from SNVs, and 75 from gene fusions (Fig. 4b and 4c). These results indicate that novel splice junctions are also the main source of neoantigens in MB cell lines.

We then investigated whether neoantigens discovered in the MB002 and D556 cell lines using our proteogenomic strategy would be recognized as immunogenic by HLA-matched donor-derived T cells (Supplementary dataset 8). We stimulated these donor-derived T cells with DCs derived from the same donor and pulsed with the novel tumor-specific peptides identified from the MB002 or D556 MB cell lines (Supplementary dataset 9). We validated these novel MB cell line-specific peptides against synthetic versions and showed that their MS/MS spectra and elution times matched those found in the cell lines (Supplementary Fig. 10 and Supplementary Fig. 11). Seven days after the 3rd stimulation, the TSA-T cells were re-stimulated with DC pulsed with MB cell line-specific peptides and assessed by anti- IFN-γ ELISpot and intracellular flow cytometric staining for IFN-γ and TNF-α) TSA-T cells stimulated with DCs pulsed with peptides discovered from the MB002 cell line were evaluated in three healthy donors. In Donor 1, a significant IFN-γ T cell response to the pool of all the MB002 peptides was observed (p-value=2.01e-4 compared to actin and p-value=1.97e-4 compared to unrelated D556 peptides), while no response was

observed to DMSO, Actin or unrelated (D556 pep) peptides (Fig. 5a). To determine whether this response was MHC-restricted, we incubated MB002 TSA-T cells with pooled MB002 peptide-pulsed DCs in the presence of anti-MHC Class I or class II blocking antibodies. No reduction was observed in the presence of anti-MHC Class I antibody. A significant reduction (80%, p-value = 6.29e-4) occurred in the presence of anti-MHC Class II antibody (Fig. 5a), which correlated with the autologous data that also demonstrated a predominant CD4+ restricted TSA-specific T cell response. To assess the individual contribution of each peptide to the T cell response, we plated MB002 TSA-T cells with DCs loaded with each individual TSA peptide. In Donor 2, a significant response to peptides 6 and 7 was observed, similar in size to the pooled peptides (p-value =2.0e-4, 7.76e-6 and 3.15e-3 for pooled peptides (MB002 Pep), Peptide 6 and Peptide 7 respectively, Fig. 5b and Supplementary Fig. 12b). Lower frequency, but significant specific responses to peptides 6 and 7 were also observed in healthy Donor 3-derived TSA-T (p-value= 0.0096, 0.0048 and 0.0082 for pooled peptides (MB002 Pep), Peptide 6 and 7 respectively, Supplementary Fig. 12a).

We further mapped the MHC class II-restricted responses identified in TSA-T cells that had been generated from healthy Donor 1 using DC pulsed with MB002-derived peptides. TSA-T were stimulated with DCs pulsed with "mini pools" (3-4 peptides/ pool) in the presence of anti-MHC class I and class II blocking antibodies (Supplementary Fig. 12c and 12d). In the presence of MHC-II-blocking antibodies, the frequency of SFU was significantly reduced by approximately 50% in peptide pool 7-10 (p-value=0.040). A similar reduction of 50% was observed in pool 4-6 however this reduction was not significant. In contrast, in the presence of MHC I-blocking antibodies the frequency of SFU in peptide pools 4-6 and 7-10 was not significantly reduced. These data further confirmed that these peptides were recognized in the context of MHC class II mimicking the autologous result. Only the pools containing peptides 6 (pool 4-6) and 7 (pool 7-10) showed robust responses, which were reduced by anti-MHC class II, confirming that this class II-restricted epitope spanned peptides 6 and 7 corresponding to sequence KASELDYITYLSIFDQLFDIPK (Fig. 5A, and Supplementary Fig. 12c and 12d).

To evaluate reproducibility beyond a single tumor cell line, TSA peptides identified from the D556 MB cell line were tested in the same way as MB002-derived TSA peptides in two healthy donors (Donors 4 and 5). IFN-γ release, indicating a positive T cell response, was likewise observed for 2 and 1 (out of 21) D556 peptides in healthy Donor 4 (Supplementary Fig. 13a) and healthy Donor 5 (Supplementary Fig. 13b) derived TSA-T products respectively. Together these results demonstrate that novel peptides identified in MB002 and D556 MB cell lines were able to prime and stimulate TSA-specific T cells in a partially HLA-matched allogeneic setting.

The phenotype of expanded TSA-T products stimulated with MB002 or D556 tumor cell line-derived peptides was evaluated by 11-colour flow cytometry. All populations comprised primarily CD3+ T cells (median: 91%; range: 77-95%) with variable compositions of CD8+ T cells (37%; 17 – 55%), CD4+

T cells (23.5%; 5 - 40%), NKT cells (37%; 18 – 64%), TCRγδ cells (8%; 0.3 – 31%), and NK cells (1.85%; 0.46 –8.5%) (Fig. 5c). The differentiation and memory status were likewise variable across donors with T effector memory (TEM) populations with a range of CD8+ (2.5 – 91%) and CD4+ (8.9 – 67%) T cells. The NKT cell proportions reflects our experience with clinical trial products manufactured using the same methods to generate tumor associated antigens (TAAs)[34]. T central memory (TCM) cells, shown to be important for long-term persistence of adoptively transferred T cells in vivo[35], also varied greatly between donors (CD8+ range: 0.43 – 67%; CD4+ range: 1.9 – 26%). Minimal numbers (< 1%) of T effector cells (TEFF) or stem cell memory T cells (TSCM) were detected (Fig. 5c).

To evaluate the polyfunctionality (TNF-α and IFN-γ-production) of TSA-T cells derived from Donor 2, we stimulated TSA-T with MB002 peptides and intracellularly labeled them with antibodies against IFN-γ and TNF-α (Fig. 5d and Supplementary Fig. 14a). In agreement with the ELISpot (Fig. 5a), Donor 2's CD4+ T cells produced IFN-γ &/or TNF-α in response to pooled MB002 peptides, and peptides 6 and 7 (7.1% and 6.3% respectively; Fig. 5d and Supplementary Fig. 14a). Thus, the intracellular cytokine data corroborated the HLA-blocking ELISpot results, supporting the evidence that the epitope spanning peptides 6 and 7 (sequence KASELDYITYLSIFDQLFDIPK) is class II-restricted.

Similarly, following stimulation with D556 cell line-specific peptides, CD4+ T (but not CD8+) cells derived from healthy donor 4 secreted IFN-γ and TNF-α in response to peptides 6 (2.95%) and 20 (4%) (Supplementary Fig. 14b and 14c). In summary, multiple CD4-mediated responses were observed in healthy donor-derived TSA-T cell products primed and expanded with DCs pulsed with peptides from MB002 and D556 cell lines. Moreover, these data confirm that we have established a robust approach to generate polyclonal and polyfunctional TSA-T cells products recognizing novel tumor specific peptides identified using our proteogenomic strategy.

TSA-T cell products are cytolytic against tumor targets in vitro

The cytolytic function of healthy donor derived TSA-T was evaluated against the MB002 cell line (Fig. 6 and Supplementary Fig. 17). The TSAT product used in this assay was derived from Donor 1 which matched the MB002 cell line at 4 HLA alleles (Supplementary dataset 8). Despite low expression of MHC Class I and Class II on MB002 cells (Supplementary Fig. 19), TSA-T induced significant lysis of MB002 cells, with only 15% of MB002 cells remaining after 94 hours co-culture compared with non-specifically activated T cells (NST; PHA blasts) (80%, p-value=0.0006) and untreated MB002 cells (287%, p-value=0.0002) (Fig. 6a and 6b). To demonstrate the reproducibility of this approach, the assay was repeated in a subsequent experiment using a TSA-T product derived from a different donor (Donor 2), which also matched at 4 HLA alleles, including HLA-C, DRB1 and DRB5 (Supplementary dataset 8). After 96 hours of co-culture with TSA-T, MB002 cells were reduced to 7% of the original population, compared to 57%

in the wells cultured with non-specific T cells (NST) derived from the same donor (Supplementary Fig. 17a and 17b). Thus, in two independent experiments, evaluating TSA-T products derived from two different donors, the specific cytotoxic capacity of the peptide-primed TSA-T was observed, which outperformed that of a non-specific, potentially allogeneic response induced by NST.

DISCUSSION

Current chemoradiotherapy treatments for malignant pediatric brain tumors suffer from insufficient efficacy coupled with significant lifelong sequelae for survivors[36]. Both constitute failures for which innovative solutions need to be envisioned. T cell-based therapies represent a class of interventions with attributes suited to the treatment of brain tumors. T cells efficiently penetrate the blood brain barrier, one of the main challenges for drug development[37-39]. They can home to areas of disease to eliminate small amounts of residual tumor. And unlike conventional therapies, T cells possess an exquisite ability to distinguish tumors from normal cells, thus eliminating a primary source of toxicity[3,40]. However, T cell therapies also have several challenges to be overcome before they can become a standard of care. Solid tumors create an immunosuppressive microenvironment characterized by exhaustion-inducing checkpoint molecules, anti-inflammatory cytokines and inhibitors of T cell migration[8]. In addition, the tumor cells themselves can downregulate MHC expression to limit antigen presentation, as well as downregulate the expression of antigens in order to escape immunosurveillance. Combinatorial immunotherapy approaches will be required to overcome these obstacles.

In considering immunoadjuvant therapies designed to magnify T cell responses, it is important to think about why these systems evolved in the first place. An activated immune system is a dual-edged sword with a fine balance to be maintained: too little and the host is unprotected from infection or invasion, while too much activity results in life-threatening inflammation. Further, inappropriate non-specific activity causes autoimmune disease. Immunoadjuvants are relatively blunt tools and thus their application does not represent an increase of the therapeutic index but rather an amplification of the cellular response, for both good and ill. Therefore, the burden of specificity lies with the T cell backbone to which these magnifiers will be applied.

Particular attributes of T cell therapies will be required to increase efficacy while decreasing off-target effects. Indeed, improved efficacy itself is a safety feature in that the more effective the T cell employed, the fewer non-specific immunoadjuvant interventions will be required. First, multi-antigen targeting will be critical to target all subclones in a heterogeneous tumor, while also making it more difficult for the tumor cell to down-regulate the targeted antigens[7]. Epitope spreading will be an important feature of a successful T cell response[41,42], but the likelihood of it taking place will depend on the magnitude of the initial immuno-activation and tumor cell lysis. Second, targeting antigens that are completely specific to the tumor and expressed nowhere else in the body both increases safety and avoids the need to overcome the central immune tolerance that protects most endogenous proteins[43]. Third, autologous T cell products can react to antigens presented by all resident HLA alleles, due to the complete HLA match between tumor

and T cells from the same host, thus eliminating disparities of access inherent in off-the-shelf HLA-restricted products such as peptide vaccines. In addition, T cells expanded from an autologous source likely include cells that have been exposed to the tumor's antigens, although ineffectively, and thus the manufacturing process can benefit from a memory response to those antigens. In our data, the TSA peptides identified in primary tumor cells elicited more robust and ubiquitous immunogenicity by ELISpot than those identified in the cell lines. Explanation for this could include the complete HLA matching and presence of memory T cell responses in the autologous as opposed to the allogenic settings. Fourth, we show that the majority of tumor-specific neoantigens are private to an individual tumor, necessitating a personalized process for neoantigen identification and T cell production. This personalization could also potentially enhance the specificity of the T cell product.

