



OPEN ACCESS

EDITED BY

Tiezheng Hou,
University College London,
United Kingdom

REVIEWED BY

Duoyi Zhao,
Fourth Affiliated Hospital of China Medical
University, China
Yijia Wang,
Nankai University, China

*CORRESPONDENCE

Trevor Clancy
✉ trevor@oncoimmunity.com

†These authors have contributed equally to
this work

RECEIVED 21 May 2023

ACCEPTED 10 July 2023

PUBLISHED 20 September 2023

CITATION

Anzar I, Malone B, Samarakoon P,
Vardaxis I, Simovski B, Fontenelle H,
Meza-Zepeda LA, Stratford R, Keung EZ,
Burgess M, Tawbi HA, Myklebost O and
Clancy T (2023) The interplay between
neoantigens and immune cells in sarcomas
treated with checkpoint inhibition.
Front. Immunol. 14:1226445.
doi: 10.3389/fimmu.2023.1226445

COPYRIGHT

© 2023 Anzar, Malone, Samarakoon,
Vardaxis, Simovski, Fontenelle, Meza-Zepeda,
Stratford, Keung, Burgess, Tawbi, Myklebost
and Clancy. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

The interplay between neoantigens and immune cells in sarcomas treated with checkpoint inhibition

Irantzu Anzar^{1,2}, Brandon Malone¹, Pubudu Samarakoon¹,
Ioannis Vardaxis¹, Boris Simovski¹, Hugues Fontenelle¹,
Leonardo A. Meza-Zepeda^{3,4}, Richard Stratford¹,
Emily Z. Keung⁵, Melissa Burgess⁶, Hussein A. Tawbi⁷,
Ola Myklebost^{3,8†} and Trevor Clancy^{1*†}

¹Oslo Cancer Cluster, NEC OncoImmunity AS, Oslo, Norway, ²Institute of Clinical Medicine, University
of Oslo, Oslo, Norway, ³Institute for Cancer Research, Oslo University Hospital, Oslo, Norway,

⁴Genomics Core Facility, Department of Core Facilities, Oslo University Hospital, Oslo, Norway,

⁵Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston,
TX, United States, ⁶Department of Medical Oncology, University of Pittsburgh Medical Center,

Pittsburgh, PA, United States, ⁷Department of Melanoma Medical Oncology, The University of Texas
MD Anderson Cancer Center, Houston, TX, United States, ⁸Department of Clinical Science, University
of Bergen, Bergen, Norway

Introduction: Sarcomas are comprised of diverse bone and connective tissue tumors with few effective therapeutic options for locally advanced unresectable and/or metastatic disease. Recent advances in immunotherapy, in particular immune checkpoint inhibition (ICI), have shown promising outcomes in several cancer indications. Unfortunately, ICI therapy has provided only modest clinical responses and seems moderately effective in a subset of the diverse subtypes.

Methods: To explore the immune parameters governing ICI therapy resistance or immune escape, we performed whole exome sequencing (WES) on tumors and their matched normal blood, in addition to RNA-seq from tumors of 31 sarcoma patients treated with pembrolizumab. We used advanced computational methods to investigate key immune properties, such as neoantigens and immune cell composition in the tumor microenvironment (TME).

Results: A multifactorial analysis suggested that expression of high quality neoantigens in the context of specific immune cells in the TME are key prognostic markers of progression-free survival (PFS). The presence of several types of immune cells, including T cells, B cells and macrophages, in the TME were associated with improved PFS. Importantly, we also found the presence of both CD8+ T cells and neoantigens together was associated with improved survival compared to the presence of CD8+ T cells or neoantigens alone. Interestingly, this trend was not identified with the combined presence of CD8+ T cells and TMB; suggesting that a combined CD8+ T cell and neoantigen effect on PFS was important.

Discussion: The outcome of this study may inform future trials that may lead to improved outcomes for sarcoma patients treated with ICI.

KEYWORDS

neoantigens, checkpoint inhibition therapy, next generation sequencing, biomarker discovery, machine learning, immune escape

Introduction

Sarcomas are a rare and heterogenous group of tumors that account for 1% of all cancers and 10–15% of solid tumors in children and young adults (1, 2). They are generally divided into soft tissue sarcomas (STS) and bone sarcomas (BS); however, sarcomas are effectively comprised of more than 150 distinct subtypes (1, 3). The majority of sarcomas are STS, while BS accounts for 15% of cases (2). In addition to the rarity of sarcomas in adults, each distinct subtype is associated with diverse genetic, molecular, anatomical, clinical and/or age related factors; making their study, diagnosis or treatment enormously challenging (4). Sarcomas are broadly considered to be “cold” tumors, with low immune cell infiltration, making them potentially challenging targets of ICI therapy (5).

Although sarcomas have had poor performance in ICI therapy clinical trials (5), the tremendous success of ICI in other cancer indications offers prospects that a path toward curative immunotherapy success may be possible for sarcoma patients. Several clinical trials evaluating ICI in sarcomas (including the SARC028/NCT02301039 trial pertaining to this study) have reported at best only modest responses, with an overall response rate (ORR) of approximately 15% (6–10). However, a small number of patients in these clinical studies had a notable positive clinical outcome. This observation, coupled with the complete remissions reported in individual case reports, warrants further research to decipher the precise mechanisms of the positive ICI therapy responses in sarcomas (11, 12), and potentially to develop predictive biomarkers for ICI and other cancer immunotherapies (6, 9, 13, 14). The tumor mutational burden (TMB), conventionally used as a predictive biomarker to ICI therapy in several cancer types (15), is an obvious immunogenomic property to investigate and potentially help elucidate the mechanisms associated to a positive clinical response. However, as we will report here and as reported in previously published studies, TMB remains inconclusive as a biomarker for ICI therapy response in sarcomas (11, 16). Therefore, in this study we were motivated to study the interplay between multiple immunogenomic properties in addition to TMB, such as neoantigen load and infiltration of several immune-cell types into the TME (11, 12).

To achieve this, we investigated 31 available samples from ICI treated sarcoma patients in the SARC028 clinical trial (6). We performed whole-exome sequencing (WES) of tumor and matched normal blood from patients and RNA-sequencing (RNA-seq) of their tumors and identified neoantigens corresponding to multiple sources of genomic variants including single nucleotide variants (SNVs), small insertion and deletions, and gene fusion events. The RNA-seq data from their tumors was used to characterize immune cell infiltration into the TME and the expression patterns of immune-related genes to improve our understanding of the immunobiology of ICI-treated sarcomas. The subsequent exploratory multivariate survival analyses revealed that the specific context of the immune cell composition of the TME and its interplay with neoantigens may be important for improved PFS. The insights gained from this analysis may guide the identification of prognostic biomarkers underlying sarcoma immunotherapy response and may be informative in future clinical trial designs and studies of ICI therapy in sarcomas.

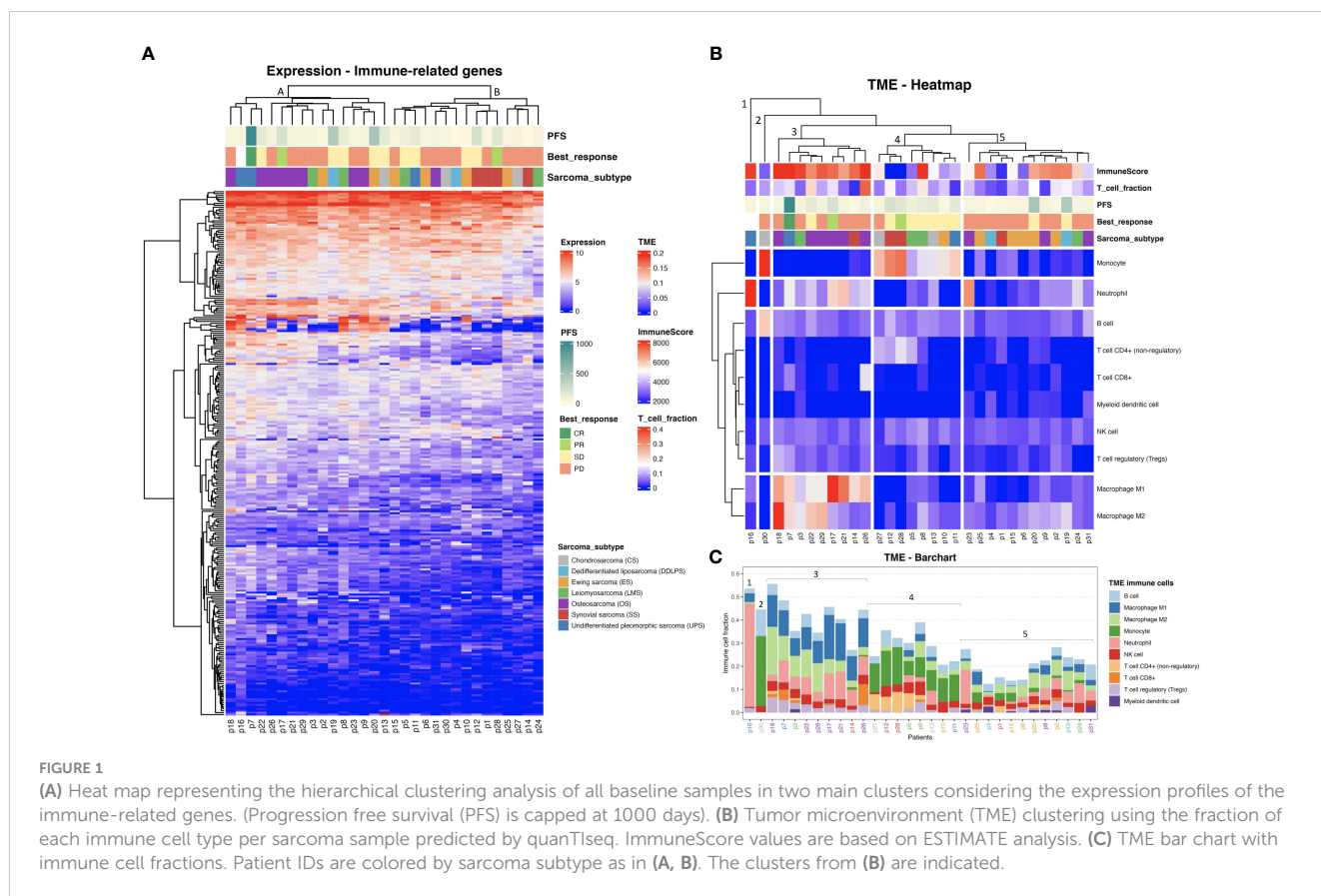
Results

Immune cell infiltration patterns in sarcoma patients treated with ICI therapy

We first analyzed tumor expression profiles by RNA-seq of bulk tumor samples (see Methods). A principal component analysis (PCA) suggested a clustering of patients that corresponded to the sarcoma subtypes (Supplementary Figure S1A). Certain subtypes, such as LMS, had a notable within-subtype heterogeneity, while other subtypes such as SS, were more similar. A subsequent PCA focused exclusively on the selected collection of immune-related genes (Supplementary Figure S1B) suggested that a distinctive difference in immunogenomic expression profile may exist between the different sarcoma subtypes treated with ICI (Supplementary Figure S1B). Interestingly, two out of the three patients responding to ICI clustered together in both PCA plots.

