

1 **Title:**
2 Immune checkpoint inhibitor response in sarcomas associates with immune infiltrates and
3 increased expression of transposable elements and viral response pathways
4
5

6 **Authors and Affiliations:**

7 Benjamin A. Nacev^{1,2,*,#,**}, Martina Bradic^{1,3,*}, Hyung Jun Woo³, Allison L. Richards³, Ciara M.
8 Kelly^{1,2}, Mark A. Dickson^{1,2}, Mrinal M. Gounder^{1,2}, Mary L. Keohan^{1,2}, Ping Chi^{1,2}, Sujana
9 Movva^{1,2}, Robert Maki^{1,2}, Emily K. Slotkin^{1,2}, Evan Rosenbaum^{1,2}, Viswatej Avutu^{1,2}, Jason E.
10 Chan^{1,2}, Lauren Banks^{1,2}, Travis Adamson^{1,2}, Samuel Singer⁵, Cristina R. Antonescu⁶, William
11 D. Tap^{1,2}, Mark T.A. Donoghue^{3,†}, Sandra P. D'Angelo^{1,2,†}
12

13 ¹Department of Medicine, Memorial Sloan Kettering Cancer Center, Manhattan, New York

14 ²Weill Cornell Medical College, New York, NY

15 ³Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering
16 Cancer Center, New York, NY

17 ⁴Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY

18 ⁵Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY

19 ⁶Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY

20 *Equal contribution

21 †Equal contribution

22 #Correspondence

23 **Current affiliations: Department of Medicine and Department of Pathology University of
24 Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; UPMC Hillman Cancer Center,
25 Pittsburgh, Pennsylvania
26

27
28 **Manuscript Information:**

29 Main text figures: 4

30 Extended Data Figures: 9

31 Extended Data Tables: 3
32

33 **ABSTRACT**

34 Response to immune checkpoint inhibition (ICI) in sarcoma is overall low and heterogeneous.
35 Understanding determinants of ICI outcomes may improve efficacy and patient selection. One
36 potential mechanism is epigenetic de-repression of transposable elements (TEs), which
37 stimulates antitumor immunity. Here, we used transcriptomic data to assign immune-hot versus
38 immune-cold status to 67 pre-treatment biopsies of sarcomas from patients treated on ICI trials.
39 Progression-free survival and overall response was superior in the immune-hot group. Expression
40 of TEs and epigenetic regulators significantly predicted immune-hot status in a regression model
41 in which specific TE subfamilies and *IKZF1*, a chromatin-interacting transcription factor, were
42 significantly contributory. TE and *IKZF1* expression positively correlated with tumor immune
43 infiltrates, inflammatory pathways, and clinical outcomes. Key findings were confirmed in a
44 validation cohort (n=190). This work suggests that TE and *IKZF1* expression warrant investigation
45 as predictive biomarkers for ICI response and as therapeutic targets in sarcomas.

46

47 INTRODUCTION

48 Sarcomas are a diverse group of more than 170 histologic entities¹ that derive from tissues of
49 mesenchymal origin. The underlying genetic causes of sarcomas^{2,3} are diverse and their biologic
50 and pathologic behavior is highly varied⁴. The clinical management of metastatic sarcomas is
51 generally palliative and relies upon systemic therapies including cytotoxic chemotherapy, targeted
52 therapies, and in some instances immune checkpoint blockade⁵. The overall response rate to
53 first-line chemotherapy in soft tissue sarcoma is approximately 20%⁶. Hence, there is a need for
54 both new treatment modalities and improved methods to select patients most likely to respond to
55 specific treatments.

56
57 Immune checkpoint inhibition (ICI) has been studied in sarcomas, and activity has been noted
58 with nivolumab (anti-PD-1) alone or in combination with ipilimumab (anti-CTLA-4)⁷ or with
59 pembrolizumab (anti-PD-1) as a single agent⁸. Efforts to enhance the activity of ICI through
60 combination with other immune-modulatory drugs or with cytotoxic therapies has revealed
61 variable response rates, which likely depend on the drug combination and sarcoma subtype
62 (recently reviewed⁹). In parallel, predictive biomarkers for ICI response in sarcoma are being
63 explored. While microsatellite instability (MSI) and high tumor mutation burden (TMB) predict
64 response in carcinomas¹⁰⁻¹⁶, TMB is relatively low in sarcomas and MSI is exceedingly rare¹⁷.
65 Alternative biomarkers such as tertiary lymphoid structures and B cell and CD8+ T cell infiltrates
66 correlate with ICI response in some soft tissue sarcomas such as undifferentiated pleomorphic
67 sarcoma (UPS)^{2,18,19}.

68
69 Another potential determinant and predictor of antitumor immunity and ICI response are
70 epigenetic states, which are determined by chemical modifications of DNA, RNA, and DNA-
71 associated proteins together with their positioning relative to specific genomic sequences²⁰. One
72 key function of epigenetic states is to regulate transcriptional programs, including those that

73 influence immune signaling. Therefore, genetic or pharmacologic perturbation of the machinery
74 that establishes or maintains epigenetic states can prime ICI response in preclinical models and
75 correlates with ICI clinical response²¹⁻²⁹. For example, epigenetic mechanisms can promote
76 immune escape through repression of antigen-presenting machinery and transposable elements
77 (TEs), epigenetically silenced sequences of viral origin that, when de-repressed, stimulate
78 antiviral immune signaling^{22,24,30}.

79
80 We therefore hypothesized that sarcoma baseline immune infiltrates and clinical outcomes
81 following immunotherapy treatment are influenced by expression of TEs and epigenetic
82 regulators. To test this, we generated and analyzed transcriptomic profiles of pre-treatment
83 biopsies from 67 unique patients enrolled in 3 ICI trials at our institution and an independent
84 validation cohort. Here, we demonstrate that the efficacy of ICIs is linked to the de-repression of
85 TEs that are normally silenced by epigenetic mechanisms and upregulation of the transcription
86 factor IKZF1, which interacts with chromatin-modifying complexes. TE and *IKZF1* upregulation in
87 turn correlate with hallmarks of tumor-intrinsic innate immune activation such as type I interferon
88 and antigen presentation, suggesting a potential mechanism for enhanced immune response
89 mediated by tumor epigenetic states.

90
91 **Results**

92
93 **Baseline immune cell populations predict response and progression-free survival in**
94 **sarcoma patients treated with immune checkpoint inhibitors**

95 To study the influence of features linked to epigenetic states on antitumor immunity, we first
96 characterized baseline immune infiltrates in tumor biopsies from 67 patients with a heterogeneous
97 set of sarcomas (>10 subtypes) who were subsequently treated on ICI clinical trials (**Extended**

98 **Data Table 1**). Twelve patients responded to ICI and 55 did not (CR/PR=12, SD=21, PD=34).
99 Baseline samples were analyzed to identify tumor characteristics that could be informative prior
100 to treatment and to eliminate confounding by varying ICI drugs and combinations used across
101 trials. We employed an RNA sequencing (RNA-seq)-based method to quantify the abundance of
102 different immune populations, MCP-counter³¹. To obtain robust clustering of samples based on
103 their profile of immune infiltrates, we used a hierarchical clustering of principal components
104 (HCPC) approach³², which integrates principal components (PCA) and hierarchical clustering.
105 This HCPC revealed two highly distinct groups, which we deemed “immune-cold” and “immune-
106 hot” (**Figure 1A, B; Extended Data Figure 1**). Except for cancer-associated fibroblasts, all cell
107 types defined by MCP-counter were significantly associated with cluster partitioning, with T cells
108 ($p= 3.41 \times 10^{-10}$) contributing the most, followed by cytotoxicity score (representative of cytotoxic
109 lymphocytes) ($p=2.01 \times 10^{-9}$), and CD8⁺ T cells ($p=6.89 \times 10^{-9}$), NK cells ($p=3.54 \times 10^{-8}$), B cells
110 ($p=3.58 \times 10^{-8}$), neutrophils ($p=9.12 \times 10^{-8}$), myeloid dendritic cells ($p=1.17 \times 10^{-7}$),
111 macrophage/monocytes ($p=1.34 \times 10^{-7}$), and endothelial cells ($p=7.88 \times 10^{-3}$). The immune-hot
112 cluster displayed, on average, greater abundance of all immune cell types in comparison to the
113 overall mean, and conversely the immune-cold cluster displayed lower abundance of the same
114 immune cell types (**Extended Data Table 2**).

115
116 Having assigned tumors to hot and cold immune groups, we next determined how these immune
117 states correlated with clinical outcomes after the 67 patients in our cohort received ICI-based
118 intervention in one of 3 clinical trials: pembrolizumab plus talmogene laherparepvec
119 (NCT03069378)³³, nivolumab plus bempedaldesleukin (NCT03282344)², and pembrolizumab
120 plus epacadostat (NCT03414229)³⁴. There were no significant differences between the 3 ICI trials
121 with respect to the number of responders and non-responders or immune-hot and -cold patients
122 (**Extended Data Table 3**). We compared overall response rates (ORR) by RECIST version 1.1³⁵
123 in immune-hot (ORR=30% [9/30]) vs. immune-cold (ORR=8.1% [3/37]) tumors. The ORR in the

124 immune-hot group was significantly greater than in the immune-cold group (Fisher's Exact Test,
125 95% CI 1.03-30.31, $p=0.02$). Furthermore, the immune hot samples were more prevalent than the
126 immune cold samples in the complete response (CR) compared to progressive disease (PD)
127 groups (Fisher's Exact Test, 95% CI 0.02-0.73, $p=0.01$), while there was no significant difference
128 in the CR versus stable disease (SD) and SD versus PD. The expression levels of immune
129 checkpoint-related genes were consistent with the patterns observed in immune infiltrates, with
130 elevated expression of *CD274* (PD-L1), *CTLA4* (two-sided t-test, $p=1.83 \times 10^{-6}$ and $p=1.18 \times 10^{-4}$,
131 respectively), and *LAG3* (two-sided t-test, $p=0.12$) in immune-hot tumors (**Figure 1C**).

132
133 To determine if the baseline immune type was prognostic for progression-free survival (PFS), we
134 performed survival analysis including the histologic sarcoma subtype as a covariate (**Figure 1D**,
135 **Extended Data Figure 2**). Median PFS among patients with immune-cold tumors was 1.7 months
136 vs. 3.65 months for immune-hot. Tumor classification as immune-hot contributed to improved PFS
137 (HR=0.43, 95% CI 0.22-0.84, $p=0.01$) (**Extended Data Figure 2**). The histologic subtypes of
138 leiomyosarcoma, myxofibrosarcoma, osteosarcoma, and small blue round cell tumors had a
139 significant effect on PFS in our cohort.

140
141 **Hot and cold immune types are analogous to previously identified sarcoma immune**
142 **classes**

143 To determine how the two immune subtypes identified in this study relate to previously described
144 sarcoma immune classes (SICs), which correlate with immune infiltrates and ICI response, we
145 classified our samples according to those 5 SIC clusters (labeled A-E, **Extended Data Figure**
146 **3**)¹⁹. In total, 47% (14/30) of the immune-hot samples from our study fell into immune-hot SICs D
147 and E. The remaining 53% (16/30) of immune-hot samples were assigned to immune-cold SIC B.
148 In contrast, the immune-cold samples from our study were almost exclusively classified into
149 immune-cold SICs A and B, with only two samples matching SIC C (**Extended Data Figure 3**).

150 In summary, the two distinct immune clusters identified in this study between which we observe
151 differences in PFS and ORR following ICI treatment are associated with SICs that are consistent
152 with immune-high and immune-low states.

153
154 In addition to the validation of our clustering through comparison with independently developed
155 classifications, we also reasoned that if the immune type clusters identified in our approach via
156 deconvolution of bulk RNA sequencing accurately reflected immune cell populations, then the
157 immune-hot cluster should contain more immune infiltrates than the immune-cold cluster,
158 resulting in lower tumor content. Concordantly, the immune-hot type displayed significantly lower
159 purity compared to the immune-cold type (two-sided t-test; $t=-3.11$, $df=64$, $p=2.7 \times 10^{-3}$)
160 (**Extended Data Figure 4A**). To further confirm this relationship, we performed a permutation test
161 randomly assigning samples to immune groups and comparing the difference between purity
162 estimates between the two groups, which was repeated 10,000 times to produce a null
163 distribution. The observed data displayed significantly greater differences in tumor purity
164 estimates compared with the null distribution ($p=3.1 \times 10^{-3}$) (**Extended Data Figure 4B**). Lastly,
165 tumor purity was inversely correlated with lymphoid and myeloid cell content (**Extended Data**
166 **Figure 4C**), which is consistent with immune cell content contributing to the non-tumor cell
167 fraction.

