Endogenous viral elements constitute a complementary source of antigens for personalized cancer immunotherapy

Christian Garde[1], Michail A. Pavlidis[1], Pablo Garces[1], Emma J. Lange[1], Sri H. Ramarathinam[2], Mateo Sokač[3,4], Kirti Pandey[2], Pouya Faridi[5], Johanne Ahrenfeldt[3,4], Shanzou Chung[2], Stine Friis[1], Daniela Kleine-Kohlbrecher[1], Nicolai J. Birkbak[3,4], Jens V. Kringelum[1], Birgitte Rønø[1], Anthony W. Purcell[2] and Thomas Trolle[1,*]

[1] Evaxion Biotech A/S, Dr Neergaards Vej 5F, Hørsholm, Denmark
[2] Department of Biochemistry and Molecular Biology & Infection and Immunity Program, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia
[3] Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark
[4] Department of Clinical Medicine, Aarhus University, Aarhus, Denmark
[5] Department of Medicine, School of Clinical Scineces, Monash University, Clayton, Victoria, Australia

*Corresponding author: Thomas Trolle, ttr@evaxion-biotech.com

SUMMARY

Cancer immunotherapies have greatly improved the treatment of several cancers, albeit a large fraction of patients still does not benefit from such treatments. Personalized therapies are envisioned to lead to improved outcomes, yet they have largely been leveraging neoepitopes arising from missense somatic mutations, and as such primarily patients with relatively high tumor mutational burden (TMB) are suitable. With the goal to enhance and broaden its application, we present a pipeline, ObsERV, for the design of personalized cancer immunotherapies based on epitopes derived from endogenous retroviral elements (EVEs). We show that EVE-derived peptides are presented as antigens on tumors and can be predicted by ObsERV. Preclinical testing of ObsERV demonstrates induction of durable poly-functional CD4+ and CD8+ T-cell responses as well as long-term tumor protection. We find signs that the EVE antigen burden may stratify low-TMB cancer patients into groups of differential overall survival, indicating that EVEs serve as a additional tumor antigen source. Furthermore, we find through gene set enrichment analysis, that the expression of Myc targets is correlated with the EVE burden for low-TMB patients indicating that Myc may regulate the expression of epigenetically primed EVEs. Finally, we demonstrate in a pan-cancer analysis that the EVE antigen burden does not correlate with TMB and appears to be cancer type specific. As such, EVEs may facilitate and improve the design of personalized immunotherapy especially for low-TMB patients.

Keywords: Endogenous retrovirus, personalized, immunotherapy, neoantigen, T-cell epitopes, biomarker, immunopeptidomics

INTRODUCTION

Neoantigen-based therapies designed to prime or amplify tumor-specific T cell responses have long been envisioned as the key to effective cancer immunotherapy[1,2]. With the regulatory approval of immune checkpoint inhibitors in multiple cancer types, immunotherapy has become standard of care for several advanced solid cancers[3]. Further development has demonstrated that augmented efficacy can be achieved through the combination of immunotherapies[4,5]. Next-generation sequencing of cancerous tumors has enabled the systematic discovery of patient-specific neoantigens and advanced immunotherapy to the stage of personalized medicine[6,7]. The design of personalized immunotherapies has predominately focused on the identification of cancer-specific epitopes arising from somatic mutations. However, its demand for a high tumor mutational burden (TMB) to select efficacious targets has generally limited the application to patients with a relatively high TMB. The hunt for novel efficacious tumor antigen sources was sparked to broaden and potentially enhance the application of personalized immunotherapy[8–12]. Endogenous viral elements (EVEs), which include endogenous retroviruses (ERVs), account for remnants of ancient infectious viruses that have integrated into the mammalian genomes. Under homeostatic conditions, expression of EVE genes in healthy tissues is generally silenced or kept at low levels likely through strict epigenetic regulation[13]. However, it has been documented that, in pathological conditions such as cancer and HIV-1 infection, loss of transcriptional regulation effectively leads to upregulation of the expression of EVE genes[14,15]. EVEs may thus constitute a source of tumor antigens to be harnessed by immunotherapy[9]. Research has indicated that T-cell-mediated immune recognition of human EVE (hEVE) peptides in healthy subjects is virtually absent, while HIV-1 and cancer patients mount functional T-cell responses against hEVE peptides[15–17]. A more recent study used DNA barcode-based pMHC multimers to confirm the absence of CD8+ T-cell-based recognition of hEVE peptides in several healthy subjects, and strikingly demonstrated hEVE peptide-specific CD8+ T-cell recognition in cancer patients[18]. To complement retrospective clinical findings, preclinical studies have demonstrated that vaccination with the EVE peptides AH1 and p15E confers tumor protection accompanied by EVE-specific T-cell responses[19–21] and that EVE-targeting T cells may mediate anti-tumor effect induced by neoepitope immunotherapies[22]. These data indicate that EVE-specific T-cell responses are detectable in the preclinical and clinical setting and may confer tumor protection. It is, however, still to be investigated whether EVEs constitute a general complementary antigen source to be leveraged for personalized cancer immunotherapy. Clinical intervention studies targeting ERVs are in the early stage of testing (NCT03354390), but there has to the authors' knowledge not been conducted any clinical vaccine studies that leverage EVE-derived epitopes.

In the current study we developed a bioinformatics pipeline, ObsERV, for the design of personalized cancer immunotherapy comprising EVE-derived peptides. We demonstrate that EVE-derived peptides are presented on tumors as MHC ligands using state-of-the-art immunopeptidomics and that they can be predicted by ObsERV. Furthermore, ObsERV-designed therapy protected mice from tumor establishment and induced both CD4+ and CD8+ T-cell responses. To support the likelihood of clinical translation, we find that the number of expressed EVEs may be predictive of overall survival of cancer patients with low TMB, suggesting that EVEs represent a complementary tumor antigen source. This is further supported by a pan-cancer analysis showing that the number of expressed EVEs is cancer type specific, yet without any observed correlation to TMB, suggesting that EVEs could facilitate the design of personalized immunotherapy.

RESULTS

EVEs are presented as MHC ligands on tumors and constitute a potential source of antigens for personalized immunotherapies

To investigate the feasibility of utilizing EVE-derived tumor antigens in a personalized immunotherapy, we developed an *in silico* bioinformatics pipeline, ObsERV. ObsERV utilizes RNA sequencing (RNA-seq) and MHC ligand predictions to identify and select EVE peptide sequences that are likely to be presented on MHC molecules (Figure 1A). The first step in the ObsERV pipeline is to leverage RNA-seq to quantify EVE expression levels in the tumor sample. For all expressed EVEs, MHC class I (MHCI) ligand predictions are generated for all 8-11mer peptides and MHC class II (MHCII) ligand predictions are generated for all 15mer peptides. MHC ligand predictions are combined, and MHC ligand hotspots are identified (see Methods). Finally, highest scoring ligand hotspots from each EVE are ranked based on the integrated MHC presentation scores, generating a list of candidate EVE-derived peptide sequences which can be selected for manufacturing and administration.

To verify that EVEs are translated, processed, and presented as MHC ligands, we performed immunopeptidomics on CT26 and B16F10 mouse cancer cells and tumor samples. For CT26, three immunopeptidomics datasets were generated using unstimulated cells, IFNy-stimulated cells and *in vivo* grown tumors. A total of 43679 unique peptide sequences without post translational modifications (PTMs) were identified across these three datasets (Figure S1A). For B16F10, two immunopeptidomics datasets were generated using IFNy-stimulated cells and *in vivo* grown tumors. Unstimulated B16F10 cells were not analyzed as previous studies have reported poor MHC presentation on unstimulated B16F10 cells[23]. A total of 7078 unique sequences without PTMs were identified across the two datasets (Figure S1G). The datasets were further filtered, keeping only peptides with a length of 8-11 amino acids, a PEAKS confidence value above 20 and expression levels above 1 TPM. Following these filtering steps, a total of 23826 MHCI ligands remained in the CT26 dataset, of which 89 originated from EVE sequences (Figure S1B). For the B16F10 dataset, a total of 4749 MHCI ligands passed the filtering criteria with 19 originating from EVEs (Figure S1H). We next clustered the MHC ligands by their sequence motifs and generated sequence logos for each cluster. For CT26, the sequence logos matched known motifs for H2-Dd, H2-Kd and H2-Ld (Figure S1D, S1E and S1F), and for B16F10, the sequence logos could be matched to H2-Db and H2-Kb, as expected (Figure S1J and S1K).

Finally, we set out to evaluate the predictive performance of the tools employed by ObsERV to predict surface presented MHCI ligands derived from expressed EVEs and endogenous

proteins, defined as EVEs/transcripts with TPM above 1. Three methods for selecting EVE ligands were benchmarked; 1) EVE expression levels alone, 2) sequence-based MHC ligand presentation predictions alone and 3) a combination of 1 and 2. For comparison, a naïve approach where MHC ligands were selected at random was also included. As expected, the integrated model of expression and ligand prediction outperforms expression and ligand prediction alone as well as the random model. For B16F10, the performance on EVE ligands was overall higher than on endogenous ligands, although this could be due to the low number of EVE ligands included in this dataset. For the CT26 datasets, the expression levels alone also performed better on the EVE ligand subset compared to the endogenous ligand subset, which was not seen with the other methods, where the performance values were similar.

Overall, this analysis shows that EVEs are expressed in mouse cancer models and that EVE-derived peptides are processed and presented by MHC molecules on the cell surface. We demonstrated that the EVE-derived ligands can be predicted using *in silico* methods already employed for prediction of endogenous MHC ligands, indicating that EVE-derived peptides are a viable source of tumor antigens for personalized immunotherapies.