A trend towards elevated expression of genes related to the immune response in UPS and OS relative to the other sarcoma subtypes was observed when using hierarchical clustering of immune-related genes as illustrated in Figure 1A (with an average transcript per million (TPM) of 94 and 84, for UPS and OS respectively, compared to the other sarcoma subtypes where the average TPM was consistently less than 60). The samples made up two main clusters: A, with almost all OS, and B, with all SS, suggesting some relation to the karyotypic subclasses. Further, a hierarchical clustering of the predicted TME composition is depicted in Figure 1B, and in Figure 1C a bar chart of the specific immune cell fractions of each cluster from Figure 1B is portrayed. Cluster 3 in Figures 1B, C consisted mostly of OS and seemed to be dominated by an elevated level of macrophages (both M1 and M2 types were detected). Interestingly, Cluster 4 was enriched in monocytes, and most of the patients in that cluster had a stable disease (SD) response to ICI therapy. Cluster 5 seemed to be the “coldest” immunological group relative to the other clusters, not surprisingly harboring most of ES samples driven by specific gene fusions. Interestingly, B cell infiltration into the TME was predicted to be a common sarcoma feature at relatively consistent levels across all sarcoma subtypes (Figure 1C). The limited availability of matched week-8 data from only seven patients, coupled with the absence of any observed therapeutic responses to ICI among them, hindered our capacity to infer TME composition changes following treatment (Supplementary Figure S2).

As observed in Figures 1B, C, CD8+ T cell infiltration was also predicted at varied levels among the sarcomas. The T cell infiltration predicted by quanTIseq was validated by TcellExTRECT tool. As can be observed in Figure 1B, TcellExTRECT results were concordant with quanTIseq, with a positive significant correlation of 0.41 and p -value = 0.022. The generic infiltration levels of non-cancer cells (*i.e.*, stroma and immune cells) across the different sarcoma samples were also evaluated using ESTIMATE toolkit. The stromal scores from ESTIMATE ranged from 1977 to 7910, while the ESTIMATE immune cell infiltration scores ranged from 574 to 9821, and finally the ESTIMATE tumor purity scores ranged from 2959 to 11086. Regarding the sarcoma subtypes, the mean and standard deviation of immune scores sorted from highest to lowest were as



follows: OS, 6870 (1626); DDLPS, 6295 (2652); UPS, 60797 (2920); LMS, 5977 (2268); ES, 5832 (1884); CS, 4139 (1201); and SS, 2358 (2838). These ESTIMATE predictions corroborated the observations in **Figures 1A, B**, in that OS had a higher immune activation compared to the other sarcoma subtypes. In **Figure 1B**, the “ImmuneScore” annotation bar represents the results of ESTIMATE tool, where it can be clearly observed that Cluster 3, consisting mostly of OS, was the cluster with the highest immune scores.

Additionally, we performed a differential expression (DE) analysis comparing the patients with a clinical response to ICI (responders) to the other patients. We identified a total of 727 differentially expressed genes (DEGs), of which 209 were up- and 518 down-regulated, with an absolute fold change larger than 1 and a p-value less than 0.05 (**Supplementary Figure S3A**). Due to the small number of responders (three), multiple test correction was not applied. Enrichment analysis of the up-regulated genes revealed a significant enrichment of several immune-related Gene Ontology (GO) terms [**Supplementary Figure S3B**, corroborating the notion that an immunologically active or “hot” TME leads to improved clinical outcome to ICI therapy (17)]. A detailed table with all DEGs and the complete list of up-regulated GO terms is provided in **Supplementary Tables 1, 2**, for DEGs and GO terms respectively. The study focuses on the immunogenomic properties of ICI-treated sarcoma patients; hence cancer hallmarks were not examined.

Sarcoma tumors exhibit a highly heterogeneous neoantigen landscape

Using a state-of-the-art somatic mutation calling framework (18), we inferred a comprehensive mutational landscape of the 31 baseline sarcoma tumor samples (see **Supplementary Figure S4** for a detailed overview). Each 9mer and 10mer peptide that had a somatic mutation was matched to the personalized HLA genotype of each individual patient to identify immunogenic neoantigens likely to be presented by the patient’s HLA alleles on their tumor cells’ surface. Such high-quality neoantigens were predicted using the NEC Immune Profiler (NIP) software (see **Methods**) (19), which uses an integrated artificial intelligence (AI) approach trained on proprietary data to predict antigen presentation (AP) scores, which can range from 0 to 1, for each candidate neoantigen. The distribution of neoantigen load (NAL) (see **Methods**) with respect to each sarcoma histological subtype was then assessed (**Figure 2A**) and revealed a highly heterogeneous NAL both between and within subtypes, ranging from 0 to 206 for intra-subtypes and a median range of 15.0 to 112.0 for inter-subtypes. DDLPS, UPS and LMS subtypes exhibited the highest NAL overall; the striking score for DDLPS is due to one case, p4, who had an extremely high number of gene fusions (**Figure 2B**). 461 (28.9%) of the mutated peptides had an AP score (see **Methods**) greater than 0.5 and were also predicted to bind more than one HLA allele in the same patient.

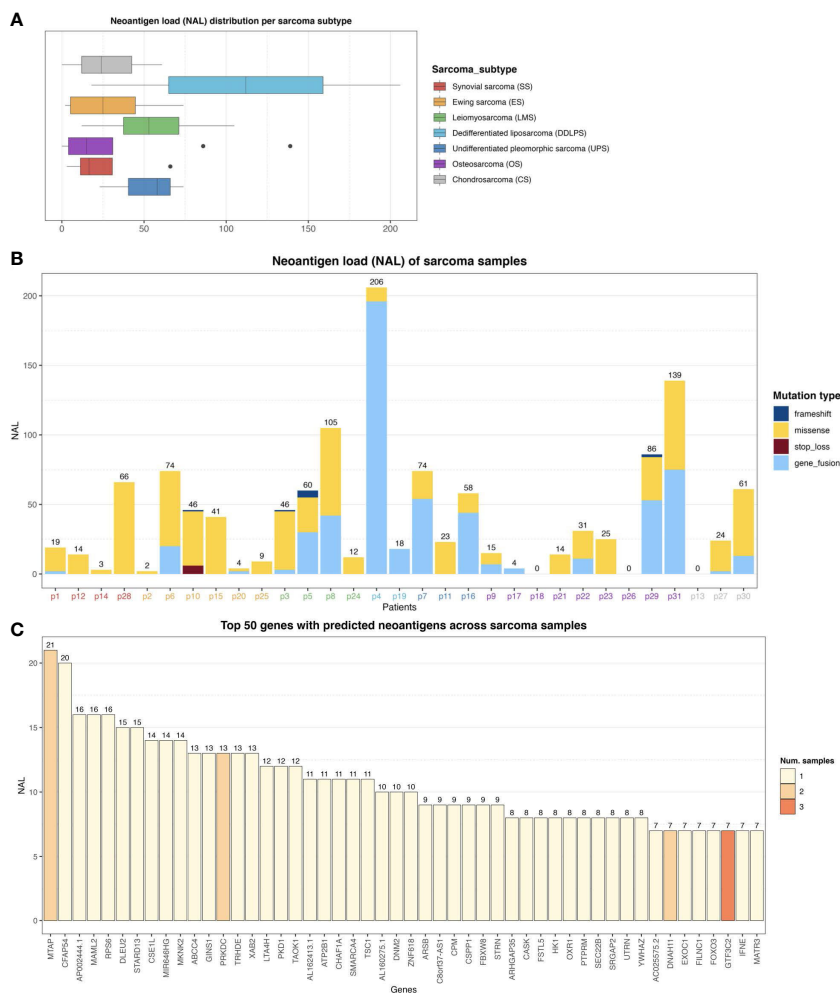


FIGURE 2 (A) Neoantigen load (NAL) distribution across the different sarcoma histological subtypes. The lines inside each box represents the median NAL value for each subtype while the dots are outliers. (B) NAL profiling of sarcoma samples and the contribution of each somatic variant type to NAL. Patient IDs are colored by sarcoma subtype as in (A). (C) Top 50 genes with predicted neoantigens across sarcoma patients. The colors represent the number of patients with neoantigen candidates arising from the given gene.

Figure 2B shows the number of neoantigens generated in each patient, stratified by the type of somatic mutation, outlining a broad diversity of potential neoantigens detected within each subtype.

In this study, no correlation between a conventional TMB and NAL was found (Spearman Rank correlation coefficient of -0.08, and p-value of 0.66). The conventional TMB calculation typically does not take gene fusions into account, and as can be observed in Figure 2B, gene fusions were one of the dominant contributors to NAL for several patients. However, a clear positive correlation between the number of gene fusions and NAL was observed, with a correlation coefficient of 0.92 and p-value of 2.65e-13. We also measured the contribution of each gene to the overall NAL across all the patients. Figure 2C shows the top 50 most frequently mutated genes that gave rise to candidate neoantigens. For instance, MTAP gene, a methylthioadenosine phosphorylase, known to be deficient in some tumors (20), contributed with 21 neoantigens (Figure 2C) across two patients (Supplementary Figure S5). PRKDC gene, encoding for a DNA-dependent protein kinase catalytic subunit, is involved in DNA repair, the establishment of immune tolerance,

and genome stability, and thus, has the potential predictive biomarker for ICI therapy (21). We detected 13 neoantigens across two patients for PRKDC gene (Figure 2C and Supplementary Figure S5). No shared neoantigens were found between the different patients, which aligns with other studies (22) and was expected here due to the highly heterogeneous mutational landscape across the sarcoma subtypes.

