SCAN-ACT: ADOPTIVE T CELL THERAPY TARGET DISCOVERY THROUGH SINGLE-CELL TRANSCRIPTOMICS

Stefano Testa[1,2], Aastha Pal[3], Ajay Subramanian[3], Sushama Varma[4], Jack Pengfei Tang[5], Danielle Graham[6], Sara Arfan[7], Minggui Pan[8], Nam Q. Bui[8], Kristen N. Ganjoo[8], Sarah Dry[9], Paul Huang[7], Matt van de Rijn[4], Wei Jiang[10], Anusha Kalbasi[3,11*], Everett J. Moding[3,11*]

[1]Department of Medicine, Stanford University, Stanford, CA, USA

[2]Current affiliation: Division of Hematology and Oncology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA

[3]Department of Radiation Oncology, Stanford University, Stanford, CA, USA

[4]Department of Pathology, Stanford University, Stanford, CA, USA

[5]Department of Pediatrics, Boston Children's Hospital, Boston, MA, USA

[6]Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[7]Division of Molecular Pathology, The Institute of Cancer Research, London, UK

[8]Division of Oncology, Department of Medicine, Stanford University, Stanford, CA, USA

[9]Department of Pathology, University of California, Los Angeles, CA, USA

[10]JWE Science Technology, Inc., Bejing, China

[11]Stanford Cancer Institute, Stanford University, Stanford, CA, USA

*Corresponding Authors

Contact Information

Everett J. Moding, M.D., Ph.D., Assistant Professor, Department of Radiation Oncology, Stanford Cancer Institute, 875 Blake Wilbur Drive, Stanford, CA 94305-5847. Email: emoding@stanford.edu. Tel: 650-498-1625.

and

27 Anusha Kalbasi, M.D., Associate Professor, Department of Radiation Oncology, Stanford
28 Cancer Institute, 875 Blake Wilbur Drive, Stanford, CA. Email: akalbasi@stanford.edu. Tel: 650-
29 498-4796.

Summary

The FDA approval of T cell receptor-engineered T cells (TCR-T) for synovial sarcoma demonstrates the potential for adoptive T cell therapies (ACTs) in solid tumors. However, the paucity of tumor-specific targets without expression in normal tissues remains a major bottleneck, especially in rare cancer subtypes. Here, we present a comprehensive computational pipeline called SCAN-ACT that leverages single cell RNA sequencing and multi-omics data from tumor and normal tissues to nominate and prioritize targets for both chimeric antigen receptor (CAR)- and TCR-T cells. For surface membrane targets, SCAN-ACT proposes target pairs for bispecific Boolean logic-gated CAR T cells. For peptide-MHC targets, SCAN-ACT proposes intracellular peptides bound to a diverse set of human leukocyte antigens. We applied the SCAN-ACT pipeline to soft tissue sarcoma (STS), analyzing 986,749 single cells to identify and prioritize 395 monospecific CAR-T targets, 14,192 bispecific CAR-T targets, and 5,020 peptide-MHC targets for TCR-T cells. Selected targets were validated experimentally by protein expression and for peptide-MHC binding. Proposed targets and target pairs reflected the mesenchymal, neuronal, and hematopoietic ontogeny of STS. This work provides a robust data repository along with a web-based and user-friendly set of analysis tools to accelerate ACT development for solid tumors.

Keywords

Cellular therapy, CAR-T cells, TCR-T cells, Peptide-centric CAR-T, logic-gated CAR-T, Single cell RNA sequencing, Soft tissue sarcoma, Immunotherapy

Introduction

Despite profound success in the treatment of hematologic malignancies, engineered adoptive T cell therapies (ACT) have had limited success in solid tumors[1–3]. While novel engineering strategies can address the major hurdles related to T cell function[4], including poor tumor

56 infiltration[5], terminal differentiation, effector dysfunction, and persistence or memory formation[6–
57 9], a lack of tumor-specific targets with uniformly high expression and low risk of on-target off-
58 tumor toxicity remains a challenge[10,11]. Advances in bispecific antigen targeting, synthetic gene
59 circuits, and Boolean logic-gating in engineered T cells have opened new opportunities to
60 improve antigen-discrimination between normal and tumor cells and increase the targetable
61 antigen pool[4,12,13]. However, there are no repositories of suitable target pairs for Boolean logic
62 CAR-T cells in solid tumors.
63
64 Prior ACT target discovery has focused on bulk RNA-sequencing and protein expression from
65 tumor and normal samples[3,14,15], which is confounded by non-malignant cells in the tumor
66 microenvironment. Similarly, analysis of cancer cell lines[16] may not reflect gene expression in
67 vivo or the heterogeneity of expression between individual cancer cells and across patients.
68 Furthermore, robust bulk transcriptomic and proteomic datasets are not available for many less
69 common human cancers. Compared to bulk RNA-seq and proteomics, scRNA-seq more
70 accurately represents the transcriptomic profile and heterogeneity of individual tumor cells
71 enabling the precise identification of tumor-specific genes[17,18].
72
73 Soft tissue sarcomas (STS) are rare and heterogeneous mesenchymal cancers with high rates
74 of metastatic disease and limited therapeutic options[19]. Although chemotherapy remains the
75 primary treatment for metastatic STS, most patients do not respond and the median overall
76 survival is less than two years[20,21]. Immune checkpoint blockers (ICBs) targeting programmed
77 cell death protein 1 (PD-1) and cytotoxic T-lymphocyte antigen-4 (CTLA-4) have shown promise
78 in specific STS histotypes[22–24], but are ineffective for most patients[25,26]. Recent studies have
79 shown promising efficacy for TCR-T cells in STS. Afamitresgene autoleucel (afami-cel) targeting
80 the cancer-testis antigen MAGE-A4 in patients with an HLA*02:01 allele achieved a 37% overall
81 response rate in heavily treated patients with synovial sarcoma (SS) and myxoid round cell

liposarcoma (MRCLS)[27,28], leading to the first FDA approval of a genetically engineered ACT in solid tumors. Likewise, the most promising results for TCR-T cell products targeting NY-ESO-1[29,30] have been reported in patients with STS[31]. However, additional targets will be critical to expand the availability of ACTs to patients with STS that do not express MAGE-A4 and NY-ESO-1. Furthermore, most patients lack the HLA-A*02:01 allele, which is most frequently expressed in Europeans, raising concerns about the equity of current ACTs[32]. Taken together, STS represent a challenging disease model with known effective ACT targets that can be used to benchmark target identification pipelines.

Here, we leveraged single cell RNA sequencing (scRNA-seq) from patient tumors and collated normal tissue and tumor genetic, transcriptomic, and proteomic data across public databases to develop a comprehensive computational pipeline for discovery of ACT targets in solid tumors, called SCAN-ACT (Single Cell Analysis Network for Adoptive Cellular therapy Targets). We applied SCAN-ACT in STS to build CellSARCTx (Cellular SARComa Therapy), a database describing and prioritizing over 20,000 TCR-T and CAR-T cell targets, including those for Boolean logic-gated and peptide-centric CAR-T cells (PC-CAR-T). Proposed targets captured the shared mesenchymal, neuronal and hematopoietic gene expression signatures of STS, reflecting the intertwined developmental ontogeny of these three lineages from the mesoderm and neural crest. We validated protein expression of select targets by IHC and binding of peptides derived from intracellular proteins to HLA molecules with a novel yeast display technology. We present two interactive web interfaces to explore and evaluate the identified STS ACT targets (CellSARCTx) and to apply SCAN-ACT to new scRNA-Seq samples (SCAN-ACT Custom). SCAN-ACT can be applied to individual patient samples to aid in therapy selection or broader panels of patients to identify conserved targets and accelerate ACT development. SCAN-ACT is freely accessible for academic use at scanact.stanford.edu.

Results

Overview of the SCAN-ACT pipeline

SCAN-ACT identifies tumor up-regulated and down-regulated genes through differential gene expression (DGE) analysis using scRNA-seq (**Figure 1A**). Tumor up-regulated genes are then compared against clinically tested CAR-T targets to identify novel targets with predicted low risk of on-target off-tumor toxicity. SCAN-ACT then defines Boolean logic-gated CAR-T target pairs by measuring tumor and normal cell/tissue co-expression patterns. Tumor-specific MHC-restricted peptides ($pMHCs$) targetable by TCR-T or PC-CAR-T cells are identified by integrating peptide immunogenicity and HLA presentation predictions with $pMHC$ cross-reactivity assessment. Lastly, SCAN-ACT prioritizes targets by integrating target homogeneity with normal tissue/cell expression, and functional, genomic, and proteomics tumor data.

scRNA-seq for cellular therapy target identification in STS

To identify ACT targets across STS histologies, we applied SCAN-ACT to newly sequenced and publicly available scRNA-seq data[17,33] from both translocation-driven and non-translocation-driven STS (**Table S1**). Our scRNA-seq dataset included 32 samples from 25 patients, totaling 128,122 cells after quality control (**Figure 1B, Figure S1A**, **S1B**, **S1C**, **S1E, S1G, S1I**). We included visceral (31%) and extremity/trunk (69%) tumor samples from male (72%) and female (28%) patients aged 22 to 77 years old. Samples included both primary (75%) and metastatic tumors (25%) collected before (84%) or after treatment (16%). Sarcoma cells were identified among the $CD45^{-}$ cells (PTPRC count = 0), using inferred copy number abnormalities[34,35] and transcriptional similarity to bulk samples of the same histology from The Cancer Genome Atlas (TCGA)[36] (Similarity Score)[33] (**Figure S1D, S1F**, **S1H**, **S1H**, **S1J**). Within the tumor microenvironment (TME), endothelial cells (ECs), myeloid cells (MCs), lymphoid cells (LCs), and cancer-associated fibroblasts (CAFs) were defined using canonical markers from the literature (**Table S2**).

134

135 **Identification of genes differentially expressed by malignant cells within the STS TME**

136 To identify tumor-specific genes, SCAN-ACT performs DGE analysis between sarcoma cells
137 and non-tumoral cells from the TME using Limma-voom[37], a pseudo-bulk method, and distinct[38],
138 a differential distribution method (**Figure 1C**). We also compared sarcoma cells with normal cell
139 clusters from 30 scRNA-seq datasets from the Human Protein Atlas (HPA)[39,40], Tabula Sapiens
140 (TS)[41], and the Human Cell Atlas (HCA)[42] covering the major normal human tissues (858,627
141 cells and 164 samples, **Figure 1D**). Unsupervised hierarchical clustering demonstrated that
142 STS were most similar to CAFs and ECs within the TME (**Figure S2A**, **S2C**) and mesenchymal
143 cells in normal tissues (**Figure S2B**, **S2D**), consistent with their putative mesenchymal origin. In
144 total, we identified 29,322 up-regulated genes and 15,812 down-regulated genes in STS cells
145 (**Table S3, S4**). SS and LMS had the most unique up-regulated and down-regulated genes,
146 respectively, while UPS, LPS, and MFS shared the most up-regulated and down-regulated
147 genes consistent with their biological similarity[36] (**Figure 1E**, **Figure S2E**).

148

149 There was concordance between Limma-voom and distinct for DGE analysis in the TME
150 (**Figure 1F**), with distinct identifying more unique differentially expressed genes. By measuring
151 differences in gene expression distribution between cell types, distinct identifies subtle
152 differences in expression[38], which can be advantageous when comparing transcriptionally
153 similar cells like STS and other mesenchymal cells in the TME. This approach identified a pool
154 of sarcoma-specific genes, including many like *FAP* or *MMP14* in UPS which are up-regulated
155 in STS compared to almost all normal cell types (**Figure 1G**). Notably, *FAP* has been targeted in
156 solid tumors due to its TME expression[43], though we found that *FAP* has intrinsic expression in
157 STS cells, particularly LPS, MFS, and UPS, consistent with IHC data from UPS patient samples
158 (**Figure 1H**), where 31/57 (54%) patients had *FAP* expression in ≥ 75% of cells (**Figure S2F,**

159 **S2G, S2H**). Other targets classically associated with the stroma that were upregulated in
160 malignant cells within the STS TME include *MRC2*, *CD248*, and *PDGFRB*.
161
162 **Monospecific CAR-T cell targets in STS**
163 The ideal monospecific CAR-T cell target would have homogeneously high expression on
164 cancer cells with low to no expression on normal cells to avoid on-target off-tumor toxicity[44].
165 However, development of CAR-T cells for solid tumors has been hampered by the scarcity of
166 antigens that meet these conditions with few targets tested in humans[10]. To address this, we
167 filtered the STS up-regulated genes to eliminate genes frequently expressed more highly in
168 normal cells than STS cells. We then selected genes coding for surface proteins with high
169 expression in STS by cross-referencing our scRNA-seq data with the TCGA STS bulk RNA-seq
170 data (TCGA SARC) and the HPA transcriptomics data in STS cell lines. Finally, we removed
171 targets with high RNA and protein expression in normal cells and tissues using CAR-T targets
172 tested in humans as references for maximum tolerable on-target off-tumor toxicity risk (**Figure**
173 **2A**, **Table S4**, **Table S5**, **Table S6**, **Table S7**)[45].
174
175 We identified targets with RNA and protein expression profiles in normal tissues that mimic the
176 normal tissue profiles of clinically tested CAR-T targets (**Figure 2B, Figure S3A**). Our approach
177 reassuringly identified several targets currently in clinical trials for sarcomas, including *IGF1R*[46],
178 *CD163*, *EPHB4*, *FGFR1, KDR*, *NCAM1*, *SMO*, *PSEN1* and *PSEN2*[47], and *NT5E*/CD73[48].
179 Overall, we found 395 unique targets, with 87 common to all sarcomas and 41 common to LPS,
180 MFS, and UPS (**Figure S3B**, **Table S8, Figure S3C**, **S3D**). LPS and SS were the histotypes
181 with the highest number of unique targets with 30 and 21, respectively.
182
183 In each sarcoma, transmembrane enzymes were the most common class (**Figure 2C**, **Table**
184 **S9**), including metalloproteases of the ADAM family[49] (*ADAM10*, *ADAM33*), metallopeptidases

185 (*MME*), serine proteases (*PRSS21*, *FAP*), and protein tyrosine phosphatases (*PTPRJ*, *PTPRB*).
186 Other common target classes were glypicans[50] (*GPC3*, *GPC6*), immune checkpoint molecules[51]
187 (*CD276*, *NECTIN2*, *BTN2A2*), protein tyrosine kinases[52] (*FGFR1*, *ERBB4*, *EFNA4*), NOTCH
188 family receptors[53] (*NOTCH2*, *NOTCH4*), G-protein coupled receptors – GPCRs[54] (*ADGRL2*,
189 *GPR107*), cell adhesion molecules of the Immunoglobulin superfamily – Ig-like CAMs (*CLMP*,
190 *NEO1*), semaphorins[55] (*SEMA4A*, *SEMA6D*), cytokine receptors (*IL1R1*, *IFNAR2*), c-type
191 lectins[56] (*MRC1*, *MRC2*), and solute transporters (*ABCA1*, *MMGT1*). Gene set over-
192 representation analysis (ORA) of the 395 targets revealed enrichment in gene sets associated
193 with cell–matrix interaction, axon and neural projections growth, cell mobility and migration, and
194 solute transport across the plasma membrane (**Figure 2D**, **Figure S4A–E, Table S10, S11**). We
195 observed moderate correlation between bulk RNA-seq expression of CAR-T targets in TCGA
196 SARC and scRNA-seq expression (**Figure S3E**). For many targets like *CLMP, PDGFRB,*
197 *NOTCH2,* and *GPC6*, TCGA SARC samples had the highest median expression compared with
198 other cancer types in the TCGA Pan-Cancer Atlas (**Figure 2E**, **Figure S4F**).
199
200 To prioritize targets, we identified those essential for sarcoma cell line growth in vitro by
201 analyzing publicly available CRISPR KO screening data from DepMap[57] (**Figure 2F**, **Table**
202 **S12**). Twenty-two genes were essential for at least one sarcoma cell line, with some being more
203 selective for specific STS, like *TSPAN31* for LPS. Other genes were essential for multiple
204 sarcoma cell lines, like *PTK2*, *VPS28*, and *UBE2D3*. We then studied the frequency of CNAs,
205 gene fusions, and SNVs and small insertions/deletions (indels) involving each target, using
206 publicly available DNA sequencing data (TCGA SARC[36] and MSK-IMPACT, **Table S13, S14,**
207 **S15, S16**)[58]. We found targets that were frequently amplified in specific STS histotypes in TCGA
208 with high correlation between their CNA level and RNA expression, indicating that these events
209 could have a driver effect[59,60] (**Figure 2G**). For example, *TSPAN31* (r = 0.89) was amplified in
210 72.4% of DDLPS, and *FGFR1* (r = 0.68) was amplified in 12.5% of UPS. Integrating data from

211 TCGA and MSK-IMPACT with the results of a search in the public Mitelman Database[61],
212 encompassing 52 STS types and 3283 unique fusions (**Figure 2H**, **Table S15**), we found 213
213 fusions across 20 STS histologies that involved at least one target. The CAR-T targets most
214 altered by SNVs or indels included *MACF1* and *AXL* in UPS/MFH (10.2% and 6.1% in TCGA,
215 **Figure 2I**, **Table S16**).
216
217 **Bispecific CAR-T targets in STS: 'AND' and 'AND–NOT' logic-gated CARs**
218 To increase the surface antigen pool targetable by CAR-T cells in solid tumors, new bispecific
219 CAR-T designs have been developed. Specifically, Boolean-logic 'AND'-gated CAR-T cells[62] like
220 the logic-gated intracellular network (LINK) CAR-T[12], and the synthetic NOTCH receptor
221 (synNOTCH) CAR-T[63] cause T-cell activation only when two separate antigens are recognized
222 on target cells. Conversely, inhibitory CAR-T cells (iCARs)[64] are an example of Boolean-logic
223 'AND-NOT'-gated CAR-T cells, where an inhibitory domain prevents T-cell activation upon
224 recognition of an antigen selectively expressed on normal but not on tumoral cells.
225
226 To discover targets for 'AND' CAR-T cells in STS, we sought to identify genes with high co-
227 expression in STS cells and low to no co-expression in normal cells and tissues. (**Figure 3A,**
228 **Figure S5A**). To this end, we measured co-expression of STS up-regulated genes in tumor cells
229 at the single cell level using CS-CORE[65] (**Figure S5B**), and in normal cells and tissues using
230 the Genotype-Tissue Expression (GTEx)[66] bulk RNA-seq and proteomics datasets, the Human
231 Proteome Map (HMP)[67] proteomics dataset, and the HPA scRNA-seq, bulk RNA-seq, and IHC
232 datasets. We then compared tumor co-expression to normal tissue co-expression and
233 calculated the $TS_{'AND'}$ score as a measure of targetability for each 'AND' logic-gated CAR-T
234 gene pair, where pairs with $TS_{'AND'} \geq 1$ were those with better targetability. Although we found
235 3608 unique 'AND' CAR-T pairs (**Figure 3B, Table S17**), only 4 (0.1%) were common to all

236 sarcoma histotypes, highlighting the histotype specificity of gene pairs as well as the challenge
237 of identifying broadly applicable target pairs for 'AND' logic-gated CAR T cells.
238
239 Most 'AND' CAR-T pairs consisted of genes with restricted expression in orthogonal normal cell
240 types but shared expression in STS. Consistent with the mesenchymal cell origin of STS[68],
241 gene pairs from "Blood & Immune Cells" and "Neuronal Cells" were the most frequent in all STS
242 histotypes reflecting the shared development of mesenchymal, hematopoietic, and neuronal
243 cells from the embryonic mesoderm and neural crest[69] (**Figure 3C, Figure S5C–F**). We
244 previously observed expression of epithelial markers in malignant STS cells[17], and gene pairs
245 from both epithelial and mesenchymal cells were identified as targets for 'AND' CAR-T cells.
246 These results highlight the potential for bispecific CAR-T cells to leverage the dysregulated
247 differentiation of malignancies as a therapeutic target[70].
248
249 For 'AND-NOT' CAR-T target pairs, the ideal combination would include an activator highly
250 expressed on cancer cells, and an inhibitor with no expression on cancer cells but always co-
251 expressed with the activator on normal cells (**Figure 3D, Figure S5A**). To find activator-inhibitor
252 target pairs for 'AND-NOT' CAR-T cells, we used highly expressed STS up-regulated genes as
253 activators and STS down-regulated genes with average scRNA-seq tumor expression $\leq 50^{th}$
254 percentile as inhibitors and measured their expression in normal cells and tissues. We
255 combined the results of the co-expression analysis in each dataset to obtain a summative
256 targetability measure, the $TS_{'AND\text{-}NOT'}$ score, with values raging from -1 to 1, where higher values
257 indicate higher co-expression in normal tissues/cells. Overall, we found a total of 10,584 'AND-
258 NOT' CAR-T pairs across all sarcomas (**Figure 3E**, **Table S18**). Of these, 79 (0.7%) pairs were
259 shared across all five STS, including *TSPAN6–CLDN17*, *PDGFC–GPR33*, and *IL17RD–ZNRF4.*
260 Pairs with genes belonging to the same HPA single cell gene signatures were the most frequent,
261 reflecting similar patterns of expression in normal cells (**Figure 3F**).

262

263 Considering the ‘AND’ and ‘AND-NOT’ CAR-T target pairs with high *TS* scores, again we found
264 that most ‘AND’ pairs tend to have similar expression profiles in the same cells and tissues,
265 while ‘AND-NOT’ pairs show opposite expression in the same tissues and cells, which was true
266 both at the protein and RNA level (**Figure 3G-H**). Examples of ‘AND’ pairs include *NCSTN*–
267 *BCAN* (“Blood & Immune Cells” – “Glial Cells”), *TM9SF1–CD81* (“Blood & Immune Cells” –
268 “Glandular Epithelial Cells”), and *NTRK2–CD99* (“Neuronal Cells” – “Mesenchymal Cells”).
269 Examples of ‘AND–NOT’ CAR-T target pairs include *PTPRC–ITGAM* (“Blood & Immune Cells”),
270 *PTPRZ1–NRXN3* (“Glial Cells” – “Neuronal Cells”), and *ADAM2–LY6K* (“Germ Cells”). Many of
271 the bispecific CAR-T targets were essential for STS cell line growth (**Figure S6A**) and/or were
272 genetically altered in human STS (**Figure S6B-H**, **Table S19-S21**), suggesting their importance
273 in STS oncogenesis.

274

275 **Targeting intracellular antigens in STS: TCR-T and PC-CAR-T cells**

276 CAR-T cells target cell surface antigens but cannot directly recognize intra-cellular proteins,
277 which constitutes a large fraction of cancer antigens[71]. To address this issue, TCR-T and
278 peptide-centric CAR-T (PC-CAR-T)[72] cells have been developed targeting peptides presented
279 by the major histocompatibility complex (MHC) on tumor cells. Most peptides targeted by TCR-T
280 cells tested in humans so far derive from cancer germline antigens (CGAs)[73], which are highly
281 expressed in cancer and reproductive organs, such as *MAGEA4* or *CTAG1B,* coding for the NY-
282 ESO-1 protein[27,30]. SCAN-ACT uses DGE analysis to identify tumor cell up-regulated genes
283 coding for either intra- or extracellular proteins, with normal tissue expression similar to CGAs
284 targeted by TCR-T cells tested in humans to find tumor-specific genes with expression restricted
285 to immune-privileged sites thus minimizing on-target off-tumor toxicity (**Figure 4A**, **Table S22**).

286

287 SCAN-ACT identified 76 genes across five STS, with some showing uniformly high expression,
288 and others more specific to one or few histologies (**Figure 4B**, **Table S23**). UPS had the highest
289 number of targets (45), followed by LMS (35), LPS (28), SS and MFS (26), with 7 shared across
290 all histologies (*ACTRT3*, *ADAM12*, *ADAM32*, *CXCL16*, *MAP10*, *MSX1*, and *SUV39H2*, **Figure**
291 **S7A**). Overall, 34/76 antigens (44.7%) were known CGAs[73], and gene set ORA identified a roles
292 of the targets in gametogenesis and gonadal cells function (**Figure 4C**). Few of the targets were
293 essential for STS cell line growth in vitro (**Figure S7B**), amplified in human STS (**Figure S7C**,
294 **Table S24**), or involved in SNVs/Indels (**Figure S7E**, **Table S25**). However, 20 gene fusions in
295 STS involved at least one target (**Figure S7D**).
296
297 Next, we identified a pool of $pMHCs$ suitable as TCR-T or PC-CAR-T targets for each antigen.
298 Specifically, we assessed the HLA presentation and immunogenicity potential of all the 9-mer
299 and 10-mer peptides derived from each antigen, using HLAthena[74] and DeepImmuno[75],
300 respectively, focusing on 9-mers and 10-mers as these account for more than 97.5% of HLA
301 class I-bound peptides[76]. Also, we chose the top 10 most frequent HLA alleles in each major
302 geographical area as reported in the Allele Frequency Net Database (AFND)[77], accounting for
303 118 HLA alleles (**Table S26**). A major limitation of TCR-T cells is their ability to recognize
304 peptides presented on one or rarely two HLA alleles, making them accessible to a small fraction
305 of the population. To address this, we combined the HLAthena and DeepImmuno predictions
306 with the average global HLA allele frequency of each $pMHC$ to obtain a combined TCR-T/PC-
307 CAR-T targetability score, $TS_{TCR/PC}$. We retained the top 5,020 $pMHCs$ formed by 348 peptides
308 and 115 HLA alleles with 322 $pMHCs$ shared between ≥ 1 antigen (**Table S27**). *ADAM32* had
309 the highest number of targetable $pMHCs$ (*n* =136), covering 64 HLA alleles (**Figure S7F**). We
310 found a higher number of targetable $pMHCs$ for HLA-C (*n* = 2,041), compared to HLA-B (*n* =
311 1,540) and HLA-A (*n* = 1,439). The top three alleles by number of targetable $pMHCs$ were

312 C*07:01 (*n* =168), C*06:02 (*n* =145), and C*07:02 (*n* =131), while C*04:01 had $pMHCs$ with the
313 highest median $TS_{TCR/PC}$ (2.1) (**Figure S7G**).
314
315 Because TCR degeneracy allows T-cells to recognize different peptides presented on the same
316 MHC molecule[78], we designed a $pMHC$ cross-reactivity prediction algorithm called PS-TRACT
317 (Peptide Similarity and TCR Reactivity Analysis for Cellular Therapies). PS-TRACT measures
318 the similarity of each tumoral peptide to peptides of the same length in the normal MHC
319 ligandome (**Supplementary Data**) while considering the tissue-specific expression and
320 likelihood of presentation of each cross-reactive peptide. Peptide similarity was measured in
321 terms of amino acid physicochemical[79] and HLA binding energy covariance similarity[80], given
322 that peptide sequence identity alone cannot fully predict T-cell cross-reactivity[81–83]. Also,
323 because not all peptide residues bear the same importance in mediating TCR recognition[84], we
324 applied an entropy-based weighting, postulating that residues with higher entropy could be more
325 relevant for $pMHC$ TCR recognition[74].
326
327 We validated PS-TRACT on experimentally observed cross-reactive peptides, including tumoral
328 and microbial epitopes[85], and the *TTN* peptide ESDPIVAQY which was found to be the off-tumor
329 target in cases of fatal myocarditis in patients receiving TCR-T cells directed at the *MAGEA3*
330 epitope EVDPIGHLY[86] (**Figure S8A**). In each known cross-reactive pair, the target peptide was
331 scored for similarity both against its known cross-reactive peptide and against all the normal
332 peptides of same length binding to the same MHC molecule, used as controls. As expected,
333 known cross-reactive peptides had peptide similarity scores (S_{sim}) ≥ 99.9th percentile
334 (**Supplementary and Source Data**). We applied PS-TRACT to each STS $pMHC$ and quantified
335 cross-reactivity scores, S_{CR}, for each tissue in the HPA/GTEx consensus database
336 (**Supplementary Data**).

To test if PS-TRACT can predict tissue-specific TCR cross-reactivity, we applied it to $pMHC$ targets for which TCR-T cells have been tested in clinical trials with safety outcome data available[11] (**Figure S8B**, **Table S28**). We selected the S_{CR} score for the $pMHC$ target of afami-cel, GVYDGREHTV–HLA-A*02:01, as the upper limit for our analyses given afami-cel's promising safety profile in clinical trials[27,28]. We found that $pMHCs$ for which cross-reactivity was observed in clinical trials[11] had higher tissue-specific S_{CR} scores compared to GVYDGREHTV–HLA-A*02:01, as well as higher overall S_{CR} profile difference ($S_{CR}Diff$) with GVYDGREHTV–HLA-A*02:01 when considering all tissues (**Figure S8C**). For example, the *MAGEA3*-derived $pMHC$ KVAELVHFL–HLA-A*02:01, targeted by TCR-T cells causing two cases of fatal necrotizing leukoencephalopathy[87], had a S_{CR} in the cerebral cortex which was 2.4 standard deviations above the S_{CR} of GVYDGREHTV–HLA-A*02:01 with the cross-reactive peptide identified as AVADLVHIV from *MCHR2,* which is highly expressed on neurons. Similarly, we observed a high cardiac S_{CR} score for EVDPIGHLY–HLA-A*01:01 driven by the known cross-reactive *TTN* epitope ESDPIVAQY[86].