The proteogenomic neoantigen identification workflow presented here represents a refinement over previously published approaches with an intent to mitigate some of the challenges inherent in T cell therapy for pediatric brain tumors. The most mature approaches combine elements of 1) affinity capture of MHC complexes from tumor cells with subsequent LC-MS/MS of bound peptides; 2) searching peptide spectra against custom peptide databases containing aberrant events (usually mutations) present in tumor genomic sequencing[44]; and 3) the use of HLA binding affinity prediction algorithms to select and rank potential neoantigen peptides. Affinity capture of MHC complexes from tumor cells requires a significant amount of tissue, far more than is available from a typical pediatric brain tumor resection. In contrast, our approach requires a very low input of tumor tissue, around 10 mg. Further, the limited efficiency of immunoprecipitation results in very few identified peptides, which hampers the ability to target the multiple antigens necessary to deal with tumor heterogeneity and antigen escape. Our approach yields an average of 9 peptides per clinical tumor sample and those peptides originating from non-coding sequences can be used to predict multiple other tumor-specific peptides. Our approach is also not dependent upon the use of HLA binding prediction algorithms which have been shown to suffer from a high degree of inaccuracy[18]. Rather, our workflow identifies longer tumor-specific peptide substrates and relies on the native peptide processing machinery of autologous dendritic cells to present the optimal epitope for T cell selection and expansion. We make use of custom databases for peptide searches but unlike most preceding efforts, our pipeline incorporates novel junctions, SNVs and fusion events rather than just the more typical coding mutations. This is necessary because pediatric brain tumors, and pediatric cancers in general, have among the lowest mutation rates of all cancers[15]. Indeed, our findings identify novel junctions as the predominant source of neoantigen peptides in medulloblastoma, a finding that we have also reproduced in pediatric low-grade glioma and ependymoma (data not shown). Novel junctions have been predicted to be a source of neoantigens, but this is the first demonstration of their dominance in primary tumor samples[26]. These novel

junctions occur in 6 different types (see Methods) with the most common source of novel peptides being junctions joining 2 non-exon regions (40%). Because some proteomic studies have reported the detection of peptides from non-coding RNAs and their potential use as neoantigens[45,46], we investigated to determine if junctions involving non-coding RNAs could explain a portion of our novel peptides. We found that approximately 20% of junctions involving non-exon regions originate from non-coding RNA transcripts found in the Noncodev6 database[47], linking our findings to previous work.

Our data demonstrates specific CD4+-restricted responses. The well-characterized anti-tumor role of CD8+ T cells has long been prioritized for immunotherapy, however the advantages of transferred cytolytic CD4+ T cells to immune activation are receiving more attention[48-50]. Most commercial peptide mixes for neoantigen adoptive cellular therapy contain 9-mer peptides and are thus optimized for MHC I-presented CD8+ epitopes. The partial enzymatic digestion strategy in our mass spec workflow is designed to identify long peptides which, when processed by autologous dendritic cells into immunogenic epitopes, should increase the selection and expansion of CD4+ cells. Several publications have shown that DCs can efficiently intake, process, and present long synthetic peptides[51-54]. Our results indicate the ability to engage CD4+ T cells, particularly for the TSA T cells targeting MB cell line-derived TSAs, which appear to elicit a preponderance of CD4+-restricted responses. Indeed, the TSA peptides that showed the strongest activation by ELISpot acted through MHC II.

The uniqueness of the majority of the novel tumor peptides we identified indicates that the genomic events that generate these peptides are unlikely be tumorigenic. We cannot disregard the possibility that some of these peptides may contribute to tumor formation as very few high frequency driver events have been identified in medulloblastoma tumors despite intensive genomic study[55,56]. Commonly held driver events in medulloblastoma occur at a relatively low frequency compared with other cancers, particularly adult cancers. The DDX31 finding discussed above may be one such event given the multiple ways in which it can be perturbed – mutation, deletion and now aberrant splicing[30]. Furthermore, the fact that aberrant splice junctions originate the majority of the tumor specific peptides identified points to the possibility that deregulation of splicing is playing a significant role in this tumor type as has been described. For example, recurrent mutations in U1 spliceosomal small nuclear RNAs have been associated with SHH medulloblastoma and correlated with changes in splicing[57]. It is therefore plausible to postulate that these unique novel peptides result from a tumor specific characteristic, such as aberrant splicing machinery, without the specific events themselves playing a role in tumorigenesis (i.e. passenger events).

In summary, our workflow identifies a robust number of neoantigens sourced from multiple types of tumor-specific genomic and transcriptomic events using very low tissue input and employing native immuno-

proteasome processing and presentation machinery to select and expand an autologous personalized T cell therapy. Such a specific, targeted T cell product could make an ideal backbone for the addition of potentiating immunoadjuvants in patients with high-risk cancers such as relapsed/refractory medulloblastoma.

MATERIALS AND METHODS

Cell lines and antibodies

MB002 and MB004 were gifts from Y.J. Cho (Oregon Health and Science University, Portland, OR, United States). D556[58] and D283 (D. Bigner, ATCC) cell lines were maintained in Eagle's Minimum Essential Medium supplemented with 10% fetal bovine serum and 100 U/mL penicillin and streptomycin (ThermoFisher Scientific). MB002 and MB004 cells were maintained in culture medium with 1:1 DMEM/F12 (Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12) and Neurobasal™-A Medium supplemented with non-essential amino acids, Sodium Pyruvate, HEPES, GlutaMax, B27, EGF, FGF, Heparin, LIF, 10% fetal bovine serum (ATCC), and 100 U/mL penicillin and streptomycin (All from ThermoFisher Scientific)). All cell lines were maintained at 37°C with 5% CO2 in a 95% humidified atmosphere. All established cell lines were verified with STR analysis (GRCF, Johns Hopkins). Antibodies used in this study are listed in Supplementary dataset 10.

Clinical tumor samples

Medulloblastoma clinical tumor samples are part of the Children's Brain Tumor Network (formerly CBTTC) study cohort CBTTC_0015a. Additional samples were sourced from the Children's National tumor bank. Informed consent of all patients was obtained under a Children's National Medical Center Institutional Review Board and CBTTC approved protocol. Additional information about clinical samples can be found in Supplementary dataset 1.

DNA/RNA extraction, library preparation and sequencing.

RNA-seq and WES data from tumor samples were provided by CBTN. For the RNA sequencing of the D556, D283, MB002 and MB004 cell lines, total RNA was extracted from cells using the RNAeasy Mini Kit (Qiagen) according to the manufacturer's protocol. Strand-specific poly-A selected RNA libraries were sequenced on Illumina HiSeq platform with 2x150 bp read length to an average of 200M reads per sample by GENEWIZ. All raw genomic data is available upon access request through the Children's Brain Tumor Network (https://cbtn.org/) and can be accessed through the Gabriella Miller Kids First Portal (https://kidsfirstdrc.org/). Additionally, cell lines RNA-seq reads have been deposited in the Sequence Read Archive (SRA) with the accession number PRJNA655511.

Bioinformatics analysis of WGS data

Somatic variant calling was done following the GATK-Mutect2 best practices[23,59,60]. Briefly, Raw read sequences were mapped to the GRCh38 reference human genome with bwa[61] and duplicates were marked with PicardTools (v2.18.1). Indel realignment and base recalibration was done with GATKv3.8. Recalibrated reads were used for variant calls using Mutectv2 following default settings and a panel of normal (PON) including the MB germline WGS cohort (Supplementary dataset 1). For tumors with no available germline WGS, the Mutectv2 tumor-only mode was used. Only variants on coding sequences were called.

Bioinformatic analysis of RNA-seq data

RNA-seq raw reads were mapped to the human reference genome GRCh38 using STAR (v.2.5.1)[62] with 2-pass alignment mode to get better alignment around novel splice junctions and Ensembl release 84 annotations (Homo_sapiens.GRCh38.84.gtf). In the case of D556, D283, MB002 and MB004 cell lines, RNA-seq reads were used for variant calls following the GATK best practices for RNA-seq. Briefly, RNA-seq were aligned to the genome using the 2-pass alignment mode of STAR (v.2.5.1). Next, duplicated reads were marked with MarkDuplicates tool (PicardTools). Then, we used the SplitNCigarReads and IndelRealigner tools from GATKv3.8[23,59,60] to split RNA-seq reads and realign reads around indels. Finally, variants were called using HaplotypeCaller tool (GATKv3.8) and filtered using the VariantFiltration tool (GATK v3.8) with the -window 35 and -cluster 3 options, and Fisher Strand values (FS > 30.0) and Qual by Depth values (QD < 2.0).

Gene Fusion calling

Gene fusions were called using the following software: Defuse[22], ericscript[21] and STAR-Fusion[20]. Default settings were used in all cases.

Database generation

Tumor-specific databases using single nucleotide variations and small insertions/deletions (SNVs), fusions and novel isoform variants were generated for each tumor and cell line. *Creation of Variant Peptide Database*: We used the R package CustomProDB[63] to generate variant peptides resulting from DNA SNVs. Coding Mutect2 calls that passed all the filters were incorporated into the genomic sequences and translated to proteins using the Ensembl release 84 transcript annotations. *Creation of Novel Junction Peptide database*: To generate protein databases from novel splice junctions, we used the R package CustomProDB[63]. This package uses as input a bed12 file with each junction found in the RNA-seq alignment bam file. The junction BED files were derived from RNA-seq alignments using the "junctions

extract" function of regtools[64]. This bed12 file contains the chromosome number, the start and end position of the junction and the block size of each exon; block sizes are calculated based in the longest read spliced, the standard format for a bed12 file. Then, CustomProDB removes any junction that is annotated in the reference annotation transcript file (ENSEMBL release 84 transcript annotation gtf file) and classifies the junctions into 6 types of novel splice junction: 1) junctions that connect two known exons, with two subtypes: a) novel alternative splicing junction where the exons belong to the same gene or b) fusion, where the exons belong to different genes, 2) junctions that connect a known exon and a region that overlaps with a known exon, 3) junctions that connect a known exon and a non-exon region, 4) junctions that connect two regions overlapping known exons, 5) junction that connect a region overlapping a known exon and a non-exon region, and 6) junctions that connect two non-exon regions. The non-exon regions could be anywhere, e.g. in intronic regions of the same gene, intronic regions of different genes or intergenic regions. Finally, each putative novel junction is translated in 3frames using the block size information in the bed file. As an expression cut off, we required at least 5 reads spanning splice junctions; novel junctions with less than 5 reads were not included in the database. *Creation of Fusion Peptide database.* Fusion breakpoint coordinates were extracted from the fusion callers and the resulting fused DNA sequences were translated in 6-frames. Stop to stop protein coding regions with more than 6 consecutive amino acids were included in the database. Finally, tumor SNVs, novel junctions and fusion peptides were merged together with human UniProt proteome (UP000005640). The databases generated contain an average of 259,048 entries, 100,179 of them corresponding to the human UniProt proteome database. The average ORF of the databases is 175.8 amino acids (including human UniProt proteome), the average ORF of each of the event types were 21.9, 47.7 and 959 amino acids for fusions, novel junctions and SNV, respectively. Detailed information for each tumor database can be found in Supplementary dataset 2.

LC-MS analysis of peptides

The tumor cells were lysed in RIPA buffer (Pierce) by homogenization followed by sonication. The lysate was centrifuged at 14,000 rpm for 30 minutes at 4°C, and the cleared supernatant collected. The protein concentration was determined by BCA assay (Pierce) and 100 µg of total protein lysate was processed for each sample. The proteins were extracted with methanol:chloroform, air dried and dissolved in 8M urea followed by dilution to 2M concentration, and digested with sequencing grade LysC enzyme (Thermo Scientific) for 4 hours at 37°C or Trypsin overnight at 37°C. The resulting peptides were desalted and fractionated into 8 fractions using the high-pH fractionation kit (Pierce). The peptide mixtures from each fraction were sequentially analyzed by LC-MS/MS using the nano-LC system (Easy nLC1000) connected to a Q Exactive HF mass spectrometer (Thermo Scientific). The platform is configured with a nano-electrospray ion source (Easy-Spray, Thermo Scientific), Acclaim PepMap 100 C18 nanoViper trap column

(3 µm particle size, 75 µm ID x 20 mm length), and EASY-Spray C18 analytical column (2 µm particle size, 75 µm ID x 500 mm length). The peptides were eluted at a flow rate of 300 nL/min using linear gradients of 7-25 % Acetonitrile (in aqueous phase and 0.1% Formic Acid) for 80 minutes, followed by 45% Acetonitrile for 25 minutes, and static flow at 90% Acetonitrile for 15 minutes. All raw proteomic data is available upon access request through the Children's Brain Tumor Tissue Consortium (https://cbttc.org/).