168 169 **TE and Ikaros (*IKZF1*) expression predict immune types in sarcoma**

170 Although the activation of immune response through increased expression of transposable
171 elements (TEs) and the involvement of epigenetic genes in the regulation of TE expression has
172 been established in many cancers^{23,24,36-39}, these processes have not been well studied in
173 sarcoma. Our analysis of expression of 1,002 intergenic TEs across the two immune types shows
174 heterogenous expression (**Extended Data Figure 5**). Thus, we next asked if expression of TEs
175 and epigenetic regulators is predictive of tumor immune types in sarcoma. Lasso logistic

176 regression models including expression of TEs ($R^2=0.29$), epigenetic regulators ($R^2=0.19$) had
177 higher R^2 values, indicating that the models including these features are a better fit for the
178 prediction of immune type than a basic model that included only sarcoma subtypes and
179 sequencing batch or models of TE and epigenetic regulator expression with randomized immune
180 type sample labels (R^2 TE shuffled=0.02, R^2 epigenetic regulators shuffled=0.01) (**Figure 2A**).
181 Furthermore, the selected models identified a small set of informative epigenetic regulator
182 genes and TEs associated with the identified immune types from a large number of genes and
183 TEs that were part of the model, i.e. signature features. Signature features with the highest
184 contribution to the model included the MER57F (ERV1), MER45A (DNA transposon), Tigger17a
185 (DNA transposon), MER61F (ERV1), LTR104_Mam (*Gypsy*), HERVL74.int (ERV1)
186 TE subfamilies, expression of which was significantly greater in the immune-hot cluster (**Figure**
187 **2B** and **2C**). In addition, *IKZF1*, a chromatin-interacting transcription factor⁴⁰ which regulates
188 three-dimensional chromatin structure⁴¹, was the only epigenetic regulator of 532 genes tested
189 as single genes to significantly contribute to the immune type prediction model and was
190 associated with B cell infiltrates (**Figure 2B, 2C** and **Extended Data Figure 6A**).

191
192 We next fitted a logistic regression model using the signature features (i.e., *IKZF1* and TE score)
193 to predict their effect on immune type. To calculate a TE score, we combined the expression
194 values of the 6 signature feature TEs, for which expression of each of which was also positively
195 correlated (**Extended Data Figure 6B**). After adjusting for sequencing batch and histology, we
196 found that both TE score ($p=2.2 \times 10^{-3}$), and *IKZF1* expression ($p=5.8 \times 10^{-3}$), were significantly
197 associated with immune type. This suggests that *IKZF1* and TE affect immune type. We used a
198 conditional independence approach (see Methods) to further investigate the potential causal
199 relationships between *IKZF1*, TEs, and immune type. Our analysis revealed that: a) given *IKZF*
200 expression, TE expression (TE score) is not conditionally independent of immune type ($p=1.17 \times$
201 10^{-5}), b) given TE expression, *IKZF1* expression is conditionally independent of immune type

202 (p=0.14), and c) TEs and *IKZF1* do not have an independent impact on immune type (p=2.35 x
203 10⁻⁶). This analysis suggests that TEs play a significant role in determining immune type, and that
204 they interact with *IKZF1* in a complex way to modulate the immune response.

205

206 **High expression of TEs and *IKZF1* is associated with immune and inflammatory pathway** 207 **signatures and progression free survival**

208 We next determined whether *IKZF1* and TE expression correlated with activation of immune and
209 inflammatory pathways using a partial Pearson correlation. Both *IKZF1* and TE score were
210 positively correlated with multiple immune pathways, while pathways related to non-immune
211 function were either significantly inversely correlated or not significantly correlated, suggesting a
212 distinct relationship between TEs and *IKZF1* expression and immune activity in sarcomas (**Figure**
213 **3A**). Specifically, TE score and *IKZF1* expression were significantly correlated with antiviral
214 response pathways such as cGAS-STING (TE score, r²=0.64, p=7.90 x 10⁻⁹; *IKZF1*, r²=0.67,
215 p=1.03 x 10⁻⁹), type I interferon (TE score, r²=0.38, p= 1.55 x 10⁻³, *IKZF1*, r²=0.32, p=7.89 x 10⁻³),
216 and type II interferon (TE score, r²=0.68, p= 2.66 x 10⁻¹⁰, *IKZF1*, r²=0.55 p=1.28 x 10⁻⁶). Moreover,
217 we observed positive correlations between TE score and *IKZF1* expression and the upregulation
218 of antigen-processing machinery (TE score, r²=0.49, p=2.99 x 10⁻⁵, *IKZF1*, r²=0.27, p=2.44 x 10⁻
219 ²) as well as the CD8+ T cell effector pathway (TE score, r²=0.54, p=2.94 x 10⁻⁶, *IKZF1*, r²=0.45,
220 p=1.17 x 10⁻⁴).

221

222 Because *CD274* (PD-L1) expression was significantly higher in the immune-hot group (**Figure**
223 **1C**), we also investigated the association between immune checkpoint-related genes and immune
224 activity and found a significant positive correlation between *CD274* and immune and inflammatory
225 pathways (CD8+ T cell effector, r²=0.36, adjusted p= 3.21 x 10⁻³, cGAS-STING, r²=0.71, p= 1.53
226 x 10⁻¹¹; type II interferon, r²=0.54, p=2.51 x 10⁻⁶). TE score and *IKZF1* also positively correlated
227 with *CD274* expression (p=4.40 x 10⁻⁹ and p= 2.5 x 10⁻¹⁰ respectively) (**Figure 3B**). We next tested

228 whether the TE score and *IKZF1* expression were predictive of PFS. Both high TE score ($p=1.65$
229 $\times 10^{-3}$) and *IKZF1* expression ($p=9.28 \times 10^{-3}$) correlated with prolonged PFS (high TE 4.4 months
230 vs. low TE 1.8 months; high *IKZF1* 5.3 months vs. low *IKZF1* 1.8 months) (**Figure 3C**). The ORR
231 based on *IKZF1* expression was 54.5% (6/11) in the high-expressing group and 10.7% (6/56) in
232 the low-expressing group ($p=2.72 \times 10^{-3}$; Fisher's exact test). ORR in the TE-high group was 40%
233 (6/15) and 11.53% (6/52) in the TE-low group ($p=0.13$; Fisher's exact test). Taken together, these
234 findings suggest that both *IKZF1* expression and TE score, which were identified in model that
235 considered subtypes as a variable, could be explored as predictive biomarkers for ICI outcomes.
236

237 **TE and *IKZF1* expression associate with immune infiltrate and inflammatory pathway** 238 **activation in a separate validation cohort of sarcoma patients**

239 To assess the replicability of our findings, we applied our analysis to gene expression data from
240 190 sarcoma samples from the TCGA³. This group was chosen as it includes 5 sarcoma subtypes,
241 DDLPS (n=49), MFS (n=17), LMS (n=80; 53 STLMS; 27 ULMS), and UPS (n=44), which were
242 prevalent in our original cohort. The immune signatures in the validation cohort segregated into
243 two distinct clusters marked by high (immune-hot) and low (immune-cold) immune infiltrates and
244 expression of immune checkpoints (**Figure 4A-C, Extended Data Figure 7**). The immune-hot
245 cluster was associated with improved overall survival ($p=1.09 \times 10^{-2}$), at a median of 37.5 months
246 vs. 25.5 months for the immune-cold cluster (**Figure 4D**).

247
248 As in the original cohort, expression of TEs and epigenetic regulators predicted immune type
249 (**Extended Data Figure 8A**) and specific TEs and *IKZF1* were identified as signature features
250 that positively correlated with immune-hot classification (**Extended Data Figure 8B, C**).
251 Furthermore, expression of *IKZF1* and TEs (again defined as a composite TE score) correlated
252 with that of immune pathways including type I and II interferon ($p<0.001$), antigen-processing
253 machinery ($p<0.001$), and immune checkpoint genes ($p<0.001$) including *CD274*, but not non-

254 immune pathways (**Extended Data Figure 9A, B**). Overall survival was greater for patients whose
255 tumor had a high TE score ($p=1.26 \times 10^{-3}$) or *IKZF1* ($p=4.94 \times 10^{-3}$) expression (**Figure 4E, F**).

256

257

258 **Discussion**

259 To address the pressing need to identify predictive biomarkers of response to ICI-based therapy
260 in sarcomas, we identified the minimal number of immune clusters that represent immune-hot and
261 -cold sarcomas and showed that the former is associated with higher ORR and longer PFS
262 following ICI treatment independent of subtype. This finding corroborates prior studies that have
263 shown a correlation between high baseline immune infiltrates and response to immune therapy¹⁹.
264 Importantly, our work demonstrates that these findings apply in a cohort with a broad spectrum of
265 sarcoma subtypes and in the setting of 3 combination ICI trials with diverse mechanisms.
266 Moreover, our analysis shows that a binary classification of tumors is sufficient to correlate with
267 clinical outcomes, indicating that immune clustering can be simplified compared to previous
268 approaches that involved more groups¹⁹. Such a simplified system could be helpful in smaller
269 studies with limited numbers of cases.

270

271 To identify specific tumor-intrinsic features that contribute to differences in immune states, we
272 focused on epigenetic regulation. Epigenetic mechanisms are known to suppress antitumor
273 immune responses and targeting epigenetic pathways has emerged as a promising therapeutic
274 strategy^{21,30,42}. Specifically, we examined the expression of epigenetic regulators and TEs, the
275 latter of which are normally epigenetically silenced (e.g., via establishment of heterochromatin)
276 and can stimulate innate immune responses when de-repressed^{27-29,43}. We observed increased
277 expression of TEs in immune-hot tumors. This is consistent with the ability of TEs to activate
278 dsRNA-sensing pathways, as has been observed in the setting of genetic lesions in epigenetic

279 regulators or pharmacologic treatment that lead to their de-repression^{23,24,30,44}. Our observation
280 of upregulated antiviral immune responses (including cGAS and type I interferon signaling) and
281 antigen-presenting pathways is consistent with this mechanism. Further investigation is needed
282 to determine whether the presentation of TE-derived neoantigens via MHC-I, as observed in the
283 loss of epigenetic TE silencing, could also contribute to the immune-hot state^{27,38}.

284
285 In addition to TEs, our analysis revealed that expression of *IKZF1* was significantly greater in
286 immune-hot tumors and associated with PD-L1 expression and B cell infiltrates, which was
287 validated in a separate cohort of 190 sarcoma samples from the TCGA. Notably, greater
288 infiltration of B lineage immune cells associates with overall survival in soft tissue sarcomas¹⁹.
289 Although Ikaros, the *IKZF1* gene product, is primarily studied as a transcription factor in
290 hematologic lineages⁴⁰, it was included in our list of epigenetic regulators given the inherent
291 interaction of transcription factors and chromatin. Recent reports also suggest an important role
292 for Ikaros in regulating higher order chromatin structure⁴¹. Notably, previous studies have
293 determined that if *IKZF1* is expressed in tumor cells and not only in immune populations, it
294 contributes to upregulation of immune infiltration and enhances the efficacy of anti-PD-1 and anti-
295 CTLA-4 immunotherapies in murine models⁴⁵. Our findings raise the possibility of a similar effect
296 in sarcomas.

297
298 It is unclear whether TE de-repression and *IKZF1* expression are directly linked mechanistically.
299 One possibility is that *IKZF1* regulates TE expression, as several TE families contain an *IKZF1*-
300 binding motif⁴⁶. It is also possible that the mechanism for loss of epigenetic silencing at TEs
301 creates a permissive chromatin state for *IKZF1* binding that allows for activation of nearby genes
302 involved in innate immune activation. Alternatively, de-repression of TEs, which can act as cis-
303 regulatory elements, could promote *IKZF1* expression. In our study, conditional independence

304 analysis supports a model in which *IKZF1* regulates TE expression, which in turn
305 determines immune types. However, functional studies are needed to test this hypothesis.

306
307 There are several limitations of this study including the relatively small sample size (n=67), the
308 heterogeneity in sarcoma subtypes, and that patients were included from 3 trials of ICI-based
309 regimens with different mechanisms. However, while this heterogeneity may have decreased our
310 ability to identify signals related to specific epigenetic genes or TE families, we were reassuringly
311 able to classify tumors into immune classes that were predictive of clinical outcomes and confirm
312 prior classification systems. Furthermore, our key findings were confirmed in a larger validation
313 cohort (n=190), which included common sarcoma subtypes also represented in the original
314 sample set. However, the validation cohort differed in that samples were from patients who had
315 not received systemic therapy, it was composed of nearly all primary tumors, and outcome was
316 overall survival and not PFS or response following ICI-based treatment. Another caveat of the
317 study is that we selected polyadenylated transcripts for RNA sequencing, which would limit
318 detection of theoretically transcribed but non-polyadenylated TEs. We were also potentially limited
319 by considering epigenetic genes as independent, when many encode proteins that form
320 complexes or functional pathways.

321
322 Increasing the effectiveness of immunotherapies and identifying predictors of ICI response would
323 both represent important advances in sarcoma. Our work presents several possibilities for
324 achieving these goals using data from pretreatment biopsies. We confirm earlier studies showing
325 that pretreatment immune status can predict ICI outcomes and propose *IKZF1* expression and
326 TE score as potential predictive biomarkers for ICI response, both of which require validation. In
327 addition, this work reveals potential avenues to enhance ICI response through stimulation of
328 immune responsiveness of baseline immune-cold tumors to convert them into an immune-hot

329 phenotype. Based on this work, promoting the de-repression of TEs by pharmacologic targeting
330 of epigenetic regulators could be explored in preclinical models.

331

332

333 **Materials and Methods**

334 Clinical data were collected and DNA and RNA sequencing of pre-treatment biopsy samples was
335 performed under Institutional Review Board oversight of 3 clinical trials performed at the Memorial
336 Sloan Kettering Cancer Center. These include pembrolizumab plus talimogene laherparepvec
337 (NCT03069378)³³, nivolumab plus bempegaldesleukin (NCT03282344)², and pembrolizumab
338 plus epacadostat (NCT03414229)³⁴. Details regarding each study's design, safety oversight, and
339 interventions can be found in referenced publications for each study.

