CAB-CAR-T: The Prioritization of Cell Surface Protein Targets for Conditionally Active Biologics to Treat All Solid Tumors

Hiba Shaban[1], Laurence Jadin[2], James Onuffer[2], Farzad Haeridazeh[2], Alissa Kerner[1], Gregory Schreiber[1], Gregory Frost[1] and Scooter Willis[1]

[1]F1 Oncology, West Palm Beach FL, USA; [2]F1 Oncology, San Diego CA, USA

5101

Key Messages

1. With the advent of Chimeric Antigen Receptor (CAR) T cell therapy to treat cancer a new category of targetable biomarkers has emerged that are associated with the surface of malignant cells and serve as targets for cytotoxic T cells
2. By modeling target mRNA expression from large scale genomic datasets, selection of targets with desirable properties can be used to maximize patient coverage
3. We showed that with 5-7 CAB-CAR-T products it is possible to treat 90% of patients with solid tumor malignancies

Introduction

Highly expressed proteins on the surface of cancer cells represent viable targets for conditionally active biologic chimeric antigen receptor T-cell (CAB-CAR-T) therapies that have preferential activity in the tumor microenvironment (see poster #3189). By modeling properties of cell surface proteins from The Cancer Genome Atlas (TCGA) datasets[1], optimal targets were identified across all TCGA cancer malignancies that provided the fewest number of CAB-CAR-T treatment options that can cover the most patients in TCGA cohorts.

Materials and methods

Cell surface proteins (n=1086) were identified from multiple public databases, and metadata specific to each gene was organized for data modeling. Multiple CAB-CAR-T properties were used to rank the cell surface proteins as therapeutic targets. Thirty-one TCGA cohorts that represented a comprehensive collection of genomic profiled tumor samples and cancer outcomes were used to rank cell surface protein candidates. Pre-determined mRNA cut-offs were used to rank cell surface proteins in each cohort to determine the percentage of patients potentially eligible for treatment.

Multiple approaches were used to filter the ranked list based on ideal CAB-CAR-T properties including:
1) High expression in the patients admitted into TCGA cancer cohorts[1].
2) High expression in cancer cell lines from the Cancer Cell Line Encyclopedia (CCLE) indicating mRNA expression is a feature of cancer cell lines[4] and
3) low expression in heart, lung, liver, brain, skeletal muscle, and stomach, which are considered to be critical tissues[7].

Different ranked lists of cell surface proteins were used to determine the number of CAB-CAR-T products required to treat 90% of patients in TCGA cohorts. A patient with the highest mRNA expression above the mean plus one standard deviation as determined across all TCGA samples was assigned to that specific protein biomarker as eligible for treatment and removed from the list of patients still to be treated. A bootstrap p-value for the ranked lists was determined by calculating the minimum number of randomly selected cell surface proteins that would give 90% coverage of the TCGA cohort.

Results

It was shown that it is feasible to find a ranked list of genes with high mRNA expression in TCGA datasets and minimum expression in off-target critical tissues such that 5-7 CAB-CAR-T products could be used to treat 90% of TCGA patients. To achieve 100% treatment coverage each additional CAB-CAR-T product added to the list had minimum inclusion of additional patients for treatment.