We next analyzed the distribution of the ranked AP scores of the top ten neoantigen candidates for each of the baseline samples. The patients were pooled according to their clinical response, consisting of three responders, 18 patients with progressive disease (PD), and 9 patients with stable disease (SD). The distribution of AP scores, and the means of the populations of each group, were then compared using Welch's test (Figure 3A). A marginally significant difference in the neoantigen scores between the responders and PD groups was observed (p-value = 0.078). There was no significant difference between the responders and SD groups (p-value = 0.63), but the comparison between the SD vs PD groups emerged with a significant difference (p-value = 0.001). Additionally, also using Welch's test, we

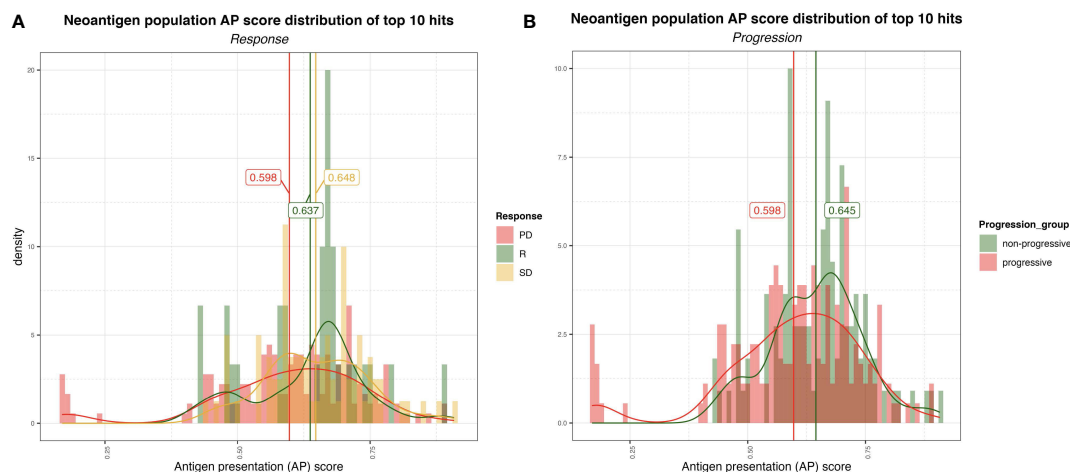


FIGURE 3

(A) Distribution of the AP scores for the top 10 (ranked by AP score) neoantigen candidates for progressive disease (PD), stable disease (SD) and responder (R) group. The vertical lines represent the mean AP score for each group. (B) Distribution of the AP scores for the top ten (ranked by AP score) neoantigen candidates for non-progressive (SD+R) and progressive (PD) groups.

compared the PD group against the remaining non-PD groups (*i.e.*, SD and responders pooled together), whereby a significant difference emerged from the analysis (p -value = 0.001) (Figure 3B). Interestingly, this analysis revealed, on average, a higher neoantigen quality for patients without PD.

The limited number of patient samples available for analysis post ICI therapy hindered our capacity to derive significant conclusions. For the few samples that were available post therapy; none responded to ICI. In addition, removal by immunoeediting of somatic mutations or immunogenic neoantigens caused by ICI therapy was not detectable among the few available ICI-treated samples (Supplementary Figure S6).

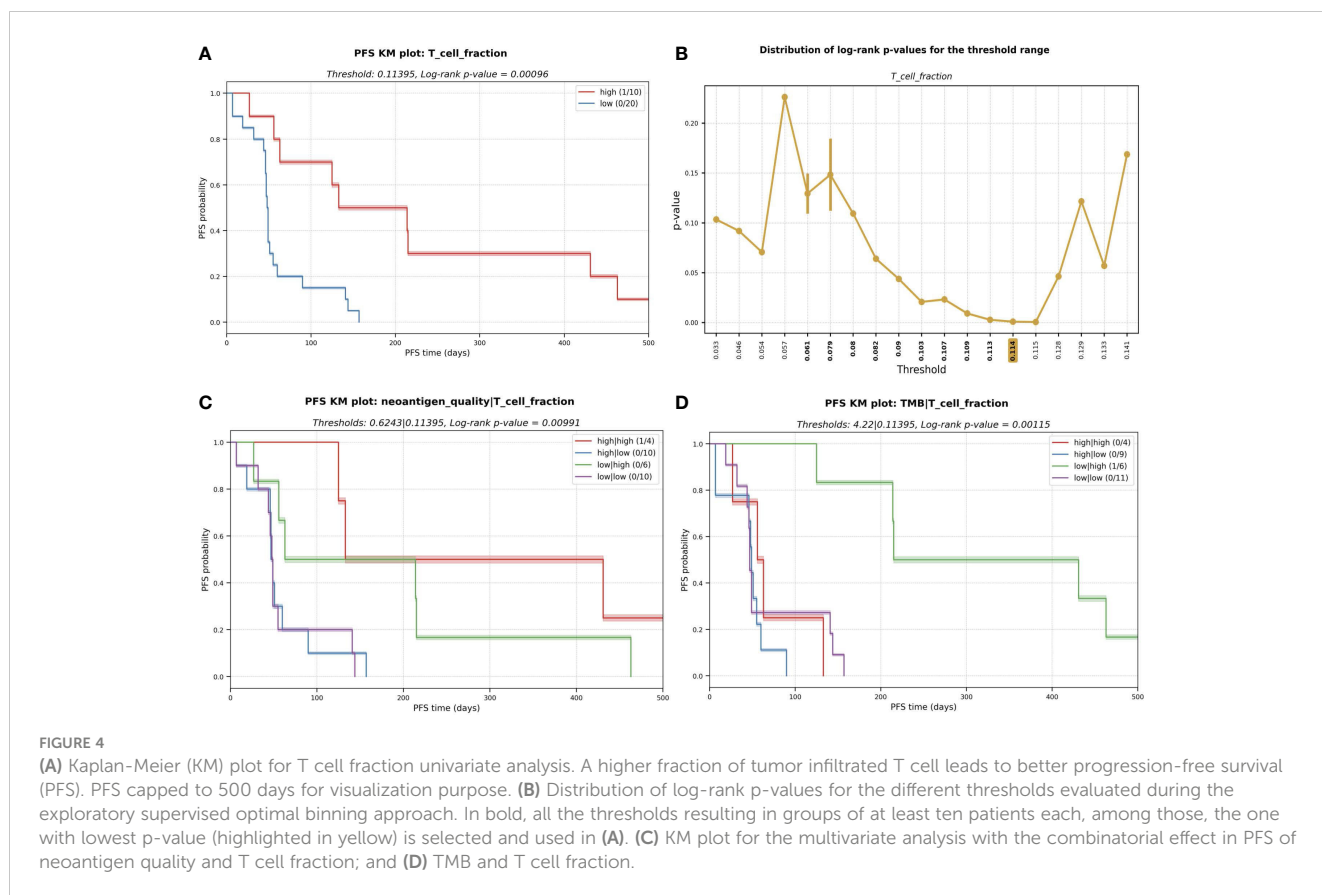
Survival and clinical response of TMB, NAL and immune cell TME in ICI treated sarcomas

We next evaluated the effect of various immunogenomic features on PFS of the patients using Kaplan-Meier (KM) survival analysis. Conventional thresholds such as mean/median or Maximally Selected Rank Statistics failed to yield significant results in most cases probably due to the considerable diversity of the dataset consisting of seven sarcoma subtypes and only 31 samples. Hence, we used a supervised optimal binning approach (see Methods) to group patients into two (low and high) groups for univariate analyses (*i.e.*, each immunogenomic feature was analyzed individually); for bivariate analysis, the univariate thresholds were used to stratify patients into four groups (*i.e.*, the immunogenomic features were analyzed in pairs of low/low, low/high, high/low, high/high). We note that the small sample size available in this study may limit the robustness of the log-rank test associated with the KM analysis. We also did not account for multiple hypothesis testing in this analysis. Consequently, we do not make strong claims of statistical significance in this preliminary study. Nevertheless, this exploratory analysis does reveal features associated with

differences in PFS among patient groups and which could guide further avenues for future studies.

We first performed a statistical interrogation of the immune cell infiltration into the TME in a univariate analysis (where a supervised optimal binning approach was utilized to independently assess the effect of the considered immunogenomic features in the survival and determine the optimal threshold cutoffs, see Methods). We identified a signal that an elevated fraction of infiltrated T cells into the TME as measured by TcellExTRECT tool (see Methods), led to improved PFS with a log-rank p -value in the KM analysis of 0.00096 (Figure 4A). The group of patients presenting a higher proportion of infiltrated T cells had a median PFS of 173 days while the lower group had a median of 48 days. Figure 4B demonstrates the suitability of the supervised optimal binning approach to find the best threshold to generate the different groups evaluated in KM survival analysis. A window of low p -values across different thresholds indicates a more robust separation of low and high groups which is not sensitive to the exact chosen threshold. An improved PFS was also observed to be associated with increased levels of macrophage M1 and M2 cell infiltration (log-rank p -values 0.047 and 0.019 for M1 and M2 cells respectively, Supplementary Figures S7A, S7B) and associated with B cells with a log-rank p -value of 0.074 (Supplementary Figures S7C).

To determine the predictive potential of the effect of combining certain immunogenomic properties for the prediction of PFS, we performed multivariate analyses. We observed that a higher neoantigen quality (see Methods) combined with a high infiltrated T cell fraction was associated with an improved PFS (median of 282 days) with a log-rank p -value of 0.006 (Figure 4C). High T cell infiltration with low neoantigen quality was still associated with good PFS (median PFS = 139 days). However, this PFS was lower than the median PFS among all patients with a high T cell fraction (173 days); this highlights that incorporating neoantigen quality with T cell fraction improves prognostic patient groupings compared to using T cell infiltration alone. Low T cell fraction was associated with poor PFS, regardless of the neoantigen



quality being high (median PFS = 49 days) or low (median PFS = 48 days; [Figure 4C](#)). The same association with improved PFS was not observed with TMB and T cell fraction ([Figure 4D](#)).