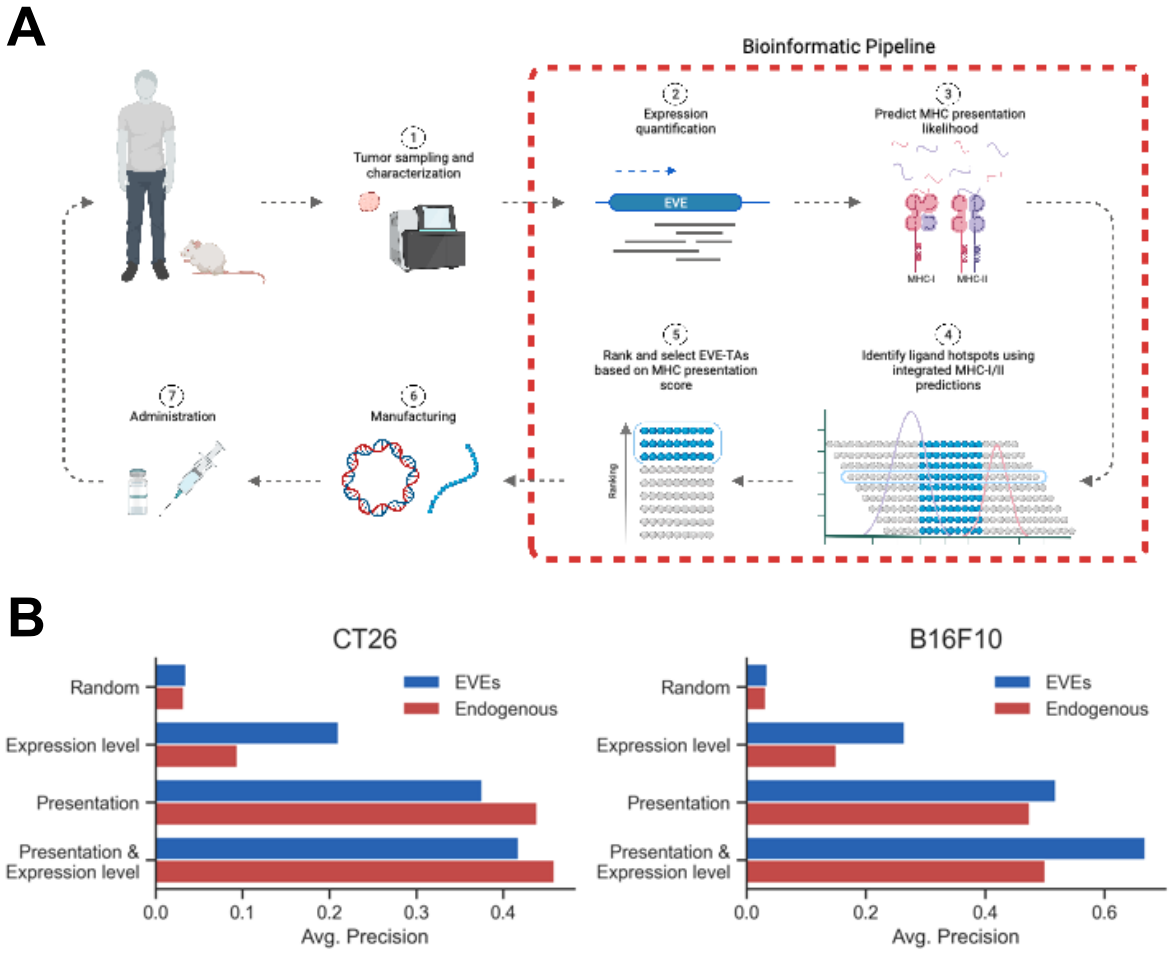

Figure 1. Pipeline overview and evaluation of predictive performance. (A) Flow chart for designing EVE-based immunotherapies. A tumor sample is collected for characterization using RNA-seq. For all expressed EVEs, the likelihood of MHC ligand presentation is predicted for all 8-11mer peptides, and a ligand hotspot is selected based on an integrated MHCI and -II presentation score. The EVE hotspots are then ranked based on the combined presentation scores and EVE expression levels, resulting in a final prioritized list of EVE-derived peptide candidates. The top scoring EVE-derived peptides are selected for manufacturing and administration. (B) Benchmark of features for selecting EVE-derived peptides in the CT26 and B16 mouse cancer models. Bars show the average precision score of the four features evaluated in the benchmark. Light blue bars show the predictive performance on EVE-derived MHC ligands and purple bars show performance on MHC ligands from endogenous proteins. Positive data points were MHC ligands identified through immunopeptidomics, filtered to only include peptide sequences from expressed EVEs/proteins with a TPM>1. The CT26 dataset consisted of 54 EVE-derived ligands and 10673 endogenous ligands identified by MS. The B16F10 dataset included 18 EVE-derived ligands and 1645 endogenous ligands identified by MS. For each dataset, negative data points were added by randomly sampling peptide sequences from other EVEs/proteins with TPM>1 at a 30:1 negative to positive ratio.

In silico designed EVE-based immunotherapy induces strong anti-tumor effect and T-cell responses in the CT26 colon carcinoma mouse model

We set out to investigate whether EVEs can serve as relevant therapeutic cancer targets for personalized immunotherapy using the preclinical cancer model, CT26. EVE expression levels in *in vivo* grown CT26 tumors were quantified using RNA-seq. Using ObsERV, we designed a CT26-specific immunotherapy based on the EVE expression levels and predicted MHC ligands. The *in silico* design comprised the 13 top ranked EVE epitopes encoded into a pTVG4 plasmid DNA (pDNA) (from hereon: ObsERV). For comparison, we designed a pDNA construct encoding 13 EVE-derived epitopes containing MHCI ligands, which we had identified in the previous immunopeptidomics experiments (from hereon: MS-EVEs). The amino acid sequences of the vaccine epitopes and their computational features are shown in Tables S3 and S4.

Groups of mice were vaccinated with ObsERV pDNA, MS-EVE pDNA or mock pDNA in one-week intervals starting two weeks prior to subcutaneous (s.c.) inoculation with CT26 cancer cells (Figure 2A). Vaccination with ObsERV and MS-EVEs prevented CT26 tumor establishment, in contrast to mock pDNA treatment which did not prevent development of tumors (Figures 2B and 2C).

We then sought to assess the immune responses elicited in vaccinated mice. For that, splenocytes from the treated mouse groups were re-stimulated with vaccine-corresponding peptide pools and each individual peptide before assessing IFNγ release by ELISPOT assay. Re-stimulation with the vaccine peptide pool in the ObsERV group yielded very strong IFNγ responses, whereas the MS-EVE group showed weaker immune reactivity towards its respective peptide pool (Figure 2D). Individual peptide re-stimulation revealed that epitopes O3, O6 and O8 were the main immunogenicity drivers in the ObsERV group, while the MS-EVE pool immune recognition was attributable to peptides M1 and M5, followed by the less immunogenic M11 and M6. Interestingly, the two most immunogenic peptides in the two groups, O3 and M1, derive from the gp70 protein of the Murine Leukemia Virus (MuLV) and contain the same MHCI minimal epitope SPSYVYHQF (AH1) but with different N- and C-terminal flanking sequences due to the different epitope selection approaches. To exclude the possibility that the higher IFNγ ELISPOT immunogenicity of peptide O1 over M3 in the two groups stems from intrinsic differences in the intracellular processing of the vaccine peptides that would affect the epitope availability, splenocytes were re-stimulated with the minimal AH1 peptide (Figure 2D). ObsERV-vaccinated mice were found to harbor stronger immune reactivity towards the minimal peptide than their MS-EVE-vaccinated counterparts.

Mock DNA-immunized mice mounted IFNγ responses upon stimulation with both vaccine peptide pools in the IFNγ ELISPOT assay. Deconvoluted peptide stimulation showed that peptides O3, M1 and M11 were immunogenic, whereas naïve mice showed no pre-existing immunity against the peptide pools (Figure 2D), indicating that this was related to tumor presence.

To gain a deeper understanding of the vaccine-induced immune responses, we sought to characterize the functional phenotype of the vaccine-specific T cells. Vaccination with ObsERV elicited polyfunctional CD8+ and CD4+ T cells reactive to the vaccine-matching peptide pool (Figure 2E). Re-stimulation with each of the ELISPOT-identified immunogenic peptides revealed that peptides O3 and O6 were strong inducers of IFNγ and TNFα production within splenic CD8+ T cells, while peptides O6 and O8 induced polyfunctional CD4+ T-cell reactivity. Interestingly, splenic CD8+ T cells from the MS-EVEs group produced a weaker polyfunctional response upon peptide pool re-stimulation, while no reactive CD4+ T cells were recorded. Similarly, all ELISPOT-identified immunogenic MS-EVE peptides stimulated cytokine release from CD8+ T cells. Finally, the mock DNA group was found to harbor CD8+ T cells reactive to the ELISPOT-identified peptides O3, M1 and M11.

Collectively, the presented data suggest that EVEs can serve as relevant cancer antigens and be targeted in the context of *in silico* designed immunotherapy, leading to strong and efficacious CD8+ and CD4+ T-cell responses.

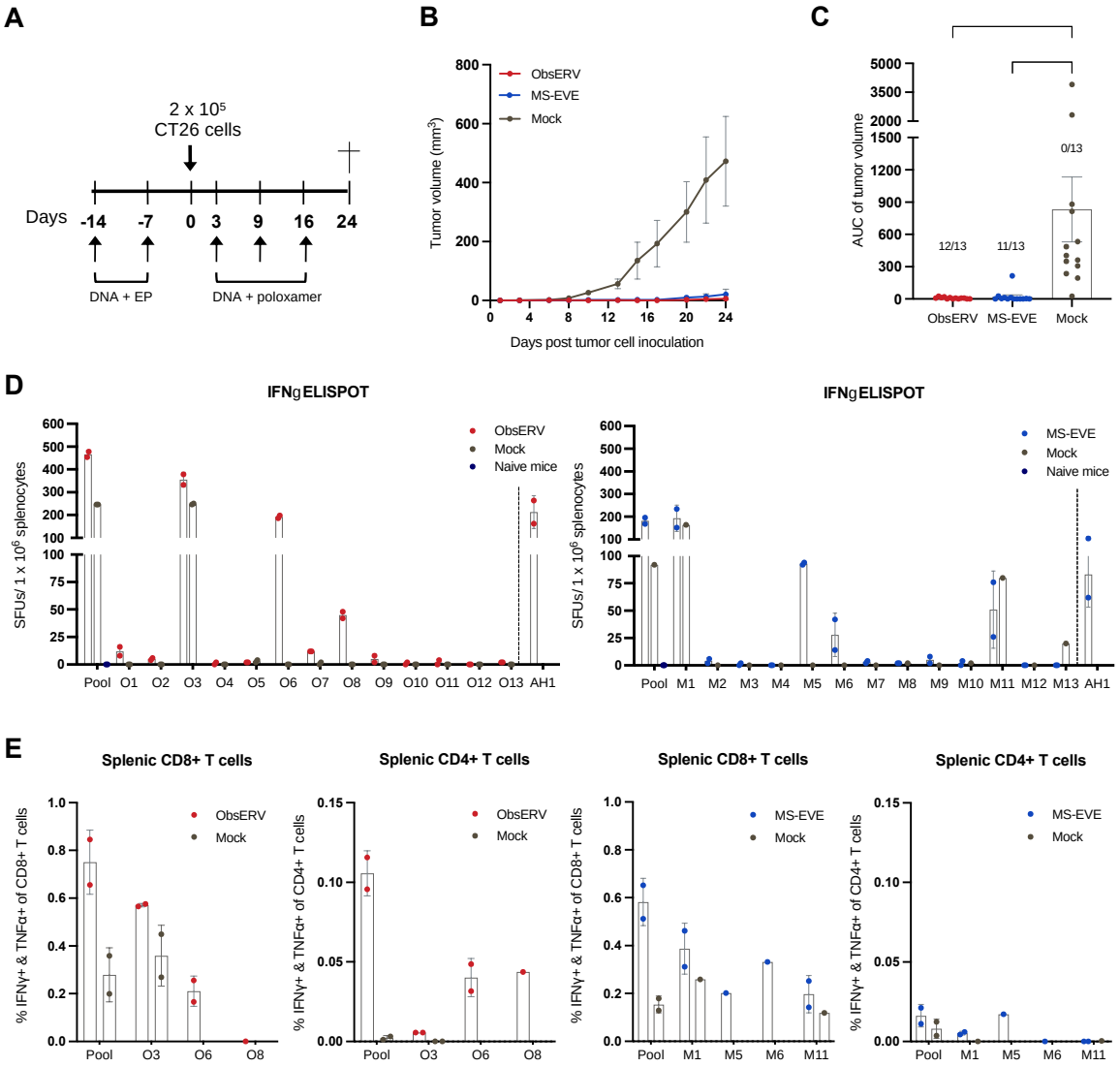

Figure 2. *In silico* designed EVE-based immunotherapy prevents CT26 tumor establishment and induces efficacious T-cell immunity. (A) Study outline of the *in vivo* experiment. (B) Group mean tumor growth curves (in mm^3) ± standard error of the mean (SEM). (C) Area under the tumor growth curve (AUC) for individual mice by group ± standard error of the mean (SEM). Fractions represent the number of mice that remained tumor-free by the end of the study. (D) Pooled bulk splenocytes (n = 6-7 mice) were tested on IFNγ ELISPOT assay for immune reactivity upon re-stimulation with vaccine-matching peptide pools and individual peptides (run in duplicates). Bars represent the mean +/- Standard Deviation (SD). Stars indicate immunogenic peptides carried on to the re-stimulation and Intracellular Cytokine Staining (ICS) assay. (E) Pooled bulk splenocytes (n = 6-7 mice per group) were re-stimulated with the vaccine-matching peptide pools and the individual peptides identified as immunogenic in IFNγ ELISPOT in (D) followed by ICS for detection of polyfunctional CD8+ and CD4+ T cells (run in duplicates). Bars represent the mean +/- Standard Deviation (SD). Background levels representing cytokine production in unstimulated samples have been subtracted. *Statistics*: Kruskal-Wallis test with Dunn's multiple comparisons correction (D), ****$p < 0.0001$

