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 and identifying tumor-specific epitopes can help elucidate the molecular determinants of therapy response. Here, we describe a pipeline, ISOTOPE (ISOform-guided prediction of epiTOPEs In Cancer), for the comprehensive identification of tumor-specific splicing-derived epitopes. Using RNA sequencing and mass spectrometry for MHC-I associated proteins, ISOTOPE identified neoepitopes from tumor-specific splicing events that are potentially presented by MHC-I complexes. Analysis of multiple samples indicates that splicing alterations may affect the production of self-epitopes and generate more candidate neoepitopes than somatic mutations. Although there was no difference in the number of splicing-derived neoepitopes between responders and non-responders to immune therapy, higher MHC-I binding affinity was associated with a positive response. Our analyses highlight the diversity of the immunogenic impacts of tumor-specific splicing alterations and the importance of studying splicing alterations to fully characterize tumors in the context of immunotherapies. ISOTOPE is available at https://github.com/comprna/ISOTOPE

Author summary

Immune cells have the ability to attack tumor cells upon the identification of tumor-specific peptides, i.e., epitopes, that are presented by the major histocompatibility complex (MHC). New cancer immunotherapies that help trigger this process provide a promising therapeutic strategy. One crucial aspect for their success is the ability to determine the molecular properties of a tumor that are

informative about the effectiveness of the therapy. Alterations in the way genes are processed to express RNA molecules could lead to the production of new peptides, with some of them potentially being presented as tumor epitopes and facilitate the attack of immune cells. It is therefore essential to facilitate the identification of these splicing-derived epitopes. In this work, we describe a computational pipeline that performs a comprehensive identification of splicing alterations in a tumor and the potential epitopes that they would produce. Analysis of tumor samples with our pipeline show that responders and non-responders to immune therapy do not show differences in the number of splicing-derived epitopes, but splicing neoepitopes have higher affinity to the MHC complex in responders. Our new pipeline facilitates the genome-scale analysis of the role of splicing alterations in shaping the molecular properties that influence response to immunotherapy.

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 estimate the antigenicity of a tumor is through the calculation of the frequency of somatic mutations as a proxy for the abundance of tumor neoantigens. This has led to the identification of an association between response to checkpoint inhibitors and tumor mutation burden (TMB) in tumors such as 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, tumor-specific transcriptome alterations have been shown to be a source of 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 tumor-specific splicing (Jayasinghe et al. 2018; Kahles et al. 2018; Smart et al. 2018; Andersen et al. 2013). 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) in tumors represents an additional potential source of cancer neoepitopes. However,

it is not clear yet whether these splicing-derived neoepitopes provide a mechanism to elicit cancer-specific immune responses and whether they may improve patient response to immune therapies.

To address these questions and expand the analysis of splicing-derived neoepitopes in cancer, we developed ISOTOPE (ISOform-guided prediction of epiTOPEs in cancer), a pipeline to exhaustively identify the immunogenic impacts from tumor-specific splicing alterations. ISOTOPE identifies splicing alterations that are specific to each tumor sample in comparison with a comprehensive set of normal samples and calculates the impact on the encoded proteins and the candidate neoepitopes. ISOTOPE also calculates native epitopes that are not present in the altered isoform, i.e., splicing-affected self-epitopes. Our analyses provide evidence that splicing alterations can modify the repertoire of epitopes in tumors and potentially impact the response to immune therapy. ISOTOPE facilitates the study of splicing alterations to fully characterize the determinants of response to immunotherapies.

