# ISOTOPE: ISOform-guided prediction of epiTOPEs in cancer

Juan L. Trincado[1], Marina Reixachs-Sole[2,3], Judith Pérez-Granado[4], Tim Fugmann[5,6], Ferran Sanz[4], Jun Yokota[7,8], Eduardo Eyras[2,3,9,10*]

[1]Josep Carreras Leukemia Research Institute, 08916 Badalona, Spain.
[2]Australian National University, 2601 Canberra, Australia.
[3]EMBL Australia Partner Laboratory Network at the Australian National University, 2601 Canberra, Australia.
[4]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Dept. of Experimental and Health Sciences, Pompeu Fabra University (UPF), 08003 Barcelona, Spain.
[5]Philochem AG, Otelfingen, Switzerland.
[6]Current address: Max Delbrück Centrum for Molecular Medicine, 13125 Berlin, Germany.
[7]Institute of Predictive and Personalized Medicine of Cancer (IMPPC), 08916 Badalona, Spain.
[8]National Cancer Center Research Institute (NCCRI), Tokyo 104-0045, Japan.
[9]Catalan Institution for Research and Advanced Studies. 08010 Barcelona, Spain
[10]Hospital del Mar Medical Research Institute (IMIM), 08003 Barcelona, Spain.

* correspondence to eduardo.eyras@anu.edu.au

## Abstract

Immunotherapies provide effective treatments for previously untreatable tumors, but the molecular determinants of response remain to be elucidated. Here, we describe a pipeline, ISOTOPE (ISOform-guided prediction of epiTOPEs In Cancer), for the comprehensive identification of cancer-specific splicing-derived epitopes. Using RNA sequencing and mass spectrometry for MHC-I associated proteins, ISOTOPE identified neoepitopes from cancer-specific splicing event types that are potentially presented by MHC-I complexes. We found that, in general, cancer-specific splicing alterations led more frequently to the depletion of potential self-antigens compared to the generation of neoepitopes. The potential loss of native epitopes was validated using MHC-I associated mass spectrometry from normal cells. Furthermore, analysis of two cohorts of melanoma patients with ISOTOPE identified that splicing-derived neoepitopes with higher MHC-I binding affinity associate with positive response to immune checkpoint blockade therapy. Additionally, we found a more frequent depletion of native epitopes in non-responders, suggesting a new mechanism of immune escape. Our analyses highlight the diversity of the immunogenic impacts of cancer-specific splicing alterations and the importance of studying splicing alterations to fully characterize the determinants of response to immunotherapies. ISOTOPE is available at https://github.com/comprna/ISOTOPE

## Introduction

Recent developments in the modulation of the immune system have revolutionized the clinical management of previously untreatable tumors. In particular, therapies targeting negative regulators

of immune response, i.e. immune checkpoint inhibitors, have shown prolonged remission in several tumor types (Ribas and Wolchok 2018). However, these therapies appear to be effective only for about one third of the patients (Antonia et al. 2016). Thus, characterizing the molecular features driving response to immune therapies is crucial to prospectively identify patients who are most likely to benefit from these agents and avoid exposing resistant patients to unnecessary and potentially harmful treatments.

The ability of the T-cells infiltrating the tumor tissue to identify and attack malignant cells relies on tumor cells maintaining sufficient antigenicity. An approach to determine the antigenicity of a tumor is through the calculation of the frequency of somatic mutations as a proxy for the production of tumor neoantigens. This has led to the identification of an association between response to checkpoint inhibitors and tumor mutation burden (TMB) in various tumor types, including melanoma (Snyder et al. 2014; Van Allen et al. 2015) urothelial carcinoma (Rosenberg et al. 2016), and lung cancer (Rizvi et al. 2015; Antonia et al. 2016). Furthermore, analysis of how somatic substitutions and indels impact the protein products in tumor cells has enabled the identification of cancer-specific neoepitopes (Wood et al. 2020; Wang et al. 2019) that can trigger the attack of the immune system against tumor cells during treatment with immune checkpoint inhibitors. However, TMB or mutation-derived neoepitopes can only explain a fraction of the responders (Goodman et al. 2017), and hence, other molecular signatures and sources of neoepitopes need be identified.

Recently, cancer-specific transcriptome alterations have been shown to be a source of tumor neoantigens that can be presented by the MHC complexes and recognized by T-cells. These include gene fusions (Zhang et al. 2016), RNA editing (Zhang et al. 2018), cryptic expression (Andersen et al. 2013; Laumont et al. 2018), and cancer-specific splicing (Andersen et al. 2013; Jayasinghe et al. 2018; Kahles et al. 2018; Smart et al. 2018). In particular, the aberrant selection of splice sites and exon-exon junctions (Jayasinghe et al. 2018; Kahles et al. 2018) or the retention of introns (Andersen et al. 2013; Smart et al. 2018) represents an additional potential source of cancer neoepitopes. However, it is not yet clear whether these splicing-derived neoepitopes provide a mechanism to elicit cancer-specific immune responses and whether they may improve patient response to immune therapy.

To address these questions, we developed ISOTOPE (ISOform-guided prediction of epiTOPEs in cancer), a new method to exhaustively identify the immunogenic impacts from cancer-specific splicing alterations. ISOTOPE identifies all cancer-specific splicing alterations including novel exon-exon junctions, aberrant splice-sites, intron retention, as well as de-novo exonizations, and calculates the impacts on the encoded proteins and the candidate epitopes. A novelty of ISOTOPE is that besides the cancer-specific epitopes, it also calculates the native epitopes potentially lost in the tumor, i.e. potential self-antigens that are depleted by the splicing alterations. Application of ISOTOPE to cancer cell lines with RNA-seq and MHC-I class associated mass spectrometry allowed us to validate a number of splicing-derived neoepitopes and to find that many more native epitopes are depleted. We confirmed this through the analysis of small cell lung cancer samples and validation with MHC-I associated peptides from normal cells. Finally, we used ISOTOPE to analyze two cohorts of melanoma patients and found that MHC-I-binding affinity associates with response to immune therapy and that self-antigens are more frequently depleted in non-responders. Our analyses suggest

that splicing alterations not only can generate novel cancer-specific epitopes, but could also potentially impact the response to immune therapy by depleting native epitopes, providing a potential mechanism of immune escape. This highlights the importance of studying splicing alterations to fully characterize the determinants of response to immunotherapies.

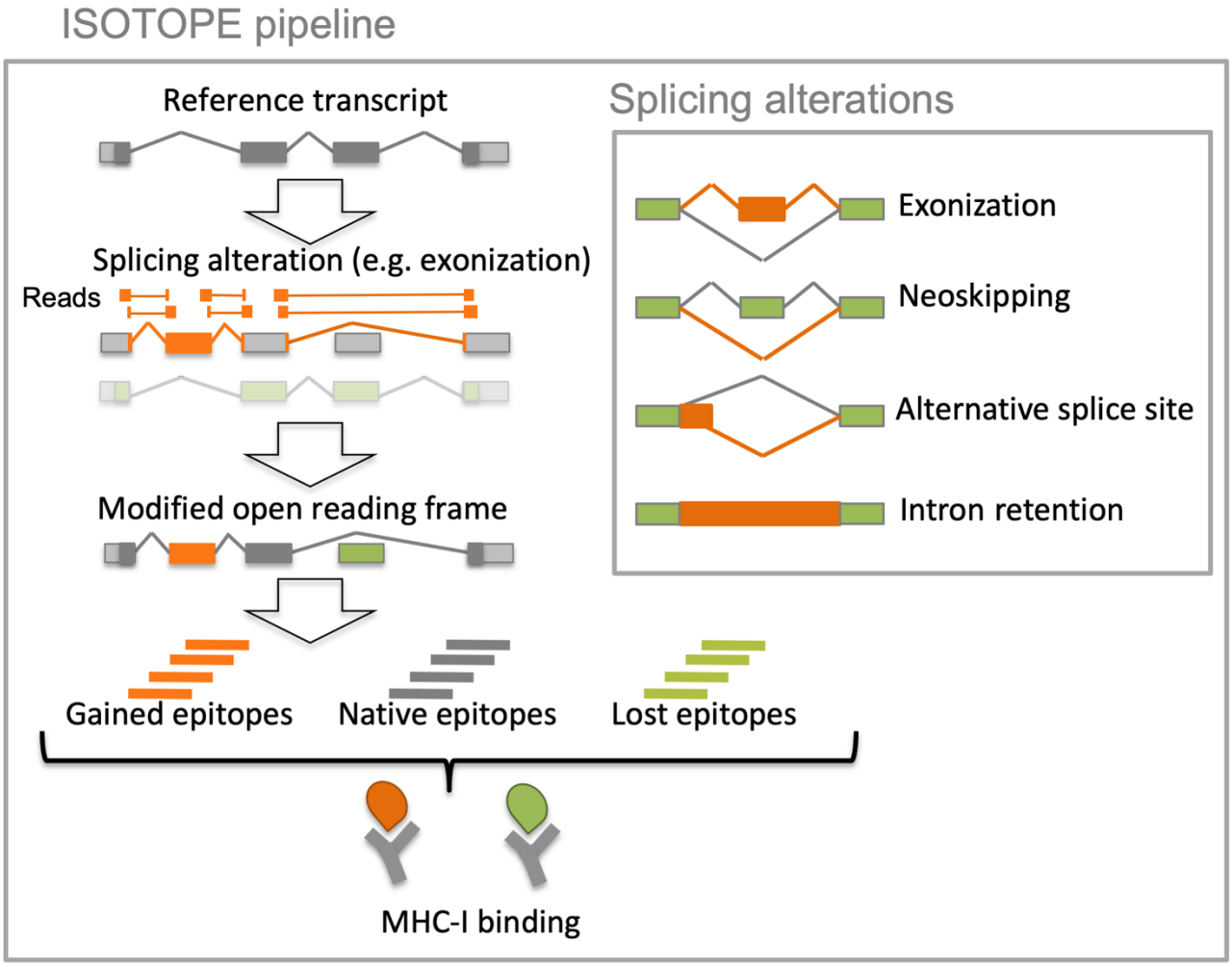


**Figure 1. ISOTOPE pipeline.** Cancer-specific splicing alterations are defined as significant variations with respect to the exon-intron structures expressed in normal samples and are classified into four different types: alternative (5'/3') splice-site, new exon skipping (neo-skipping), *de novo* exonizations and intron retention. The figure illustrates an exonization. ISOTOPE calculates the modified open reading frame (ORF) from the reference open reading frame from the annotated transcript with the highest expression in control samples using the splicing alterations, and test resulting modified peptides for affinity with the MHC-I class (see Methods for details). Candidate MHC-I binders are further searched in mass-spectrometry data. ISOTOPE also calculates MHC-I binding peptides that are encoded by the transcript expressed in the control samples but are depleted due to the splicing alteration.