In silico designed EVE-based immunotherapy induces durable tumor protection and recall T-cell responses in the CT26 model

Motivated by the strong immunogenicity in vaccinated mice, we proceeded to examine whether these responses were long-lasting. For that reason, a subset of mice vaccinated with ObsERV and MS-EVEs that remained tumor-free along with age-matched, treatment-naive mice were interrogated for T-cell immune recognition of the minimal epitope SPSYVYHQF (AH1) encoded by the vaccine epitopes O3 and M1 79 days after the CT26 tumor challenge using the MHCI dextramer methodology (Figure 3A). Mice vaccinated with ObsERV displayed stronger AH1 recognition by CD8+ T cells than the MS-EVEs group. Tumor re-challenge 85 days after the primary tumor challenge revealed that vaccinated mice preserved their capacity to prevent CT26 tumor establishment, as almost all vaccinated mice remained tumor-free. In contrast, untreated mice developed and grew tumors (Figures 3C and 3D).

Further immune analyses demonstrated that mice vaccinated with ObsERV preserved 80% of the immune reactivity to the vaccine peptide pool measured in the primary study and sustained their capacity to recall strong IFNγ responses against the same vaccine epitopes identified as immunogenic previously (Figure 3E). In contrast, the IFNγ response yielded in the MS-EVEs group upon re-stimulation with the vaccine peptide pool was contracted by approximately 70%. Similar diminishment was observed in the functional response towards the lead immunogenic peptide M1, while the previously immunogenic peptides M5 and M6 did not elicit immune recognition. These data indicate that although both groups show an equally strong anti-tumor effect upon re-challenge, the ObsERV-designed therapy leads to retainment of strong and broad immune recognition compared to the MS-EVEs.

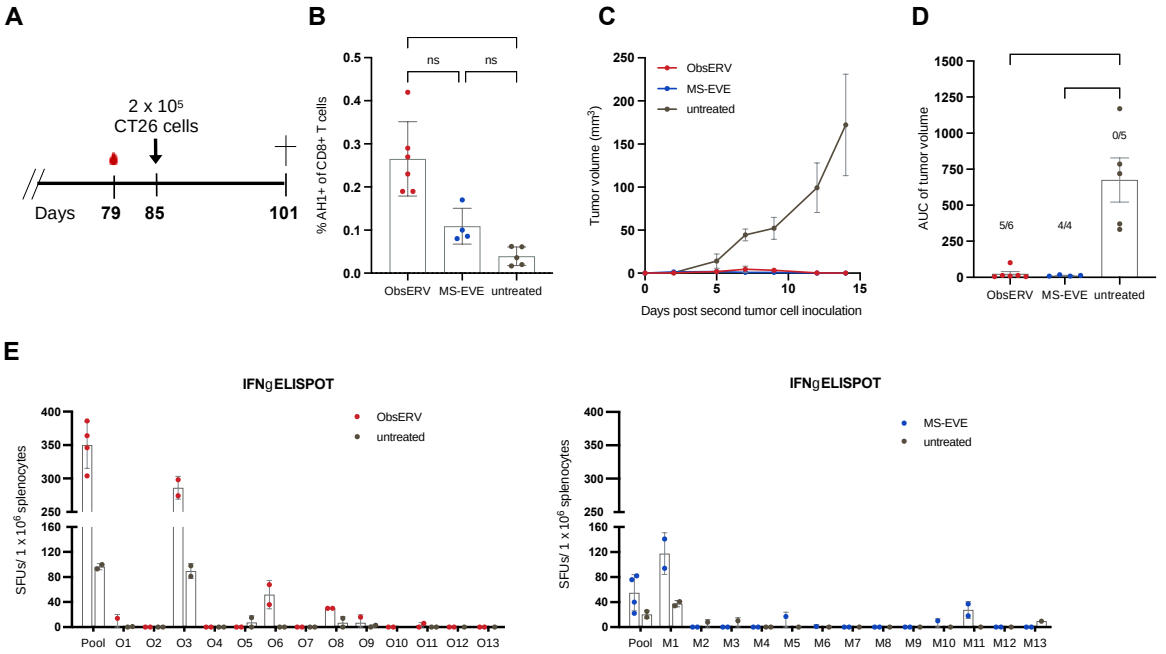

Figure 3. *In silico* designed EVE-based immunotherapy induces long-lasting T-cell responses and tumor protection upon tumor re-challenge. (A) Study outline of the *in vivo* experiment. (B) % presence of vaccine epitope AH1-specific CD8+ T cells in tail vein blood of vaccinated mice and age-matched untreated mice 79 days after the primary CT26 tumor challenge +/- Standard Deviation (SD). Blood was stained with MHCI dextramer loaded with the H2-Ld-restricted minimal epitope SPSYVYHQF (AH1) derived from gp70 protein encoded by vaccine epitopes O3 and M1 for quantification of AH1-specific CD8+ T cells. (C) Group mean tumor growth curves (in mm^3) ± standard error of the mean (SEM). (D) Area under the tumor growth curve (AUC) for individual mice by group ± standard error of the mean (SEM). Fractions represent the number of mice that remained tumor-free by the end of the study. (E) Bulk splenocytes (n = 4-6 mice per group) were pooled group-wise and re-stimulated on IFNγ ELISPOT assay with vaccine-matching pools and peptides (run in duplicates) +/- Standard Deviation (SD). Experiment was run twice. *Statistics*: Kruskal-Wallis test with Dunn's multiple comparisons correction (D), *p < 0.05, **p < 0.01.

The EVE antigen burden stratifies low TMB cancer patients receiving immunotherapy into groups of differential overall survival

Next, we explored if EVEs constitute a general source of tumor antigens. We let the number of expressed EVEs in a tumor represent the EVE antigen burden (hereon known as EVE burden), and then studied whether the EVE burden could stratify cancer patients receiving immunotherapy into groups with differential overall survival. We hypothesized that a higher EVE burden would result in a longer survival for cancer patients in a similar fashion as is observed for TMB. Three studies of metastatic melanoma patients receiving immunotherapy were included in the initial investigation, namely Hugo et al.[24], Lauss et al.[25] and Riaz et al.[26]. The dataset was subset to the patients with pre-therapy tumor biopsies characterized by RNA-seq and whole-exome sequencing (WES) and a healthy tissue biopsy characterized by WES. The TMB was represented by the number of missense somatic mutations based on the matched tumor/healthy WES, and the RNA-seq was analyzed for EVE expression levels to compute the EVE burden. The patients were assigned group labels based on their TMB (High TMB, Low TMB) and EVE burden (High EVE Burden, Low EVE Burden). Across all three studies, the EVE burden appears to modestly stratify the patients into groups with differential overall survival (Figure S3). Interestingly, this stratification was not displayed by the patients with High TMB (Figure S3B and S3H). In contrast, the patients with Low TMB showed promising signs of differential overall survival based on the EVE burden across the three studies with longer overall survival for the High EVE Burden group compared to the Low EVE Burden group (Figure S3C, S3F and S3I). Increasing the number of patients through combination of all three studies enhances that the EVE burden stratification is conditioned on the TMB (Figures 4A, 4B and 4C). To verify these initial findings, we analyzed a cohort of 245 bladder cancer patients treated with anti-PDL1 and tumor samples characterized by RNA-seq and WES[27]. We observe that the EVE burden is lower for the bladder cancer patients as compared to the malignant melanoma patients, yet trends of the same stratification patterns are preserved, i.e., differential stratification by the EVE burden for low-TMB patients but not high TMB patients (Figures 4D, 4E, 4F).

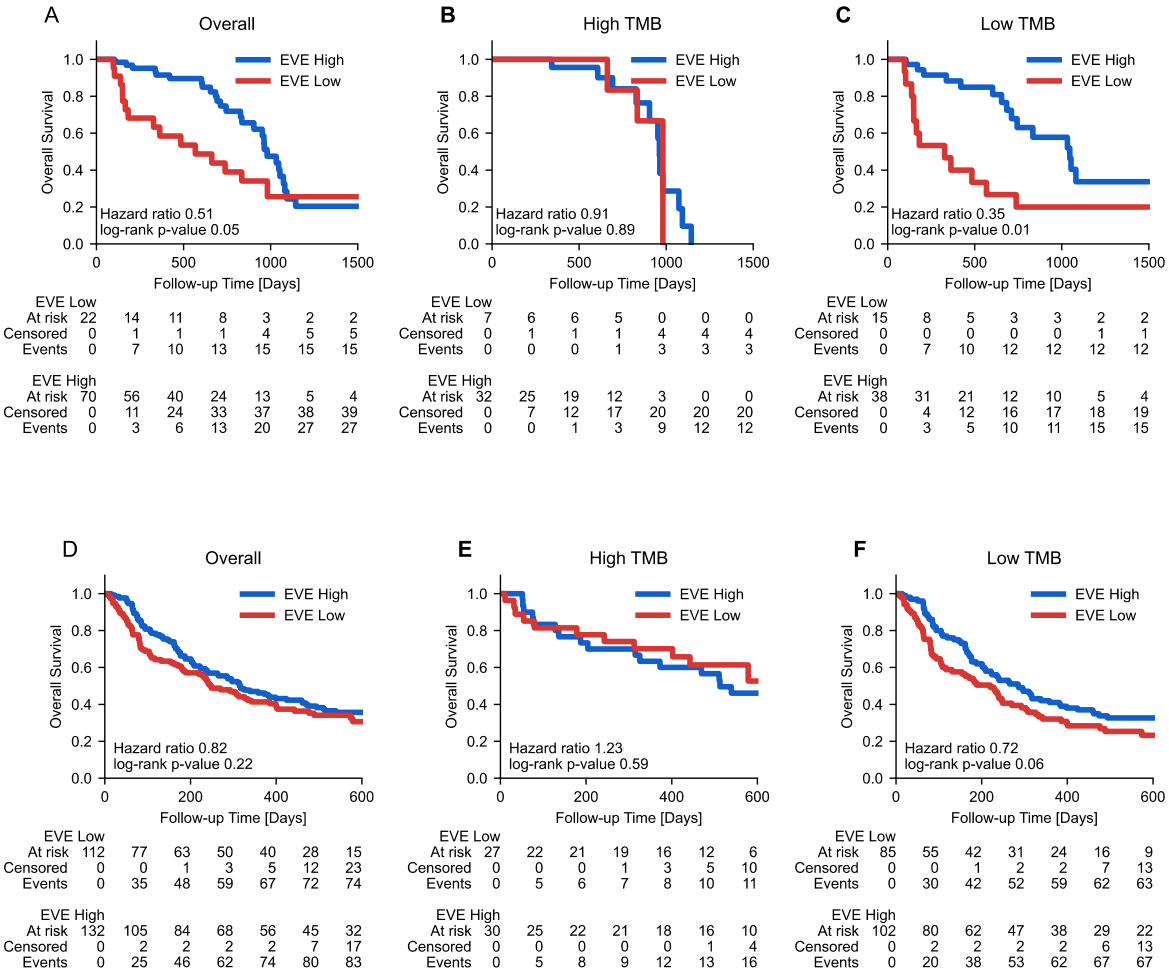

Figure 4. EVE antigen burden stratifies low-TMB patients.
(A-C) Stratification of melanoma patients using the EVE burden: (A) all patients, (B) high TMB patients, (C) low TMB patients. (D-F) Stratification of bladder cancer patients using the EVE burden: (D) all patients, (E) high TMB patients, (F) low TMB patients.