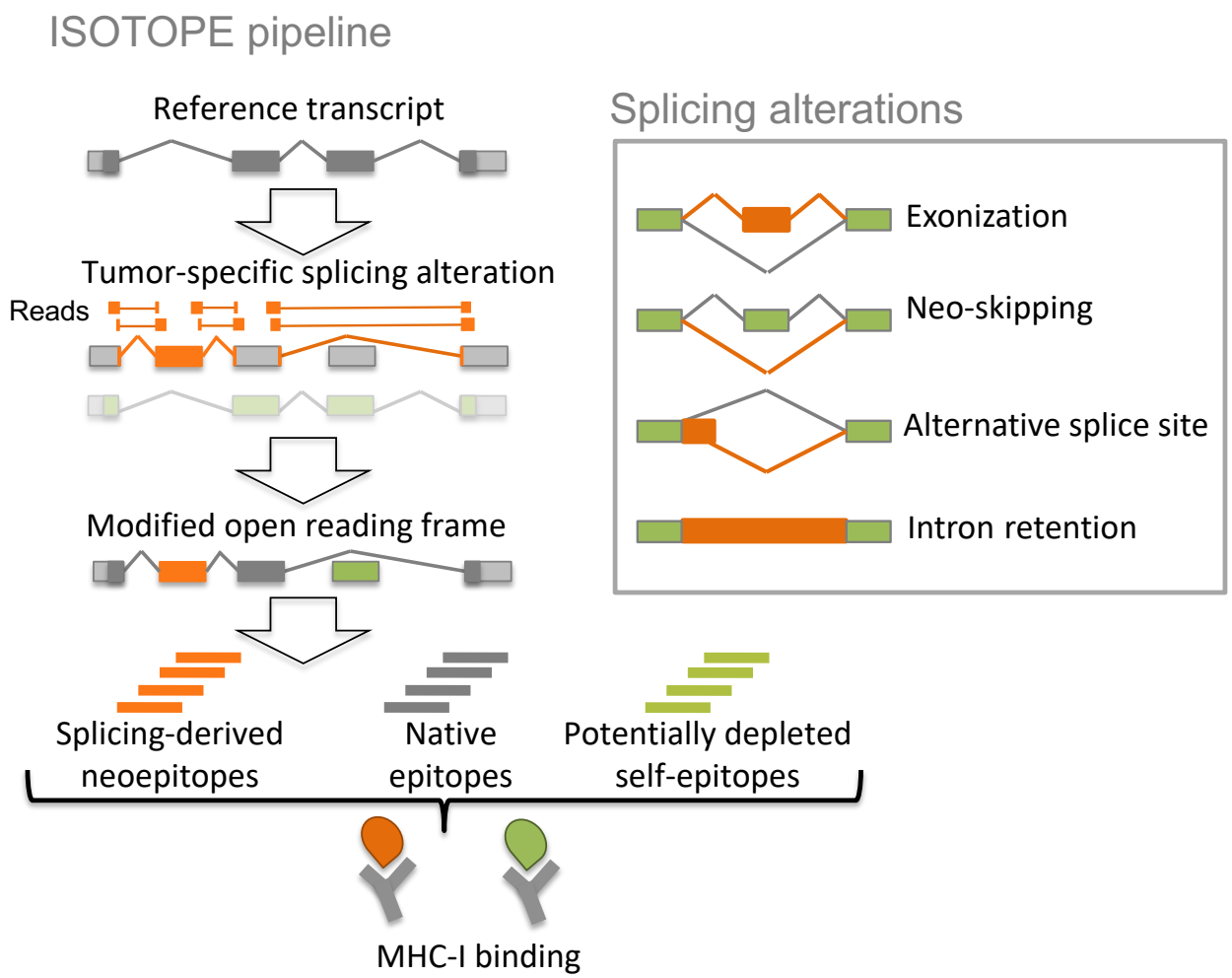

Figure 1. ISOTOPE pipeline. Tumor-specific splicing alterations are defined as significant variations with respect to the exon-intron structures expressed in normal samples and are classified into four different types: *de novo* exonization, new exon skipping (neoskipping), alternative (5'/3') splice site, and intron retention. ISOTOPE calculates the modified open reading frame (ORF) from the reference ORF using the splicing alterations, and identifies the candidate splicing-derived neoepitopes and self-epitopes encoded by the

reference transcript that would not be present in the modified ORF as a consequence of the splicing alteration. These candidate peptides are then tested for affinity with the MHC complexes (see Methods for details).

Results

Comprehensive identification of tumor-specific splicing-derived epitopes

ISOTOPE identifies tumor-specific splicing alterations by generating a catalogue of all exon-exon junctions calculated from RNA sequencing (RNA-seq) reads from each individual tumor sample, filtering out those that appear in any of the samples from a comprehensive set of normal controls (Fig. 1). The remaining junctions are classified into one of four possible types: *de novo* exonization, new exon skipping event, alternative splice site, and intron retention. ISOTOPE performs an empirical test to establish the significance of each candidate splicing alteration taking into account the read support of the event and the coverage and splicing variation in the same gene locus. This test ensures the robustness of the events detected. Changes in the protein products are predicted through the impact of the splicing alterations on the open reading frames (ORF) of the reference transcriptome. This reference transcriptome is obtained by selecting from each gene the transcript with the highest mean expression in the control normal samples. 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 splicing-derived neoepitopes as MHC binders that are expressed in the tumor sample but not in the control normal samples, i.e., *splicing-neoepitopes*. Additionally, MHC binders that are expressed in the control sample but are potentially removed by the change in the ORF through the tumor-specific splicing alteration are also calculated and referred to as splicing-affected self-epitopes, *self-epitopes* for short. Further details are provided in the Methods section.

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

ISOTOPE operates on individual tumor samples, without necessarily having a matched normal sample. We thus first tested 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). However, PHLAT showed greater consistency across most of the HLA types and recovered above 80% of cases for HLA-I and between 65% and 90% for HLA-II types (Supp. Table 1). We thus decided to use PHLAT for further analyses with ISOTOPE.

To test the ability of ISOTOPE 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). *De novo* exonization was the least common of all splicing event types, whereas new junctions skipping one or more exons, i.e., neoskipping, aberrant splice-sites, and intron retentions were more frequent (Fig. 2b). Although most of the splicing alterations did not affect the

encoded ORF, neoskipping events impacted more frequently the ORF compared with the other event types (Fig. 2b).