## Results

### Comprehensive identification of cancer-specific splicing-derived epitopes

ISOTOPE identifies cancer-specific splicing alterations by generating a catalogue of all exon-exon junctions calculated from RNA sequencing (RNA-seq) reads from tumor samples, filtering out those

that appear in any of the samples from a large dataset of normal controls (Fig. 1). The remaining junctions are classified into one of four possible types: alternative splice site, new exon skipping event, intron retention, and *de novo* exonization. This last type has not been previously considered in the context of neoepitope generation. Additionally, ISOTOPE performs a bootstrapping test to establish the significance of the read support of the candidate splicing alterations. This type of test has not been performed in previous similar analyses of splicing-derived neoepitopes and ensures the robustness of the events detected. Changes in the open reading frame (ORF) induced by the splicing alterations are predicted using a reference transcriptome obtained by selecting for each gene the transcript with highest mean expression across control samples (Methods). ISOTOPE identifies potential epitopes by calculating the binding affinity of the encoded peptides to the major histocompatibility complex class I (MHC-I) or II (MHC-II) using NetMHC-4.0.0 (Andreatta and Nielsen 2016). When the human leukocyte antigen (HLA) type for a patient is not available, this is calculated directly from the tumor RNA-seq sample. ISOTOPE defines candidate tumor neoepitopes as the MHC binders that are expressed in the tumor sample but not in the control samples. Additionally, ISOTOPE identifies depleted self-antigens as MHC binders expressed in the control sample but not in the tumor sample (see Methods for details).

To validate the accuracy of predicting HLA types directly from the tumor RNA-seq, we calculated HLA-I and HLA-II types from RNA-seq reads from tumor and matched normal samples for 24 small cell lung cancer (SCLC) patient samples (Rudin et al. 2012) using PHLAT (Bai et al. 2014) and Seq2HLA (Boegel et al. 2018). Both methods showed an overall agreement between the HLA predictions from the tumor and the normal RNA-seq data (Fig. 2a). In particular, PHLAT recovered above 80% of cases for HLA-I and between 65% and 90% for HLA-II types (Supp. Table 1). As PHLAT showed greater consistency across most of the HLA types, we decided to use this method for further analyses with ISOTOPE.

**Detection of cancer-specific splicing-derived epitopes in MHC-I mass-spectrometry**

To validate ISOTOPE ability to identify potential neoepitopes, we analyzed RNA-seq data and MHC-I associated proteomics data for the cancer cell lines CA46, HL-60 and THP-I (Barretina et al. 2012; Ritz et al. 2016). We found that *de novo* exonization events are the least common of all types, and that new junctions skipping one or more exons, i.e. neo-skipping events, and aberrant splice-sites were approximately as frequent as intron retention (Fig. 2b). Further, we found that the majority of splicing alterations did not affect the encoded protein (Fig. 2b). However, neo-skipping events impacted more frequently the protein product compared with the other event types (Fig. 2b).

In total we found 2108 genes with predicted alterations in the protein product due to cancer cell specific splicing alterations, with a similar number of protein-affecting splicing changes in each cell line CA46: 1368, HL-60: 1043, and THP-I: 1700. Further, the predicted HLA-types from the RNA-seq data with PHLAT matched those previously reported (Ritz et al. 2016). We then predicted candidate MHC-I binding peptides (binding affinity ≤ 500nM) with NetMHC on all peptides, keeping only those splicing-derived peptides that are not encoded in the reference transcriptome (Methods). This produced 830 (CA46), 461 (HL-60) and 2072 (THP-1) candidate neoepitopes (Fig. 2c) (Supp. Table 2). Neoskipping events produced more candidate neoepitopes in all cell lines compared with

the other event types (Fig. 2c). On the other hand, despite being less frequent, *de novo* exonizations produced a similar number of neoepitopes similar compared to intron retention events. Intriguingly, candidate self-antigens that were depleted by splicing alterations were more common than neoepitopes (Fig. 2c). These were peptides with high affinity to the MHC-I complex encoded in the reference transcript but that would be depleted or lost by the splicing alteration. We further separated candidate epitopes (spliced-derived neoepitopes and depleted epitopes) according to HLA-types (Fig 2d). The distribution follows closely that of the cell lines, indicating an agreement of the MHC affinity of the peptides found in the cell lines and the HLA class predicted.

To test the potential therapeutic implications of these gained and depleted epitopes, we performed enrichment analysis on the genes involved using two databases of treatment-associated responses (Methods) (Supp. Table 3). We observed a significant association with lymphoma in genes depleted of self-antigens, including *ATM*, *EZH2*, *REL*. Alterations on these genes have been connected with therapy response, highlighting the potential clinical relevance of these detected splicing changes.

To validate the splicing-derived neoepitopes we used MHC-1 associated mass-spectrometry (MS) data available for the same cell lines (Ritz et al. 2016). We were able to detect three neoepitopes on three different samples, all three generated by neoskipping events in the genes *TOP1* (KRFEPLGMQK), *ERF* (IPAPDHPAL) and *IFRD2* (RTALGGMSW) (Fig. 2e). These were different from the three peptides detected previously using the same datasets and only analyzing intron retention (Smart et al. 2018). This disparity was possibly due to the different criteria used in the selection of relevant events: we performed a bootstrapping analysis to keep only events with significant read support (Methods) and studied other splicing alterations types beyond intron retention. Our results suggest that new types of splicing alteration can potentially produce tumor neoepitopes.

To further validate our approach, we applied ISOTOPE to RNA-seq data and MHC-I associated mass spectrometry data for ten breast cancer cell lines (MCF7, T47D, LY2, BT549, CAMA1, HCC1395, HCC1419, HCC1428, HCC1569, HCC1806) (Barretina et al. 2012; Rozanov et al. 2018). The most frequent splicing alterations found were IR events, except for cell lines HCC1569 and LY2, for which neo-skipping events were the most frequent (Fig. 3a). However, from all types, neo-skipping events produced the largest number of changes in protein in all cell lines. As before, we predicted the MHC-I binding potential for candidate epitopes, either splicing-derived neoepitopes or splicing-depleted native epitopes. Events of type neoskipping produced the largest number of neoepitopes (Fig. 3b) (Supp. Table 4). As before, we observed more depleted candidate epitopes than splicing-derived neoepitopes. Separating by HLA-type, we observed that candidate neoepitopes were more frequently associated to HLA type A (Fig 3c). Additionally, we identified a significant association of genes involved in treatment response in breast cancer with the gain and loss of epitopes, i.e. *ERBB2*, *ESR1*, *TIMP1*, *ABCC3*, and with self-antigen depletion, i.e. *AKT1*, *CCNE1*, *RET*, *TFF3* (Supp. Table 3). Next we searched the predicted neoepitopes using MHC-I associated mass spectrometry data for the same cell types (Rozanov et al. 2018) (Methods). This search only identified one significant peptide in all cell lines analyzed, which was generated by a neoskipping event resulting in a frameshift in the gene *SIL1* in BT549 (LPAAPLPLCPA, HLA-B) (Fig. 3d).

