Repurposing CAR targets in solid tumors via a pan-cancer single-cell transcriptome analysis

Sanna Madan[1,2], Alejandro A. Schäffer[2,**], Eytan Ruppin[2,**]

[1] Dept. of Computer Science, University of Maryland, College Park MD 20742 USA
[2] Cancer Data Science Laboratory, National Cancer Institute, Bethesda, MD 20892 USA

**To whom correspondence may be addressed at alejandro.schaffer@nih.gov or eytan.ruppin@nih.gov

Abstract

Chimeric antigen receptor (CAR) T cell therapy is a powerful and promising tool for unleashing lasting antitumor immunity, which is currently studied in numerous blood and solid cancers. Here, mining single cell transcriptomics datasets of solid tumors, we survey many existing CAR targets, aiming to new solid tumor types in which these targets are differentially expressed in tumor cells and not on non-tumor cells within the tumor microenvironment, thus providing new repurposing opportunities. Quite surprisingly, we find that many existing solid cancer CAR targets have a better discriminatory power in uncharted cancer types than in the ones they have been developed or being tested for. Furthermore, we show that several existing liquid cancer CAR targets may also hold promise in specific solid tumors. Taken together, this analysis uncovers a large potential of repurposing current CARs that is currently yet untapped.

Introduction

Chimeric antigen receptor (CAR) T cell therapy is a powerful and promising tool for unleashing lasting antitumor immunity against tumor cells. After a patient's T cells are extracted, they are engineered to recognize cancer cell-specific antigens via the addition of a CAR protein on the cell membrane and infused back into the patient. These therapies have shown significant promise in the clinic, and have led to durable remissions in patients, particularly in those with B cell malignancies.[1,2]

As of September 10, 2021, TrialTrove has recorded 1,469 clinical trials that have either taken place, are currently in progress, or are proposed/planned, involving cells engineered to target various membrane surface proteins or other antigens in specific indications.[3] Most of these trials involve modified CAR T cells and some involve NK cells; because our analysis applies to either T or NK cells, we use "CAR" without a suffix in most places. The CAR targets that are currently FDA approved or being tested include CD19 and CD22 in B-cell malignancies, and HER2, IL-13 receptor α2, EGFRvIII, carbonic anhydrase IX, and MUC1 in various solid tumors[1].

Here, we ask whether mining single cell transcriptomics datasets of solid tumors can enable us to identify CAR targets that are differentially expressed in tumor cells and not on non-tumor cells within the tumor microenvironment (TME). We use the term "indication" to refer to a pair of (cancer type, CAR target) that has been proposed in at least one clinical trial; we use the opposite "non-indication" to mean a (cancer type, CAR target) for which we could not find any proposed clinical trial (**Methods**). Our analysis proceeds in two key steps: First, given the reasoning that, a viable CAR target should be expressed more highly on tumor cells than on non-tumor cells, we analyze single cell expression data of known indications of current CAR targets in solid tumors to determine a threshold parameter characterizing their discriminatory power between tumor and nontumor TME cells. Second, based on this analysis, we ask if we can find solid tumor types that match or exceed this discriminatory threshold, thus consequently forming promising indications to which existing CARs aiming at these targets may be repurposed.

Results and Discussion

To address the question at the end of the **Introduction**, we analyzed the scRNA-seq Tumor Immune Single-cell Hub (TISCH) collection, spanning colorectal cancer, glioma, head and neck cancer, liver cancer, non-small cell lung cancer, ovarian cancer, pancreatic adenocarcinoma, melanoma, and stomach cancer[4] (**Methods**). As a criterion for identifying targets which are repurposing candidates, we employed the "area under the curve" (AUC) metric, a proxy of a given clinical protein CAR target's power to discriminate between tumor and non-tumor cells in the tumor microenvironment (TME); *i.e.*, these AUCs were calculated based on the distribution of gene expression levels of each respective protein in the single cell data (**Methods**). Step I of our analysis establishes the threshold AUC for a given target's existing indications. Based on that, Step II of our analysis aims to identify new potential indications for each existing CAR by identifying those whose tumor-vs-non-tumor AUC exceeded the one found in Step I, suggesting such cancer types as promising candidates for repurposing.