The EVE antigen burden is correlated with expression of Myc target genes

Based on the observation that EVE burden stratifies patients receiving immunotherapy into groups of differential survival times, we set out to investigate the functional correlates of the expressed EVEs. We performed a gene co-expression analysis[28] on a combined RNA-seq dataset, comprising the three melanoma cohorts[24,25,26] analyzed previously, identifying 16 functional gene modules. We found that four of the functional modules correlated significantly with the EVE burden (Figure 5A) and subjected them to gene set enrichment analysis (GSEA) using the hallmark gene sets from the Molecular Signatures DataBase[29]. Strikingly, all four modules were enriched for genes that are annotated as transcriptionally regulated targets of the protooncogene *MYC*[30] (Figure 5B). To confirm these findings, we performed differential gene expression analyses comparing the patients with high EVE burden to those with low burden. The analyses were conducted on the complete melanoma dataset as well as low- and high TMB subsets (Figure S5). The GSEA on significantly up- and downregulated genes confirmed that Myc target genes were differentially expressed and showed that the effect was most significant in the low TMB subset (Figure 5C). Furthermore, Myc targets tend to have larger fold changes as compared to the other genes (Figure 5D).

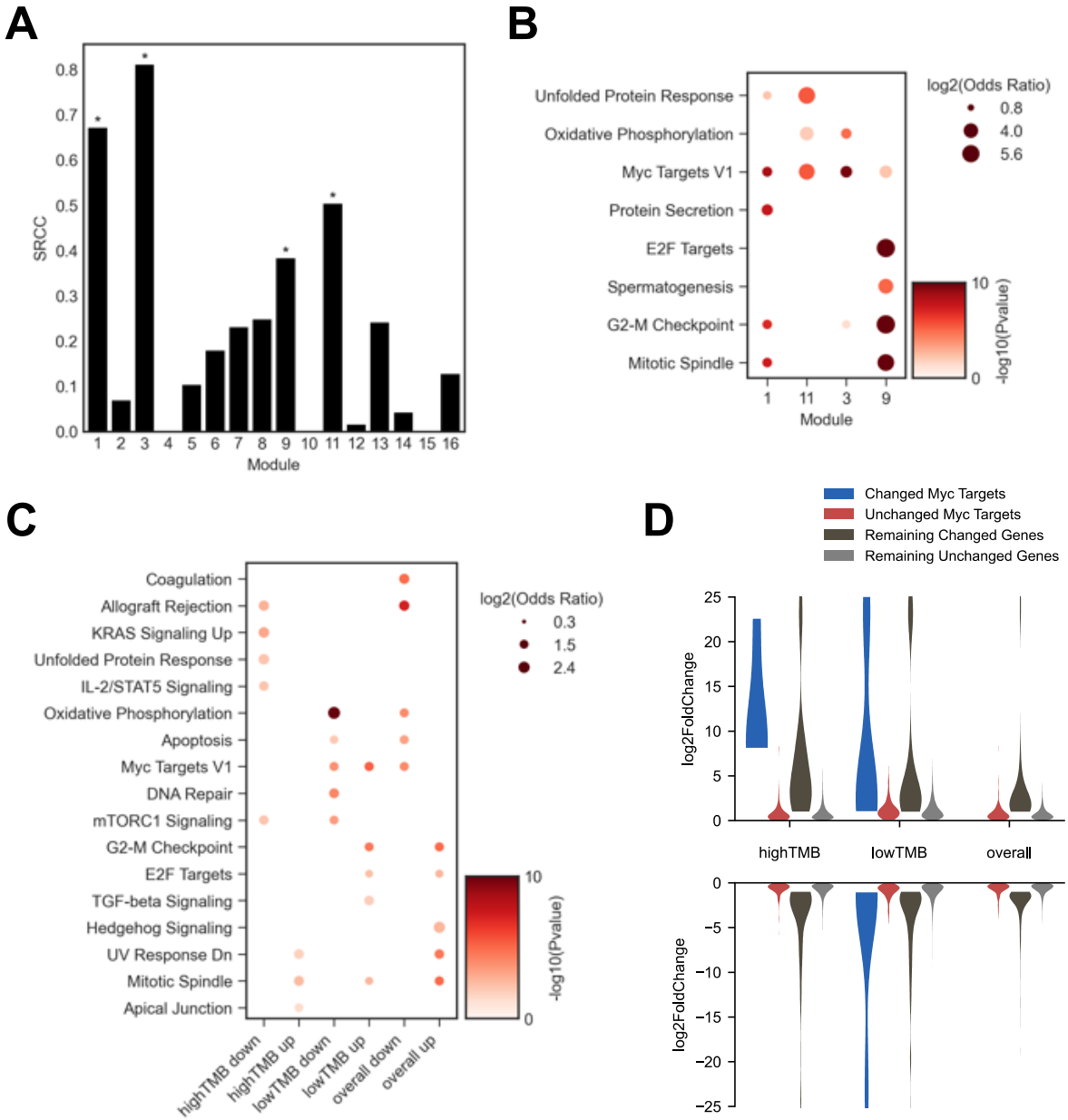

Figure 5. Functional inference of EVE expression in melanoma. (A) Spearman rank correlation between EVE burden and WGCNA module eigengenes. *p<0.001. (B and C) Gene set enrichment analysis performed on (B) genes included in functional modules identified by gene co-expression and (C) DEGs. The analysis was performed using MSigDB hallmark gene sets. Dot sizes depict the odds ratio and dot are colored by adjusted p-value. Results are only shown for significantly enriched gene sets (adjusted p-value < 0.05). (D) Distribution plots of log-fold changes for the Myc targets in comparison to the remaining genes.

EVEs may facilitate personalized immunotherapy of low TMB cancer patients

Next, we investigated whether targeting EVEs could facilitate personalized immunotherapy of cancer patients with low TMB by serving as a complementary antigen source. Matched TCGA deposited RNA-seq and WES of tumor biopsies from different cancer types were analyzed for EVE expression levels and somatic mutations (Figure 6A and 6B). We observe that the EVE burden differs across cancer types and does not correlate with TMB (Figure 6C and 6D). Interestingly, we find that bladder cancer has the lowest EVE burden of the analyzed cancer types, while esophageal cancer and stomach adenocarcinoma display considerately higher EVE burdens than the other cancer types. Overall, EVEs appear to be able to facilitate the design of personalized immunotherapy by augmenting the pool of epitopes.

Notably, the EVE burdens observed in the TCGA deposited melanoma and bladder cancer datasets are comparable to those observed in the immunotherapy treated cohorts subjected to survival analysis above[24–27].

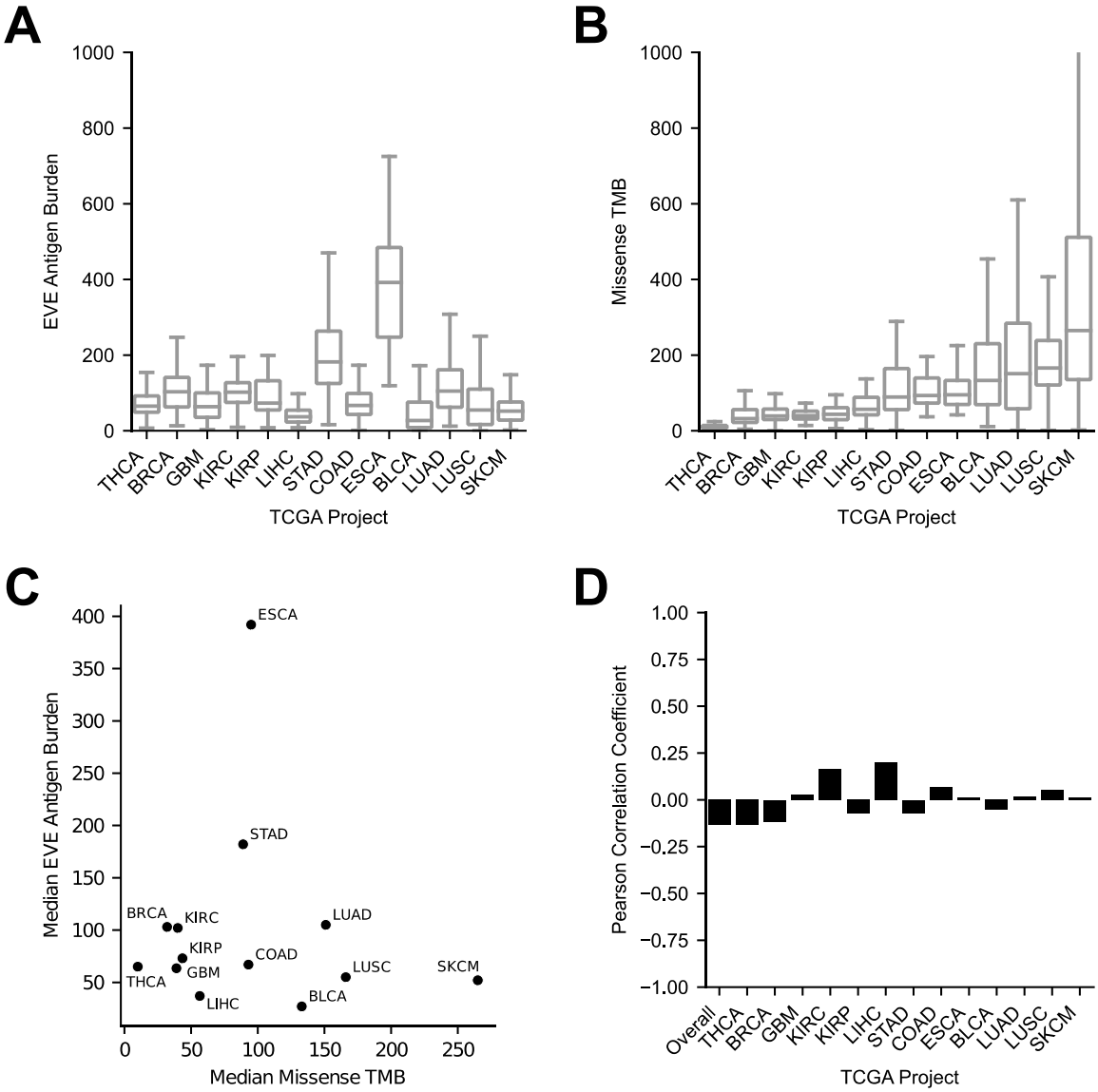

Figure 6. EVE antigen burden and tumor mutational burden across cancer types. (A) Distributions of EVE burden. (B) Distributions of the number of missense somatic variants. (C) Scatterplot of cancer type specific EVE burden versus the number of missense somatic variants. (D) Barplot showing the Pearson correlation coefficient between TMB and EVE burden. None of the correlation coefficients were found statistically significant despite high n-sizes (all p-values > 0.05). Abbreviations: THCA - Thyroid carcinoma, BRCA – Breast invasive carcinoma, GBM – Glioblastoma multiforme, KIRC – Kidney renal clear cell carcinoma, KIRP - Kidney renal clear cell carcinoma, LIHC - Liver hepatocellular carcinoma, STAD, COAD - Colon adenocarcinoma, ESCA - Esophageal carcinoma, BLCA - Bladder Urothelial Carcinoma, LUAD - Lung adenocarcinoma, LUSC - Lung squamous cell carcinoma, SKCM - Skin cutaneous melanoma.