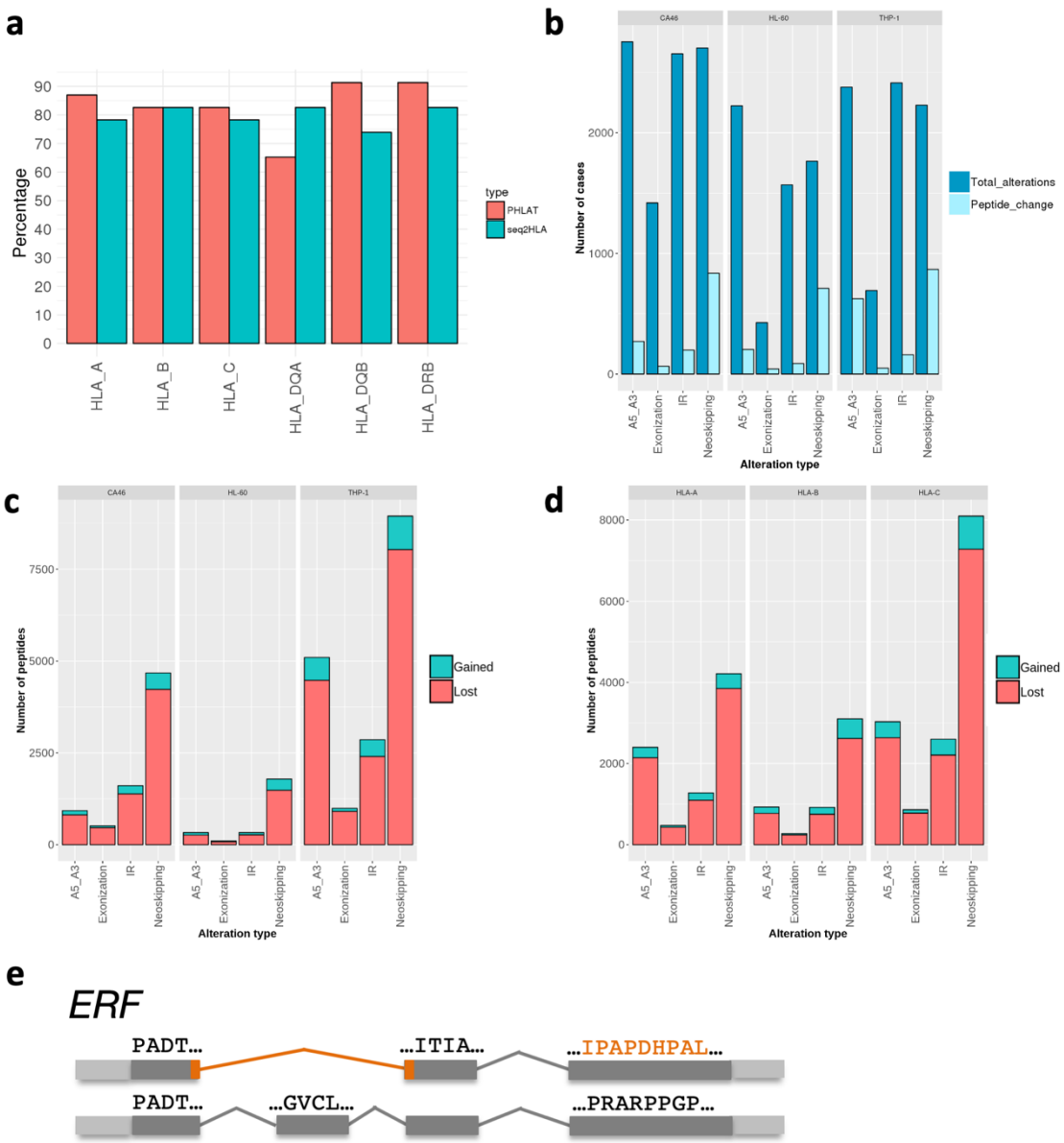


**Figure 2.** **(a)** Validation of the HLA type prediction from tumor RNA-seq data. We show the predictions for MHC Class I (HLA-A, HLA-B, HLA-C) and II (HLA-DQA, HLA-DQB, HLA-DRB) with PHLAT (red) or SeqHLA (blue). Each bar corresponds to the proportion of samples (over a total of 24 small cell lung cancer samples) for which the prediction on the tumor sample coincides with the prediction on the matched normal sample **(b)** For each cell line, CA46, HL-60 and THP-1, we show the number of different splicing alterations measured (dark blue) and the number of cases leading to a change in the encoded open reading frame (light blue). Alterations shown are alternative (5'/3') splice-site (A5_A3), *de novo* exonizations (Exonization), intron retentions (IR), and new exon skipping events (Neoskipping). **(c)** Number of candidate splicing-derived neoepitopes (red) and depleted native epitopes (blue) detected for each of the splicing alterations in each of the cell lines CA46, HL-60 and THP-1, separated by type of splicing alteration. **(d)** Number of candidate splicing-derived neoepitopes (red) and depleted native epitopes (blue) detected for each of the splicing alterations in each of the cell lines CA46, HL-60 and THP-1, separated by HLA-type. **(e)** Example of a splicing-derived neoepitope validated with MHC-I associated mass spectrometry data and derived from a neoskipping event in the gene *ERF*. The peptides are given in the same orientation as the 5' to 3' direction of the gene.

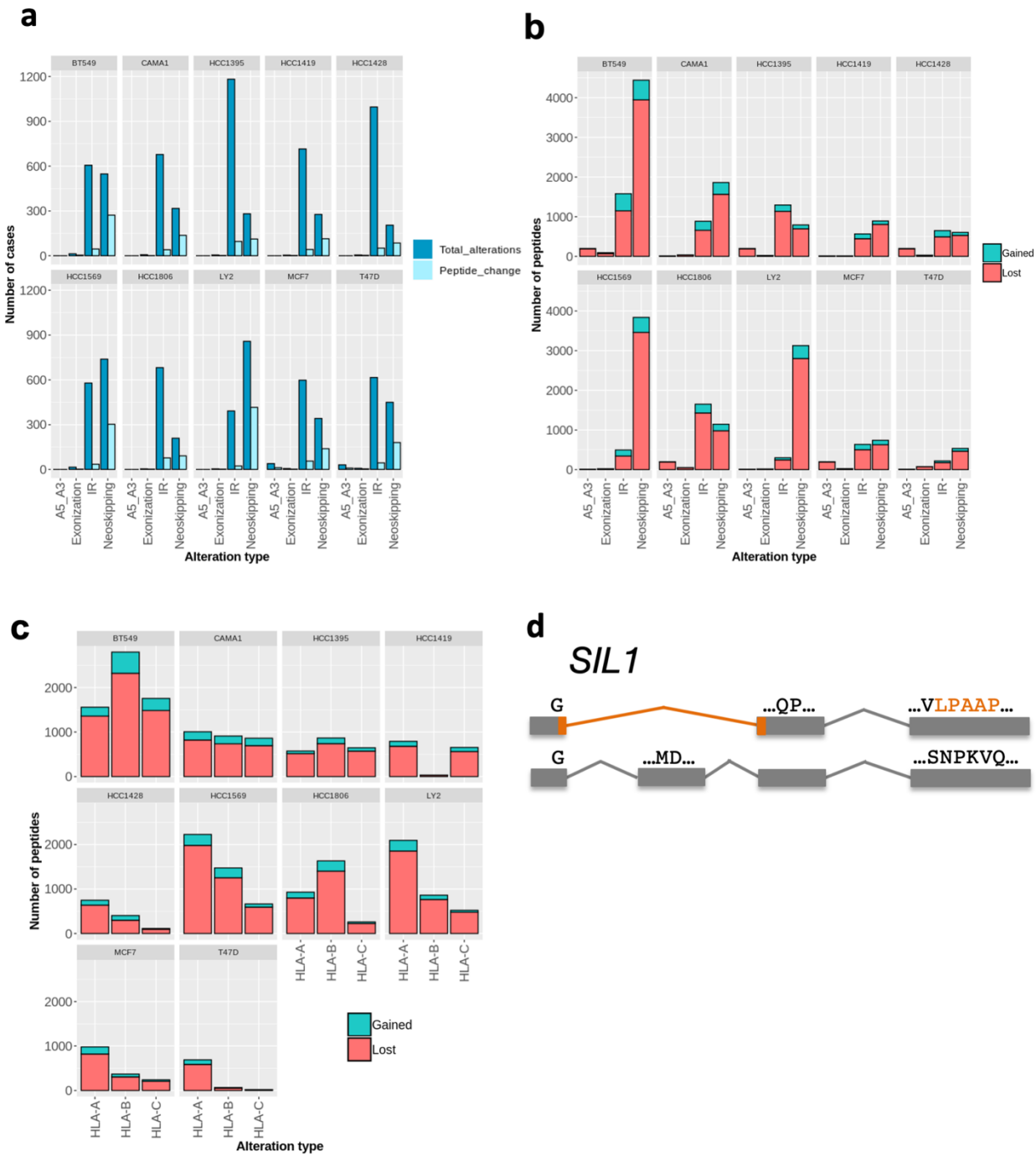


**Figure 3. (a)** For each breast cancer cell line analyzed, the barplots show the number of splicing alterations measured and the number of cases leading to a change in the encoded open reading frame relative to the reference proteome. Alterations shown are alternative 5' or 3' splice-site (A5_A3), *de novo* exonizations (exonization), intron retentions (IR), and new exon skipping events (neoskipping). **(b)** Number of candidate splicing-derived neoepitopes (blue) and depleted native epitopes (red) detected for each of the splicing alterations in each of the breast cancer cell lines tested, separated by type of splicing alteration. **(c)** Number of candidate splicing-derived neoepitopes (blue) and depleted native epitopes (red) detected for each of the splicing alterations in each of the breast cancer cell lines tested separated by HLA-type. **(d)** Example of a splicing-derived neopitope validated with MHC-I associated mass spectrometry data and derived from a neoskipping of gene *SIL1* in cell line BT549.

**Cancer-specific splicing alterations frequently deplete native epitopes**

Tumors show frequent depleted expression of normal transcript isoforms (Sebestyén et al. 2015; Climente-González et al. 2017). Intriguingly, we observed here that these changes lead to the depletion of peptides that are candidate binders of MHC-I. That is, splicing alterations could lead to the removal of a piece of the open reading frame expressed in normal samples that could function as a native epitope. To further investigate this, we analyzed a dataset of 123 small cell lung cancer (SCLC) patients (George et al. 2015; Peifer et al. 2012; Iwakawa et al. 2015; Rudin et al. 2012). SCLC is the most aggressive type of lung cancer, with a very early relapse after chemotherapy treatment and an average survival of 5% after 5 years of diagnosis (Cooper and Spiro 2006). SCLC is one of the cancer types with the largest TMB (Fig. 4a), which has been shown to associated with the efficacy of immune therapy in SCLC (Hellmann et al. 2018). Additionally, SCLC presents a significantly higher density of mutations in introns compared to exons (Fig. 4a), which may be indicative of a widespread impact on RNA-processing. Accordingly, SCLC represents an interesting tumor type to investigate how splicing alterations may contribute to neoepitope burden and impact the antigenicity of tumor cells.