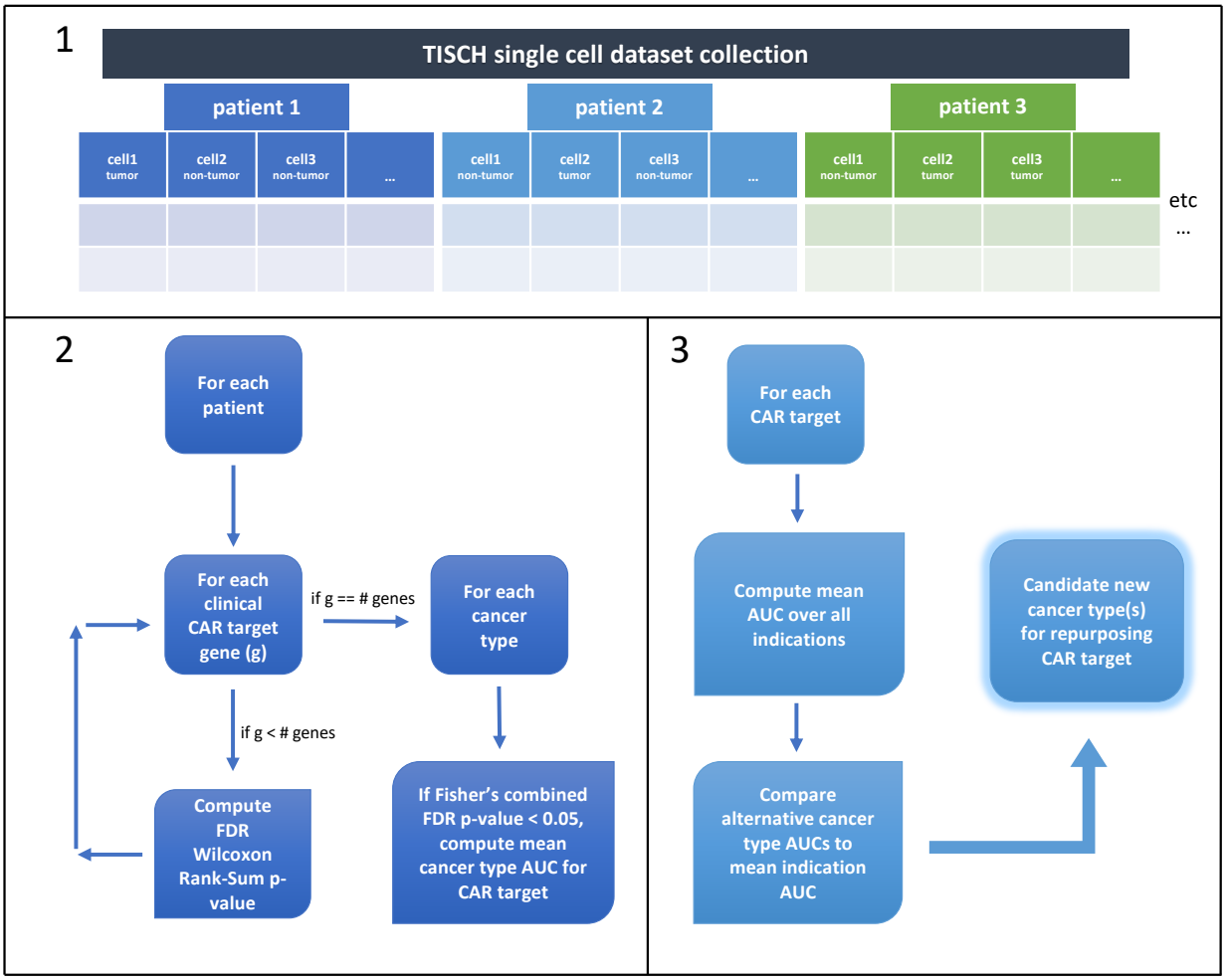

Figure 1. The analysis workflow. First, data is partitioned by patients and quality-control is performed. Next, each clinical CAR target's AUC is computed for each cancer type, along with an FDR-corrected Wilcoxon Rank-Sum p-value denoting the significance of the separation. Second, indication AUCs are compared with non-indication AUCs for each clinical CAR target, and non-indications exceeding the indication AUC thresholds are deemed candidates for repurposing.