DISCUSSION

We developed an *in silico* platform, ObsERV, for the design of personalized cancer immunotherapy based on EVE derived peptides. We demonstrated that EVEs are presented by MHC class I on the cell surfaces of murine tumors, validating that EVE derived peptides are eligible targets for T-cell mediated therapy. These observations are in line with previous studies that have found that peptides from various transposable elements are presented on MHC molecules and may elicit immune responses[9,10,18,31]. ObsERV was benchmarked and found capable of predicting the EVE derived MHC ligands for both the CT26 and B16F10 murine models, motivating the advancement of ObsERV to preclinical testing.

Preclinical testing demonstrated that a DNA-based cancer vaccine encoding ObsERV-selected EVE epitopes prevents CT26 tumor establishment mirrored by robust polyfunctional CD4+ and CD8+ T-cell responses. The engagement of both T-cell subsets validates the ObsERV ranking strategy that seeks to leverage both CD4+ and CD8+ T cells. Comparable anti-tumor protection was also attained by DNA-based vaccination with MS-EVEs, a selection of EVE peptides containing experimentally confirmed MHCI ligands. It is thus not surprising that this group elicited a primarily CD8+ T-cell response against the immunogenic vaccine epitopes. Remarkably, both immunotherapies demonstrated capacity to preserve tumor protection 85 days after the initial tumor challenge with no boost vaccination. While cases of preclinical anti-tumor protection of endogenous retroviral therapy have been reported for select model antigens[19,20,21], this is to our knowledge the first time that an *in silico* designed immunotherapy comprising EVE derived epitopes was found to confer long-lasting tumor protection and elicit a broad and strong poly-functional T cell response.

We found that ObsERV selected epitopes induced both CD8+ and CD4+ T cells, while MS-EVE epitopes primarily induced CD8+ T cells. The engagement of tumor-specific, Th1-polarized CD4+ T cells induced by ObsERV selected epitopes may be the cause for the improved immunogenicity of the ObsERV group compared to the MS-EVE group. Firstly, we hypothesize that CD4+ T-cell help may be responsible for the improved CD8+ T-cell-mediated functional responses. It is well documented that CD4+ T cells play a crucial role in potentiating cytotoxic CD8+ T-cell anti-tumor responses in preclinical models[32] by enabling their expansion and differentiation during the priming phase[33]. Consequently, we hypothesize that the increased immune reactivity against the vaccine-encoded AH1 minimal peptide is explained by the provision of CD4+ T-cell help on Antigen Presenting Cells (APCs) during the priming of CD8+ T cells specific to this antigen, which have been shown to require CD4+ T-cell help to be maximally effective[19]. Secondly, vaccine-induced involvement of the CD4+ T-cell compartment may contribute to increased persistence of vaccine-specific CD8+ T cells, as CD4+ T cells regulate the memory CD8+ T-cell differentiation[34]. We speculate that this

underlies the increased frequency of a persistent population of circulating AH1-specific CD8+ T cells in the ObsERV group compared to MS-EVE group and, importantly, contributes to the significant retainment of vaccine epitope immunogenicity in the splenic compartment of tumor re-challenged mice from the ObsERV group. This highlights the importance of incorporating CD4+ T-cell epitopes in the vaccine design to achieve a robust and durable T-cell response profile. Further studies are needed to validate these hypotheses, e.g., a tumor challenge study with early CD4+ T-cell depletion.

Based on published melanoma and bladder cancer cohorts receiving immunotherapy, we find support for EVEs constituting a complementary tumor antigen source to be leveraged by personalized immunotherapy. We find signs that the number of expressed EVEs specifically stratify cancer patients with a low number of somatic mutations. One may hypothesize that to achieve a T-cell mediated clinical response, there must be a certain number of targetable epitopes. For cancer patients with high TMB, there is an abundance of potential neoepitope targets and clinical response to immunotherapy is likely not blunted by the lack of good epitope targets. In contrast, for patients with low TMB, the addition of EVE-derived peptides may increase the combined epitope reservoir to levels that facilitate a clinical response. The EVE antigen burden could potentially be used in combination with TMB to improve the identification of cancer patients that would benefit from immunotherapy. Smith et al. (2018)[35] identified an ERV signature that stratifies renal cell carcinoma patients, supporting that EVEs can serve as biomarkers for stratification of cancer patients. Furthermore, we found an enrichment for targets of the protooncogene *MYC* among genes whose expression levels correlate with the EVE burden. *MYC* plays a key role in regulating cell growth and proliferation, and is frequently found to be amplified in multiple different cancers, including melanoma[36,37]. Interestingly, *in vitro* studies of cell lines have associated the expression of ERVs to tumorigenesis[38] and others have suggested that this effect is mediated via de-repression of *MYC*[39]. While our results suggest that there is a correlation between EVE expression and increased Myc activity, they do not demonstrate a causal relationship. Another explanation for the observed correlation could be that *MYC* deregulation disturbs the transcriptional regulation of EVEs, leading to their upregulation. If Myc is a master regulator of the expression of epigenetically primed EVEs, activation of Myc could potentially induce the expression of several EVE-derived epitopes and enhance the effect of immunotherapy, especially for patients with few antigen targets. Such an effect could potentially be further enhanced by co-administration of epigenetic modulators[40–42].

Finally, we found that the EVE antigen burden is cancer type specific and that it does not correlate with TMB. As such, EVEs may be used as a supplementary antigen source to pave the way for the application of personalized immunotherapy in low-TMB patients and cancer types that are considered unsuitable target indications, e.g., esophageal cancer and stomach

adenocarcinoma. Collectively, we have demonstrated that EVEs are potent anti-tumor targets that can be leveraged to facilitate and potentially enhance personalized immunotherapy.

ACKNOWLEDGEMENTS

We would like to acknowledge Mads Lausen Nielsen, Anne Lund and Marianne Blirup Jensen (Evaxion Biotech) for their invaluable assistance in performing the *ex vivo* experiments. Furthermore, we wish to thank Henriette Løwe Press, Heidii Mazur and Rikke Sølberg (Evaxion Biotech) for their excellent execution of preclinical studies and meticulous monitoring of the mice well-being. NJB is a fellow of the Lundbeck Foundation (R272-2017-4040) and acknowledges funding from Aarhus University Research Foundation (AUFF-E-2018-7-14), and the Novo Nordisk Foundation (NNF21OC0071483). AWP is a National Health and Medical Research Council of Australia Investigator Fellow. PF was granted the Victorian Cancer Agency Fellowship.

AUTHOR CONTRIBUTIONS

CG, MAP, PG, EJL, BR and TT conceived the study and designed the experiments. CG, MAP, PG and TT wrote the manuscript. CG, PG, MS, JA and TT performed computational analysis. MAP, SHR, KP, PF, SC and SF performed experiments. DKK, NJB, JVK and AWP provided further intellectual input. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

CG, MAP, PG, EJL, SF, DKK, JVK, BR and TT are employed by Evaxion Biotech A/S that holds IP for identifying neoepitopes and personalized immunotherapy. AWP is on the scientific advisory board of Evaxion Biotech A/S. NJB has a patent application (PCT/GB2020/050221) on methods for cancer prognostication and a patent on methods for predicting anti-cancer response (US14/466,208). The remaining authors declare no competing interests.

REFERENCES

1. Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J.M., Desrichard, A., Walsh, L.A., Postow, M.A., Wong, P., Ho, T.S., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med *371*, 2189–2199. 10.1056/NEJMoa1406498.

2. Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science *348*, 124–128. 10.1126/science.aaa1348.

3. Twomey, J.D., and Zhang, B. (2021). Cancer Immunotherapy Update: FDA-Approved Checkpoint Inhibitors and Companion Diagnostics. AAPS J *23*, 39. 10.1208/s12248-021-00574-0.

4. Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J.-J., Rutkowski, P., Lao, C.D., Cowey, C.L., Schadendorf, D., Wagstaff, J., Dummer, R., et al. (2019). Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. N Engl J Med *381*, 1535–1546. 10.1056/NEJMoa1910836.

5. Tawbi, H.A., Schadendorf, D., Lipson, E.J., Ascierto, P.A., Matamala, L., Castillo Gutiérrez, E., Rutkowski, P., Gogas, H.J., Lao, C.D., De Menezes, J.J., et al. (2022). Relatlimab and Nivolumab versus Nivolumab in Untreated Advanced Melanoma. N Engl J Med *386*, 24–34. 10.1056/NEJMoa2109970.

6. Ott, P.A., Hu-Lieskovan, S., Chmielowski, B., Govindan, R., Naing, A., Bhardwaj, N., Margolin, K., Awad, M.M., Hellmann, M.D., Lin, J.J., et al. (2020). A Phase Ib Trial of Personalized Neoantigen Therapy Plus Anti-PD-1 in Patients with Advanced Melanoma, Non-small Cell Lung Cancer, or Bladder Cancer. Cell *183*, 347-362.e24. 10.1016/j.cell.2020.08.053.

7. Mørk, S.K., Kadivar, M., Bol, K.F., Draghi, A., Westergaard, M.C.W., Skadborg, S.K., Overgaard, N., Sørensen, A.B., Rasmussen, I.S., Andreasen, L.V., et al. (2022). Personalized therapy with peptide-based neoantigen vaccine (EVX-01) including a novel adjuvant, CAF®09b, in patients with metastatic melanoma. Oncoimmunology *11*, 2023255. 10.1080/2162402X.2021.2023255.

8. Yadav, M., Jhunjhunwala, S., Phung, Q.T., Lupardus, P., Tanguay, J., Bumbaca, S., Franci, C., Cheung, T.K., Fritsche, J., Weinschenk, T., et al. (2014). Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature *515*, 572–576. 10.1038/NATURE14001.

9. Kong, Y., Rose, C.M., Cass, A.A., Williams, A.G., Darwish, M., Lianoglou, S., Haverty, P.M., Tong, A.-J., Blanchette, C., Albert, M.L., et al. (2019). Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. Nat Commun *10*, 5228. 10.1038/s41467-019-13035-2.