Mass spectrometry data analysis

The LC-MS/MS data were collected in data-dependent mode switching between one full scan MS mode (m/z 380-1400, resolution 60K, AGC 3e6, max ion time 20 ms), and 10 MS/MS scans (resolution 15K, AGC 1e5, max ion time 120 ms, nCE 27) of the top 10 target ions. The ions were sequenced once and then dynamically excluded from the list for 30 seconds. The MS raw data sets were analyzed using Thermo Proteome Discoverer Software (version 2.3). The spectrum files were recalibrated using Trypsin or LysC digested indexed Human UniProt database, and peptide spectrums were searched against a tumor-specific custom database using the Sequest HT algorithm at precursor mass tolerance of 10 ppm, and fragment mass tolerance of 0.02 Da. Methionine oxidation and N-terminus acetylation were specified as dynamic modifications. For each tumor a fully- and a partially- digested search was performed. Peptides and proteins were filtered using a Percolator at a target FDR of 0.01 and a Xcorr > 1.

Peptide filtering

A highly stringent filtering strategy was developed in order to filter out previously annotated and unannotated but normal peptides (Figure S1). This strategy was divided into 2 steps. (i) To identify previously annotated peptides, we used the BLASTP tool[25] to remove tumor-specific peptides that matched any of the following protein databases: UniProtKB/Swiss-Prot including isoforms, NCBI human non-redundant sequences (including all non-redundant GenBank CDS translations, PDB, SwissProt, PIR and PRF excluding environmental samples from WGS projects) and neXtProt. (ii) To identify unannotated normal peptides, we used 2 different approaches, one based on proteomics and another based on genomics. First, we performed proteomic profiling of 5 healthy childhood cerebellum samples using the same methods as the tumor samples. These proteomic raw files were searched against the tumor-specific databases and each novel/non-annotated peptide identified both in the normal cerebellum and in the tumor tissue was removed, leaving only novel peptides identified exclusively in the tumor tissue. Second, for each novel peptide identified from fusion or junction events, we evaluated if the fusion or junction events were also detected in a collection of related tissue RNA-seq files from the Genotype-Tissue Expression (GTEx)

Project (Supplementary dataset 1). For fusion events, the exact breakpoint genomic coordinates in each arm were compared. For example, if we detected a peptide arising from a fusion event with the breakpoint coordinates chr1:15,908,861 and chr5:38,702,49, and if the same breakpoints were detected in any of the GTEx normal tissues analyzed, this peptide was removed. Similarly, for junctions, exact genomic coordinates were compared. For example, a peptide arising from a junction with coordinates chr9:132618441-132642004 would be removed if the same junction is detected in any of the GTEx tissues used. The GTEx data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424.v2.p1.

Peptide hydrophobicity index prediction.

Peptide sequence specific hydrophobicity index (HI) was calculated with the SSRCalc vQ tool[27]. The parameters were set to 100Å C18 column, 0.1% Formic Acid separation system and only unmodified peptides were included. Observed retention times were collected from Proteome Discoverer PSM files. If a peptide was detected multiple times (multiple psm) the average retention time was used. Retention times were plotted against the predicted HI and fitted to a linear model using the R function “lm”. R squared and p-value was calculated using the same “lm” function.

Retention time prediction

Retention times were predicted using the deep learning algorithm AutoRT[28]. A training model (MB model) was created with all non-modified peptides detected in the cohort using AutoRT default settings. For peptides identified multiple times, an average retention time (RT) was calculated. RT was predicted for 4000 random normal UniProt matched peptides and for all tumor specific peptides identified in the 46 MB tissues. Alternatively, the model used as example in the AutoRT publication (model PXD006109, using pepides data from the PXD006109 dataset[29]) was used to calculate RT for a subset of normal peptides matched to UniProt or for all tumor specific peptides in the 46 MB tumors. Experimental retention times were plotted against the predicted retention times and fitted to a linear model using the R function “lm”. R squared and p-value was calculated using the same “lm” function.

Synthetic peptides

Peptides for spectra validation and T cell stimulation were synthesized by GenScript with >98% purity and TFA removal.

A common approach to the manufacture of antigen specific T cells is to identify open reading frames (ORFs), either novel or annotated, and then generate iterative overlapping peptides attempting to find

antigenic peptide sequences that will bind to the MHC complex in the context of a particular patient's HLA type. This is advantageous where the same peptide mixture can be used "off the shelf" to manufacture a T cell product for every patient irrespective of their HLA type[34]. However, in order to target a personalized unique antigen set with complete HLA specificity, our approach is to instead identify longer peptides and rely upon autologous antigen presenting cells to process them into the proper length and sequence to bind MHC I and II molecules in the proper HLA context. Such protocols using DCs pulsed with overlapping peptide pools (15mers overlapping by 11 amino acids) have been used for the manufacture of tumor associated antigen (TAA) specific T cells that have been used clinically to treat patients with solid tumors[34]. Long synthetic peptides are rapidly and much more efficiently processed by DCs, resulting in an increased presentation to $CD4^+$ and $CD8^+$ T cells. Long synthetic peptides are detected very rapidly in an endolysosome-independent manner after internalization by DCs, followed by proteasome processing, transport and Ag processing-dependent MHC class I and Class II presentation[51-54].

Generation of Antigen Presenting Cells

Dendritic cells (DC) and peripheral blood mononuclear cells (PBMC) from partially HLA-matched healthy donors were derived from 30-40 mL sodium heparin anti-coagulated blood from local donors, commercially available cryopreserved PBMCs (Stemcell Technologies, Vancouver, Canada) and patient 7316-3778. PBMCs were purified on day 0 by Ficoll density gradient centrifugation (SepMate, Stemcell Technologies, Vancouver, Canada) according to the manufacturer's protocol. Red blood cells were lysed in ACK buffer and PBMCs incubated at 37°C, 5% CO_2 (6 well plate, 4 mL DC medium, 10-15e6/well). After 2 hours, the non-adherent fraction was removed by gentle flushing followed by two rinses with PBS to remove lymphocytes. The adherent fraction was cultured in DC medium (CellGenix® GMP DC Medium + 1% L-Glutamine (200 mM) + IL-4 (1000 U/mL; R&D) and GM-CSF (800 U/mL; R&D). The non-adherent cells (NAC) were cryopreserved in freezing medium (50% RPMI-1640, 40% FBS, 10% DMSO). Differentiating monocytes/DCs were fed on day 3 or day 4 by half medium removal and replacement with fresh DC medium + IL-4 and GM-CSF; cytokines at 2X concentration to account for final volume. On day 7, immature DCs were harvested and incubated with cell line/tumor-specific peptides that had been resuspended at 10 µg/ml in DC medium containing IL-4 and GM-CSF. DCs were resuspended in 100 µL of peptide solution in a 15 mL tube with a loose lid (37°C, 5% CO_2). After 4-6 hours, 2.5 mL DC maturation medium was added (CellGenix® GMP DC Medium + L-Glut) + IL-4 + GM-CSF + IL-1β (10 ng/ml) + IL-6 (10 ng/ml) + TNF-α (10 ng/ml) + LPS (30 ng/mL) and DCs were transferred to a 24 well plate for 16-18 hours (37°C, 5% CO_2).

Generation of PHA Blasts

For phytohemagglutinin (PHA) blast generation, PBMC or NAC were stimulated with 5 mg/mL of the mitogen PHA (Sigma-Aldrich) to promote blast formation (PHA blasts). PHA blasts were initially cultured in RPMI-1640 supplemented with 5% human serum (Valley Biomedical, Inc), 2 mmol/L Gluta-Max. Every 1-3 days, PHA blasts were split 1 in 2 and/or fed with medium containing IL-2 (100 U/mL; R&D).

Induction of de novo T cell response in vitro

At a maximum of 16-18 hours following maturation, peptide-loaded DCs were irradiated (25 Gy) and co-cultured with non-adherent cells at a ratio of 1:10 in 24-well plates (1e5 DC: 1e6 NAC). T cell medium for stimulation 1 comprised RPMI-1640 (60%) + Click's Medium (40%) + Human serum (10%) + GlutaMax (1%) + IL-6 (100 ng/mL) + IL-7 (10 ng/mL) + IL-12 (10 ng/mL) + IL-15 (5 ng/mL). Cells were fed with complete RPMI-1640 (no cytokines) on day 3 and day 6 if required based on medium color and cell density. On day 6, a fresh batch of DCs was peptide loaded, matured overnight and used to restimulate the proliferating cells on day 7. Stimulation 2 medium contained IL-7 (10 ng/mL) and IL-2 (100 U/mL). Restimulation with DCs was repeated on day 14 with medium containing IL-2 only. On day 21, expanded cells were harvested and used for ELISpot assays and flow cytometric analysis. Cells used for cytotoxicity assays were either plated fresh or following overnight rest after cryopreservation. Cells used for TCR sequencing were cryopreserved.

Anti-IFN-γ Enzyme-Linked Immuno-spot (ELISpot) Assay

Peptide recognition by expanded cells was assessed by anti-IFN-γ ELISpot. Multi-Screen HTS filter plate membranes (Millipore) were activated with 70% EtOH, washed with PBS, coated with IFN-γ capture antibody (10 mg/mL; Mabtech) and incubated overnight at 4°C. The next day plates were washed with PBS and blocked with T cell medium for 1 hour at 37°C to control for non-specific protein binding. Expanded cells were washed, resuspended at 1e6 cells/mL in T cell medium and 5e4-1e5 cells added to appropriate wells, in the presence or absence of peptide-loaded DCs as appropriate. Pooled or single peptides (100 µL; 0.2 mg/mL) were added to appropriate wells along with actin (100 µL; 0.2 mg/mL; irrelevant peptide control) and Staphylococcal enterotoxin B (SEB; positive control). Plates were incubated for 24 hours at 37°C. Plates were developed by washing 6 times in PBS/0.05%Tween 20 followed by incubation with biotinylated IFN-γ detection antibody (0.5 mg/mL; Mabtech; 2 hours, 37°C). This was followed by a further 6 washes in PBS/0.05%Tween 20 and incubation with avidin DH-coupled biotinylated peroxidase H complex (Vectastain Elite ABC Kit; Vector Laboratories) for 1 hour in the dark at room temperature (RT). Following 3 washes in PBST and 3 washes in PBS, spot formation was detected by incubation with 3-Amino-9-Ethylcarbazole (AEC) substrate for 4 minutes in the dark. Spot-forming units (SFU) were counted

and evaluated by Zellnet Consulting using an automated plate reader system (Zeiss). Recognition of pooled neoantigen peptides and individual peptides was compared against no peptide and actin. In the absence of antigen-presenting cells (APC) a positive T cell response was defined as a minimum of 5 SFU and a fold increase of ≥2.5 over actin. In the presence of APCs, due to increased background, a more stringent positive response was defined as a fold increase of ≥2 over actin or T cell+APC, whichever was higher.