Using PS-TRACT, we found 1113 STS $pMHCs$ with $S_{CR}Diff \leq 0$ across 182 peptides, 69 antigens, and 113 HLA alleles, and of these 74 were shared between ≥ 2 antigens (**Table S27**). We found several peptides predicted to form $pMHCs$ with $S_{CR}Diff \leq 0$ with ≥ 10 different HLA alleles for either the class A, class B, or class C locus (**Figure 4D**). Such MHC-promiscuous peptides could be targetable with a PC-CAR-T recognizing the same peptide presented on different MHC molecules as long as the molecular surface of the resulting $pMHCs$ is similar. For example, YVVDGQIIIY derived from both *PSG8* and *PSG11* bound to 17 class C (total allele frequency: 70%), 10 class B (total allele frequency: 8%), and 4 class A alleles (total allele frequency: 6%). Lastly, we found peptides forming $pMHCs$ with very low risk of on-target off-

362 tumor toxicity ($S_{CR}Diff \leq 10^{th}$ percentile) with highly prevalent HLA alleles (**Figure 4E**). These
363 peptides could be ideal TCR-T targets, given the increased TCR dependency on HLA restriction,
364 and could be targeted in a larger fraction of the population due to the high prevalence of their
365 HLA alleles. For example, YFKDKRVHHF from *DAZL* and QFYDNKGITY from *SUV39H2* were
366 both targetable on C*07:01 (allele frequency: 14%).
367

368 **Validation of peptide-MHC binding for high-priority targets**

369 Most $pMHC$ binding prediction tools rely on datasets from mass spectrometric measurements of
370 eluted peptide ligands (EL method), which require further verifications using the actual binding
371 measurements (BA method) to eliminate false identifications[88]. Here, we adopted a recently
372 developed empirical system, IMMUNE (Identifying Massive MHC Utilized Novel Epitopes), to
373 validate the peptide-MHC binding of targets identified with PS-TRACT. IMMUNE employs a
374 yeast display system enabling fast cloning as well as mammal-like post-translational
375 modifications for the ex-vivo expression of complex HLA heterodimers[89,90]. Through anchorage
376 to agglutinin (formed by Aga1p and Aga2p), a peptide-receptive format of HLA-A, B, or C non-
377 covalent heterodimer can be assembled on the cell surface for antibody detection and peptide
378 accommodation (**Figure 5A**). Using A*02:01, A*11:01, B*35:01, and C*12:03 as examples, we
379 observed the entire heavy chain fusion protein represented by its C-terminal HA staining, the
380 soluble form of β2m in association with the heavy chain, and the correctly folded heavy
381 chain/β2m heterodimer represented by the HLA-A, B, C conformation-specific antibody staining,
382 on the surface of correspondingly transformed yeast strains (**Figure S9A**). The successful
383 detection of each component of the separately expressed heavy chain and β2m without any
384 covalent linkers allowed further development of on-yeast peptide binding assays. Without the
385 necessity to purify HLA proteins, the IMMUNE system substantially expedited BA evaluation and
386 synergized with our computational prediction for rapid TCR-T or PC-CAR-T target discovery.
387

388 By comprehensively analyzing the epitope repertoire of each HLA in the immune epitope
389 database (IEDB), we designed indicator peptides that could be biotinylated without affecting
390 HLA binding. Co-staining of HLA molecules and the exogenously loaded biotinylated indicators
391 confirmed specific binding on the surface of each yeast strain (**Figure S9B**). The capability of
392 IMMUNE to display peptide-receptive HLA mono-alleles represents an ideal system for
393 competitive binding analysis against the indicator. Based on overall S_{CR} profile and $TS_{TCR/PC}$, we
394 selected 9 peptides forming 14 $pMHC$ targets for testing. All test peptides bound A*02:01,
395 including RLHRIIPKI from SSX (**Figure 5B**, **5C**). However, other than AVLDGLDVLL and
396 VLYPVPLESY from PRAME, most peptides showed low to moderately competitive binding for
397 B*35:01 and C*12:03. These results suggest that IMMUNE can be utilized to prioritize pMHCs
398 for targeting.
399
400 **Target expression in STS Proteomics Datasets**
401 Next, we analyzed target expression in STS clinical proteomics databases[91,92], sarcoma cell
402 lines[93,94], and STS tissue micro-arrays (TMAs). Since none of the TCR-T/PC-CAR-T target
403 antigens were measured in the clinical STS proteomics datasets, we analyzed their expression
404 in STS cell lines, as reported in the Cancer Cell Line Encyclopedia (CCLE)[93] and in the ProCan-
405 DepMapSanger[94] repositories. Overall, 33/76 (43.4%) targets were expressed in 21 cell lines
406 spanning 13 STS histotypes (**Figure 5D, Figure S9C**). Of all the targets assessable for
407 expression, *SSX2*/*SSX2B*, *LIN28B*, and *MAGEA4* had the most cell lines with medium-high
408 expression in ProCan-DepMapSanger at 32%, 18% and 14%, respectively. In the Tang et al.
409 STS clinical proteomics dataset[91], we found overall good correlation between scRNA-seq and
410 protein expression for the STS up-regulated surfaceome genes (**Figure S9D**, **S9E**). This
411 correlation was consistent across histotypes when considering the monospecific CAR-T targets,
412 particularly only those with protein expression ≥ 50th percentile (**Figure 5E**, **Figure S9F**). We

413 found that many targets had high expression in a large fraction of samples (**Figure 5F, Figure**
414 **S10A, S10C**). Similar results were observed in the Burns et al. dataset (**Figure S10B**).
415
416 For the 'AND' CAR-T targets with measurable expression of both genes in the Tang et al.
417 dataset, we measured co-expression across all the STS samples (**Table S29**). Of the 1207
418 'AND' pairs, 274 showed positive correlation in protein expression between the two targets
419 across all STS samples (**Figure 5G**). Of the 438 'AND' CAR-T pairs found in the Burns et al.
420 dataset, 79 had positive correlation in expression across all samples (**Table S29**). Notably, we
421 were unable to distinguish tumor and normal cell protein expression from the bulk analysis,
422 which may confound the correlation.
423
424 Lastly, we measured expression of selected targets using in-house TMAs of LMS, SS, MFS,
425 UPS, and MPNST samples (**Figure 5H**). *GPC6* was expressed in 100% (*n* = 10) of SS, 81.4%
426 (*n* = 59) of UPS, and 95% (*n* = 537) of LMS samples, similarly to *MRC2* which was expressed in
427 all SS samples, 84.6% LMS, 79.7% UPS, and 76.1% LPS (*n* = 21) samples. LMS had the
428 highest proportion of positive samples for *KITLG* (91.0%), while SS had the highest number of
429 samples positive for *CLMP* (90.0%), followed by LMS (89.4%). These antigens were selected
430 for IHC assessment due to their overall high STS expression and low normal expression profile.
431
432 **Target Prioritization**
433 We calculated a Target Priority Index (TPI) for each cell therapy target class, integrating normal
434 cell/tissue expression with functional, genomics, and proteomics STS data, while considering
435 single-cell tumor expression homogeneity and existing immunotherapeutic agents (**Table S30-**
436 **S33**). We defined scRNA-seq expression homogeneity as the product of average expression
437 and fraction of cells with non-zero expression (H_{RNA} score). For monospecific and bispecific
438 CAR-T targets, we used the average H_{RNA} score of *CD276* across all histotypes as the

439 homogeneity reference, and the average fraction of samples in the Tang et al. proteomics
440 dataset with *CD276* z-score ≥ 0 as the protein expression reference because *CD276* has been
441 tested as a CAR-T target in STS[95]. Using LPS as an example (**Figure 6A**, **6B**), notable
442 monospecific CAR-T targets in the top 100 TPI_{CAR} included *GPC6* (high protein expression and
443 H_{RNA}), *FGFR1* (essential in LPS cell lines and with ADCs in clinical development[96]), and
444 *TSPAN31* (essential for LPS cell lines and involved by gene fusions and amplifications in LPS).
445
446 For bispecific CAR-T targets, we integrated the average $TS_{,AND,}$ and $TS_{,AND-NOT,}$ for all the
447 targetable pairs in which each gene appears into the TPI, prioritizing genes suitable both as
448 'AND' CAR-T targets and 'AND-NOT' CAR-T activators. In LMS as an example (**Figure 6C**, **6D**,
449 **6E**), notable targets with antibodies developed clinically or pre-clinically include *GPNMB*[97],
450 *RTN4*[98] and *CD46*[99]. Considering all the monospecific and bi-specific CAR-T targets identified,
451 69 had a clinically developed monoclonal antibody (mAb), antibody-drug conjugate (ADC),
452 single-chain variable fragments (scFv), fusion protein, nanobody, or fragment antigen binding
453 region (Fab), while 36 had at least a mAb or ADC in pre-clinical development. Also, of all the
454 CAR-T targets described, 26 and 32 have been previously tested as CAR-T targets either
455 clinically or pre-clinically, respectively.
456
457 Finally, we calculated a TPI for TCR-T and PC-CAR-T target antigens, using *MAGEA4* H_{RNA} in
458 SS as the homogeneity reference, given afami-cel's efficacy in SS. In addition, we considered
459 protein expression in STS cell lines, the total number of target $pMHCs$, and the HLA allele
460 frequency per antigen (**Table S33**). Top targets included *PRAME* and *LY6K*, which have more
461 homogenous expression than *MAGEA4* in SS and a higher total HLA allele frequency. In
462 addition, *SSX1* is recurrently involved in pathogenic fusions in SS and has a higher predicted
463 number of targetable $pMHCs$ than *MAGEA4* (**Figure 6F**, **6G**). Given the large number of targets

464 predicted by SCAN-ACT, the TPI will be helpful to identify the most promising tumor-specific
465 antigens for further testing.
466
467 **Discussion**
468 The major bottleneck for development of ACTs in solid tumors continues to be the lack of tumor-
469 specific antigens with low expression on normal cells[100,101]. Differently from hematological
470 malignancies, solid tumors lack lineage-specific antigens with homogeneous high-density
471 expression. Also, it remains challenging to ascertain an acceptable level of target expression in
472 normal tissues. To address this, we developed SCAN-ACT, a computational pipeline leveraging
473 scRNA-seq to discover targets for TCR-T cells, and monospecific, bispecific, and peptide-centric
474 CAR-T cells, targeting both intracellular and surface proteins.
475
476 We utilized SCAN-ACT to screen for targets in STS, a disease with limited therapeutic options
477 where ACT has proven potential. The characteristic heterogeneity of STS is also an ideal small-
478 scale representation of solid tumor heterogeneity, thus providing a high-bar test case for SCAN-
479 ACT. We screened sarcoma-specific antigens using single-cell DGE analysis between STS
480 cells, non-tumoral cells in the TME, and normal cells across 30 human tissues. scRNA-seq can
481 identify genes that are more selectively expressed in tumoral cells compared to non-tumoral
482 cells in the TME, as opposed to bulk transcriptomics and proteomics[17,18]. Additionally, scRNA-
483 seq allows measurement of target expression heterogeneity, a crucial barrier to ACT efficacy in
484 solid tumors[1], and the availability of normal tissue scRNA-seq data allows for better estimation
485 of the on-target off-tumor toxicity risk[41,42]. Previous studies demonstrated that target expression
486 even on rare cellular populations in vital organs can mediate CAR-T cell toxicity[102], and scRNA-
487 seq may better predict such off-tumor effects than bulk RNA-seq[103].
488

489 We extensively validated the off-tumor toxicity risk of the identified targets using publicly
490 available multiomics data in normal cells and tissues, including scRNA-seq, bulk RNA-seq, IHC,
491 and proteomics, with targets already tested in humans as reference. Through this approach, we
492 identified targets with lower predicted risk of on-target off-tumor toxicity than those currently
493 available. We assessed the relevance of the identified targets to STS biology at the
494 transcriptomic, genomic, and proteomic level, as well as functionally, through CRISPR KO
495 screens. This highlighted *TSPAN31* on chromosome 12, which is frequently involved by CNAs in
496 DDLPS and co-amplified with *MDM2* in both WDLPS and DDLPS[104], indicating a potential
497 oncogenic driver role in LPS.
498
499 We found targets not previously studied in STS like *GPC6,* a member of the glypican family
500 which is pathogenic in sarcomas[105], or *CLMP,* described as tumor suppressor in colon cancer[106].
501 We show targetable pathways and protein families in STS, including semaphorins and plexins
502 that are physiologically involved in neural development[107] and poorly studied in STS[108]; the
503 NOTCH pathway, previously targeted in STS[109,110]; the EPH/Ephrin system, described as
504 pathogenic in sarcomas[111] and physiologically involved in neuron migration, axon guidance, and
505 vascular remodeling[112]; and the WNT/β-Catenin pathway, implicated in sarcomagenesis and
506 previously targeted in STS[113]. Also, we identify stromal antigens, previously targeted to
507 manipulate the TME, as tumor cell intrinsic in STS like *FAP*[43], *FOLR2*[114], or *CD248*[115].
508
509 Here we also provide a tool for screening safe and effective $pMHC$ targets for ACT in solid
510 tumors. We expand the pool of known CGAs targetable in STS[116], identifying antigens not
511 previously described such as *SUV39H2* or *MEIOC*, while confirming the role of known CGAs
512 like *PRAME*, *SSX1*, and those of the MAGE and GAGE families[116]. Targetable $pMHCs$ were
513 identified using computational tools measuring peptide HLA presentation and immunogenicity.
514 Despite being an in silico analysis, this approach prioritizes peptides likely to trigger cytotoxic T-

515 cell responses while maximizing the number of patients who could benefit from targeting a
516 specific $pMHC$. Almost all TCR-T cells tested in humans so far recognize peptides presented on
517 HLA-A*02:01[11], which is highly prevalent in Caucasians, and less so in African Americans and
518 Asians and Pacific Islanders[77]. This creates equity and generalizability issues for HLA-restricted
519 ACTs[32]. To address this, we integrated allele frequency into $pMHC$ target prioritization, including
520 the most frequent HLA alleles from each major geographical area[77]. We curated a normal MHC
521 ligandome database of 46x10^6 $pMHCs$, including 630,022 experimentally validated peptides,
522 and developed PS-TRACT, a tool for $pMHC$ T-cell cross-reactivity assessment. PS-TRACT
523 overcomes limitations of current methods[72,117], allowing predictions for 9-mer and 10-mer
524 peptides over 223 HLA alleles and measuring peptide similarity through biophysical properties
525 while applying an entropy-based weighting that prioritizes amino acids critical for TCR
526 recognition. By using a large pool of experimentally observed and computationally predicted
527 normal $pMHCs$, PS-TRACT provides a multi-tissue cross-reactivity risk estimation for each
528 tumoral $pMHC$. Finally, we experimentally measured HLA binding of selected target peptides
529 using a novel yeast display technology, IMMUNE, especially critical for validation of peptide-
530 binding to less common HLA alleles where in silico predictions may be inaccurate.
531
532 Compared to prior CAR-T target discovery efforts that either performed DGE analysis between
533 tumoral and normal bulk RNA-seq samples[14,15] or analyzed patterns of protein expression in
534 tumor samples[45] or cancer cell lines[16], we leveraged single-cell DGE analysis with a large cohort
535 of non-tumoral reference cells, allowing a more precise screening for tumor-specific genes.
536 Unlike prior studies on bispecific CAR-T targets[14,16,45], we defined tumoral target co-expression
537 at the single cell level and directly measured co-expression in normal cells and tissues using
538 correlation analyses, including normal tissue scRNA-seq and proteomics data. Lastly, compared
539 to prior scRNA-seq studies for CAR-T target discovery[18], we analyzed a larger number of tumor

540 samples, included more normal reference cells, applied robust single-cell DGE analyses less
541 prone to false discovery[118], validated the identified targets with genomics, proteomics,
542 transcriptomics, and functional assays in STS, and identified TCR-T/PC-CAR-T and Boolean-
543 logic gated CAR-T cell targets.
544
545 In summary, we developed a comprehensive tool for the identification of ACT targets in solid
546 tumors enabling the targeting of both intracellular and membrane proteins. By applying the tool
547 for STS, we nominate and prioritize novel putative targets to expand and accelerate the early
548 success of ACT in this disease. Our pipeline and sarcoma ACT target database are freely
549 accessible for academic use at scanact.stanford.edu.
550

551 **Acknowledgments**
552 We thank the patients and families who participated in this study. This work was supported by
553 the My Blue Dots organization (E.J.M.) and the Tad and Diane Taube Family Foundation
554 (M.v.R., E.J.M.). This work used the UCLA Technology Center for Genomics and Bioinformatics
555 and the UCLA Translational Pathology Core Lab. The schematics in Figure 1, 2, 3, 4, and Figure
556 S5 were created with BioRender.com.
557

558 **Author contributions**
559 S.T., A.K., and E.J.M. were responsible for conceptualization. S.T., A.K., and E.J.M. were
560 responsible for the methodology. S.T. and E.J.M. were responsible for software. S.T., A.P., A.K.,
561 and E.J.M were responsible for formal analysis. S.T., A.S., A.K., and E.J.M. were responsible for
562 data curation. S.T., A.K., and E.J.M. were responsible for writing the original draft. W.J. and A.K.
563 were responsible for the IMMUNE validation and related methods and results. S.T., A.P., A.S.,
564 S.V., M.P., N.Q.B, K.N.G., M.v.R., P.H., D.G., S.D., W.J., A.K., and E.J.M. were responsible for
565 review and editing. S.T. A.P., S.A., P.T., S.A., D.G. and E.J.M. were responsible for visualization.

P.H., M.v.R., A.K., and E.J.M. were responsible for supervision. A.K. and E.J.M. were responsible for funding acquisition.

Declaration of Interests

E.J.M. has served as a consultant for GLG and Guidepoint. A.K. serves on the advisory board and holds stock for Dispatch Therapeutics, and Certis Oncology and consults for Sastra Cell Therapy. W.J. serves as CSO for JWE Science Technology, Inc. who owns the intellectual property of the IMMUNE technology. The other authors declare no financial interests.

Data and Code Availability

The processed scRNA-seq data new to this paper have been deposited to GEO with accession code GSE279852 (reviewer token: qbejwckcjnkvbgv). The raw scRNA-seq data new to this project is available on SRA with project code PRJNA1174211 (reviewer link: https://dataview.ncbi.nlm.nih.gov/object/PRJNA1174211?reviewer=rb9d7oupgql7feeple4ml0l7fl). The previously published scRNA-seq datasets reanalyzed in this project are available in GEO with accession codes GSE212527 and GSE131309. The SCAN-ACT pipeline and code used to create the images in the manuscript are available on Github at https://github.com/StefTesta/SCAN_ACT (private repository during submission) and as supplementary material. The SCAN-ACT pipeline and the CellSARCTx interactive interface are fully accessible for academic use at https://scanact.stanford.edu (reviewer/editor username: editor@stanford.edu, password: editor123).

Supplemental Information

Document S1. Figures S1-S10.

Document S2. Excel File containing Supplementary Tables S1-S34.

592 **Methods**

593 **Human subjects**

594 All samples analyzed in this study were collected with informed consent from subjects enrolled
595 on protocols approved by the Institutional Review Boards at Stanford University and University
596 of California Los Angeles.

597

598 **Single cell RNA sequencing of STS samples**

599 STS belonging to the following STS histotypes were included in the study: LMS, MFS, UPS,
600 DDLPS, WDLPS, and SS. Pathological diagnosis was confirmed through intuitional pathologist
601 review for the samples analyzed at Stanford University and University of California Los Angeles,
602 and as reported in the original study for the SS samples[33]. For the MFS2 sample, initial tumor
603 biopsy was diagnosed as MFS, while final surgical specimen was classified as UPS. Due to the
604 ambiguity of diagnosis, and the lower representation of MFS samples in our cohort, we decided
605 to analyze this sample as MFS. Also, the UPS6 sample was classified as Spindle Cell Sarcoma
606 versus UPS and was hence analyzed with the other UPS samples in the study.

607

608 For samples processed at Stanford University, fresh tumors were minced and enzymatically
609 dissociated using the Human Tumor Dissociation Kit (Miltenyi Biotec) using a GentleMACS Octo
610 Dissociator with heaters (Miltenyi Biotec) following the manufacturer's instructions. The
611 dissociated cell suspension was then applied to a 70 um MACS SmartStrainer (130-098-462;
612 Miltenyi Biotec). The quality of the suspension was checked by generating a hematoxylin and
613 eosin (H&E) smear and analyzed using a TC10 automated cell counter (BioRad). Samples were
614 further processed by the Stanford Genomics Shared Resource or the using the Chromium 3'
615 Gene Expression Solution v3.1 (10x Genomics) per the manufacturer's instructions and
616 sequenced on a NovaSeq 6000 (Illumina). For samples processed at the University of California
617 Los Angeles (MFS1/BO112_sarc04, UPS1/BO112_sarc10, UPS2/BO112_sarc11,

618 UPS3/BO112_sarc13, UPS4/BO112_sarc15, UPS5/BO112_sarc17, LPS1/BO112_sarc07,
619 LPS2/BO112_sarc09, LPS2sx/BO112_sarc09_sx, LPS3sx/BO112_sarc16_sx), these were
620 dissociated by treatment with collagenase D (100mg/mL) and DNase I (100mg/mL) using a 100
621 μm cell strainer, following the manufacturer's instructions. After digestion, RBC lysis was
622 performed by adding 1mL of BD lysing buffer (BD Biosciences) and incubating for 5 mins. Then,
623 dead cells were eliminated using a dead cell removal kit (Miltenyi Biotec). Cells were counted
624 using a Countess II cell counter (Thermo Scientific). Samples were then further processed by
625 the UCLA Technology Center for Genomics & Bioinformatics (TCGB) using the Chromium 3'
626 Gene Expression Solution v3.1 (10x Genomics) per the manufacturer's instructions and
627 sequenced on a NexSeq500 High Output instrument (Illumina).
628
629 **Pre-processing of scRNA-seq STS samples**
630 Sample demultiplexing, read alignment to the GRCh38 human reference genome, and
631 generation of feature-barcode matrices were performed using CellRanger v6.0.0 (10x
632 Genomics). Raw data was pre-processed in R (version 4.2.2.)[119] using the Seurat package
633 (version 4.3.0)[120]. Quality control was performed separately for each sample and consisted of
634 excluding cells with number of expressed genes < 200 and percentage of mitochondrial genes >
635 15%. We then identified and excluded doublets using the R package DoubletFinder[121] (version
636 2.0.3) applying the standard workflow as described by the developer.
637
638 **Sample Integration and cell annotation**
639 Log-normalization was performed independently for each sample using the Seurat function
640 NormalizeData. After this step, the Seurat objects for each independent sample were merged
641 preserving the individual normalization of the data. Sample integration and clustering were
642 performed separately for each histotype, accounting for batch effects and differences in
643 sequencing technologies. The top 2000 highly variable genes of the merged and normalized

644 expression matrices were identified using the function FindVariableFeatures, and then centered
645 and scaled using the ScaleData function. Principal component analysis was then run using the
646 top 2000 highly variable genes previously identified. To integrate multiple samples, the harmony
647 R package[122] (version 0.1.0) was employed for batch correction. Harmony was ran using both
648 the sample of origin and the sequencing technology variables as arguments of "group.by.vars",
649 and based on the first 25 PCA dimensions previously identified. UMAP embeddings were then
650 obtained using the first 25 Harmony dimensions, and clusters were obtained by calculating the
651 k-nearest neighbors (k-NN) and a shared nearest-neighbor graph using the Louvain algorithm
652 implemented in the FindClusters function of the Seurat package with a resolution of 1.0. For the
653 LPS, MFS, UPS, and LMS samples, endothelial cells, myeloid cells, lymphoid cells, and CAFs
654 were annotated using the differentially expressed genes discovered with the FindMarkers
655 function in Seurat, which was run with the default parameters except for only.pos = T and
656 min.pct = 0.25. Canonical cell identity markers obtained from the literature were used for cell
657 type annotations. A list of the marker genes for each cell type in our samples is provided in
658 **Table S2**. A small number of cells could not be uniquely assigned to a specific cell identity and
659 were removed from further downstream analysis. UMAP plots were generated with Seurat and
660 scCustomize[123]. Density plots were generated with the R package Nebulosa[124].
661
662 **Public scRNA-seq Datasets**
663 *Synovial Sarcoma Samples*
664 Previously published synovial sarcoma 10x Genomics and SMART-seq2 scRNA-seq data from
665 Jerby-Arnon et al.[33] were downloaded from GEO (GSE131309). After applying quality control
666 filtering as described above, the Seurat objects for each sample were integrated using the
667 harmony[122] pipeline as previously described. Author-provided annotations were used to classify
668 cells as sarcoma cells, endothelial cells, lymphoid cells, myeloid cells, and CAFs.
669

670 *Normal Tissue scRNA-seq Datasets*

671 A list of normal tissue scRNA-seq samples of reference used for differential gene expression
672 was obtained from the HPA website
673 (https://www.proteinatlas.org/about/assays+annotation#singlecell_rna). If for any of the
674 reference datasets listed on the HPA dataset, only one sample was available, additional data
675 was obtained from either Tabula Sapiens[41] or from the Human Cell Atlas[42]. For each dataset, the
676 raw count matrices were downloaded from the publication-specific repositories as follows:
677 Adipose Tissue (GEO: accession code GSE155960); Bone Marrow (GEO: accession codes
678 GSE159929-GSM4850584, and Tabula Sapiens: https://cellxgene.cziscience.com/e/4f1555bc-
679 4664-46c3-a606-78d34dd10d92.cxg/); Brain (The Allen Brain Map: https://portal.brain-
680 map.org/atlases-and-data/rnaseq/human-m1-10x); Breast (GEO: accession code GSE164898,
681 except for D6-D10 samples); bronchus (figshare: accession code fig11981034); Colon (GEO:
682 accession code GSE116222); Endometrium (GEO: accession code GSE111976); Esophagus
683 (GEO: accession codes GSE159929-GSM4850580 and HCA:
684 https://cellxgene.cziscience.com/e/b07fb54c-d7ad-4995-8bb0-8f3d8611cabe.cxg/); Eye (GEO:
685 accession code GSE137537); Fallopian Tube (GEO: accession code GSE178101, excluding
686 hydrosalpinx samples); Heart Muscle (GEO: accession code GSE109816 and Tabula Sapiens:
687 https://cellxgene.cziscience.com/e/e6a11140-2545-46bc-929e-da243eed2cae.cxg/); Kidney
688 (GEO: accession code GSE131685); Liver (GEO: accession code GSE115469); Lung (Tabula
689 Sapiens: https://cellxgene.cziscience.com/e/0d2ee4ac-05ee-40b2-afb6-ebb584caa867.cxg/);
690 Lymph Node (GEO: accession codes GSE159929-GSM4850583 and Tabula Sapiens:
691 https://cellxgene.cziscience.com/e/18eb630b-a754-4111-8cd4-c24ec80aa5ec.cxg/); Ovary
692 (EMBL-EBI: accession code E-MTAB-8381); Pancreas (GEO: accession code GSE131886);
693 PBMC (GEO: accession code GSE112845 and Tabula Sapiens:
694 https://cellxgene.cziscience.com/e/983d5ec9-40e8-4512-9e65-a572a9c486cb.cxg/); Placenta
695 (EMBL-EBI: accession code E-MTAB-6701); Prostate (Tabula Sapiens:

696 https://cellxgene.cziscience.com/e/d77ec7d6-ef2e-49d6-9e79-05b7f8881484.cxg/); Rectum
697 (GEO: accession code GSE125970); Salivary Gland (Tabula Sapiens:
698 https://cellxgene.cziscience.com/e/f01bdd17-4902-40f5-86e3-240d66dd2587.cxg/); Skeletal
699 Muscle (GEO: accession code GSE143704); Skin (GEO: accession code GSE130973); Small
700 Intestine (GEO: accession code GSE125970); Spleen (GEO: accession codes GSE159929-
701 GSM4850589 and Tabula Sapiens: https://cellxgene.cziscience.com/e/cee11228-9f0b-4e57-
702 afe2-cfe15ee56312.cxg/); Thymus (Tabula Sapiens:
703 https://cellxgene.cziscience.com/e/0ced5e76-6040-47ff-8a72-93847965afc0.cxg/); Testis (GEO:
704 accession code GSE120508); Tongue (Tabula Sapiens:
705 https://cellxgene.cziscience.com/e/55cf0ea3-9d2b-4294-871e-bb4b49a79fc7.cxg/); Vascular
706 (Tabula Sapiens: https://cellxgene.cziscience.com/e/a2d4d33e-4c62-4361-b80a-
707 9be53d2e50e8.cxg/). After download, each dataset was pre-processed in a similar fashion as
708 our own STS samples. The pre-processed samples for a given tissue were then integrated with
709 the harmony pipeline[122] using the sample variable, and for the SS samples the sequency
710 technology variable as well, for batch correction, as previously described. Then, unsupervised
711 clustering using the Louvain algorithm with a resolution of 1 was applied for all datasets.
712 Exceptions to this workflow include the brain dataset where the dimensionality reduction and
713 clustering from HPA were used, as well as the endometrium and the bronchus datasets where
714 the author-provided clusters and annotations were used. Author-provided annotations were
715 used whenever possible, and if these were not available, the package SingleR[125] and common
716 gene markers were used for cluster cell-type annotation, where each cluster was assigned to its
717 most frequent cell type. In SingleR[125], the Encode[126] and Blueprint epigenomics[127] datasets
718 were used for annotation of all datasets except for the pancreas for which the scRNA-seq
719 reference data from Baron et al.[128] was used instead.
720
721 **Identification of tumor cells based on InferCNV**

722 CNVs were calculated using the R package inferCNV[35]. Each sample was independently
723 analyzed using as reference cells the previously annotated CAFs, endothelial cells, myeloid
724 cells, and lymphoid cells from the TME. These were identified based on expression of known
725 marker genes available from the literature as described above. The following parameters were
726 used to run the inferCNV function: analysis_mode="subclusters";
727 tumor_subcluster_partition_method="qnorm"; cluster_by_groups= FALSE; k_obs_groups = 3,
728 denoise=T, HMM=TRUE. All the other parameters were set as default. Only for UPS
729 k_obs_groups was set to 2 since this resulted in better overlap with the classification based on
730 comparison to bulk sarcoma tumors (see below). To define a cell as potentially malignant, we
731 used the CNA states across all genes calculated using a hidden Markov model (HMM) as
732 implemented in the inferCNV package. To reduce the number of false positive CNVs identified,
733 a Bayesian latent mixture model was applied with a default p-value threshold of 0.1 for all
734 samples. For each gene, inferCNV reports a CNA state as one of the following: state 1
735 corresponding to a complete allele loss, state 2 corresponding to loss of one allele, state 3
736 corresponding to the neutral state with both allelic copies intact, state 4 corresponding to gain of
737 one allele, state 5 corresponding to gain of 2 alleles, and state 6 corresponding to gain of more
738 than two alleles. For each sample, CNA states of 6 in the output data matrix obtained in the 19th
739 step of the inferCNV workflow were first converted to a state of 5, and then the CNA states were
740 normalized to values ranging from -1 to +1 and centered on 0 using the following formula:
741 $(\text{CNA State} - 3) / 2$. The matrix values were then squared to give the same weight in the
742 analysis to copy gains and copy losses. Then we calculated the overall CNA level of each cell.
743 For each sample, we identified, the CD45^{-} cells with the top 10% CNA level. Cells with
744 normalized PTPRC RNA count $= 0$ were defined as CD45^{-}. The average CNA level for these
745 cells was defined as the CNA profile of the tumor. A spearman correlation coefficient was then
746 calculated for each cell between its overall CNA level and the CNA profile of the tumor. Cells
747 whose spearman correlation coefficient was above the 25th percentile of all the cells in each

748 sample were considered to have detected CNAs and labeled as potentially malignant. The
749 output of the inferCNV analysis used to identify tumor cells is provided in the supplementary
750 data.
751
752 **Bulk sarcoma similarity score**
753 Bulk RNA-seq of STS samples from TCGA were downloaded (Tatlow and Piccolo, 2016) and
754 scaled to transcripts per million (TPM) as previously described and used as the bulk tumor
755 references. Cells from scRNA-seq data available from Jerby-Arnon et al.[33] that were annotated
756 as normal fibroblasts, macrophages, and endothelial cells were used as the normal reference
757 cells. First, the sarcoma histotypes of reference were extracted from TCGA, then the expression
758 matrices for both the bulk RNA-seq sarcoma samples (TPM) and the scRNA-seq samples (raw
759 UMI counts) were subsampled to include only genes in common between each dataset to
760 compare. Then, two spearman correlation coefficients were calculated for a given cell of a
761 scRNA-seq sample, respectively comparing its gene expression profile to the average
762 expression profile of all the bulk RNA-seq tumor samples of reference (ρ Tumor) and to the
763 average gene expression profile of all the normal cells in the scRNA-seq dataset used as
764 normal control (ρ Normal). Based on these, we defined the similarity score as
765 $\rho\ \mathrm{Tumor} - \rho\ \mathrm{Normal}$. Cells with similarity score > 0 were classified as potentially malignant. The
766 InferCNV and similarity to bulk analyses were combined so that a cell was labeled as malignant
767 if it was CD45^{-} (PTPRC RNA count = 0) and had either detected CNV or a similarity score > 0.
768 Cells that could not be classified into a specific cell type based on any of the above methods
769 were labeled as “unknown” and excluded from further downstream analysis.
770
771 **Differential Expression Analysis**
772 To limit false discovery of differentially expressed genes in STS using scRNA-seq[118], we used a
773 pseudobulk DGE method, limma-voom[37], which has been shown to be statistically more robust

774 and less prone to false discovery than other methods when applied to single cell
775 transcriptomics. Also, for the STS cells versus non-tumoral cells in TME, we complemented the
776 limma-voom results with a new differential distribution testing method, distinct[38], which was
777 specifically designed for DGE using scRNA-seq.
778
779 *Pseudobulk Limma-Voom*
780 Sarcoma cells and normal cells were identified as above. For each STS histotype, the tumor
781 cells were compared to endothelial cells, myeloid cells, lymphoid cells, and CAFs in the TME, as
782 well as with each normal cell cluster of the normal tissues scRNA-seq samples of reference. For
783 each comparison, first a single integrated Seurat object was created and from this, the raw non-
784 normalized mRNA count matrix was extracted. To minimize variability in DE analysis and reduce
785 false discoveries, the synovial sarcoma samples analyzed with plate sequencing technology
786 and those analyzed with droplet-based technology were tested separately. Also, when
787 comparing tumor samples to normal samples from HPA-TS-HCA, only genes that were present
788 in both gene expression matrices were included in the differential expression test, to avoid
789 identification of falsely up-regulated or down-regulated genes. Based on the information in the
790 Seurat object metadata, the total mRNA counts for each gene were calculated by summing
791 across all the cells of a certain cell type belonging to a specific sample, obtaining a unique
792 pseudobulk for each cell type belonging to a given sample in the analysis. The pseudobulk gene
793 expression matrix was then converted to a DGEList object using the DGEList function in the
794 edgeR[129] package (version 3.99.4), and subsequently normalized using the calcNormFactors
795 function. Before voom transformation using the voom function from the R package limma[37]
796 (version 3.57.11), lowly expressed genes were filtered from the expression matrix, using a cutoff
797 of 1 CPM. Counts data was then voom-transformed and a limma workflow was implemented
798 using the sample variable as the blocking variable to control for batch effect. The limma model
799 was then fitted using the patient sex as a covariate. Contrasts were then compared between

each normal cell cluster and each sarcoma cluster separately (UPS, MFS, LMS, LPS, and SS). Genes with a Benjamini-Hochberg (BH) corrected p-value < 0.25 and $\log_2(FC) \neq 0$ were considered as significantly differentially expressed. We used a more lenient BH-corrected p-value cutoff given the screening and exploratory nature of this part of our analysis.