We applied ISOTOPE to RNA-seq from 123 small cell lung cancer (SCLC) patients and derived an exhaustive compendium of splicing alterations (Supp. Fig. 1) (Supp. Table 5). We filtered out all SCLC junctions and events that appeared in normal samples comparing with a large comprehensive dataset including normal lung tissues (Pertea et al. 2018; Nellore et al. 2016; GTEx Consortium 2015; Rudin et al. 2012) (Methods). We found a total 14643 aberrant splice sites, 7039 intron retentions, 1311 aberrant exon skipping events, and 290 *de novo* exonizations that are SCLC specific, affecting 2955, 149, 620, and 169 genes, respectively (Fig. 4b) (Supp. Table 5). Furthermore, these alterations distributed homogeneously across all samples and showed no association to tumor mutation burden, mutations on spliceosome factors or overexpression of MYC genes (Supp. Fig. 2). However, the level of detection of aberrant splice sites, new skipping events and exonizations showed some dependency with the number of spliced reads obtained in a sample (Supp. Fig. 2). We focused on the SCLC-specific events that occurred within an ORF and could therefore alter the protein product: 3890 (27%) of the aberrant splice sites, 804 (61%) of the new skipping events, 753 (10%) of the intron retentions and 85 (29%) of the new exonizations. Using a reference proteome for the SCLC samples, 26% of aberrant splice sites, 61% of the new skipping events, 10% of intron retentions, and 29% of new exonizations were predicted to produce a protein change (Fig. 4b).

To evaluate the immunogenic impacts induced by these splicing alterations, we predicted HLA-I and HLA-II types from the RNA-seq for the SCLC samples using PHLAT. We next used the altered and reference ORFs and searched for candidate MHC-I binders (binding affinity $\leq$ 500nM) that were specific to SCLC. We identified a total of 47,088 candidate splicing-derived neoepitopes, with the majority (60%) associated to intron retention events (Supp. Table 5). On the other hand, we identified a total of 254,125 candidate native epitopes depleted from splicing alterations. Thus, as observed before for cell lines, splicing alterations in SCLC patients have a greater number of depleted native epitopes than neoepitopes as a consequence of cancer-specific splicing (Figs. 4c-d). This imbalance towards the elimination of epitopes occurred at the level of the number of predicted immunogenic

peptides as well as the number of events producing immunogenic peptides. Moreover, this effect was not specific of any type of splicing alteration or HLA-type (Figs. 4c-d) (Supp. Fig. 3).

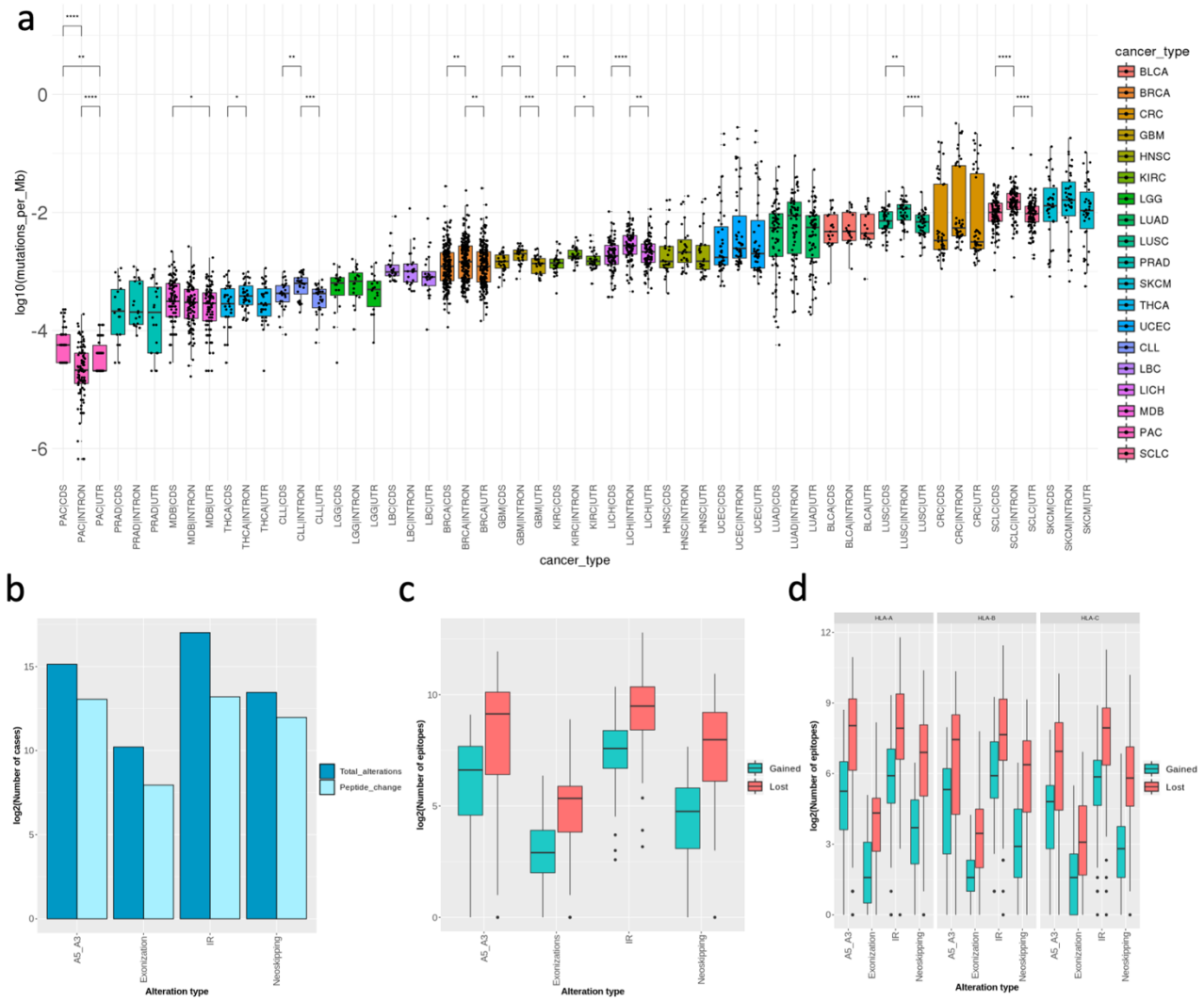


**Figure 4. (a)** Mutation burden calculated separately for introns, coding exons and non-coding exons from whole genome sequencing (WGS) data for a number of tumor types, including small cell lung cancer (SCLC), and separated by region: coding exons (CDS), introns (INTRON) and non-coding exonic regions (UTR). We indicate those comparisons between distributions that were significantly different using a Wilcoxon test (* p-val <0.05, ** p-val <0.01, *** p-val<0.001, **** p-val<0.0001). **(b)** Distribution of splicing alterations according to each type. All and those that impact protein. **(c)** Distribution of gained and lost epitopes according to splicing alteration type. **(d)** Candidate splicing-derived neoepitopes in SCLC patients, separated by HLA-type.

We could not detect any significant association of response biomarkers specific of SCLC but we did observe multiple significant associations in other tissues (Supp Table 7). These results are specially relevant in SCLC, for which no alteration has been yet described as therapeutically targeteable. We did not have an MHC-I associated mass-spectrometry data available for these SCLC samples. However, using mass-spectrometry data for MHC-I associated proteins in lymphoblasts (Lanoix et al. 2018) we were able to validate 1458 (11.7%) of the candidate depleted native epitopes. To test the

significance of the association of the predicted epitopes to the mass spectrometry data, a randomized comparison was performed. We took 1000 random peptides from those with good affinity and 1000 peptides from all the set and checked how many from these 2 random sets would be validated by mass-spectrometry. We repeated this process 100 times and tested the difference of the two distributions. This analysis yielded that the number of depleted candidate native epitopes validated were significantly higher from the complete set (Kolmogorov-Smirnov p-value = 3.44e-13).