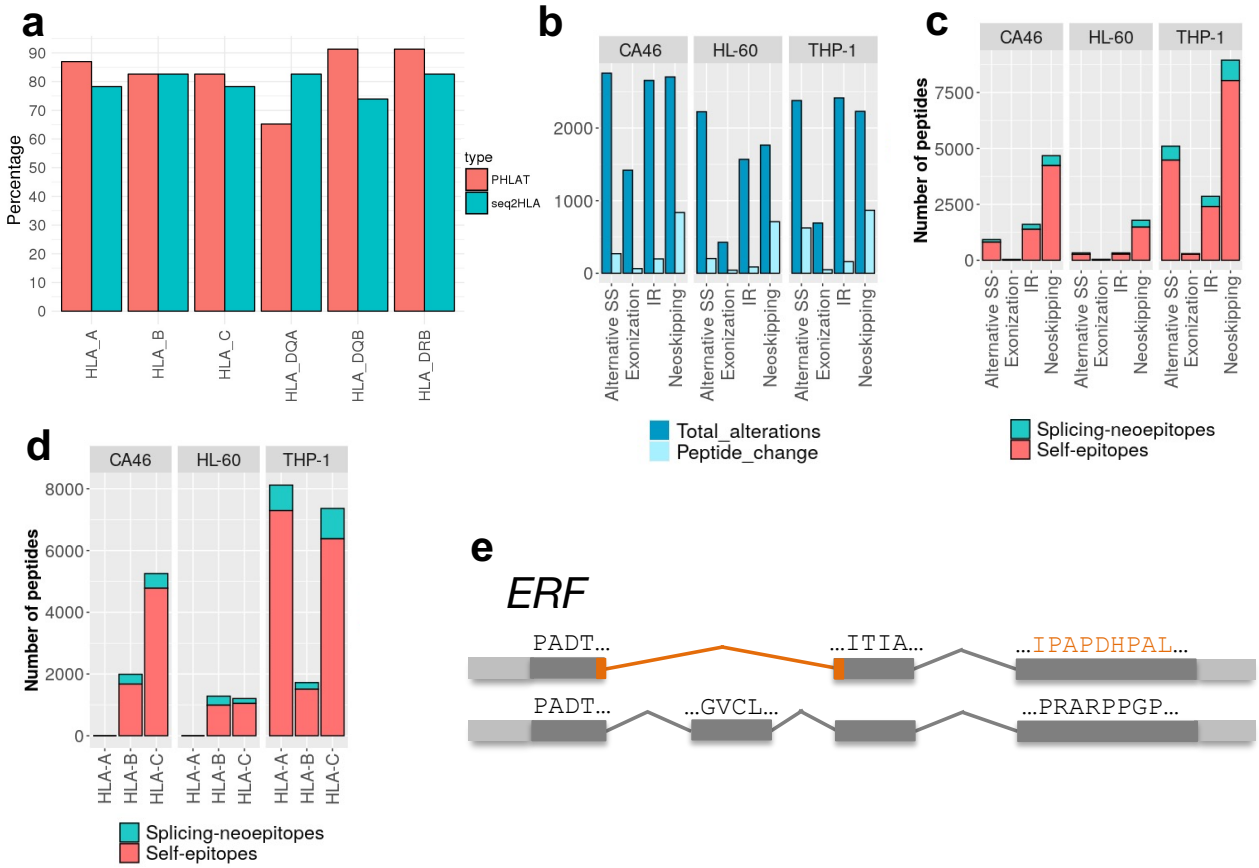

Figure 2. Initial testing of ISOTOPE. (a) Validation of the HLA type prediction from tumor RNA-seq data. We show the predictions for MHC Class I (HLA-A, HLA-B, HLA-C) and II (HLA-DQA, HLA-DQB, HLA-DRB) with PHLAT (red) and SeqHLA (blue). Each bar corresponds to the proportion of samples (over a total 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 splicing-derived neoepitopes (splicing-neoepitopes) (red) and splicing-affected self-epitopes (self-epitopes) (blue) detected for each of the splicing alterations in each of cell lines analyzed (CA46, HL-60 and THP-1). **(d)** as in (c) but separated by HLA-type. **(e)** Example of a splicing-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.

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. Moreover, 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 were not encoded in the reference transcriptome. 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 compared to intron retention events. Candidate self-epitopes that were affected by the splicing alteration were more common than the splicing-epitopes (Fig. 2c). Moreover, separation of these candidate splicing-neoepitopes and self-epitopes according to HLA-types followed closely the results by cell line (Fig 2d), indicating an agreement in the MHC affinity of the peptides found in the cell lines and the HLA class predicted.

To test the potential therapeutic implications of these findings, we tested whether genes from two databases of treatment-associated responses were significantly represented in the set of genes with splicing-neoepitopes or splicing-affected self-epitopes. We observed genes linked with therapy response in lymphoma were significantly represented in the set of self-epitopes (Supp. Table 3). This result suggests a possible role of the splicing alterations detected in the involvement of these genes in therapy response. To validate the candidate splicing-derived neoepitopes we used MHC-1 associated mass-spectrometry (MS) data available for the same cell lines (Ritz et al. 2016).(Ritz et al. 2016) We identified three neoepitopes, all of them generated by neoskipping events in the genes *TOP1* (KRFEPLGMQK), *ERF* (IPAPDHPAL) and *IFRD2* (RTALGGMSW) (Fig. 2e). These are different from the three peptides detected previously using the same datasets but only analyzing intron retention (Smart et al. 2018). This disparity is possibly due to the different criteria used in the selection of relevant events. We performed an empirical test to keep only events with significant read support and considered other splicing alteration types beyond intron retention. Our results indicate that new types of splicing alteration can potentially produce tumor neoepitopes.

To further test ISOTOPE, we analyzed RNA-seq and MHC-I associated mass spectrometry data from ten breast cancer cell lines (MCF7, T47D, LY2, BT549, CAMA1, HCC1395, HCC1419, HCC1428, HCC1569, HCC1806) (Rozanov et al. 2018; Barretina et al. 2012). The most frequent splicing alterations found were IR events, except for cell lines HCC1569 and LY2, for which neoskipping events were the most frequent (Fig. 3a). However, for all types, neoskipping events produced the largest number of changes in ORFs in all cell lines. As before, we predicted the MHC-I binding potential for candidate epitopes, either splicing-derived neoepitopes or splicing-affected self-epitopes. Neoskipping events produced the largest number of neoepitopes (Fig. 3b) (Supp. Table 4). As before, we observed more potentially affected self-epitopes than splicing-derived neoepitopes. Separating by HLA-type, splicing-derived 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 genes producing splicing-derived neoepitopes (*ERBB2*, *ESR1*, *TIMP1*, *ABCC3*) or potentially depleted self-epitopes (*AKT1*, *CCNE1*, *RET*, *TFF3*). Next we searched the MHC-I associated mass spectrometry data for the same breast cancer cell lines (Rozanov et al. 2018) for the predicted neoepitopes. We only identified one significant peptide match in the ten 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).