Gene	acc	blca	brca	cesc	chol	coadread	dlbc	esca	gbm	hnsc	kich	kirc	kirp	lgg	lihc	luad	lusc	meso	ov	paad	pcpg	prad	sarc	skcm	stad	tgct	thca	thym	ucec	ucs	uvm	Running %
VP1000	0.0	0.7	5.9	1.3	16.7	7.7	45.8	15.2	2.6	0.2	90.9	19.1	60.0	9.5	9.7	12.0	1.0	11.5	4.0	18.5	10.6	29.2	5.0	79.8	23.9	1.3	18.0	4.2	8.0	0.0	100.0	19.8
VP1001	89.9	11.8	16.9	12.8	5.6	44.1	0.0	12.0	8.5	11.0	3.0	5.8	1.4	1.6	1.6	15.5	14.6	13.8	21.5	5.6	64.8	51.5	14.3	4.8	4.8	5.3	4.4	0.0	60.2	35.1	0.0	37.2
VP1002	0.0	2.7	43.4	11.5	2.8	0.3	0.0	24.5	0.7	27.9	0.0	14.4	3.8	2.9	0.0	5.2	29.9	10.3	15.2	0.6	4.5	5.4	10.0	3.8	1.9	10.7	17.6	20.8	1.1	8.8	0.0	46.3
VP1003	1.3	3.7	7.9	4.9	5.6	8.4	6.2	7.6	2.0	9.4	3.0	0.4	0.3	0.2	65.5	10.5	9.0	0.0	1.7	1.7	12.8	2.8	4.6	2.9	14.5	2.0	0.0	0.0	0.0	0.0	0.0	52.4
VP1004	0.0	11.8	11.7	32.2	0.0	0.5	0.0	21.7	0.0	38.7	0.0	19.7	8.3	0.0	0.3	35.1	28.9	1.1	34.7	5.1	0.0	0.0	0.0	1.0	3.6	0.0	1.0	0.8	10.2	1.8	0.0	61.0
VP1005	0.0	0.5	2.3	0.0	0.0	0.3	0.0	3.3	1.3	0.4	0.0	0.8	0.0	5.6	1.3	0.4	0.8	0.0	2.0	0.0	0.0	0.0	3.1	0.0	2.7	54.7	0.0	0.0	0.0	15.8	0.0	64.1
VP1006	0.0	6.9	0.4	3.0	16.7	20.8	0.0	5.4	0.0	0.4	0.0	0.2	0.7	0.0	0.3	2.9	5.8	3.4	2.0	9.6	0.0	0.0	0.0	0.0	22.7	0.0	0.8	0.8	0.6	0.0	0.0	67.5
VP1007	0.0	26.7	0.5	3.3	2.8	0.0	0.0	1.6	0.0	1.7	0.0	4.9	1.7	0.0	1.1	0.0	0.6	1.1	1.7	0.6	0.6	0.0	4.2	1.9	0.5	0.7	0.6	0.0	2.3	1.8	0.0	69.4
VP1008	0.0	3.2	2.2	3.9	5.6	2.9	0.0	0.5	16.3	1.7	0.0	1.3	3.4	22.7	0.8	4.7	1.4	2.3	1.3	27.5	0.0	0.2	4.2	0.0	5.1	12.0	7.6	0.0	2.3	1.8	0.0	73.8
VP1009	3.8	6.4	0.3	1.6	38.9	1.3	0.0	1.6	60.8	0.8	0.0	28.7	14.8	55.2	5.7	0.4	0.2	4.6	0.3	1.7	2.2	0.0	4.2	1.0	5.3	8.7	0.0	3.3	2.8	12.3	0.0	82.4
Total	95.0	74.4	91.5	74.5	94.7	86.3	52.0	93.4	92.2	92.2	96.9	95.3	94.4	97.7	86.3	86.7	92.2	48.1	84.4	70.9	95.5	89.1	49.6	95.2	85.0	95.4	50.0	29.9	87.5	77.4	100.0	

Table 1. Percentage coverage of patients in TCGA cohorts when cutoff for inclusion is mean + 1 Stdev. Each gene is a possible surface protein target for CAR-T therapy selected for by our modeling criteria, and each column is a TCGA cohort abbreviation. Running percentage column shows the total percentage of patients that fall under coverage cutoff. The last row is the sum of percentage coverage in each cancer cohort.