Investigation of immune escape parameters: antigen presentation machinery, personalized HLA-typing, and the tumor-specific HLA status

The polymorphic nature of HLA alleles and its association to tumor-immune escape called for accurately typing and evaluating the mutation and expression status of each HLA allele in the patients ([23](#)). The status of the HLA locus in the different sarcoma subtypes was evaluated using NeoOncoHLA ([24](#)) (see Methods). Using this personalized HLA-typing approach, the somatic mutations and tumor-specific expression of each patient-specific HLA allele were described. A total of 18 somatic variants affecting HLA class I alleles were detected among 14 patients (45% of the patients presented at least one somatic mutation affecting one HLA allele). Seven of those 18 were non-synonymous variants, and six of those seven affected the peptide binding regions of the alleles. [Figure 5](#) depicts the expression of each HLA allele estimated as TPM and the NAL associated with each sample's HLA-A and HLA-B alleles. The number of good neoantigen candidates for the HLA-A*02:01 allele in p9 increased from nine at baseline to 24 in week-8, while its expression (measured as TPM) decreased from 861 to 350. This could indicate a possible

immune escape mechanism in this patient, through the downregulation of this HLA allele's expression (see [Figure 5](#)). Although p31 showed numerous good neoantigen candidates, the expression of the HLA alleles seemed to be slightly downregulated, particularly for HLA-A*11:01. We did not consistently observe this putative immune escape pattern through HLA downregulation as a global trend across all the patients, and we had too few treated samples to draw conclusions. HLA expression did not demonstrate statistical significance in the survival study. Nonetheless, in the multivariate KM survival analysis, we found a significant separation for HLA-C (log-rank p-value = 0.001) expression in conjunction with the predicted NK cell fraction. Patients with a high NK cell infiltration combined with low HLA-C expression had the best PFS (median PFS = 157 days), while a high NK and HLA-C values resulted in short survival (median PFS = 50 days) ([Figure 6A](#)). This trend was not observed when combining HLA expression with T cell infiltration into the TME, where the high HLA expression and high T cell infiltration group resulted in similar PFS times as low HLA expression and high T cell infiltration group (see [Figures 6B-D](#)). This pattern could theoretically be owed to the combined benefit of enhanced NK cell activity due to the decreased HLA-C expression and the increased presence of T cells in the TME modulating tumor cell killing through HLA-A and/or -B antigen presentation to T cells; in the backdrop of improved PFS with high neoantigen quality with high T cell fraction reported above (see [Figure 4C](#)).

Importantly, the expression profile of genes linked with the antigen presentation machinery (APM), in addition to HLA

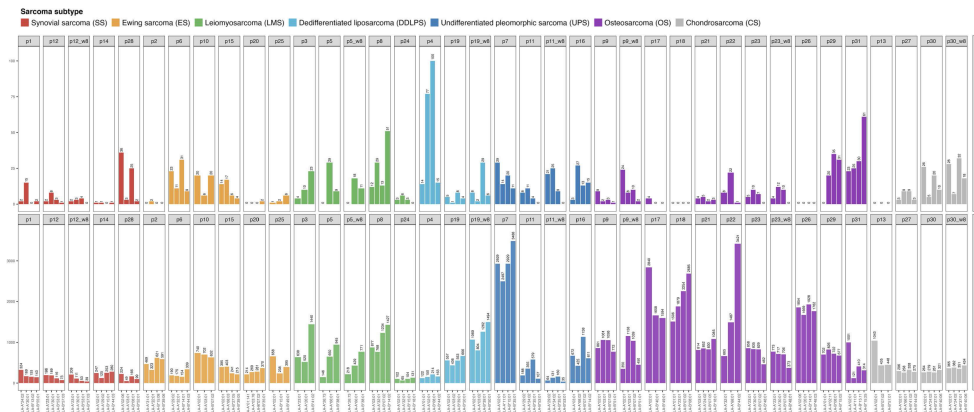


FIGURE 5 HLA-A and B allele-specific expression calculated as TPM across the different ICI-treated sarcoma samples (bottom histogram) and the correspondent NAL for each allele (top histogram).

expression, revealed probable patterns of immune evasion in sarcoma. Among the APM genes we profiled (see Methods), we found that a decreased expression of beta-2-microglobulin (B2M), MHC class II transactivator CIITA, endoplasmic reticulum aminopeptidase 2 (ERAP2), transporter 2 (TAP2) and TAP binding protein like (TAPBPL) were significantly associated with a shorter PFS (Figure 7). Consistent with previous research, these findings underline the utility of APM profiling as a potential biomarker of tumor cell antigen presentation status and immune escape (25–27).

Discussion

The efficacy of pembrolizumab in patients with advanced bone and soft-tissue sarcoma (STS) was assessed in the SARC028 trial (NCT02301039) and demonstrated that only some of the many different histological subclasses of sarcomas had a positive clinical response to ICI (28). In general, the results were modest, with promising results within some of the STS subtypes (6). Many studies have reported the importance of immune cell infiltration into the TME and their complex interplay with tumor cells in the

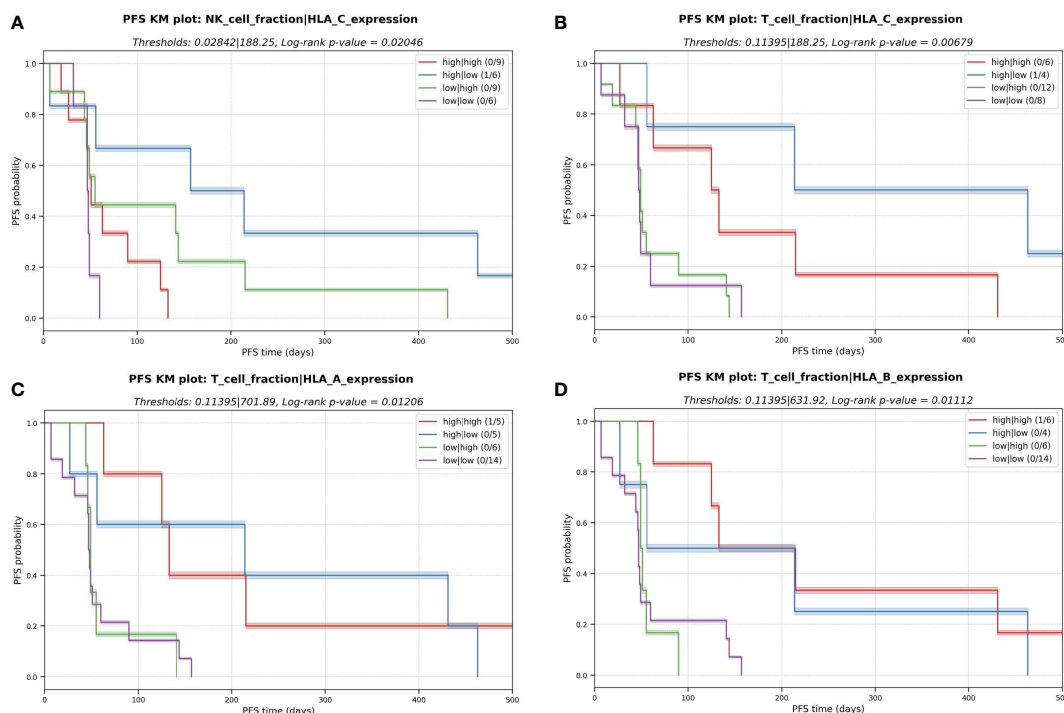


FIGURE 6 Kaplan-Meier (KM) plots for the multivariate analysis with the combinatorial effect in PFS of HLA expression and immune cell fractions in the TME. (A) NK cell fraction and expression of HLA-C. (B) T cell fraction and expression of HLA-C. (C) T cell fraction and expression of HLA-A (D) T cell fraction and expression of HLA-B.

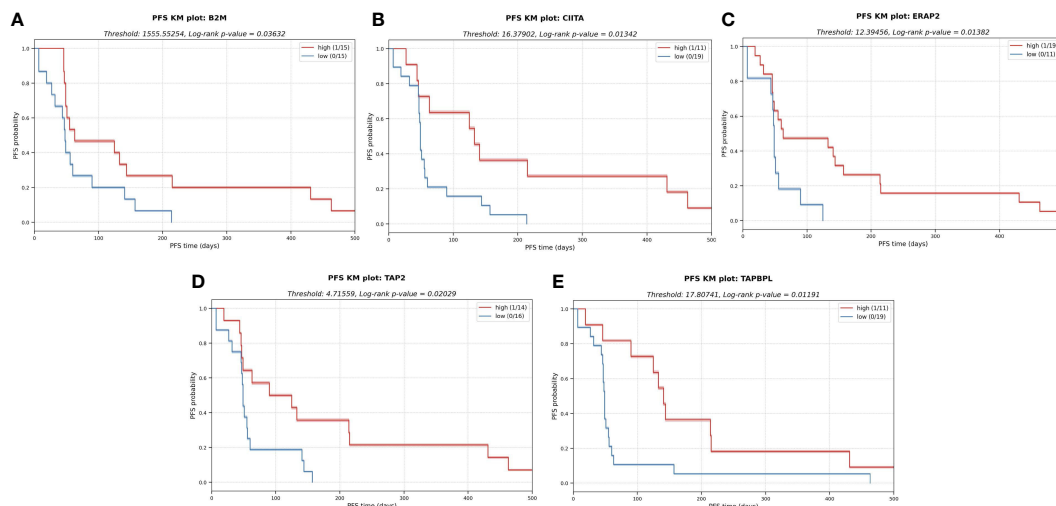


FIGURE 7

KM plots for several APM components analyzed in a univariate analysis. The downregulation of these components in the tumor cells might impair antigen presentation and subsequent escape from immune system. (A) Beta-2-microglobulin (B2M). (B) MHC class II transactivator CIITA. (C) Endoplasmic reticulum aminopeptidase 2 (ERAP2) (D) Transporter 2 (TAP2) (E) TAP binding protein like (TAPBPL).

clinical outcome of cancer patients (29–31). Furthermore, in recent years, neoantigens have emerged as important immunotherapy targets that play a central role in the HLA-restricted T-cell response and have been linked to the clinical efficiency of ICI therapy in several cancer indications (32, 33). Moderate responses have been found in clinical trials of ICI in sarcomas, and only a subset of patients has benefited with durable clinical outcomes (28). Due to this modest response to ICI therapy, a comprehensive immunogenomic analysis of sarcomas guided by state-of-the-art AI tools to predict neoantigens and their interplay with respect to immune cells in the TME was warranted in this study to help elucidate some of the mechanisms of possible immune escape and resistance of sarcomas to ICI therapy and to enable prior identification of patients most likely to respond.

We used a series of computational and AI methods to investigate several tumor immunogenomic properties, including the somatic mutational landscape, neoantigens, expression of key immune-related genes (e.g., APM genes), and the TME immune cell composition from the available SARCO28 trial samples. The analysis was conducted on NGS data from 31 sarcoma patients, comprising WES from matched tumor and blood samples and tumor RNA-seq data. In terms of ICI response, three patients had a clinical benefit from pembrolizumab (one complete responder and two partial responders), nine had stable disease (SD), 19 had progressive disease (PD), and one did not have response information available.