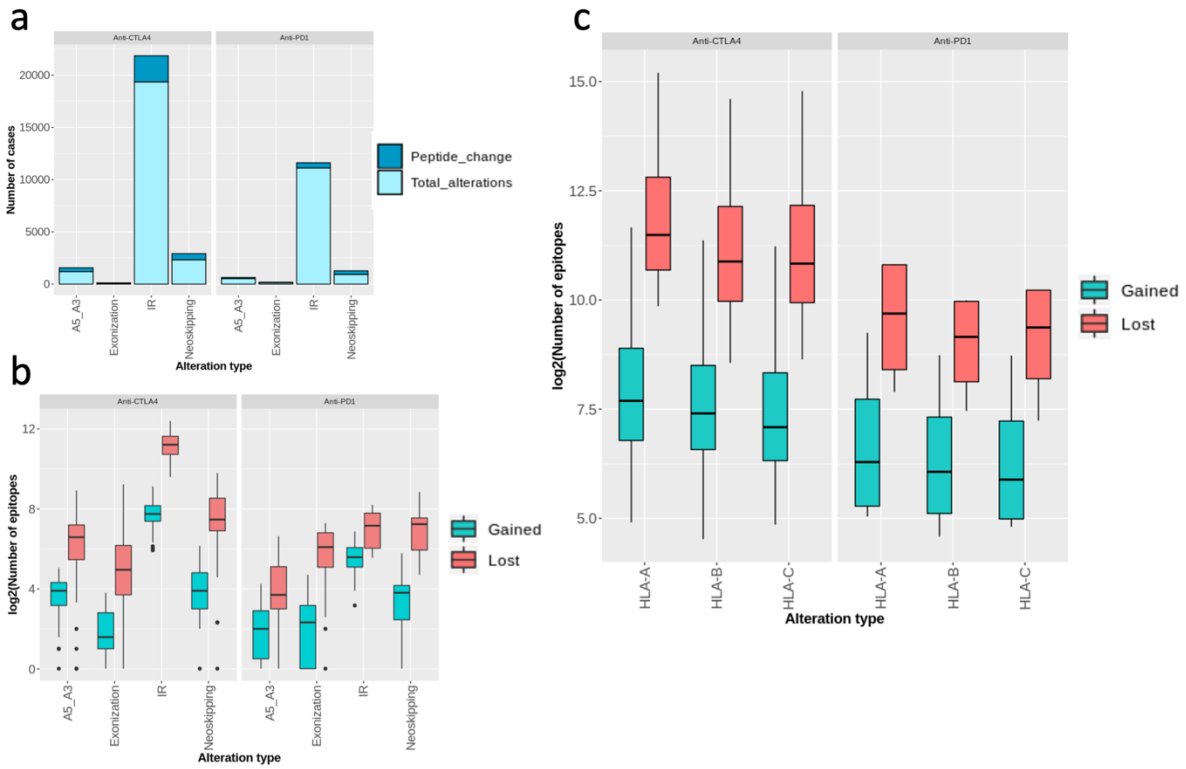


**Figure 5. (a)** Number of events and protein-affecting events per patient for the melanoma cohorts treated with anti-CTLA4 and with anti-PD1. **(b)** Number of candidate cancer-specific splicing-derived neoepitopes per patient separated by event type. **(c)** Number of candidate cancer-specific splicing-derived neoepitopes per patient separated by HLA-type.

## Association of cancer-specific splicing-derived epitopes with response to immune checkpoint inhibitors

To test whether splicing-derived neoepitopes may be associated to the patient response to immune therapy, we applied ISOTOPE to RNA-seq data from melanoma samples prior to treatment with anti-*CTLA4* (Van Allen et al. 2015) or anti-*PD1* (Hugo et al. 2016) (Supp. Fig. 4) (Supp. Tables 6 and 7). As before, we first calculated all the cancer-specific splicing alterations occurring in the patient samples by removing all events that also occurred in any of the samples from GTEX or in normal samples from SRA (Methods) (Fig. 5). We compared the predicted neoepitopes in both sets with the annotated clinical response of the patient to the immunotherapy: responder or non-responder (Van Allen et al. 2015; Hugo et al. 2016). The number of candidate splicing-derived neoepitopes in responders to anti-*CTLA4* (mean 346) and non-responders (mean 375) were not significantly

different. Similarly, there was no significant difference between the total number of splicing-derived neoepitopes between responders (mean 46) and non-responders (mean 62.6) in the anti-*PD1* cohort.

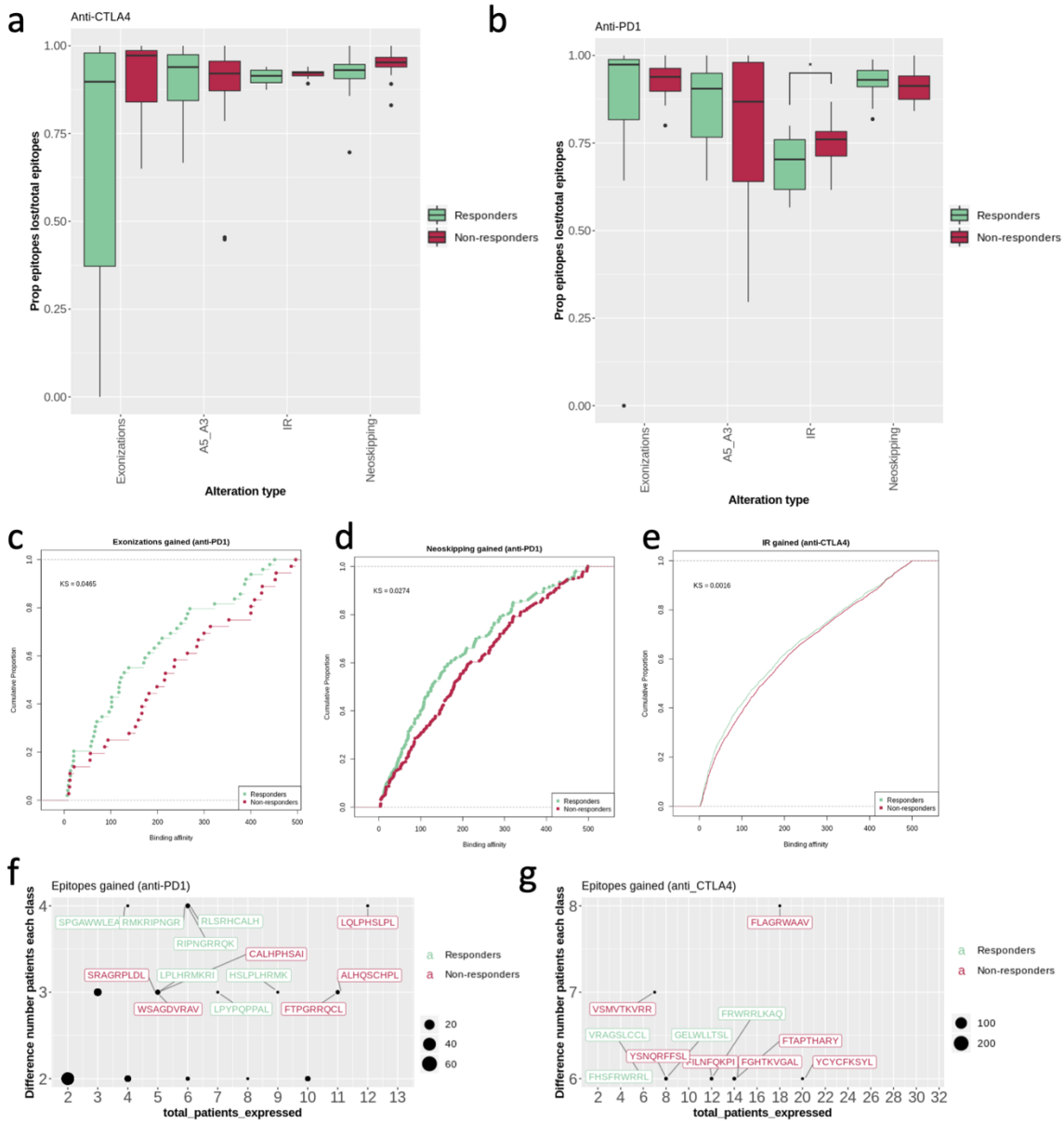


**Figure 6. (a)** Proportion of lost epitopes over the total of gained and lost epitopes through cancer-specific splicing alterations (y axis) for patients treated with anti-*CTLA4,* separated by type of splicing alteration (x axis) and by patient response: responder (green) or non-responder (red). **(b)** As in (a) but for melanoma patients treated with anti-*PD1*. **(c)** Cumulative plots of the binding affinities (x axis) of epitopes gained in melanoma tumors from exonization events separated in responders (green) and non-responders (red) to anti-PD1 therapy. We also give the Kolmogorov-Smirnov test p-value (KS) = 0.0465 **(d)** As in (c), for splicing-derived neoepitopes from new skipping events (neoskipping) events, KS = 0.0274. **(e)** Cumulative plots of the affinities of epitopes gained in melanoma tumors from intron retention events separated in responders (green) and non-responders (red) to anti-CTLA4 therapy, KS = 0.0016. **(f)** Frequency of splicing-derived neoepitopes according to the total number of patients in which are predicted (x axis, *total_patients_expressed*), and according to the absolute difference between responders and non-responders to anti-*PD1* therapy (y axis, *Difference number patients each class*). Epitopes are indicated in green if they are more frequent in responders, and in red otherwise. The size of the point indicates the number of cases. **(g)** The same as in (f) but for responders and non-responders to anti-*CTLA4* therapy.

Separating the splicing alterations by type, we found in general a higher proportion of lost native epitopes in all patients (Figs. 6a and 6b). Furthermore, we did not observe any differences in the proportion of epitopes between responders and non-responders to anti-*CTLA4* (Fig. 6a). On the other hand, responders to anti-*PD1* therapy had more lost native epitopes from intron retention events compared to non-responders, whereas other splicing alterations did not show any significant differences (Fig. 6b). To further test the potential role of gained and lost epitopes in the response to immunotherapy, we studied their binding affinities. We found stronger MHC-I interaction (lower values of binding affinity) for splicing-derived neoepitopes in responders. In particular, gained epitopes from *de novo* exonizations and new skipping events had stronger interactions in anti-PD1 responders (Figs. 6c and 6d), and those gained from intron-retention events had stronger interactions in anti-CTLA4 responders (Fig. 6e). Incidentally, lost native epitopes from exonizations in anti-*PD1* responders, and from new skipping events and intron retentions in anti-*CTLA4* responders also showed stronger MHC-I binding (Supp. Fig. 5).