Differential Distribution Testing with distinct

The R package distinct[38] (version 1.10.2) was used to perform differential gene distribution testing between sarcoma cells and non-tumoral cells of the TME. Since distinct compares empirical cumulative distribution functions for a given gene across cell types, it allows identification of subtle differences in expression that are not captured by other methods which might consider only measures of central tendency in the expression distribution. We hypothesized that this could be particularly useful in comparing transcriptionally similar cells, as STS and non-tumoral mesenchymal cells in TME. Due to its computational load, it resulted unfeasible to use distinct for the HPA-TS-HCA normal cells comparisons. The raw gene expression matrices of sarcoma cells and non-tumoral cells were first processed in Seurat as described above to create a merged Seurat object. Afterwards, a single cell experiment object specific to each tumor versus normal comparison was created using the SingleCellExperiment R package[130] (version 1.20.1). The expression data was then log normalized using the logNormCounts function from the R package Scuttle[131] (version 1.8.4). A “batch” variable indicating the sex of the patient was included in the design matrix as a covariate, unless patients were of the same sex. The distinct_test function was then used to run differential distribution testing using default parameters. To obtain the $\log_2$ FC of the gene expression between sarcoma and normal cells, counts were first converted to CPM using the calculateCPM function from Scuttle[131], and these were used as input for the log2_FC function in distinct[38]. Then the $\log_2$FC was then calculated as $\log_2 \frac{(\text{average sarcoma cells CPM} + 0.01)}{(\text{average normal cells CPM} + 0.01)}$, where the average RNA

825 expression in the normal and sarcoma cells were calculated through the distinct function
826 log2_FC. Genes with a $\log_2 FC \neq 0$ and a "p_adj.glb" value < 0.25 were considered as
827 significantly differentially expressed between sarcoma and normal cells.
828
829 *Identification of differentially expressed genes in STS*
830 For each STS histotype, we defined a gene as up-regulated, if it resulted as differentially up-
831 regulated ($\log_2 FC > 0$, FDR < 0.25) in sarcoma cells in at least one of the comparisons made,
832 either against reference cells in the TME (limma-voom or distinct), or in the HPA-TS-HCA
833 datasets (limma-voom). STS down-regulated genes were instead defined as those that resulted
834 as down-regulated in sarcoma cells ($\log_2 FC < 0$, FDR < 0.25) in at least one of the comparisons
835 made, and that were never up-regulated in sarcoma cells, separately considering the TME and
836 the HPA-TS-HCA tests. For genes with more than one alternative transcript, these were tested
837 separately.
838
839 **Hierarchical Clustering of STS versus non-tumoral cells in the TME and normal tissues**
840 To perform hierarchical clustering between STS and non-tumoral cells in the TME as well as
841 normal cells in the HPA-TS-HCA datasets, we used the R package factoextra (version 1.0.7)
842 and base R functions. First, we filtered the CPM normalized scRNA count matrices to retain only
843 genes defined as differentially expressed in STS (either up- or down-regulated), considering the
844 TME and HPA-TS-HCA comparisons separately. Then we merged the count matrices from
845 different samples/tissues, averaged gene expression per cell type and $\log_2(+1)$ transformed the
846 expression values. Pearson correlation was then used to obtain pairwise distances between cell
847 types, separately analyzing STS cells versus non-tumoral cells in the TME, and STS cells
848 versus normal cells in HPA-TS-HCA (supplementary data). The Pearson distance matrices were
849 used to create a dendrogram with the hclust base R function, setting the agglomeration method
850 to "ward.D2", and this was then visualized using the fviz_dend function from factoextra. The

851 optimal number of clusters, k, was obtained by calculating the silhouette score to estimate how
852 similar cells within the same cluster are to each other and how different they are from cells in
853 other clusters, for $k \in \{2, \ldots, n-1\}$, where n is the number of cells considered. The silhouette
854 score peaked at $k = 3$ for the TME comparisons, and around $k = 5$ and $k = 6$, for the HPA-TS-
855 HCA comparisons. We chose $k = 6$ for the latter as it provided a more biologically relevant
856 separation of the different cell types.
857
858 **List of Surface Proteins**
859 To obtain a comprehensive list of genes coding for proteins that are known to or predicted to be
860 expressed on the plasma membrane, we combined The Cancer Surfaceome Atlas[14] with the in
861 silico human surfaceome (SURFY)[132] after conversion of gene names to standard HGNC
862 symbols (**Table S4**).
863
864 **Normal Tissue Expression Datasets**
865 To assess the suitability of a given gene as a cellular therapy target, we used several normal
866 tissue and cell datasets to measure RNA and protein expression.
867 Normal tissue RNA expression was assessed using the bulk RNA-seq dataset from the Human
868 Protein Atlas[39] (version 22.0, Ensembl version 103.38) accessed at
869 https://www.proteinatlas.org/about/download (file name: rna_tissue_hpa.tsv.zip), and the bulk
870 RNA-seq dataset from the GTEx[66] accessed at https://gtexportal.org/home/datasets
871 (expression file: GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_tpm.gct; annotation
872 file: GTEx_Analysis_v8_Annotations_SampleAttributesDS.txt).
873 RNA expression in normal cells was evaluated through the HPA single cell RNA-seq dataset
874 (version 22.0, Ensembl version 103.38), accessed at
875 https://www.proteinatlas.org/about/download (file name: rna_single_cell_type.tsv.zip).
876

877 Protein expression in normal tissues was evaluated using three separate datasets:

878 1. The HPA dataset (version 22.0, Ensembl version 103.38) obtained at (file name:
879 normal_tissue.tsv.zip), which contains protein expression data across 44 normal human
880 tissues obtained through immunohistochemistry.
881 2. The Human Proteome Map[67] (HPM) dataset obtained at
882 http://www.humanproteomemap.org/download.php , which reports protein expression
883 data in 30 normal human tissues obtained through LC-MS/MS proteomics.
884 3. The GTEx Proteomics dataset[133] which reports protein expression data in 32 normal
885 human tissues, obtained through tandem mass tag (TMT) proteomics. Data was
886 obtained from the Supplemental data (Table S2D) of the corresponding manuscript.
887 Specifically, we used the $\log_2$ transformed protein abundances relative to the average of
888 the reference in the same run.

889

890 *Dataset pre-processing*

891 Gene names across datasets were standardized using the R package mygene[134] (version
892 1.38.0) by converting the gene name aliases to the corresponding standardized HGNC symbol.
893 For the bulk RNA-seq datasets (GTEx and HPA), TPM expression values were converted to
894 $\log_2$(TPM +1). Then, the expression data was processed differently based on the target
895 discovery pipeline considered. For the permissive on-target off-tumor toxicity filter of the
896 monospecific CAR target discovery pipeline, RNA expression from the GTEx dataset was first
897 averaged across all samples belonging to a given tissue type. Then tissues were organized in
898 tissue groups in a similar fashion as done by MacKay et al.[45], and the average gene expression
899 per tissue group was calculated (**Tables S5**, **S6**). Similarly, for the HPA dataset, the RNA
900 expression was averaged across tissues belonging to same tissue group. Then, tissue
901 expression in both the GTEx and HPA dataset was categorized as “Medium–High” (M-H) for
902 $\log_2(\text{TPM}+1) \geq 4$, and “Not Detected–Low” (ND-L) for $\log_2(\text{TPM}+1) < 4$. In all the other

903 pipelines, the expression matrices of the bulk RNA-seq GTEx and HPA dataset were median-
904 centered across tissues (HPA) or samples (GTEx), and then z-scored within genes, before use
905 for downstream analysis.
906
907 In the HPA IHC protein expression dataset, the protein expression is reported for each cell type
908 in each tissue as either “Not Detected”, “Low”, “Medium”, or “High”. For use in the permissive
909 on-target off-tumor toxicity filter of the monospecific CAR target discovery pipeline, protein
910 expression for each tissue was classified as follows: “ND-L” if all the cells in a tissue had only
911 “Not Detected” or “Low” expression; “M-H” if all the cells in a tissue had “Medium” or “High”
912 expression, or only “High” expression; “Medium” if all the cells in a tissue had “Medium”
913 expression; “Variable” if the cells in a tissue had either “Not Detected”, “Low”, “Medium” or
914 “High” expression. Tissue group expression was then calculated as the highest tissue
915 expression within a tissue group using the following hierarchy: M-H > Variable > Medium > ND-
916 L. Again, tissue groups were defined as above. For use in all the other pipelines, we first
917 converted the cell type protein expression level from a categorical to a numerical score, where
918 “Not Detected” = 1, “Low” = 2, “Medium” = 3, and “High” = 4. Then, we calculated the average
919 expression value per tissue group for each gene, and then obtained a gene per tissue group
920 expression matrix that was median-centered across tissue groups and z-scored within genes for
921 downstream analysis.
922
923 In the HPA single cell RNA-seq dataset, gene expression is reported as normalized TPM
924 (nTPM) for each normal cell cluster in 31 tissues. We first, converted the nTPM values to
925 $\log_2$(nTPM+1) and then obtained a gene per cell cluster matrix, where columns represented cell
926 clusters in each normal tissue in the dataset. The $\log_2$(nTPM+1) expression values were then
927 median-centered across cell clusters and z-scored within genes, and this matrix was used for
928 downstream analysis.

929

930 The LC-MS/MS proteomics expression data from the HPM was $\log_2(+1)$ transformed and then
931 median-centered across tissues and z-scored within genes, before further analysis. Lastly, in
932 the GTEx TMT proteomics dataset the relative $\log_2$ transformed protein abundance data
933 provided by the authors was median-centered across tissue samples and z-scored within
934 genes, similarly to the other datasets. All the processed normal tissue and cell expression
935 datasets used in the study are provided in the supplementary data.

936

937 **Sarcoma Cell Lines RNA Expression**

938 The HPA cancer cell line expression was downloaded from HPA (version 22.0, Ensembl version
939 103.38) at https://www.proteinatlas.org/about/download (file: rna_celline_cancer.tsv.zip). The
940 dataset was subsetted to retain only data corresponding to sarcoma cell lines including: GCT,
941 HT-1080, MES-SA, RD, Rh18, Rh30, Rh41, RKN, S-117, SK-LMS-1, SK-UT-1, TE 441.T, TE
942 617.T, and U-2197. For each gene, HPA reports the average expression as nTPM for all the cell
943 lines of a given cancer type. The nTPM values were averaged for those genes with alternative
944 transcripts, and then converted to $\log_2(\text{nTPM}+1)$ before use for downstream analyses. The
945 processed data is provided in the supplementary data.

946

947 **Monospecific CAR-T Cell Discovery Pipeline**

948 To define genes suitable for a monospecific CAR-T cell design, first we removed from our initial
949 pool of STS up-regulated genes any genes that were up-regulated more often in normal cells
950 than in sarcoma cells, considering all the comparisons in which that gene appeared. Then this
951 gene pool was filtered to retain only those present in the surfaceome. HLA class I and II genes
952 were also removed from the pool.

953

954 To increase tumor-specificity of our candidate targets, we filtered the gene pool to retain only
955 those with high tumor expression in both our scRNA-seq data, in TCGA SARC, and in the HPA
956 sarcoma cell lines (bulk RNA-seq). Tumor expression in our scRNA-seq data was calculated as
957 the average $\log_2$(CPM+1) across the sarcoma cells of a given STS histotype, where CPM
958 indicates RNA expression in counts per million, while RNA expression in TCGA SARC was
959 calculated as the average $\log_2$(TPM+1) across all the samples of a given STS histotype. Then,
960 we filtered our candidate pool to retain any gene that had above-median expression in both the
961 SARC TCGA data and in the scRNA-seq data, or above-median expression in the scRNA-seq
962 data and in the HPA sarcoma cell lines. Genes without expression data in SARC TCGA or
963 sarcoma HPA cell lines but with scRNA-seq expression $\geq$ the 50th percentile were also retained.
964
965 Then we assessed the potential for on-target off-tumor toxicity of each candidate target by
966 measuring its RNA and protein expression in normal tissues and cells and comparing it to that of
967 targets for which CAR-T cells are currently available for clinical use or in clinical development.
968 The list of tested CAR-T targets was obtained from the previous work from MacKay et al[45]
969 (**Table S7**). Specifically, we identified genes at risk for on-target off-tumor toxicity by applying
970 both a *permissive* and a *stringent filter*.
971
972 For our *permissive filter*, we used the HPA IHC protein database, as well as the GTEx and HPA
973 bulk RNA-seq databases. Specifically, protein and RNA expression were categorized as
974 described above, and then the number of tissue groups (TGs) with medium-high (M-H)
975 expression was measured for the tested CAR-T targets. We used the 90th percentile of the
976 number of TGs with M-H expression for the tested CAR-T targets in each of the three datasets
977 as cutoffs. A gene was considered to have passed the filter in a dataset if it had number of TGs
978 with M-H expression below the respective cutoff.
979

980 For our *stringent filter*, we assessed the risk of on-target off-tumor toxicity both at the single cell
981 level, using the HPA scRNA-seq dataset, and by measuring protein expression at the tissue
982 level, using the HPA, HPM, and the GTEx protein expression datasets. For this analysis, we
983 used the HPA protein numerical scoring classification obtained as described above. Also, for the
984 HPA scRNA-seq dataset we used the average cell type expression across all tissues. Again, we
985 used the normal cell/tissue expression of tested CAR-T targets as a reference to score our
986 candidate genes against.
987
988 For a gene to have an acceptable on-target off-tumor toxicity risk based on its expression in the
989 HPA scRNA-seq dataset, it had to meet two conditions. First, we categorized gene expression
990 as "high" for z-score> 1, and "low" for z-score ≤ 1. Based on this, genes from our pool that had
991 "high" expression in ≤14 cell types, which corresponds to the 75th percentile of the number of
992 cell types with "high" expression for the tested CAR-T targets, were initially retained. Then, we
993 calculated the maximum expression value per cell type for the tested CAR-T targets and
994 retained only those genes that had RNA expression below this threshold in each cell type. The
995 following cell types were not considered: lymphoid immune cells (T-cells, NK-cells, B-cells,
996 plasma cells), due to the potentially higher tolerability of their depletion; fibroblasts, adipocytes,
997 and endometrial stromal cells due to potential high transcriptional similarity to the STS in the
998 study; and cells in the gonads and placenta (ovarian stromal cells, granulosa cells, oocytes,
999 peritubular cells (testis), cytotrophoblasts, extravillous trophoblasts, Hofbauer cells,
1000 syncytiotrophoblasts, Leydig cells, spermatocytes, late spermatids, spermatogonia, early
1001 spermatids, and Sertoli cells) being sites of immune privilege. Only those genes that met both
1002 criteria were retained based on their single cell RNA expression in normal cells.
1003
1004 For each of the 3 protein expression datasets, we calculated the maximum expression values
1005 per tissue (GTEx and HPM) or tissue group (HPA) for the tested CAR-T targets and used these

1006 as thresholds so that only the candidate targets with protein expression below the threshold in
1007 each of the tissues or tissue groups studied were retained. The following tissues and tissue
1008 groups were not considered: fetal tissues, male or female reproductive tissues, placenta, CD4
1009 T-cells, CD8 T-cells, NK-cells, B-cells, spleen, immune cells and immune tissue. Notably, for the
1010 GTEx proteomics dataset, we used the average z-scored protein expression across all samples
1011 belonging to a given tissue.
1012
1013 For each candidate gene (*g*), we then combined the results of both filters by calculating a
1014 monospecific CAR Targetability Score, TS_{CAR}, as follows:

1015 $$TS_{CAR}(g) = \sum_{i=1}^{2} m_i + \sum_{j=1}^{4} m_j + m_k$$

1016 Where:

1017 - m_i represents the score from the 2 bulk RNA-seq datasets (HPA and GTEx categorical)
1018 and is calculated as +1 if the gene passes the filter, –1 if it does not, and 0 if the gene is
1019 not present in the dataset.
1020 - m_j represents the score from any of the 4 protein expression datasets (HPA categorical,
1021 GTEx proteomics, HPM, and HPA numeric) and is calculated as +2 if the gene passes
1022 the filter, –2 if it does not, and 0 if it is not present in the dataset.
1023 - m_k represents the score from the HPA scRNA-seq dataset and is calculated as +1.5 if
1024 the gene passes the filter, –1.5 if it does not, and 0 if it is not present in the dataset.

1025 Any gene with $TS_{CAR} \geq 5$ which corresponds to the 25th percentile of the TS_{CAR} for the tested
1026 CAR-T targets was considered as a suitable target for a monospecific CAR-T cell. The results of
1027 the pipeline are reported in **Table S8**.
1028
1029 **Gene set over-representation analyses**

1030 Gene set over-representation analysis (ORA) was performed using the R packages,
1031 clusterProfiler[135] (version 4.10.0), ReactomePA[136,137] (version 1.46.0), and org.Hs.eg.db[138]
1032 (version 3.18.0). Before analysis, the gene symbols were converted to Entrez gene IDs. Then,
1033 we analyzed all the candidate targets for monospecific CAR-T cells for each STS, using the
1034 function enrichGO for Gene Ontology (GO) term enrichment, the function enrichKEGG for Kyoto
1035 Encyclopedia of Genes and Genomes (KEGG) pathways enrichment, the function
1036 enrichPathway for enrichment of Reactome pathways, and the function enrichWP for
1037 enrichment of WikiPathways terms. The p-value cutoff for significance was set to 0.05 and the
1038 Bonferroni-Hochberg method was used for FDR correction, while minimum and maximum gene
1039 sets sizes were set respectively to 10 and 500. We then performed a clustered ORA on the
1040 monospecific CAR-T targets by STS histotype. In this case we used the compareCluster
1041 function from clusterProfiler and within this we used the same enrichment functions as
1042 described above, depending on the reference database used, again setting FDR cutoff to 0.05
1043 and minimum and maximum gene sets sizes to 10 and 500. When performing the cluster ORA
1044 analysis on the TCR-T and PC-CAR targets, the FDR cutoff was instead set to 0.1 due to
1045 smaller gene pool to test, while the minimum and maximum gene sets sizes were set to 2 and
1046 500, respectively.
1047
1048 **TCGA Pan-Cancer Atlas**
1049 RNA expression data for the Pan-Cancer Atlas TCGA[139] samples (batch normalized RSEM
1050 values) was obtained from cBioPortal[140,141] for the genes of interest. The downloaded data is
1051 available in the Source Data. Data for the following cancer types was analyzed: KICH: Kidney
1052 Chromophobe; ACC: Adrenocortical carcinoma; LIHC: Liver hepatocellular carcinoma; UVM:
1053 Uveal Melanoma; PCPG: Pheochromocytoma and Paraganglioma; KIRP: Kidney renal papillary
1054 cell carcinoma; LAML: Acute Myeloid Leukemia; THYM: Thymoma; SKCM: Skin Cutaneous
1055 Melanoma; KIRC: Kidney renal clear cell carcinoma; READ: Rectum Adenocarcinoma; LGG:

1056 Brain Lower Grade Glioma; CHOL: Cholangiocarcinoma; COAD: Colon Adenocarcinoma;
1057 PRAD: Prostate Adenocarcinoma; STAD: Stomach Adenocarcinoma; UCEC: Uterine Corpus
1058 Endometrial Carcinoma; CESC: Cervical squamous cell carcinoma and endocervical
1059 adenocarcinoma; BLCA: Bladder Urothelial Carcinoma; ESCA: Esophageal Carcinoma; DLBC:
1060 Diffuse Large B-Cell Lymphoma; TGCT: Testicular Germ Cell Tumors; OV: Ovarian serous
1061 cystadenocarcinoma; LUSC: Lung Squamous Cell Carcinoma; LUAD: Lung Adenocarcinoma;
1062 HNSC: Head and Neck Squamous Cell Carcinoma; BRCA: Breast Invasive Carcinoma; PAAD:
1063 Pancreatic Adenocarcinoma; GBM: Glioblastoma Multiforme; THCA: Thyroid Carcinoma; UCS:
1064 Uterine Carcinosarcoma; SARC: Sarcoma; MESO: Mesothelioma.
1065
1066 **Cell Line Gene Dependency**
1067 The DepMap project[57,71] reports the results of CRISPR KO experiments on 1100 human cell
1068 lines. To assess the essentiality of our candidate targets for sarcoma cell lines (including both
1069 soft tissue sarcomas as well as chondrosarcoma, osteosarcoma and Ewing's sarcoma), we
1070 used the "CRISPRGeneEffect.csv" file obtained from the DepMap website, containing data
1071 regarding the effect of the KO of each gene in each cell line. Gene effects less than –1 indicate
1072 a significant reduction in cell growth and proliferation, as this was the median gene effect
1073 estimate of genes known to be essential for cell growth and survival.
1074
1075 **Soft Tissue Sarcoma Genomics Data**
1076 Data on copy number alterations (CNAs) and mutations in soft tissue sarcomas were obtained
1077 from the TCGA SARC firehose database[36] and from the MSK-IMPACT[58] sarcoma study. Both
1078 datasets were downloaded from cBioPortal[140,141]. The file named "data_linear_cna.txt" was used
1079 to analyze linear CNA data from TCGA firehose, while the file named "data_cna.txt" was used to
1080 determine amplification status of a given gene in each sample for both the TCGA and MSK-
1081 IMPACT databases. For the Pearson correlation analysis between linear CNA and RNA

1082 expression, the TCGA SARC bulk RNA-seq expression data obtained as above, was used.
1083 Correction for multiple hypothesis was obtained using the BH method setting the threshold for
1084 significance to 0.01.
1085
1086 Data regarding single nucleotide variants (SNVs) and small insertions/deletions (Indels) was
1087 retrieved from the "data_mutations.txt" files for both the TCGA and the MSK-IMPACT cohorts.
1088 Lastly, data on gene fusions in soft tissue sarcomas was obtained by integrating data in the
1089 "data_sv.txt" file obtained from cBioPortal for the MSK-IMPACT cohort, data from the study from
1090 Gao et al.[142] on recurrent gene fusions in TCGA tumor samples, and from a search through the
1091 Mitelman Database[61,143] (https://mitelmandatabase.isb-cgc.org/mb_search) that included the
1092 following sarcoma types: Ewing sarcoma; Atypical lipomatous tumor/atypical lipoma/well-
1093 differentiated liposarcoma; Alveolar rhabdomyosarcoma; Myxoinflammatory fibroblastic
1094 sarcoma; Undifferentiated pleomorphic sarcoma; Liposarcoma, dedifferentiated; Inflammatory
1095 myofibroblastic tumor/myofibroblastic sarcoma; Soft tissue tumor, NOS; Clear cell sarcoma;
1096 Undifferentiated round cell sarcoma; Undifferentiated sarcoma; Mesenchymal tumor, NOS;
1097 Fibrosarcoma; Embryonal rhabdomyosarcoma; Spindle cell/sclerosing rhabdomyosarcoma;
1098 Myxofibrosarcoma; Endometrial stromal sarcoma; Pleomorphic rhabdomyosarcoma;
1099 Liposarcoma, NOS; Alveolar soft part sarcoma; Leiomyosarcoma; Sclerosing epithelioid
1100 fibrosarcoma; Malignant soft tissue tumor, NOS; Angiosarcoma; Rhabdomyosarcoma, NOS;
1101 Epithelioid sarcoma; Dermatofibrosarcoma protuberans/Bednar tumor/giant cell fibroblastoma;
1102 Low-grade fibromyxoid sarcoma; Synovial sarcoma; Histiocytic sarcoma/malignant histiocytosis;
1103 Liposarcoma, myxoid/round cell; Liposarcoma, pleomorphic. The combined fusion data in STS
1104 is available from **Table S15**.
1105
1106 **Logic-gated 'AND' CAR Targets Discovery Pipeline**

To discover gene pairs suitable for a bispecific logic-gated 'AND' CAR design, first we selected genes up-regulated in sarcoma cells, separately for each STS, as described above. Next, we removed from this pool of STS up-regulated genes any genes that were up-regulated more often in normal cells than in sarcoma cells, considering all the comparisons in which that gene appeared. Then, we filtered this gene pool to include only highly expressed genes, as defined in the monospecific CAR-T cell target discovery pipeline. Then, for the selected genes we computed all possible pairs and measured gene co-expression at the single cell level by running the CS-CORE algorithm from the R package CSCORE[65] (version 0.0.0.9). Specifically, we ran CS-CORE on all the tumor cells of a given sarcoma histotype and corrected p-values for multiple hypothesis testing using the BH method. Any non-significant correlations were converted to a score of 0, using an adjusted p-value significance threshold of 0.05. For a given gene pair (g_i, g_j), CS-CORE computes a correlation score, which we will refer to as $\rho_{Tumor}(g_i, g_j)$, and which can assume values between –1 (lowest likelihood of co-expression) and +1 (highest likelihood of co-expression).

Then, for each gene pair we assessed co-expression in normal cells and tissues, with the caveat that genes for an ideal target pair of an 'AND' logic-gated CAR would have mutually exclusive expression across all normal cells and tissues. To assess this, we used, the GTEx and HPA bulk RNA-seq databases, the HPA scRNA-seq database, the HPA IHC protein expression database, as well as the TMT GTEx proteomics and the HPM LC-MS/MS proteomics datasets. For the bulk RNA-seq datasets (GTEx and HPA), the median-centered, z-scored numerical gene expression matrices were used, where expression per each tissue (HPA), or tissue sample (GTEx) was measured. In this case, we used Pearson correlation to assess gene co-expression across normal tissues. Correction for multiple hypothesis testing was performed using the BH method, and for pairs with a corrected p-value ≥ 0.05 the correlation coefficient ρ was set to 0.

Similarly, for the scRNA-seq HPA database, we used Pearson correlation to measure gene co-expression across normal cell clusters, with non-significant correlations set to $\rho = 0$, using a BH-corrected p-value threshold of 0.05. For the GTEx TMT proteomics database, we first removed any genes with more than 25% missing values across all tissue samples analyzed. Then, we applied the Pearson correlation as above to assess protein co-expression across normal tissue samples, again using a BH-corrected p-value threshold of 0.05 for significance and setting $\rho = 0$ for any non-significant correlations.