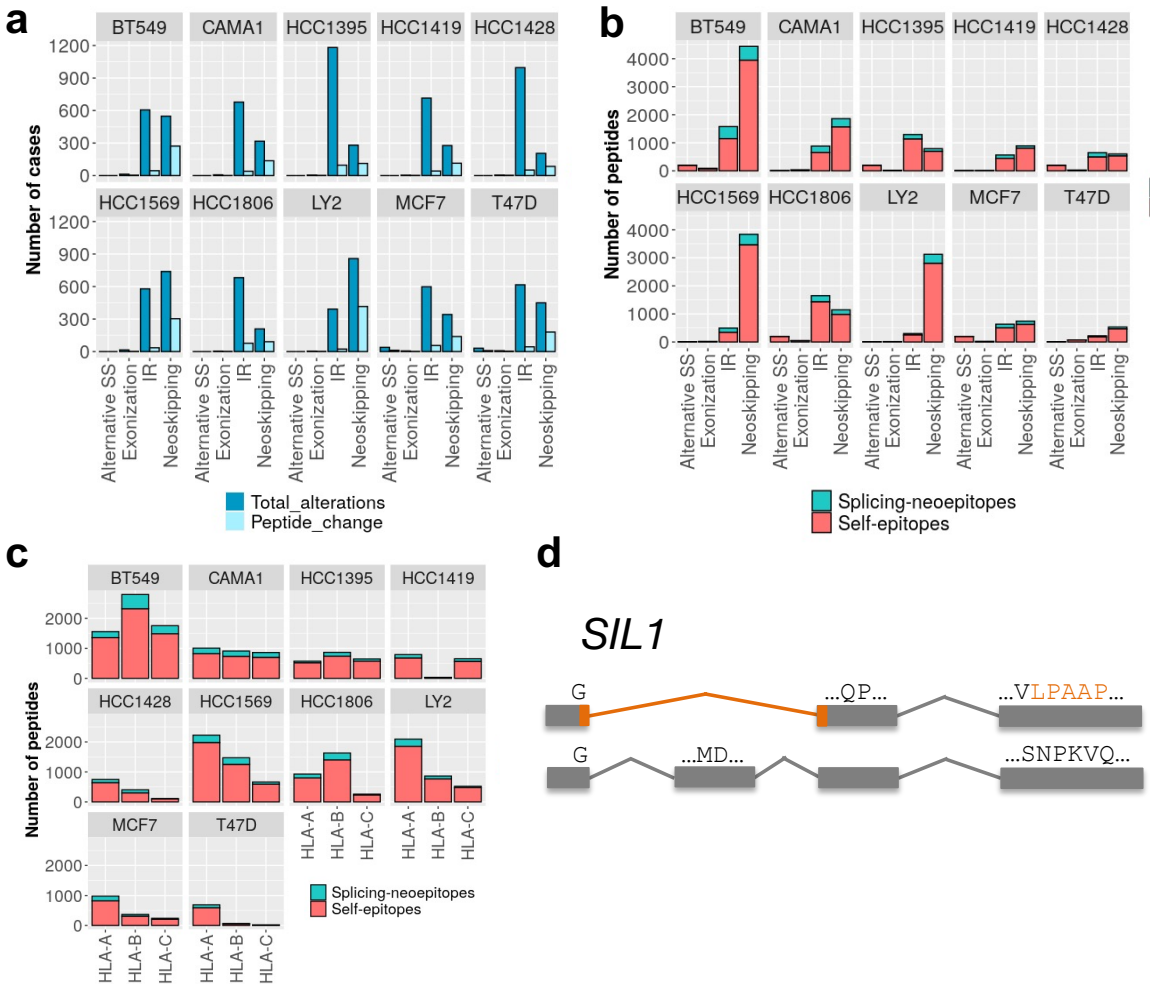

Figure 3. Splicing epitopes in ten breast cancer cell lines. (a) For each breast cancer cell line analyzed, the bar plots show the number of splicing alterations measured and the number of cases leading to a change in the reference protein. Alterations shown are alternative 5' or 3' splice-site (A5_A3), *de novo* exonizations (exonization), intron retentions (IR), and new exon skipping events (neoskipping). **(b)** Number of splicing-derived neoepitopes (splicing-neoepitopes) (red) and splicing-affected self-epitopes (self-epitopes) (blue) for each of the splicing alterations in each of the breast cancer cell lines tested. **(c)** as in (c) but separated by HLA-type. **(d)** Example of a splicing-derived neoepitope from a neoskipping event in the gene *SIL1* validated with MHC-I associated mass spectrometry in the cell line BT549.

Tumor-specific splicing alterations impact self-epitopes and leads to more neoepitopes than mutations.

We described above that tumor-specific splicing alterations potentially affect part of the open reading frame expressed in normal samples that could function as a self-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, which has been associated with its response to immune therapy (Hellmann et al. 2018). Interestingly, SCLC

presents a significantly higher density of mutations in introns compared to exons (Fig. 4a), which may associate with a widespread impact on RNA-processing. Accordingly, SCLC represents an interesting tumor type to investigate how splicing alterations may contribute to neoepitope burden in tumor cells.

We applied ISOTOPE to RNA-seq from 123 small cell lung cancer (SCLC) patients. We derived an exhaustive compendium of SCLC-specific splicing alterations by filtering out all SCLC junctions that appeared in a comprehensive dataset of normal splice junctions (Supp. Fig. 1). We found a total 14643 aberrant splice sites, 7039 intron retentions, 1311 neoskipping events, and 290 *de novo* exonizations that were SCLC specific, and were affecting 2955, 149, 620, and 169 genes, respectively (Fig. 4b) (Supp. Table 5). The identified SCLC-specific splicing alterations distributed homogeneously across all samples and showed no association to mutations on spliceosomal factors or overexpression of MYC genes (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 (Fig. 4b) (Supp. Fig. 3).

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 ≤ 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 splicing-affected self-epitopes (Figs. 4c and 4d). This imbalance towards the potential elimination of self-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 and 4d).

We could not detect any significant association with SCLC-specific response biomarkers but did observe multiple significant associations of SCLC-specific splicing-derived neoepitopes with response biomarkers from other tissues (Supp. Table 3). These results are especially relevant in SCLC, for which no alteration has been yet described as therapeutically targetable. We did not have access to MHC-I associated mass-spectrometry data 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 self-epitopes predicted to be potentially depleted in the altered isoform. To test the significance of the association of the predicted epitopes to the mass spectrometry data, we performed a randomized comparison. We took 1000 random peptides predicted with high affinity (≤ 500nM) and 1000 peptides from the entire set of self-epitopes 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 a significantly higher number of validations for the candidate self-epitopes with high affinity (Kolmogorov-Smirnov p-value = 3.44e-13).