Using the RNA-seq data from the tumor samples, a DE analysis revealed that several immune-related GO terms were significantly enriched in the responders, suggesting that an immunologically active TME may lead to an improved clinical response to ICI therapy. A further interrogation revealed three distinct TME immune profiles based on a hierarchical clustering of the tumor transcriptome data; (1) an immune active cluster (cluster 3), formed mostly by OS, and enriched in macrophages and CD8+ expression,

(2) a cluster enriched in monocytes and with patients with SD response (cluster 4), and (3) a slightly immune desert cluster consisting of all ES patients except one (cluster 5). Importantly, two of the three ICI responders, including the patient experiencing a complete clinical response, were included in the TME immune active group (cluster 3). This was consistent with the DE analysis which showed that responders were associated with a pre-existing immune activity in the TME. The RNA-seq data was also used to deconvolute the immune cell composition in the TME, and then applied to a univariate KM survival analysis to identify immune cell infiltration patterns associated with improved PFS. We observed that higher fractions of T cells and macrophages were associated with longer PFS. These observations were consistent with previous studies that examined the importance of the immune cell composition of the TME for ICI response in sarcomas (34), although contradictory results have been reported in the literature. For example, B cell signatures correlate with longer survival times (17), whereas the converse has been reported for CD8+ T cells (35–38). Additionally, our observation of elevated macrophages in OS was consistent with previous studies (39). Similarly, using RNA-seq data, we found that a lower expression of APM-related genes, including B2M, CIITA, ERAP2, TAP2 and TAPBPL, was also associated with a shorter PFS, reflecting the increasing interest in APM biomarkers for clinical outcome and immune escape in cancer (40). Finally, RNA-seq data revealed a pattern whereby a decreased allele-specific HLA-C expression, combined with high NK cell infiltration, was linked with improved PFS. This is consistent with the well-established trend of downregulation of class I HLA allele expression correlating with NK cell activity (41).

The somatic mutation profile of the sarcomas was highly variable both within and between histological sarcoma subtypes. OS samples had the highest TMB score, followed closely by UPS, with CS having the lowest and least variable TMB. While many

genes harbored mutations across the different patients, the top four most commonly mutated genes were all members of the mucin (MUC) gene family, known to play a role for epithelial tissues and reported in some neoplastic lesions (42). However, this trend may be explainable by the high degree of exon repeats in the MUC family of genes among individuals (43). In addition to small variants, copy-number variations and chromosomal rearrangements were found in many patients, including the known ES driver gene fusions of EWSR1 and SS18 genes. An interesting observation here was the high number of fusion-generated neoantigens in DDLPS compared to the other subtypes. Due to the complex karyotypes of DDLPS, it would be expected to contain many genome rearrangements, but so would OS and UPS, which did not show this pattern. This observed difference is challenging to explain but could contribute to the better response seen in patients with DDLPS compared to OS but does not explain the better response of UPS cases (30).

Increased neoantigen load (NAL) has previously been positively correlated with TMB (44, 45), and therefore described to correlate with ICI therapy response (46–48) in certain cancer indications. In this study, however, we did not find a correlation between TMB and NAL. This finding reinforces the argument that higher TMB does not necessarily always equate to higher NAL. It is important to note that the conventional TMB calculation does not consider gene fusions and therefore bypasses powerful sources of neoantigens. Additionally, the conventional TMB calculation does not consider some key determinants of antigen presentation or immunogenicity, such as antigen processing, HLA binding and the expression of the somatically mutated peptides, in addition to their distance from self (that is, the “wild type” protein).

In our univariate analysis, neither NAL nor neoantigen quality was associated with improved PFS. However, when combining neoantigen quality with the presence of T cells, we observed a striking joint behavior associated with longer PFS compared to the presence of T cells alone. When the T cell fraction was low, PFS was always low (median PFS = 48 days) regardless of neoantigen quality. This was consistent with the finding that immune T cell desert TMEs are not associated with good outcomes for ICI therapy (49). Additionally, it is reasonable that high-quality neoantigens are not effective if there are no T cells in the surrounding environment to recognize them. In the case of a high T cell count but low quality neoantigens, PFS was modestly improved (median PFS = 139 days) compared to the low T cell count patients. Remarkably, patients with both high CD8+ T cell count and high-quality neoantigens had improved PFS (median PFS=282 days). The observed improved PFS when combining the tumor infiltrated CD8+ T cell fraction and neoantigen quality was not replicated when combining T cell fraction and TMB. In this case, a good PFS was observed only for low TMB with high T cell fraction, suggesting that the quality and not quantity of neoantigens might be more relevant in certain settings for clinical benefit (50). Overall, the comparison of NAL vs TMB in the context of T cell infiltration indicated that the AI neoantigen prediction platform used to identify neoantigens is reliably predicting mutated peptides presented on the tumor cell surface that are potentially immunogenic. For B cells, a similar picture emerged, except it was required for both the neoantigen AP scores and the B cell fraction in the TME to be elevated for longer

PFS. This finding is reflective of the new landscape emerging on the importance of B cell tumor infiltration, prognosis, and response to immunotherapy (17). Overall, the findings of the KM survival analysis of this study were consistent with the TME clustering which found that responders were associated with a pre-treatment, immune-inflamed TME.

The KM survival analysis based on the supervised optimal binning approach we used to arrive at the insights described in this study has several limitations and caveats. First, each set of thresholds could be considered as a hypothesis in the sense of statistical testing; appropriate multiple test corrections, such as Benjamini-Hochberg would ideally be applied if the p-value was to be interpreted for true statistical significance. In this exploratory work on a limited number of samples, though, we simply aimed to identify the best bins for the data and did not intend to make broad robust statistical claims. Thus, we believe the approach is still justified as it provides informative insights from a small patient cohort, but we acknowledge the need to validate the results in a larger patient cohort. Secondly, some choices of thresholds would have resulted in very small groups; the assumptions of many statistical tests do not hold in such cases. In this work, we limited this problem by only considering univariate thresholds which result in at least ten individuals in all groups.

In addition, it is important to note that there are several properties not considered in this study, due to the experimental data not being available for the SARC028 samples, which may also be important to help our understanding and prediction of the response to cancer immunotherapy. For example, there is increasing evidence that the gut microbiota (and related metabolites such as butyrate and cholic acid), can influence the modulation of CD8+ T cell activity and immune cell infiltration into the TME (51, 52).

In conclusion, to the best of our knowledge, this is the first study to exhaustively profile the immune cell TME of sarcomas with its interplay with immunogenic neoantigens under the context of ICI therapy, in a manner that uses advanced computational AI tools to comprehensively capture this important interplay. While the sample size for this study was small, the insights gained were suggestive that the interplay between neoantigens and immune cell infiltration patterns into the TME is a key prognostic marker of clinical response to ICI and PFS. This therefore warrants further clinical and biomarker studies with larger sarcoma cohorts.

Material and methods

SARC028 trial cohort data description

The SARC028 trial (NCT02301039) recruited 86 patients. Of those, 80 patients (40 with advanced STS and 40 with bone sarcomas) were eligible for the ICI therapy (pembrolizumab, an anti-PD-1 antibody) (6). The eligibility criteria for the patients included: (a) underwent at least one previous systemic therapy; (b) metastatic STS or bone sarcoma diagnosis histologically confirmed by pathological expert in accordance with the WHO Classification of Tumors and Soft Tissue and Bone; (c) had at least one measurable lesion by RECIST 1.1 and

one biopsy accessible lesion; (d) 12 years or older; (d) at least 12 weeks of life expectancy (6). STS group included the following subtypes: undifferentiated pleomorphic sarcoma (UPS), dedifferentiated liposarcoma (DDLPS), synovial sarcoma (SS) and leiomyosarcoma (LMS). The bone sarcoma group included: osteosarcoma (OS), Ewing sarcoma (ES) and dedifferentiated chondrosarcoma (CS). All the patients received intravenously a dose of 200 mg of pembrolizumab every three weeks. According to the protocol blood and tumor samples were to be collected before pembrolizumab treatment (receiving the name of “baseline” samples) and 8 weeks after the start of treatment (receiving the name of “week-8” samples). For this specific study, there were unfortunately only seven patient samples available for analysis post ICI therapy. All the research and ethical approvals and permits together with the written informed consents from all the patients were obtained prior to sample collection. The 31-sarcoma patient cohort consisted of 13 patients with STS, including four synovial sarcomas (SS), four leiomyosarcomas (LMS), two dedifferentiated liposarcomas (DDLPS), three undifferentiated pleomorphic sarcomas (UPS); and 18 patients with bone sarcomas, including six Ewing sarcomas (ES), nine osteosarcomas (OS) and three dedifferentiated chondrosarcomas.

The import and analysis of the available samples, 31 baseline and seven week-8 (Supplementary Table 3), to Norway was approved by the Committee for Medical Ethics in Southeastern Norway #17866. All available samples from participating centers were collected and DNA and/or RNA were purified by the trial organization and shipped to Oslo University Hospital for sequencing analysis. The data were stored and analyzed by NEC OncoImmunity in the computing infrastructure specially designed for high-level protection of sensitive personal data at the University of Oslo (53).

Whole exome sequencing on sarcoma tumor tissue and matched PBMCs and whole transcriptome sequencing on sarcoma tumor tissue

Whole exome libraries were prepared at the Oslo University Hospital Genomics Core Facility from 100 ng of genomic DNA using the Twist Core Exome enrichment system (Twist Bioscience) following the manufacturer’s protocol. RNA sequencing libraries were constructed using the KAPA RNA Hyper kit to generate a total RNA library, which was further captured using the Twist Core Exome probe set. Variable input amounts of RNA were used depending on the availability of material (from 20-100 ng). Exomes and RNA libraries were sequenced paired-end 2x75bp using the Illumina HiSeq 4000 instrument, and FASTQ files were generated using Illumina’s bcl2fastq conversion software.