For the HPA IHC database (numerical score data) and for the HPM proteomics database, given the low number of comparisons, we did not use Pearson correlation but instead calculated a modified odds ratio (OR) for each gene pair (g_i, g_j) as follows:

$$OR_{mod(g_i,g_j)} = \frac{d}{b \times c}$$

Where:

- *d* is the number of tissues/cells where both g_i and g_j have z-score > 1
- *b* is the number of tissues/cells where g_i has z-score > 1 and g_j has z-score ≤ 1
- *c* is the number of tissues/cells where g_j has z-score > 1 and g_i has z-score ≤ 1

In this case, pairs with mutually exclusive protein expression were considered those where $OR_{mod} = 0$. Of note, if both *d* and $b \times c$ were 0, the OR was set to 0.

Then, for each gene pair (g_i, g_j) we combined results from each normal tissue/cell database to obtain a score that estimates gene co-expression across normal tissues and cells, $N(g_i, g_j)$, as follows:

$$N(g_i, g_j) = \sum_{k=1}^{2} r_{g_i,g_j}^{k} + \sum_{l=1}^{2} r_{g_i,g_j}^{l} + r_{g_i,g_j}^{m} + r_{g_i,g_j}^{n}$$

Where:

- $r^k_{g_i,g_j}$ represents the score from the 2 bulk RNA-seq datasets and is calculated as:

$$\begin{cases} +1 \ \ if \ \rho_{g_i,g_j} < 0 \text{ and } p_{g_i,g_j} < 0.05 \\ -1 \times \ if \ \rho_{g_i,g_j} > 0 \text{ and } p_{g_i,g_j} < 0.05 \\ 0 \ if \ p_{g_i,g_j} \geq 0.05 \text{ or } if \text{ either } g_i \ or \ g_j \ is \ not \ in \ the \ dataset \ or \ if \ g_i \ or \ g_j \ SD = 0 \end{cases}$$

- $r^l_{g_i,g_j}$ represents the score from the HPA IHC and the HPM proteomics datasets and is calculated as:

$$\begin{cases} +2 \ \ if \ OR_{mod(g_i,g_j)} = 0 \\ -2 \times if \ OR_{mod(g_i,g_j)} \neq 0 \\ 0 \ if \text{ either } g_i \ or \ g_j \ is \ not \ in \ the \ dataset \end{cases}$$

- $r^m_{g_i,g_j}$ represents the score from the HPA scRNA-seq dataset and is calculated as:

$$\begin{cases} +1.5 \ \ if \ \rho_{g_i,g_j} < 0 \text{ and } p_{g_i,g_j} < 0.05 \\ -1.5 \times \ if \ \rho_{g_i,g_j} > 0 \text{ and } p_{g_i,g_j} < 0.05 \\ 0 \ if \ p_{g_i,g_j} \geq 0.05 \text{ or } if \text{ either } g_i \ or \ g_j \ is \ not \ in \ the \ dataset \ or \ if \ g_i \ or \ g_j \ SD = 0 \end{cases}$$

- $r^n_{g_i,g_j}$ represents the score from the GTEx proteomics dataset and is calculated as:

$$\begin{cases} +2 \ \ if \ \rho_{g_i,g_j} < 0 \text{ and } p_{g_i,g_j} < 0.05 \\ -2 \times \ if \ \rho_{g_i,g_j} > 0 \text{ and } p_{g_i,g_j} < 0.05 \\ 0 \ if \ p_{g_i,g_j} \geq 0.05 \text{ or } if \text{ either } g_i \ or \ g_j \ is \ not \ in \ the \ dataset \ or \ if \ g_i \ or \ g_j \ SD = 0 \end{cases}$$

- ρ_{g_i,g_j} is the Pearson correlation coefficient in a dataset between the two genes in a pair and p_{g_i,g_j} is the respective BH-corrected p-value.

The raw score $N(g_i, g_j)$ was then normalized and scaled as follows:

$$N_{norm}(g_i, g_j) = -\left\{\left(\frac{N(g_i, g_j) \ - min(N)}{max(N) \ - min(N)} \times \ 2\right) - 1\right\}$$

Where:

- $N(g_i, g_j)$ is the raw score for a given gene pair (g_i, g_j)

- min(N) and max(N) are the maximum and minimum raw scores across all gene pairs for a given STS histotype, respectively.

This normalization ensures that a score between 0 and 1 indicates higher likelihood of gene co-expression across normal tissues/cells and conversely a score between -1 and 0 indicates a lower likelihood of co-expression.

Finally, for each gene pair (g_i, g_j) we calculated an 'AND' logic-gated Targetability Score, called $TS_{'\mathrm{AND}'}$ score, as follows:

$$TS_{'\mathrm{AND}'}(g_i, g_j) = \rho_{Tumor}(g_i, g_j) - N_{norm}(g_i, g_j)$$

Gene pairs with $TS_{'\mathrm{AND}'}$ score ≥ 1 were considered to have higher fitness as targets for an 'AND'-gated CAR design. The results of the pipeline by STS histotype are reported in **Table S17**.

Logic-gated 'AND–NOT' CAR Targets Discovery Pipeline

To discover gene pairs for an 'AND–NOT' bispecific CAR design, we applied a similar logic to that used for discovery of gene pairs for an 'AND'-gated CAR. In this case, we looked for genes with high expression in sarcoma cells that would function as activators, and for genes with low to no expression in sarcoma cells but high co-expression with the paired activator in normal cells/tissues that would function as inhibitors, thus preventing on-target off-tumor toxicity.

To do this, first we obtained a list of highly expressed sarcoma up-regulated genes for each STS histotype, removing any genes up-regulated in normal cells more often than STS cells, as described above, and considered these as candidate activators. Then, to define a list of possible inhibitors with little to no expression in sarcoma cells, for each STS histotype we

selected surfaceome genes with scRNA-seq expression below the 50th percentile among the sarcoma down-regulated genes as defined in our differential gene expression pipeline. Single cell RNA expression was again measured as the average $\log_2$(CPM+1) across all the tumor cells in a sarcoma histotype.

We then calculated all the possible unique pairs between the candidate activators and inhibitors and measured their co-expression in normal cells/tissues using the same datasets used for the 'AND'-gated CAR targets discovery pipeline. Specifically, for the GTEx and HPA bulk RNA-seq datasets, as well as for the scRNA-seq HPA and the GTEx TMT proteomics datasets we used Pearson correlation to measure co-expression of the two genes in each pair, and applied BH multiple hypothesis testing correction setting the threshold for significance to a p-value < 0.05. For non-significant correlations, we then set $\rho = 0$. For the GTEx TMT proteomics dataset, again we first removed any genes with more than 25% missing values across all tissue samples analyzed before performing correlation testing.

For the HPA IHC and the HPM proteomics databases, instead of performing Pearson correlation analysis we calculated a modified odds ratio (OR) for each gene pair (g_a, g_i) between an activator, g_a, and an inhibitor, g_i as follows:

$$OR_{mod(g_a,g_i)} = \frac{d}{b}$$

Where:

- *d* is the number of tissues where both the activator g_a and the inhibitor g_i have z-score > 1
- *b* is the number of tissues where the activator g_a has z-score > 1 and the inhibitor g_i has z-score ≤ 1

Pairs where $OR_{mod} = Inf$ were considered as suitable for an 'AND–NOT' CAR design. In cases where both *d* and *b* were 0, we set OR_{mod} to Inf.

Then, for each gene pair (g_a, g_i) we combined results from each normal tissue/cell database to calculate an 'AND–NOT' logic-gated Targetability Score, $TS_{,AND-NOT,}$ as follows:

$$TS_{,AND-NOT,}(g_a, g_i) = \sum_{k=1}^{2} S^k_{g_a,g_i} + \sum_{l=1}^{2} S^l_{g_a,g_i} + S^m_{g_a,g_i} + S^n_{g_a,g_i}$$

Where:

- $S^k_{g_a,g_i}$ represents the score from the 2 bulk RNA-seq datasets and is calculated as:

$$\begin{cases} +1 \times \left(1 + I_{g_a,g_i}\right) \; if \; \rho_{g_a,g_i} > 0 \text{ and } p_{g_a,g_i} < 0.05 \\ -1 \times \left(1 + A_{g_a,g_i}\right) \; if \; \rho_{g_a,g_i} < 0 \text{ and } p_{g_a,g_i} < 0.05 \\ 0 \; if \; p_{g_a,g_i} \geq 0.05 \text{ or } if \text{ either } g_a \; or \; g_i \; is \; not \; in \; the \; dataset \; or \; if \; g_a \; or \; g_i \; SD = 0 \end{cases}$$

- $S^l_{g_a,g_i}$ represents the score from the HPA IHC and the HPM proteomics datasets and is calculated as:

$$\begin{cases} +2 \times \left(1 + I_{g_a,g_i}\right) \; if \; OR_{mod(g_a,g_i)} = Inf \\ -2 \times \left(1 + A_{g_a,g_i}\right) \; if \; OR_{mod(g_a,g_i)} \neq Inf \\ 0 \; if \text{ either } g_a \; or \; g_i \; is \; not \; in \; the \; dataset \end{cases}$$

- $S^m_{g_a,g_i}$ represents the score from the HPA scRNA-seq dataset and is calculated as:

$$\begin{cases} +1.5 \times \left(1 + I_{g_a,g_i}\right) \; if \; \rho_{g_a,g_i} > 0 \text{ and } p_{g_a,g_i} < 0.05 \\ -1.5 \times \left(1 + A_{g_a,g_i}\right) \; if \; \rho_{g_a,g_i} < 0 \text{ and } p_{g_a,g_i} < 0.05 \\ 0 \; if \; p_{g_a,g_i} \geq 0.05 \text{ or } if \text{ either } g_a \; or \; g_i \; is \; not \; in \; the \; dataset \; or \; if \; g_a \; or \; g_i \; SD = 0 \end{cases}$$

- $S^n_{g_a,g_i}$ represents the score from the GTEx proteomics dataset and is calculated as:

$$\begin{cases} +2 \times \left(1 + I_{g_a,g_i}\right) \; if \; \rho_{g_a,g_i} > 0 \text{ and } p_{g_a,g_i} < 0.05 \\ -2 \times \left(1 + A_{g_a,g_i}\right) \; if \; \rho_{g_a,g_i} < 0 \text{ and } p_{g_a,g_i} < 0.05 \\ 0 \; if \; p_{g_a,g_i} \geq 0.05 \text{ or } if \text{ either } g_a \; or \; g_i \; is \; not \; in \; the \; dataset \; or \; if \; g_a \; or \; g_i \; SD = 0 \end{cases}$$

- I_{g_a,g_i} is the proportion of comparisons in a dataset where the expression of the inhibitor is equal or above that of the activator ($z - score_{g_i} \geq z - score_{g_a}$)

1242 • A_{g_a,g_i} is the proportion of comparisons in a dataset where the expression of the activator
1243 is above that of the activator ($z - score_{g_a} > z - score_{g_i}$)
1244 • ρ_{g_a,g_i} is the Pearson correlation coefficient in a dataset between the activator and the
1245 inhibitor and p_{g_a,g_i} is the respective BH-corrected p-value.
1246
1247 The $TS_{'AND-NOT'}$ scores were then normalized as follows:

1248 $$TS_{'AND-NOT'}(g_a, g_i) = \left\{\left(\frac{TS_{'AND-NOT'}(g_a, g_i) - min(TS_{'AND-NOT'})}{max(TS_{'AND-NOT'}) - min(TS_{'AND-NOT'})} \times 2\right) - 1\right\}$$

1249 Where:
1250 • $TS_{'AND-NOT'}(g_a, g_i)$ is the raw score for a given gene pair (g_a, g_i)
1251 • $min(TS_{'AND-NOT'})$ and $max(TS_{'AND-NOT'})$ are the maximum and minimum raw scores
1252 across all gene pairs for a given STS histotype, respectively.
1253 In each sarcoma histotype, gene pairs with values above the 99th percentile of the normalized
1254 $TS_{'AND-NOT'}$ score were considered as those suitable for an 'AND-NOT'-gated CAR design. The
1255 results of the pipeline by STS histotype are reported in **Table S18**.
1256
1257 **HPA Single Cell Gene Clusters and Signatures**
1258 HPA categorizes genes based on their expression profile across different normal cell types in
1259 their scRNA-seq dataset[144]. Specifically, genes are categorized into 80 clusters, with one or
1260 more clusters corresponding to a specific cell type. We downloaded the gene cluster
1261 assignments directly from the HPA website and grouped gene clusters specific to a given cell
1262 type together to obtain cell type-specific gene signatures that were used to annotate genes in
1263 each bispecific CAR target pair. Also, the expression of these gene signatures in our STS
1264 scRNA-seq data was measured using the Seurat function AddModuleScore and shown in

1265 **Figure S7**. The HPA single cell gene clusters and signatures are provided in the Supplementary
1266 Data.
1267
1268 **T cell receptor-engineered T cell and Peptide-Centric CAR Target Discovery Pipeline**
1269 To identify targets for TCR-T or PC-CAR-T cells in STS, we obtained a list of up-regulated
1270 genes for each of the STS histotype studied. This was defined as any gene, either coding for
1271 intracellular or surface proteins, that was up-regulated in at least one of the DET comparisons
1272 made. Within this gene pool we then identified genes with high STS expression, defined in this
1273 case as those with expression $\geq 25^{th}$ percentile in both our scRNA-seq data and in the SARC
1274 TCGA samples, or those genes with expression $\geq 25^{th}$ percentile in both the sarcoma HPA cell
1275 lines and our scRNA-seq data. Genes without expression data in SARC TCGA or sarcoma HPA
1276 cell lines but with scRNA-seq expression $\geq$ the 25th percentile were also considered to have
1277 high STS expression.
1278
1279 Then we compared the expression in normal cells and tissues of each of the up-regulated
1280 genes to that of known TCR-T cell therapy targets that are currently in clinical development
1281 identified through a search on clinicaltrials.gov with the term “solid tumor” for the
1282 Condition/disease field, and “TCR-T cells” or “cellular therapy” for the Intervention/treatment
1283 field (**Table S22**). Specifically, only genes belonging to the cancer germline antigen (CGA) class
1284 were chosen as reference due to their preferential expression in immune-privileged sites like the
1285 gonads.
1286
1287 Normal cell and tissue expression was measured again using the HPA scRNA-seq dataset
1288 (average cell type expression), the HPA and GTEx bulk RNA-seq datasets (average tissue
1289 group expression), the GTEx TMT proteomics dataset (average tissue expression), the HPM
1290 LC-MS/MS proteomics dataset, or the HPM IHC data, obtained as previously described.

1291

1292 For each gene (g) we calculated a summative score $N(g)$, representing its normal cell/tissue
1293 RNA and protein expression as compared to that of the CGAs used as TCR-T cell targets, and
1294 calculated as follows:

1295 $$N(g) = \sum_{i=1}^{2} t_i + \sum_{j=1}^{3} t_j + t_k$$

1296 Where:

1297 - t_i is the score calculated by comparing the candidate target expression to that of the
1298 reference targets in the HPA and GTEx bulk RNA-seq datasets and is calculated as $+1$ if
1299 the candidate has expression below the maximum expression value for the reference
1300 targets in each tissue group, 0 if the candidate has expression above the maximum
1301 value for the reference targets in at least one of the tissue groups or if the gene is not
1302 present in the parent dataset;
1303 - t_j is the score calculated by comparing the candidate target expression to that of the
1304 reference targets in the HPA IHC, GTEx, and HPM proteomics datasets and is
1305 calculated as $+2$ if the candidate has expression below the maximum expression value
1306 for the reference targets in each tissue or tissue group, 0 if the candidate has expression
1307 above the maximum value for the reference targets in at least one of the tissues or
1308 tissue groups or if the gene is not present in the parent dataset;
1309 - t_k is the score calculated by comparing the candidate target expression to that of the
1310 reference targets in the HPA scRNA-seq dataset and is calculated as $+1.5$ if the
1311 candidate has expression below the maximum expression value for the reference targets
1312 in each cell type and 0 if the candidate has expression above the maximum value for the
1313 reference targets in at least one of the cell types or if the gene is not present in the
1314 parent dataset.

1315

1316 Genes with a $N(g)$ score ≥ 3.5 and with high STS expression as defined above were retained
1317 for further analysis. Due to the low number of genes with $N(g)$ ≥ 3.5 meeting the combined high
1318 expression criteria, possibly related to RNA expression dilution in bulk tissue analysis (TCGA
1319 SARC) and differences between in vivo and in vitro analysis (HPA cell lines), genes that had
1320 $N(g)$ score ≥ 3.5 and scRNA-seq expression ≥ the 25th percentile, regardless of TCGA SARC
1321 or HPA cell line expression, were also considered for further analysis as below (**Table S23**).

1322

1323 *Global HLA Allele Frequency Estimation*

1324 To estimate the global HLA allele frequencies at the HLA-B, HLA-C, and HLA-A loci, we
1325 downloaded allele frequencies for all the population studies available in the Allele Frequency
1326 Net Database (AFND)[77]. Then, we used the python package HLAfreq[145] to combine allele
1327 frequencies across populations, first within countries and then across countries to obtain an
1328 average world allele frequency. Analyses were ran using default parameters. The population
1329 size of each country was use as weight in the global allele frequency calculation, as described
1330 by the developer. Population sizes by country updated to 2022 were downloaded from the World
1331 Bank (https://data.worldbank.org/indicator/SP.POP.TOTL?most_recent_year_desc=false), or
1332 from the United Nations Populations Division (Martinique, Turkey, Taiwan), or from the European
1333 Union Commission (Madeira, Azores).

1334

1335 *Peptide-MHC targetability assessment*

1336 Next for each candidate target antigen, we used the MSiC model from HLAthena[74] to infer
1337 presentation and binding to HLA class I molecules of all the possible 9-mer and 10-mer peptides
1338 derived from each target. The amino acid sequence of each candidate target in FASTA format
1339 was obtained from UniProt[146]. For our analysis, we chose the top 10 most frequent HLA class I
1340 alleles by geographical region worldwide as reported in the Allele Frequency Net Database

1341 (AFND)[77] https://www.allelefrequencies.net/top10freqs.asp, and then applied DeepImmuno[75] to
1342 measure the predicted immunogenicity, the ability to elicit a $CD8^+$ immune response, of each
1343 peptide-MHC pair ($pMHC$). Then, for each HLA allele we calculated the average worldwide
1344 allele frequency across all the populations in which that allele was detected in the AFND (**Table**
1345 **S26**). Finally, for each $pMHC$ of each candidate target gene *g*, a targetability score, $TS_{TCR/PC}$,
1346 was calculated to assess its suitability as a target for either TCR-T or PC-CAR-T cells as
1347 follows:
1348

1349 $$TS_{TCR/PC}(pMHC_i) = \left(S_{HLAthena} + S_{DeepImmuno}\right) \times (1 + F_{allele})$$

1350

1351 Where $pMHC_i$ represents the ith $pMHC$ pair of a target gene *g*, $S_{HLAthena}$ is the score from the
1352 MSiC HLAthena model for $pMHC_i$, $S_{DeepImmuno}$ is the immunogenicity score assigned by
1353 DeepImmuno for $pMHC_i$, and F_{allele} is the average worldwide frequency for the HLA allele in
1354 $pMHC_i$. For each target antigen, we then ranked peptides based on their cumulative HLA allele
1355 frequency calculated as the sum of the allele frequencies of the HLA alleles in all of the $pMHCs$
1356 formed by each peptide, and then retained for cross-reactivity assessment only those $pMHCs$
1357 originating from peptides with the top five highest cumulative HLA allele frequency, and with
1358 $TS_{TCR/PC} \geq$ the 99th percentile, and $S_{HLAthena} \geq 0.85$. The results for HLAthena and
1359 DeepImmuno for all the target antigens, as well as the $TS_{TCR/PC}$ scores, are available in the
1360 supplementary data.
1361

1362 *Peptide-MHC cross-reactivity assessment*

1363 For the $pMHCs$ of each target antigen selected as described above, we assessed the likelihood
1364 of cross-reactivity with $pMHCs$ derived from other proteins expressed in normal cells and
1365 tissues. To do this, we developed an algorithm called PS-TRACT (Peptide Similarity and TCR

1366 Reactivity Analysis for Cellular Therapies), that considers the similarity of each tumoral $pMHC$ to
1367 a pool of normal $pMHCs$ of reference while accounting for the likelihood of presentation of the
1368 cross-reactive $pMHCs$ in normal tissues. Specifically, first we obtained a list of $pMHCs$
1369 experimentally detected in normal tissues/cells (*observed normal ligandome*) by integrating data
1370 from IEDB[147], systeMHC[148], and the HLA Ligand Atlas[149]. For IEDB, the full dataset was
1371 downloaded from https://iedb.org/database_export_v3.php (file: "mhc_ligand_full.csv.csv"), and
1372 epitopes were filtered to retain only linear peptides derived from human proteins and binding to
1373 MHC class I, with any qualitative binding measurement other than "Negative", and of either 9 or
1374 10 amino acids in length. Epitopes where MHC binding was measured with any of the following
1375 methods were also removed: half maximal effective concentration (EC50), dissociation constant
1376 KD (~EC50), and dissociation constant KD (~IC50). Lastly, only epitope binding data for alleles
1377 in standard four-digit typing HLA allele nomenclature were retained. For the HLA Ligand Atlas,
1378 the 2020.12 data release was obtained from https://hla-ligand-atlas.org/downloads. Epitope data
1379 was filtered to retain only 9-mer and 10-mer peptides binding to HLA class I or HLA class I and
1380 II, categorized as either strong or weak binders, and only those binding to HLA alleles reported
1381 in standard four-digit typing nomenclature. For systeMHC, data for the following SYSMHCIDs
1382 were obtained from https://systemhc.sjtu.edu.cn/datasets: SYSMHC00002, SYSMHC00003,
1383 SYSMHC00004, SYSMHC00007, SYSMHC00026, SYSMHC00028, SYSMHC00029,
1384 SYSMHC00031, SYSMHC00037, SYSMHC00039, SYSMHC00040, SYSMHC00044,
1385 SYSMHC00047, SYSMHC00052, SYSMHC00053, SYSMHC00055, SYSMHC00057,
1386 SYSMHC00064, SYSMHC00073, SYSMHC00088, and SYSMHC00089. In this case, only 9-
1387 mer and 10-mer peptides binding HLA class I and derived from human proteins were retained.
1388 To expand the pool of $pMHCs$ presented in normal tissues, we ran MHCflurry[150] for all the
1389 possible 9-mer and 10-mer peptides derived from the whole human proteome (*predicted normal*
1390 *ligandome*), using all the HLA alleles tested in our $pMHC$ targetability pipeline as described
1391 above. Protein FASTA sequences for 20,409 proteins total were obtained from UniProt and

1392 mhcflurry-predict-scan was used to obtain binding, processing, and presentation predictions,
1393 with results filtered for affinity percentile < 2.0. Data from the observed and predicted normal
1394 ligandomes were combined to obtain a unified non-redundant reference dataset of unique
1395 $pMHC$/protein combinations (*unified normal ligandome*, supplementary data). Those $pMHCs$
1396 with MHCflurry presentation score ≤ 0.5 were removed from the unified normal ligandome
1397 unless they were listed in the observed ligandome. Then for each selected tumoral peptide we
1398 measured its similarity with normal peptides accounting for both their physicochemical and HLA
1399 binding properties. Specifically, given a tumoral peptide p_t and a normal reference peptide p_n
1400 both binding to an HLA allele A, peptide similarity S_{sim} was calculated as:
1401

1402 $$S_{sim}(A, p_t, p_n) = \frac{simPMBEC - min(simPMBEC)}{max(simPMBEC) - min(simPMBEC)} + \frac{simAAindex - min(simAAindex)}{max(simAAindex) - min(simAAindex)}$$

1403

1404 Where $simPMBEC$ is a similarity score based on the peptide:MHC binding energy covariance
1405 (PMBEC) matrix[80] which measures amino acid similarity based on their effect on peptide HLA
1406 binding. Specifically, $simPMBEC$ was calculated as:
1407

1408 $$\frac{\sum_{i=1}^{L}\left(\frac{1}{1 + distPMBEC(p_{ti}, p_{ni})} \times (1 + H_{Ai})\right)}{L}$$

1409 Where $distPMBEC = max(PMBEC) - PMBEC$, $PMBEC$ is the value in the PMBEC matrix for the
1410 pair of amino acids at the ith position of p_t and p_n, L is the peptide length (9 or 10 amino acids),
1411 and H_{Ai} is the entropy at the i^{th}peptide residue calculated for all the peptides of same length as
1412 p_t binding to the HLA allele A in both the observed and predicted ligandome. This was obtained
1413 using the MolecularEntropy function from the R package HDMD[151] (version 1.2), with values
1414 ranging from 0 to 1, indicating respectively no amino acid sequence variability and maximum
1415 sequence variability at the ith peptide residue.

1416

1417 On the other end, $simAAindex$ was calculated as follows:

1418
$$\frac{\sum_{i=1}^{L}\left(\frac{1}{1+distAAindex(p_{ti},p_{ni})}\times(1+H_{Ai})\right)}{L}$$

1419 Where $distAAindex$ is the Euclidean distance calculated between the pair of amino acids at the
1420 i^{th} position of p_t and p_n, considering the first nine principal components of a PCA dimensionality
1421 reduction of 554 physicochemical amino acid properties available from AAindex[79] and obtained
1422 through the R package ftrCOOL[152] (version 2.0.0), L is the peptide length, and H_{Ai} is the residue
1423 entropy calculated as above. The results of the peptide similarity analysis for the candidate
1424 $pMHCs$ for each target antigen are available in the supplementary data.

1425

1426 To define a significant threshold for S_{sim}, we tested PS-TRACT on known cross-reactive peptide
1427 pairs[85] and found that a similarity score of 0.95 was the lowest observed for cross-reactive
1428 peptides. Also, we ran PS-TRACT between the *MAGEA3*-derived peptide EVDPIGHLY and the
1429 *TTN*-derived peptide ESDPIVAQY, which was found to be the off-tumor target leading to cases
1430 of fatal myocarditis in patients that received EVDPIGHLY-targeted TCR-T cells[86], obtaining a
1431 peptide similarity score of 1.19. Based on these results, we chose a threshold of $S_{sim} > 1.15$ as
1432 this provided the highest specificity in identifying known cross-reactive peptides (**Figure S11A**).
1433 Thus, after running PS-TRACT on each candidate tumoral $pMHC$, we selected only the
1434 corresponding normal $pMHCs$ with $S_{sim} > 1.15$ for further cross-reactivity assessment.
1435 Specifically, for each of these highly similar normal $pMHCs$ we calculated a tissue-specific
1436 cross-reactivity score, S_{CR}, that considers the tissue abundance of the antigen from which the
1437 cross-reactive peptide originates, the likelihood of presentation based on MHCflurry, and the
1438 presence of the peptide in the observed ligandome:

1439

$$S_{CR}(pMHC_y, pMHC_i) = (P_{MHCflurry} + 0.5) \times (E_{HPA/GTEx}) \times \begin{cases} 1.25 \; if \; pMHC_y \in observed \; ligandome \\ 1 \; if \; pMHC_y \notin observed \; ligandome \end{cases}$$

Where $pMHC_y$ is the yth cross-reactive $pMHC$ for a given tumoral $pMHC_i$, $P_{MHCflurry}$ is the MHCflurry presentation score for $pMHC_y$, and $E_{HPA/GTEx}$ is the median-centered, z-scored RNA expression for the parent antigen of the $pMHC_y$ peptide in the HPA and GTEx bulk RNA-seq consensus datasets (source: https://www.proteinatlas.org/about/download, file: rna_tissue_consensus.tsv.zip). Then for each tumoral $pMHC$ in each of the 50 normal tissues, we selected the maximum S_{CR} score among all the filtered cross-reactive normal $pMHCs$, and this vector of S_{CR} scores was used to define the cross-reactivity profile of the tumoral target $pMHC$ considered.