340

341 **Samples**

342 A total of 67 baseline samples from twelve sarcoma subtypes (angiosarcoma (ANGS) =4, alveolar
343 soft part sarcoma (ASPS)=1, chondrosarcoma (CHS)=6, epithelioid hemangioendothelioma
344 (EHE)=8, leiomyosarcoma (LMS)=11, liposarcoma (LPS)=8, myxofibrosarcoma (MFS)=2,
345 osteosarcoma (OS)=4, Other=7, sarcoma not otherwise specified (SARCNOs)=2, small blue
346 round cell sarcoma (SBRC)=4, and undifferentiated pleomorphic sarcoma (UPS)=8, representing
347 twelve responders and 55 non-responders (CR/PR=12, SD=21, PD=34) were transcriptionally
348 profiled (**Extended Data Table 1**).

349

350 **RNA sequencing, and quantification of TEs and genes**

351 After quantification of RNA using RiboGreen and quality control using the Agilent BioAnalyzer,
352 469-500 ng of total RNA with RNA integrity values ranging from 6.8–10 underwent polyA selection
353 and TruSeq library preparation following the instructions provided by Illumina (TruSeq Stranded

354 mRNA LT Kit, catalog #RS-122-2102), with 8 cycles of PCR. The resulting samples were
355 barcoded and run on a HiSeq 4000 at 100 paired-end reads, using the HiSeq 3000/4000 SBS
356 Kit (Illumina), generating an average of 41 million paired reads per sample. Ribosomal reads
357 represented 0.9–5.9% of the total reads generated and the percent of mRNA bases averaged
358 64%.

359
360 The obtained FASTQ files were processed using the REdiscoverTE³⁸ workflow, which allowed for
361 quantification based on transcript levels. Gene transcripts were aggregated to obtain individual
362 gene quantification. Read counts for each individual transposable element (TE) were then
363 gathered to the level of TE subfamily, family, and class, as defined by the human Repeatmasker
364 Hg38. TE expression was further divided into inter- and intragenic regions as defined by Gencode
365 GTF/GFF and implemented in REdiscoverTE. Downstream analysis considered only intergenic
366 expression of 1002 out of a total of 1052 TE subfamilies that were expressed. Gene-based
367 normalization factors were calculated using the 'RLE' algorithm in edgeR⁴⁷, as determined by
368 REdiscoverTE. The data was further variance-stabilized using the voom function from edgeR.

369

370 **RNAseq deconvolution and generation of immune clusters**

371 We quantified immune cell populations from variance-stabilized RNAseq data using the
372 immunedeconv R package and its deconvolute function, along with the MCPcounter option
373 3.6.3³¹. Batch effects due to sequencing run were removed using removeBatchEffect() function
374 from the limma R package⁴⁸. To reduce the dimensionality of the immune cell proportion data, we
375 first performed a principal component analysis, followed by hierarchical clustering on principal
376 components (HCPC) using the FactoMineR package⁴⁹. Cluster types were visualized using the
377 factoextra R package (<https://CRAN.R-project.org/package=factoextra>).

378

379 Heatmaps of expression in each cluster were generated based on the scaled (Z-scores) immune
380 cell proportions. Z-scores were calculated using the formula $z = (x-\mu)/\sigma$, where x is the raw cell
381 fraction, μ is the mean of all samples, and σ is the standard deviation for all samples.

382
383 To obtain the cellularity enrichment scores for 64 cell types, from which lymphoid and myeloid cell
384 type proportions can be derived, we used the xCellAnalysis function in the xCell R package
385 (<https://github.com/dviraran/xCell>). Total lymphoid content was calculated as the sum of 21
386 lymphoid cell scores, including CD8 + T cells, NK cells, CD4 + naive T cells, B cells, CD4+ T cells,
387 CD8+ Tem, Tregs, plasma cells, CD4 + Tcm, CD4+ Tem, memory B cells, CD8+ Tcm, naive B-
388 cells, CD4+ memory T cells, pro B cells, class-switched memory B cells, Th2 cells, Th1 cells,
389 CD8+ naive T cells, NKT, and Tgd cells. Total myeloid content was expressed as the sum of 13
390 cell scores, including monocytes, macrophages, dendritic cells (including activated, conventional,
391 interstitial, and plasmacytoid), neutrophils, eosinophils, M1 macrophages, M2 macrophages,
392 basophils, and mast cells.

393
394 To test if immune type predicts survival in sarcoma, we performed Cox regression analysis that
395 included histology to control for subtype-specific differences in outcomes. We compared survival
396 between groups using the Kaplan-Meier survival curve and the Cox proportional-hazards
397 regression mode. Differences were considered significant if the p-value was less than 0.05 for the
398 tested group.

399
400 **Exome sequencing and purity estimation**

401 Vially frozen cells were thawed and pelleted and incubated for at least 30 min in 360 μ L Buffer
402 ATL + 40 μ L proteinase K at 55°C. DNA was isolated using the DNeasy Blood & Tissue Kit
403 (QIAGEN catalog #69504) according to the manufacturer's protocol with 1 h of incubation at 55°C
404 for digestion. DNA was eluted in 0.5X Buffer AE.

405

406 After PicoGreen quantification and quality control by Agilent BioAnalyzer, 100-250 ng of DNA was
407 used to prepare libraries using the KAPA Hyper Prep Kit (Kapa Biosystems KK8504) with 8 cycles
408 of PCR. After sample barcoding, 100-500 ng of library DNA was captured by hybridization using
409 the xGen Exome Research Panel v1.0 (IDT) according to the manufacturer's protocol. Post-
410 capture libraries were amplified using 8 PCR cycles. Samples were run on a HiSeq 4000 at 100
411 paired-end reads using the HiSeq 3000/4000 SBS Kit (Illumina). Normal and tumor samples were
412 covered to an average of 102X and 219X, respectively.

413

414 FASTQ files were aligned and processed using the in-house workflow Tempo
415 (<https://github.com/mskcc/tempo>)⁵⁰. Briefly, reads were aligned using Burroughs-Wheeler Aligner
416 (BWA)-MEM⁵¹ to the GRCh37 reference genome and base recalibration was performed using
417 Genome Analysis Toolkit (GATK) best practices. Somatic genome variants were called using the
418 union of Mutect2 and Strelka2. Variants were then filtered based on the following criteria: tumor
419 read depth of 20, variant allele frequency $< 0.5 \times$ the tumor alternate read count of 3, and normal
420 read depth of 10. In addition, repeated regions from RepeatMasker⁵² and variants that appear at
421 allele frequencies > 0.01 in GNOMAD⁵³ were filtered out. Somatic copy number alterations were
422 analyzed using FACETS (Fraction and Allele-Specific Copy Number Estimates from Tumor
423 Sequencing) v0.5.14⁵⁴. Each tumor and matched normal pair was processed in two steps: a first
424 run for ploidy and purity estimation followed by a second run for detection of focal events. Each
425 fit was reviewed manually to minimize false positives and to estimate the quality of the fit. Purity
426 estimates from facets were used in the subsequent analysis.

427

428 **Lasso association between immune types and genomic features**

429 To identify genomic features that significantly differed between the two immune types, we used
430 lasso logistic regression via penalized maximum likelihood using the R package *glmnet*⁶⁵. To
431 account for potential variations due to sequencing batch or subtypes, we included these
432 parameters in the basic model. Other models were further built with either normalized expression
433 of 1002 intergenic TEs, shuffled TE expression, normalized expression of epigenetic modulators
434 (532 genes), or shuffled epigenetic modulator expression.

- 435 1. Basic model: Immune types~ batch + sarcoma subtypes
- 436 2. Basic model + TE: Immune types~ batch + sarcoma subtypes + 1002 TEs
- 437 3. Basic model + TE shuffled: Immune types~ batch + sarcoma subtypes + 1002 TEs
438 shuffled
- 439 4. Basic model + epigenetic genes: Immune types~ batch + sarcoma subtypes + 532
440 epigenetic genes
- 441 5. Basic model + epigenetic genes shuffled: Immune types~ batch + sarcoma subtypes +
442 532 epigenetic genes shuffled

443 TEs in the models represent intergenic TEs, and shuffled TE or epigenetic genes data represents
444 randomly assigned TE or epigenetic genes expression to the samples.

445
446 Tenfold cross-validation was performed for each regression, and lasso coefficients at one
447 standard error of the minimum mean cross-validation errors (λ_{1se}) were used. Each lasso
448 fit returned a small number of predictors, i.e. variables with non-zero coefficients, matching
449 genomic features with significant contributions to difference between the two immune types.
450 R^2 values for each model were calculated from the fraction of deviance explained and averaged
451 across the 10 rounds of cross-validation. R^2 values were then used to determine the model with
452 the best performance. To identify notable features associated with immune type, we extracted
453 non-zero coefficients of the final best models.

454

455 To further test the relationship between significant TE and epigenetic features determined by the
456 glmnet model, we used logistic glm regression in which immune type represented a dependent
457 variable, while TE score and *IKZF1* expression represented independent variables. The model
458 was corrected for batch and histology covariates. The TE score was calculated by generating a
459 z-score for the expression of 8 TEs found to be significant in the glmnet analysis and that exhibited
460 positive correlation with each other (**Extended Data Figure 6B**). Z-score was generated using
461 the *gsva* function of the GSVA package⁵⁶.

462

463 **Logistic regression and conditional independence test**

464 To further confirm the relationship between selected TEs and *IKZF1* expression with respect to
465 immune cluster, we performed a logistic regression test. TE score and *IKZF1* were used as
466 independent variables to assess their association with immune type. Batch and histology were
467 used as covariates in the model.

468

469 Conditional independence (mutual information) tests to identify causal relationships between TEs,
470 *IKZF1*, and immune-hot/-cold phenotype were performed using the *bnlearn* package in R
471 (<https://www.bnlearn.com/>). Three hypotheses were tested:

472 a) TE score \rightarrow *IKZF1* \rightarrow Immune type: Immune type is conditionally independent of TE given
473 *IKZF1*; TEs regulate *IKZF1* and do not directly regulate immune type.

474 b) *IKZF1* \rightarrow TE score \rightarrow Immune type: Immune type is conditionally independent of *IKZF1*
475 given TE; *IKZF1* regulates TEs and does not directly regulate immune type.

476 c) *IKZF1* \rightarrow Immune type \leftarrow TE score: *IKZF1* and TEs are conditionally independent of
477 immune type.

478

479 **Gene signature calculations**

480 Genes for immune/inflammatory and other signatures used to determine the correlation of
481 significant features found to be predictive of immune type were defined as previously described
482 in literature and summarized in Kong et al.³⁸ (except the cGAS pathway, which was downloaded
483 from KEGG
484 <https://www.gseamsigdb.org/gsea/msigdb/cards/>). The ssGSEA algorithm was used to
485 comprehensively assess gene signature expression of each⁵⁷. The correlation between gene
486 signatures and normalized expression of significant features was assessed by partial Pearson
487 correlation analysis with batch and histology as covariates. P values were corrected using the
488 Benjamini-Hochberg correction.

489

490 **Comparison with previously reported immune classes**

491 To compare our immune clusters with formerly derived 5 sarcoma immune classes (SIC)
492 previously defined by Petitprez et al.,¹⁹ we obtained centroid infiltration scores for each of 4 cell
493 types (i.e. T cells, cytotoxic scores, B lineage, endothelial cells) of the 5 clusters derived from
494 MCP-counter analysis from the authors of the paper. We then calculated Euclidian distance
495 (distance = $\sqrt{\sum(A_i - B_i)^2}$) between centroids of four cell types (i.e. T cells, cytotoxic scores, B lineage,
496 endothelial cells) from each SIC (i.e. A,B,C,D,E) and the Z-score scaled MCP-counter proportions
497 from the same four cell types in our data. Each sample was assigned to SIC type based on the
498 lowest Euclidian distance with the 4 centroid infiltration scores for each SIC. Z-score-scaled
499 immune cell proportions were then plotted using the Complex heatmap package in R, and the
500 comparison with our Immune hot and cold clusters was performed.

501

502 **TCGA data analysis**

503 RNA sequencing data and phenotypic information were obtained from dbGaP for 190 TCGA
504 samples from 5 sarcoma subtypes, including DDLPS (n=49), MFS (n=17), LMS (n=80; 53 STLMS
505 +27 ULMS), and UPS (n=44). The REdiscoverTE pipeline was used to quantify gene and TE

506 expression. Batch effect information was downloaded from the TCGA Batch Effects Viewer
507 (<https://bioinformatics.mdanderson.org/public-software/tcga-batch-effects/>) and considered in the
508 subsequent data analysis. RNA sequencing was deconvoluted, immune clusters identified, and
509 lasso associations between immune types and genomic features and overall survival were
510 analyzed as described above. For the Kaplan- Meier analysis of this dataset our Cox regression
511 analysis included histology and tumor size. The latter was included since the TCGA dataset
512 comprises nearly all primary cases in which tumor size can be an more important prognostic
513 factor.

514

515 **Data Availability**

516 All RNA sequencing data, where informed consent has been obtained from the patient, is publicly
517 available via dbGaP (accession number: phs003284). Three samples are not publicly available
518 due to lack of consent for their release. All exome recapture sequencing data will be available via
519 dbGaP under accession number phs001783 by the time of publication.