HLA-blocking experiments

HLA-restriction of antigen recognition was tested in anti-IFN-γ ELISpot using autologous DCs loaded with the relevant peptide or without peptide (negative control) and blocking antibodies against HLA class I and II (both Dako). DCs were incubated with peptides overnight and the next day anti-HLA mABs were added to ELISpot plates for 1 hour prior to the addition of expanded cells at a ratio of 1 DC to 50 expanded cells. ELISpot plates were incubated and developed as described above.

T cell phenotyping by flow cytometry

Freshly harvested expanded cells were washed with PBS, FcR blocked and stained for viability (LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit, ThermoFisher Scientific, MA, USA). Cell surface molecules were labeled with optimally titrated mABs, then cells were fixed (Cytofix, BD Biosciences), permeabilized (Perm Wash, BD Biosciences) and labeled with mABs against intracellular targets according to manufacturer's instructions. Cell populations were defined as follows: T central memory (TCM): CD3+ CD4/8+CD45RO+CCR7+CD62L+/-; T effector memory (TEM): CD3+CD4/8+CD45RO+ CCR7-CD62L- ; T effector (TEFF): CD3+CD4/8+CD45RO-CCR7-CD62L-; putative natural killer cell (NK): CD3- CD56+ CD16+/-; putative natural killer T cell (NKT): CD3+CD56+; T stem cell memory (TSCM): CD3+CD4/8+CD45RO-CR7+CD62L+CD95+CD127+. Acquisition was performed on a Beckman Coulter CytoFlex S using CytExpert version 2.2.0.97 software. Analysis was performed using FlowJo v10.5.
The gating tree for detecting CD3+CD4+CD8+TCRγδ+, T_{REG}s, NK and NKT cells in a single panel was as follows (Figure S10A): 1. Time/FSH (events collected during stable Flow and excludes debris). 2. FSC/SSC (cell distribution light scatter based on size and intracellular composition, respectively) 3. FSC- height/FSC- area (pulse geometry allows exclusion of events that fall outside single cell range) 4. Live/Dead Aqua/CD3+ (identifies T cells, putative NK cells and live cells). 5. CD4-BV605/CD8-BV421 (identified CD4+CD8+CD4-CD8-). 6. Gamma delta T cells were classified as CD3+CD4-CD8-TCRγδ+CD56+/- 7. CD4+ cells were further interrogated by bivariate examination of CD25 and CD127 to identify putative T_{REG}s, classified as CD3+CD4+CD25+CD127dim. 8. Putative NKT cells were classified simply as CD3+CD56+ and CD8+ T cells were classified as CD3+CD8+CD56-. 9. Putative NK cells were classified

as CD3-CD56+CD16+/-. The gating tree for detecting T central memory (TCM), T effector memory (TEM), T effector (TEFF) & T stem cell memory (TSCM) cells in a single panel was as follows (Figure S10B): 1-5 identical to S10A. Antigen-experienced cells (CD4+/8+CD45RO+) were further assessed as TCM (CD62L+CCR+/- or TEM (CD62L-CCR7-. (Antigen-inexperienced cells (CD45RO-) were further interrogated for TEFF (CD62L-CCR7-) and TSCM (CD62L+CCR7+CD95+CD127+) status. Where > 10% of events fell on an axis bi-exponential scaling was used to visualize all cells on the plot.

Analysis of IFN-γ and TNF-α release by flow cytometry

Freshly harvested expanded cells were incubated with pooled medulloblastoma cell line/tumor-specific peptides (2 ng/μL; 100 uL/well) and anti-CD28/CD49 (5 μL/well); (FastImmune, BD Biosciences, CA, USA). Unstimulated cells (anti-CD28/CD49 only), and actin-stimulated (anti-CD28/CD49 + actin 200 ng/well) cells served as negative and irrelevant antigen controls. After 2 hours (37°C, 5% CO_2), the protein transport inhibitor Brefeldin A (Golgi Plug, BD Biosciences, CA, USA) was added to inhibit cytokine release from the cells. Cells were incubated for a further 4 hours, then washed with PBS, FcR blocked and stained for viability (LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit, ThermoFisher Scientific, MA, USA). Cell surface molecules were labeled with optimally titrated mABs, fixed (Cytofix, BD Biosciences), permeabilized (Perm Wash, BD Biosciences) and labeled with mABs against intracellular targets according to manufacturer's instructions. Acquisition was performed on a Beckman Coulter CytoFlex S using CytExpert version 2.2.0.97 software. Analysis was performed using FlowJo v10.5.

Cytotoxicity Assay

Cryopreserved, expanded cells were quickly thawed in a 37°C water bath, immediately transferred to 10 ml 37°C RPMI-1640 medium and centrifuged (RT, 400g, 5 minutes) to remove DMSO from the cell suspension. The cell pellet was resuspended in T cell medium and counted using a Luna dual fluorescence automated cell counter (Logos Biosystems). Cells were rested overnight in T cell medium + IL-2 (100 U/mL) at 37°C+5% CO2. Target medulloblastoma tumor cells were switched to T cell medium (no IL-2) for 96 hours prior to commencement of co-culture with T cells to allow tumor cells to acclimatise to T cell medium. This reduced spontaneous tumor cell death at the 24-hour time that resulted from the change of medium. PHA blasts were used fresh following the 7-day generation.

On the day of co-culture initiation target tumor cells and were stained with Cell Trace Violet (ThermoFisher Scientific) cell labeling dye as follows. Working in a hood with the light off, a 5 mM stock solution of CellTrace™ was prepared immediately prior to use according to manufacturer's instructions. Target cells were resuspended at 1e6 cells/mL in PBS (no FBS). CellTrace™ solution was added (1 μL/mL) for a final

working solution of 5 µM. Cells were briefly vortexed to ensure even staining and incubated for 20 minutes at 37°C, protected from light. Following incubation, culture medium (5 X the staining volume) containing at least 1% human serum was added and cells incubated for 5 minutes to remove any free dye remaining in the solution. Cells were pelleted by centrifugation, resuspended at 1e6 cells/mL in fresh pre-warmed medium, and incubated for at least 10 minutes to allow cells to undergo acetate hydrolysis before proceeding with co-culture.

While tumor targets were incubating, expanded T cells were pelleted, counted and resuspended at 1e6 cells/mL in medium+IL-2. T cells were dispensed in triplicate into 4 X 96-well U-bottom plates (D0, 48, and 96 hours) in ratio target:TSA-T 1:10.

Following staining, 1e4 target cells were added to the appropriate wells and medium added for a final volume of 200 µL. Targets only were plated as a background viability control. T cell, tumor cell and blast-only wells were plated for unstained and single stain controls for the viability Cell Trace dyes. OneComp eBeads (Invitrogen) were used for single stain controls for anti-CD3 BV785 (BioLegend; Clone: OKT4).

As soon as possible after plating (D0), and at 48 and 96 hours, co-cultures were stained and fixed as follows. For D556 cells, to halt activation and detach adherent cells, 20 µL/well of 20 mM EDTA in PBS was added, mixed by pipetting, incubated for 15 min at room temperature, then mixed again by pipetting to resuspend adhered cells fully. Plates were centrifuged (RT, 400g, 4 minutes), supernatants aspirated and 5 µL FcR block (Miltenyi; Fc Receptor Blocking Reagent human; 130-059-901) added to cells in residual liquid for 5 minutes to control against non-specific binding. Without washing 100 µL Fixable Live Dead Green viability dye (1/1000; ThermoFisher Scientific) was added and the cells incubated for 15 minutes at RT in the dark. Cells were washed once by adding 100 µL FACS Buffer (PBS + 0.5% FBS) and centrifuging (RT, 400g, 4 minutes). After aspiration, cells were gently pipetted to resuspend the pellet and 50 µL of pre-prepared antibody cocktails added to the residual liquid (1 minute, RT, in the dark). FACS Buffer (50 µL) was added to unstained cells. Following AB labeling, cells were washed once by the addition of 100 µL FACS Buffer and centrifugation (RT, 400g, 4 minutes). Cells were fixed by adding 100 µL 4% paraformaldehyde to the residual liquid, incubating at 4°C for 15 minutes and washed by addition of 100 µL FACS Buffer followed by centrifugation (RT, 400g, 4 minutes). Cells were resuspended in 100 µL FACS Buffer, the plates sealed, foil-wrapped and stored at 4°C until acquisition.

Flow cytometry acquisition and data analysis

Flow cytometry acquisition was performed on a CytoFlex S (Beckman Coulter) equipped with 405 nm, 488 nm, 561 nm and 638 nm lasers and calibrated weekly. The instrument was set to acquire a minimum of 50,000 single, live CD3+ cells on a low flow rate. The time parameter was included to detect fluidics issues during sample acquisition and doublets were excluded using fluorescence peak integral versus height.

Spectral overlap was compensated for using beads labeled with the same mAB used in the panel (OneComp Beads, eBioscience), CellTrace-labeled cells (in cytotoxicity assays) and dead cells (viability dye). Post-acquisition compensation was applied and samples analyzed following application of flow stability, pulse geometry and viability gates to exclude debris, doublets and dead cells prior to analysis of cells of interest. Unstained cells, single-stain positive controls and fluorescence minus one (FMO) controls were used to determine background auto-fluorescence and set positive and negative gates. Data were analyzed using FlowJo version 10 analysis software (BD Biosciences, Ashland, OR). To determine specific T cell cytokine responses, all samples were compared to the respective controls and percent positive and/or fold change (FC) was calculated. In samples with low numbers of positive events, a positive response was defined as a fold change ≥ 10. The gating strategy and sample dot plots are shown in Supplementary Figures S14 (ICS; function), S15A (cell populations/phenotype), S15B (differentiation status) and S17 (cytotoxicity).

Immunosequencing

RNA was extracted from peptide-stimulated T cells, using an RNA Easy mini kit, (Qiagen). TCR Vβ CDR3 sequencing was performed by Adaptive Biotechnologies using the survey level resolution Immunoseq platforms (Adaptive Biotechnologies, Seattle, WA). Analysis and compilation of sequence results was performed on Adaptive software.

DATA VALILABILITY.

RNA-seq and WGS reads from medulloblastoma tumors are available upon access request through the Children's Brain Tumor Network (https://cbtn.org/) and can be accessed through the Gabriella Miller Kids First Portal (https://kidsfirstdrc.org/). Proteomic raw spectra files, databases and identified peptides are available upon access request through the Children's Brain Tumor Network (https://cbtn.org/). Additionally, cell line RNA-seq reads have been deposited in the Sequence Read Archive (SRA) with the accession number PRJNA655511.

ACKNOWLEDGEMENTS

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. This research was conducted using data (WGS and RNA-seq) and tumor specimens made available by The Children's Brain Tumor Network (CBTN). Funding for these experiments was provided by the Chance for Life foundation, The Children's Cancer Foundation, The Jeff Gordon Children's Foundation, The Children's Brain Tumor Foundation, and the Lilabean Foundation.

REFERENCES

1 Shih, D. J. *et al.* Cytogenetic prognostication within medulloblastoma subgroups. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **32**, 886-896, doi:10.1200/JCO.2013.50.9539 (2014).

2 Agudo, J. *et al.* GFP-specific CD8 T cells enable targeted cell depletion and visualization of T-cell interactions. *Nature biotechnology* **33**, 1287-1292, doi:10.1038/nbt.3386 (2015).

3 Mount, C. W. *et al.* Potent antitumor efficacy of anti-GD2 CAR T cells in H3-K27M(+) diffuse midline gliomas. *Nat Med* **24**, 572-579, doi:10.1038/s41591-018-0006-x (2018).

4 Heslop, H. E. *et al.* Long-term outcome of EBV-specific T-cell infusions to prevent or treat EBV-related lymphoproliferative disease in transplant recipients. *Blood* **115**, 925-935, doi:10.1182/blood-2009-08-239186 (2010).