Performing this two-step analysis, we identified eight existing CARs that are currently indicated and in clinical trials in various solid tumors for which we find alternative cancer types that show greater tumor-vs-non-tumor discrimination than those observed in the cancer types in which they are currently tested (in which, to the best of our knowledge, they are currently not being tested, as determined by our searches on ClinicalTrials.gov). In a similar manner, we evaluated the AUCs of clinically used CAR targets that are currently only indicated in liquid tumors in our collection of solid tumor datasets. Remarkably, five such targets had exceptionally

high AUCs in at least one solid tumor type and are thus additional worthy targets of exploration for repurposing. In the following, we discuss a few of our most interesting specific findings in more detail. The mean AUCs for each indication tested are shown in **Table 1**, while the full results, including the distribution of mean AUC values obtained for each cancer type in each study cohort are provided in **Supp. Figure 1.**

One top emerging CAR new indication in our results involves the Epithelial cell adhesion molecule, encoded by the *EPCAM* gene. A recent clinical trial has been proposed, but has not yet recruited any patients, for pancreatic, colorectal, gastric, and lung cancers (NCT04151186). In our data, *EPCAM* has a mean AUC of 0.64 in its current indications (Fisher's combined FDR $p < 1.0\times10^{-309}$). Strikingly, it has mean AUCs of 0.79 in NSCLC (where it has only been proposed for but not yet studied) (Fisher's combined FDR $p < 1.0\times10^{-309}$), 0.81 in ovarian cancer (Fisher's combined FDR $p=1.01260\times10^{-78}$), and 0.68 in HNSC (Fisher's combined FDR $p=1.08725\times10^{-101}$). A second result involves the Mucin 1 protein, encoded by the *MUC1* gene, which is in Phase I and II studies in pancreatic, brain, liver, lung, colon, breast, stomach, esophagus, sarcoma, and cervical cancers. Its mean AUC in its current indications is 0.59 (Fisher's combined FDR $p < 1.0\times10^{-309}$), while its AUC in ovarian cancer is 0.86 (Fisher's combined FDR $p=5.49937\times10^{-110}$), highlighting this indication. Thirdly, the protein c-Met (also known as the tyrosine-protein kinase MET), encoded by *MET*, has been targeted in Phase I and II CAR clinical trials in liver, colon, breast, ovarian and renal cancers, and melanoma, where it has a mean AUC of 0.61 in its indications (Fisher's combined FDR $p=1.30910\times10^{-240}$). *MET* has an AUC of 0.78 in HNSC (Fisher's combined FDR $p=1.08725\times10^{-101}$), 0.62 in NSCLC (Fisher's combined FDR $p < 1.0\times10^{-309}$) and 0.67 in PAAD (Fisher's combined FDR $p < 1.0\times10^{-309}$). Fourthly, the transmembrane 4 L6 family member 1, encoded by *TM4SF1*, has been proposed to be used as a CAR target in pancreatic, colorectal, gastric, and lung cancers in a clinical trial, but this trial has not yet begun recruiting patients (NCT04151186). Its mean AUC in its proposed indications is 0.80 (Fisher's combined FDR $p < 1.0\times10^{-309}$), and its AUC in HNSC 0.80 (Fisher's combined FDR $p=1.49486\times10^{-191}$). Fifth, cluster of differentiation 276 (also known as B7 Homolog 3), encoded by *CD276*, is a CAR target in early Phase I, Phase I, and Phase II clinical trials in osteosarcoma, neuroblastoma, and gastric, lung, and brain cancers. It has a mean AUC of 0.52 in its specified indications (Fisher's combined FDR $p=2.56128\times10^{-178}$), and AUCs of 0.71 in