Finally, we benchmarked the cross-reactivity profile of each candidate tumoral $pMHC$ pair against that of the $pMHC$ targeted by afami-cel (GVYDGREHTV–HLA-A*02:01) which was used as reference given afami-cel's promising safety and tolerability profile observed in clinical trials[27,28]. Specifically, for each candidate tumoral peptide-MHC pair $pMHC_i$, we calculated the difference $S_{CR}Diff(i)$ between its S_{CR} cross-reactivity profile and that of GVYDGREHTV–HLA-A*02:01, as follows:

$$S_{CR}Diff(i) = \frac{1}{n}\sum_{j=1}^{n} \begin{cases} \left(\left(S_{CR}^{ij} - S_{CR}^{refj}\right) / \sigma_j\right) \times \alpha \; if \; S_{CR}^{ij} > S_{CR}^{refj} \\ \left(S_{CR}^{ij} - S_{CR}^{refj}\right) / \sigma_j \qquad otherwise \end{cases}$$

Where S_{CR}^{ij} is the S_{CR} score in the jth tissue of $pMHC_i$, while S_{CR}^{refj} is the S_{CR} score in the jth tissue of the reference $pMHC$ (GVYDGREHTV–HLA-A*02:01), n is the number of tissues

1463 considered, σ_j is the standard deviation of the S_{CR} scores for all the $pMHC$ pairs in the jth tissue,
1464 and $\alpha = 2$. Target tumoral $pMHCs$ with differential S_{CR} profile compared to GVYDGREHTV–HLA-
1465 A*02:01 ($S_{CR}Diff$) ≤ 0 were considered to have acceptable risk of cross-reactivity.
1466
1467 **Clinical STS proteomics datasets**
1468 Two clinical STS MS proteomics datasets were used to validate the expression of the identified
1469 cell therapy targets at the protein level, one from Tang et al.[91] and one from Burns et al.[92] The
1470 dataset from Tang et al. was obtained from the supplementary material of the relative paper
1471 ("Proteomic_matrix(tumor)" and "Clinical_information" tables). The relative fraction of total (Fot)
1472 protein intensities for the Tang et al. dataset were $\log_2(+1)$ transformed and then median
1473 centered across samples and z-scored across all proteins for standardization. Protein symbols
1474 were standardized to the current HGNC symbols before analysis. The proteomics dataset from
1475 Burns et al. was obtained from the supplementary material (Supplementary Data 2). The
1476 expression matrix in this case was used as is given that protein expression values were already
1477 log-transformed, and median centered across samples and z-scored across proteins. Protein
1478 symbols were standardized to the current HGNC symbols before analysis. For SCAN-ACT
1479 implementation in other cancers, we obtained pre-processed mass spectrometry-based
1480 proteomics data from either LinkedOmicsKB[153], the Proteomic Data Commons (PDC)[154], or from
1481 a literature search (**Table S34**). Most of the data was derived from studies conducted by Clinical
1482 Proteomic Tumor Analysis Consortium (CPTAC) or the International Cancer Proteogenome
1483 Consortium (ICPC). For data obtained from LinkedOmicsKB, which was provided as $\log_2$
1484 reference-normalized protein abundance, we applied median-centering across samples and z-
1485 scoring across proteins. In contrast, for data from the PDC, where the $\log_2$ reference-normalized
1486 protein abundance was already standardized across proteins, we performed only median-
1487 centering across samples. For the gastric adenocarcinoma dataset, expression data from

1488 different technical replicates of the same patient was averaged before analysis. For other
1489 cancers, expression data provided in the supplementary materials of the related publications
1490 was used[155–158]. For the prostate cancer dataset, which was provided as log2 reference-
1491 normalized protein abundance, we applied median-centering across samples and z-scoring
1492 across proteins. For the melanoma and esophageal cancer datasets, only median-centering
1493 across samples was necessary, as the data was already standardized across proteins. The
1494 bladder cancer expression data was used as provided by the authors, as it was already
1495 standardized across both samples and proteins. Prior to analysis, gene symbols across all
1496 datasets were updated to the current HGNC version. Proteins with missing values in 50% or
1497 more samples were removed, and the remaining missing values were imputed with a value of
1498 1e-6. For genes with multiple protein isoforms, the isoform with the highest expression was
1499 used.
1500
1501 **STS cell line proteomics data**
1502 Protein expression data in soft tissue sarcoma cell lines was obtained from the Cancer Cell Line
1503 Encyclopedia[93] (CCLE) and the ProCan-DepMapSanger repository. For CCLE, the normalized
1504 $\log_2$ transformed relative protein intensities were obtained from
1505 https://gygi.hms.harvard.edu/publica Shtions/ccle.html (file:
1506 protein_quant_current_normalized.csv). Gene symbols were standardized to the current HGNC
1507 symbols, and for different protein isoforms for a given gene symbol we retained the isoform with
1508 maximum expression (supplementary data). Then, data was median-centered across all cell
1509 lines and z-scored within genes before use for downstream analysis. The following “soft tissue”
1510 cell lines were considered in the CCLE dataset for further analysis: "A-204", "RD", "RH-41", "HT-
1511 1080", "RH-30", and "KYM-1". The ProCan-DepMapSanger dataset was obtained from the
1512 figshare repository associated with the corresponding publication:
1513 https://doi.org/10.6084/m9.figshare.19345397 (file: ProCan-

1514 DepMapSanger_protein_matrix_8498_averaged.txt). Similarly to the CCLE dataset, first we
1515 standardized gene symbols to the current HGNC nomenclature, then for genes with more than
1516 one protein isoform we retained the isoform with highest expression (supplementary data).
1517 Since the data was already provided as $\log_2$ transformed relative protein intensities, the protein
1518 per cell line expression matrix was median-centered across cell lines and z-scored within genes
1519 before further analysis. The following soft tissue sarcoma cell lines were considered for
1520 downstream analysis: "A204" , "ESS-1", "G-402", "GCT", "Hs-633T", "HT-1080" , "KYM-1",
1521 "MES-SA", "MFH-ino", "RD", "RH-18", "RH-41", "RKN", "SJRH30", "SK-LMS-1", "SK-UT-1" ,
1522 "SKN", "SW684", "SW872", "SW982", "TE-441-T", and "VA-ES-BJ".
1523
1524 **IHC of Selected CAR-T Targets**
1525 IHC was performed by the Stanford Human Pathology/Histology Service Center for ACVR1,
1526 AXL, CLMP, GPC6, KITLG, and MRC2. Previously obtained tissue microarrays constructed
1527 using one to seven 1-mm FFPE cores per patient were used for the analysis, including UPS,
1528 LPS, LMS, SS, and MPNST samples. The following antibodies were used for the analysis:
1529 ACVR1 Rabbit IgG (cat. HPA007505, Sigma Aldrich, MO, USA), AXL Rabbitt IgG (cat.
1530 HPA037422, Sigma Aldrich, MO, USA), GPC6 (cat. HPA017671, Sigma Aldrich, MO, USA),
1531 KITLG (cat. HPA070395, HPA017671), CLMP (cat. 16127-1-AP, Thermo Fisher Scientific, MA,
1532 USA), and MRC2 (cat. HPA041991, Sigma Aldrich, MO, USA). All the antibodies were used at a
1533 dilution titer of 1:50 except for the CLMP antibody (1:400), and the KITLG antibody which was
1534 used at a 1:100, 1:200, or 1:50 dilution depending on the TMA analyzed. Tumor cell staining
1535 intensity was scored using a three-point scale (0, negative; 1, weakly positive; and 2, strongly
1536 positive). Staining for FAP was performed at the UCLA pathology department on a tissue
1537 microarray of 60 undifferentiated pleomorphic sarcomas in triplicate (no previous treatment), of
1538 which 57 samples were evaluable. FAP immunostaining was performed on Leica Bond platform
1539 using a polyclonal rabbit anti-FAP alpha antibody (Abcam, ab207178). All slides were scanned

1540 at 20X magnification. The percentage of FAP expressing cells was quantified using the HALO®
1541 image analysis platform based on automatic cell segmentation (Indica Labs). Missing or folded
1542 tissue cores were excluded. Replicate cores were averaged.
1543
1544 **Measurement of peptide-MHC binding through IMMUNE**
1545 *Plasmid construction*
1546 To construct plasmids driving the expression of non-covalent HLA-I heterodimers, the
1547 commercially available backbone vector pRS315 was modified[89]. Briefly, the entire expression
1548 cassette containing genes encoding the Aga2 leader sequence, the ectodomain of a particular
1549 HLA-I heavy chain allele (A*02:01 or A*11:01 or B*35:01 or C*12:03), the Aga2p yeast native
1550 protein subunit, an HA epitope tag, and the MFα terminator was cloned to the downstream of
1551 the GAL10 promoter in pRS315, so that the expressed HLA-I heavy chain ectodomain was
1552 fused to the N-terminus of Aga2p for yeast surface display. Subsequently, a second expression
1553 cassette containing genes encoding the Aga2 leader sequence, β2m, a C-myc epitope tag and
1554 the MFα terminator was cloned to the downstream of the GAL1 promoter in the same pRS315-
1555 based vector, allowing the secretion of the β2m domain for HLA heavy chain/β2m dimerization
1556 and surface detection. In addition, the Aga2p subunit was further removed from the above
1557 plasmids, which resulted in control plasmids yielding no HLA detection on the surface of yeast.
1558 Gene synthesis and subcloning services were provided by GenScript (Jiangsu, China).
1559
1560 *Electrotransformation of yeast cells and expression of HLA*
1561 The *Saccharomyces cerevisiae* yeast parent strain, EBY100[159] was used for generation of
1562 daughter strains displaying each of the aforementioned HLA alleles. The constructed plasmids
1563 were transformed into EBY100 one at a time by electroporation, following the MicroPulser
1564 electroporation mannual (BioRad). Transformed cells that received the corresponding plasmid
1565 were recovered on selected dextrose-based (SD) medium agar plates minus the nutrient

1566 corresponding to the selection marker gene within the plasmid. Single colonies on SD agar
1567 plates appeared after 2-3 days of incubation at 30°C. To induce yeast expression of target
1568 proteins, a single yeast colony was inoculated with liquid SD medium and cultured in a shaking
1569 tube at 30°C until it reached the exponential growth phase. A sufficient amount of yeast cells
1570 was then transferred from the SD medium to selected galactose-based (SG) medium for protein
1571 expression under the promotion of GAL1 or GAL10. Typically, an overnight shaking and
1572 induction at 30°C can promote effective expression of the protein of interest in yeast for surface
1573 detection by flow cytometry.
1574
1575 *Immunofluorescent labeling and flow cytometric analysis*
1576 A sufficient amount of galactose-induced yeast cells was collected by centrifugation and labeled
1577 by primary antibodies in PBS + 0.5% bovine serum albumin (BSA) for 1 h on ice. The primary
1578 antibody was rabbit anti-HA tag monoclonal antibody (mAb clone C29F4, Cell signaling), mouse
1579 anti-human β2m mAb (clone 2M2, Biolegend), or mouse anti-HLA-A, B, C conformation-specific
1580 mAb (clone W6/32, Biolegend). After the primary staining, the cells were washed at least twice
1581 with ice cold PBS + 0.5% BSA and labeled by secondary antibodies on ice for another 1 h. The
1582 secondary antibody was Alexa Fluor 488 conjugated goat anti-rabbit IgG (H+L) highly cross-
1583 adsorbed secondary antibody, Alexa Fluor 488 or Alexa Fluor 647 conjugated goat anti-mouse
1584 IgG (H+L) highly cross-adsorbed secondary antibody (Thermo Fisher Scientific). After the
1585 secondary staining, cells were washed with ice cold PBS + 0.5% BSA and resuspended in PBS
1586 + 0.5% BSA for flow cytometric analysis. A FACSCalibur flow cytometer (BD biosciences) was
1587 used to acquire flow data. At least 10000 cell events were collected for each sample and gated
1588 through forward and side scattering. FL1 and FL4 fluorescence channels were selected for
1589 Alexa Fluor 488 and Alexa Fluor 647, respectively, according to the fluorophore selected at the
1590 secondary labeling step as mentioned above. Flowjo (BD biosciences) was used for data
1591 analysis.

1592

1593 *Peptide design and synthesis*

1594 To design indicator peptides appropriate for evaluation of HLA peptide binding on the surface of
1595 yeast cells and for downstream competitive assay development, the reported epitope repertoire
1596 for each aforementioned HLA allele from the Immune Epitope Database (https://www.IEDB.org)
1597 was carefully analyzed. First, an epitope peptide was considered only if it contained a cysteine
1598 residue at P4 or P5 or P6 anchor position available for site-specific modification, i.e.,
1599 biotinylation or fluorophore conjugation. After that, the chosen peptide was further analyzed and
1600 validated 1) based on its predicted %Rank_EL (<2) and %Rank_BA scores (<2) from the
1601 NetMHCpan-4.1 webserver, and 2) based on its sequence similarity to the peptides with
1602 available X-ray structure in complex with the given HLA. For example, Bio-KLT (amino acids:
1603 KLTPL{Cys-Biotin}VTL), Bio-AIF (aa: AIFQ{Cys-Biotin}SMTK), Bio-RYP (aa: RYPV{Cys-
1604 Biotin}KFPSL), and Bio-FAF (aa: FAF{Cys-Biotin}RITSF) were designed as the indicator peptide
1605 for A*02:01, A*11:01, B*35:01, and C*12:03, respectively. All peptides were synthesized
1606 (GenScript) with over 95% purify.

1607 *On yeast peptide binding*

1608 About half a million yeast cells displaying a peptide-receptive HLA were collected by
1609 centrifugation, resuspended in 40 µl of PBS, and incubated with 0.5 uM biotinylated indicator
1610 peptides ([peptide]>>[HLA]) overnight prior to fluorescent labeling for flow cytometry. After
1611 peptide binding reached equilibrium, yeast cells were washed twice with ice-cold PBS + 0.5%
1612 BSA and then simultaneously stained for bound biotinylated indicator peptides and HLA
1613 molecules on the surface using Alexa Fluor 647 conjugated streptavidin and HLA conformation-
1614 specific mAb W6/32 followed by Alexa Fluor 488 conjugated goat anti-mouse IgG (H+L) highly
1615 cross-adsorbed secondary antibody (Thermo Fisher Scientific). Double stained cells were then
1616 washed and resuspended in ice-cold PBS + 0.5% BSA for analysis on a BD FACSCalibur. Flow
1617 cytometric data was analyzed using FlowJo.

1618

1619 *On yeast peptide competition*

1620 To evaluate the competitive binding of a test peptide against the indicator peptide to HLA on the
1621 surface of yeast cells, about half a million yeast cells were collected and incubated with the
1622 biotinylated peptide as mentioned above, but with addition of a 20-fold more concentrated test
1623 peptide in its non-biotinylated format. The peptide competition assay was also performed under
1624 a similar condition as mentioned above, allowing reaction to reach equilibrium. Yeast cells were
1625 then collected, washed, and co-stained for bound biotinylated indicator peptides and HLA
1626 molecules using the aforementioned fluorescent labeling method for flow cytometric analysis.
1627 The %Binding in the presence of competitors was quantified as $[(MFI_{with\ competitor}-BG)/(MFI_{no}$
1628 $_{competitor}-BG)]\times 100\%$, where $MFI_{with\ competitor}$ represents the average fluorescence intensity in the
1629 presence of test peptide (a competitor), $MFI_{no\ competitor}$ represents the average fluorescence
1630 intensity in the absence of competitors, and BG represents the background fluorescence
1631 intensity. The %Competition was calculated as 100%−%Binding to reveal binding of the test
1632 peptide competitor to HLA molecules displayed on yeast. MFI was determined using
1633 appropriately gated cell populations.

1634

1635 **Target Priority Index**

1636 For each target discovery pipeline, we calculated a final Target Priority Index (TPI) that
1637 integrates normal tissues/cells target expression, scRNA-seq tumor expression and
1638 homogeneity, STS protein expression, DepMap sarcoma cell line essentiality, incidence of
1639 SNVs/Indels, gene fusions and CNAs in STS involving the target considered, and the presence
1640 of clinically or pre-clinically developed immunotherapeutic agents against each target (for
1641 monospecific and bispecific CAR-T cells targets only).

1642

1643 Homogeneity of expression at the single-cell level in each STS was calculated as the product of
1644 the average tumor expression in $\log_2$(CPM+1) and the percentage of tumor cells with
1645 $\log_2$(CPM+1) > 0, which we defined as H_{RNA} score:

1646 $$H_{RNA} = \left(\overline{log_2(CPM+1)}\right) \times \left(\frac{n\ cells\ log_2(CPM+1) > 0}{total\ cells} \times 100\right)$$

1647 To assess if a gene has been previously studied as an immunotherapy target, we focused on
1648 therapeutic monoclonal antibodies (mAb), ADCs, single-chain variable fragments (scFv), fusion
1649 proteins, nanobodies, and fragment antigen binding regions (Fab), as well as CAR-T cells.
1650 Specifically, we downloaded a list of clinically developed mAbs, ADCs, scFvs, Fabs, nanobodies
1651 and fusion proteins from Thera-SAbDab[160] (https://opig.stats.ox.ac.uk/webapps/sabdab-
1652 sabpred/therasabdab, accessed on 07/16/2024, Supplementary Data), retaining only agents
1653 developed for human use, and removing all those agents with clinical development phase
1654 reported as "TBC", "Preregistration", or "Preregistration (w)". We complemented Thera-SAbDab
1655 with a PubMed search for pre-clinically developed therapeutic antibodies or ADCs against any
1656 genes in the surfaceome, using the following search query: "gene AND antibody, OR gene AND
1657 antibody drug conjugate, OR gene AND ADC OR gene AND antibody-drug conjugate" (**Table**
1658 **S30**). Targets with clinical development phase reported as "TBC", "Preregistration", or
1659 "Preregistration (w)" in Thera-SAbDab were included as pre-clinical targets, unless they were
1660 described as clinical targets. As far CAR-T cells, for clinically developed targets we used the
1661 previously defined list of targets as per MacKay et al[45], while for pre-clinically developed targets
1662 we performed a PubMed search for all the genes in the surfaceome using the following search
1663 query: "gene AND CAR-T OR gene AND chimeric antigen receptor T-cells OR gene AND
1664 chimeric antigen receptor", **Table S30**).
1665
1666 *Monospecific CAR-T Cell Target TPI Calculation*

1667 To obtain the TPI for monospecific CAR-T targets, we used the list of targets defined based on
1668 the TS_{CAR} as described above. Then, for these targets we normalized the TS_{CAR} to assume
1669 values between 0 and 1 as follows:

1670 $$TS_{CAR}(g)(norm) = \frac{TS_{CAR}(g) - \min(TS_{CAR})}{\max(TS_{CAR}) - \min(TS_{CAR})}$$

1671 Where $TS_{CAR}(g)$ is the TS_{CAR} for a target g, and $\min(TS_{CAR})$ and $\max(TS_{CAR})$ represents the
1672 minimum and maximum TS_{CAR} values, respectively, for all the final targets being considered for
1673 the STS histotype analyzed.
1674
1675 Then, for each target g, we obtained a summative validation score, V_{CAR}, calculated as the sum
1676 of different scores assigned for each validation feature, defined as follows:

1677 - If the H_{RNA} score for g is equal or greater than the average H_{RNA} for *CD276* across the
1678 five STS studied, g is assigned a score of 2, otherwise g is assigned a score of 0.
1679 *CD276* was used as reference as it has been clinically (NCT04433221) and pre-
1680 clinically[95] studied as a CAR-T target in STS.
1681 - For each gene g and STS histotype we measured the percentage of samples in the STS
1682 proteomics Tang et al. dataset with medium-high (M-H) expression (z-score ≥ 0).
1683 DDLPS and WDLPS samples were considered as one group (LPS) for this analysis. If
1684 the percentage of samples of the STS analyzed with M-H expression for g was greater
1685 than or equal to the average percentage of samples with M-H expression for CD276
1686 across SS, MFS, UPS, LMS and LPS, then we assigned a score of 2, otherwise we
1687 assigned a score of 0.
1688 - If g was reported as target for clinically developed mAbs, ADCs, scFv, Fabs, nanobodies
1689 or fusion proteins, we assigned a score of 1. If instead g was only studied preclinically as
1690 target for therapeutic antibodies or ADCs, we assigned a score of 0.5. Otherwise, we
1691 assigned a score of 0.

- If g was reported as target for clinically developed CAR-T cells, we assigned a score of 1, while we assigned a score of 0.5 if g was only explored as a CAR-T target preclinically, and a score of 0 if it was never tested as a CAR-T target.
- If g resulted as essential (Chronos DepMap score ≤ -1) in sarcoma cell lines derived from the STS histotype studied, we assigned a score of 1, otherwise we assigned a score of 0.5 if g was essential in any sarcoma cell line, or a score of 0 if g was not essential in any sarcoma cell lines.
- If g was involved in at least one gene fusion in the STS histotype studied we assigned a score of 1, otherwise we assigned a score of 0.5 if g was involved in at least one gene fusion in any STS, or a score of 0 if g was never involved in gene fusions in STS.
- If g was involved by at least one SNV or Indel in the STS histotype studied we assigned a score of 1, otherwise we assigned a score of 0.5 if g was involved by at least one SNV/Indel in any STS histotype, or a score of 0 if g was never involved by SNVs/Indels.
- If g was involved by CNAs in the STS histotype studied and there was a positive correlation between its linear CNV level and RNA expression in TCGA, we assigned a score of 1, otherwise if it was involved by CNAs either in TCGA or MSK-IMPACT in the same STS histotype studied we assigned a score of 0.5. Instead, if g was never involved by CNAs in STS we assigned a score of 0.

We then normalized V_{CAR} to assume values ranging from 0 to 1 as follows:

$$V_{CAR}(g)(norm) = \frac{V_{CAR}(g) - \min(V_{CAR})}{\max(V_{CAR}) - \min(V_{CAR})}$$

Finally, we calculated the TPI for each gene g as follows:

$$TPI_{CAR} = TS_{CAR}(g)(norm) + V_{CAR}(g)(norm)$$

The TPI for the monospecific CAR-T targets is reported in **Table S31**.

Bispecific CAR-T Cells Targets TPI Calculation

1717 For bispecific CAR-T cells targets we obtained a TPI for logic-gated 'AND' CAR-T cells

1718 ($TPI_{'AND'}$), and a separate TPI for 'AND-NOT' CAR-T cells ($TPI_{'AND-NOT'}$). Specifically, we

1719 defined suitable target pairs for both 'AND' and 'AND-NOT' CAR-T cells as described above,

1720 and then for each 'AND' CAR-T target and 'AND-NOT' CAR-T activator we calculated the

1721 average $TS_{'AND'}$ and $TS_{'AND-NOT'}$ scores considering all the target pairs in which that gene

1722 appears. We then normalized the average $TS_{'AND'}$ and $TS_{'AND-NOT'}$ for each gene g to assume

1723 values ranging from 0 to 1 as follows:

1724 $$TS_{'AND'}(g)(norm) = \frac{\overline{TS_{'AND'}}(g) - \min(\overline{TS_{'AND'}})}{\max(\overline{TS_{'AND'}}) - \min(\overline{TS_{'AND'}})}$$

1725

1726 $$TS_{'AND-NOT'}(g)(norm) = \frac{\overline{TS_{'AND-NOT'}}(g) - \min(\overline{TS_{'AND-NOT'}})}{\max(\overline{TS_{'AND-NOT'}}) - \min(\overline{TS_{'AND-NOT'}})}$$

1727 Then, for each target g we calculated $V_{CAR}(g)$ and $V_{CAR}(g)(norm)$ as described above, and

1728 finally obtained the TPIs as follows:

1729 $$TPI_{'AND'} = TS_{'AND'}(g)(norm) + V_{CAR}(g)(norm)$$

1730 $$TPI_{'AND-NOT'} = TS_{'AND-NOT'}(g)(norm) + V_{CAR}(g)(norm)$$

1731 The TPIs for the bispecific CAR-T targets are reported in **Table S32**.

1732

1733 *TCR and PC-CAR-T Cells Targets TPI Calculation*

1734 For each of the defined target antigens g for TCR-T and PC-CAR-T cells in each STS, we first

1735 normalized the $N(g)$ score, indicative of the normal tissue and cell expression of g as previously

1736 defined, to assume values between 0 and 1:

1737 $$N(g)_{norm} = \frac{N(g) - \min(N)}{\max(N) - \min(N)}$$

1738 Then, for each target g we obtained a summative validation score, $V_{TCR/PC}$, calculated as the

1739 sum of the following scores:

1740 • If the H_{RNA} score for g is equal or greater than the H_{RNA} for *MAGEA4* in SS, g is
1741 assigned a score of 2, otherwise is assigned a score of 0. *MAGEA4* was used as
1742 reference in this case since it is the source antigen for the target of afami-cel, the first
1743 TCR-T cell therapy to receive FDA approval in soft tissue sarcomas.
1744 • If g had a z-score ≥ 0 in at least two separate STS cell lines in either CCLE or in the
1745 ProCan-DepMapSanger datasets, we assigned a score of 1, otherwise we assigned a
1746 score of 0.
1747 • If g resulted as essential (Chronos DepMap score ≤ -1) in sarcoma cell lines derived
1748 from the STS histotype studied, we assigned a score of 1, otherwise we assigned a
1749 score of 0.5 if g was essential in any sarcoma cell line, or a score of 0 if g was not
1750 essential in any sarcoma cell lines.
1751 • If g was involved in at least one gene fusion in the STS histotype studied we assigned a
1752 score of 1, otherwise we assigned a score of 0.5 if g was involved in at least one gene
1753 fusion in any STS, or a score of 0 if g was never involved in gene fusions in STS.
1754 • If g was involved by at least one SNV or Indel in the STS histotype studied we assigned
1755 a score of 1, otherwise we assigned a score of 0.5 if g was involved by at least one
1756 SNV/Indel in any STS histotype, or a score of 0 if g was never involved by SNVs/Indels.
1757 • If g was involved by CNAs in the STS histotype studied and there was a positive
1758 correlation between its linear CNV level and RNA expression in TCGA, we assigned a
1759 score of 1, otherwise if it was involved by CNAs either in TCGA or MSK-IMPACT in the
1760 same STS histotype studied we assigned a score of 0.5. Instead, if g was never involved
1761 by CNAs in STS we assigned a score of 0.
1762 We then normalized $V_{TCR/PC}$ to assume values ranging from 0 to 1 as follows:

1763 $$V_{TCR/PC}(g)(norm) = \frac{V_{TCR/PC}(g) - \min(V_{TCR/PC})}{\max(V_{TCR/PC}) - \min(V_{TCR/PC})}$$

Next, for each target antigen we considered the target $pMHCs$ defined based on their $TS_{TCR/PC}$ as reported in **Table S27**. Specifically, for each target antigen g we calculated the average HLA allele frequency at each locus, by considering all the peptides derived from each antigen binding to HLA-A, HLA-B, or HLA-C class alleles, as follows:

$$HLA_A(f_{TOT}) = \frac{\sum_{i=1}^{pA} HLA_A(f)}{pA}$$

$$HLA_B(f_{TOT}) = \frac{\sum_{i=1}^{pB} HLA_B(f)}{pB}$$

$$HLA_C(f_{TOT}) = \frac{\sum_{i=1}^{pC} HLA_C(f)}{pC}$$

Where pA, pB, pC are respectively the number of $pMHC$ pairs including class A, class B and class C alleles for the target g, while $HLA_A(f), HLA_B(f)$, and $HLA_C(f)$ represent the HLA allele frequencies of the $i^{th} pMHC$ considered. We then obtained the total HLA allele frequency, $HLA_{TOT}(g)$ for each target g by adding $HLA_A(f_{TOT})$, $HLA_B(f_{TOT})$, and $HLA_C(f_{TOT})$, and normalized this to obtain values ranging from 0 to 1, as follows:

$$HLA_{TOT}(g)(norm) = \frac{HLA_{TOT}(g) - \min(HLA_{TOT})}{\max(HLA_{TOT}) - \min(HLA_{TOT})}$$

Next, for each target g we calculated the total number of $pMHC$ targets ($pMHC_{TOT}$) and the number of $pMHC$ targets with $S_{CR}Diff \leq 0$ ($pMHC_{SCR}$). We divided these by the target antigen total number of amino acids (L_{AA}) to account for differences in protein size, and then normalized them to obtain values ranging from 0 to 1, as follows:

$$pMHC_{TOT} = \frac{n(pMHC_{TOT})}{L_{AA}}$$

$$pMHC_{TOT}(g)(norm) = \frac{pMHC_{TOT}(g) - \min(pMHC_{TOT})}{\max(pMHC_{TOT}) - \min(pMHC_{TOT})}$$

$$pMHC_{SCR} = \frac{n(pMHC_{SCR})}{L_{AA}}$$

1785 $$pMHC_{SCR}(g)(norm) = \frac{pMHC_{SCR}(g) - \min(pMHC_{SCR})}{\max(pMHC_{SCR}) - \min(pMHC_{SCR})}$$

1786 Finally, we calculated the TPI for each target g as follows:

1787 $$TPI_{TCR/PC} = N(g)_{norm} + V_{TCR/PC}(g)(norm) + HLA_{TOT}(g)(norm) + pMHC_{TOT}(g)(norm)$$

1788 $$+ pMHC_{SCR}(g)(norm)$$

1789 The TPIs for the TCR-T and PC-CAR-T cells targets are reported in **Table S33**.

1790

1791 **References**

1792 1. Wagner, J., Wickman, E., DeRenzo, C., and Gottschalk, S. (2020). CAR T Cell Therapy for

1793 Solid Tumors: Bright Future or Dark Reality? Mol. Ther. J. Am. Soc. Gene Ther. *28*, 2320–

1794 2339. https://doi.org/10.1016/j.ymthe.2020.09.015.

1795 2. Neelapu, S.S., Locke, F.L., Bartlett, N.L., Lekakis, L.J., Miklos, D.B., Jacobson, C.A.,

1796 Braunschweig, I., Oluwole, O.O., Siddiqi, T., Lin, Y., et al. (2017). Axicabtagene Ciloleucel

1797 CAR T-Cell Therapy in Refractory Large B-Cell Lymphoma. N. Engl. J. Med. *377*, 2531–

1798 2544. https://doi.org/10.1056/NEJMoa1707447.

1799 3. Maude, S.L., Laetsch, T.W., Buechner, J., Rives, S., Boyer, M., Bittencourt, H., Bader, P.,

1800 Verneris, M.R., Stefanski, H.E., Myers, G.D., et al. (2018). Tisagenlecleucel in Children

1801 and Young Adults with B-Cell Lymphoblastic Leukemia. N. Engl. J. Med. *378*, 439–448.

1802 https://doi.org/10.1056/NEJMoa1709866.

1803 4. Labanieh, L., and Mackall, C.L. (2023). CAR immune cells: design principles, resistance

1804 and the next generation. Nature *614*, 635–648. https://doi.org/10.1038/s41586-023-

1805 05707-3.

1806 5. Adachi, K., Kano, Y., Nagai, T., Okuyama, N., Sakoda, Y., and Tamada, K. (2018). IL-7 and
1807 CCL19 expression in CAR-T cells improves immune cell infiltration and CAR-T cell
1808 survival in the tumor. Nat. Biotechnol. *36*, 346–351. https://doi.org/10.1038/nbt.4086.
1809 6. Lynn, R.C., Weber, E.W., Sotillo, E., Gennert, D., Xu, P., Good, Z., Anbunathan, H., Lattin,
1810 J., Jones, R., Tieu, V., et al. (2019). c-Jun overexpression in CAR T cells induces
1811 exhaustion resistance. Nature *576*, 293–300. https://doi.org/10.1038/s41586-019-1805-
1812 z.
1813 7. Rafiq, S., Yeku, O.O., Jackson, H.J., Purdon, T.J., van Leeuwen, D.G., Drakes, D.J., Song,
1814 M., Miele, M.M., Li, Z., Wang, P., et al. (2018). Targeted delivery of a PD-1-blocking scFv by
1815 CAR-T cells enhances anti-tumor efficacy in vivo. Nat. Biotechnol. *36*, 847–856.
1816 https://doi.org/10.1038/nbt.4195.
1817 8. Narayan, V., Barber-Rotenberg, J.S., Jung, I.-Y., Lacey, S.F., Rech, A.J., Davis, M.M.,
1818 Hwang, W.-T., Lal, P., Carpenter, E.L., Maude, S.L., et al. (2022). PSMA-targeting TGFβ-
1819 insensitive armored CAR T cells in metastatic castration-resistant prostate cancer: a
1820 phase 1 trial. Nat. Med. *28*, 724–734. https://doi.org/10.1038/s41591-022-01726-1.
1821 9. Doan, A.E., Mueller, K.P., Chen, A.Y., Rouin, G.T., Chen, Y., Daniel, B., Lattin, J.,
1822 Markovska, M., Mozarsky, B., Arias-Umana, J., et al. (2024). FOXO1 is a master regulator
1823 of memory programming in CAR T cells. Nature *629*, 211–218.
1824 https://doi.org/10.1038/s41586-024-07300-8.
1825 10. Larson, R.C., and Maus, M.V. (2021). Recent advances and discoveries in the
1826 mechanisms and functions of CAR T cells. Nat. Rev. Cancer *21*, 145–161.
1827 https://doi.org/10.1038/s41568-020-00323-z.