520

521 **Code Availability**

522 Custom code used for analysis is publicly available here:

523 https://github.com/BradicM/Sarcoma_TE_paper_analysis

524

525 **Acknowledgements**

526 This work was supported by Merck, Amgen, NEKTAR, Incyte, Bristol Myers Squibb, Cycle for
527 Survival, and Witherwax Fund. BAN received support from the NCI K08CA245212, the
528 Connective Tissue Oncology Society Basic Science Research Award (with WDT), and the Damon
529 Runyon Clinical Investigator Award. Additional support was provided by the Memorial Sloan
530 Kettering Cancer Center Support Grant (P30 CA008748) and Hillman Cancer Center Support
531 Grant (P30 CA047904) provided additional support. We acknowledge the use of the Integrated

532 Genomics Operation Core, funded by the NCI Cancer Center Support Grant (CCSG, P30
533 CA08748), and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

534

535 **Competing Interests**

536 Sujana Movva: research funding from Ascentage Pharma, Tracon, Hutchinson Medi-pharma,
537 Pfizer/Trillium and research support from Merck, Clovis, and Bristol Meyers Squibb. Jason Chan:
538 research support from Ono pharmaceuticals. Mark Dickson: Research funding (to institution) from
539 Eli Lilly, Aadi Bioscience, and Sumitomo Pharma. Mrinal Gounder: Personal Honoraria/Advisory
540 Boards and/or Associated Research Paid to Institution from Aadi, Ayala, Bayer, Boehringer
541 Ingelheim, Daiichi, Epizyme, Karyopharm, Regeneron, Rain, Springworks, Tracon and TYME;
542 OTHER: Guidepoint, GLG, Third Bridge; Flatiron Health CME Honoraria: Medscape, More Health,
543 Physicians Education Resource, MJ LifeSciences and touchIME; ROYALTIES: Wolters Kluwer;
544 patents with MSKCC (GODDESS PRO); uncompensated research with Foundation Medicine
545 GRANTS from Food and Drug Administration (R01 FD005105) and the National Cancer Institute,
546 National Institutes of Health (P30CA008748)—core grant (CCSG shared resources and core
547 facility). Ping Chi: personal honoraria/advisory boards/consulting from Deciphera, NingboNewBay
548 Medical Technology; institutional research funding from Pfizer/Array, Deciphera, Ningbo NewBay
549 Medical Technology. Robert Maki: consulting fees from AADi, Bayer, Deciphera, Presage,
550 Springworks, American Board of Internal Medicine, American Society for Clinical Oncology and
551 UptoDate. Ciara Kelly: Institutional research funding from Merck, Amgen, Servier, Regeneron,
552 Xencor, Curadev pharma; Consulting for Kartos pharmaceuticals, Deciphera. Sandra D'Angelo:
553 Consulting or Advisory Role for Aadi Bioscience Adaptimmune, Adicet Bio, GI Innovations,
554 GlaxoSmithKline, Incyte, Medendi, Medidata, Nektar, Pfizer, Rain Therapeutics, Servier;
555 Research Funding from EMD Serono, Amgen, Merck, Incyte, Nektar, Britsol-Meyers Squibb,
556 Deciphera; Travel, Accommodations, Expenses from Adaptimmune, EMD Serono, Nektar;
557 Participation on a DataSafety Monitoring Board or Advisory Board for GlaxoSmithKline, Nektar,

558 Adaptimmune, Merck. William Tap: Consulting, Advisory Role, Honoraria: Aadi Biosciences,
559 Abbisko, Amgen, AmMAx Bio, Avacta, Ayala Pharmaceuticals, Bayer, BioAlta, Boehringer
560 Ingelheim, C4 Therapeutics, Cogent Biosciences, Curadev, Daiichi Sankyo, Deciphera, Eli Lilly,
561 Epizyme Inc (Nexus Global Group), Foghorn Therapeutics, Ikena Oncology, IMGT, Inhibix Inc.,
562 Ipsen Pharma, Jansen, Kowa Research Inst., Medpacto, Novo Holdings, PER, Servier, Sonata
563 Therapeutics; research funding from Novartis, Eli Lilly, Plexxikon, Daiichi Sankyo, Tracon
564 Pharma, Blueprint Medicines, Immune Design, BioAlta, Deciphera; Patents, Royalties, Other
565 Intellectual Property: Companion Diagnostics for CDK4 inhibitors (14/854,329), Stock and Other
566 Ownership Interests: Certis Oncology Solution, Atropos. All other authors declare no competing
567 interests.

568

569 **Extended Data Tables**

570

571 **Extended Data Table 1. Patient and sample characteristics.**

572

573 **Extended Data Table 2. Summary of cell type contribution to two immune clusters**

574 **resulting from hierarchical clustering of principal components (HCPC) analysis.** Overall
575 mean of cell proportion per cluster.

576

577 **Extended Data Table 3. Summary of immune type counts and clinical responses per**
578 **clinical protocol.** P values derived by Fisher's exact test.

579

580 **References**

581

- 582 1. WHO. *Soft tissue and bone tumours*, (International Agency for Research on Cancer,
583 Lyon (France), 2020).
- 584 2. D'Angelo, S.P., *et al.* Pilot study of bempegaldesleukin in combination with nivolumab in
585 patients with metastatic sarcoma. *Nat Commun* **13**, 3477 (2022).
- 586 3. Cancer Genome Atlas Research, N. Comprehensive and Integrated Genomic
587 Characterization of Adult Soft Tissue Sarcomas. *Cell* **171**, 950-965 e928 (2017).
- 588 4. Siegel, R.L., Miller, K.D. & Jemal, A. Cancer statistics, 2018. *CA Cancer J Clin* **68**, 7-30
589 (2018).
- 590 5. von Mehren, M., *et al.* Soft Tissue Sarcoma, Version 2.2022, NCCN Clinical Practice
591 Guidelines in Oncology. *J Natl Compr Canc Netw* **20**, 815-833 (2022).
- 592 6. Seddon, B., *et al.* Gemcitabine and docetaxel versus doxorubicin as first-line treatment
593 in previously untreated advanced unresectable or metastatic soft-tissue sarcomas
594 (GeDDiS): a randomised controlled phase 3 trial. *The Lancet Oncology* **18**, 1397-1410
595 (2017).
- 596 7. D'Angelo, S.P., *et al.* Nivolumab with or without ipilimumab treatment for metastatic
597 sarcoma (Alliance A091401): two open-label, non-comparative, randomised, phase 2
598 trials. *The Lancet Oncology* **19**, 416-426 (2018).
- 599 8. Tawbi, H.A., *et al.* Pembrolizumab in advanced soft-tissue sarcoma and bone sarcoma
600 (SARC028): a multicentre, two-cohort, single-arm, open-label, phase 2 trial. *The Lancet*
601 *Oncology* **18**, 1493-1501 (2017).
- 602 9. Banks, L.B. & D'Angelo, S.P. The Role of Immunotherapy in the Management of Soft
603 Tissue Sarcomas: Current Landscape and Future Outlook. *J Natl Compr Canc Netw* **20**,
604 834-844 (2022).

- 605 10. Carbone, D.P., *et al.* First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung
606 Cancer. *N Engl J Med* **376**, 2415-2426 (2017).
- 607 11. Hellmann, M.D., *et al.* Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor
608 Mutational Burden. *N Engl J Med* **378**, 2093-2104 (2018).
- 609 12. Rosenberg, J.E., *et al.* Atezolizumab in patients with locally advanced and metastatic
610 urothelial carcinoma who have progressed following treatment with platinum-based
611 chemotherapy: a single-arm, multicentre, phase 2 trial. *The Lancet* **387**, 1909-1920
612 (2016).
- 613 13. Van Allen, E.M., *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic
614 melanoma. *Science* **350**, 207-211 (2015).
- 615 14. Rizvi, N.A., *et al.* Cancer immunology. Mutational landscape determines sensitivity to
616 PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124-128 (2015).
- 617 15. Samstein, R.M., *et al.* Tumor mutational load predicts survival after immunotherapy
618 across multiple cancer types. *Nat Genet* **51**, 202-206 (2019).
- 619 16. Le, D.T., *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1
620 blockade. *Science* **357**, 409-413 (2017).
- 621 17. Nacev, B.A., *et al.* Clinical sequencing of soft tissue and bone sarcomas delineates
622 diverse genomic landscapes and potential therapeutic targets. *Nat Commun* **13**, 3405
623 (2022).
- 624 18. Italiano, A., *et al.* Pembrolizumab in soft-tissue sarcomas with tertiary lymphoid
625 structures: a phase 2 PEMBROSARC trial cohort. *Nat Med* **28**, 1199-1206 (2022).
- 626 19. Petitprez, F., *et al.* B cells are associated with survival and immunotherapy response in
627 sarcoma. *Nature* **577**, 556-560 (2020).
- 628 20. Allis, C.D., Caparros, M.-L., Jenuwein, T. & Reinberg, D. *Epigenetics*, (CSH Press, Cold
629 Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2015).

- 630 21. Que, Y., *et al.* Frequent amplification of HDAC genes and efficacy of HDAC inhibitor
631 chidamide and PD-1 blockade combination in soft tissue sarcoma. *J Immunother Cancer*
632 **9**(2021).
- 633 22. Krug, B., *et al.* Pervasive H3K27 Acetylation Leads to ERV Expression and a
634 Therapeutic Vulnerability in H3K27M Gliomas. *Cancer Cell* **35**, 782-797 e788 (2019).
- 635 23. Chiappinelli, K.B., *et al.* Inhibiting DNA Methylation Causes an Interferon Response in
636 Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162**, 974-986 (2015).
- 637 24. Topper, M.J., *et al.* Epigenetic Therapy Ties MYC Depletion to Reversing Immune
638 Evasion and Treating Lung Cancer. *Cell* **171**, 1284-1300 e1221 (2017).
- 639 25. Sheng, W., *et al.* LSD1 Ablation Stimulates Anti-tumor Immunity and Enables
640 Checkpoint Blockade. *Cell* **174**, 549-563.e519 (2018).
- 641 26. Macfarlan, T.S., *et al.* Endogenous retroviruses and neighboring genes are coordinately
642 repressed by LSD1/KDM1A. *Genes & Development* **25**, 594-607 (2011).
- 643 27. Griffin, G.K., *et al.* Epigenetic silencing by SETDB1 suppresses tumour intrinsic
644 immunogenicity. *Nature* (2021).
- 645 28. Zhang, S.M., *et al.* KDM5B promotes immune evasion by recruiting SETDB1 to silence
646 retroelements. *Nature* **598**, 682-687 (2021).
- 647 29. Hu, H., *et al.* Targeting the Atf7ip-Setdb1 Complex Augments Antitumor Immunity by
648 Boosting Tumor Immunogenicity. *Cancer Immunol Res* **9**, 1298-1315 (2021).
- 649 30. Burr, M.L., *et al.* An Evolutionarily Conserved Function of Polycomb Silences the MHC
650 Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. *Cancer*
651 *Cell* **36**, 385-401 e388 (2019).
- 652 31. Sturm, G., *et al.* Comprehensive evaluation of transcriptome-based cell-type
653 quantification methods for immuno-oncology. *Bioinformatics* **35**, i436-i445 (2019).
- 654 32. Lê S, J.J., Husson F. FactoMineR: A Package for Multivariate Analysis. *Journal of*
655 *Statistical Software* **25**, 1–18 (2008).

- 656 33. Kelly, C.M., Antonescu, C. R., Bowler, T., Munhoz, R., Chi, P., Dickson, M. A., Gounder,
657 M. M., Keohan, M. L., Movva, S., Dholakia, R., Ahmad, H., Biniakewitz, M., Condy, M.,
658 Phelan, H., Callahan, M., Wong, P., Singer, S., Ariyan, C., Bartlett, E. K., Crago, A.,
659 Yoon, S., Hwang, S., Erinjeri, J.P., Qin, L.X., Tap, W.D., D'Angelo, S. P. Objective
660 Response Rate Among Patients With Locally Advanced or Metastatic Sarcoma Treated
661 With Talimogene Laherparepvec in Combination With Pembrolizumab: A Phase 2
662 Clinical Trial. . *JAMA Oncol*, 402-408 (2020).
- 663 34. Kelly, C.M., *et al.* A Phase II Study of Epcadostat and Pembrolizumab in Patients with
664 Advanced Sarcoma. *Clin Cancer Res* **29**, 2043-2051 (2023).
- 665 35. Eisenhauer, E.A., *et al.* New response evaluation criteria in solid tumours: revised
666 RECIST guideline (version 1.1). *Eur J Cancer* **45**, 228-247 (2009).
- 667 36. Grundy, E.E., Diab, N. & Chiappinelli, K.B. Transposable element regulation and
668 expression in cancer. *FEBS J* **289**, 1160-1179 (2022).
- 669 37. Anwar, S.L., Wulaningsih, W. & Lehmann, U. Transposable Elements in Human Cancer:
670 Causes and Consequences of Deregulation. *Int J Mol Sci* **18**(2017).
- 671 38. Kong, Y., *et al.* Transposable element expression in tumors is associated with immune
672 infiltration and increased antigenicity. *Nat Commun* **10**(2019).
- 673 39. Griffin, G.K., *et al.* Epigenetic silencing by SETDB1 suppresses tumour intrinsic
674 immunogenicity. *Nature* **595**, 309-314 (2021).
- 675 40. Churchman, M.L. & Mullighan, C.G. Ikaros: Exploiting and targeting the hematopoietic
676 stem cell niche in B-progenitor acute lymphoblastic leukemia. *Exp Hematol* **46**, 1-8
677 (2017).
- 678 41. Hu, Y., *et al.* Lineage-specific 3D genome organization is assembled at multiple scales
679 by IKAROS. *Cell* **186**, 5269-5289 e5222 (2023).