To further assess the potential relevance of the candidate splicing-derived epitopes, we studied whether the identified peptides occurred in multiple patients and in association with response to the immunotherapy. The most frequent splicing-derived neoepitope produced in responders for anti-*PD1* therapy was produced from an intron retention event in the proto-oncogene *PIM3* (SPGAWWLEA) and occurred in 4 out of 14 patients (29%), with HLA types HLA-B0702 (2 of them), HLA-B5501 and HLA-B5601 (Fig. 6f). The most frequent splicing-derived neoepitopes in responders to anti-*CTLA4* therapy occurred in 6 out of 18 cases (33%) and were produced from intron retention events in the genes *SPTAN1* (FHSFRWRRL) and *GNAS* (VRAGSLCCL). All patients for both epitopes were HLA-C0701 (Fig. 6g).

## Discussion

We have described a comprehensive analysis of how cancer-specific splicing alterations may impact the antigenicity of tumors. Our approach, implemented in the pipeline ISOTOPE, presents a number of novelties and advantages with respect to previous approaches. It is exhaustive in the type of alterations tested, e.g. it includes *de novo* exonizations, which have not been previously characterised. Thus, making possible an assessment at unprecedented scale of the splicing-derived neoepitope burden. Although tumor-associated intron retention is quite common (Dvinge and Bradley 2015), we observed that neo-skipping events showed in greater proportion a disruption of the encoded proteins and led to more potential candidate neoepitopes. Additionally, as ISOTOPE only requires RNA-seq data from a tumor sample, it is applicable in the absence of DNA sequencing data from the tumor, or from normal controls. Unlike previous studies (Jayasinghe et al. 2018; Kahles et al. 2018; Smart et al. 2018), our analysis describes cancer-specific alterations by comparing to a large compendium of normal samples and performs a bootstrapping test to ensure that the cancer-specific splicing alteration considered is supported by significantly more reads than any other splicing alterations in the same locus. In our analyses we also tested potential MHC-II neo-epitopes. Although these predictions are generally less reliable, MHC-II associated neo-epitopes may also be relevant for immunotherapy (Sun

et al. 2017; Lu et al. 2017). In summary, ISOTOPE novel approach enables a robust and exhaustive survey of the immunogenic impacts of splicing in cancer.

In agreement with previous studies, only a small proportion of predicted neoepitopes were validated with MHC-I associated mass-spectrometry. There are several possible reasons for that. The RNA-seq data used might have not been of sufficient depth to be able to robustly identify all relevant splicing alterations. This is suggested by the overall low recurrence of the cancer-specific splicing alterations found across patients. Although many of these might be accidental transcripts produced in a tumor, they still can change the identity of the tumor cell and shape their fitness. An additionally reason may related to the interpretation of the proteomics data. MHC-I associated mass spectrometry does not use the enzymatic digestion standard in unbiased proteomics. Thus, to ensure that matches were reliably detected, we built a control dataset containing a large reference set of peptides, which could lead to a low detection rate. Additionally, we relied on candidate epitopes predicted from RNA-seq. However, a more sensitive approach might be based on the identification of splicing-derived neoepitopes directly from the MHC-I associated mass spectrometry.

Recently, it was shown that for some tumor types, the occurrence of splicing alterations associates with higher expression of PD1 and PD-L1 (Jayasinghe et al. 2018). It was then suggested that these tumors could benefit from immune checkpoint inhibitor therapy due to the presence of a higher content of splicing-derived neoepitopes. However, in a recent study of neoepitopes derived from intron retention (Smart et al. 2018), no association was found between neoepitope count and the response to checkpoint inhibitors. Here, we extended this comparison to all other types of splicing alterations. Using two cohorts of patients treated with immune checkpoint inhibitors, and testing all splicing alterations, the number of splicing-derived neoepitopes does not differ significantly between responders and non-responders. However, we observed differences in the predicted affinity to the MHC-I complex. Indeed, the overall interaction strength predicted for neoepitopes in responders is larger, possibly indicating a better recognition of tumor cells in the immune response triggered by the treatment. This raises the possibility that splicing-derived neoepitopes may contribute to the positive response to the therapy. We showed before that splicing alterations and their functional impacts anti-correlate with the frequency of mutations and the presence of mutational drivers (Climente-González et al. 2017). Thus, it is possible that splicing-derived neoepitopes play a role in tumors with fewer mutations. Given the high mutation rate and heterogeneity of the melanoma tumors, we would expect that splicing-derived neoepitopes will only have a small contribution to the patient response. The mechanisms here described might therefore be more prominent in tumors with low mutation rate, for which other types of somatic alterations play a relevant role, like fusion-derived neoepitopes in pediatric tumors (Chang et al. 2017).

Additionally, we have described how splicing alterations could potentially deplete the expression of peptides that would otherwise be recognized as self-epitopes. Although T-cell selection in the thymus can remove some of these self-reactive specificities, it is known that this could be incomplete or suboptimal (Sewell 2012). Self-peptides might not bind equally well to MHC molecules, which would then compromise the efficiency of negative selection of self-reactive T-cells. As a consequence, potentially self-reactive T-cells can be found in circulation in healthy individuals (Su et al. 2013), and tolerance to some self-antigens could often rely on the additional control of regulatory T-cells

expressing *CTLA4* (Legoux et al. 2015). This suggests an intriguing possibility. Upon treatment with immune-checkpoint inhibitors, among the immunocompetent T-cells that are freed to act against the tumor cells, there may be some with self-reactive capabilities. These may help destroying tumor cells could also be a potential trigger of autoimmune responses. Indeed, treatment with immune-checkpoint inhibitors has led to serious and sometimes fatal autoimmune reactions in patients (Cappelli et al. 2016; Amos et al. 2011; Ferrari et al. 2019). T-cells may attack the tumor cells via neoepitopes as well as self-epitopes but could trigger immune responses through the reactivity against self-epitopes in normal cells. On the other hand, a depletion of self-epitopes may lead to a reduced response. We have observed that some melanoma patients with self-epitope depletion show no response to the treatment. Also, when we characterized the splicing alterations in small-cell lung cancer, a tumor type with low survival and with limited response to immune therapy, self-epitope depletion occurs much more frequently than neoepitope production, and we could validate a large number of them from MHC-associated mass spectrometry in lymphocytes. Cancer-specific splicing alterations could generate neoepitopes, but as shown here, could potentially also deplete self-epitopes. These alterations may not necessarily prevent the self-reactivity in normal cells but could reduce the recognition and destruction of tumor cells, thereby hindering the effect of the immune therapy. This suggests the interesting hypothesis that cancer-specific splicing alterations may contribute to the escape of tumor cells to immune-checkpoint inhibitor treatment.

As the ability of the immune system to identify malignant cells relies on the tumor cells maintaining sufficient antigenicity, it is thus essential to exhaustively explore all potential immunogenic impacts through the variety of splicing alterations that may arise specifically in tumors. Our method ISOTOPE facilitates this exploration, thereby helping in the identification of molecular markers of response to immunotherapy.

## Methods

### Datasets

RNA sequencing (RNA-seq) data for the cell lines analysed was collected from (Barretina et al. 2012) (GEO accession number GSE36139). We also collected RNA-seq data from 38 melanoma patients pre anti-CTLA4 treatment classified as responder (18 cases) and non-responder (20 cases) (Van Allen et al. 2015), available from dbGAP (https://www.ncbi.nlm.nih.gov/gap) under accession phs000452.v2.p1: and RNA-seq data from 27 melanoma patients pre anti-PD1 treatment (Hugo et al. 2016), available at SRA (https://www.ncbi.nlm.nih.gov/sra) under accession SRP070710, also labeled as responder (14 cases) or non-responder (13 cases). Additionally, we gathered RNA-seq data from 123 SCLC patients (George et al. 2015) (EGA accession EGAS00001000925) (Rudin et al. 2012) (EGA accession EGAD00001000223), and (Iwakawa et al. 2015). For 26 of the SCLC patients (Rudin et al. 2012) we also had normal controls. Samples with more than 30% of junctions present in other samples but had missing value in them, were filtered out. We estimated the tumor purity of every tumor sample from the gene expression data using ESTIMATE R package (v.1.0.13) (Yoshihara et al. 2013).

**Identification of cancer-specific splicing alterations**

All RNA-seq samples were mapped to the genome (hg19) using STAR (Dobin et al. 2013) and were processed as described before (Singh et al. 2018): Mapped spliced reads with at least a common splice site across two o more samples were clustered with LeafCutter (Li et al. 2018), with a minimum of 30 reads per cluster and a minimum fraction of reads of 0.01 in a cluster supporting a junction. Read-counts per junction were normalized over the total of reads in a cluster. Junction clusters were defined across all patients but normalized read counts were calculated per patient. Junctions were classified as novel if either or both of the splice-sites were not present in the annotation (Gencode v19) (Harrow et al. 2012), they had at least 10 supporting reads in at least one tumor sample, and did not appear in any of the 7859 normal samples from 18 tissues from the GTEX consortium (GTEx Consortium 2015), or in normal samples from Intropolis (Nellore et al. 2016). Novel junctions in clusters were classified as aberrant splice-site, new exon skipping (neo-skipping), or *de novo* exonization. To define exonizations, we considered all pairs of junctions not present in normal samples (GTEx Consortium 2015; Nellore et al. 2016; Rudin et al. 2012) that would define a potential new internal exon not longer than 500nt, with flanking canonical splice site motifs (AG-GT). We kept only cases with more than 5 reads validating each splice site. For tumor specific neo-skippings we considered those new junctions that skipped known exons and defined new connections between adjacent exons. To define retained introns we used KMA (Pimentel et al. 2015) to extend the Gencode (v19) transcriptome with potential retained introns (RIs), which we quantified for each RNA-seq sample with Kallisto (Bray et al. 2016). To filter out RIs that were not tumor specific, we calculated RI events with SUPPA (Trincado et al. 2018) from the Gencode (Harrow et al. 2012) and the CHESS 2.0 (Pertea et al. 2018) annotations, and removed KMA-predicted RIs that appeared in these annotations. To control for confounding signals due to overall lack of transcript processing, for each splicing alteration we compared the expression of the alterations with 100 randomly selected cases from the same gene using an Empirical Cumulative Distribution Function (ECDF) test. Junctions were compared with other junctions, exonizations were compared with genic regions of similar length, and retained introns were compared with other introns. Cell line data was processed in a similar way, but without removing the alterations in normal samples, as those tests were focused on the presentation of splicing-derived neo-epitopes.