SKCM (Fisher's combined FDR $p=2.56495\times10^{-207}$), 0.66 in HNSC (Fisher's combined FDR $p=1.65618\times10^{-33}$), and 0.53 in PAAD (Fisher's combined FDR p=0.00336). The cluster of differentiation 147 protein, encoded by *BSG*, has been indicated in liver and brain cancers in early Phase I and Phase I clinical trials. Its mean AUC in these indications is 0.56 (Fisher's combined FDR $p < 1.0\times10^{-309}$, but its AUCs are 0.79 in SKCM (Fisher's combined FDR $p=1.43031\times10^{-305}$), 0.66 in HNSC (Fisher's combined FDR $p=1.14712\times10^{-75}$), 0.59 in NSCLC (Fisher's combined FDR $p < 1.0\times10^{-309}$, and 0.57 in CRC (Fisher's combined FDR $p=1.48558\times10^{-23}$). Additionally, the melanoma-associated antigen 4 (encoded by *MAGEA4*), has been a CAR target in a Phase I and Phase II clinical trial in lung cancer (NCT03356808), with a mean AUC of 0.58 in NSCLC (Fisher's combined FDR $p < 1.0\times10^{-309}$). Its AUCs are 0.74 in HNSC (Fisher's combined FDR $p=2.07492\times10^{-185}$) and 0.66 in SKCM (Fisher's combined FDR $p=1.15409\times10^{-142}$). Lastly, CARs targeting the human epidermal growth factor receptor 2 (HER2), encoded by *ERBB2*, have been used in numerous Phase I and II clinical trials across solid tumors, including pancreatic, brain, liver, lung, colon, breast, ovarian, stomach, prostate, esophageal, head and neck, bladder, endometrial, and gastric cancers, and sarcomas. *ERBB2* has a mean AUC of 0.59 in its indications (Fisher's combined FDR $p < 1.0\times10^{-309}$) and an AUC of 0.60 in SKCM (Fisher's combined FDR $p=5.47906\times10^{-53}$).

Beyond studying new indications for CARs already indicated for solid tumors, we additionally examined if CAR designed for blood cancers may be repurposed for targeting solid tumors (**Table 2**). CARs for liquid tumors generally are not designed to target proteins expressed significantly higher on cancer cells than on their non-cancer counterparts, with the expected killing of non-cancer cells (*e.g.*, hematopoietic stem and progenitor cells (HSPCs) and B lymphocytes) being tolerated. Indeed, the tumor-vs-non-tumor discriminatory power of the Phase I, II, III liquid tumor CAR targets CD19, CD33, and CD123 (encoded by *IL3RA*) was quite poor in acute myeloid leukemia (AML), with AUCs of 0.50, 0.51, and 0.52, respectively. However, notably, some of the liquid CAR targets do have strikingly high AUCs in various solid tumors. These include Interleukin 3 receptor alpha (also known as CD123), encoded by *IL3RA*, with an AUC of 0.72 in STAD (Fisher's combined $p= 3.92365\times10^{-81}$). The transmembrane heparan sulfate proteoglycan syndecan 1 (also known as CD138), encoded by the *SDC1* gene, had a strikingly high AUC of 0.92 in HNSC (Fisher's combined FDR $p < 1.0\times10^{-309}$), 0.74 AUC in

PAAD (Fisher's combined FDR $p < 1.0\times10^{-309}$), 0.70 AUC in NSCLC (Fisher's combined FDR $p < 1.0\times10^{-309}$), 0.74 AUC in LIHC (Fisher's combined FDR $p= 1.18384\times10^{-82}$), and 0.65 AUC in CRC (Fisher's combined FDR $p=5.79612\times10^{-63}$). The transmembrane phosphoglycoprotein, encoded by *CD34,* had an AUC of 0.73 in STAD. The neural cell adhesion molecule (also called CD56), encoded by the *NCAM1* gene, had an AUC of 0.66 in GBM (Fisher's combined FDR $p < 1.0\times10^{-309}$). Finally, *IL1RAP*, encoding the interleukin 1 receptor accessory protein, had an AUC of 0.69 in SKCM (Fisher's combined FDR $p=8.21198\times10^{-163}$) and 0.75 in HNSC (Fisher's combined FDR $p=1.38429\times10^{-120}$).