- 680 42. Soldi, R., *et al.* The novel reversible LSD1 inhibitor SP-2577 promotes anti-tumor
681 immunity in SWItch/Sucrose-NonFermentable (SWI/SNF) complex mutated ovarian
682 cancer. *PLOS ONE* **15**, e0235705 (2020).
- 683 43. Voon, H.P., *et al.* ATRX Plays a Key Role in Maintaining Silencing at Interstitial
684 Heterochromatic Loci and Imprinted Genes. *Cell Rep* **11**, 405-418 (2015).
- 685 44. Stone, M.L., *et al.* Epigenetic therapy activates type I interferon signaling in murine
686 ovarian cancer to reduce immunosuppression and tumor burden. *Proc Natl Acad Sci U S*
687 *A* **114**, E10981-E10990 (2017).
- 688 45. Chen, J.C., Perez-Lorenzo, R., Saenger, Y.M., Drake, C.G. & Christiano, A.M. IKZF1
689 Enhances Immune Infiltrate Recruitment in Solid Tumors and Susceptibility to
690 Immunotherapy. *Cell Syst* **7**, 92-103 e104 (2018).
- 691 46. Deniz, O., *et al.* Endogenous retroviruses are a source of enhancers with oncogenic
692 potential in acute myeloid leukaemia. *Nat Commun* **11**, 3506 (2020).
- 693 47. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for
694 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-
695 140 (2010).
- 696 48. Ritchie, M.E., *et al.* limma powers differential expression analyses for RNA-sequencing
697 and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
- 698 49. Barrero, M.J. Epigenetic Regulation of the Non-Coding Genome: Opportunities for
699 Immuno-Oncology. *Epigenomes* **4**(2020).
- 700 50. MIT. Tempo: CCS Research Pipeline for Whole-Genome and Whole-Exome Sequencing
701 (2019).
- 702 51. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
703 (2013).
- 704 52. Smit, A.H., R; Green, P. RepeatMasker Open-4.0. (2013-2015).

- 705 53. Karczewski, K.J., *et al.* The mutational constraint spectrum quantified from variation in
706 141,456 humans. *Nature* **581**, 434-443 (2020).
- 707 54. Shen, R. & Seshan, V.E. FACETS: allele-specific copy number and clonal heterogeneity
708 analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* **44**, e131 (2016).
- 709 55. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear
710 Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
- 711 56. Lee, E., Chuang, H.Y., Kim, J.W., Ideker, T. & Lee, D. Inferring pathway activity toward
712 precise disease classification. *PLoS Comput Biol* **4**, e1000217 (2008).
- 713 57. Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for
714 microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
- 715

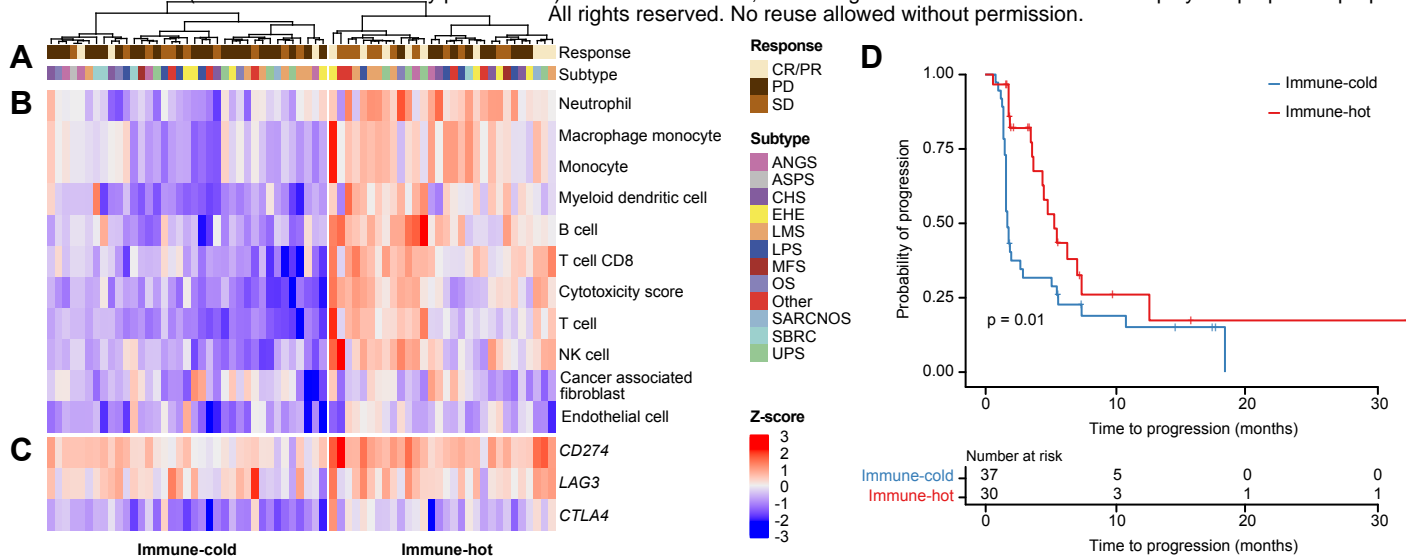


Figure 1. Clustering of immune cell fractions groups tumors into two distinct types. **A.** Color bars at the top of the heatmap label samples by response (SD, stable disease; PD, progressive disease; CR, complete response; PR, partial response), and histological subtype. Angiosarcoma, ANGS; alveolar soft part sarcoma, ASPS; chondrosarcoma, CHS; epithelioid hemangioendothelioma, EHE; leiomyosarcoma, LMS; liposarcoma, LPS; myxofibrosarcoma, MFS; osteosarcoma, OS; sarcoma not otherwise specified, SARCNOs; small blue round cell sarcoma, SBRC; undifferentiated pleomorphic sarcoma, UPS. **B.** Heatmap of immune and stromal cell fractions and cytotoxicity score determined by MCP-counter Z-scores. **C.** Immune checkpoint gene expression Z-scores. **D.** Kaplan-Meier plot representing progression-free survival (PFS) probability of immune-hot and -cold types. Tick marks indicate censoring. The P values on the Kaplan-Meier plots represent that output from cox proportional model that includes histology as a covariate.

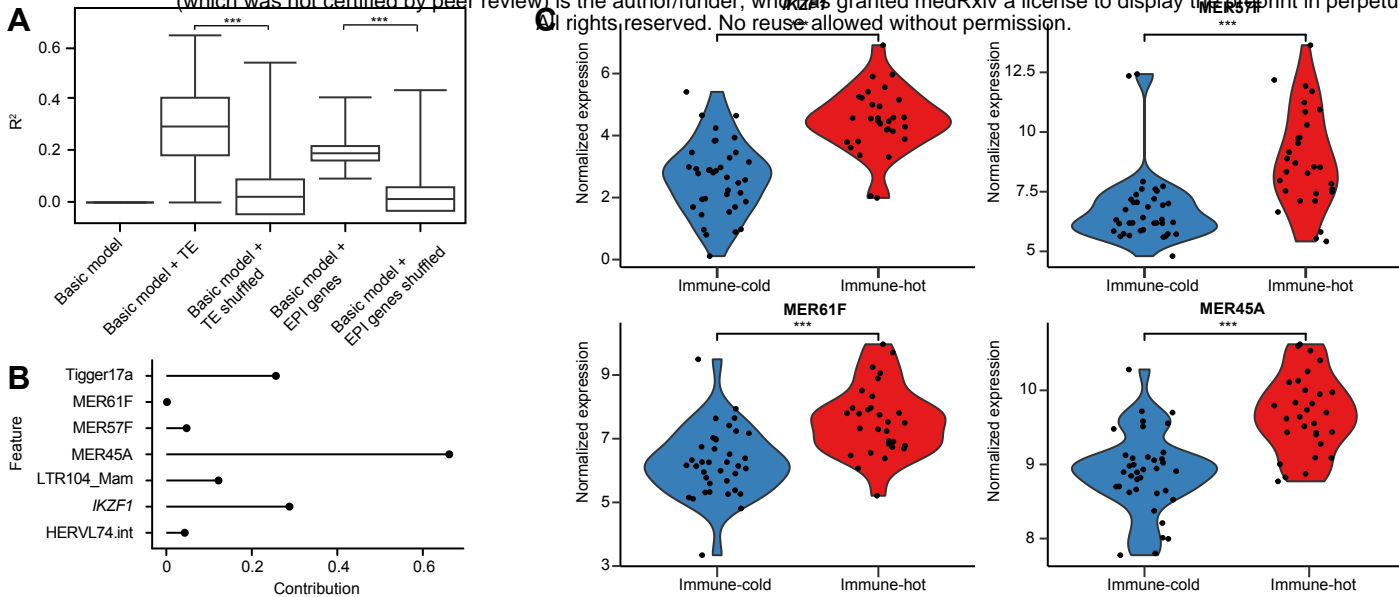


Figure 2. Transposable element and IKZF1 expression predict tumor immune groups. **A.** Comparison of lasso logistic regression model performances (R^2) of the 5 tested models for prediction of immune type. P values determined by t-test; $*** < 2.2 \times 10^{-16}$. **B.** Contribution of significant features from the TE and epigenetic models (models with the highest R^2) represented as non-zero coefficients. The size and sign of contribution (coefficients) indicate the direction and strength of the feature's effect on the outcome (immune cluster). **C.** Violin plots of normalized expression of transcripts identified as significant features in the regression model in immune-hot and -cold clusters. $***$, $p < 0.001$ as determined by one sided t-test.

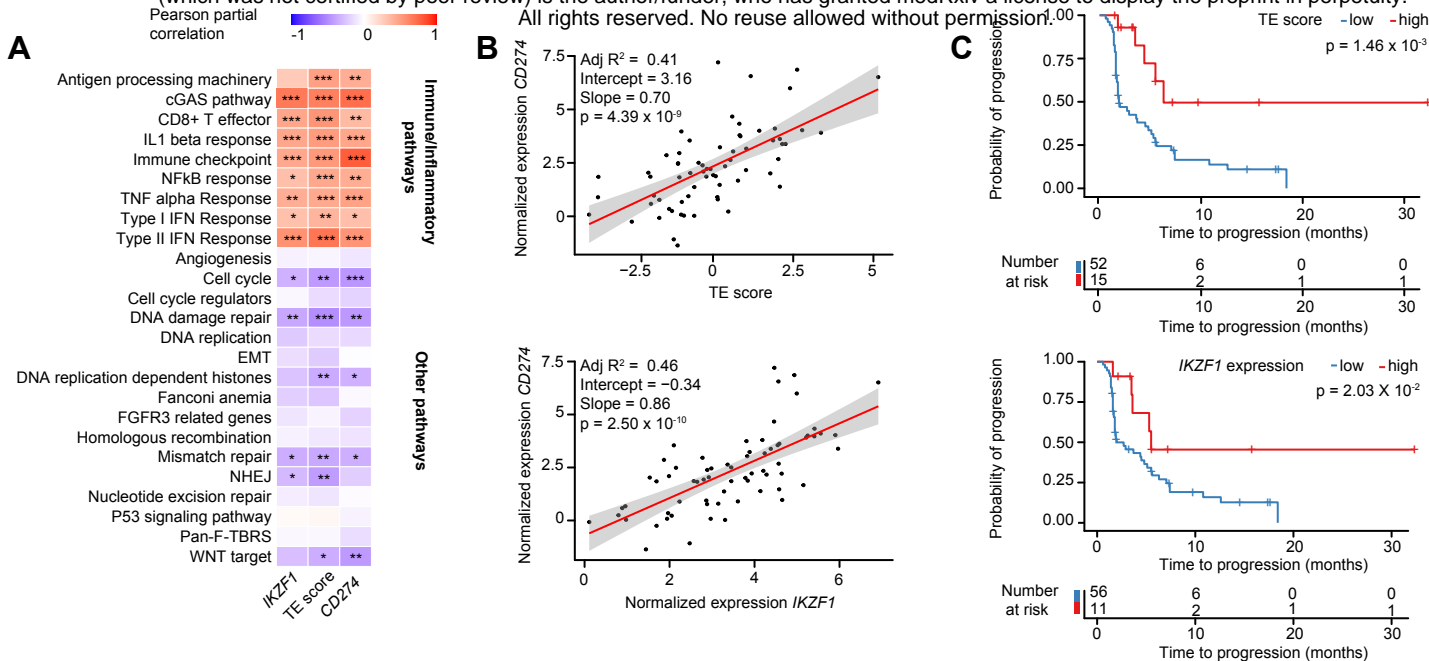


Figure 3. Immune pathway activation and progression-free survival following ICI treatment are associated with increased expression of multiple TE families and IKZF1. **A.** Heatmap of partial Pearson correlation including batch and histology as covariates. Scale from -1 (inverse correlation, blue) to 1 (positive correlation, red). Asterisks indicate Benjamini-Hochberg-corrected p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **B.** Correlation between CD274 (PD-L1) gene expression and TE score, and CD274 and IKZF1 expression. **C.** Kaplan-Meier curves representing progression-free survival probability according to high vs. low TE scores and IKZF1 expression. The P values on the Kaplan-Meier plots represent that output from cox proportional model that includes histology as a covariate.