5 Gerdemann, U. *et al.* Cytotoxic T lymphocytes simultaneously targeting multiple tumor-associated antigens to treat EBV negative lymphoma. *Mol Ther* **19**, 2258-2268, doi:10.1038/mt.2011.167 (2011).

6 Weber, G. *et al.* Generation of tumor antigen-specific T cell lines from pediatric patients with acute lymphoblastic leukemia--implications for immunotherapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**, 5079-5091, doi:10.1158/1078-0432.CCR-13-0955 (2013).

7 Chowell, D. *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582-587, doi:10.1126/science.aao4572 (2018).

8 Kvistborg, P. & Yewdell, J. W. Enhancing responses to cancer immunotherapy. *Science* **359**, 516-517, doi:10.1126/science.aar6574 (2018).

9 Fratta, E. *et al.* The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Molecular oncology* **5**, 164-182, doi:10.1016/j.molonc.2011.02.001 (2011).

10 Lee, M. H. *et al.* Expression of cancer-testis genes in brain tumors. *Journal of Korean Neurosurgical Society* **43**, 190-193, doi:10.3340/jkns.2008.43.4.190 (2008).

11 Lee, P. P. *et al.* Characterization of circulating T cells specific for tumor-associated antigens in melanoma patients. *Nat Med* **5**, 677-685, doi:10.1038/9525 (1999).

12 Manzo, T., Heslop, H. E. & Rooney, C. M. Antigen-specific T cell therapies for cancer. *Hum Mol Genet* **24**, R67-73, doi:10.1093/hmg/ddv270 (2015).

13 Mohme, M. & Neidert, M. C. Tumor-Specific T Cell Activation in Malignant Brain Tumors. *Front Immunol* **11**, 205, doi:10.3389/fimmu.2020.00205 (2020).

14 Warren, E. H. *et al.* Therapy of relapsed leukemia after allogeneic hematopoietic cell transplantation with T cells specific for minor histocompatibility antigens. *Blood* **115**, 3869-3878, doi:10.1182/blood-2009-10-248997 (2010).

15 Grobner, S. N. *et al.* The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321-327, doi:10.1038/nature25480 (2018).

16 Blaeschke, F. *et al.* Low mutational load in pediatric medulloblastoma still translates into neoantigens as targets for specific T-cell immunotherapy. *Cytotherapy* **21**, 973-986, doi:10.1016/j.jcyt.2019.06.009 (2019).

17 Hu, Z., Ott, P. A. & Wu, C. J. Towards personalized, tumour-specific, therapeutic vaccines for cancer. *Nat Rev Immunol* **18**, 168-182, doi:10.1038/nri.2017.131 (2018).

18 Bassani-Sternberg, M. *et al.* Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* **7**, 13404, doi:10.1038/ncomms13404 (2016).

19 Zhao, W. & Sher, X. Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes. *PLoS computational biology* **14**, e1006457, doi:10.1371/journal.pcbi.1006457 (2018).

20 Haas, B. J. *et al.* STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv*, 120295, doi:10.1101/120295 (2017).

21 Benelli, M. *et al.* Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics* **28**, 3232-3239, doi:10.1093/bioinformatics/bts617 (2012).

22 McPherson, A. *et al.* deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS computational biology* **7**, e1001138, doi:10.1371/journal.pcbi.1001138 (2011).

23 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* **43**, 491-498, doi:10.1038/ng.806 (2011).

24 Burrows, S. R., Rossjohn, J. & McCluskey, J. Have we cut ourselves too short in mapping CTL epitopes? *Trends Immunol* **27**, 11-16, doi:10.1016/j.it.2005.11.001 (2006).

25 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

26 Shen, L., Zhang, J., Lee, H., Batista, M. T. & Johnston, S. A. RNA Transcription and Splicing Errors as a Source of Cancer Frameshift Neoantigens for Vaccines. *Sci Rep* **9**, 14184, doi:10.1038/s41598-019-50738-4 (2019).

27 Krokhin, O. V. & Spicer, V. Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal Chem* **81**, 9522-9530, doi:10.1021/ac9016693 (2009).

28 Wen, B., Li, K., Zhang, Y. & Zhang, B. Cancer neoantigen prioritization through sensitive and reliable proteogenomics analysis. *Nat Commun* **11**, 1759, doi:10.1038/s41467-020-15456-w (2020).

29 Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J. & Mann, M. BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nature methods* **15**, 440-448, doi:10.1038/s41592-018-0003-5 (2018).

30 Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43-48, doi:10.1038/nature11213 (2012).

31 Adamson, D. C. *et al.* OTX2 is critical for the maintenance and progression of Shh-independent medulloblastomas. *Cancer research* **70**, 181-191, doi:10.1158/0008-5472.CAN-09-2331 (2010).

32 Di, C. *et al.* Identification of OTX2 as a medulloblastoma oncogene whose product can be targeted by all-trans retinoic acid. *Cancer research* **65**, 919-924 (2005).

33 Danilova, L. *et al.* The Mutation-Associated Neoantigen Functional Expansion of Specific T Cells (MANAFEST) Assay: A Sensitive Platform for Monitoring Antitumor Immunity. *Cancer Immunol Res* **6**, 888-899, doi:10.1158/2326-6066.CIR-18-0129 (2018).

34 Hont, A. B. *et al.* Immunotherapy of Relapsed and Refractory Solid Tumors With Ex Vivo Expanded Multi-Tumor Associated Antigen Specific Cytotoxic T Lymphocytes: A Phase I Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **37**, 2349-2359, doi:10.1200/JCO.19.00177 (2019).

35 Restifo, N. P., Dudley, M. E. & Rosenberg, S. A. Adoptive immunotherapy for cancer: harnessing the T cell response. *Nat Rev Immunol* **12**, 269-281, doi:10.1038/nri3191 (2012).

36 Northcott, P. A. *et al.* Medulloblastoma. *Nat Rev Dis Primers* **5**, 11, doi:10.1038/s41572-019-0063-6 (2019).

37 Galea, I. *et al.* An antigen-specific pathway for CD8 T cells across the blood-brain barrier. *J Exp Med* **204**, 2023-2030, doi:10.1084/jem.20070064 (2007).

38 Ransohoff, R. M., Kivisakk, P. & Kidd, G. Three or more routes for leukocyte migration into the central nervous system. *Nat Rev Immunol* **3**, 569-581, doi:10.1038/nri1130 (2003).

39 Sayour, E. J. & Mitchell, D. A. Immunotherapy for Pediatric Brain Tumors. *Brain Sci* **7**, doi:10.3390/brainsci7100137 (2017).

40 Platten, M. & Reardon, D. A. Concepts for Immunotherapies in Gliomas. *Semin Neurol* **38**, 62-72, doi:10.1055/s-0037-1620274 (2018).

41 Hu, Z. *et al.* Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. *Nat Med* **27**, 515-525, doi:10.1038/s41591-020-01206-4 (2021).

42 Lo, J. A. *et al.* Epitope spreading toward wild-type melanocyte-lineage antigens rescues suboptimal immune checkpoint blockade responses. *Science translational medicine* **13**, doi:10.1126/scitranslmed.abd8636 (2021).

43 Han, X. J. *et al.* Progress in Neoantigen Targeted Cancer Immunotherapies. *Front Cell Dev Biol* **8**, 728, doi:10.3389/fcell.2020.00728 (2020).

44 Polyakova, A., Kuznetsova, K. & Moshkovskii, S. Proteogenomics meets cancer immunology: mass spectrometric discovery and analysis of neoantigens. *Expert Rev Proteomics* **12**, 533-541, doi:10.1586/14789450.2015.1070100 (2015).

45 Chong, C. *et al.* Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun* **11**, 1293, doi:10.1038/s41467-020-14968-9 (2020).

46 Laumont, C. M. *et al.* Noncoding regions are the main source of targetable tumor-specific antigens. *Science translational medicine* **10**, doi:10.1126/scitranslmed.aau5516 (2018).

47 Zhao, L. *et al.* NONCODEV6: an updated database dedicated to long non-coding RNA annotation in both animals and plants. *Nucleic acids research* **49**, D165-D171, doi:10.1093/nar/gkaa1046 (2021).

48 Alspach, E. *et al.* MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696-701, doi:10.1038/s41586-019-1671-8 (2019).

49 Borst, J., Ahrends, T., Babala, N., Melief, C. J. M. & Kastenmuller, W. CD4(+) T cell help in cancer immunology and immunotherapy. *Nat Rev Immunol* **18**, 635-647, doi:10.1038/s41577-018-0044-0 (2018).

50 Brightman, S. E., Naradikian, M. S., Miller, A. M. & Schoenberger, S. P. Harnessing neoantigen specific CD4 T cells for cancer immunotherapy. *J Leukoc Biol* **107**, 625-633, doi:10.1002/JLB.5RI0220-603RR (2020).

51 Bijker, M. S. *et al.* Superior induction of anti-tumor CTL immunity by extended peptide vaccines involves prolonged, DC-focused antigen presentation. *Eur J Immunol* **38**, 1033-1042, doi:10.1002/eji.200737995 (2008).

52 Ma, W. *et al.* Long-Peptide Cross-Presentation by Human Dendritic Cells Occurs in Vacuoles by Peptide Exchange on Nascent MHC Class I Molecules. *J Immunol* **196**, 1711-1720, doi:10.4049/jimmunol.1501574 (2016).

53 Menager, J. *et al.* Cross-presentation of synthetic long peptides by human dendritic cells: a process dependent on ERAD component p97/VCP but Not sec61 and/or Derlin-1. *PloS one* **9**, e89897, doi:10.1371/journal.pone.0089897 (2014).

54 Rosalia, R. A. *et al.* Dendritic cells process synthetic long peptides better than whole protein, improving antigen presentation and T-cell activation. *Eur J Immunol* **43**, 2554-2565, doi:10.1002/eji.201343324 (2013).

55 Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311-317, doi:10.1038/nature22973 (2017).

56 Waszak, S. M. *et al.* Spectrum and prevalence of genetic predisposition in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial cohort. *The Lancet. Oncology* **19**, 785-798, doi:10.1016/S1470-2045(18)30242-0 (2018).

57 Suzuki, H. *et al.* Recurrent noncoding U1 snRNA mutations drive cryptic splicing in SHH medulloblastoma. *Nature* **574**, 707-711, doi:10.1038/s41586-019-1650-0 (2019).

58 Aldosari, N. *et al.* Comprehensive molecular cytogenetic investigation of chromosomal abnormalities in human medulloblastoma cell lines and xenograft. *Neuro-oncology* **4**, 75-85, doi:10.1093/neuonc/4.2.75 (2002).

931 59 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
932 generation DNA sequencing data. *Genome research* **20**, 1297-1303, doi:10.1101/gr.107524.110
933 (2010).
934 60 Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome
935 Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11 10 11-33,
936 doi:10.1002/0471250953.bi1110s43 (2013).
937 61 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
938 *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
939 62 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21,
940 doi:10.1093/bioinformatics/bts635 (2013).
941 63 Wang, X. & Zhang, B. customProDB: an R package to generate customized protein databases
942 from RNA-Seq data for proteomics search. *Bioinformatics* **29**, 3235-3237,
943 doi:10.1093/bioinformatics/btt543 (2013).
944 64 Cotto, K. C. *et al.* RegTools: Integrated analysis of genomic and transcriptomic data for the
945 discovery of splicing variants in cancer. *bioRxiv*, 436634, doi:10.1101/436634 (2020).