Screening of mutational landscape of sarcoma through a comprehensive variant calling approach

The screening and characterization of single nucleotide variants (SNVs) and small insertions and deletion (indels) sculpturing the

tumor genome was conducted using NeoMutate (18), a tool previously published by our group yielding very high validation rates (18). After an initial thorough inspection of the raw sequencing data, including quality control and adapter trimming, the high quality paired-end reads were mapped to the human genome (GRCh38) using BWA-MEM (54)(v0.7.17-r1188). The output BAM files were treated according to the genome analysis toolkit (GATK, v4.0.6.0) best practices (55) (including PCR duplicate marking and realignment optimization). NeoMutate incorporates an ensemble of six independent state-of-the-art somatic variant calling tools enabling the capture of the full mutational profile from tumor-normal matched WES data. Only those variant candidates detected with high confidence at least by two out of the six tools were kept avoiding false positive mutation calls. The variants were additionally filtered according to different quality thresholds (including minimum read depth (DP) of 10 reads for both tumor and normal data, more than three reads supporting the mutation in the tumor sample at the variant position). Ensembl Variant Effect Predictor (VEP) (56) toolkit was used to annotate the functional effect of the detected variants on the resulting gene product. VEP was also exploited to identify the non-synonymous variants, in other words, those mutations altering the amino acid sequence of the tumor proteome, underpinning the neoantigen landscape of tumors (see Methods section “Characterization of sarcomas immunogenicity through neoantigen prediction”). Importantly, the expression of the somatic variations was evaluated using RNA-seq data, and only the expressed somatic variants were retained for the neoantigen prediction step. This is because the altered peptides need to be expressed by the tumor to result in the production of a neoantigen.

In addition to somatic variant identification, GATK-HaplotypeCaller (v4.0.6) (57) was used for germline variant identification, and VEP (56) for the variant annotation. Importantly, the combined effect of proximal (nearby) variants (either germline or somatic) altering the same protein, and therefore, the same neoantigenic peptide, was evaluated. Haplotype phasing is the bioinformatics process of statistical estimation of haplotypes from genotype data. WhatsHap (58) (v0.17) was used giving as input both tumor WES and RNA-seq data to assess the phase relationship between proximal variants, in other words, to evaluate whether nearby variants were affecting the same allele in the same tumor subclone. After the phasing, HaploSaurus (59) (included in VEP package), was called to assess the joined functional impact of the phased variants and fully reconstruct the mutated protein sequence. Phasing step ensures that the selected neoantigen repertoire arising from the mutated proteins correctly represents the patient’s genome, increasing the chances of anti-tumor response to immunotherapy.

Characterization of sarcoma gene fusions

It is widely recognized that large chromosomal rearrangements and fusion genes play a critical role in underpinning and driving the sarcomagenesis course in specific morphological subtypes, making them a valuable diagnostic marker (60, 61). The accurate identification of sarcoma fusion genes helps to understand its

pathogenesis and the development of specific treatment strategies against the targetable fusions. In addition, gene fusions represent an incredibly valuable source of potentially immunogenic neoantigens that can mediate the anticancer immune response to ICI, even in those tumors with low TMB (62). We predicted the gene fusions from RNA-seq data using Arriba (63), the winner method for gene fusion detection in the ICGC-TCGA DREAM Somatic Mutation Calling-RNA Challenge. Arriba (63), developed for clinical research setting, is based in the ultrafast STAR aligner, and it computes a confidence score (low, medium or high) reflecting the likelihood of the fusion being generated due to an underlying genomic rearrangement specific to the tumor and not due to a sequencing artifact. Low confidence gene fusions were discarded from downstream analysis.

Tumor mutational burden calculation

TMB was defined as the number of non-synonymous somatic SNV and indels with a VAF of at least %5 per megabase in the coding area of the cancer genome, as recommended by the guidelines of the Friends of Cancer Research TMB Harmonization Project (64). Fusions, CNVs, non-coding and synonymous mutations were discarded for TMB calculation.

Human leukocyte antigen typing

HLA alleles of each patient were inferred *in silico* using OncoHLA (65) providing peripheral blood mononuclear cell (PBMC) WES data as input. OncoHLA uses an integer linear programming algorithm together with prior probabilities of the allelic ethnic frequencies to determine the closest-matched HLA allele from the IPD-IMGT/HLA Database (66) (v.3.41.2). The output includes the HLA types for both class I and class II up to four field of resolution and the associated HLA gene, transcripts, and protein sequences.

HLA expression quantification and HLA somatic variant screening in sarcoma samples

It is well established that cancer cells can exploit several HLA-associated immune evasion mechanisms to hijack the immune system (67, 68). A comprehensive scrutiny of the HLA status in the tumor was conducted using a previously developed method by our group (24). Using the typed HLA alleles, an exhaustive profiling of the somatic mutations affecting each individual HLA allele was carried out, and their functional impact in the corresponding HLA protein sequences was annotated. In addition, the expression (abundance), reported as transcripts per million mapped reads (TPM), of each allele was quantified by mapping the RNA-seq reads to the inferred HLA sequences, to evaluate whether any allele was downregulated – a well-known immune escape mechanism in tumor development.

Isoform and gene-level expression quantification using RNA-seq data from sarcoma samples

Gene isoform expression profiling from the RNA-seq data of both STS and bone sarcoma was carried out utilizing Kallisto (69) (v0.43.1), based on pseudoalignments of the reads and expectation-maximization (EM) algorithm to conduct isoform-level expression quantification. The reference transcriptome for GRCh38 genome, required as input, was obtained from Ensembl database version 95. Kallisto reports the abundance transcript level measured TPM.

The expression values of each transcript were used for several analyses that will be further detailed in the following sections, including: (1) Calculation of the abundance of the potential neoantigens generated from the proteins affected by non-synonymous variants in the tumor; (2) Differential expression and enrichment analysis; (3) TME profiling; (4) Manual inspection of immune-related genes expression.

Characterization of sarcomas immunogenicity through neoantigen prediction analysis

The immunogenicity of the mutated peptides derived from tumor-specific alterations was assessed using NEC Immune Profiler (NIP) modular neoantigen pipeline developed by NEC OncoImmunity, comprising several proprietary T cell epitope machine-learning (ML) prediction algorithms. Neoantigen predictions for HLA-A and -B were conducted for each patient with peptides of lengths 9 and 10. Due to the lack of HLA-C validated data influencing the accuracy of the ML models, HLA-C was not evaluated. The pipeline considers several features determining the immunogenicity of a neoantigen, including:

1. The binding affinity of the peptide for the MHC/HLA molecule. NIP exploits three distinct binding affinity ML predictors that compute IC50 (nM) scores for each mutated peptide. The lower the IC50 score, the stronger the binding of the peptide to the specific HLA molecule.
2. The peptide's ability to be efficiently digested by the antigen processing machinery (APM). An ensemble of 13 Support Vector Machines (SVM) included in NIP and trained on validated mass spectrometry immunopeptidome datasets determine which peptides have the optimal features to be efficiently processed by the APM, which include cleavage probability by the proteasome and antigen processing transport (TAP) efficiency.
3. The expression level of the mutated peptide. The expression of each neoantigen candidate was computed by summing the expression values (TPMs) of all the isoforms coding for the specific peptides under consideration, which is critical for the accurate prediction of immunogenic neoantigens. To determine the specific abundance of the mutated peptide, the sum of the expression levels of all the

isoforms containing the peptide was adjusted according to the variant allele frequency (VAF) computed at RNA level.

- The relative uniqueness of the candidate neoantigens compared to the normal peptides of healthy tissue was evaluated to avoid cross-reaction with self-antigen sequences.

The result is summarized in a single score ranging from 0 to 1, known as the antigen presentation score (AP), which indicates the T cell recognition probability, ranging from 0 to 1, with 1 signifying the highest likelihood that the given neoantigen is immunogenic. Neoantigen load (NAL) was computed by calculating the number of neoantigens with an AP higher than 0.5. Neoantigen quality in this study was defined as the mean AP score for the top ten neoantigen candidates ranked by AP score.

Tumor microenvironment profiling

QuantIseq (70) deconvolution algorithm was utilized to analyze the immune cell composition of each sarcoma sample. QuantIseq has demonstrated a robust overall performance (71) and is one of the very few methods generating an absolute score that can be interpreted as a cell fraction, which allows both inter- and intra-sample comparisons. It takes as input RNA-seq reads and quantifies *via* deconvolution of cell fractions based on constrained least squares regression the proportion of ten different immune cell phenotypes, including B cells, M1 and M2 macrophages, CD8+, CD4+ and regulatory T cells, natural killer cells (NK), among others.

In addition to QuantIseq, TcellExTRACT (72) R package was applied to estimate infiltrated T cell fraction. The tool employs WES data and makes use of a signal based on somatic copy number from V(D)J recombination to directly quantify the proportion of T cells.

Further, the ESTIMATE algorithm was applied to calculate the stromal and immune scores for each sarcoma sample using the normalized gene expression values as input (73).

Selection of immune-related genes

We conducted an exhaustive literature research and compiled a list of 282 genes known to be related with immune system interaction and response (74–76). The full immune-related gene list is provided in [Supplementary Table 4](#).

Differential expression analysis and enrichment analysis

In order to characterize the differentially expressed genes (DEG) between week-8 and baseline samples, the DESeq2 (77) (v1.30.1) R package was selected. “Tximport” (78) (1.18.0) R package was utilized to import the transcript-level expression estimates

generated by Kallisto and produce gene-level count matrices and normalizing offsets, as required by DESeq2 (77). DESeq2 (77) models the counts using normalization factors to account for differences in the library depth, estimates the gene-wise dispersions and uses shrinkage of effect size to remove the low count genes. Then, it fits a negative binomial model and performs Wald test hypothesis testing. DEGs were obtained by applying a filter of $p\text{-value} < 0.05$.

The results of the DE analysis were expanded by conducting an enrichment analysis using an overrepresentation analysis (ORA) to associate the expression data with specific biological processes (BP). The R package called “goseq” (79) (1.42.0) conducted ORA of Gene Ontology (GO) terms, which corrects the results based on gene length bias of DEGs. DEGs were first separated into up and down-regulated genes attending to their fold change (FC) (greater than 1 or smaller than -1, for up and down-regulated genes, respectively), and then goseq (79) was used to detect overrepresented up and down-regulated BPs.

Statistical analysis

The statistical analyses in this study were conducted using R (4.0.3) and python (3.8) programming languages. The criteria for the annotation of the different determinants of each patient to ICI therapy, such as progression-free survival (PFS) and overall survival (OS), has been previously described in (6). Survival curves were plotted using the Kaplan–Meier (KM) functionality within “lifelines” python library to compare PFS for a different set of individual covariates (univariate analysis) and in conjunction (multivariate analysis). Differences in median PFS were assessed using the log-rank test and multivariate long-rank test.