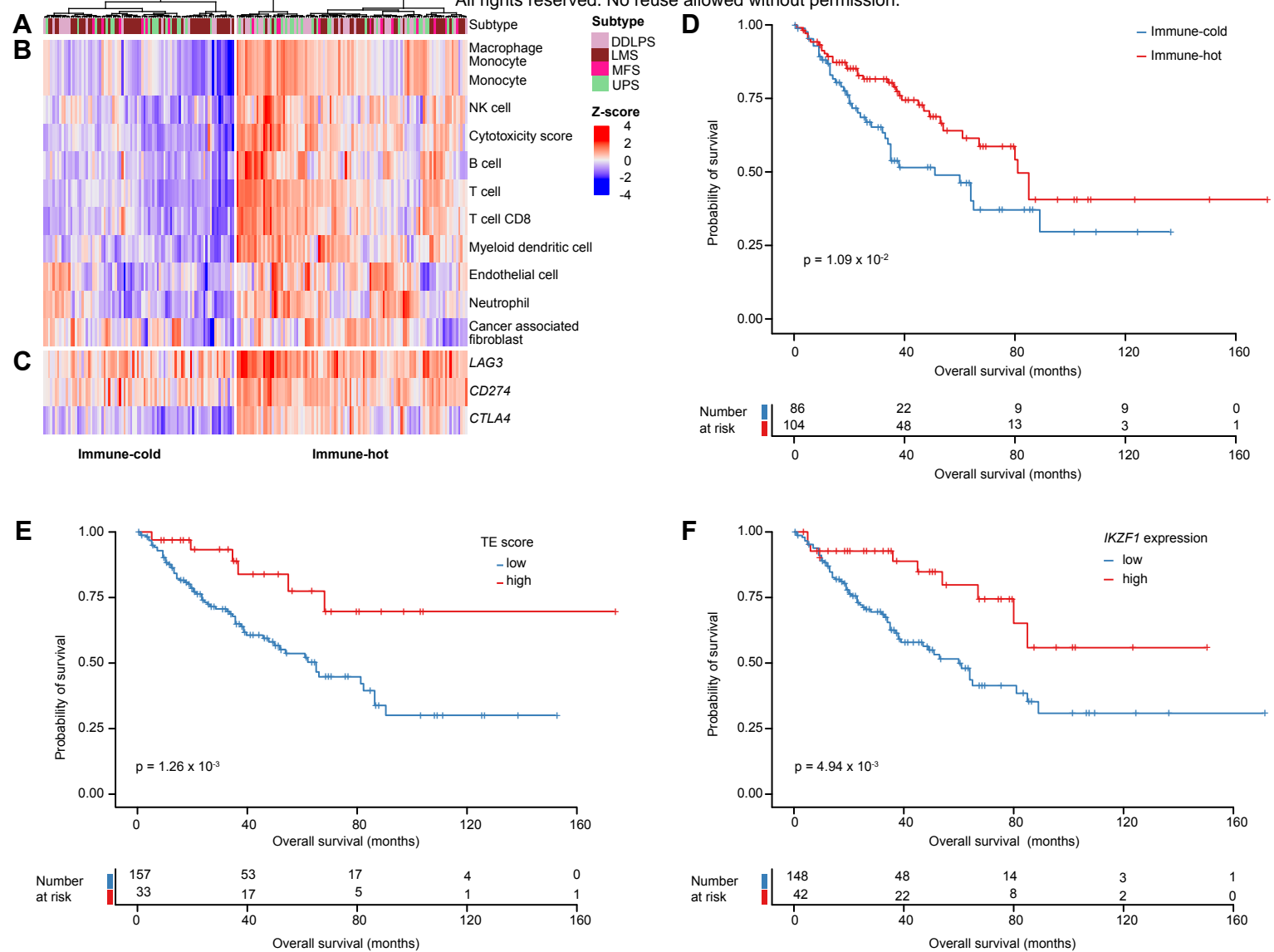
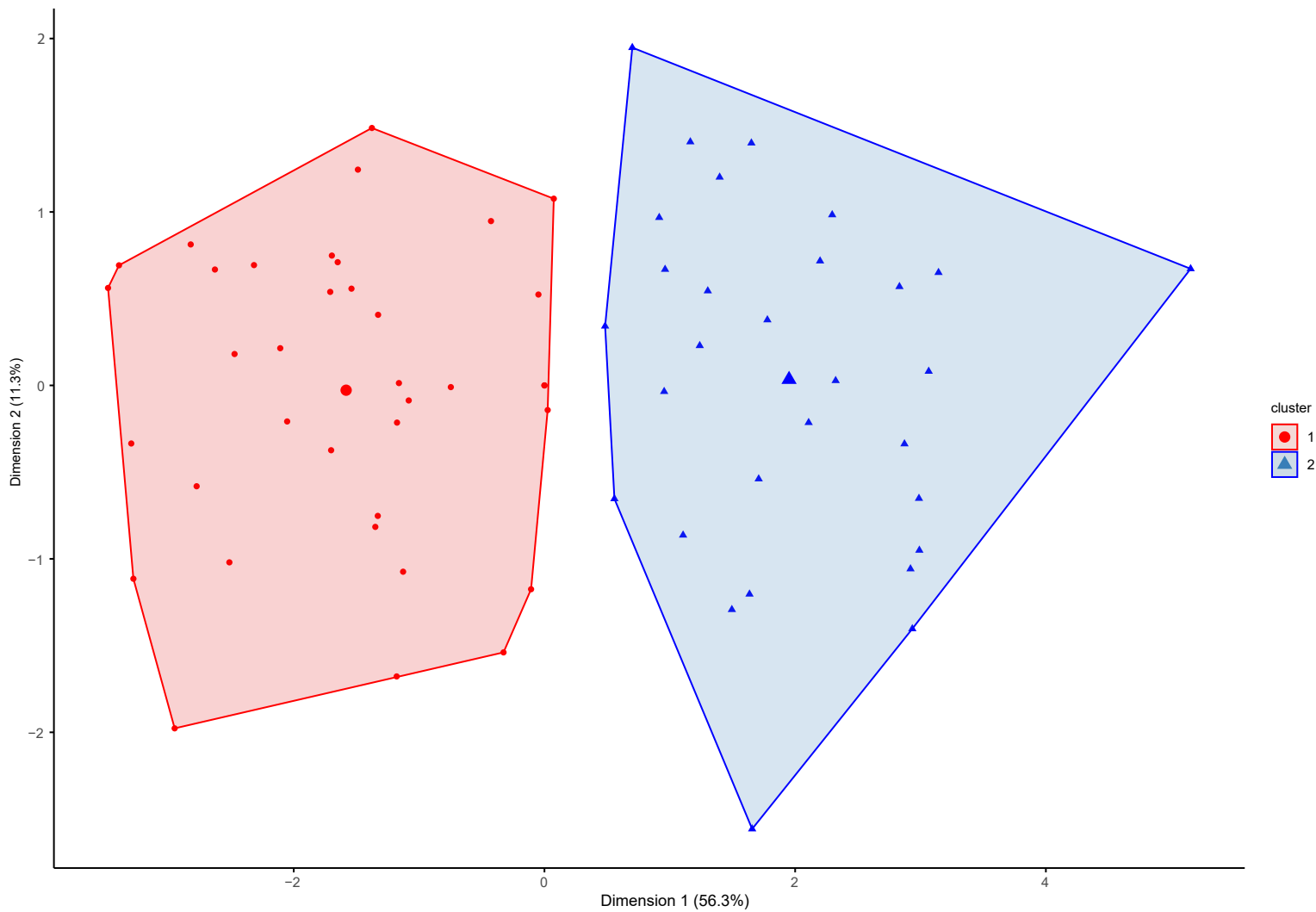
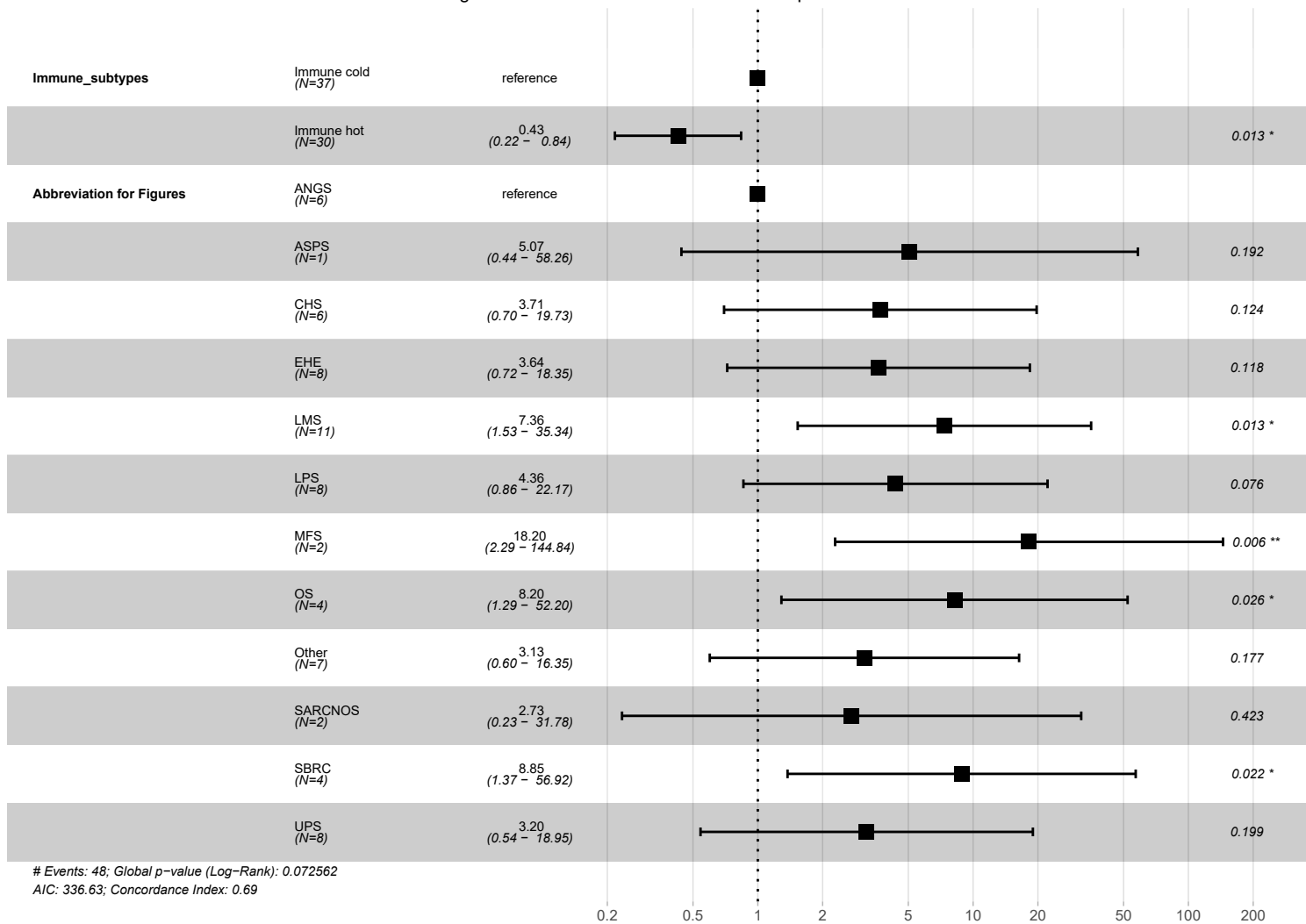