10. Peltonen, K., Feola, S., Umer, H.M., Chiaro, J., Mermelekas, G., Ylösmäki, E., Pesonen, S., Branca, R.M.M., Lehtiö, J., and Cerullo, V. (2021). Therapeutic Cancer Vaccination with Immunopeptidomics-Discovered Antigens Confers Protective Antitumor Efficacy. Cancers *13*, 3408. 10.3390/cancers13143408.

11. Bonté, P.-E., Arribas, Y.A., Merlotti, A., Carrascal, M., Zhang, J.V., Zueva, E., Binder, Z.A., Alanio, C., Goudot, C., and Amigorena, S. (2022). Single-cell RNA-seq-based proteogenomics identifies glioblastoma-specific transposable elements encoding HLA-I-presented peptides. Cell Rep *39*, 110916. 10.1016/j.celrep.2022.110916.

12. Capietto, A.-H., Hoshyar, R., and Delamarre, L. (2022). Sources of Cancer Neoantigens beyond Single-Nucleotide Variants. Int J Mol Sci *23*, 10131. 10.3390/ijms231710131.

13. Grandi, N., and Tramontano, E. (2018). Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. Front Immunol *9*, 2039. 10.3389/fimmu.2018.02039.

14. Bonaventura, P., Shekarian, T., Alcazer, V., Valladeau-Guilemond, J., Valsesia-Wittmann, S., Amigorena, S., Caux, C., and Depil, S. (2019). Cold Tumors: A Therapeutic Challenge for Immunotherapy. Frontiers in Immunology *10*.

15. Garrison, K.E., Jones, R.B., Meiklejohn, D.A., Anwar, N., Ndhlovu, L.C., Chapman, J.M., Erickson, A.L., Agrawal, A., Spotts, G., Hecht, F.M., et al. (2007). T cell responses to human endogenous retroviruses in HIV-1 infection. PLoS Pathog *3*, e165. 10.1371/journal.ppat.0030165.

16. Thonnard, J., Colau, D., Boon, T., and Coulie, P. (2002). A human endogenous retroviral sequence encoding an antigen recognized on melanoma by cytolytic T lymphocytes. Cancer research *62*, 5510–5516.

17. Rycaj, K., Plummer, J.B., Yin, B., Li, M., Garza, J., Radvanyi, L., Ramondetta, L.M., Lin, K., Johanning, G.L., Tang, D.G., et al. (2015). Cytotoxicity of human endogenous retrovirus K-specific T cells toward autologous ovarian cancer cells. Clin Cancer Res *21*, 471–483. 10.1158/1078-0432.CCR-14-0388.

18. Saini, S.K., Ørskov, A.D., Bjerregaard, A.-M., Unnikrishnan, A., Holmberg-Thydén, S., Borch, A., Jensen, K.V., Anande, G., Bentzen, A.K., Marquard, A.M., et al. (2020). Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers. Nat Commun *11*, 5660. 10.1038/s41467-020-19464-8.

19. Casares, N., Lasarte, J.J., de Cerio, A.L., Sarobe, P., Ruiz, M., Melero, I., Prieto, J., and Borrás-Cuesta, F. (2001). Immunization with a tumor-associated CTL epitope plus a tumor-related or unrelated Th1 helper peptide elicits protective CTL immunity. Eur J Immunol *31*, 1780–1789. 10.1002/1521-4141(200106)31:6<1780::AID-IMMU1780>3.0.CO;2-I.

20. Stringhini, M., Spadafora, I., Catalano, M., Mock, J., Probst, P., Spörri, R., and Neri, D. (2021). Cancer therapy in mice using a pure population of CD8+ T cell specific to the AH1 tumor rejection antigen. Cancer Immunol Immunother *70*, 3183–3197. 10.1007/s00262-021-02912-9.

21. Neukirch, L., Nielsen, T.K., Laursen, H., Daradoumis, J., Thirion, C., and Holst, P.J. (2019). Adenovirus based virus-like-vaccines targeting endogenous retroviruses can eliminate growing colorectal cancers in mice. Oncotarget *10*, 1458–1472. 10.18632/oncotarget.26680.

22. Kreiter, S., Vormehr, M., Van De Roemer, N., Diken, M., Löwer, M., Diekmann, J., Boegel, S., Schrörs, B., Vascotto, F., Castle, J.C., et al. (2015). Mutant MHC class II epitopes drive therapeutic immune responses to cancer. Nature *520*, 692–696. 10.1038/nature14426.

23. Böhm, W., Thoma, S., Leithäuser, F., Möller, P., Schirmbeck, R., and Reimann, J. (1998). T cell-mediated, IFN-gamma-facilitated rejection of murine B16 melanomas. J Immunol *161*, 897–908. 10.4049/jimmunol.161.2.897.

24. Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. Cell *165*, 35–44. 10.1016/j.cell.2016.02.065.

25. Lauss, M., Donia, M., Harbst, K., Andersen, R., Mitra, S., Rosengren, F., Salim, M., Vallon-Christersson, J., Törngren, T., Kvist, A., et al. (2017). Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. Nat Commun *8*, 1738. 10.1038/s41467-017-01460-0.

26. Riaz, N., Havel, J.J., Makarov, V., Desrichard, A., Urba, W.J., Sims, J.S., Hodi, F.S., Martín-Algarra, S., Mandal, R., Sharfman, W.H., et al. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. Cell *171*, 934-949.e16. 10.1016/j.cell.2017.09.028.

27. Mariathasan, S., Turley, S.J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, E.E., Koeppen, H., Astarita, J.L., Cubas, R., et al. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature *554*, 544–548. 10.1038/nature25501.

28. van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J.P. (2018). Gene co-expression analysis for functional classification and gene-disease predictions. Briefings in bioinformatics *19*, 575–592. 10.1093/bib/bbw139.

29. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst *1*, 417–425. 10.1016/j.cels.2015.12.004.

30. Dang, C.V. (2012). MYC on the path to cancer. Cell *149*, 22–35. 10.1016/j.cell.2012.03.003.

31. Attig, J., Young, G.R., Hosie, L., Perkins, D., Encheva-Yokoya, V., Stoye, J.P., Snijders, A.P., Ternette, N., and Kassiotis, G. (2019). LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. Genome Res *29*, 1578–1590. 10.1101/gr.248922.119.

32. Ossendorp, F., Mengedé, E., Camps, M., Filius, R., and Melief, C.J.M. (1998). Specific T helper cell requirement for optimal induction of cytotoxic T lymphocytes against major histocompatibility complex class II negative tumors. Journal of Experimental Medicine *187*, 693–702. 10.1084/jem.187.5.693.

33. Bedoui, S., Heath, W.R., and Mueller, S.N. (2016). CD4(+) T-cell help amplifies innate signals for primary CD8(+) T-cell immunity. Immunol Rev *272*, 52–64. 10.1111/imr.12426.

34. Laidlaw, B.J., Craft, J.E., and Kaech, S.M. (2016). The multifaceted role of CD4(+) T cells in CD8(+) T cell memory. Nat Rev Immunol *16*, 102–111. 10.1038/nri.2015.10.

35. Smith, C.C., Beckermann, K.E., Bortone, D.S., De Cubas, A.A., Bixby, L.M., Lee, S.J., Panda, A., Ganesan, S., Bhanot, G., Wallen, E.M., et al. (2018). Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. J Clin Invest *128*, 4804–4820. 10.1172/JCI121476.

36. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905. 10.1038/nature08822.

37. Kraehn, G.M., Utikal, J., Udart, M., Greulich, K.M., Bezold, G., Kaskel, P., Leiter, U., and Peter, R.U. (2001). Extra c-myc oncogene copies in high risk cutaneous malignant melanoma and melanoma metastases. Br J Cancer *84*, 72–79. 10.1054/bjoc.2000.1535.

38. Zhou, F., Li, M., Wei, Y., Lin, K., Lu, Y., Shen, J., Johanning, G.L., and Wang-Johanning, F. (2016). Activation of HERV-K Env protein is essential for tumorigenesis and metastasis of breast cancer cells. Oncotarget *7*, 84093–84117. 10.18632/oncotarget.11455.

39. Denne, M., Sauter, M., Armbruester, V., Licht, J.D., Roemer, K., and Mueller-Lantzsch, N. (2007). Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. J Virol *81*, 5607–5616. 10.1128/JVI.02771-06.

40. Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., et al. (2015). Inhibiting DNA Methylation Causes an

Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell *162*, 974–986. 10.1016/j.cell.2015.07.011.

41. Brocks, D., Schmidt, C.R., Daskalakis, M., Jang, H.S., Shah, N.M., Li, D., Li, J., Zhang, B., Hou, Y., Laudato, S., et al. (2017). DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. Nat Genet *49*, 1052–1060. 10.1038/ng.3889.

42. Liu, M., Thomas, S.L., DeWitt, A.K., Zhou, W., Madaj, Z.B., Ohtani, H., Baylin, S.B., Liang, G., and Jones, P.A. (2018). Dual Inhibition of DNA and Histone Methyltransferases Increases Viral Mimicry in Ovarian Cancer Cells. Cancer Res *78*, 5754–5766. 10.1158/0008-5472.CAN-17-3953.

43. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res *47*, D442–D450. 10.1093/nar/gky1106.

44. McIlroy, D., Barteau, B., Cany, J., Richard, P., Gourden, C., Conchon, S., and Pitard, B. (2009). DNA/amphiphilic block copolymer nanospheres promote low-dose DNA vaccination. Mol Ther *17*, 1473–1481. 10.1038/mt.2009.84.

45. Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. Nucleic acids research *48*, D682–D688. 10.1093/nar/gkz966.

46. Nakagawa, S., and Takahashi, M.U. (2016). gEVE: a genome-based endogenous viral element database provides comprehensive viral protein-coding sequences in mammalian genomes. Database : the journal of biological databases and curation *2016*, baw087. 10.1093/database/baw087.

47. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12. 10.14806/ej.17.1.200.

48. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. 10.1093/bioinformatics/btu170.

49. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. 10.1093/bioinformatics/bts635.

50. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323. 10.1186/1471-2105-12-323.

51. Purcell, A.W., Ramarathinam, S.H., and Ternette, N. (2019). Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. Nat Protoc *14*, 1687–1707. 10.1038/s41596-019-0133-y.

52. Pandey, K., Ramarathinam, S.H., and Purcell, A.W. (2021). Isolation of HLA Bound Peptides by Immunoaffinity Capture and Identification by Mass Spectrometry. Curr Protoc *1*, e92. 10.1002/cpz1.92.

53. UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res *49*, D480–D489. 10.1093/nar/gkaa1100.

54. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research *12*, 2825–2830.

55. Lorenz, R., Bernhart, S.H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. Algorithms for Molecular Biology *6*, 26. 10.1186/1748-7188-6-26.

56. Cancer Genome Atlas Research Network, Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J.M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet *45*, 1113–1120. 10.1038/ng.2764.

57. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 10.48550/arXiv.1303.3997.

58. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*, giab008. 10.1093/gigascience/giab008.

59. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. 10.1101/gr.107524.110.

60. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biology *17*, 122. 10.1186/s13059-016-0974-4.

61. Davidson-Pilon, C. (2019). lifelines: survival analysis in Python. Journal of Open Source Software *4*, 1317. 10.21105/joss.01317.

62. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559. 10.1186/1471-2105-9-559.

63. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods *17*, 261–272. 10.1038/s41592-019-0686-2.

64. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. 10.1186/s13059-014-0550-8.