**Somatic mutation data**

We used somatic mutations from whole genome sequencing for 505 tumor samples from 14 tumor types (Fredriksson et al. 2014): bladder carcinoma (BLCA) (21 samples), breast carcinoma (BRCA) (96 samples), colorectal carcinoma (CRC) (42 samples), glioblastoma multiforme (GBM) (27 samples), head and neck squamous carcinoma (HNSC) (27 samples), kidney chromophobe (KICH) (15 samples), kidney renal carcinoma (KIRC) (29 samples), low grade glioma (LGG) (18 samples), lung adenocarcinoma (LUAD) (46 samples), lung squamous cell carcinoma (LUSC) (45 samples), prostate adenocarcinoma (PRAD) (20 samples), skin carcinoma (SKCM) (38 samples), thyroid carcinoma (THCA) (34 samples), and uterine corpus endometrial carcinoma (UCEC) (47 samples). Additionally, we used whole-genome somatic mutation calls for SCLC from (George et al. 2015) (EGA accession EGAS00001000925). We only used substitutions and discarded those that overlapped with frequent (>1% allele frequency) SNPs (dbSNP 144).

**Protein impact of the splicing alterations**

For each analyzed cohort, we built a reference transcriptome using the largest mean expression per gene across samples, using only those cases with mean > 1 transcript per million (TPM). Transcript abundance was calculated using Salmon (Patro et al. 2017). A reference proteome was defined from these reference transcripts. For each splicing alteration, a modified transcript was then built using as scaffold the reference transcript exon-intron structure. Unless the splicing alteration only affected the untranslated region (UTR), an altered protein was calculated from the longest open reading frame (ORF) (start to stop) predicted on the modified transcript. Each splicing alteration was considered only if an ORF was predicted. If the splicing alteration deleted the region of the start codon, the closest downstream start codon was used. Further, if the stop codon in the altered ORF was located further than 50nt to a downstream splice site, the case was discarded as potential NMD target. Software to run this analysis and selection of novel splicing junctions is available at http://github.com/comprna/ISOTOPE

**Prediction of splicing-derived neoepitopes**

Unless available, we inferred the HLA-type from the tumor RNA-seq using PHLAT (Bai et al. 2014). From all proteins derived from the splicing alterations and from the reference proteome, we predicted potential MHC-I binders with NetMHC-4.0.0 (Andreatta and Nielsen 2016), and with NetMHCpan-4.0 (Jurtz et al. 2017) for the classes missing in NetMHC-4.0.0. Those peptides in common between the reference and the altered protein were discarded. Peptides in the altered protein with binding affinity ≤ 500nM, but not present in the reference proteome were considered candidate neo-epitopes. We performed the same analysis for MHC-II binders using predictions from NetMHCII-2.3, and complementing them with the predictions from NetMHCIIpan-3.2 for the missing types (Jensen et al. 2018).

**Validation of neoepitope prediction with MHC-I mass-spectrometry**

MHC-I associated mass-spectrometry data was analyzed following the approach from (Murphy et al. 2017). We tested all the candidate neoepitopes using as a control database all candidate MHC-I binders from Uniprot. Using as a control all candidate MHC-I binders from the reference proteome yielded similar results. Using candidate binders rather than the entire Uniprot reduces the search space and takes into account that MHC-I proteomics involves unspecific digestion. We matched the mass spectra to the joined set of control and candidate cancer-specific peptides, and with the control set alone. Candidate matches for both sets were compared to calculate their significance. For the analysis of the MHC-I associated data for the cell lines CA46, HL-60 and THP-I, we used the same procedures as described before for these datasets (Smart et al. 2018). To test the significance of the identification of the predicted epitopes in the mass-spectrometry data from (Lanoix et al. 2018), a randomized comparison was performed. We took two random sets of 1000 random predicted epitopes each, one set with cases of good affinity (≤ 500nM) and one set from all the set of predicted neoepitopes (with or without good affinity). We then checked how many of these 2 random sets are validated with mass-spectrometry data from (Lanoix et al. 2018). We repeated this process 100 times and compared with

a Kolomogorov-Simrnoff test whether the 2 distributions of the number of peptides validated were significantly different. For the splicing-lost epitopes in SCLC we found a p-value of 3.44e-13, whereas for the spicing-gained epitopes there was no significant difference.

**Biomarkers enrichment analysis**

The Clinical Interpretation of Variants in Cancer (CIViC) (Griffith et al. 2017) and the Cancer Genome Interpreter (CGI) (Tamborero et al. 2018) databases were used to identify biomarkers, meaning genetic alterations, associated to the treatment response to cancer. Genes with characterized splicing alterations in each cohort were assessed for enrichment in biomarkers by means of a Fisher Test and multiple test correction (FDR estimation).

## Author Contributions

E.E. and J.L.T designed the study. J.L.T. developed the software and performed the analyses. M.R-S. contributed with the analyses of the tumor mutation burden. J.P-G. performed the biomarker enrichment analysis. T.F. contributed with part of the proteomics analyses. E.E. and F.S. supervised the analyses. E.E. and J.L.T. wrote the manuscript with essential inputs from J.P-G. and F.S.

## Acknowledgements

We thank the authors from the different studies used in this article for facilitating access to their RNA-seq datasets: Van Allen et al (phs000452.v2.p1), Hugo et al. (SRP070710), George et al. EGAS00001000925), Rudin et al. (EGAD00001000223). This work was supported by the Spanish Government and FEDER with grant BIO2017-85364-R, by AGAUR with grants SGR2017-1020, and by the Instituto de Salud Carlos III (ISCIII) - Fondo Social Europeo (FI18/00034). The Research Programme on Biomedical Informatics (GRIB) is a member of the Spanish National Bioinformatics Institute (INB), funded by ISCIII and FEDER (PT17/0009/0014). The DCEXS is a 'Unidad de Excelencia María de Maeztu', funded by the AEI (CEX2018-000782-M). The GRIB is also supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), Generalitat de Catalunya (2017 SGR 00519).

## Author Contributions

The authors declare no competing interests

## References

Amos SM, Duong CPM, Westwood JA, Ritchie DS, Junghans RP, Darcy PK, Kershaw MH. 2011. Autoimmunity associated with immunotherapy of cancer. *Blood* **118**: 499–509. https://ashpublications.org/blood/article/118/3/499/28926/Autoimmunity-associated-with-immunotherapy-of.

Andersen RS, Andersen SR, Hjortsø MD, Lyngaa R, Idorn M, Køllgård TM, Met O, Thor Straten P, Hadrup SR. 2013. High frequency of T cells specific for cryptic epitopes in melanoma

patients. *Oncoimmunology* **2**: e25374.
Andreatta M, Nielsen M. 2016. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**: 511–7.
Antonia SJ, López-Martin JA, Bendell J, Ott PA, Taylor M, Eder JP, Jäger D, Pietanza MC, Le DT, de Braud F, et al. 2016. Nivolumab alone and nivolumab plus ipilimumab in recurrent small-cell lung cancer (CheckMate 032): a multicentre, open-label, phase 1/2 trial. *Lancet Oncol* **17**: 883–895.
Bai Y, Ni M, Cooper B, Wei Y, Fury W. 2014. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genomics* **15**: 325.
Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov G V, Sonkin D, et al. 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**: 603–7.
Boegel S, Bukur T, Castle JC, Sahin U. 2018. In Silico Typing of Classical and Non-classical HLA Alleles from Standard RNA-Seq Reads. *Methods Mol Biol* **1802**: 177–191. http://www.ncbi.nlm.nih.gov/pubmed/29858809.
Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525–7.
Cappelli LC, Shah AA, Bingham CO. 2016. Cancer immunotherapy-induced rheumatic diseases emerge as new clinical entities. *RMD Open* **2**: e000321. http://rmdopen.bmj.com/lookup/doi/10.1136/rmdopen-2016-000321.
Chang T-C, Carter RA, Li Y, Li Y, Wang H, Edmonson MN, Chen X, Arnold P, Geiger TL, Wu G, et al. 2017. The neoepitope landscape in pediatric cancers. *Genome Med* **9**: 78. http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-017-0468-3.
Climente-González H, Porta-Pardo E, Godzik A, Eyras E. 2017. The Functional Impact of Alternative Splicing in Cancer. *Cell Rep* **20**: 2215–2226.
Cooper S, Spiro SG. 2006. Small cell lung cancer: treatment review. *Respirology* **11**: 241–8.
Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21.
Dvinge H, Bradley RK. 2015. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med* **7**: 45.
Ferrari SM, Fallahi P, Elia G, Ragusa F, Ruffilli I, Patrizio A, Galdiero MR, Baldini E, Ulisse S, Marone G, et al. 2019. Autoimmune Endocrine Dysfunctions Associated with Cancer Immunotherapies. *Int J Mol Sci* **20**: 2560. https://www.mdpi.com/1422-0067/20/10/2560.
Fredriksson NJ, Ny L, Nilsson JA, Larsson E. 2014. Systematic analysis of noncoding somatic mutations and gene expression alterations across 14 tumor types. *Nat Genet* **46**: 1–7.
George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, Leenders F, Lu X, Fernández-Cuesta L, Bosco G, et al. 2015. Comprehensive genomic profiles of small cell lung cancer. *Nature* **524**: 47–53.
Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, Stephens PJ, Daniels GA, Kurzrock R. 2017. Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther* **16**: 2598–2608. http://mct.aacrjournals.org/lookup/doi/10.1158/1535-7163.MCT-17-0386.
Griffith,M., Spies,N.C., Krysiak,K., et al. (2017) CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nature Genetics, **49**, 170–174.
Gumy-Pause, F., Wacker, P. & Sappino, A. *ATM* gene and lymphoid malignancies. *Leukemia* **18,** 238–242 (2004).
GTEx Consortium. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**: 648–60.
Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The