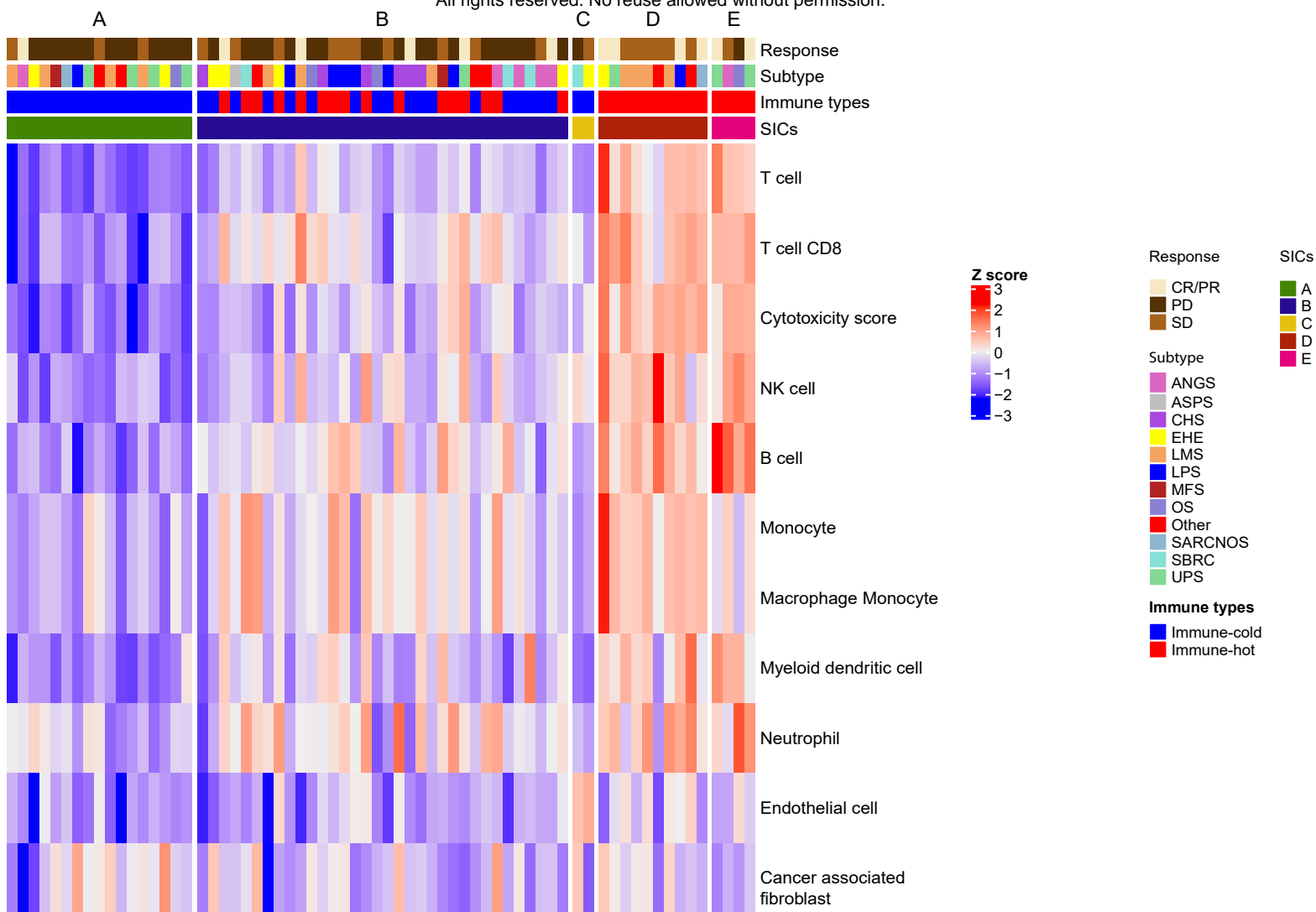
Figure 4. TE score and IKZF1 expression associate with improved survival in a validation cohort. A. Clustering of samples into immune-hot and immune-cold types. Color bar at top labels samples by histological subtype. DDLPS, dedifferentiated liposarcoma; LMS, leiomyosarcoma; MFS, myxofibrosarcoma; UPS, undifferentiated pleomorphic sarcoma. **B.** Heatmap of immune and stromal cell fractions and cytotoxicity score determined by MCP-counter Z-scores. **C.** Immune checkpoint gene expression Z-scores. **D.** Kaplan-Meier plot of overall survival probability of patients with immune-hot and -cold type tumors. **E-F.** Kaplan-Meier curves representing overall survival probability of high (red) and low (blue) **E.** TE scores and **F.** high (red) and (low) IKZF1 expression. The P values on the Kaplan-Meier plots represent that output from cox proportional model that includes histology and tumor size as covariates.



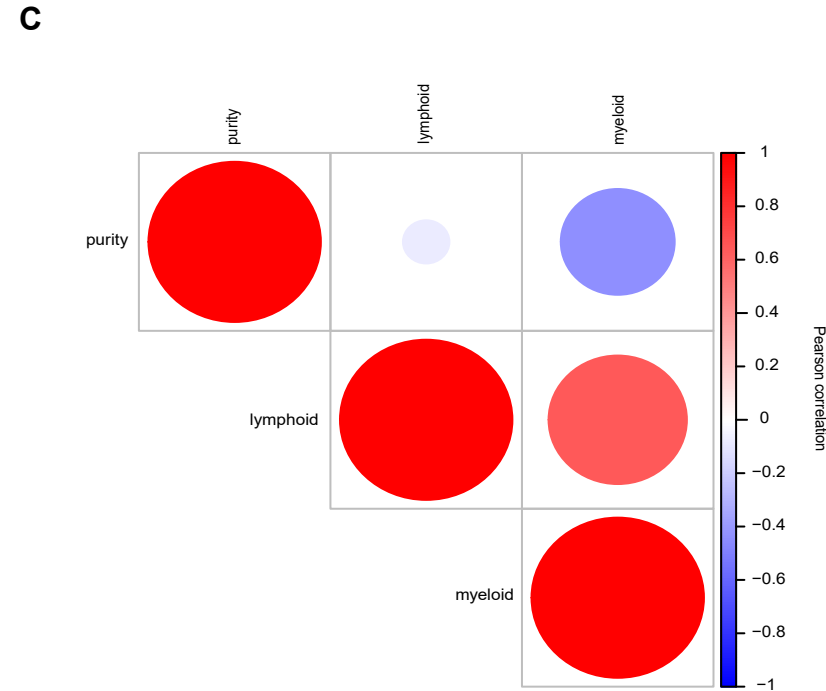
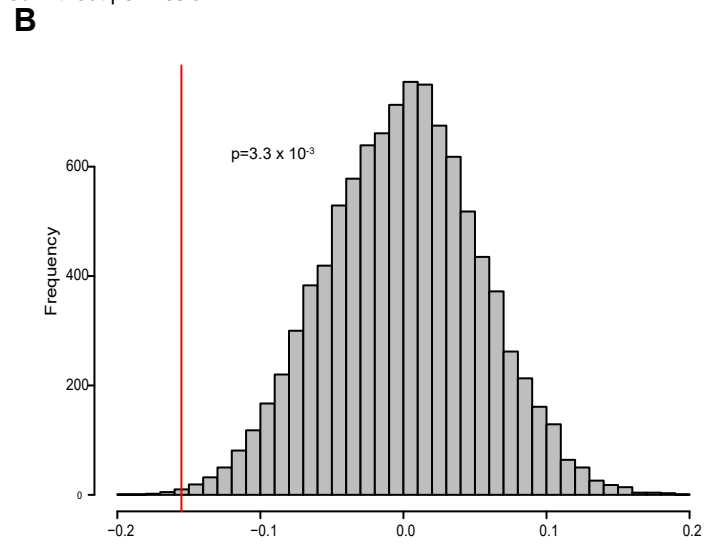
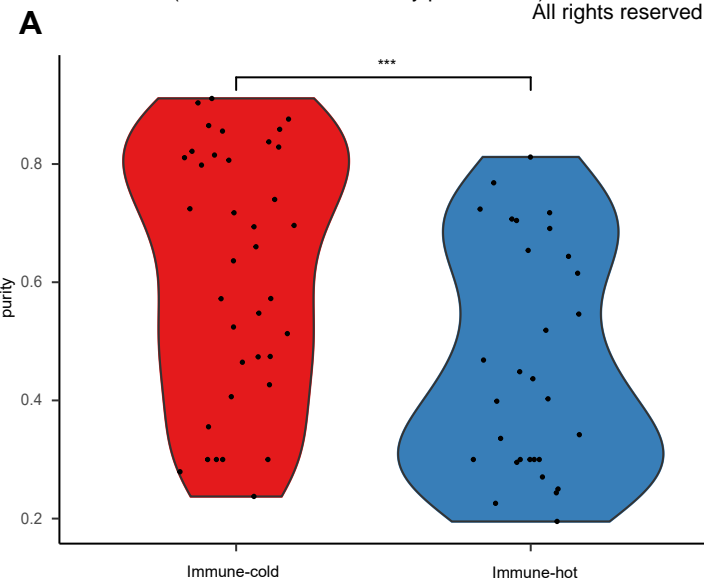
Extended Data Figure 1. Determination of immune clusters from MCP-counter-based immune-deconvoluted cell proportions. Factor map representing two clusters based on hierarchical clustering of principal components. Each dot represents an individual patient sample.



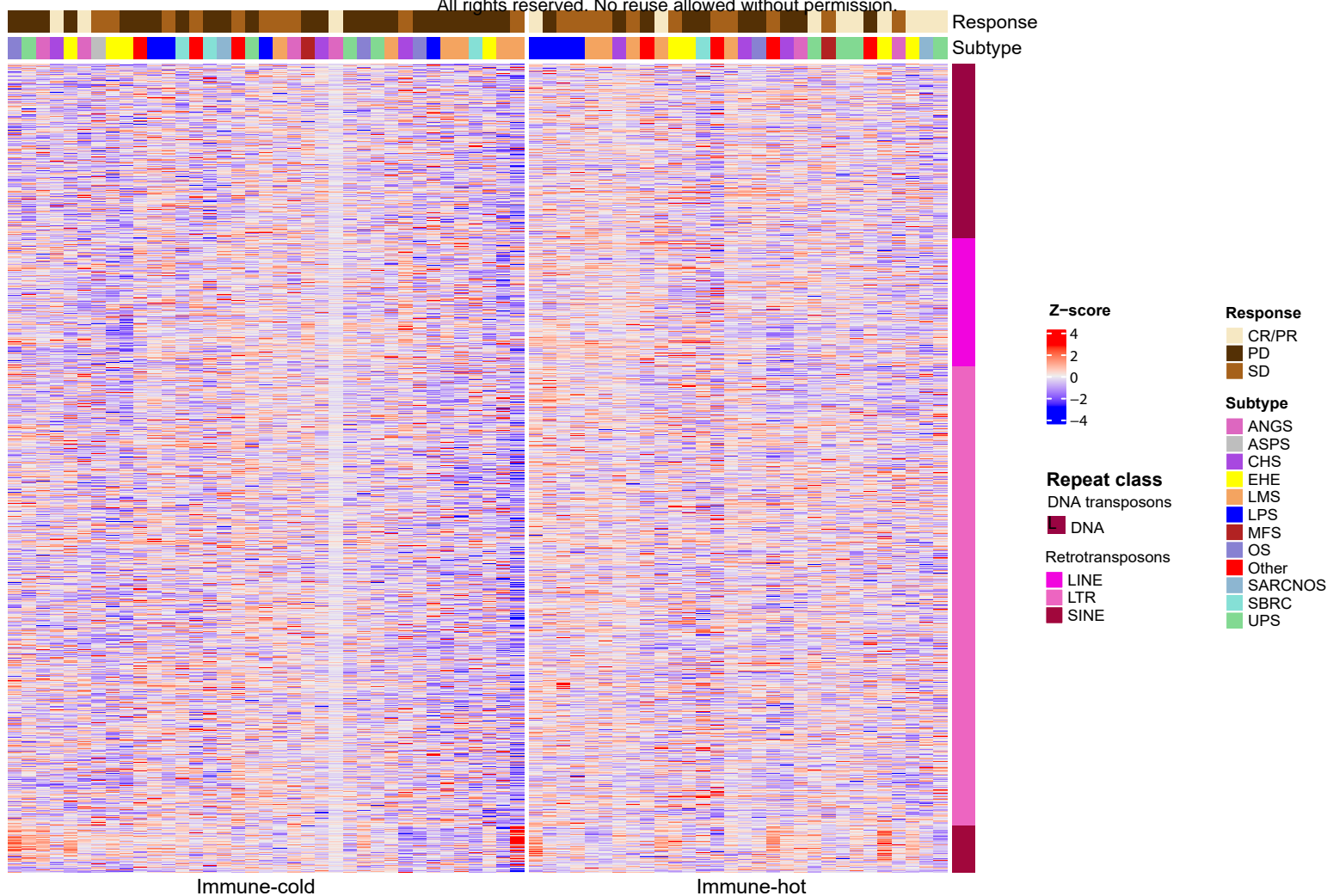
Extended Data Figure 2. Forest plots showing multivariable Cox regression analysis of contribution of immune cluster, histology, and clinical protocol to risk of progression. p-values calculated using Cox proportional hazards analysis (*, $p \leq 0.05$).



Extended Data Figure 3. Hot and cold immune types are related to previously identified sarcoma immune classes. SIC types of each sample in study cohort as determined by the Euclidian distance between centroids of the proportions of 4 cell types (i.e. T cells, cytotoxic scores, B lineage, endothelial cells) between each SIC (i.e., A, B, C, D, E)19 and proportions in each sample. Color bars at top of heatmap label the samples by response, histological subtype, SIC, and immune type. Heatmap indicates immune cell fraction determined by MCP-counter Z-score.

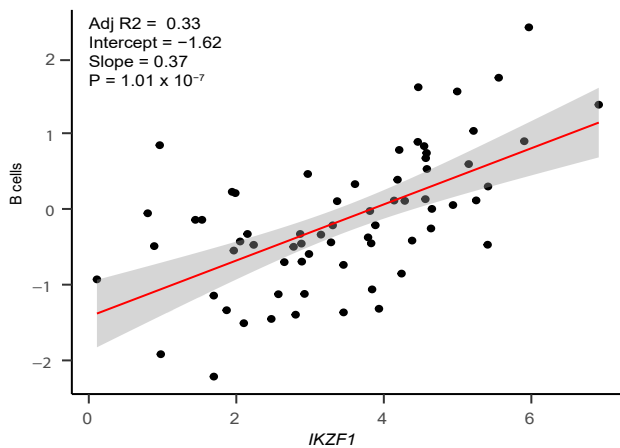


Extended Data Figure 4. Immune infiltrates are inversely related to sample purity. **A.** Violin plot comparing tumor purity between the two immune types. **B.** Distribution of 10,000 permutations of tumor purity shuffling between “immune-cold” and “immune-hot”. The histogram shows the simulated absolute permuted differences in means. The vertical red line represents the observed value for the original two samples (immune-cold and immune-hot). **C.** Correlation between purity, lymphoid content, and myeloid cell content. Scale from -1 (inverse correlation, blue), to 1 (positive correlation, red). Areas of circles represent the absolute value of corresponding correlation coefficients.