Protein	Gene	Solid tumor indications	NCT IDs	Mean indication AUC	Repurposing candidate AUCs
Mucin 1	*MUC1*	pancreatic, brain, liver, lung, colon, breast, stomach, esophagus, sarcoma, cervical	NCT03706326, NCT03633773, NCT03525782, NCT03356795, NCT03356808, NCT03356782, NCT03267173, NCT03198052, NCT03179007, NCT03170141, NCT02839954, NCT02617134, NCT02587689, NCT04842812	0.59	OV: 0.86
EPCAM	*EPCAM*	pancreatic, liver, colon, breast, stomach, esophageal, and nasopharyngeal, proposed: gastric and lung	NCT03563326, NCT03013712, NCT02915445, NCT02725125, NCT02729493	0.64	NSCLC: 0.79 OV: 0.81 HNSC: 0.68
c-Met	*MET*	liver, colon, breast, ovarian, renal, and melanoma	NCT03672305, NCT03638206, NCT03060356, NCT01837602	0.61	HNSC: 0.78 PAAD: 0.67 NSCLC: 0.62
TM4SF1	*TM4SF1*	proposed: pancreatic, colorectal, gastric, and lung cancers	NCT04151186	0.80	HNSC: 0.80
B7-H3	*CD276*	osteosarcoma, neuroblastoma, and gastric, lung, brain	NCT04691713, NCT04432649, NCT04433221, NCT04637503, NCT04385173, NCT04077866, NCT04842812	0.52	SKCM: 0.71 HNSC: 0.66 PAAD: 0.53

CD147	*BSG*	liver, brain	NCT04045847, NCT03993743	0.56	SKCM: 0.79 HNSC: 0.66 NSCLC: 0.59 CRC: 0.57
MAGE-A4	*MAGEA4*	lung	NCT03356808	0.58	SKCM: 0.66 HNSC: 0.74
HER2	*ERBB2*	pancreatic, brain, liver, lung, colon, breast, ovarian, stomach, prostate, esophageal, sarcoma, head and neck, bladder, endometrial, gastric	NCT03740256, NCT03696030, NCT03389230, NCT03500991, NCT03423992, NCT03383978, NCT03267173, NCT03198052, NCT02442297, NCT02713984, NCT02547961, NCT01935843, NCT00902044, NCT01109095, NCT00889954, NCT00924287, NCT04511871, NCT04650451, NCT04660929, NCT04842812, NCT04995003, NCT04433221, NCT04430595	0.59	SKCM: 0.60

Table 1. Overview of key solid CAR repurposing indications in solid tumors. Each row contains each protein CAR target, its encoding gene name, its indications, NCT IDs, mean AUC of the current indication(s) and those of the repurposing indication candidates.

Protein	Gene	Solid tumors with mean AUC > 0.65
CD123	*IL3RA*	STAD: 0.72
CD138	*SDC1*	HNSC: 0.92, LIHC: 0.74, PAAD: 0.74, NSCLC: 0.70, CRC: 0.65
CD56	*NCAM1*	GBM: 0.66
CD34	*CD34*	STAD: 0.73
IL-1-RAP	*IL1RAP*	HNSC: 0.75, SKCM: 0.69

Table 2. Overview of key blood cancer CAR repurposing indications in solid tumors. Similarly, this Table highlights key liquid tumor CAR targets which have high (> 0.65) AUCs in at least one solid tumor.