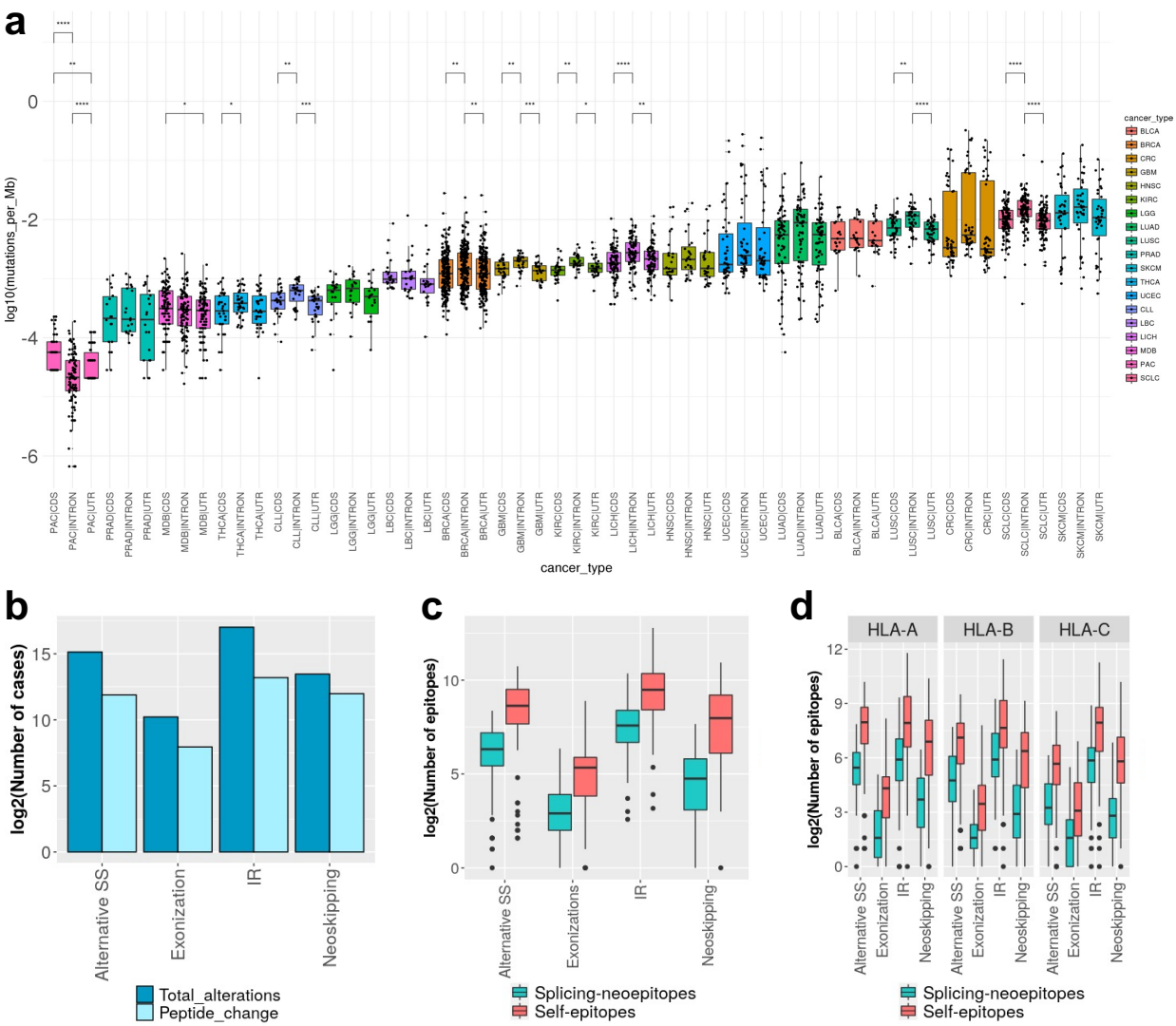

Figure 4. Splicing epitopes in small cell lung cancer. (a) Mutation burden (y axis) calculated separately for introns (INTRON), coding exons (CDS) and non-coding exonic regions in protein-coding genes (UTR) calculated from whole genome sequencing (WGS) data for several tumor types (x axis), including small cell lung cancer (SCLC). We indicate the pairs of 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)** Number of splicing alterations (y axis) according to event type (x axis), indicating all alterations and the subset that impact the open reading frame (ORF). **(c)** Distribution of splicing-derived neoepitopes (splicing-neoepitopes) and splicing-affected self-epitopes (self-epitopes), separated by splicing alteration type. **(d)** Same as (c) but separated by HLA-type.

We further tested the association with tumor mutation burden (TMB). Overall, we found no association between the TMB and the number of splicing alterations or the number of epitopes (splicing-neoepitopes or self-epitopes). However, there was some association for the neoskipping events across all samples (Supp. Fig. 4). We next tested the association of splicing-derived neoepitopes with mutational neoepitopes. In the subset of SCLC samples with whole genome sequencing (WGS) data (George et al. 2015; Peifer et al. 2012) there were more splicing- neoepitopes than mutational ones, and there was a weak correlation between their numbers in patients (Spearman rho=0.397, p-value = 0.003) (Supp. Fig. 5). On the other hand, none of the splicing-neoepitopes matched any of the mutational neoepitopes.