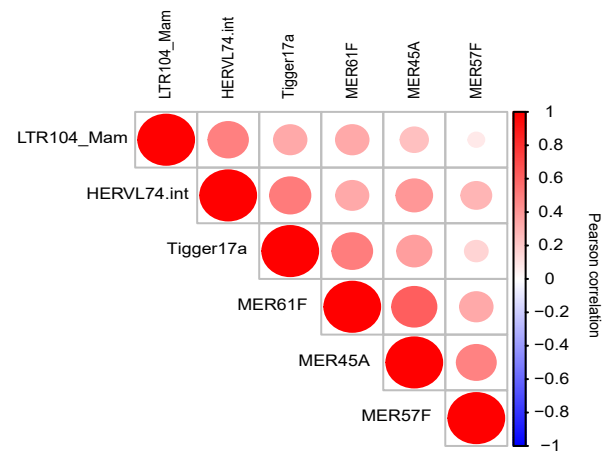


Extended Data Figure 5. TE expression is heterogeneous across sarcoma samples. Expression of all 1002 intergenic TEs expressed in the studied samples. Color bars at top of heatmap label the samples by response and histological subtype. Color bar at right labels repeat classes; LINE- Long interspersed nuclear elements, LTR-long terminal repeats, SINE- short interspersed nuclear elements. TE expression represented as Z- score; batch effect was removed.

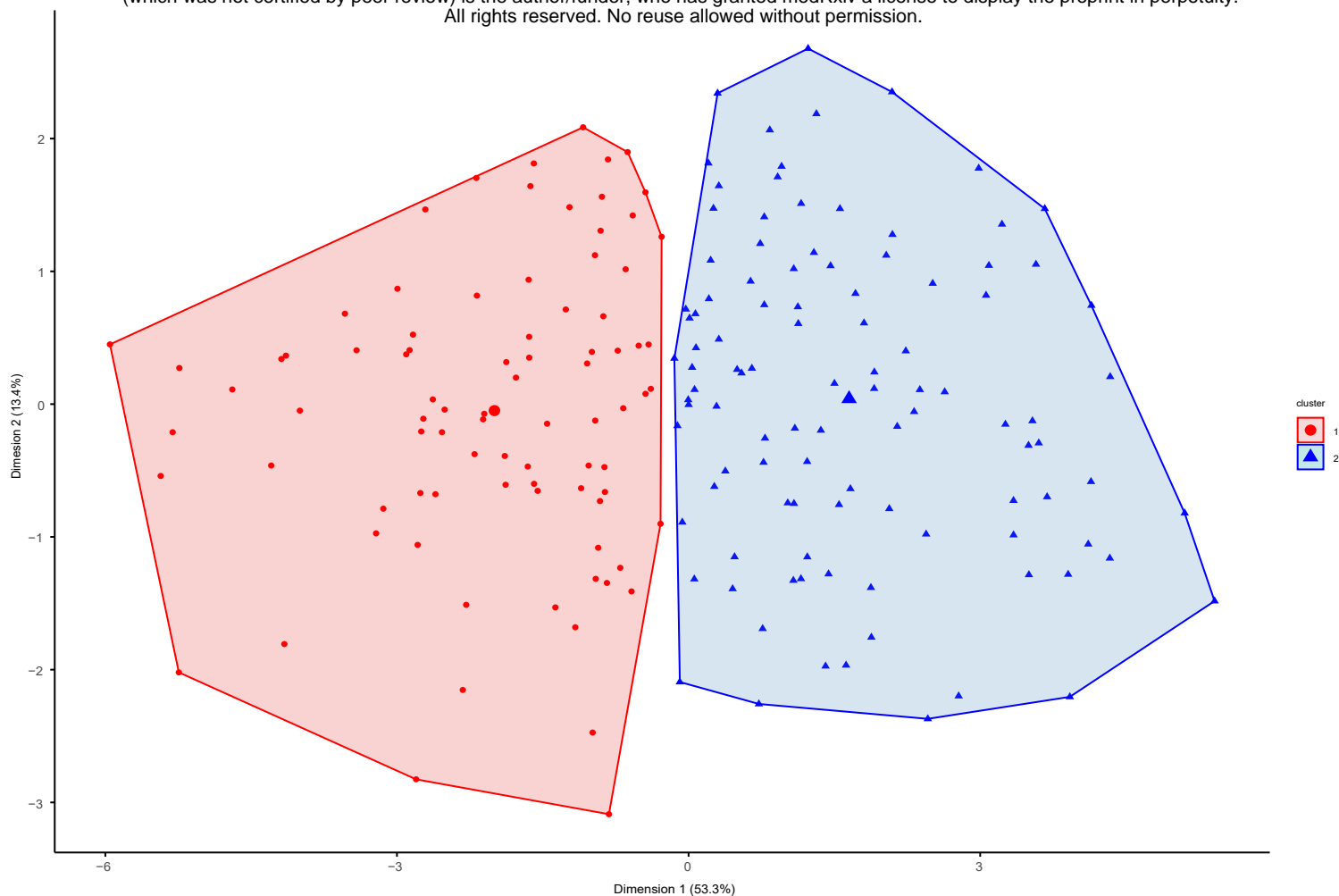
A



B

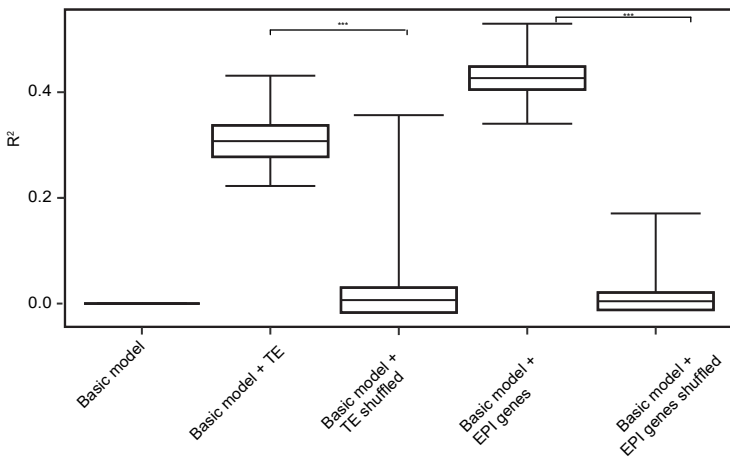


Extended Data Figure 6. Correlation between IKZF1 expression and B cell infiltrates and between eight significant TEs detected in our model. **A.** Correlation between IKZF1 expression and B cell infiltrates. **B.** Pearson correlation among expression of 6 TEs. Scale from -1 (inverse correlation, blue) to 1 (positive correlation, red). Areas of circles represent the absolute value of corresponding correlation coefficients.

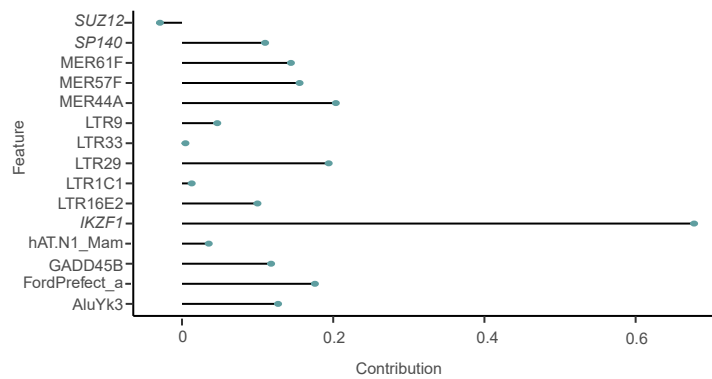


Extended Data Figure 7. Determination of immune clusters from MCP-counter-based immune deconvoluted cell proportions in the TCGA cohort. Factor map representing two clusters based on hierarchical clustering of principal components. Each dot represents an individual patient sample.

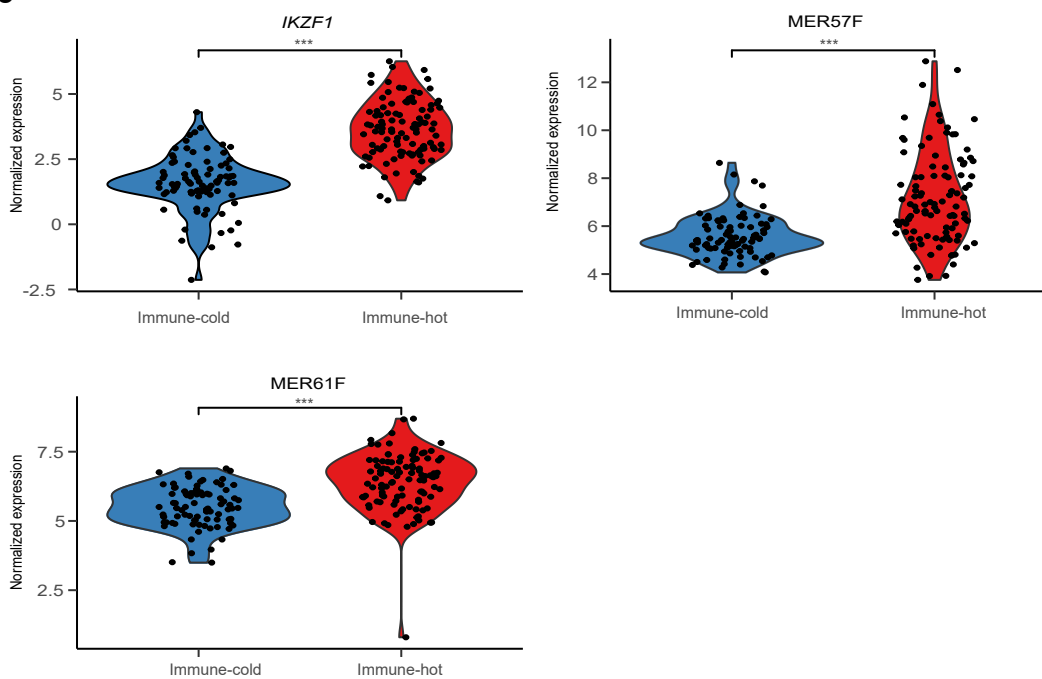
A



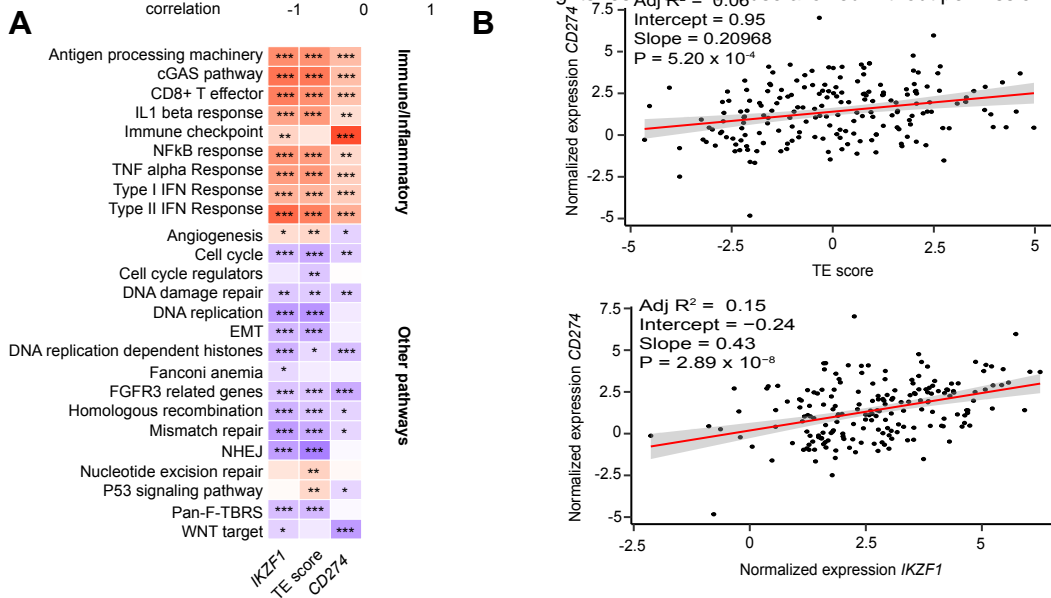
B



C



Extended Data Figure 8. TEs and IKZF1 expression predict immune groups in the TCGA cohort. A. Comparison of performances (R^2) of 5 lasso logistic regression model models. Each boxplot represents a different model (basic model, bootstrapped basic model + TE, and bootstrapped basic model + Epigenetic genes (EPI), and error bars represent 95% confidence intervals. Shuffled models for TE and EPI are also shown. Difference between bootstrapped and shuffled model is shown as result of t-test (***, p -value < $2.2e-16$). **B.** Contribution of significant features from the TE and epigenetic models (models with the highest R^2) represented as non-zero coefficients. Significant TE families and epigenetic genes (italicized) are shown. **C.** Violin plots of normalized expression of transcripts identified as significant features in the regression model in immune-hot and -cold clusters. *** represents $p < 0.001$ as determined by two-sided t-test.



Extended Data Figure 9. TEs and IKZF1 expression correlate with immune pathway expression in the TCGA cohort. **A.** Heatmap of partial Pearson correlation including batch and histology as covariates. Scale from -1 (inverse correlation, blue), to 1 (positive correlation, red). Asterisks indicate Benjamini-Hochberg-corrected p-values: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **B.** Correlation between CD274 (PD-L1) gene expression and TE score (top) and IKZF1 expression (bottom).