Supervised optimal binning

Kaplan–Meier (KM) survival analysis entails splitting individuals into two or more groups and comparing survival times between individuals in the groups. For a numeric covariate, a threshold is required to perform such a split; that is, we must select a threshold to bin individuals into groups. Prior work (80) has shown that for such binning problems, the only thresholds which change group composition are the values that actually occur in the dataset. For example, when splitting individuals into groups based on TMB, the only TMB thresholds which change to which group individuals are assigned are the actual TMB values of the individuals. Thus, we determined an “optimal” threshold with respect to KM analysis by evaluating the log-rank test p -value for all possible binning of individuals in the dataset; we implemented this by simply trying each observed value for the numeric covariate of interest in the patient dataset. Only threshold cutoffs generating bins with at least 10 individuals each were evaluated.

Data availability statement

All the datasets presented in this study are available through the SARC clinical trial data repository: <https://sarctrials.org/clinical-trials/sarc028/>. The data is not readily available for direct download from SARC, as it is classed as patient sensitive genetic information. However, please contact SARC at sarc028@sarctrials.org and instructions on downloading the data will be subsequently provided. The accession numbers for each of the samples used in this study can be found within the article supplementary materials (Supplementary Table 5).

Ethics statement

The studies involving human participants were reviewed and approved by Regional Committees for Medical Research Ethics South East Norway. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Conceptualization, TC, OM, IA. Methodology, IA and BM. Data analysis, IA, BM, PS, BS and HF. Data investigation, LM-Z, RS, EK, MB, HT, OM, TC. Draft manuscript preparation, IA and TC. Manuscript review and editing, all authors. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grant MISP # 53061 from Merck and from the Norwegian Research Council.

References

- Kallen ME, Hornick JL. The 2020 WHO classification: what's new in soft tissue tumor pathology? *Am J Surg Pathol* (2021) 45:e1–23. doi: 10.1097/PAS.0000000000001552
- Burningham Z, Hashibe M, Spector L, Schiffman JD. The epidemiology of sarcoma. *Clin Sarcoma Res* (2012) 2:14. doi: 10.1186/2045-3329-2-14
- Anderson WJ, Doyle LA. Updates from the 2020 World Health Organization classification of soft tissue and bone tumours. *Histopathology* (2021) 78:644–57. doi: 10.1111/his.14265
- Koumariou A, Duran-Moreno J. The sarcoma immune landscape: emerging challenges, prognostic significance and prospective impact for immunotherapy approaches. *Cancers (Basel)* (2021) 13(3):363. doi: 10.3390/cancers13030363
- Rytlewski J, Milhem MM, Monga V. Turning 'Cold' tumors 'Hot': immunotherapies in sarcoma. *Ann Transl Med* (2021) 9:1039. doi: 10.21037/atm-20-6041
- Tawbi HA, Burgess M, Bolejack V, Van Tine BA, Schuetz SM, Hu J, et al. Pembrolizumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. *Lancet Oncol* (2017) 18:1493–501. doi: 10.1016/S1470-2045(17)30624-1
- D'Angelo SP, Mahoney MR, Van Tine BA, Atkins J, Milhem MM, Jahagirdar BN, et al. Nivolumab with or without ipilimumab treatment for metastatic sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2 trials. *Lancet Oncol* (2018) 19:416–26. doi: 10.1016/S1470-2045(18)30006-8
- Ben-Ami E, Barysaukas CM, Solomon S, Tahlil K, Malley R, Hohos M, et al. Immunotherapy with single agent nivolumab for advanced leiomyosarcoma of the uterus: Results of a phase 2 study. *Cancer* (2017) 123:3285–90. doi: 10.1002/cncr.30738
- Somaiah N, Conley AP, Parra ER, Lin H, Amini B, Solis Soto L, et al. Durvalumab plus tremelimumab in advanced or metastatic soft tissue and bone sarcomas: a single-centre phase 2 trial. *Lancet Oncol* (2022) 23:1156–66. doi: 10.1016/S1470-2045(22)00392-8
- Wilky BA, Trucco MM, Subhawong TK, Florou V, Park W, Kwon D, et al. Axitinib plus pembrolizumab in patients with advanced sarcomas including alveolar soft-part sarcoma: a single-centre, single-arm, phase 2 trial. *Lancet Oncol* (2019) 20:837–48. doi: 10.1016/S1470-2045(19)30153-6
- Nacev BA, Sanchez-Vega F, Smith SA, Antonescu CR, Rosenbaum E, Shi H, et al. Clinical sequencing of soft tissue and bone sarcomas delineates diverse genomic landscapes and potential therapeutic targets. *Nat Commun* (2022) 13:3405. doi: 10.1038/s41467-022-30453-x
- van Oost S, Meijer DM, Kuijjer ML, Bovee J, de Miranda N. Linking immunity with genomics in sarcomas: is genomic complexity an immunogenic trigger? *Biomedicines* (2021) 9(8):1048. doi: 10.3390/biomedicines9081048
- Marcrom S, De Los Santos JF, Conry RM. Complete response of mediastinal clear cell sarcoma to pembrolizumab with radiotherapy. *Clin Sarcoma Res* (2017) 7:14. doi: 10.1186/s13569-017-0079-1

Acknowledgments

We are grateful for the services of the OUH Genomics Core Facility (oslo.genomics.no)¹ and the UiO Services for sensitive data², and for NN and MM in collecting, preparing, and shipping DNA and RNA samples.

Conflict of interest

IA, BM, PS, IV, BS, HF, RS, and TC were employed by the company NEC OncoImmunity.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1226445/full#supplementary-material>

¹ <https://genomics.no>.

² <https://www.uio.no/english/services/it/research/sensitive-data/about/index.html>.