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

To test whether tumor-specific splicing-derived neoepitopes may be associated to the patient response to immune therapy, we applied ISOTOPE to RNA-seq data from two cohorts of melanoma patient samples prior to treatment with anti-*CTLA4* (Van Allen et al. 2015) or anti-*PD1* (Hugo et al. 2016) (Supp. Tables 6 and 7). We calculated all the tumor-specific splicing alterations in each patient sample by removing all events that also occurred in a large set of control normal samples analyzed. Intron retention was the most abundant alteration but the impact on the encoded protein was not equally abundant in both cohorts (Fig. 5a). Despite these differences, there was an overall decrease of the ORF lengths as a consequence of the splicing alterations (Supp. Fig. 6), in agreement with previous studies showing a reduction of ORF lengths expressed in tumors (Climente-González et al. 2017).

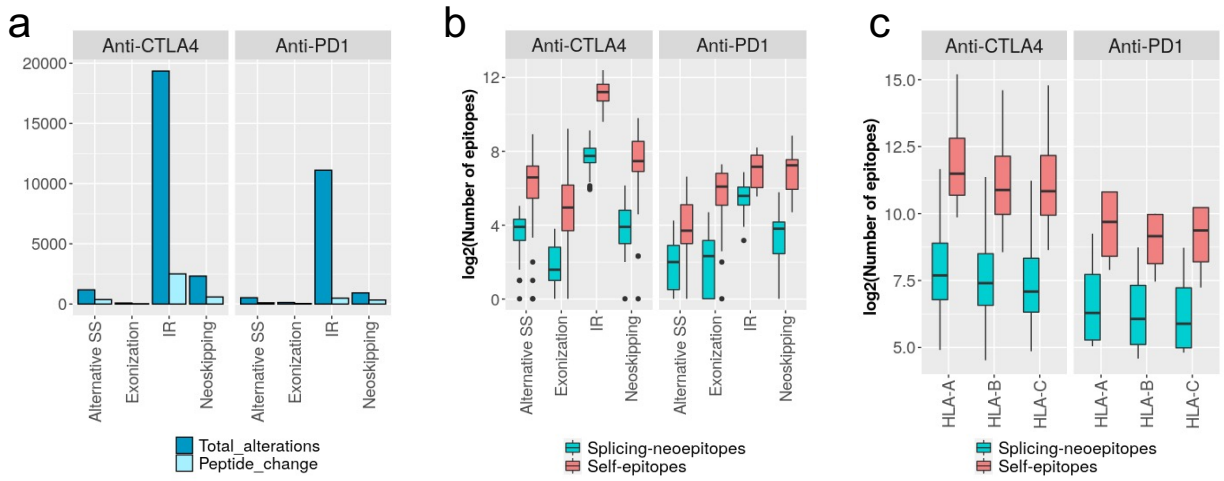

Figure 5. Splicing epitopes in two melanoma cohorts. (a) Total number of events and subset of protein-affecting events in the melanoma cohorts treated with anti-CTLA4 and with anti-PD1. **(b)** Distribution of the number of candidate tumor-specific splicing-derived neoepitopes (splicing-epitopes) and self-epitopes that would be depleted in the altered isoform (self-epitopes). **(c)** Distribution of the number of candidate epitopes from (b), separated by HLA-type.

As for other samples tested, the overall number of splicing-affected self-epitopes was overall higher than the splicing-derived neoepitopes, with larger numbers of self-epitopes affected by intron retention events in the anti-*CTLA4* cohort (Fig. 5b). Moreover, the anti-*CTLA4* cohort presented more epitopes from both classes for all HLA-types (Fig. 5c). These results did not change when we used ≤300nM to define the candidate epitopes (Supp. Fig. 7). 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 splicing-derived neoepitopes in responders and non-responders in anti-*CTLA4* or anti-*PD1* showed no significant difference (Supp. Fig. 8). Separating the splicing alterations by type, we found in general a higher proportion of self-epitopes affected by splicing in all patients.

We did not observe any differences in the proportion of epitopes between responders and non-responders to anti-*CTLA4* (Fig. 6a). However, responders to anti-*PD1* therapy had more splicing-

affected self-epitopes from intron retention events compared to non-responders. Other splicing alterations did not show any significant differences (Fig. 6b). We found similar results using the threshold 300nM to define candidates (Supp. Fig. 8). To further test the potential role of splicing-derived neoepitopes and splicing-affected self-epitopes in the response to immunotherapy, we studied their binding affinities. Splicing-derived neoepitopes showed stronger MHC-I interaction (lower values of binding affinity) in responders. In particular, splicing-derived neoepitopes from *de novo* exonizations (Fig. 6c) and neoskipping events (Fig. 6d) had stronger interactions in anti-*PD1* responders, and those associated with intron-retention events had stronger interactions in anti-*CTLA4* responders (Fig. 6e). Incidentally, splicing-affected self-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. 9).

When we considered the threshold of 300nM to define candidates, not all comparisons remained significant, but we observed similar trends (Supp. Fig. 10). We looked at various properties in the samples to test for possible confounding effects with the therapy response. The number of sequencing reads, and the overall distribution of transcript expression did not vary significantly between responders and non-responders (Supp. Fig. 11). We also observed no general association of response with the immune infiltration or purity of samples (Supp. Fig. 12). Moreover, although we observed a significant association between response and estimated stromal content in the anti-PD1 cohort, there was no correlation between the number of splicing-neoepitopes and the estimated stromal content (Supp. Fig. 12).

Finally, 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 type 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 of HLA type HLA-C0701 (Fig. 6g).