65. Xie, Z., Bailey, A., Kuleshov, M.V., Clarke, D.J.B., Evangelista, J.E., Jenkins, S.L., Lachmann, A., Wojciechowicz, M.L., Kropiwnicki, E., Jagodnik, K.M., et al. (2021). Gene Set Knowledge Discovery with Enrichr. Curr Protoc *1*, e90. 10.1002/cpz1.90.

66. Fang, Z., Liu, X., and Peltz, G. (2022). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics, btac757. 10.1093/bioinformatics/btac757.

67. Elias, J.E., and Gygi, S.P. (2010). Target-Decoy Search Strategy for Mass Spectrometry-Based Proteomics. In Proteome Bioinformatics Methods in Molecular Biology™., S. J. Hubbard and A. R. Jones, eds. (Humana Press), pp. 55–71. 10.1007/978-1-60761-444-9_5.

METHODS

Mouse cell lines

The BALB/c syngeneic colon cancer cell line CT26 (#CRL2638) was purchased from ATCC and cultured in R10 medium prepared from RPMI (Gibco #72400-021) supplemented with 10% heat inactivated fetal calf serum (FCS, Gibco # 10500-064) at 37 °C and 5% CO2 as per supplier's instructions. The murine melanoma cell line B16F10 derived from a C57BL/6 mouse (ATCC, cat# CRL-6475) was purchased from ATCC and grown in DMEM (Sigma, cat# D6546) supplemented with 10% Fetal Bovine Serum (Gibco, cat# 10500-064) and 1% Glutamax (Gibco, cat# 35050-061). Cells were grown to 60-70% confluency, trypsinized and washed 2x in serum free RPMI in preparation for inoculation in mice.

Animal studies

Animals were maintained in the University of Copenhagen, Panum 10.3, Animal Facility, Copenhagen, Denmark or at the animal facility at Evaxion Biotech, Hørsholm, Denmark. All experiments were conducted under license 2017-15-0201-01209 from the Danish Animal Experimentation Inspectorate in accordance with the Danish Animal Experimentation Act (BEK nr. 12 of 7/01/2016), which is compliant with the European directive (2010/63/EU).

6-8 week old female BALB/cJRj and C57BL/6JRj SPF mice were acquired from Janvier Labs (France). The mice were acclimated for one week before initiation of experiments. BALB/cJRj and C57BL/6JRj mice were subcutaneously (s.c.) inoculated with 2 x 10^5 CT26 or B16F10 tumor cells, respectively, on the left flank in a volume of 100 µl serum free RPMI. Upon establishment, tumors were measured three times a week using a digital caliper and tumor volumes, *V*, were calculated using the following formula:

$$V = \pi * (d_1 * d_2)^{\frac{3}{2}}$$

where d_1 and d_2 are the orthogonal diameters of the tumor. B16F10 tumors were harvested on day 15 and CT26 tumors were harvested on day 19-21. Tumors were immersed in RNAlater for RNA-seq or snap frozen in an ethanol bath on dry ice and maintained at -80 °C before immunopeptidomics analysis.

RNA sequencing of mouse tumors

Total RNA was extracted from three CT26 tumors and three B16 tumors using the RNeasy Plus Mini Kit (Qiagen #74134) and shipped to GenomeScan (Leiden, The Netherlands) for sequencing. RNA-seq libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB #E7760S/L) following the kit protocol. Briefly, mRNA was isolated from total RNA using oligo-dT magnetic beads, fragmented and used for cDNA synthesis. Sequencing adapters were ligated to the resulting cDNA and PCR amplified to create the final sequencing libraries. The libraries were sequenced on an Illumina NovaSeq 6000, generating ~40M clusters of 150bp PE reads for each tumor sample.

Quantification of EVE expression

Reference genomes and associated GTF annotation files were downloaded from Ensembl[45] for mouse (GRCm38.102) and human (GRCh38.104). Mouse- and human EVE nucleotide and amino acid sequences and GTF files were retrieved from the gEVE database[46] v1.1. EVEs mapping to alt contigs not present in the Ensembl reference genomes were removed from the gEVE GTFs.

For both mouse and human, a combined reference + EVE GTF was created by concatenating the reference GTF files with modified gEVE GTF files where each feature field had been modified to "transcript" and "exon". The combined GTF and reference genome fasta files were used to create STAR and RSEM indexes.

Raw fasta files for each tumor sample were preprocessed using cutadapt[47] v1.18 and trimmomatic[48] v0.38 to remove adapter sequences and poor quality bases. Preprocessed RNA reads were mapped to the GRCm38/GRCh38 reference using STAR[49] v2.7.9a and transcript expression was quantified using RSEM[50] v1.3.1. For the mouse samples, combined expression profiles were calculated using the median TPM for each transcript.

Immunoprecipitation of pMHC complexes

Cells were lysed in a lysis buffer containing 1% IGEPAL (Sigma), 50 mM Tris pH 8, 150 mM NaCl and protease inhibitors (cOmplete™, EDTA-free Protease Inhibitor Cocktail tablets, Roche Molecular Biochemicals) as described previously[51,52]. Lysates were cleared by centrifugation and MHC-peptide complexes were isolated using antibodies specific to the MHC molecule(s) of interest. Kd-, Ld- and Dd-associated peptides were selectively immunopurified for CT26 cell lines using the SF1.1.10, 28-14-8s and 34-1-2s antibodies, and a pan-MHCI

approach was employed for the CT26 tumor and both B16 experiments where all three antibodies were mixed in equal proportions. The HLA-peptide complexes were then eluted using 10% acetic acid and subsequently subjected to molecular weight cut-off filters (small scale) or fractionated using a reversed-phase C18 end capped HPLC column (Chromolith SpeedRod, Merck), running a mobile phase buffer A of 0.1% trifluoroacetic acid (TFA) and buffer B of 80% acetonitrile (ACN) and 0.1% TFA. The peptides were separated using a gradient of 2-40% buffer B for 4 minutes and 40-45% for another 4 min. Fractions collected were then concentrated using centrifugal evaporation and reconstituted in 2% ACN in 0.1% formic acid (FA) prior to analysis by LC-MS/MS.

Data acquisition by LC-MS/MS

Reconstituted peptides were analyzed using a timsTOF Pro mass spectrometer, where reconstituted peptides were loaded onto an IonOpticks Aurora 25 cm C18 column using a nanoElute liquid chromatography system (Bruker Daltonics). The peptides were separated using a gradient of buffer B (0.1% FA in ACN) against buffer A (0.1% FA, 2% ACN in water) initially to 17% over 60 min then to 25% over 30 min then to 37% over 10 min with a flow rate of 300 nl/min. Data dependent acquisition (DDA) was performed with the following settings: mz range: 100-1700mz, capillary voltage: 1600V, target intensity of 30000, TIMS ramp of 0.60 to 1.60 Vs/cm2 for 166 ms.

Identification of MHC ligand sequences

LC-MS/MS data was searched with PEAKs Xpro v10.6 (Bioinformatic solutions) against the Uniprot annotated proteome[53] (2020_06 release for mouse and 2021_04 release for human) appended with the gEVE database[46] for the corresponding organism. The search was carried out using the following settings: Enzyme: none, parent mass tolerance 20 ppm, fragment mass tolerance 0.02 Da, Variable modifications: Deamidation (NQ), Oxidation (M). For the B16F10 datasets, peptides with C-terminal lysine- and arginine residues were filtered to remove potential tryptic peptide contamination. Incorrect peptide identification was estimated based on the target-decoy search strategy described previously[67] where the false discovery rate (FDR) variable indicates the proportion of likely incorrect results contained at any given export file. All searches were exported with an FDR of 5%.

Benchmarking of MHC ligand predictions

For each MHC ligand identified by immunopeptidomics, source protein/EVE expression levels were determined as described previously. MHC ligand predictions were generated for all relevant MHC alleles (H2-Dd, H2-Kd, H2-Ld for CT26 and H2-Db and H2-Kb for B16) using a proprietary MHC ligand prediction tool developed by Evaxion Biotech. As input, the tool was provided with the ligand sequences and the source protein/EVE expression values. As output, the tool calculated two MHC ligand probability scores based on: i) the amino acid sequence alone and ii) the amino acid sequence and the expression level.

To generate positive datapoints for the benchmark datasets, the MHC ligand sequences determined immunopeptidomics were filtered to only keep ligands that passed the following criteria: i) -log10(p) > 20, ii) peptide length 8-11, iii) derived from a source transcript/EVE with TPM>1 and iv) not included in the training dataset of the MHC ligand prediction tool. Negative datapoints were generated by sampling peptide sequences from the Ensembl reference proteome and the gEVE database. Predictions were generated for all negative datapoints as described above. Negative datapoints were subject to the same filters as those applied to the positive datapoints. The final benchmark datasets were compiled by sampling negative datapoints at a 30:1 negative to positive ratio. For each datapoint in the benchmark datasets a random value was sampled from a uniform distribution between 0-1 using Numpy v1.23.5. Average precision scores were calculated using scikit-learn[54] v1.1.3.

Selection of EVE-derived epitopes

EVE epitope sequences were selected by first predicting MHC ligands. Predictions were generated for all peptide sequences in the gEVE database of lengths 8-11 for MHCI molecules and of length 15 for MHCII molecules. For each EVE amino acid sequence, all unique 27mers were extracted and the best MHCI and -II ligands were determined based on predicted ligand probabilities. A combined MHC ligand score, *S*, was calculated for each 27mer using the following equation:

$$S = 1 - \sqrt{(1 - P_{MHCI}) * (1 - P_{MHCII})}$$

Where P_{MHCI} and P_{MHCII} are the ligand probabilities for MHC class I and -II respectively. The list of 27mers was sorted by their combined scores and filtered to keep only the highest scoring 27mer from each EVE and the top ranking 27mers from the resulting sorted- and filtered list were selected.

For preclinical testing with the CT26 cancer model, MHC ligands were predicted for the BALB/c MHC type, i.e., H2-Dd, H2-Kd, H2-Ld and H2-IAd as described above. The top 13 ranking EVE 27mer sequences were selected. For testing immune responses in the clinical samples, MHC ligand predictions were generated against the respective patients' HLA type for the loci HLA-A, HLA-B, HLA-C and HLA-DRB1. The top 20 ranking EVE 27mers were selected and split into two peptide pools according to their ranking, I.e., pools with peptides ranking 1-10 and 11-20, respectively.

Immunopeptidomics-driven selection of MHC ligands was carried out by first identifying all peptides that were found across all three CT26 samples; the IFNγ-treated cell pellet, the untreated cell pellet, and the pool of five CT26 tumors. All peptides of length 8-13 amino acids, without post-translational modifications (PTMs) were included. Further, peptides were only included if PEAKS had assigned the peptide an area value above zero. Next, peptides were mapped to their transcript to determine their expression level. Since EVE ligands often map to more than one EVE gene, each multi-mapper ligand was only paired to the associated gene with the highest expression level. The prioritization of the candidate peptides was based on the sum of their chromatographic peak area for the three biological samples (two cell line samples and one tumor sample), the mean of the identification confidence and mean expression level with a cut-off of 1 TPM. The top 13 EVE candidates were selected for preclinical testing, allowing up to two ligands per gene.