946

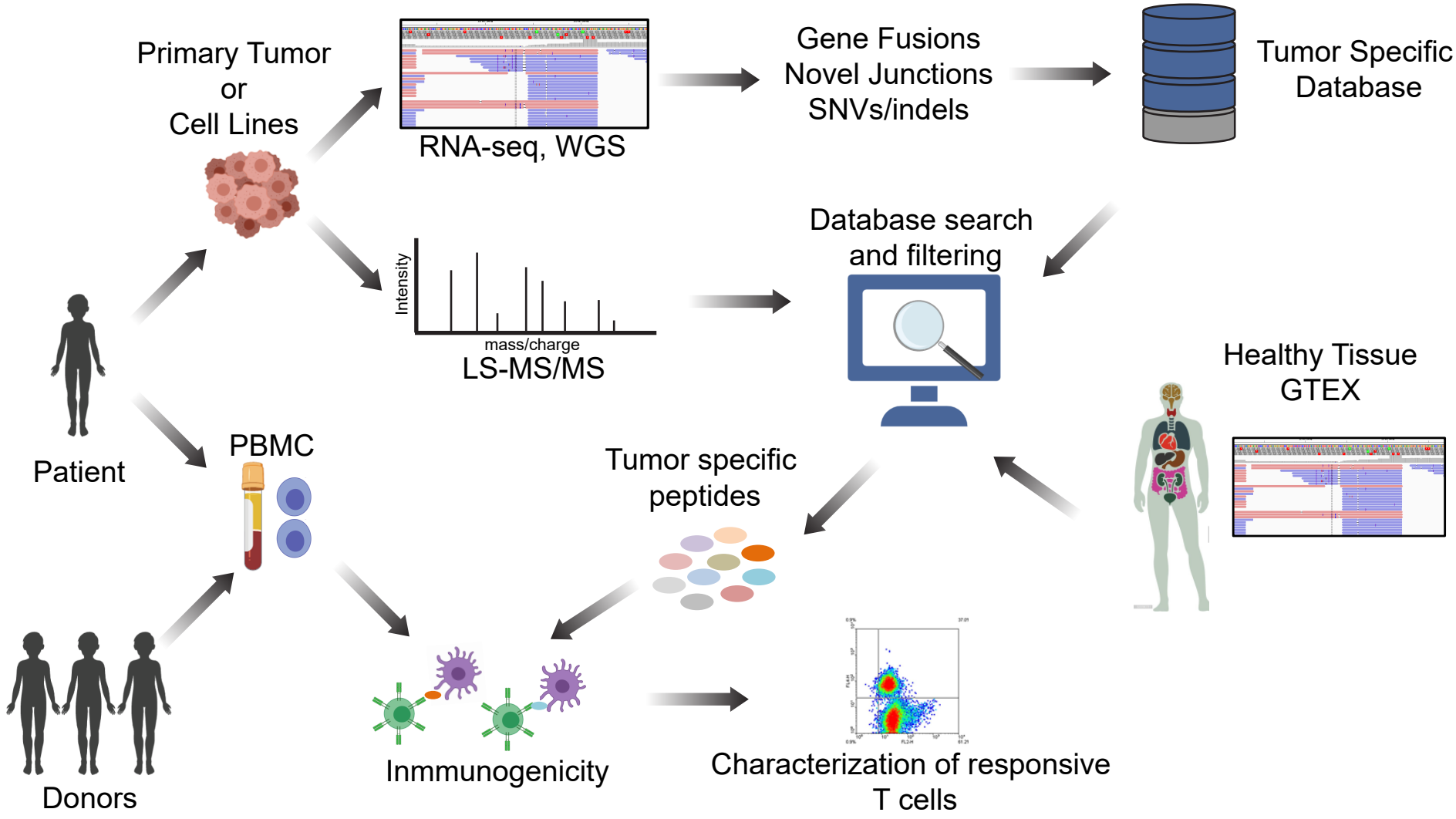
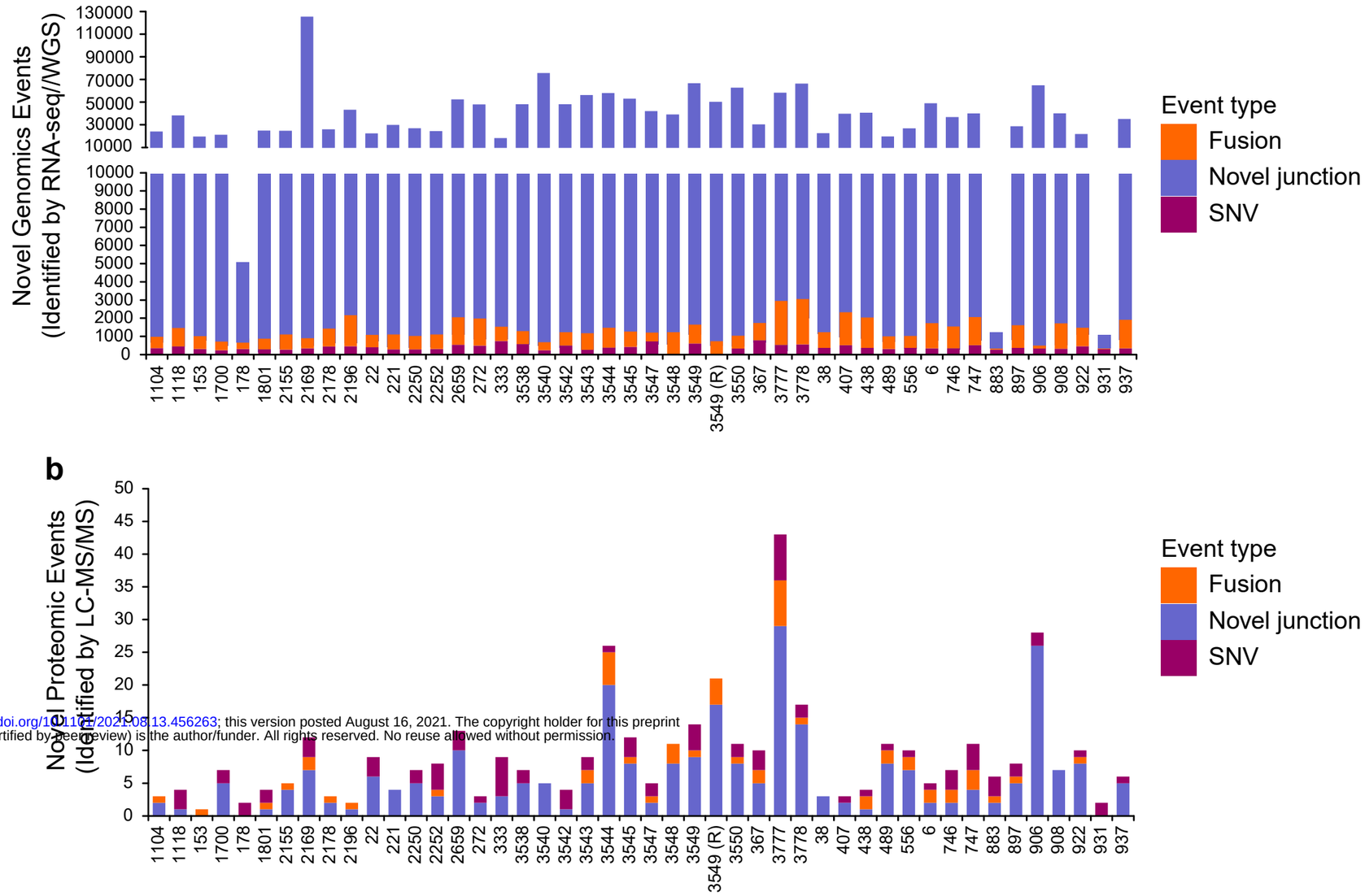

Primary Tumor
or
Cell Lines
RNA-seq, WGS
Gene Fusions
Novel Junctions
SNVs/indels
Tumor Specific
Database
Intensity
mass/charge
LS-MS/MS
Database search
and filtering
Patient
PBMC
Tumor specific
peptides
Healthy Tissue
GTEX
Donors
Inmmunogenicity
Characterization of responsive
T cells

Fig. 1. Proteogenomic approach to identify tumor-specific antigens in pediatric brain tumors. Schematic representation of the entire workflow is shown. Tumor tissue samples were obtained from patients, and WGS and RNA-seq were performed to identify tumor-specific genomic aberrations (SNV/indels, novel junctions and fusions). Protein lysates were subjected to LC-MS/MS shotgun proteomics and spectra were searched against tumor-specific databases originating from tumor WGS and RNA-seq. MS-identified peptides were filtered using genomic and proteomic data from normal tissues to eliminate potential non-annotated normal proteins. Finally, to evaluate tumor-specific novel peptides for immunogenicity, autologous and allogeneic T cells were selected and expanded against the peptides and characterized for phenotype and function.

a



b

c

Fig. 2. Novel genomic and proteomic events in 46 medulloblastoma tumors. The number of (**a**) genomic (detected by RNA-seq/WGS) and (**b**) proteomic (detected by LC-MS/MS) novel events are shown for 46 medulloblastoma tumors. The type of tumor-specific genomic events is indicated: SNV/indels (purple), novel junctions (blue) and gene fusions (orange). The number of identified novel peptides ranged from 1 to 43 peptides per tumor with a mean of 9 peptides per tumor. **c** Tile plot depicting the number of novel peptides identified in medulloblastoma tumors. Black tiles indicate unannotated peptides identified in healthy cerebellum or normal tissues from GTEX. SNVs, fusions and novel junctions are shown in purple, orange, and blue respectively. Each gray tile represents a peptide, although some peptides are found in multiple tumors, the vast majority are tumor-specific.

Fig. 3. Autologous tumor-specific peptides induce a specific IFN-γ response. **a** Tile plot depicting the number of novel peptides identified in patient 7316-3778's tumor. Black tiles indicate the unannotated peptides identified in healthy cerebellar or normal tissues from GTEX. SNVs, fusions and novel junctions are displayed in purple, orange, and blue respectively. **b** Dendritic cells derived from patient 7316-3778 were loaded with peptides identified in the subject's tumor by our proteogenomic pipeline. DCs were co-cultured with non-adherent cells as described. Following 3 stimulations, peptide-specific responses were assessed by anti-IFN-γ ELISpot. In the presence of peptide-loaded DCs, a significant anti-IFN-γ response was observed against 13/15 peptides. One-sided t-test was used to calculate the p-values, n=4. p-values are indicated by stars, *p-value<0.05, **p-value< 0.01, ***p-value<0.001, n.s.=non-significant. **c** Summary data of patient TSAT phenotype, memory and differentiation status. Gating strategy TSAT populations and phenotypes (Figure S15). **d** To assess CD4- and CD8-specific cytokine function, TSAT were incubated in the presence of pooled peptides or peptide-loaded DCs. Summary of intracellular staining data showing specific CD4+ and CD8+ responses to 7316-3778 peptides in the presence of peptide-loaded DCs. Gating strategy for intracellular staining (Figure S15). **e** Results of TCR Vβ CDR3 sequencing on the 7316-3778 TSA T product. Pie chart of the top 10 clonotypes in the TSA T. Clonotypes are listed in Table S5. SFU: spot-forming units; 1 SFU = 1 T cell secreting IFN-γ. Actin: specific peptide control; PMA/Ionomycin positive control. Error bars represent one standard deviation (SD).