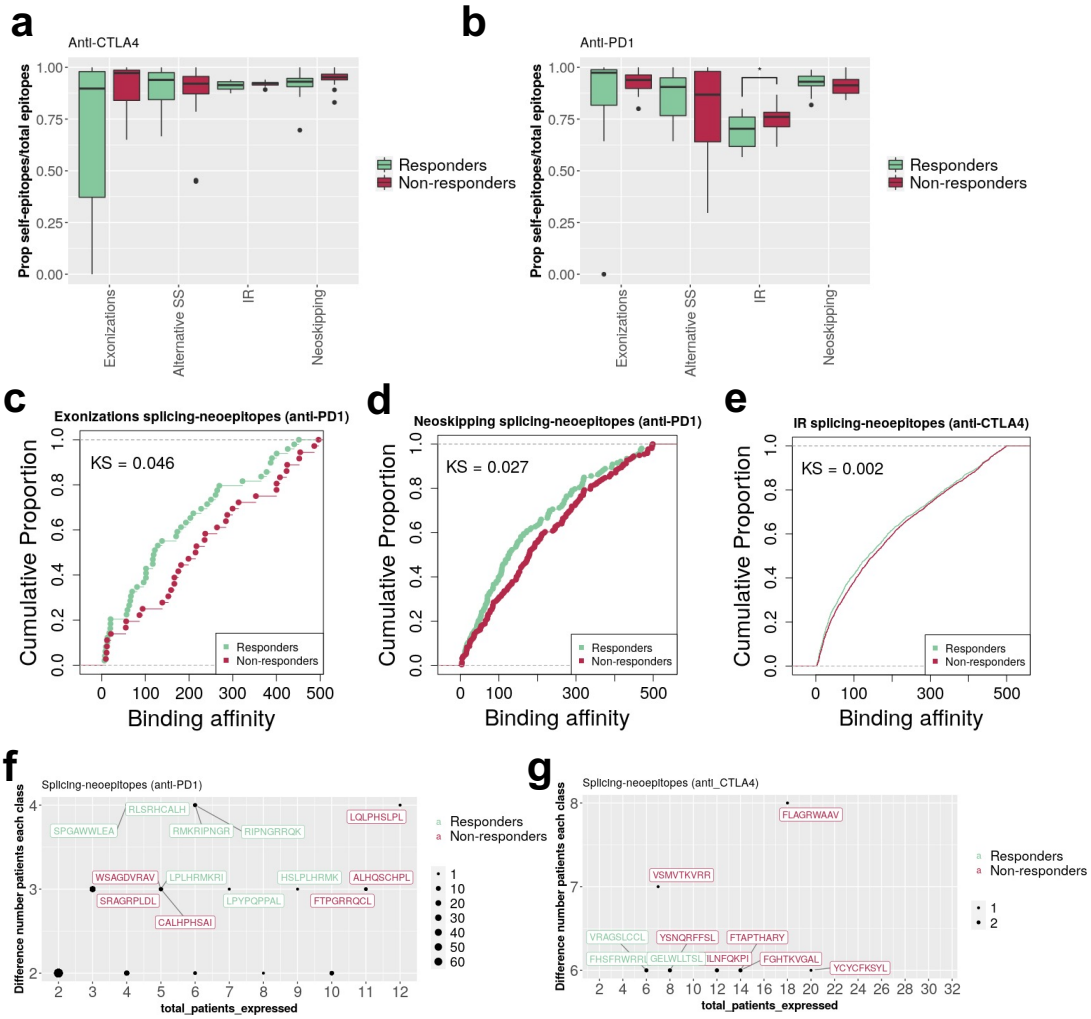

Figure 6. Splicing epitopes and response to immune therapy. (a) Proportion of splicing-affected self-epitopes (self-epitopes) over the total of epitopes, i.e., splicing-derived neoepitopes (splicing-neoepitopes) plus self-epitopes, (y axis) for patients treated with anti-*CTLA4,* separated by type of splicing alteration (x axis) and by patient response: responder (green) or non-responder (red). **(b)** As in (a) but for a different cohort of melanoma patients treated with anti-*PD1*. **(c)** Cumulative plots of the binding affinities (x axis) of splicing-neoepitopes in melanoma tumors from exonization events separated in responders (green) and non-responders (red) to anti-PD1 therapy. Kolmogorov-Smirnov test p-value (KS) = 0.0465 **(d)** As in (c), for splicing-derived neoepitopes from neoskipping events, KS = 0.0274. **(e)** Cumulative plots of the affinities of splicing-neoepitopes 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-neoepitopes represented according to the total number of patients in which are predicted (x axis, *total_patients_expressed*) and to the absolute count-difference in 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.

Discussion

Our comprehensive analysis cancer-specific splicing alterations indicate that splicing changes of any kind may potentially contribute to the immunopeptidome, hence they should be considered in studies of cancer and immunotherapy. Our approach, implemented in the pipeline ISOTOPE (https://github.com/comprna/ISOTOPE), presents several 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 characterized. Thus, making possible an assessment at unprecedented scale of candidate splicing-derived neoepitopes. Although tumor-associated intron retention is quite common in tumors (Dvinge and Bradley 2015), we observed that neoskipping 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 matched controls. Moreover, unlike previous studies (Jayasinghe et al. 2018; Kahles et al. 2018; Smart et al. 2018), our analysis describes tumor-specific alterations by comparing to a large compendium of normal samples and performs an empirical 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). Finally, we provide the pipeline and instructions to run the analyses. This was not made available in previous studies. In summary, ISOTOPE enables a robust and exhaustive survey of the immunogenic impacts of splicing in cancer.

The low validation of splicing-neoepitopes with MHC-associated proteomics would suggest that there is a small contribution of tumor-specific splicing to novel epitopes, in agreement with previous studies. 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 tumor-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 additional reason may be 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. It is also possible that most peptides associated to the immune recognition of tumor samples may be produced through other mechanisms or may not be novel from the point of view of the expression pattern.

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 we found no differences in the number of splicing-derived neoepitopes 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. It has been shown before that response tend to correlate with tumor mutation burden. Our analysis indicated a weak correlation of the number of splicing-neoepitopes with tumor mutation burden. Moreover, splicing-neoepitopes were generally more abundant that mutational neoepitopes and showed no overlap between them. This suggests the exciting possibility that splicing-neoepitopes may represent powerful biomarkers of immunogenicity independently of the mutational patterns.