Our study has a few limitations that we would like to acknowledge. First and foremost, we have used gene expression as a proxy for protein expression, but it is ultimately proteins that are targeted by CARs in the clinic. When advances in high-throughput single-cell proteomics technology emerge, it will be pertinent to apply the approach presented in this study here to evaluate the discriminatory potential of membrane protein receptors as viable CAR targets. In the meantime, with the accumulating plethora of large scRNA-seq datasets of tumors, gene expression is the best estimate available to us. Second, the use of the AUC metric to evaluate target discrimination potential may bear some limitations. In practice, optimal discrimination relies on knowing antigen detection thresholds of CAR-T cells. Currently, our control in setting antigen detection density of CARs is poor, but emerging efforts may allow us to tune antigen detection cutoffs[5]. This would allow us to evaluate CAR target viability more practically. Finally, in future studies, it would be of great interest to evaluate novel CAR targets, dual CAR targets, and multi-targeted logical CAR targets, based on the approach outlined herewith.

In summary, our analysis offers repurposing indications to several current CAR treatments that are in use or are being tested in solid tumors. Our analysis shows that in these indications, these CAR targets actually have a better discriminatory power than in the ones they are being tested. Additionally, we show that existing liquid cancer CAR targets may also hold promise in specific solid tumors. Taken together, this analysis uncovers a large potential of repurposing current CARs, that is currently yet untapped.

Methods

Clinical trial curation

To obtain a list of CAR clinical trials and targets in solid tumors, we first obtained a list of single clinical CAR targets from Mason *et al.*[3], who carried out their ClinicalTrials.gov query

on March 3, 2019. To obtain a complete up-to-date list of clinical trials since then, an additional search was carried out on ClinicalTrials.gov on August 20, 2021 for CAR-T clinical trials posted on March 3, 2019 and after. The following query was used: "*CAR-T OR CAR OR chimeric antigen receptor OR chimeric immunoreceptors OR artificial T-cell receptors | Recruiting, Not yet recruiting, Active, not recruiting, Completed, Enrolling by invitation, Suspended, Terminated Studies | Interventional Studies | (cancer OR carcinoma OR solid tumor OR solid tumors OR melanoma OR sarcoma) NOT (hematological OR lymphoid OR liquid cancers OR lymphoma OR leukemia OR CD19 OR B cell malignancy OR myeloma OR Hematologic OR blood cancer) | First posted from 03/03/2019 to 08/20/2021*". CAR trials and targets were further identified through manual curation, and clinical trial statuses were cross-checked on TrialTrove (https://pharmaintelligence.informa.com/products-and-services/data-and-analysis/trialtrove) to ensure consistency.

CAR protein targets were then selected for further analyses if they had a corresponding single gene that could be analyzed for gene expression. The following molecules were left out of the analysis: Claudin 18.2, CD44v6, EGFRvIII, k-IgG, LeY, CEA, and GD2 because i) they are not proteins or ii) they are protein isoforms whose presence and abundance cannot easily be estimated from single cell gene expression data. For each target, its indications were identified as the specific tumor types which were mentioned in its ClinicalTrials.gov records, excluding 'solid tumors' and 'metastases'. The following mapping of clinical trial indications to our datasets was used: brain cancer or glioma → glioblastoma; pancreatic cancer → pancreatic adenocarcinoma; stomach cancer or gastric cancer → stomach adenocarcinoma; liver or hepatocellular carcinoma → liver hepatocellular carcinoma; head and neck cancer → head and neck squamous cell carcinoma; melanoma → skin cutaneous melanoma; colon cancer or colorectal cancer → colorectal cancer; ovarian cancer → ovarian serous cystadenocarcinoma; lung cancer or non-small cell lung cancer (but not small cell lung cancer) → non-small cell lung cancer; breast cancer → breast invasive carcinoma. Liquid tumor CAR targets were taken from MacKay *et al.*[3]

Single cell data collection

We downloaded single cell transcriptomics datasets from the Tumor Immune Single-cell Hub (TISCH) database[2]. This resource contains 76 uniformly processed scRNA-seq datasets

spanning 28 cancer types and containing data from over 2 million cells. We selected datasets from human tissues which contained *both* malignant and non-malignant cells. This yielded a total of 26 unique solid tumor datasets: two of breast cancer (BRCA), one of colorectal cancer (CRC), eleven of glioblastoma (GBM), one of head and neck squamous cell carcinoma (HNSC), one of liver hepatocellular carcinoma (LIHC), four of non-small cell lung cancer (NSCLC), one of ovarian cancer (OV), two of pancreatic adenocarcinoma (PAAD), two of skin cutaneous melanoma (SKCM), and one of stomach adenocarcinoma (STAD). Additionally, an acute myeloid leukemia (AML) dataset was analyzed. These data were all uniformly processed by TISCH as $\log_2(TPM/10+1)$ values (TPM stands for transcripts per million).