1828 11. Baulu, E., Gardet, C., Chuvin, N., and Depil, S. (2023). TCR-engineered T cell therapy
1829 in solid tumors: State of the art and perspectives. Sci. Adv. *9*, eadf3700.
1830 https://doi.org/10.1126/sciadv.adf3700.
1831 12. Tousley, A.M., Rotiroti, M.C., Labanieh, L., Rysavy, L.W., Kim, W.-J., Lareau, C.,
1832 Sotillo, E., Weber, E.W., Rietberg, S.P., Dalton, G.N., et al. (2023). Co-opting signalling
1833 molecules enables logic-gated control of CAR T cells. Nature *615*, 507–516.
1834 https://doi.org/10.1038/s41586-023-05778-2.
1835 13. Choe, J.H., Watchmaker, P.B., Simic, M.S., Gilbert, R.D., Li, A.W., Krasnow, N.A.,
1836 Downey, K.M., Yu, W., Carrera, D.A., Celli, A., et al. (2021). SynNotch-CAR T cells
1837 overcome challenges of specificity, heterogeneity, and persistence in treating
1838 glioblastoma. Sci. Transl. Med. *13*, eabe7378.
1839 https://doi.org/10.1126/scitranslmed.abe7378.
1840 14. Hu, Z., Yuan, J., Long, M., Jiang, J., Zhang, Y., Zhang, T., Xu, M., Fan, Y., Tanyi, J.L.,
1841 Montone, K.T., et al. (2021). The Cancer Surfaceome Atlas integrates genomic, functional
1842 and drug response data to identify actionable targets. Nat. Cancer *2*, 1406–1422.
1843 https://doi.org/10.1038/s43018-021-00282-w.
1844 15. Shaw, T.I., Wagner, J., Tian, L., Wickman, E., Poudel, S., Wang, J., Paul, R., Koo, S.C.,
1845 Lu, M., Sheppard, H., et al. (2024). Discovery of immunotherapy targets for pediatric
1846 solid and brain tumors by exon-level expression. Nat. Commun. *15*, 3732.
1847 https://doi.org/10.1038/s41467-024-47649-y.
1848 16. Perna, F., Berman, S.H., Soni, R.K., Mansilla-Soto, J., Eyquem, J., Hamieh, M.,
1849 Hendrickson, R.C., Brennan, C.W., and Sadelain, M. (2017). Integrating Proteomics and

1850 Transcriptomics for Systematic Combinatorial Chimeric Antigen Receptor Therapy of
1851 AML. Cancer Cell *32*, 506-519.e5. https://doi.org/10.1016/j.ccell.2017.09.004.
1852 17. Subramanian, A., Nemat-Gorgani, N., Ellis-Caleo, T.J., Van IJzendoorn, D.G.P., Sears,
1853 T.J., Somani, A., Luca, B.A., Zhou, M.Y., Bradic, M., Torres, I.A., et al. (2024). Sarcoma
1854 microenvironment cell states and ecosystems are associated with prognosis and predict
1855 response to immunotherapy. Nat. Cancer *5*, 642–658. https://doi.org/10.1038/s43018-
1856 024-00743-y.
1857 18. Gottschlich, A., Thomas, M., Grünmeier, R., Lesch, S., Rohrbacher, L., Igl, V.,
1858 Briukhovetska, D., Benmebarek, M.-R., Vick, B., Dede, S., et al. (2023). Single-cell
1859 transcriptomic atlas-guided development of CAR-T cells for the treatment of acute
1860 myeloid leukemia. Nat. Biotechnol. *41*, 1618–1632. https://doi.org/10.1038/s41587-023-
1861 01684-0.
1862 19. Lochner, J., Menge, F., Vassos, N., Hohenberger, P., and Kasper, B. (2020). Prognosis
1863 of Patients with Metastatic Soft Tissue Sarcoma: Advances in Recent Years. Oncol. Res.
1864 Treat. *43*, 613–619. https://doi.org/10.1159/000509519.
1865 20. Judson, I., Verweij, J., Gelderblom, H., Hartmann, J.T., Schöffski, P., Blay, J.-Y., Kerst,
1866 J.M., Sufliarsky, J., Whelan, J., Hohenberger, P., et al. (2014). Doxorubicin alone versus
1867 intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic
1868 soft-tissue sarcoma: a randomised controlled phase 3 trial. Lancet Oncol. *15*, 415–423.
1869 https://doi.org/10.1016/S1470-2045(14)70063-4.
1870 21. Tap, W.D., Papai, Z., Van Tine, B.A., Attia, S., Ganjoo, K.N., Jones, R.L., Schuetze, S.,
1871 Reed, D., Chawla, S.P., Riedel, R.F., et al. (2017). Doxorubicin plus evofosfamide versus

1872 doxorubicin alone in locally advanced, unresectable or metastatic soft-tissue sarcoma
1873 (TH CR-406/SARC021): an international, multicentre, open-label, randomised phase 3
1874 trial. Lancet Oncol. *18*, 1089–1103. https://doi.org/10.1016/S1470-2045(17)30381-9.
1875 22. Tawbi, H.A., Burgess, M., Bolejack, V., Van Tine, B.A., Schuetze, S.M., Hu, J.,
1876 D'Angelo, S., Attia, S., Riedel, R.F., Priebat, D.A., et al. (2017). Pembrolizumab in
1877 advanced soft-tissue sarcoma and bone sarcoma (SARC028): a multicentre, two-cohort,
1878 single-arm, open-label, phase 2 trial. Lancet Oncol. *18*, 1493–1501.
1879 https://doi.org/10.1016/S1470-2045(17)30624-1.
1880 23. Burgess, M.A., Bolejack, V., Schuetze, S., Van Tine, B.A., Attia, S., Riedel, R.F., Hu,
1881 J.S., Davis, L.E., Okuno, S.H., Priebat, D.A., et al. (2019). Clinical activity of
1882 pembrolizumab (P) in undifferentiated pleomorphic sarcoma (UPS) and
1883 dedifferentiated/pleomorphic liposarcoma (LPS): Final results of SARC028 expansion
1884 cohorts. J. Clin. Oncol. *37*, 11015–11015.
1885 https://doi.org/10.1200/JCO.2019.37.15_suppl.11015.
1886 24. Somaiah, N., Conley, A.P., Parra, E.R., Lin, H., Amini, B., Solis Soto, L., Salazar, R.,
1887 Barreto, C., Chen, H., Gite, S., et al. (2022). Durvalumab plus tremelimumab in
1888 advanced or metastatic soft tissue and bone sarcomas: a single-centre phase 2 trial.
1889 Lancet Oncol. *23*, 1156–1166. https://doi.org/10.1016/S1470-2045(22)00392-8.
1890 25. Keung, E.Z., Burgess, M., Salazar, R., Parra, E.R., Rodrigues-Canales, J., Bolejack, V.,
1891 Van Tine, B.A., Schuetze, S.M., Attia, S., Riedel, R.F., et al. (2020). Correlative Analyses of
1892 the SARC028 Trial Reveal an Association Between Sarcoma-Associated Immune

1893 Infiltrate and Response to Pembrolizumab. Clin. Cancer Res. Off. J. Am. Assoc. Cancer
1894 Res. *26*, 1258–1266. https://doi.org/10.1158/1078-0432.CCR-19-1824.
1895 26. Petitprez, F., De Reyniès, A., Keung, E.Z., Chen, T.W.-W., Sun, C.-M., Calderaro, J.,
1896 Jeng, Y.-M., Hsiao, L.-P., Lacroix, L., Bougoüin, A., et al. (2020). B cells are associated
1897 with survival and immunotherapy response in sarcoma. Nature *577*, 556–560.
1898 https://doi.org/10.1038/s41586-019-1906-8.
1899 27. D'Angelo, S.P., Araujo, D.M., Abdul Razak, A.R., Agulnik, M., Attia, S., Blay, J.-Y.,
1900 Carrasco Garcia, I., Charlson, J.A., Choy, E., Demetri, G.D., et al. (2024). Afamitresgene
1901 autoleucel for advanced synovial sarcoma and myxoid round cell liposarcoma
1902 (SPEARHEAD-1): an international, open-label, phase 2 trial. The Lancet *403*, 1460–1471.
1903 https://doi.org/10.1016/S0140-6736(24)00319-2.
1904 28. Hong, D.S., Van Tine, B.A., Biswas, S., McAlpine, C., Johnson, M.L., Olszanski, A.J.,
1905 Clarke, J.M., Araujo, D., Blumenschein, G.R., Kebriaei, P., et al. (2023). Autologous T cell
1906 therapy for MAGE-A4+ solid cancers in HLA-A*02+ patients: a phase 1 trial. Nat. Med. *29*,
1907 104–114. https://doi.org/10.1038/s41591-022-02128-z.
1908 29. D'Angelo, S.P., Melchiori, L., Merchant, M.S., Bernstein, D., Glod, J., Kaplan, R.,
1909 Grupp, S., Tap, W.D., Chagin, K., Binder, G.K., et al. (2018). Antitumor Activity Associated
1910 with Prolonged Persistence of Adoptively Transferred NY-ESO-1 c259T Cells in Synovial
1911 Sarcoma. Cancer Discov. *8*, 944–957. https://doi.org/10.1158/2159-8290.CD-17-1417.
1912 30. D'Angelo, S.P., Druta, M., Van Tine, B.A., Liebner, D.A., Schuetze, S., Hasan, A.N.,
1913 Holmes, A.P., Huff, A., Kapoor, G.S., Zajic, S., et al. (2021). Safety and efficacy of
1914 letetresgene autoleucel (lete-cel; GSK3377794) in advanced myxoid/round cell

liposarcoma (MRCLS) following high lymphodepletion (Cohort 2): Interim analysis. J. Clin. Oncol. *39*, 11521–11521. https://doi.org/10.1200/JCO.2021.39.15_suppl.11521.

31. Robbins, P.F., Morgan, R.A., Feldman, S.A., Yang, J.C., Sherry, R.M., Dudley, M.E., Wunderlich, J.R., Nahvi, A.V., Helman, L.J., Mackall, C.L., et al. (2011). Tumor regression in patients with metastatic synovial cell sarcoma and melanoma using genetically engineered lymphocytes reactive with NY-ESO-1. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *29*, 917–924. https://doi.org/10.1200/JCO.2010.32.2537.

32. Smithy, J.W., Blouin, A., Diamond, L.C., and Postow, M. (2022). Ensuring equity in the era of HLA-restricted cancer therapeutics. J. Immunother. Cancer *10*, e005600. https://doi.org/10.1136/jitc-2022-005600.

33. Jerby-Arnon, L., Neftel, C., Shore, M.E., Weisman, H.R., Mathewson, N.D., McBride, M.J., Haas, B., Izar, B., Volorio, A., Boulay, G., et al. (2021). Opposing immune and genetic mechanisms shape oncogenic programs in synovial sarcoma. Nat. Med. *27*, 289–300. https://doi.org/10.1038/s41591-020-01212-6.

34. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401. https://doi.org/10.1126/science.1254257.

35. Timothy Tickle, Itay Tirosh, Christophe Georgescu, Maxwell Brown, Brian Haas inferCNV of the Trinity CTAT Project.. Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. (Bioconductor).

https://doi.org/10.18129/B9.BIOC.INFERCNV

https://doi.org/10.18129/B9.BIOC.INFERCNV.

36. Abeshouse, A., Adebamowo, C., Adebamowo, S.N., Akbani, R., Akeredolu, T., Ally, A., Anderson, M.L., Anur, P., Appelbaum, E.L., Armenia, J., et al. (2017). Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. Cell *171*, 950-965.e28. https://doi.org/10.1016/j.cell.2017.10.014.

37. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

38. Tiberi, S., Crowell, H.L., Samartsidis, P., Weber, L.M., and Robinson, M.D. distinct: a novel approach to differential distribution analyses.

39. Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.

40. Pontén, F., Gry, M., Fagerberg, L., Lundberg, E., Asplund, A., Berglund, L., Oksvold, P., Björling, E., Hober, S., Kampf, C., et al. (2009). A global view of protein expression in human cells, tissues, and organs. Mol. Syst. Biol. *5*, 337. https://doi.org/10.1038/msb.2009.93.

41. The Tabula Sapiens Consortium*, Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula

1957 Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*,
1958 eabl4896. https://doi.org/10.1126/science.abl4896.
1959 42. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E.,
1960 Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell
1961 Atlas. eLife *6*, e27041. https://doi.org/10.7554/eLife.27041.
1962 43. Bughda, R., Dimou, P., D'Souza, R.R., and Klampatsa, A. (2021). Fibroblast
1963 Activation Protein (FAP)-Targeted CAR-T Cells: Launching an Attack on Tumor Stroma.
1964 ImmunoTargets Ther. *Volume 10*, 313–323. https://doi.org/10.2147/ITT.S291767.
1965 44. June, C.H., and Sadelain, M. (2018). Chimeric Antigen Receptor Therapy. N. Engl. J.
1966 Med. *379*, 64–73. https://doi.org/10.1056/NEJMra1706169.
1967 45. MacKay, M., Afshinnekoo, E., Rub, J., Hassan, C., Khunte, M., Baskaran, N., Owens,
1968 B., Liu, L., Roboz, G.J., Guzman, M.L., et al. (2020). The therapeutic landscape for cells
1969 engineered with chimeric antigen receptors. Nat. Biotechnol. *38*, 233–244.
1970 https://doi.org/10.1038/s41587-019-0329-2.
1971 46. DuBois, S.G., Krailo, M.D., Glade-Bender, J., Buxton, A., Laack, N., Randall, R.L.,
1972 Chen, H.X., Seibel, N.L., Boron, M., Terezakis, S., et al. (2023). Randomized Phase III Trial
1973 of Ganitumab With Interval-Compressed Chemotherapy for Patients With Newly
1974 Diagnosed Metastatic Ewing Sarcoma: A Report From the Children's Oncology Group. J.
1975 Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. *41*, 2098–2107.
1976 https://doi.org/10.1200/JCO.22.01815.
1977 47. Gounder, M.M., Rosenbaum, E., Wu, N., Dickson, M.A., Sheikh, T.N., D'Angelo, S.P.,
1978 Chi, P., Keohan, M.L., Erinjeri, J.P., Antonescu, C.R., et al. (2022). A Phase Ib/II

Randomized Study of RO4929097, a Gamma-Secretase or Notch Inhibitor with or without Vismodegib, a Hedgehog Inhibitor, in Advanced Sarcoma. Clin. Cancer Res. *28*, 1586–1594. https://doi.org/10.1158/1078-0432.CCR-21-3874.

48. Luke, J.J., Powderly, J.D., Merchan, J.R., Barve, M.A., Hotson, A.N., Mobasher, M., Kwei, L., Luciano, G., Buggy, J.J., Piccione, E., et al. (2019). Immunobiology, preliminary safety, and efficacy of CPI-006, an anti-CD73 antibody with immune modulating activity, in a phase 1 trial in advanced cancers. J. Clin. Oncol. *37*, 2505–2505. https://doi.org/10.1200/JCO.2019.37.15_suppl.2505.

49. Edwards, D., Handsley, M., and Pennington, C. (2008). The ADAM metalloproteinases. Mol. Aspects Med. *29*, 258–289. https://doi.org/10.1016/j.mam.2008.08.001.

50. Li, N., Spetz, M.R., and Ho, M. (2020). The Role of Glypicans in Cancer Progression and Therapy. J. Histochem. Cytochem. *68*, 841–862. https://doi.org/10.1369/0022155420933709.

51. Qin, S., Xu, L., Yi, M., Yu, S., Wu, K., and Luo, S. (2019). Novel immune checkpoint targets: moving beyond PD-1 and CTLA-4. Mol. Cancer *18*, 155. https://doi.org/10.1186/s12943-019-1091-2.

52. Du, Z., and Lovly, C.M. (2018). Mechanisms of receptor tyrosine kinase activation in cancer. Mol. Cancer *17*, 58. https://doi.org/10.1186/s12943-018-0782-4.

53. Zhou, B., Lin, W., Long, Y., Yang, Y., Zhang, H., Wu, K., and Chu, Q. (2022). Notch signaling pathway: architecture, disease, and therapeutics. Signal Transduct. Target. Ther. *7*, 95. https://doi.org/10.1038/s41392-022-00934-y.

54. Dorsam, R.T., and Gutkind, J.S. (2007). G-protein-coupled receptors and cancer. Nat. Rev. Cancer *7*, 79–94. https://doi.org/10.1038/nrc2069.

55. Neufeld, G., Mumblat, Y., Smolkin, T., Toledano, S., Nir-Zvi, I., Ziv, K., and Kessler, O. (2016). The role of the semaphorins in cancer. Cell Adhes. Migr. *10*, 652–674. https://doi.org/10.1080/19336918.2016.1197478.

56. Yan, H., Kamiya, T., Suabjakyong, P., and Tsuji, N.M. (2015). Targeting C-Type Lectin Receptors for Cancer Immunity. Front. Immunol. *6*. https://doi.org/10.3389/fimmu.2015.00408.

57. Pacini, C., Dempster, J.M., Boyle, I., Gonçalves, E., Najgebauer, H., Karakoc, E., Van Der Meer, D., Barthorpe, A., Lightfoot, H., Jaaks, P., et al. (2021). Integrated cross-study datasets of genetic dependencies in cancer. Nat. Commun. *12*, 1661. https://doi.org/10.1038/s41467-021-21898-7.

58. Nacev, B.A., Sanchez-Vega, F., Smith, S.A., Antonescu, C.R., Rosenbaum, E., Shi, H., Tang, C., Socci, N.D., Rana, S., Gularte-Mérida, R., et al. (2022). Clinical sequencing of soft tissue and bone sarcomas delineates diverse genomic landscapes and potential therapeutic targets. Nat. Commun. *13*, 3405. https://doi.org/10.1038/s41467-022-30453-x.

59. Woo, H.G., Park, E.S., Lee, J.-S., Lee, Y.-H., Ishikawa, T., Kim, Y.J., and Thorgeirsson, S.S. (2009). Identification of Potential Driver Genes in Human Liver Carcinoma by Genomewide Screening. Cancer Res. *69*, 4059–4066. https://doi.org/10.1158/0008-5472.CAN-09-0164.

2022 60. Pollack, J.R., Sørlie, T., Perou, C.M., Rees, C.A., Jeffrey, S.S., Lonning, P.E., Tibshirani,
2023 R., Botstein, D., Børresen-Dale, A.-L., and Brown, P.O. (2002). Microarray analysis reveals
2024 a major direct role of DNA copy number alteration in the transcriptional program of
2025 human breast tumors. Proc. Natl. Acad. Sci. U. S. A. *99*, 12963–12968.
2026 https://doi.org/10.1073/pnas.162471999.
2027 61. Wang, J., Zheng, J., Lee, E.E., Aguilar, B., Phan, J., Abdilleh, K., Taylor, R.C.,
2028 Longabaugh, W., Johansson, B., Mertens, F., et al. (2023). A cloud-based resource for
2029 genome coordinate-based exploration and large-scale analysis of chromosome
2030 aberrations and gene fusions in cancer. Genes. Chromosomes Cancer *62*, 441–448.
2031 https://doi.org/10.1002/gcc.23128.
2032 62. Kloss, C.C., Condomines, M., Cartellieri, M., Bachmann, M., and Sadelain, M.
2033 (2013). Combinatorial antigen recognition with balanced signaling promotes selective
2034 tumor eradication by engineered T cells. Nat. Biotechnol. *31*, 71–75.
2035 https://doi.org/10.1038/nbt.2459.
2036 63. Roybal, K.T., Williams, J.Z., Morsut, L., Rupp, L.J., Kolinko, I., Choe, J.H., Walker, W.J.,
2037 McNally, K.A., and Lim, W.A. (2016). Engineering T Cells with Customized Therapeutic
2038 Response Programs Using Synthetic Notch Receptors. Cell *167*, 419-432.e16.
2039 https://doi.org/10.1016/j.cell.2016.09.011.
2040 64. Fedorov, V.D., Themeli, M., and Sadelain, M. (2013). PD-1- and CTLA-4-based
2041 inhibitory chimeric antigen receptors (iCARs) divert off-target immunotherapy
2042 responses. Sci. Transl. Med. *5*, 215ra172.
2043 https://doi.org/10.1126/scitranslmed.3006597.

2044 65. Su, C., Xu, Z., Shan, X., Cai, B., Zhao, H., and Zhang, J. (2023). Cell-type-specific co-
2045 expression inference from single cell RNA-sequencing data. Nat. Commun. *14*, 4846.
2046 https://doi.org/10.1038/s41467-023-40503-7.
2047 66. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—
2048 Analysis Working Group, Statistical Methods groups—Analysis Working Group,
2049 Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH,
2050 NIH/NIDA, Biospecimen Collection Source Site—NDRI, et al. (2017). Genetic effects on
2051 gene expression across human tissues. Nature *550*, 204–213.
2052 https://doi.org/10.1038/nature24277.
2053 67. Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R.,
2054 Madugundu, A.K., Kelkar, D.S., Isserlin, R., Jain, S., et al. (2014). A draft map of the
2055 human proteome. Nature *509*, 575–581. https://doi.org/10.1038/nature13302.
2056 68. Xiao, W., Mohseny, A.B., Hogendoorn, P.C.W., and Cleton-Jansen, A.-M. (2013).
2057 Mesenchymal stem cell transformation and sarcoma genesis. Clin. Sarcoma Res. *3*, 10.
2058 https://doi.org/10.1186/2045-3329-3-10.
2059 69. Ferretti, E., and Hadjantonakis, A.-K. (2019). Mesoderm specification and
2060 diversification: from single cells to emergent tissues. Curr. Opin. Cell Biol. *61*, 110–116.
2061 https://doi.org/10.1016/j.ceb.2019.07.012.
2062 70. Shekhani, M.T., Jayanthy, A.-S., Maddodi, N., and Setaluri, V. (2013). Cancer stem
2063 cells and tumor transdifferentiation: implications for novel therapeutic strategies. Am. J.
2064 Stem Cells *2*, 52–61.

2065 71. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S.,
2066 Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer
2067 Dependency Map. Cell *170*, 564-576.e16. https://doi.org/10.1016/j.cell.2017.06.010.
2068 72. Yarmarkovich, M., Marshall, Q.F., Warrington, J.M., Premaratne, R., Farrel, A., Groff,
2069 D., Li, W., Di Marco, M., Runbeck, E., Truong, H., et al. (2023). Targeting of intracellular
2070 oncoproteins with peptide-centric CARs. Nature *623*, 820–827.
2071 https://doi.org/10.1038/s41586-023-06706-0.
2072 73. Gjerstorff, M.F., Andersen, M.H., and Ditzel, H.J. (2015). Oncogenic cancer/testis
2073 antigens: prime candidates for immunotherapy. Oncotarget *6*, 15772–15787.
2074 https://doi.org/10.18632/oncotarget.4694.
2075 74. Sarkizova, S., Klaeger, S., Le, P.M., Li, L.W., Oliveira, G., Keshishian, H., Hartigan,
2076 C.R., Zhang, W., Braun, D.A., Ligon, K.L., et al. (2020). A large peptidome dataset
2077 improves HLA class I epitope prediction across most of the human population. Nat.
2078 Biotechnol. *38*, 199–209. https://doi.org/10.1038/s41587-019-0322-9.
2079 75. Li, G., Iyer, B., Prasath, V.B.S., Ni, Y., and Salomonis, N. (2021). DeepImmuno: deep
2080 learning-empowered prediction and generation of immunogenic peptides for T-cell
2081 immunity. Brief. Bioinform. *22*, bbab160. https://doi.org/10.1093/bib/bbab160.
2082 76. Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S.,
2083 Røder, G., Peters, B., Sette, A., Lund, O., et al. (2007). NetMHCpan, a method for
2084 quantitative predictions of peptide binding to any HLA-A and -B locus protein of known
2085 sequence. PloS One *2*, e796. https://doi.org/10.1371/journal.pone.0000796.

2086 77. Gonzalez-Galarza, F.F., McCabe, A., Santos, E.J.M.D., Jones, J., Takeshita, L., Ortega-
2087 Rivera, N.D., Cid-Pavon, G.M.D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., et al.
2088 (2019). Allele frequency net database (AFND) 2020 update: gold-standard data
2089 classification, open access genotype data and new query tools. Nucleic Acids Res.,
2090 gkz1029. https://doi.org/10.1093/nar/gkz1029.
2091 78. Sewell, A.K. (2012). Why must T cells be cross-reactive? Nat. Rev. Immunol. *12*,
2092 669–677. https://doi.org/10.1038/nri3279.
2093 79. Kawashima, S., Pokarowski, P., Pokarowska, M., Kolinski, A., Katayama, T., and
2094 Kanehisa, M. (2007). AAindex: amino acid index database, progress report 2008. Nucleic
2095 Acids Res. *36*, D202–D205. https://doi.org/10.1093/nar/gkm998.
2096 80. Kim, Y., Sidney, J., Pinilla, C., Sette, A., and Peters, B. (2009). Derivation of an amino
2097 acid similarity matrix for peptide:MHC binding and its application as a Bayesian prior.
2098 BMC Bioinformatics *10*, 394. https://doi.org/10.1186/1471-2105-10-394.
2099 81. Antunes, D.A., Rigo, M.M., Freitas, M.V., Mendes, M.F.A., Sinigaglia, M., Lizée, G.,
2100 Kavraki, L.E., Selin, L.K., Cornberg, M., and Vieira, G.F. (2017). Interpreting T-Cell Cross-
2101 reactivity through Structure: Implications for TCR-Based Cancer Immunotherapy. Front.
2102 Immunol. *8*, 1210. https://doi.org/10.3389/fimmu.2017.01210.
2103 82. Bentzen, A.K., Such, L., Jensen, K.K., Marquard, A.M., Jessen, L.E., Miller, N.J.,
2104 Church, C.D., Lyngaa, R., Koelle, D.M., Becker, J.C., et al. (2018). T cell receptor
2105 fingerprinting enables in-depth characterization of the interactions governing recognition
2106 of peptide-MHC complexes. Nat. Biotechnol. https://doi.org/10.1038/nbt.4303.

2107 83. Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J.,
2108 Ozkan, E., Davis, M.M., Wucherpfennig, K.W., and Garcia, K.C. (2014). Deconstructing
2109 the peptide-MHC specificity of T cell recognition. Cell *157*, 1073–1087.
2110 https://doi.org/10.1016/j.cell.2014.03.047.
2111 84. Singh, N.K., Riley, T.P., Baker, S.C.B., Borrman, T., Weng, Z., and Baker, B.M. (2017).
2112 Emerging Concepts in TCR Specificity: Rationalizing and (Maybe) Predicting Outcomes.
2113 J. Immunol. *199*, 2203–2213. https://doi.org/10.4049/jimmunol.1700744.
2114 85. Tagliamonte, M., Cavalluzzo, B., Mauriello, A., Ragone, C., Buonaguro, F.M.,
2115 Tornesello, M.L., and Buonaguro, L. (2023). Molecular mimicry and cancer vaccine
2116 development. Mol. Cancer *22*, 75. https://doi.org/10.1186/s12943-023-01776-0.
2117 86. Linette, G.P., Stadtmauer, E.A., Maus, M.V., Rapoport, A.P., Levine, B.L., Emery, L.,
2118 Litzky, L., Bagg, A., Carreno, B.M., Cimino, P.J., et al. (2013). Cardiovascular toxicity and
2119 titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma. Blood *122*,
2120 863–871. https://doi.org/10.1182/blood-2013-03-490565.
2121 87. Morgan, R.A., Chinnasamy, N., Abate-Daga, D., Gros, A., Robbins, P.F., Zheng, Z.,
2122 Dudley, M.E., Feldman, S.A., Yang, J.C., Sherry, R.M., et al. (2013). Cancer regression and
2123 neurological toxicity following anti-MAGE-A3 TCR gene therapy. J. Immunother.
2124 Hagerstown Md 1997 *36*, 133–151. https://doi.org/10.1097/CJI.0b013e3182829903.
2125 88. Xie, N., Shen, G., Gao, W., Huang, Z., Huang, C., and Fu, L. (2023). Neoantigens:
2126 promising targets for cancer therapy. Signal Transduct. Target. Ther. *8*, 9.
2127 https://doi.org/10.1038/s41392-022-01270-x.

2128 89. Jiang, W., and Boder, E.T. (2010). High-throughput engineering and analysis of
2129 peptide binding to class II MHC. Proc. Natl. Acad. Sci. U. S. A. *107*, 13258–13263.
2130 https://doi.org/10.1073/pnas.1006344107.
2131 90. Liu, R., Jiang, W., and Mellins, E.D. (2021). Yeast display of MHC-II enables rapid
2132 identification of peptide ligands from protein antigens (RIPPA). Cell. Mol. Immunol. *18*,
2133 1847–1860. https://doi.org/10.1038/s41423-021-00717-5.
2134 91. Tang, S., Wang, Y., Luo, R., Fang, R., Liu, Y., Xiang, H., Ran, P., Tong, Y., Sun, M., Tan,
2135 S., et al. (2024). Proteomic characterization identifies clinically relevant subgroups of
2136 soft tissue sarcoma. Nat. Commun. *15*, 1381. https://doi.org/10.1038/s41467-024-
2137 45306-y.
2138 92. Burns, J., Wilding, C.P., Krasny, L., Zhu, X., Chadha, M., Tam, Y.B., Ps, H.,
2139 Mahalingam, A.H., Lee, A.T.J., Arthur, A., et al. (2023). The proteomic landscape of soft
2140 tissue sarcomas. Nat. Commun. *14*, 3834. https://doi.org/10.1038/s41467-023-39486-2.
2141 93. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S.,
2142 Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line
2143 Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*,
2144 603–607. https://doi.org/10.1038/nature11003.
2145 94. Gonçalves, E., Poulos, R.C., Cai, Z., Barthorpe, S., Manda, S.S., Lucas, N., Beck, A.,
2146 Bucio-Noble, D., Dausmann, M., Hall, C., et al. (2022). Pan-cancer proteomic map of
2147 949 human cell lines. Cancer Cell *40*, 835-849.e8.
2148 https://doi.org/10.1016/j.ccell.2022.06.010.