14. Guram K, Nunez M, Einck J, Mell LK, Cohen E, Sanders PD, et al. Radiation therapy combined with checkpoint blockade immunotherapy for metastatic undifferentiated pleomorphic sarcoma of the maxillary sinus with a complete response. *Front Oncol* (2018) 8:435. doi: 10.3389/fonc.2018.00435
15. Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* (2019) 51:202–6. doi: 10.1038/s41588-018-0312-8
16. He M, Abro B, Kaushal M, Chen L, Chen T, Gondim M, et al. Tumor mutation burden and checkpoint immunotherapy markers in primary and metastatic synovial sarcoma. *Hum Pathol* (2020) 100:15–23. doi: 10.1016/j.humpath.2020.04.007
17. Petitprez F, de Reynies A, Keung EZ, Chen TW, Sun CM, Calderaro J, et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature* (2020) 577:556–60. doi: 10.1038/s41586-019-1906-8
18. Anzar I, Sverchkova A, Stratford R, Clancy T. NeoMutate: an ensemble machine learning framework for the prediction of somatic mutations in cancer. *BMC Med Genomics* (2019) 12:63. doi: 10.1186/s12920-019-0508-5
19. Malone B, Simovski B, Moline C, Cheng J, Gheorghe M, Fontenelle H, et al. Artificial intelligence predicts the immunogenic landscape of SARS-CoV-2 leading to universal blueprints for vaccine designs. *Sci Rep* (2020) 10:22375. doi: 10.1038/s41598-020-78758-5
20. Alhalabi O, Chen J, Zhang Y, Lu Y, Wang Q, Ramachandran S, et al. MTAP deficiency creates an exploitable target for antifolate therapy in 9p21-loss cancers. *Nat Commun* (2022) 13:1797. doi: 10.1038/s41467-022-29397-z
21. Chen Y, Li Y, Xiong J, Lan B, Wang X, Liu J, et al. Role of PRKDC in cancer initiation, progression, and treatment. *Cancer Cell Int* (2021) 21:563. doi: 10.1186/s12935-021-02229-8
22. Xie N, Shen G, Gao W, Huang Z, Huang C, Fu L. Neoantigens: promising targets for cancer therapy. *Signal Transduct Target Ther* (2023) 8:9. doi: 10.1038/s41392-022-01270-x
23. Dhatchinamoorthy K, Colbert JD, Rock KL. Cancer immune evasion through loss of MHC class I antigen presentation. *Front Immunol* (2021) 12:636568. doi: 10.3389/fimmu.2021.636568
24. Anzar I, Sverchkova A, Samarakoon P, Ellingsen EB, Gaudernack G, Stratford R, et al. Personalized HLA typing leads to the discovery of novel HLA alleles and tumor-specific HLA variants. *HLA* (2022) 99:313–27. doi: 10.1111/tan.14562
25. Zhao Y, Cao Y, Chen Y, Wu L, Hang H, Jiang C, et al. B2M gene expression shapes the immune landscape of lung adenocarcinoma and determines the response to immunotherapy. *Immunology* (2021) 164:507–23. doi: 10.1111/imm.13384
26. Henle AM, Nassar A, Puglisi-Knutson D, Youssef B, Knutson KL. Downregulation of TAP1 and TAP2 in early stage breast cancer. *PLoS One* (2017) 12:e0187323. doi: 10.1371/journal.pone.0187323
27. Leon Machado JA, Steimle V. The MHC class II transactivator CIITA: not (Quite) the odd-one-out anymore among NLR proteins. *Int J Mol Sci* (2021) 22. doi: 10.3390/ijms22031074
28. Eulo V, Van Tine BA. Immune checkpoint inhibitor resistance in soft tissue sarcoma. *Cancer Drug Resist* (2022) 5:328–38. doi: 10.20517/cdr.2021.127
29. Roulleaux Dugae M, Nassif EF, Italiano A, Bahleda R. Improving immunotherapy efficacy in soft-tissue sarcomas: a biomarker driven and histotype tailored review. *Front Immunol* (2021) 12:775761. doi: 10.3389/fimmu.2021.775761
30. Dyson KA, Stover BD, Grippin A, Mendez-Gomez HR, Lagmay J, Mitchell DA, et al. Emerging trends in immunotherapy for pediatric sarcomas. *J Hematol Oncol* (2019) 12:78. doi: 10.1186/s13045-019-0756-z
31. Zhu MMT, Shenasa E, Nielsen TO. Sarcomas: Immune biomarker expression and checkpoint inhibitor trials. *Cancer Treat Rev* (2020) 91:102115. doi: 10.1016/j.ctrv.2020.102115
32. Yi M, Qin S, Zhao W, Yu S, Chu Q, Wu K. The role of neoantigen in immune checkpoint blockade therapy. *Exp Hematol Oncol* (2018) 7:28. doi: 10.1186/s40164-018-0120-y
33. Yarchoan M, Johnson BA3rd, Lutz ER, Laheru DA, Jaffee EM. Targeting neoantigens to augment antitumor immunity. *Nat Rev Cancer* (2017) 17:209–22. doi: 10.1038/nrc.2016.154
34. Keung EZ, Burgess M, Salazar R, Parra ER, Rodrigues-Canales J, Bolejack V, et al. Correlative analyses of the SARC028 trial reveal an association between sarcoma-associated immune infiltrate and response to Pembrolizumab. *Clin Cancer Res* (2020) 26:1258–66. doi: 10.1158/1078-0432.CCR-19-1824
35. Klaver Y, Rijnders M, Oostvogels A, Wijers R, Smid M, Grunhagen D, et al. Differential quantities of immune checkpoint-expressing CD8 T cells in soft tissue sarcoma subtypes. *J Immunother Cancer* (2020) 8. doi: 10.1136/jitc-2019-000271
36. Sorbye SW, Kilvaer TK, Valkov A, Donnem T, Smeland E, Al-Shibli K, et al. Prognostic impact of peritumoral lymphocyte infiltration in soft tissue sarcomas. *BMC Clin Pathol* (2012) 12:5. doi: 10.1186/1472-6890-12-5
37. Sorbye SW, Kilvaer T, Valkov A, Donnem T, Smeland E, Al-Shibli K, et al. High expression of CD20+ lymphocytes in soft tissue sarcomas is a positive prognostic indicator. *Oncimmunology* (2012) 1:75–7. doi: 10.4161/onci.1.1.17825
38. Kim JR, Moon YJ, Kwon KS, Bae JS, Wagle S, Kim KM, et al. Tumor infiltrating PD1-positive lymphocytes and the expression of PD-L1 predict poor prognosis of soft tissue sarcomas. *PLoS One* (2013) 8:e82870. doi: 10.1371/journal.pone.0082870
39. Deng C, Xu Y, Fu J, Zhu X, Chen H, Xu H, et al. Reprogramming the tumor immunologic microenvironment using neoadjuvant chemotherapy in osteosarcoma. *Cancer Sci* (2020) 111:1899–909. doi: 10.1111/cas.14398
40. Wiecezorek E, Jablonowski Z, Lesicka M, Jablonska E, Kutwin P, Reszka E, et al. Genetic contributions of MHC class I antigen processing and presentation pathway to bladder cancer risk and recurrence. *Neoplasia* (2022) 69:443–55. doi: 10.4149/neo_2021_210805N1113
41. Carrillo-Bustamante P, Kesmir C, de Boer RJ. Can selective MHC downregulation explain the specificity and genetic diversity of NK cell receptors? *Front Immunol* (2015) 6:311. doi: 10.3389/fimmu.2015.00311
42. King RJ, Yu F, Singh PK. Genomic alterations in mucins across cancers. *Oncotarget* (2017) 8:67152–68. doi: 10.18632/oncotarget.17934
43. Svensson F, Lang T, Johansson MEV, Hansson GC. The central exons of the human MUC2 and MUC6 mucins are highly repetitive and variable in sequence between individuals. *Sci Rep* (2018) 8:17503. doi: 10.1038/s41598-018-35499-w
44. Maleki Vareki S. High and low mutational burden tumors versus immunologically hot and cold tumors and response to immune checkpoint inhibitors. *J Immunother Cancer* (2018) 6:157. doi: 10.1186/s40425-018-0479-7
45. Howitt BE, Shukla SA, Sholl LM, Ritterhouse LL, Watkins JC, Rodig S, et al. Association of polymerase e-mutated and microsatellite-instable endometrial cancers with neoantigen load, number of tumor-infiltrating lymphocytes, and expression of PD-1 and PD-L1. *JAMA Oncol* (2015) 1:1319–23. doi: 10.1001/jamaonc.2015.2151
46. Ma X, Zhang Y, Wang S, Yu J. Predictive value of tumor mutation burden (TMB) with targeted next-generation sequencing in immunotherapy inhibitors for non-small cell lung cancer (NSCLC). *J Cancer* (2021) 12:584–94. doi: 10.7150/jca.48105
47. Ning B, Liu Y, Wang M, Li Y, Xu T, Wei Y. The predictive value of tumor mutation burden on clinical efficacy of immune checkpoint inhibitors in melanoma: a systematic review and meta-analysis. *Front Pharmacol* (2022) 13:748674. doi: 10.3389/fphar.2022.748674
48. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* (2016) 351:1463–9. doi: 10.1126/science.aaf1490
49. Bonaventura P, Shekarian T, Alcazar V, Valladeau-Guilemond J, Valsesia-Wittmann S, Amigorena S, et al. Cold tumors: A therapeutic challenge for immunotherapy. *Front Immunol* (2019) 10:168. doi: 10.3389/fimmu.2019.00168
50. McGranahan N, Swanton C. Neoantigen quality, not quantity. *Sci Transl Med* (2019) 11. doi: 10.1126/scitranslmed.aax7918
51. Zhao H, Wang D, Zhang Z, Xian J, Bai X. Effect of gut microbiota-derived metabolites on immune checkpoint inhibitor therapy: enemy or friend? *Molecules* (2022) 27. doi: 10.3390/molecules27154799
52. Shim JA, Ryu JH, Jo Y, Hong C. The role of gut microbiota in T cell immunity and immune mediated disorders. *Int J Biol Sci* (2023) 19:1178–91. doi: 10.7150/ijbs.79430
53. Øvreid E, Bygstad B, Thomassen G. TSD: A research platform for sensitive data. *Proc Comput Sci* (2021) 181:127–34. doi: 10.1016/j.procs.2021.01.112
54. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997* (2013), 1–3. doi: 10.48550/arXiv.1303.3997
55. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* (2011) 43:491–8. doi: 10.1038/ng.806
56. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol* (2016) 17:122. doi: 10.1186/s13059-016-0974-4
57. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Auwera GAVD, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv* (2018), 201178. doi: 10.1101/201178
58. Patterson M, Marschall T, Pisanti N, van Iersel L, Stougie L, Klau GW, et al. WhatsHap: weighted haplotype assembly for future-generation sequencing reads. *J Comput Biol* (2015) 22:498–509. doi: 10.1089/cmb.2014.0157
59. Spooner W, McLaren W, Slidel T, Finch DK, Butler R, Campbell J, et al. HaploSaurus computes protein haplotypes for use in precision drug design. *Nat Commun* (2018) 9:4128. doi: 10.1038/s41467-018-06542-1
60. Taylor BS, Barretina J, Maki RG, Antonescu CR, Singer S, Ladanyi M. Advances in sarcoma genomics and new therapeutic targets. *Nat Rev Cancer* (2011) 11:541–57. doi: 10.1038/nrc3087
61. Zhu G, Benayed R, Ho C, Mullaney K, Sukhadia P, Rios K, et al. Diagnosis of known sarcoma fusions and novel fusion partners by targeted RNA sequencing with identification of a recurrent ACTB-FOSB fusion in pseudomyogenic hemangioendothelioma. *Mod Pathol* (2019) 32:609–20. doi: 10.1038/s41379-018-0175-7
62. Yang W, Lee KW, Srivastava RM, Kuo F, Krishna C, Chowell D, et al. Immunogenic neoantigens derived from gene fusions stimulate T cell responses. *Nat Med* (2019) 25:767–75. doi: 10.1038/s41591-019-0434-2
63. Uhrig S, Ellermann J, Walther T, Burkhardt P, Frohlich M, Hutter B, et al. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res* (2021) 31:448–60. doi: 10.1101/gr.257246.119

64. Merino DM, McShane LM, Fabrizio D, Funari V, Chen SJ, White JR, et al. Establishing guidelines to harmonize tumor mutational burden (TMB): in silico assessment of variation in TMB quantification across diagnostic platforms: phase I of the Friends of Cancer Research TMB Harmonization Project. *J Immunother Cancer* (2020) 8. doi: 10.1136/jitc-2019-000147
65. Sverchkova A, Anzar I, Stratford R, Clancy T. Improved HLA typing of Class I and Class II alleles from next-generation sequencing data. *HLA* (2019) 94:504–13. doi: 10.1111/tan.13685
66. Robinson J, Malik A, Parham P, Bodmer JG, Marsh SG. IMGT/HLA database—a sequence database for the human major histocompatibility complex. *Tissue Antigens* (2000) 55:280–7. doi: 10.1034/j.1399-0039.2000.550314.x
67. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* (2015) 33:1152–8. doi: 10.1038/nbt.3344
68. Montesin M, Murugesan K, Jin DX, Sharaf R, Sanchez N, Guria A, et al. Somatic HLA class I loss is a widespread mechanism of immune evasion which refines the use of tumor mutational burden as a biomarker of checkpoint inhibitor response. *Cancer Discov* (2021) 11:282–92. doi: 10.1158/2159-8290.CD-20-0672
69. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* (2016) 34:525–7. doi: 10.1038/nbt.3519
70. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med* (2019) 11:34. doi: 10.1186/s13073-019-0638-6
71. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics* (2019) 35:i436–45. doi: 10.1093/bioinformatics/btz363
72. Bentham R, Litchfield K, Watkins TBK, Lim EL, Rosenthal R, Martinez-Ruiz C, et al. Using DNA sequencing data to quantify T cell fraction and therapy response. *Nature* (2021) 597:555–60. doi: 10.1038/s41586-021-03894-5
73. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
74. Kelly A, Trowsdale J. Genetics of antigen processing and presentation. *Immunogenetics* (2019) 71:161–70. doi: 10.1007/s00251-018-1082-2
75. Mizuno S, Yamaguchi R, Hasegawa T, Hayashi S, Fujita M, Zhang F, et al. Immunogenomic pan-cancer landscape reveals immune escape mechanisms and immunoeediting histories. *Sci Rep* (2021) 11:15713. doi: 10.1038/s41598-021-95287-x
76. Li S, Liu Q, Zhou H, Lu H, Wang X. Subtyping of sarcomas based on pathway enrichment scores in bulk and single cell transcriptomes. *J Transl Med* (2022) 20:48. doi: 10.1186/s12967-022-03248-3
77. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) 15:550. doi: 10.1186/s13059-014-0550-8
78. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* (2015) 4:1521. doi: 10.12688/f1000research.7563.1
79. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol* (2010) 11:R14. doi: 10.1186/gb-2010-11-2-r14
80. Kontkanen P, Myllymäki P. MDL histogram density estimation. In: *Artificial Intelligence and Statistics 219-226 (PMLR, 2007)* (Helsinki, Finland: PMLR (Proceedings of Machine Learning Research)) (2007).