Single cell data analysis

In-house programs written in Python 3 were used to carry out analyses on the scRNA-seq datasets. A cell was labeled 'Tumor' if it was labeled 'Malignant' by TISCH, and 'Non-tumor' otherwise. We then partitioned the cells in each dataset by patient (if such annotation was available); the partition by patient is so that each patient counts equally in the downstream analysis regardless of how many cells were sampled from that patient. As quality control, only patient samples which contained at least 25 tumor cells and 25 non-tumor cells were selected for analysis. After this filtering step, only the cancer types where we had samples from at least three patients remaining were subsequently considered, excluding breast cancer entirely from our analysis, for which we only had two patients. While the key results presented in the main text were generated using the 25-cell-per-class threshold, each result was only included if it held robustly across increasing cell-count thresholds of 25, 35, 45, 55, 65, 75, 85, 95, and 105 cells-per-class. Notably, all key findings that emerged when using the 25-cell-per-class threshold held robustly across every threshold tested.

Next, we generated patient-specific-AUCs of tumor-vs-non-tumor cell discrimination for each gene whose respective protein was listed in a CAR clinical trial, along with an associated Wilcoxon Rank-Sum p-value indicating the statistical significance of the tumor-vs-non-tumor separation by the gene. Bonferroni FDR p-value correction was performed for each patient. Then, for each clinical CAR target, a mean AUC over all patients for each cancer type was computed, with the Fisher's combined probability test used to aggregate the FDR p-values

associated with each patient AUC. A p-value cutoff of 0.05 denoted the significance of the mean AUC. Next, for each target, we computed a mean AUC across the tumor types in our data in which the target has been already clinically indicated, their p-values again aggregated with the Fisher's combined probability test. Hereafter, only mean AUCs with aggregated p-values less than 0.05 were considered. The AUC in each cancer type not in the set of indications for a given clinical CAR target was compared against the mean AUC threshold established over the existing known indications of that respective target. If a clinical target AUC in an alternative cancer type was equal to or higher than the threshold AUC established in its indications, it was considered a candidate for repurposing to that tumor type.

Acknowledgements

This research is supported in part by the Intramural Research Program of the National Institutes of Health, NCI, CCR. S.M. is supported by the NCI-UMD Partnership for Integrative Cancer Research Program. The NIH HPC Biowulf cluster (http://hpc.nih.gov) was used for computational analyses.

References

1. Milone, M. C., Xu, J., Chen, S. J., Collins, M. A., Zhou, J., Powell, D. J., & Melenhorst, J. J. (2021). Engineering-enhanced CAR T cells for improved cancer therapy. Nature Cancer, 2(8), 780-793.

2. Klebanoff, C. A., Rosenberg, S. A., & Restifo, N. P. (2016). Prospects for gene-engineered T cell immunotherapy for solid cancers. Nature medicine, 22(1), 26-36.

3. MacKay, M., Afshinnekoo, E., Rub, J., Hassan, C., Khunte, M., Baskaran, N., ... & Mason, C. E. (2020). The therapeutic landscape for cells engineered with chimeric antigen receptors. Nature biotechnology, 38(2), 233-244.

4. Sun, D., Wang, J., Han Y., Dong, X., Ge, J., Zheng, R., … &Wang, C. (2021) TISCH: a comprehensive web resource enabling interactive single-cell transcriptome visualization of tumor microenvironment, Nucleic Acids Research, 49(D1):D1420-D1430.

5. Williams, J. Z., Allen, G. M., Shah, D., Sterin, I. S., Kim, K. H., Garcia, V. P., ... & Lim, W. A. (2020). Precise T cell recognition programs designed by transcriptionally linking multiple receptors. Science, 370(6520), 1099-1104.