95. Majzner, R.G., Theruvath, J.L., Nellan, A., Heitzeneder, S., Cui, Y., Mount, C.W., Rietberg, S.P., Linde, M.H., Xu, P., Rota, C., et al. (2019). CAR T Cells Targeting B7-H3, a Pan-Cancer Antigen, Demonstrate Potent Preclinical Activity Against Pediatric Solid Tumors and Brain Tumors. Clin. Cancer Res. *25*, 2560–2574. https://doi.org/10.1158/1078-0432.CCR-18-0432.

96. Yoshida, K., Poon, V., Dash, A., Kunder, R., Chinn, L., and Kågedal, M. (2024). Simulation-based evaluation of personalized dosing approaches for ANTI-FGFR / KLB bispecific antibody fazpilodemab. CPT Pharmacomet. Syst. Pharmacol. *13*, 544–550. https://doi.org/10.1002/psp4.13111.

97. Vahdat, L.T., Schmid, P., Forero-Torres, A., Blackwell, K., Telli, M.L., Melisko, M., Möbus, V., Cortes, J., Montero, A.J., Ma, C., et al. (2021). Glembatumumab vedotin for patients with metastatic, gpNMB overexpressing, triple-negative breast cancer (“METRIC”): a randomized multicenter study. Npj Breast Cancer *7*, 57. https://doi.org/10.1038/s41523-021-00244-6.

98. Meininger, V., Genge, A., Van Den Berg, L.H., Robberecht, W., Ludolph, A., Chio, A., Kim, S.H., Leigh, P.N., Kiernan, M.C., Shefner, J.M., et al. (2017). Safety and efficacy of ozanezumab in patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled, phase 2 trial. Lancet Neurol. *16*, 208–216. https://doi.org/10.1016/S1474-4422(16)30399-4.

99. Wang, X., Qi, L., Chen, M., Zhang, Y., Gao, X., and Cai, Y. (2024). Feasibility study of ADCs targeting TROP-2, HER2, and CD46 in Ductal Adenocarcinoma and Intraductal

2170 Carcinoma of the prostate. World J. Urol. *42*, 404. https://doi.org/10.1007/s00345-024-
2171 05109-8.
2172 100. Rafiq, S., Hackett, C.S., and Brentjens, R.J. (2020). Engineering strategies to
2173 overcome the current roadblocks in CAR T cell therapy. Nat. Rev. Clin. Oncol. *17*, 147–
2174 167. https://doi.org/10.1038/s41571-019-0297-y.
2175 101. Rodriguez-Garcia, A., Palazon, A., Noguera-Ortega, E., Powell, D.J., and Guedan, S.
2176 (2020). CAR-T Cells Hit the Tumor Microenvironment: Strategies to Overcome Tumor
2177 Escape. Front. Immunol. *11*, 1109. https://doi.org/10.3389/fimmu.2020.01109.
2178 102. Parker, K.R., Migliorini, D., Perkey, E., Yost, K.E., Bhaduri, A., Bagga, P., Haris, M.,
2179 Wilson, N.E., Liu, F., Gabunia, K., et al. (2020). Single-Cell Analyses Identify Brain Mural
2180 Cells Expressing CD19 as Potential Off-Tumor Targets for CAR-T Immunotherapies. Cell
2181 *183*, 126-142.e17. https://doi.org/10.1016/j.cell.2020.08.022.
2182 103. Jing, Y., Liu, Y., Li, Q., Ye, Y., Diao, L., Huang, Y., Zhou, Y., Green, M.R., Mills, G.B., and
2183 Han, L. (2021). Expression of chimeric antigen receptor therapy targets detected by
2184 single-cell sequencing of normal cells may contribute to off-tumor toxicity. Cancer Cell
2185 *39*, 1558–1559. https://doi.org/10.1016/j.ccell.2021.09.016.
2186 104. Lee, A.T.J., Thway, K., Huang, P.H., and Jones, R.L. (2018). Clinical and Molecular
2187 Spectrum of Liposarcoma. J. Clin. Oncol. *36*, 151–159.
2188 https://doi.org/10.1200/JCO.2017.74.9598.
2189 105. Cassinelli, G., Zaffaroni, N., and Lanzi, C. (2016). The heparanase/heparan sulfate
2190 proteoglycan axis: A potential new therapeutic target in sarcomas. Cancer Lett. *382*,
2191 245–254. https://doi.org/10.1016/j.canlet.2016.09.004.

106. Wu, Z., Zhang, X., An, Y., Ma, K., Xue, R., Ye, G., Du, J., Chen, Z., Zhu, Z., Shi, G., et al. (2023). CLMP is a tumor suppressor that determines all-trans retinoic acid response in colorectal cancer. Dev. Cell *58*, 2684-2699.e6. https://doi.org/10.1016/j.devcel.2023.10.006.

107. Limoni, G., and Niquille, M. (2021). Semaphorins and Plexins in central nervous system patterning: the key to it all? Curr. Opin. Neurobiol. *66*, 224–232. https://doi.org/10.1016/j.conb.2020.12.014.

108. Worzfeld, T., and Offermanns, S. (2014). Semaphorins and plexins as therapeutic targets. Nat. Rev. Drug Discov. *13*, 603–621. https://doi.org/10.1038/nrd4337.

109. Gounder, M., Ratan, R., Alcindor, T., Schöffski, P., van der Graaf, W.T., Wilky, B.A., Riedel, R.F., Lim, A., Smith, L.M., Moody, S., et al. (2023). Nirogacestat, a γ-Secretase Inhibitor for Desmoid Tumors. N. Engl. J. Med. *388*, 898–912. https://doi.org/10.1056/NEJMoa2210140.

110. Mir, O., Azaro, A., Merchan, J., Chugh, R., Trent, J., Rodon, J., Ohnmacht, U., Diener, J.T., Smith, C., Yuen, E., et al. (2018). Notch pathway inhibition with LY3039478 in soft tissue sarcoma and gastrointestinal stromal tumours. Eur. J. Cancer Oxf. Engl. 1990 *103*, 88–97. https://doi.org/10.1016/j.ejca.2018.08.012.

111. Hadjimichael, A.C., Pergaris, A., Kaspiris, A., Foukas, A.F., Kokkali, S., Tsourouflis, G., and Theocharis, S. (2022). The EPH/Ephrin System in Bone and Soft Tissue Sarcomas' Pathogenesis and Therapy: New Advancements and a Literature Review. Int. J. Mol. Sci. *23*, 5171. https://doi.org/10.3390/ijms23095171.

2213 112. Kania, A., and Klein, R. (2016). Mechanisms of ephrin–Eph signalling in
2214 development, physiology and disease. Nat. Rev. Mol. Cell Biol. *17*, 240–256.
2215 https://doi.org/10.1038/nrm.2015.16.
2216 113. Martinez-Font, E., Pérez-Capó, M., Vögler, O., Martín-Broto, J., Alemany, R., and
2217 Obrador-Hevia, A. (2021). WNT/β-Catenin Pathway in Soft Tissue Sarcomas: New
2218 Therapeutic Opportunities? Cancers *13*, 5521.
2219 https://doi.org/10.3390/cancers13215521.
2220 114. Rodriguez-Garcia, A., Lynn, R.C., Poussin, M., Eiva, M.A., Shaw, L.C., O'Connor, R.S.,
2221 Minutolo, N.G., Casado-Medrano, V., Lopez, G., Matsuyama, T., et al. (2021). CAR-T cell-
2222 mediated depletion of immunosuppressive tumor-associated macrophages promotes
2223 endogenous antitumor immunity and augments adoptive immunotherapy. Nat.
2224 Commun. *12*, 877. https://doi.org/10.1038/s41467-021-20893-2.
2225 115. Ash, S.L., Orha, R., Mole, H., Dinesh-Kumar, M., Lee, S.P., Turrell, F.K., and Isacke,
2226 C.M. (2024). Targeting the activated microenvironment with endosialin (CD248)-directed
2227 CAR-T cells ablates perivascular cells to impair tumor growth and metastasis. J.
2228 Immunother. Cancer *12*, e008608. https://doi.org/10.1136/jitc-2023-008608.
2229 116. Wei, R., Dean, D.C., Thanindratarn, P., Hornicek, F.J., Guo, W., and Duan, Z. (2020).
2230 Cancer testis antigens in sarcoma: Expression, function and immunotherapeutic
2231 application. Cancer Lett. *479*, 54–60. https://doi.org/10.1016/j.canlet.2019.10.024.
2232 117. Fonseca, A.F., and Antunes, D.A. (2023). CrossDome: an interactive R package to
2233 predict cross-reactivity risk using immunopeptidomics databases. Front. Immunol. *14*,
2234 1142573. https://doi.org/10.3389/fimmu.2023.1142573.

118. Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat. Commun. *12*, 5692. https://doi.org/10.1038/s41467-021-25960-2.

119. R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

120. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

121. McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. *8*, 329-337.e4. https://doi.org/10.1016/j.cels.2019.03.003.

122. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.

123. Samuel Marsh, Maëlle Salmon, and Paul Hoffman (2024). samuel-marsh/scCustomize: Version 2.1.2. Version v2.1.2 (Zenodo). https://doi.org/10.5281/ZENODO.5706430 https://doi.org/10.5281/ZENODO.5706430.

124. Alquicira-Hernandez, J., and Powell, J.E. (2021). Nebulosa recovers single-cell gene expression signals by kernel density estimation. Bioinforma. Oxf. Engl. *37*, 2485–2487. https://doi.org/10.1093/bioinformatics/btab003.

125. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172. https://doi.org/10.1038/s41590-018-0276-y.

126. The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

127. Stunnenberg, H.G., International Human Epigenome Consortium, and Hirst, M. (2016). The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell *167*, 1145–1149. https://doi.org/10.1016/j.cell.2016.11.007.

128. Baron, M., Veres, A., Wolock, S.L., Faust, A.L., Gaujoux, R., Vetere, A., Ryu, J.H., Wagner, B.K., Shen-Orr, S.S., Klein, A.M., et al. (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst. *3*, 346-360.e4. https://doi.org/10.1016/j.cels.2016.08.011.

129. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

130. Amezquita, R.A., Lun, A.T.L., Becht, E., Carey, V.J., Carpp, L.N., Geistlinger, L., Marini, F., Rue-Albrecht, K., Risso, D., Soneson, C., et al. (2020). Orchestrating single-cell analysis with Bioconductor. Nat. Methods *17*, 137–145. https://doi.org/10.1038/s41592-019-0654-x.

131. McCarthy, D.J., Campbell, K.R., Lun, A.T.L., and Wills, Q.F. (2017). Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics *33*, 1179–1186. https://doi.org/10.1093/bioinformatics/btw777.

132. Bausch-Fluck, D., Goldmann, U., Müller, S., Van Oostrum, M., Müller, M., Schubert, O.T., and Wollscheid, B. (2018). The in silico human surfaceome. Proc. Natl. Acad. Sci. *115*. https://doi.org/10.1073/pnas.1808790115.

133. Jiang, L., Wang, M., Lin, S., Jian, R., Li, X., Chan, J., Dong, G., Fang, H., Robinson, A.E., Snyder, M.P., et al. (2020). A Quantitative Proteome Map of the Human Body. Cell *183*, 269-283.e19. https://doi.org/10.1016/j.cell.2020.08.036.

134. Adam Mark, R.T. (2017). mygene. (Bioconductor). https://doi.org/10.18129/B9.BIOC.MYGENE https://doi.org/10.18129/B9.BIOC.MYGENE.

135. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov. Camb. Mass *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

136. Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol. Biosyst. *12*, 477–479. https://doi.org/10.1039/C5MB00663E.

137. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J. Integr. Biol. *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

138. Carlson M (2019). org.Hs.eg.db: Genome wide annotation for Human. R package version 3.8.2.

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

140. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. *2*, 401–404. https://doi.org/10.1158/2159-8290.CD-12-0095.

141. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. *6*, pl1. https://doi.org/10.1126/scisignal.2004088.

142. Gao, Q., Liang, W.-W., Foltz, S.M., Mutharasu, G., Jayasinghe, R.G., Cao, S., Liao, W.-W., Reynolds, S.M., Wyczalkowski, M.A., Yao, L., et al. (2018). Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. Cell Rep. *23*, 227-238.e3. https://doi.org/10.1016/j.celrep.2018.03.050.

143. Mertens F, Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (2024). Mitelman F, Johansson B and Mertens F (Eds.), https://mitelmandatabase.isb-cgc.org.

144. Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single–cell type transcriptomics map of human tissues. Sci. Adv. *7*, eabh2169. https://doi.org/10.1126/sciadv.abh2169.

145. Wells, D.A., and McAuley, M. (2023). HLAfreq: Download and combine HLA allele frequency data. Preprint, https://doi.org/10.1101/2023.09.15.557761 https://doi.org/10.1101/2023.09.15.557761.

146. The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bye-A-Jee, H., et al. (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

147. Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., and Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. *47*, D339–D343. https://doi.org/10.1093/nar/gky1006.

148. Huang, X., Gan, Z., Cui, H., Lan, T., Liu, Y., Caron, E., and Shao, W. (2024). The SysteMHC Atlas v2.0, an updated resource for mass spectrometry-based immunopeptidomics. Nucleic Acids Res. *52*, D1062–D1071. https://doi.org/10.1093/nar/gkad1068.

149. Marcu, A., Bichmann, L., Kuchenbecker, L., Kowalewski, D.J., Freudenmann, L.K., Backert, L., Mühlenbruch, L., Szolek, A., Lübke, M., Wagner, P., et al. (2021). HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. J. Immunother. Cancer *9*, e002071. https://doi.org/10.1136/jitc-2020-002071.

150. O'Donnell, T.J., Rubinsteyn, A., and Laserson, U. (2020). MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing. Cell Syst. *11*, 42-48.e7. https://doi.org/10.1016/j.cels.2020.06.010.

151. McFerrin L (2013). HDMD: Statistical Analysis Tools for High Dimension Molecular Data (HDMD).

152. Amerifar S (2021). ftrCOOL: Feature Extraction from Biological Sequences. Version version 2.0.0.

153. Liao, Y., Savage, S.R., Dou, Y., Shi, Z., Yi, X., Jiang, W., Lei, J.T., and Zhang, B. (2023). A proteogenomics data-driven knowledge base of human cancer. Cell Syst. *14*, 777-787.e5. https://doi.org/10.1016/j.cels.2023.07.007.

154. Thangudu, R.R., Rudnick, P.A., Holck, M., Singhal, D., MacCoss, M.J., Edwards, N.J., Ketchum, K.A., Kinsinger, C.R., Kim, E., and Basu, A. (2020). Abstract LB-242: Proteomic Data Commons: A resource for proteogenomic analysis. Cancer Res. *80*, LB-242-LB-242. https://doi.org/10.1158/1538-7445.AM2020-LB-242.

155. Sinha, A., Huang, V., Livingstone, J., Wang, J., Fox, N.S., Kurganovs, N., Ignatchenko, V., Fritsch, K., Donmez, N., Heisler, L.E., et al. (2019). The Proteogenomic Landscape of Curable Prostate Cancer. Cancer Cell *35*, 414-427.e6. https://doi.org/10.1016/j.ccell.2019.02.005.

156. Liu, W., Xie, L., He, Y.-H., Wu, Z.-Y., Liu, L.-X., Bai, X.-F., Deng, D.-X., Xu, X.-E., Liao, L.-D., Lin, W., et al. (2021). Large-scale and high-resolution mass spectrometry-based proteomics profiling defines molecular subtypes of esophageal cancer for therapeutic targeting. Nat. Commun. *12*, 4961. https://doi.org/10.1038/s41467-021-25202-5.

157. Beck, L., Harel, M., Yu, S., Markovits, E., Boursi, B., Markel, G., and Geiger, T. (2021). Clinical Proteomics of Metastatic Melanoma Reveals Profiles of Organ Specificity and

Treatment Resistance. Clin. Cancer Res. *27*, 2074–2086. https://doi.org/10.1158/1078-0432.CCR-20-3752.

158. Groeneveld, C.S., Sanchez-Quiles, V., Dufour, F., Shi, M., Dingli, F., Nicolle, R., Chapeaublanc, E., Poullet, P., Jeffery, D., Krucker, C., et al. (2024). Proteogenomic Characterization of Bladder Cancer Reveals Sensitivity to Apoptosis Induced by Tumor Necrosis Factor–related Apoptosis-inducing Ligand in FGFR3-mutated Tumors. Eur. Urol. *85*, 483–494. https://doi.org/10.1016/j.eururo.2023.05.037.

159. Boder, E.T., and Wittrup, K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. *15*, 553–557. https://doi.org/10.1038/nbt0697-553.

160. Raybould, M.I.J., Marks, C., Lewis, A.P., Shi, J., Bujotzek, A., Taddese, B., and Deane, C.M. (2020). Thera-SAbDab: the Therapeutic Structural Antibody Database. Nucleic Acids Res. *48*, D383–D388. https://doi.org/10.1093/nar/gkz827.

Figures

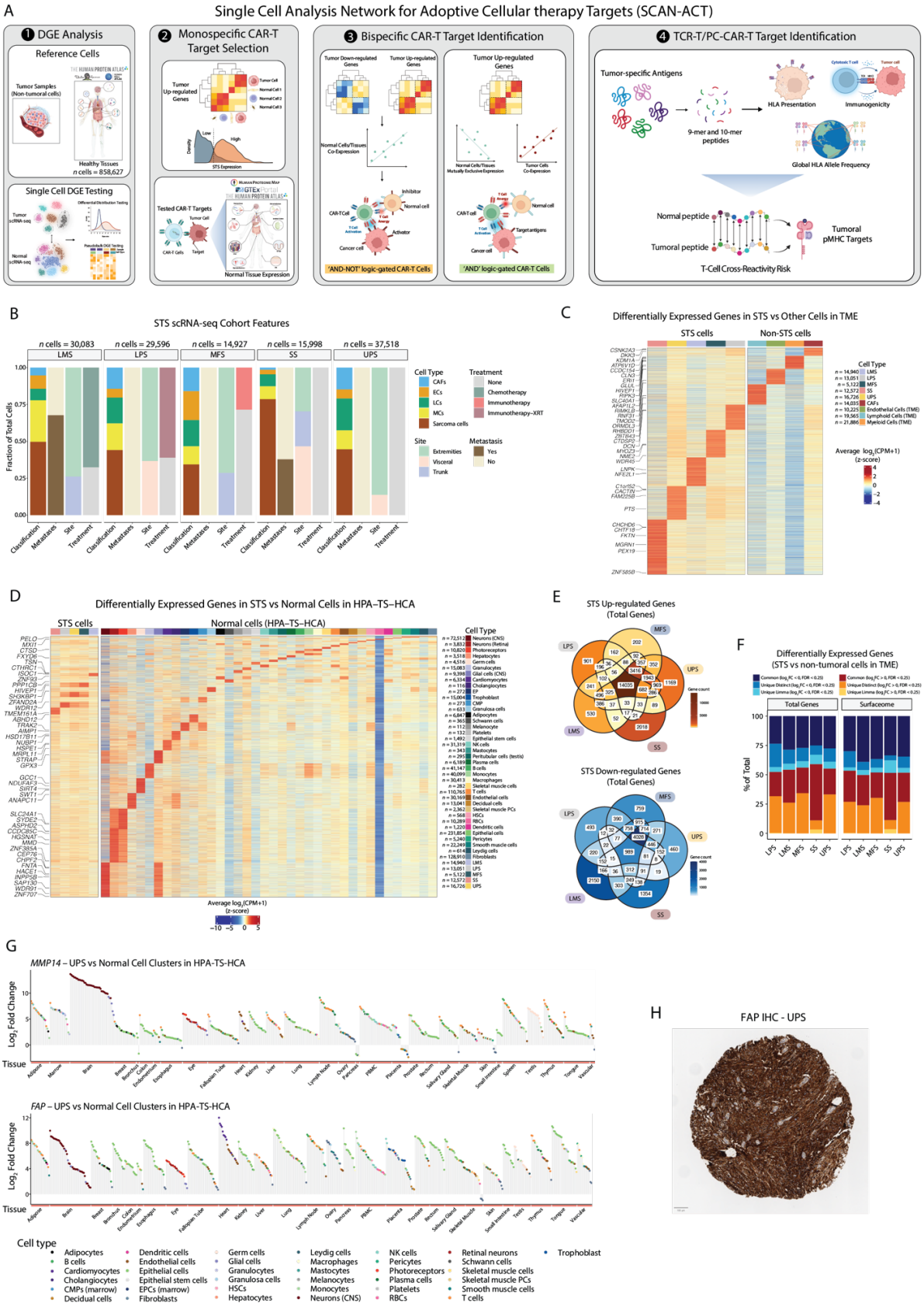
A
Single Cell Analysis Network for Adoptive Cellular therapy Targets (SCAN-ACT)
1 DGE Analysis
Reference Cells
Tumor Samples (Non-tumoral cells)
Healthy Tissues
n cells = 858,627
Single Cell DGE Testing
2 Monospecific CAR-T Target Selection
Tumor Up-regulated Genes
Tumor Cell
Normal Cell 1
Normal Cell 2
Normal Cell 3
Density
Low
High
STS Expression
Tested CAR-T Targets
Tumor Cell
CAR-T Cells
Target
Normal Tissue Expression
3 Bispecific CAR-T Target Identification
Tumor Down-regulated Genes
Tumor Up-regulated Genes
Normal Cells/Tissues Co-Expression
Inhibitor
Normal cell
CAR-T cell
Activator
Cancer cell
'AND-NOT' logic-gated CAR-T Cells
Tumor Up-regulated Genes
Normal Cells/Tissues Mutually Exclusive Expression
Tumor Cells Co-Expression
Target antigens
'AND' logic-gated CAR-T Cells
4 TCR-T/PC-CAR-T Target Identification
Tumor-specific Antigens
9-mer and 10-mer peptides
HLA Presentation
Immunogenicity
Global HLA Allele Frequency
Normal peptide
Tumoral peptide
Tumoral pMHC Targets
T-Cell Cross-Reactivity Risk
B
STS scRNA-seq Cohort Features
n cells = 30,083
n cells = 29,596
n cells = 14,927
n cells = 15,998
n cells = 37,518
LMS
LPS
MFS
SS
UPS
Fraction of Total Cells
Classification
Metastasis
Site
Treatment
Cell Type
CAFs
ECs
LCs
MCs
Sarcoma cells
Treatment
None
Chemotherapy
Immunotherapy
Immunotherapy–XRT
Site
Extremities
Visceral
Trunk
Metastasis
Yes
No
C
Differentially Expressed Genes in STS vs Other Cells in TME
STS cells
Non-STS cells
Cell Type
Average log2(CPM+1) (z-score)
D
Differentially Expressed Genes in STS vs Normal Cells in HPA–TS–HCA
STS cells
Normal cells (HPA–TS–HCA)
Cell Type
Average log2(CPM+1) (z-score)
E
STS Up-regulated Genes (Total Genes)
STS Down-regulated Genes (Total Genes)
LPS
MFS
UPS
LMS
SS
Gene count
F
Differentially Expressed Genes (STS vs non-tumoral cells in TME)
Total Genes
Surfaceome
% of Total
LPS
LMS
MFS
SS
UPS
G
MMP14 – UPS vs Normal Cell Clusters in HPA-TS-HCA
Log2 Fold Change
Tissue
FAP – UPS vs Normal Cell Clusters in HPA-TS-HCA
Cell type
Adipocytes
B cells
Cardiomyocytes
Cholangiocytes
CMPs (marrow)
Decidual cells
Dendritic cells
Endothelial cells
Epithelial cells
Epithelial stem cells
EPCs (marrow)
Fibroblasts
Germ cells
Glial cells
Granulocytes
Granulosa cells
HSCs
Hepatocytes
Leydig cells
Macrophages
Mastocytes
Melanocytes
Monocytes
Neurons (CNS)
NK cells
Pericytes
Photoreceptors
Plasma cells
Platelets
RBCs
Retinal neurons
Schwann cells
Skeletal muscle cells
Skeletal muscle PCs
Smooth muscle cells
T cells
Trophoblast
H
FAP IHC - UPS

Figure 1. **scRNA-seq for cellular therapy target identification**.

(A). Flow diagram showing the full SCAN-ACT pipeline. Stacked bar charts showing the fraction of cells identified as tumoral, endothelial, myeloid, lymphoid, or CAFs, sample site of origin, prior treatment, and metastatic or primary tumors. **(C, D)** Heatmaps showing average RNA expression in STS versus non-tumoral cells in the TME (**C**) and normal cells in the HPA, TS, and HCA (**D**). RNA expression ($\log_2$(CPM+1)) was averaged and median-centered across cell types, and z-scored across genes. The number of cells analyzed is shown in the legend. Only genes differentially expressed in each STS and measured in all samples are shown. **(E)** Venn diagrams showing the number of STS up-regulated and down-regulated genes shared between histotypes. **(F)** Stacked bar chart showing the fraction of DEGs in STS versus non-tumoral cells in the TME, identified with Limma-voom, distinct, or both methods, considering surfaceome and total genes separately. **(G)** Lollipop plots showing the $\log_2$FC in expression for *MMP14* and *FAP* between UPS cells and the normal cell clusters in the HPA, TS, and HCA using Limma-voom. Only statistically significant comparisons are shown (FDR $<$ 0.25). (**H**) FAP IHC expression on tissue microarrays of UPS samples. CAFs: cancer associated fibroblasts; ECs: endothelial cells; LCs: lymphoid cells; MCs: myeloid cells; Mφ: Macrophages.

2380

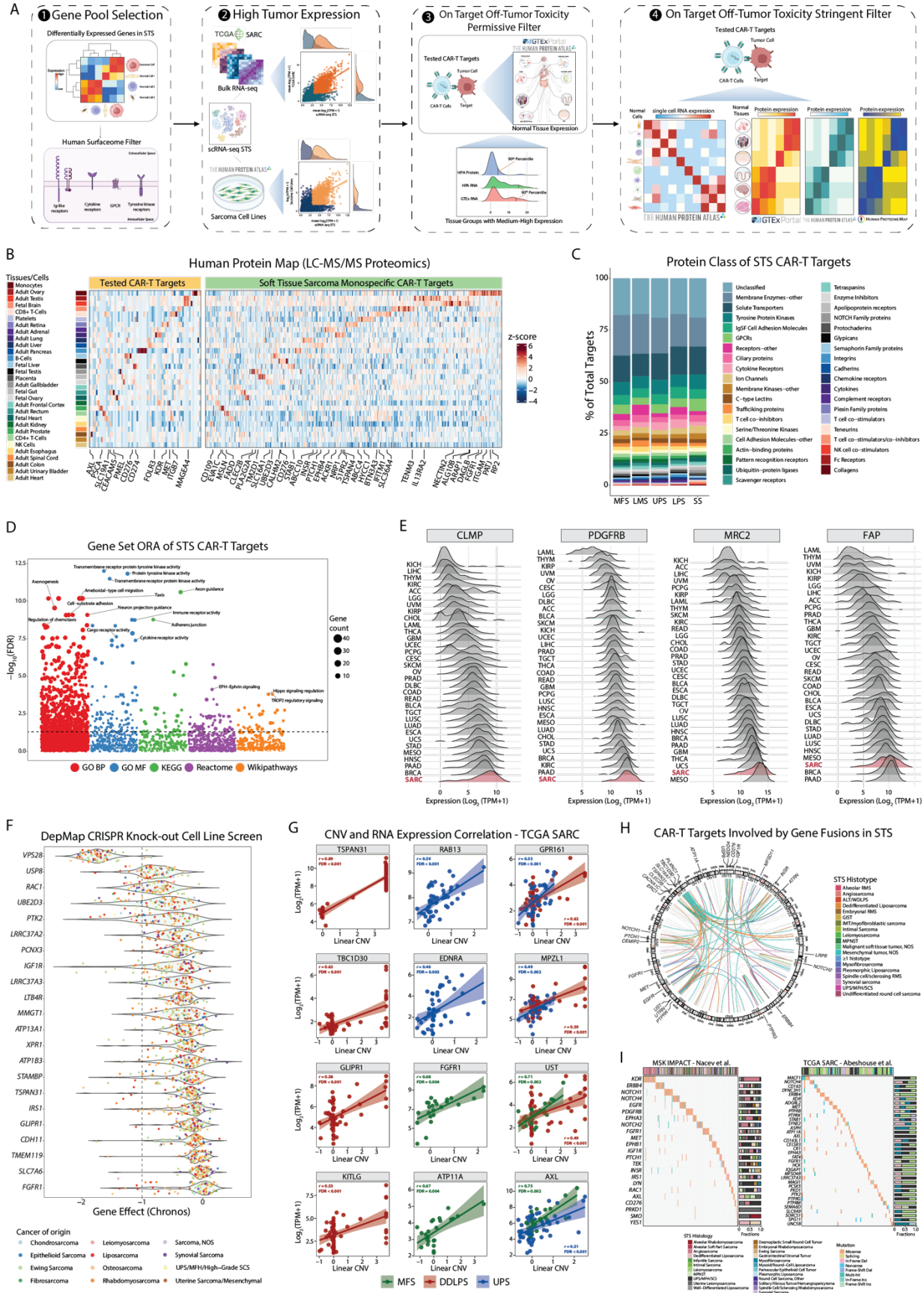
A
1 Gene Pool Selection
Differentially Expressed Genes in STS
Human Surfaceome Filter
2 High Tumor Expression
TCGA SARC
Bulk RNA-seq
scRNA-seq STS
Sarcoma Cell Lines
3 On Target Off-Tumor Toxicity Permissive Filter
Tested CAR-T Targets
Normal Tissue Expression
Tissue Groups with Medium-High Expression
4 On Target Off-Tumor Toxicity Stringent Filter
Tested CAR-T Targets
B
Human Protein Map (LC-MS/MS Proteomics)
Tissues/Cells
Tested CAR-T Targets
Soft Tissue Sarcoma Monospecific CAR-T Targets
z-score
C
Protein Class of STS CAR-T Targets
% of Total Targets
MFS LMS UPS LPS SS
D
Gene Set ORA of STS CAR-T Targets
Gene count
GO BP GO MF KEGG Reactome Wikipathways
E
CLMP
PDGFRB
MRC2
FAP
Expression (Log2 (TPM+1))
F
DepMap CRISPR Knock-out Cell Line Screen
Gene Effect (Chronos)
Cancer of origin
G
CNV and RNA Expression Correlation - TCGA SARC
TSPAN31
RAB13
GPR161
TBC1D30
EDNRA
MPZL1
GLIPR1
FGFR1
UST
KITLG
ATP11A
AXL
Linear CNV
MFS DDLPS UPS
H
CAR-T Targets Involved by Gene Fusions in STS
STS Histotype
I
MSK IMPACT - Nacev et al.
TCGA SARC - Abeshouse et al.