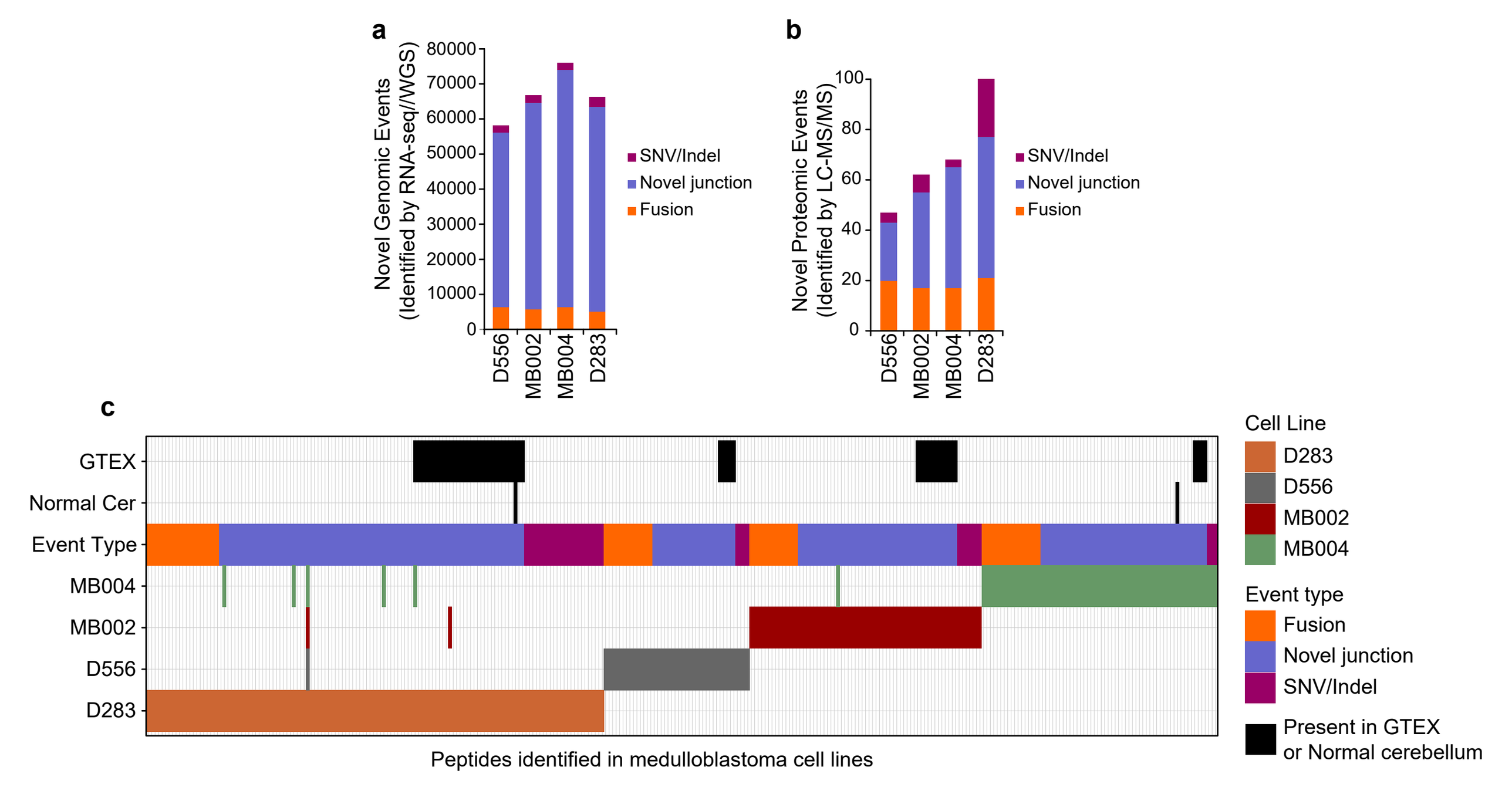
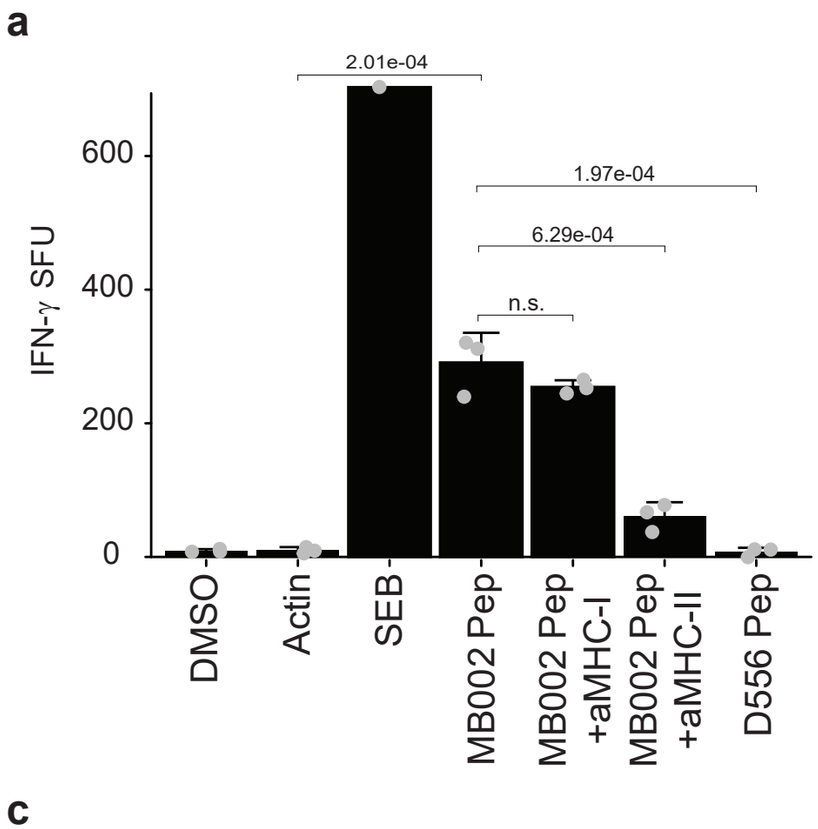
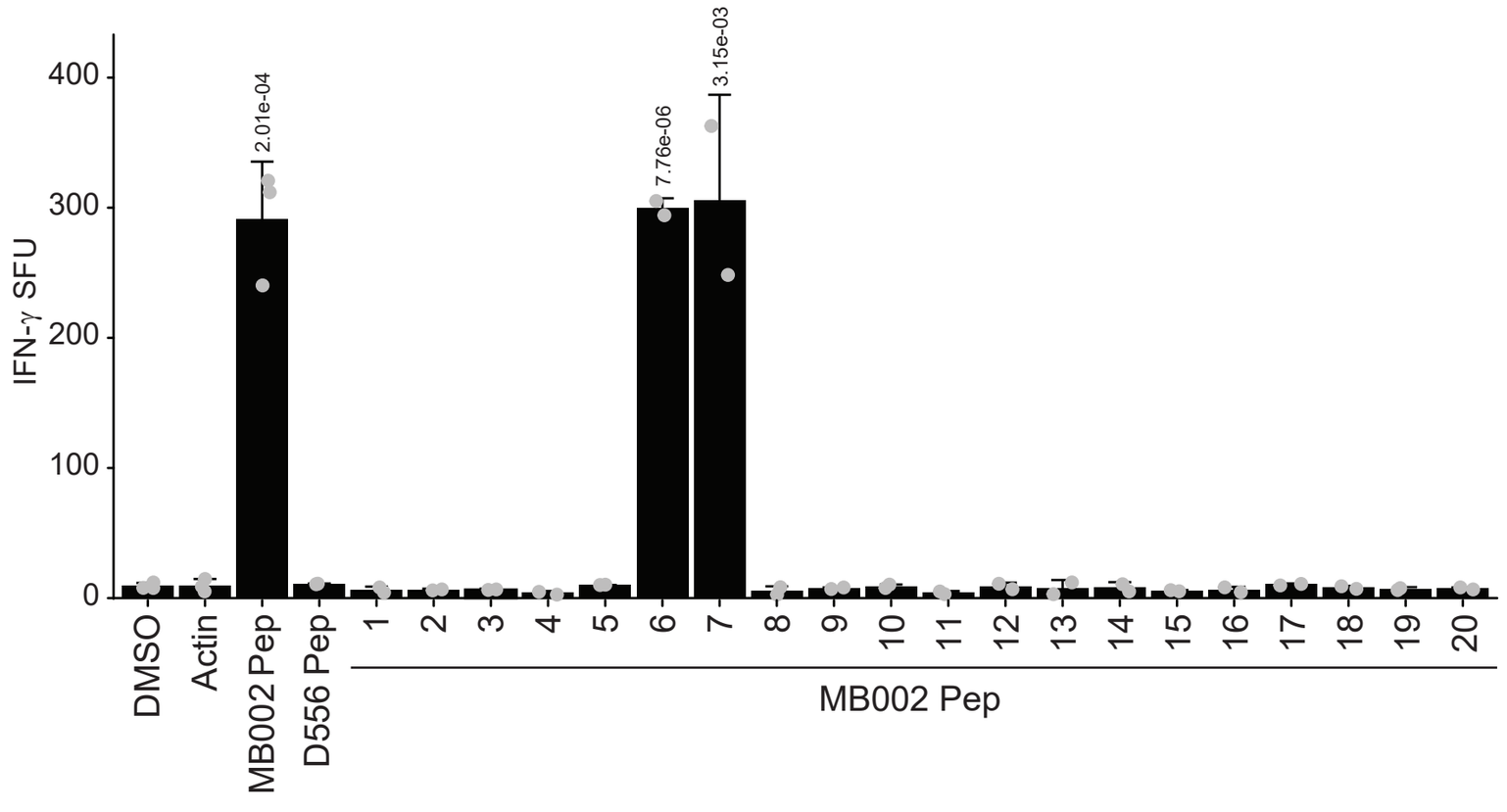
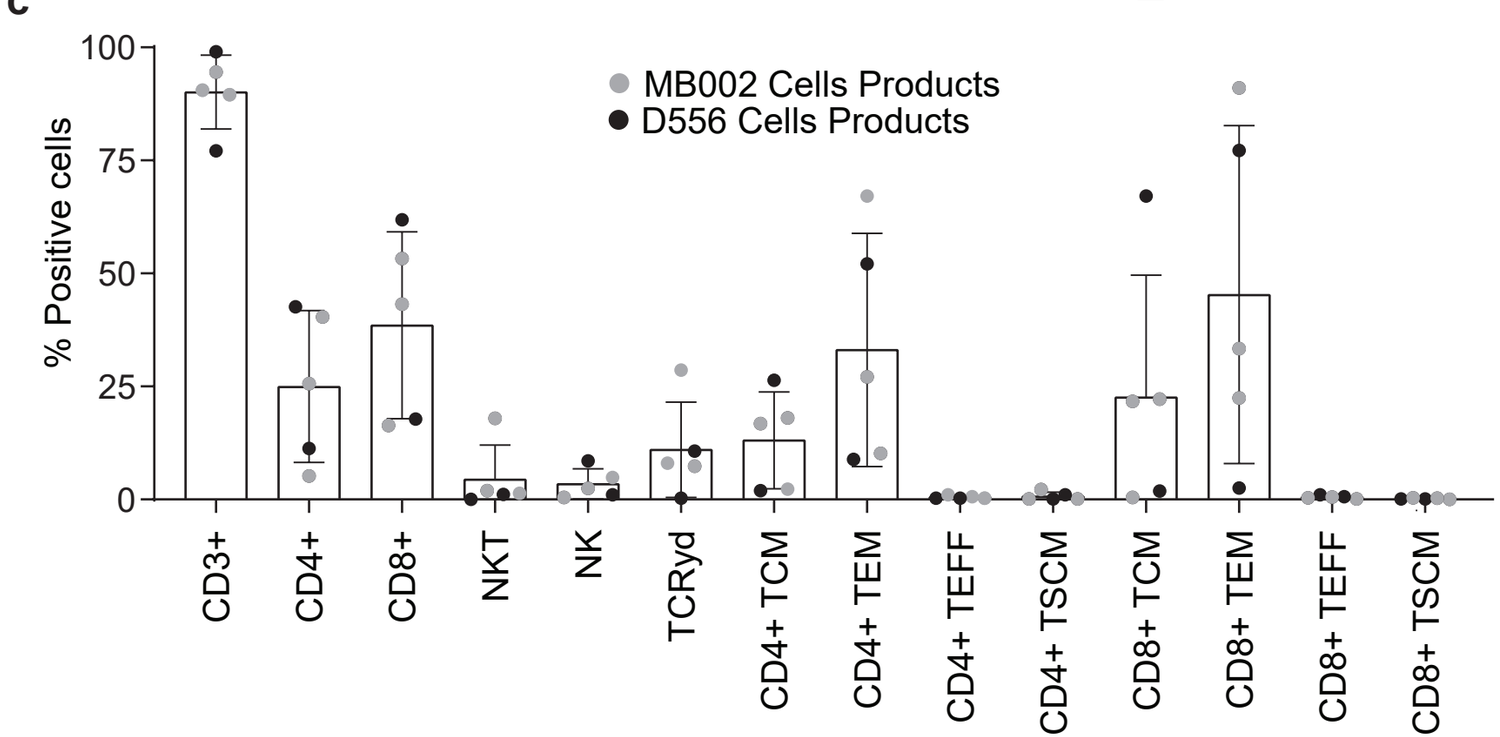

a
Novel Genomic Events
(Identified by RNA-seq//WGS)
80000
70000
60000
50000
40000
30000
20000
10000
0
D556
MB002
MB004
D283
SNV/Indel
Novel junction
Fusion
b
Novel Proteomic Events
(Identified by LC-MS/MS)
100
80
60
40
20
0
D556
MB002
MB004
D283
SNV/Indel
Novel junction
Fusion
c
GTEX
Normal Cer
Event Type
MB004
MB002
D556
D283
Peptides identified in medulloblastoma cell lines
Cell Line
D283
D556
MB002
MB004
Event type
Fusion
Novel junction
SNV/Indel
Present in GTEX
or Normal cerebellum