ENCODE Project. *Genome Res* **22**: 1760–74.
Hellmann MD, Callahan MK, Awad MM, Calvo E, Ascierto PA, Atmaca A, Rizvi NA, Hirsch FR, Selvaggi G, Szustakowski JD, et al. 2018. Tumor Mutational Burden and Efficacy of Nivolumab Monotherapy and in Combination with Ipilimumab in Small-Cell Lung Cancer. *Cancer Cell* **33**: 853-861.e4.
Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, 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. http://www.ncbi.nlm.nih.gov/pubmed/26997480.
Iwakawa R, Kohno T, Totoki Y, Shibata T, Tsuchihara K, Mimaki S, Tsuta K, Narita Y, Nishikawa R, Noguchi M, et al. 2015. Expression and clinical significance of genes frequently mutated in small cell lung cancers defined by whole exome/RNA sequencing. *Carcinogenesis* **36**: 616–21.
Jayasinghe RGRG, Cao S, Gao Q, Wendl MC, Vo NSNS, Reynolds SMSM, Zhao Y, Climente-González H, Chai S, Wang F, et al. 2018. Systematic Analysis of Splice-Site-Creating Mutations in Cancer. *Cell Rep* **23**: 270-281.e3.
Jensen KK, Andreatta M, Marcatili P, Buus S, Greenbaum JA, Yan Z, Sette A, Peters B, Nielsen M. 2018. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* **154**: 394–406.
Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. 2017. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* **199**: 3360–3368.
Kahles A, Lehmann K-V, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, Stegle O, Kohlbacher O, Sander C, Cancer Genome Atlas Research Network, et al. 2018. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer Cell* **34**: 211-224.e6.
Lanoix J, Durette C, Courcelles M, Cossette É, Comtois-Marotte S, Hardy M-P, Côté C, Perreault C, Thibault P. 2018. Comparison of the MHC I Immunopeptidome Repertoire of B-Cell Lymphoblasts Using Two Isolation Methods. *Proteomics* **18**: e1700251.
Laumont CM, Vincent K, Hesnard L, Audemard É, Bonneil É, Laverdure J-P, Gendron P, Courcelles M, Hardy M-P, Côté C, et al. 2018. Noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl Med* **10**: eaau5516. https://stm.sciencemag.org/lookup/doi/10.1126/scitranslmed.aau5516.
Legoux FP, Lim J-B, Cauley AW, Dikiy S, Ertelt J, Mariani TJ, Sparwasser T, Way SS, Moon JJ. 2015. CD4 + T Cell Tolerance to Tissue-Restricted Self Antigens Is Mediated by Antigen-Specific Regulatory T Cells Rather Than Deletion. *Immunity* **43**: 896–908. https://linkinghub.elsevier.com/retrieve/pii/S1074761315004367.
Li, B., Chng, W. EZH2 abnormalities in lymphoid malignancies: underlying mechanisms and therapeutic implications.*J Hematol Oncol* **12,**118 (2019).
Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, Pritchard JK. 2018. Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**: 151–158.
Lu Y-C, Parker LL, Lu T, Zheng Z, Toomey MA, White DE, Yao X, Li YF, Robbins PF, Feldman SA, et al. 2017. Treatment of Patients With Metastatic Cancer Using a Major Histocompatibility Complex Class II-Restricted T-Cell Receptor Targeting the Cancer Germline Antigen MAGE-A3. *J Clin Oncol* **35**: 3322–3329.
Murphy JP, Konda P, Kowalewski DJ, Schuster H, Clements D, Kim Y, Cohen AM, Sharif T, Nielsen M, Stevanovic S, et al. 2017. MHC-I Ligand Discovery Using Targeted Database Searches of Mass Spectrometry Data: Implications for T-Cell Immunotherapies. *J Proteome Res* **16**: 1806–1816. https://pubs.acs.org/doi/10.1021/acs.jproteome.6b00971.
Nellore A, Jaffe AE, Fortin J-P, Alquicira-Hernández J, Collado-Torres L, Wang S, Phillips Iii RA, Karbhari N, Hansen KD, Langmead B, et al. 2016. Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the Sequence Read Archive.

*Genome Biol* **17**: 266.
Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*.
Peifer M, Fernández-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, et al. 2012. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet* **44**: 1104–1110.
Pertea M, Shumate A, Pertea G, Varabyou A, Chang Y-C, Madugundu AK, Pandey A, Salzberg S. 2018. Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. *bioRxiv*.
Pimentel H, Conboy JG, Pachter L. 2015. Keep Me Around: Intron Retention Detection and Analysis.
Ribas A, Wolchok JD. 2018. Cancer immunotherapy using checkpoint blockade. *Science (80- )* **359**: 1350–1355. https://www.sciencemag.org/lookup/doi/10.1126/science.aar4060.
Ritz D, Gloger A, Weide B, Garbe C, Neri D, Fugmann T. 2016. High-sensitivity HLA class I peptidome analysis enables a precise definition of peptide motifs and the identification of peptides from cell lines and patients' sera. *Proteomics* **16**: 1570–80.
Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, et al. 2015. Mutational landscape determines sensitivity to PD-1 blockade in non–small cell lung cancer. *Science (80- )* **348**: 124–128. https://www.sciencemag.org/lookup/doi/10.1126/science.aaa1348.
Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar A V, Necchi A, Dawson N, O'Donnell PH, Balmanoukian A, Loriot Y, et al. 2016. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* **387**: 1909–1920. https://linkinghub.elsevier.com/retrieve/pii/S0140673616005614.
Rozanov D V, Rozanov ND, Chiotti KE, Reddy A, Wilmarth PA, David LL, Cha SW, Woo S, Pevzner P, Bafna V, et al. 2018. MHC class I loaded ligands from breast cancer cell lines: A potential HLA-I-typed antigen collection. *J Proteomics* **176**: 13–23. http://www.ncbi.nlm.nih.gov/pubmed/29331515.
Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, Bergbower EA, Guan Y, Shin J, Guillory J, et al. 2012. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. *Nat Genet* **44**: 1111–6.
Sebestyén E, Zawisza M, Eyras E. 2015. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res* **43**.
Sewell AK. 2012. Why must T cells be cross-reactive? *Nat Rev Immunol* **12**: 669–677. http://www.nature.com/articles/nri3279.
Singh B, Trincado JL, Tatlow PJ, Piccolo SR, Eyras E. 2018. Genome sequencing and RNA-motif analysis reveal novel damaging noncoding mutations in human tumors. *Mol Cancer Res* **16**.
Smart AC, Margolis CA, Pimentel H, He MX, Miao D, Adeegbe D, Fugmann T, Wong K-K, Van Allen EM. 2018. Intron retention is a source of neoepitopes in cancer. *Nat Biotechnol* **36**: 1056–1058. http://www.ncbi.nlm.nih.gov/pubmed/30114007.
Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS, et al. 2014. Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N Engl J Med* **371**: 2189–2199. http://www.nejm.org/doi/10.1056/NEJMoa1406498.
Su LF, Kidd BA, Han A, Kotzin JJ, Davis MM. 2013. Virus-Specific CD4+ Memory-Phenotype T Cells Are Abundant in Unexposed Adults. *Immunity* **38**: 373–383. https://linkinghub.elsevier.com/retrieve/pii/S1074761313000526.
Sun Z, Chen F, Meng F, Wei J, Liu B. 2017. MHC class II restricted neoantigen: A promising target in tumor immunotherapy. *Cancer Lett* **392**: 17–25.

Tamborero,D., Rubio-Perez,C., Deu-Pons,J., et al. (2018) Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Medicine 2018 10:1, 10, 25.

Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyras E. 2018. SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol* **19**.

Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MHG, Goldinger SM, et al. 2015. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**: 207–211.

Wang T-Y, Wang L, Alam SK, Hoeppner LH, Yang R. 2019. ScanNeo: identifying indel-derived neoantigens using RNA-Seq data ed. I. Birol. *Bioinformatics* **35**: 4159–4161. https://academic.oup.com/bioinformatics/article/35/20/4159/5384488.

Wood MA, Nguyen A, Struck AJ, Ellrott K, Nellore A, Thompson RF. 2020. neoepiscope improves neoepitope prediction with multivariant phasing ed. I. Birol. *Bioinformatics* **36**: 713–720. https://academic.oup.com/bioinformatics/article/36/3/713/5551338.

Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine D a, et al. 2013. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* **4**: 2612.

Zhang J, Mardis ER, Maher CA. 2016. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. *Bioinformatics* btw674. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btw674.

Zhang M, Fritsche J, Roszik J, Williams LJ, Peng X, Chiu Y, Tsou C-C, Hoffgaard F, Goldfinger V, Schoor O, et al. 2018. RNA editing derived epitopes function as cancer antigens to elicit immune responses. *Nat Commun* **9**: 3919. http://www.nature.com/articles/s41467-018-06405-9.