Figure 2. **Monospecific CAR-T cell targets in STS**.

(A) Diagram showing discovery pipeline for monospecific STS CAR-T targets. **(B)** Heatmap showing the protein expression in normal tissues and cells or tissue groups for CAR-T targets tested in humans (left panel) and for the STS targets (right panel) in the HPM dataset. **(C)** Bar chart showing the proportion of STS CAR-T targets for each protein class/family. **(D)** Scatter plot showing gene set ORA on the 449 CAR-T targets, using the gene ontology (GO) molecular function (MF) and biological process (BP), as well as the KEGG, Reactome, and WikiPathways terms. The x-axis shows the $-\log_{10}$(FDR) while the dots size indicates the number of targets per gene set. The horizontal line intersects at the significance threshold (FDR = 0.05). **(E)** Ridge plots showing RNA expression ($\log_2$(TPM+1)) in the Pan-Cancer TCGA Atlas for STS CAR-T targets. STS samples (SARC) are highlighted in bold red. **(F)** Violin plots showing the effect of CAR-T targets KO on sarcoma cell line growth from DepMap, quantified using the Chronos algorithm. Negative values indicate genes essential for cell growth. The vertical line indicates the median Chronos value for known essential genes (–1). Dots are colored based on the cell line cancer of origin. **(G)** Pearson correlation between linear CNV level and RNA expression in TCGA SARC for monospecific CAR-T targets. P-values were adjusted using the BH method setting significance to FDR < 0.01. **(H)** Circos plot showing gene fusions involving a CAR-T target in STS from the Mitelman database, MSK-IMPACT, and TCGA SARC. Colored links show fusions for the top 30 most recurrently fused genes, with link colors indicating the STS histotype. **(I)** Oncoplots showing recurring SNVs/Indels for monospecific CAR-T targets in MSK-IMPACT (*left*) and TCGA SARC (*right*). ALT: Atypical Lipomatous Tumor; GIST: Gastrointestinal Stromal Tumor; MPNST: Malignant Peripheral Nerve Sheath Tumor; IMT: Inflammatory Myofibroblastic Tumor; MFH: Malignant Fibrous Histiocytoma; SCS: High-grade Spindle Cell Sarcoma.

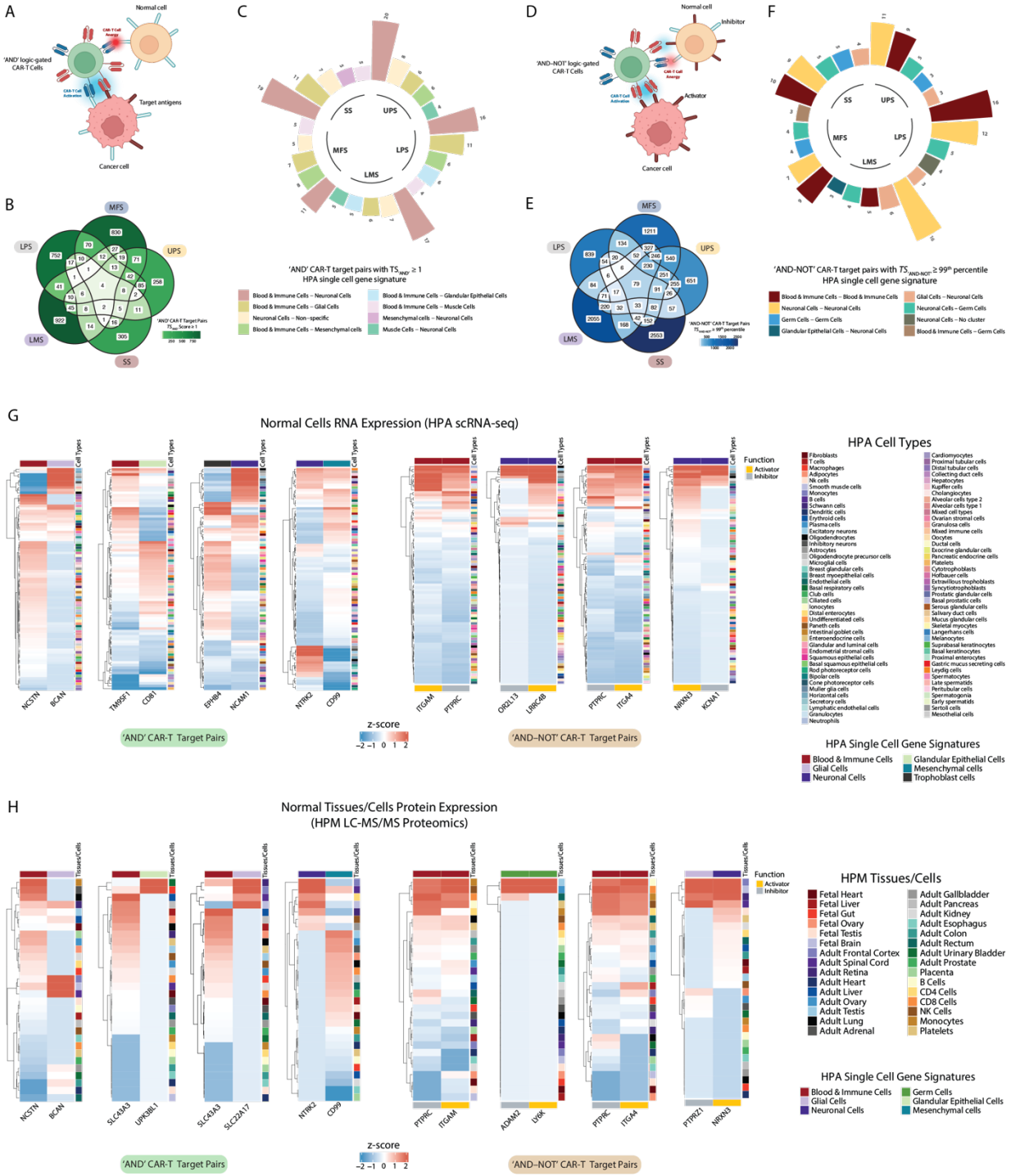
A
Normal cell
'AND' logic-gated CAR-T Cells
Target antigens
Cancer cell
B
MFS
UPS
LPS
LMS
SS
C
'AND' CAR-T target pairs with TS AND ≥ 1
HPA single cell gene signature
Blood & Immune Cells – Neuronal Cells
Blood & Immune Cells – Glial Cells
Neuronal Cells – Non-specific
Blood & Immune Cells – Mesenchymal cells
Blood & Immune Cells – Glandular Epithelial Cells
Blood & Immune Cells – Muscle Cells
Mesenchymal cells – Neuronal Cells
Muscle Cells – Neuronal Cells
D
Normal cell
Inhibitor
'AND–NOT' logic-gated CAR-T Cells
Activator
Cancer cell
E
F
'AND–NOT' CAR-T target pairs with TS AND–NOT ≥ 99th percentile
HPA single cell gene signature
Blood & Immune Cells – Blood & Immune Cells
Neuronal Cells – Neuronal Cells
Germ Cells – Germ Cells
Glandular Epithelial Cells – Neuronal Cells
Glial Cells – Neuronal Cells
Neuronal Cells – Germ Cells
Neuronal Cells – No cluster
Blood & Immune Cells – Germ Cells
G
Normal Cells RNA Expression (HPA scRNA-seq)
NCSTN
BCAN
TMPSF1
CD81
EPHB4
NCAM1
NTRK2
CD99
ITGAM
PTPRC
ORZL13
LRRC4B
PTPRC
ITGA4
NRXN3
KCNA1
'AND' CAR-T Target Pairs
z-score
'AND–NOT' CAR-T Target Pairs
Function
Activator
Inhibitor
HPA Cell Types
HPA Single Cell Gene Signatures
Blood & Immune Cells
Glial Cells
Neuronal Cells
Glandular Epithelial Cells
Mesenchymal cells
Trophoblast cells
H
Normal Tissues/Cells Protein Expression
(HPM LC-MS/MS Proteomics)
NCSTN
BCAN
SLC43A3
UPK3BL1
SLC43A3
SLC22A17
NTRK2
CD99
PTPRC
ITGAM
ADAM12
LY6K
PTPRC
ITGA4
PTPRZ1
NRXN3
'AND' CAR-T Target Pairs
z-score
'AND–NOT' CAR-T Target Pairs
Function
Activator
Inhibitor
HPM Tissues/Cells
Fetal Heart
Fetal Liver
Fetal Gut
Fetal Ovary
Fetal Testis
Fetal Brain
Adult Frontal Cortex
Adult Spinal Cord
Adult Retina
Adult Heart
Adult Liver
Adult Ovary
Adult Testis
Adult Lung
Adult Adrenal
Adult Gallbladder
Adult Pancreas
Adult Kidney
Adult Esophagus
Adult Colon
Adult Rectum
Adult Urinary Bladder
Adult Prostate
Placenta
B Cells
CD4 Cells
CD8 Cells
NK Cells
Monocytes
Platelets
HPA Single Cell Gene Signatures
Blood & Immune Cells
Glial Cells
Neuronal Cells
Germ Cells
Glandular Epithelial Cells
Mesenchymal cells

Figure 3. **Bispecific CAR-T targets in STS: 'AND' and 'AND–NOT' logic-gated CARs**. **(A)** Illustration of an 'AND' CAR-T cell activated by two different tumor antigens. **(B)** Venn diagram showing the number of 'AND' pairs in each STS. **(C)** Bar chart showing the top 5 most frequent 'AND' pairs by STS. Genes are classified based on their HPA single cell type signature. **(D)** Illustration of an 'AND-NOT' CAR-T cell activated by a tumor antigen, and inhibited by a normal cell antigen. **(E)** Venn diagram showing the number of shared and unique 'AND-NOT' pairs in each STS. **(F)** Bar chart showing the top 5 most frequent 'AND-NOT' pairs by STS. Genes are classified based on their HPA single cell signature, with the first indicating the activator and the second the inhibitor. **(G, H)** Heatmaps showing selected 'AND' and 'AND-NOT' pairs expression in the HPA scRNA-seq **(G)**, and the HPM **(H)** datasets. Genes are annotated based on the HPA single-cell signature they belong to (top annotation), and their functional role as activators or inhibitors (bottom annotation, 'AND-NOT' pairs).

2383

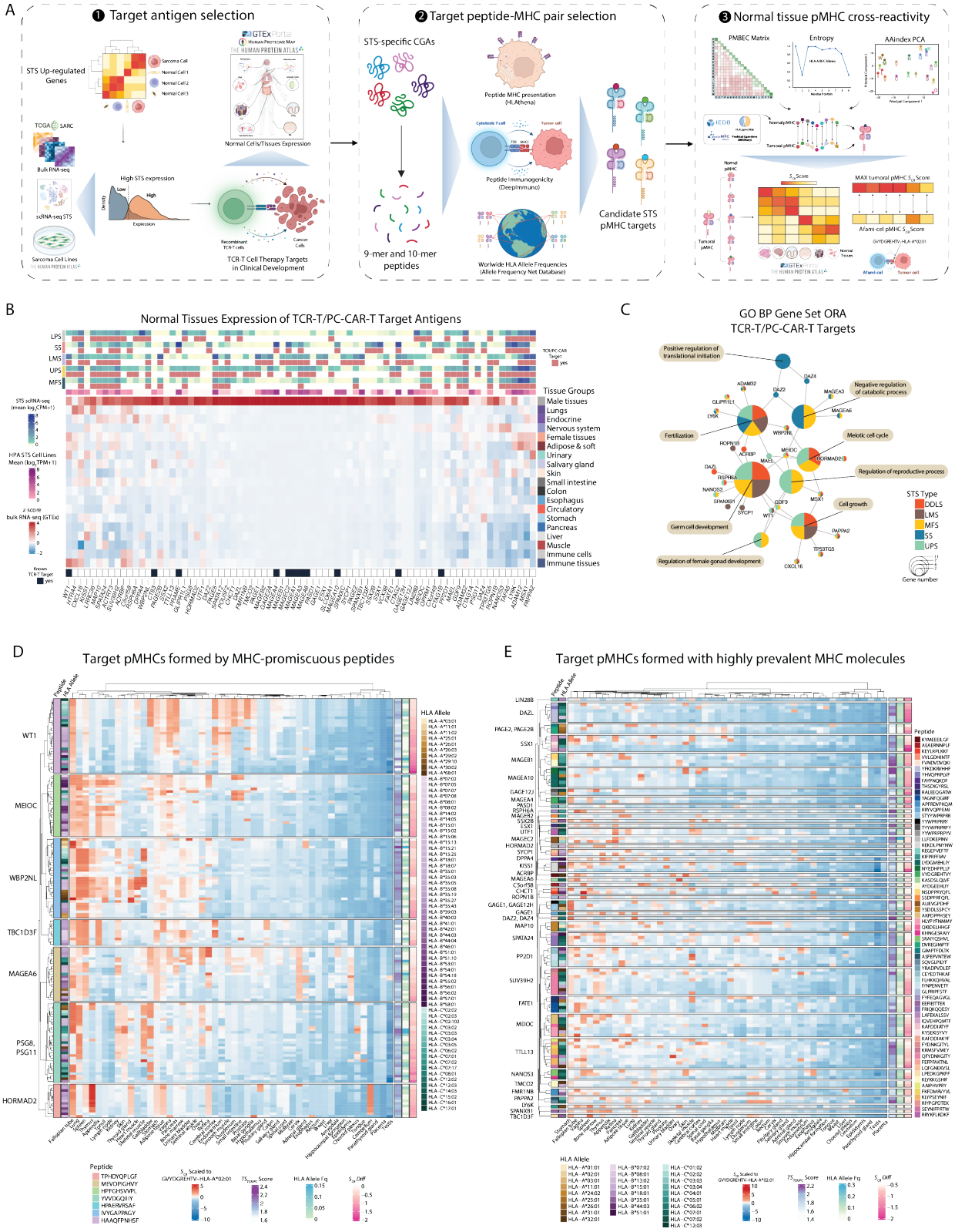
A
❶ Target antigen selection
STS Up-regulated Genes
Normal Cells/Tissues Expression
Bulk RNA-seq
High STS expression
Sarcoma Cell Lines
TCR-T Cell Therapy Targets in Clinical Development
❷ Target peptide-MHC pair selection
STS-specific CGAs
9-mer and 10-mer peptides
Peptide MHC presentation (HLAthena)
Peptide Immunogenicity (DeepImmuno)
Worlwide HLA Allele Frequencies (Allele Frequency Net Database)
Candidate STS pMHC targets
❸ Normal tissue pMHC cross-reactivity
PMBEC Matrix
Entropy
AAindex PCA
MAX tumoral pMHC
Afami cel pMHC
B
Normal Tissues Expression of TCR-T/PC-CAR-T Target Antigens
LPS
SS
LMS
UPS
MFS
Tissue Groups
Male tissues
Lungs
Endocrine
Nervous system
Female tissues
Adipose & soft
Urinary
Salivary gland
Skin
Small intestine
Colon
Esophagus
Circulatory
Stomach
Pancreas
Liver
Muscle
Immune cells
Immune tissues
C
GO BP Gene Set ORA
TCR-T/PC-CAR-T Targets
Positive regulation of translational initiation
Negative regulation of catabolic process
Fertilization
Meiotic cell cycle
Regulation of reproductive process
Cell growth
Germ cell development
Regulation of female gonad development
STS Type
DDLS
LMS
MFS
SS
UPS
Gene number
D
Target pMHCs formed by MHC-promiscuous peptides
WT1
MEIOC
WBP2NL
TBC1D3F
MAGEA6
PSG8, PSG11
HORMAD2
HLA Allele
Peptide
E
Target pMHCs formed with highly prevalent MHC molecules
LIN28B
DAZL
PAGE2, PAGE2B
SSX1
MAGEB1
MAGEA10
GAGE12J
MAGEA4
PASD1
RSPH6A
MAGEB2
SSX2B
ESX1
UTF1
MAGEC2
HORMAD2
SYCP1
DPPA4
KISS1
ACRBP
MAGEA6
C5orf58
CHCT1
ROPN1B
GAGE1, GAGE12H
GAGE1
DAZ2, DAZ4
MAP10
SPATA24
PP2D1
SUV39H2
FATE1
MEIOC
TTLL13
NANOS3
TMCO2
FMR1NB
PAPPA2
LY6K
SPANXB1
TBC1D3F
HLA Allele

Figure 4. **Targeting intracellular antigens in STS: TCR-T and PC-CAR-T cells**.

(A) TCR-T and PC-CAR-T $pMHC$ target discovery pipeline. **(B)** Heatmap showing bulk RNA-seq GTEx expression for the 76 target antigens. Top annotation indicates predicted antigen in each STS, with average expression in STS scRNA-seq data and HPA STS cell lines. **(C)** Gene set ORA using GO BP terms for PC-CAR-T and TCR-T target antigens (FDR < 0.1, minimum and maximum gene set sizes of 2 and 500). **(D**, **E)** Heatmaps showing HPA/GTEx tissue-specific S_{CR} scores for target $pMHCs$ scaled to the S_{CR} scores of afami-cel's target, GVYDGREHTV–HLA-A*02:01, obtained as follows: $\left(\frac{\left(S_{CR}^{j}(target)-S_{CR}^{j}(reference)\right)}{\sigma_j}\right) \times 2\ if\ S_{CR}^{j}(target) > S_{CR}^{j}(reference)$, where σ_j is the standard deviation of the S_{CR} values for the 5,020 target $pMHCs$ in the jth tissue. In **(D)**, $pMHCs$ with $S_{CR}Diff \leq 0$, and formed by peptides that form $pMHCs$ with ≥ 10 different HLA alleles in either locus are shown. In **(E)**, only $pMHCs$ with $S_{CR}Diff \leq 10^{th}$ percentile and formed by one of the top 10 most prevalent HLA alleles at either locus are shown.

2384

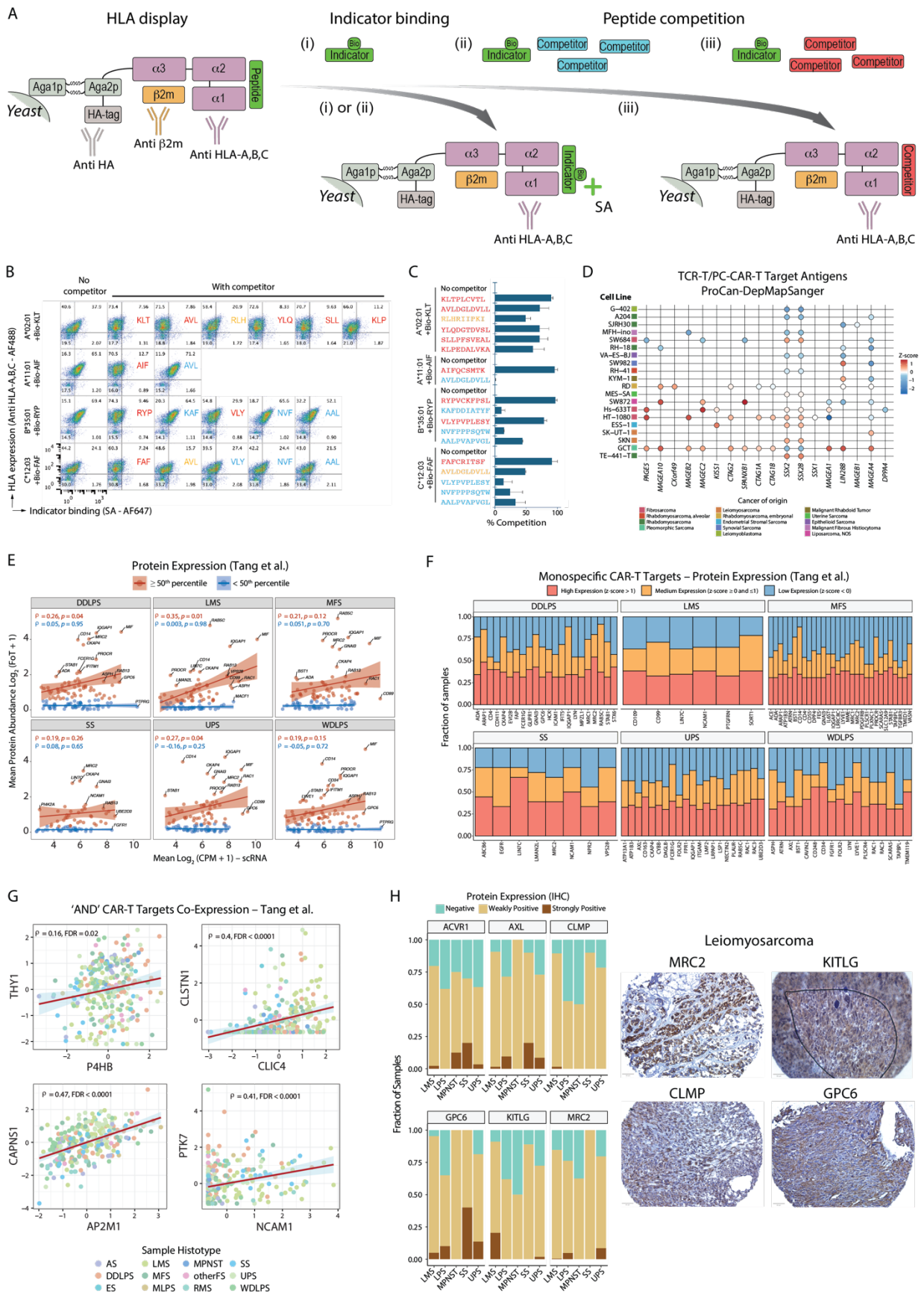
A
HLA display
Indicator binding
Peptide competition
Yeast
Aga1p
Aga2p
HA-tag
α3
α2
β2m
α1
Peptide
Anti HA
Anti β2m
Anti HLA-A,B,C
Indicator
Competitor
SA
B
No competitor
With competitor
HLA expression (Anti HLA-A,B,C - AF488)
Indicator binding (SA - AF647)
C
No competitor
% Competition
D
TCR-T/PC-CAR-T Target Antigens
ProCan-DepMapSanger
Cell Line
Cancer of origin
Z-score
E
Protein Expression (Tang et al.)
≥ 50th percentile
< 50th percentile
DDLPS
LMS
MFS
SS
UPS
WDLPS
Mean Protein Abundance Log2 (FoT + 1)
Mean Log2 (CPM + 1) – scRNA
F
Monospecific CAR-T Targets – Protein Expression (Tang et al.)
High Expression (z-score > 1)
Medium Expression (z-score ≥ 0 and ≤1)
Low Expression (z-score < 0)
Fraction of samples
G
'AND' CAR-T Targets Co-Expression – Tang et al.
THY1
P4HB
CLSTN1
CLIC4
CAPNS1
AP2M1
PTK7
NCAM1
Sample Histotype
AS
DDLPS
ES
LMS
MFS
MLPS
MPNST
otherFS
RMS
SS
UPS
WDLPS
H
Protein Expression (IHC)
Negative
Weakly Positive
Strongly Positive
ACVR1
AXL
CLMP
GPC6
KITLG
MRC2
Fraction of Samples
Leiomyosarcoma
MRC2
KITLG
CLMP
GPC6

Figure 5. **Target validation**.

(A) Vignette demonstrating the IMMUNE workflow for peptide-MHC binding assessment. **(B)** IMMUNE results for 14 $pMHC$ targets identified through PS-TRACT, including KAFDDIATYF and RLHRIIPKI from SSXI, AVLDGLDVLL and VLYPVPLESY from PRAME, NVFPPPSQTW and YLQDGTDVSL from ADAM12, and SLLPFSVEAL, KLPEDALVKA and AALPVAPVGL from MSX1. The scatterplots report the mean fluorescence intensity (MFI) for the peptide-HLA binding indicator (x-axis) and for the HLA molecule tested (y-axis). Each peptide is identified by its first three amino acids. Strong binders are indicated in red, while moderate binders are shown in yellow, and weak binders in light blue. The KLT, AIF, RYP, and FAF peptides are used as positive controls for known strong binders for the HLA considered. **(C)** Bar chart showing the % Competition as measured with IMMUNE, for each tested $pMHC$ target. Strong binders are indicated in red, moderate binders in yellow, and weak binders in light blue. **(D)** Dot plot showing protein expression in STS cell lines in ProCan-DepMapSanger. **(E)** Spearman correlation of scRNA-seq expression and protein abundance in the Tang et al. dataset. Only monospecific CAR-T targets for each histotype are plotted. LPS scRNA-seq samples were compared to DDLPS and WDLPS proteomics samples separately. **(F)** Bar charts showing the fraction of samples in each STS with high, medium, and low expression of selected monospecific CAR-T targets in the Tang et al. dataset. **(G)** Spearman correlation of protein expression in the Tang et al. dataset for 'AND' target pairs. **(H)** Protein expression by IHC in STS tissue micro-arrays. The bar chart shows the fraction of samples in each STS with negative, weakly positive, and strongly positive expression. Images show tissue sections of LMS samples scored as strongly positive for GPC6, MRC2, CLMP, and KITLG. MPNST: malignant peripheral nerve sheath tumor; MLPS: mucinous liposarcoma; others: other fibrosarcoma; ES: Ewing sarcoma; RMS: rhabdomyosarcoma; AS: angiosarcoma.

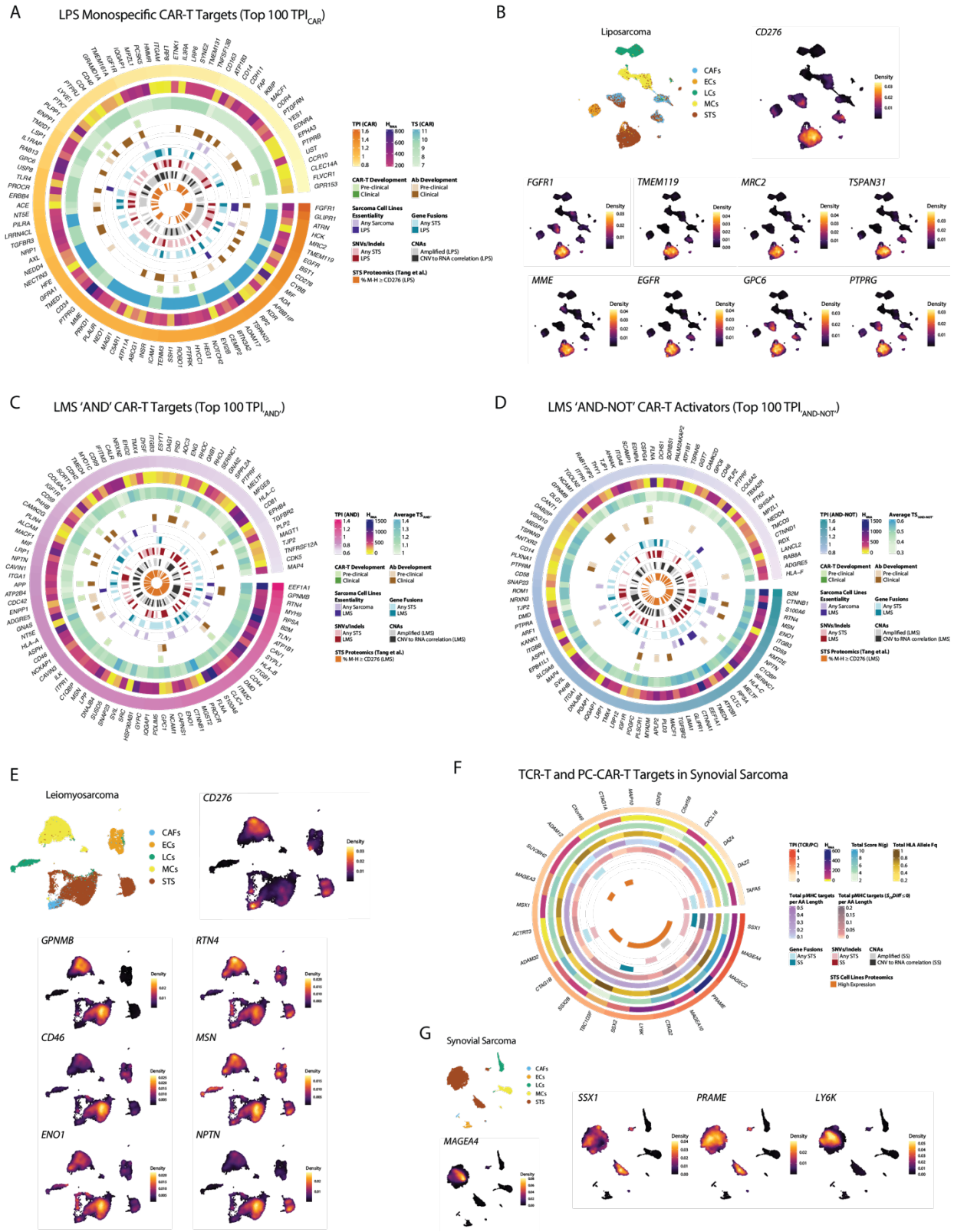
A
LPS Monospecific CAR-T Targets (Top 100 TPI$_{CAR}$)
B
Liposarcoma
CAFs
ECs
LCs
MCs
STS
CD276
FGFR1
TMEM119
MRC2
TSPAN31
MME
EGFR
GPC6
PTPRG
C
LMS 'AND' CAR-T Targets (Top 100 TPI$_{AND}$)
D
LMS 'AND-NOT' CAR-T Activators (Top 100 TPI$_{AND-NOT}$)
E
Leiomyosarcoma
CD276
GPNMB
RTN4
CD46
MSN
ENO1
NPTN
F
TCR-T and PC-CAR-T Targets in Synovial Sarcoma
G
Synovial Sarcoma
MAGEA4
SSX1
PRAME
LY6K

Figure 6. Integrated target prioritization.

(A, **C**, **D)** Circular heatmaps showing the top 100 CAR-T targets ranked by TPI for LPS (**A**, monospecific CAR-T targets), and LMS (**C**, **D**, bispecific CAR-T targets). (**B**, **E**, **G**) UMAP plots showing expression density of selected monospecific CAR-T targets in LPS (**B**), bispecific CAR-T targets in LMS (**E**), and TCR-T/PC-CAR-T target antigens in SS (**G**). In one UMAP plot per panel, cells are colored based on identity. (**F**) Circular heatmap showing TPI scoring for TCR-T and PC-CAR-T target antigens in SS.

2387