Additionally, we have described how splicing alterations could affect the open reading frame in such a way that certain self-epitopes are no longer produced. This raises the interesting question about the impacts that the lack of these self-epitopes might have. 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 but 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 splicing-neoepitope production, and we could validate many of them from MHC-associated mass spectrometry in lymphocytes. Thus, tumor-specific splicing alterations could generate neoepitopes, but could also potentially 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 tumor-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, including the variety of splicing alterations that may arise 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 analyzed was collected from the cancer cell line encyclopedia (CCLE) (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 classified 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 the SCLC patients from (Rudin et al. 2012) we also obtained the matched normal controls. Samples with more than 30% of junctions present in other samples but with missing value in them, were filtered out. We estimated the stromal content, immune infiltrate, and tumor purity of every sample from gene expression information using the ESTIMATE R package (v.1.0.13) (Yoshihara et al. 2013).

Identification of tumor-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 or 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 human 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 normal samples from a comprehensive dataset collected from multiple sources: 7859 normal samples from 18 tissues from the GTEX consortium (Consortium 2015), normal samples from Intropolis (Nellore et al. 2016), CHESS 2.0 (Pertea et al. 2018), and 24 matched normal samples from lung (Rudin et al. 2012).

ISOTOPE classifies the novel junctions in clusters as one of the following types: aberrant splice-site, new exon skipping (neoskipping), or *de novo* exonization. To define exonizations, we considered all pairs of spliced junctions that were not present in normal samples (see above) 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 neoskippings, we considered those new junctions that skipped known exons and defined new

connections between exons. To define retained introns (RIs) we used KMA (Pimentel et al. 2015) to extend the Gencode (v19) transcriptome with potential retained introns (RIs), which we quantified in 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 human Gencode (Harrow et al. 2012) and the CHESS 2.0 (Pertea et al. 2018) annotations, and removed KMA-predicted RIs that appeared in the SUPPA RI annotations. To control for confounding effects due to defects in pre-mRNA processing across the entire gene locus, 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. Candidate 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.

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 from 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

ISOTOPE calculates two types of epitopes. One type corresponds to tumor-specific splicing-derived neoepitopes. These are peptides with affinity to the MHC-I complex that are not encoded in the wild-type transcripts but are encoded in the altered ORF as a consequence of the tumor-specific splicing alteration. The second type corresponds to splicing-affected self-epitopes. These are peptides with affinity to the MHC-I complex that are encoded in the wild-type transcripts but would not be encoded in the altered ORF, i.e. potentially depleted, as a consequence of the tumor-specific splicing alteration.

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 splicing neoepitopes, 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 tested with a Kolomogorov-Simrnoff test whether the 2 distributions of the number of peptides validated were significantly different. For the self-epitopes in SCLC was significant (p-value = 3.44e-13), whereas for splicing-neoepitopes there was no significant difference.

Somatic mutation data and detection of mutation-derived neoepitopes

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).

Mutation-derived epitopes were calculated with pVACtools (Hundal et al. 2020), using whole genome sequencing data (WGS) for two SCLC cohorts (Peifer et al. 2012; George et al. 2015). The identification of splicing-derived neoepitopes was carried out with ISOTOPE using the RNA-seq data from the same patient samples. The candidate epitopes were calculated in both cases using the same tools, NetMHC and NetMHCPan, with the same parameters, testing peptides with amino acid length from 8 to 11, and selecting candidates with binding affinities less or equal than 500nM.

Biomarker 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, in the form of genetic alterations, associated to the treatment response to anti-cancer therapy. Genes with predicted splicing-neoepitopes or self-epitopes 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., J.Y., and F.S. supervised the analyses. E.E. and J.L.T. wrote the manuscript with essential inputs from J.Y., 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–e25374.

Andreatta M, Nielsen M. 2016. Gapped sequence alignment using artificial neural networks:

application to the MHC class I system. *Bioinformatics* **32**: 511–517.
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**: 603–607.
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**: 525–527.
Cappelli LC, Shah AA, Bingham CO. 2016. Cancer immunotherapy-induced rheumatic diseases emerge as new clinical entities. *RMD Open* **2**: e000321–e000321. http://rmdopen.bmj.com/lookup/doi/10.1136/rmdopen-2016-000321.
Climente-González H, Porta-Pardo E, Godzik A, Eyras E. 2017. The Functional Impact of Alternative Splicing in Cancer. *Cell Rep* **20**: 2215–2226.
Consortium Gte. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**: 648–660.
Cooper S, Spiro SG. 2006. Small cell lung cancer: treatment review. *Respirology* **11**: 241–248.
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 NC, Krysiak K, McMichael JF, Coffman AC, Danos AM, Ainscough BJ, Ramirez CA, Rieke DT, Kujan L, et al. 2017. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* **49**: 170–174. http://www.ncbi.nlm.nih.gov/pubmed/28138153.
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**: 1760–1774.
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.

Hundal J, Kiwala S, McMichael J, Miller CA, Xia H, Wollam AT, Liu CJ, Zhao S, Feng Y-Y, Graubert AP, et al. 2020. pVACtools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens. *Cancer Immunol Res* **8**: 409–420. http://www.ncbi.nlm.nih.gov/pubmed/31907209.

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**: 616–621.

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, Network CGAR, 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–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–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 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, Breitwieser FP, Chang Y-C, Madugundu AK, Pandey A, Salzberg SL. 2018. CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol* **19**: 208. http://www.ncbi.nlm.nih.gov/pubmed/30486838.
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**: 1570–1580.
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**: 1111–1116.
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, Schroeder MP, Vivancos A, Rovira A, Tusquets I, Albanell J, Rodon J, Tabernero J, et al. 2018. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med* **10**: 25.

http://www.ncbi.nlm.nih.gov/pubmed/29592813.
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–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.