Generation of DNA plasmids encoding EVE epitopes

DNA sequences were generated for each EVE epitope using reverse translation through weighted random sampling of codons based on observed codon frequencies in the mouse genome. The EVE epitope DNA sequences were then combined with linker sequences encoding 10 amino acid glycine-serine linkers to form a DNA insert sequence. This process was repeated to generate 10 000 DNA insert sequences. The list of DNA sequences was then filtered to retain sequences with the lowest GC content and highest codon adaptation index (CAI). The GC content and CAI thresholds were defined so that at least 20% of the sequences were retained when each threshold was applied individually, and at least 10% of the sequences when applied together. The resulting list of DNA sequences was further filtered to remove any containing Kozak sequences. The predicted minimum free energy (MFE) of the corresponding RNA molecule to each remaining DNA sequence was calculated using ViennaRNA[55] v2.4.18 and the best scoring insert was selected. A Kozak sequence and restriction sites were added to final DNA insert before it was manufactured and cloned into a pTVG4 DNA plasmid. The pTVG4 DNA vector was generated from the standard plasmid pUMVC3 acquired from Aldevron (North Dakota, USA, #4010) by cloning two copies of a 36-

bp CpG-rich immunostimulatory DNA sequence (ISS) containing the 5'-GTCGTT-3' motif downstream of the multi-cloning site. The EVE epitope containing pTVG4 DNA plasmid was upscaled by Aldevron. An empty pTVG4 DNA vector with no cloned EVE epitope DNA insert was also ordered to serve as a mock control.

Tumor challenge experiments

Mice from the different experimental groups (13 mice per group) were distributed across different cages to avoid potential cage effects. Mice were vaccinated weekly in left and right tibialis anterior muscles (i.m.) with 100 µg of research-grade DNA for a total of five immunizations. Vaccination commenced two weeks prior to subcutaneous CT26 cell inoculation (defined as study day 0). In the first two immunizations, 2 x 50 µL vaccine comprising DNA formulated in PBS was administered using Electroporation (EP) (BTX Agile pulse, VWR). In the remaining three immunizations, DNA was formulated with block co-polymer poloxamer 188 (gifted by BASF, Germany) to a final concentration of 3% in PBS and administered in 2 x 75 µL vaccine solution. Poloxamer has been shown to increase the longevity of the pDNA after injection and thereby increase antigen expression and exposure[44]. At the day of primary CT26 cell inoculation, 2 x 10^5 of *in vitro* expanded CT26 cells were inoculated subcutaneously (s.c.) in the right flank of the mice. For the tumor re-challenge experiment, the CT26 cells were inoculated in the opposite (left) flank. Upon establishment, tumors were measured as described above. Mice were euthanized through cervical dislocation when the majority of tumors in the control groups reached the maximum allowed size of 15 mm diameter in either direction or upon reaching humane endpoints.

Detection of immunogenic epitopes with IFNγ ELISPOT assay

T-cell reactivity against vaccine-encoded epitopes was measured *ex vivo* using the Enzyme-Linked ImmunoSPOT (ELISPOT) assay. In brief, 5 x 10^5 of viable splenocytes suspended in R10 medium [RPMI 1640 (Fisher Scientific #72400021) + 10% FBS (Fisher Scientific #10500064)] were plated on an ELISPOT plate (Merck Millipore #MAIP4510) coated overnight with 5 µg/mL anti-murine IFNγ capture antibody (BD #51-2525KZ). Cells were stimulated with 5 µg/mL of synthetic vaccine-corresponding 27mer peptides, peptide pools or left unstimulated (R10 medium only). Cells were stimulated for 20 hours at 37 °C and 5% CO_2. IFNγ produced upon re-activation of peptide-specific T cells and captured on anti-IFNγ capture antibody was visualized via sequential addition of 2 µg/mL biotinylated anti-IFNγ detection antibody (BD #51-1818KA), Streptavidin-Horseradish Peroxidase (HRP) (BD #557630) and AEC chromogen substrate (BD #551951) as per manufacturer's protocol. After overnight drying, the ELISPOT plates were imaged in an ELISPOT reader (Cellular

Technology, Ltd) and bound IFNγ was calculated as Spot Forming Units (SFUs) per 1 x 10^6 splenocytes.

MHC class I dextramer staining for AH1-specific T cells

To monitor T-cell immune recognition of vaccine epitopes in blood of mice, the MHCI dextramer staining methodology was used. MHCI dextramers comprised H2-Ld molecules loaded with the restricted minimal peptide SPSYVYHQF derived from the MuLV gp70 protein and encoded by the vaccine epitopes MS-EVE 1 (ELWPRVTYHSPSYVYHQFERRAKYKRE) and ObsERV 3 (TYHSPSYVYHQFERRAKYKREPVSLTL). MHCI dextramer was acquired from Immudex (#JG3294) and fluorescently labelled with PE. In brief, 50 µL of tail vein blood was collected in EDTA-coated tubes (Sarstedt #20.1278.100) and then transferred in a 96 deep-well plate (Sigma #575653). Blood cells were then incubated with the FcR-blocking anti-murine CD16/32 antibody (Biolegend #101301) for 10 minutes at 4 °C. Prior to use, the dextramer reagent was centrifuged at 3300 g for 5 minutes. 10 µL of the dextramer reagent was added in each blood sample and incubated at 37 °C for 15 minutes. Cells were subsequently stained with fluorescently labelled antibodies specific for murine CD3e (FITC, Biolegend #100305), CD4 (PE-Cy7, BD #552775), and CD8 (BV786, BD #563332) at 4 °C for 30 minutes. Finally, cells were subjected to one-step fixation/red blood cell lysis (eBioscience #00-5333-57) before acquisition on FACS Celesta (BD). Determination of AH1 reactive cells out of CD8+ T cells were determined using FlowJo Software (version 10.8.0). The gating strategy is presented in Supplementary Figure S2A.

Stimulation and Intracellular cytokine staining (ICS) of epitope-specific T cells

To characterize the functional T-cell responses elicited against vaccine epitopes, 2 x 10^6 viable splenocytes in single cell suspension were plated in round bottomed, 96-well culture plates (Corning #3799) in RPMI 1640 medium (Fisher Scientific #72400021) supplemented with heat-inactivated FBS (Fisher Scientific #10500064). Cells were stimulated with synthetic, 5 µg/mL of vaccine-corresponding 27mer peptides, peptide pools or left unstimulated (medium only). Protein transport was inhibited by addition of brefeldin A (GolgiPlug BD #555029) and monensin (GolgiStop BD #554724) two hours after the initiation of stimulation, followed by overnight incubation at 37 °C and 5% CO_2. Cells were then washed and incubated with the FcR-blocking anti-murine CD16/32 antibody (Biolegend #101301) at 4 °C for 10 minutes, followed by staining for murine CD3e (FITC, Biolegend #100305), CD4 (PE-Cy7, BD #552775), CD8 (BV786, BD #563332), and viability dye (GloCell™ Fixable Viability Dye, Stem Cell Tech #75010) at 4 °C for 30 minutes. Cells were then fixed (eBioscience #00-8222-49) at 4 °C for 20 minutes and permeabilization

(eBioscience #00-8333-56) before staining for intracellular IFNγ (BV650, BD #563854) and TNFα (BV421, BD # 566287) at 4 °C for 30 minutes. Stained cells were acquired on FACS Celesta (BD) and frequencies of cytokine-producing CD4+ and CD8+ T cells were determined in FlowJo Software (version 10.8.0). The gating strategy is presented in Supplementary Figure S2B.

Preclinical statistical analyses

GraphPad Prism 9 for Mac OS X was used for graphing, statistical analyses, and tools. Data were subjected to Kolmogorov-Smirnov test for normality (alpha = 0.05). Parametric data were analyzed by ordinary ANOVA with Sidak's multiple comparison correction. Non-parametric data were analyzed by Mann-Whitney test (if two comparisons) or Kruskal-Wallis test with Dunn's multiple comparison correction (if more than two comparisons). For all preclinical results, the following levels of statistical significance are applied: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.

Patient stratification based on EVE burden

RNA-seq of tumor biopsies were retrieved from Hugo et al.[24], Lauss et al.[25] , Riaz et al.[26] and Mariathasan et al.[27]. Transcript expression levels were quantified as described above (Section: Quantification of EVE expression) and the EVE burden defined as the number of EVE transcripts with expression levels above 1 transcripts per million (TPM). A threshold of 50 expressed EVEs was applied to group the melanoma patients into EVE High and EVE Low strata, while a threshold of 15 expressed EVEs was applied for the bladder cancer patients. RNA-seq of tumor biopsies from select cancer types were retrieved from the TCGA database[56] and the EVE burden computed following the same approach.

WES of matched tumor and healthy tissue biopsy was retrieved from Hugo et al.[24] and Mariathasan et al.[27]. The WES reads were mapped to the human reference genome (GRCh38) with BWA-MEM[57] v0.7.17-r1188. Read duplicates were marked and removed using SAMtools[58] v1.13 followed by somatic variant calling with GATK-Mutect2[59] v4.2.2.0. Somatic variants were annotated using the variant effect predictor[60] v104 and the Ensembl GRCh38.104 annotation reference. Furthermore, annotated somatic variants were retrieved from the supporting material of Lauss et al.[25] and Riaz et al.[26]. The tumor mutational burden (TMB) was computed as the number of missense somatic variants. For Hugo et al.[24], the median TMB was used to group the patients into TMB High and TMB Low strata, while a threshold of 150 was used for Mariathasan et al.[27]. The reported TMB High and TMB Low strata in the source studies were used for Lauss et al.[25] and Riaz et al.[26].

Survival analysis was conducted with patient stratification based on TMB and EVE burden using the python package lifelines v0.27.3[61].

Pan cancer analysis

Matched TCGA deposited RNA-seq and Mutect2 somatic variants calls were retrieved. The EVE burden was computed as described above, while the Mutect2 variant calls were subset to the coding regions followed by annotation with the variant effect predictor[60] v104.

Gene co-expression analysis

Transcript expression levels calculated from the Hugo et al.[24], Lauss et al.[25] and Riaz et al.[26] datasets were combined and filtered to remove all transcripts that were not expressed with at least 1 TPM in at least 10% of the samples. Gene co-expression analysis was then performed using WGCNA[62] v1.71 with a soft thresholding power of 20. Module eigengenes were correlated to EVE burden using the Spearman rank correlation coefficient function from SciPy[63] v1.9.3.

Differential expression analysis

Transcript read counts were calculated using RSEM[50] v1.3.1 for the the Hugo et al.[24], Lauss et al.[25] and Riaz et al.[26] datasets. Differential expression analysis was conducted using the R package DESeq2[64] v1.26.0 with the formula '~ *eve_burden* + *cohort'.* P-values were adjusted for multiple testing using the Benjamin-Hochberg procedure and differential expression defined by a false discovery rate below 0.05 and log2 fold change amplitude larger than 1.

Gene set enrichment

Transcript IDs were converted to gene names using pyensembl v2.1.0 and duplicate gene names were collapsed. The resulting gene lists were then used for GSEA using the Enrichr[65] module from GSEApy[66] v0.14.0.

Data and code availability

- Raw RNA sequencing data generated in this study have been deposited in the NCBI GEO database under accession number GSE223515.

- Raw mass spectrometry proteomics data have been deposited in ProteomeXchange Consortium via the PRIDE[43] partner repository under accession numbers PXD040085 and PXD040165
- Accessions for datasets used in this study include: GSE78220, GSE100797, GSE91061 and PRJNA307199