Fig. 4. Novel genomic and proteomic events in medulloblastoma cells lines. The number of (**a**) genomic (detected by RNA-seq/WGS) and (**b**) proteomic (detected by LC-MS/MS) novel events are shown for D556, MB002, MB004 and D283 medulloblastoma cell lines. The type of tumor-specific genomic events is indicated: SNV/indels (purple), novel junctions (blue) and gene fusions (orange). **c** Tile plot depicting the number of novel peptides identified in medulloblastoma cell lines. Black tiles indicate unannotated peptides identified in healthy cerebellum or normal tissues from GTEX. SNVs, fusions and novel junctions are shown in green, red, and blue respectively.


a
IFN-γ SFU
2.01e-04
1.97e-04
6.29e-04
n.s.
DMSO
Actin
SEB
MB002 Pep
MB002 Pep +aMHC-I
MB002 Pep +aMHC-II
D556 Pep


b
IFN-γ SFU
2.01e-04
7.76e-06
3.15e-03
DMSO
Actin
B002 Pep
D556 Pep
MB002 Pep


c
% Positive cells
MB002 Cells Products
D556 Cells Products
CD3+
CD4+
CD8+
NKT
NK
TCRyd
CD4+ TCM
CD4+ TEM
CD4+ TEFF
CD4+ TSCM
CD8+ TCM
CD8+ TEM
CD8+ TEFF
CD8+ TSCM

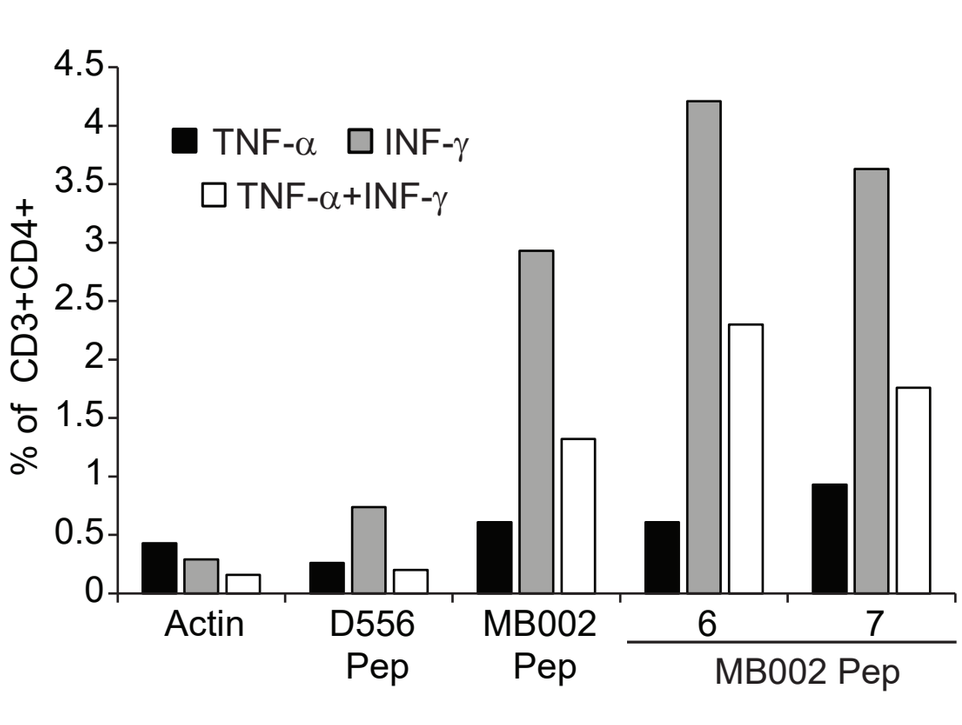

d
% of CD3+CD4+
TNF-α
INF-γ
TNF-α+INF-γ
Actin
D556 Pep
MB002 Pep
6
7
MB002 Pep

Fig. 5. MB002 cell line specific peptides induce peptide-specific, poly-functional Class II-mediated T cell IFN-γ responses *in vitro*. Dendritic cells were loaded with MB002-specific peptides and co-cultured with non-adherent PBMCs. After 3 stimulations, peptide-specific T cell responses were analyzed by anti-IFN-γ ELISpot. **a** Summary data of IFN-γ response to pooled MB002 peptides in healthy donor 1. One-sided t-test was used to calculate the p-values, n=3. p-values are indicated the figure. **b** Summary data showing IFN-γ response to individual MB002 peptides in healthy Donor 2. One-sided t-test was used to calculate the p-values, n=2. p-values are indicated the figure. **c** Summary data of MB002 and D556 TSATs populations and phenotypes (Error bars: mean + SD of 5 independent experiments). **d** Summary data showing CD4+ cytokine response to the pooled MB002 peptides and to the individual peptides 6 and 7 in the healthy donor 2. SFU: spot-forming units; 1 SFU = 1 T cell secreting IFN-γ; TSA T: Tumor-specific antigen T cell; a-MHC-I/II: anti-MHC Class I/II blocking antibodies; MB002 pep: pooled MB002 peptides; D556 pep: pooled D556 peptides; DMSO: dimethyl sulfoxide (peptide solvent; unstimulated control); actin: peptide specificity control; SEB: staphylococcus enterotoxin B (positive control). Gating strategy for intracellular staining (Supplementary Fig. 15). Gating strategy TSA T populations and phenotypes (Supplementary Fig.16). Error bars represent one standard deviation (SD).

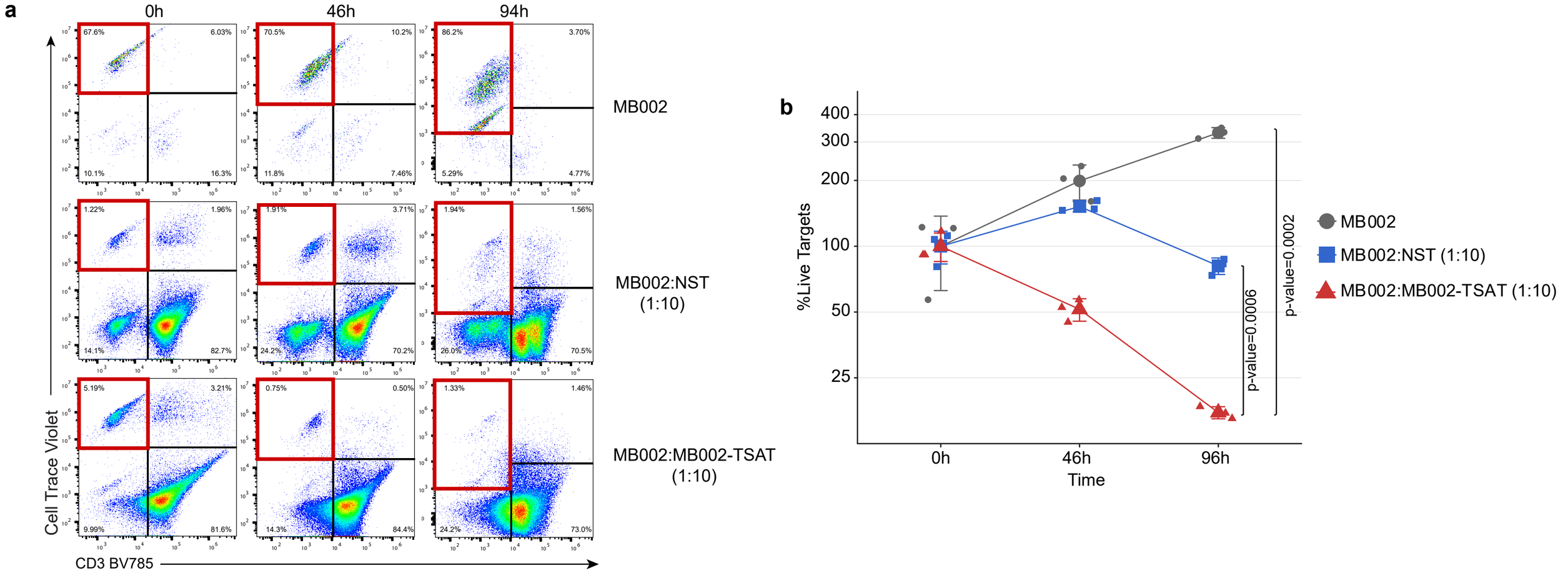

a
0h
46h
94h
MB002
MB002:NST
(1:10)
MB002:MB002-TSAT
(1:10)
Cell Trace Violet
CD3 BV785
67.6%
6.03%
10.1%
16.3%
70.5%
10.2%
11.8%
7.46%
86.2%
3.70%
5.29%
4.77%
1.22%
1.96%
14.1%
82.7%
1.91%
3.71%
24.2%
70.2%
1.94%
1.56%
26.0%
70.5%
5.19%
3.21%
9.99%
81.6%
0.75%
0.50%
14.3%
84.4%
1.33%
1.46%
24.2%
73.0%
b
%Live Targets
400
300
200
100
50
25
0h
46h
96h
Time
p-value=0.0002
p-value=0.0006
MB002
MB002:NST (1:10)
MB002:MB002-TSAT (1:10)

Fig. 6. MB002 TSA T cells specifically lyse partially HLA-matched tumor cells. To assess cytotoxic function, cryopreserved TSA T were thawed and rested overnight prior to plating with tumor targets. At the indicated time points, co-cultures were harvested and acquired as described in Materials & Methods. **a** Representative dot plots from one healthy donor showing proliferation of tumor targets in the absence of TSA T (top row), moderate reduction of tumor targets in the presence of non-specifically activated T cells (PHA blasts; middle row) and robust lysis of tumor targets in the presence of TSA T (bottom row). Lysis was determined based on the disappearance of targets from quadrant 1 (red border). **b** Summary data of a. NST: non-specific T cells (PHA blasts). One-way ANOVA p-values are shown. Values at each time point were normalized to 0 hours (100%). One-sided ANOVA test was used to calculate the p-values, n=3. p-values are indicated the figure. Gating strategy for expanded TSA T tumor cell cytotoxicity assays shown in Supplementary Fig.18. Error bars represent one standard deviation (SD).