Gene	acc	blca	brca	cesc	chol	coadread	dlbc	esca	gbm	hnsc	kich	kirc	kirp	lgg	lihc	luad	lusc	meso	ov	paad	pcpg	prad	sarc	skcm	stad	tgct	thca	thym	ucec	ucs	uvm	Running %
VP1000	3.8	8.3	55.9	12.2	86.1	78.6	75.0	45.1	60.1	10.2	95.5	86.3	97.9	83.1	82.5	68.3	28.5	64.4	49.8	94.4	56.4	95.0	32.0	94.2	83.4	27.3	98.8	28.3	48.3	21.1	100.0	60.3
VP1001	96.2	39.0	27.5	28.9	2.8	13.2	0.0	33.7	12.4	37.7	3.0	8.8	1.4	2.1	1.9	15.9	38.3	21.8	30.7	2.8	43.6	4.6	44.0	1.9	6.0	20.7	0.2	25.8	44.3	59.6	0.0	81.9
VP1002	0.0	27.0	13.3	40.5	2.8	2.6	0.0	17.9	5.9	46.3	1.5	3.8	0.7	13.2	0.3	7.8	28.9	9.2	14.5	1.1	0.0	0.2	14.3	1.9	3.4	42.0	0.8	28.3	2.3	14.0	0.0	93.0
VP1003	0.0	18.9	2.8	15.1	5.6	5.3	10.4	2.2	19.6	5.6	0.0	0.8	0.0	1.0	14.8	7.6	3.4	1.1	4.0	1.1	0.0	0.2	4.2	1.9	5.8	6.7	0.0	0.0	4.5	1.8	0.0	97.7
VP1004	0.0	3.7	0.4	3.0	2.8	0.3	2.1	1.1	0.7	0.0	0.0	0.2	0.0	0.2	0.3	0.4	0.8	1.1	1.0	0.6	0.0	0.0	0.4	0.0	1.2	2.0	0.0	0.8	0.6	0.0	0.0	98.5
VP1005	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.7	0.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	0.0	0.0	1.3	0.0	0.0	0.0	0.0	0.0	98.6
VP1006	0.0	2.7	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	2.3	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.2	0.8	0.0	0.0	0.0	98.9
VP1007	0.0	0.2	0.0	0.0	0.0	0.0	4.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.0	0.0	0.0	0.0	15.0	0.0	3.5	0.0	99.6
VP1008	0.0	0.0	0.0	0.0	0.0	0.0	6.2	0.0	0.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9
VP1009	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9
Total	100.0	99.8	100.0	100.0	100.1	100.0	97.9	100.0	100.1	100.0	100.0	100.1	100.0	100.0	99.8	100.0	99.9	99.9	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	99.0	100.0	100.0	100.0	

Table 2. Percentage coverage of patients in TCGA cohorts when cutoff for inclusion is mean.

Results – continued

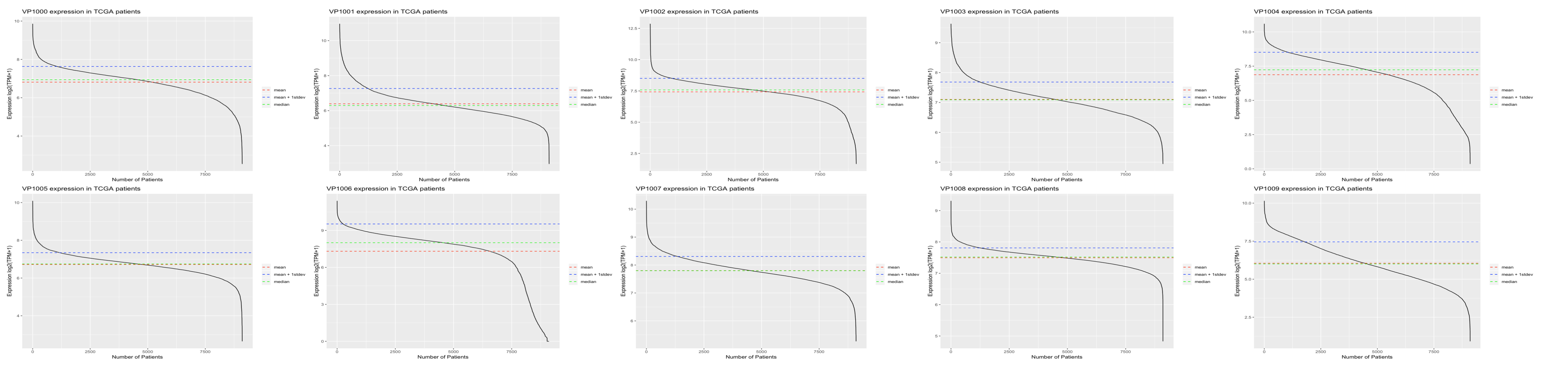

Fig 1. Line graphs showing the distribution of expression of the top 10 genes in approximately 9,000 TCGA patients. Each graph represents a gene. The red and blue horizontal lines show two cutoffs used for patient inclusion, mean and mean + 1stdev. The green horizontal line is the median. Expression values are from TCGA datasets[1] and graphs were generated using R[6].

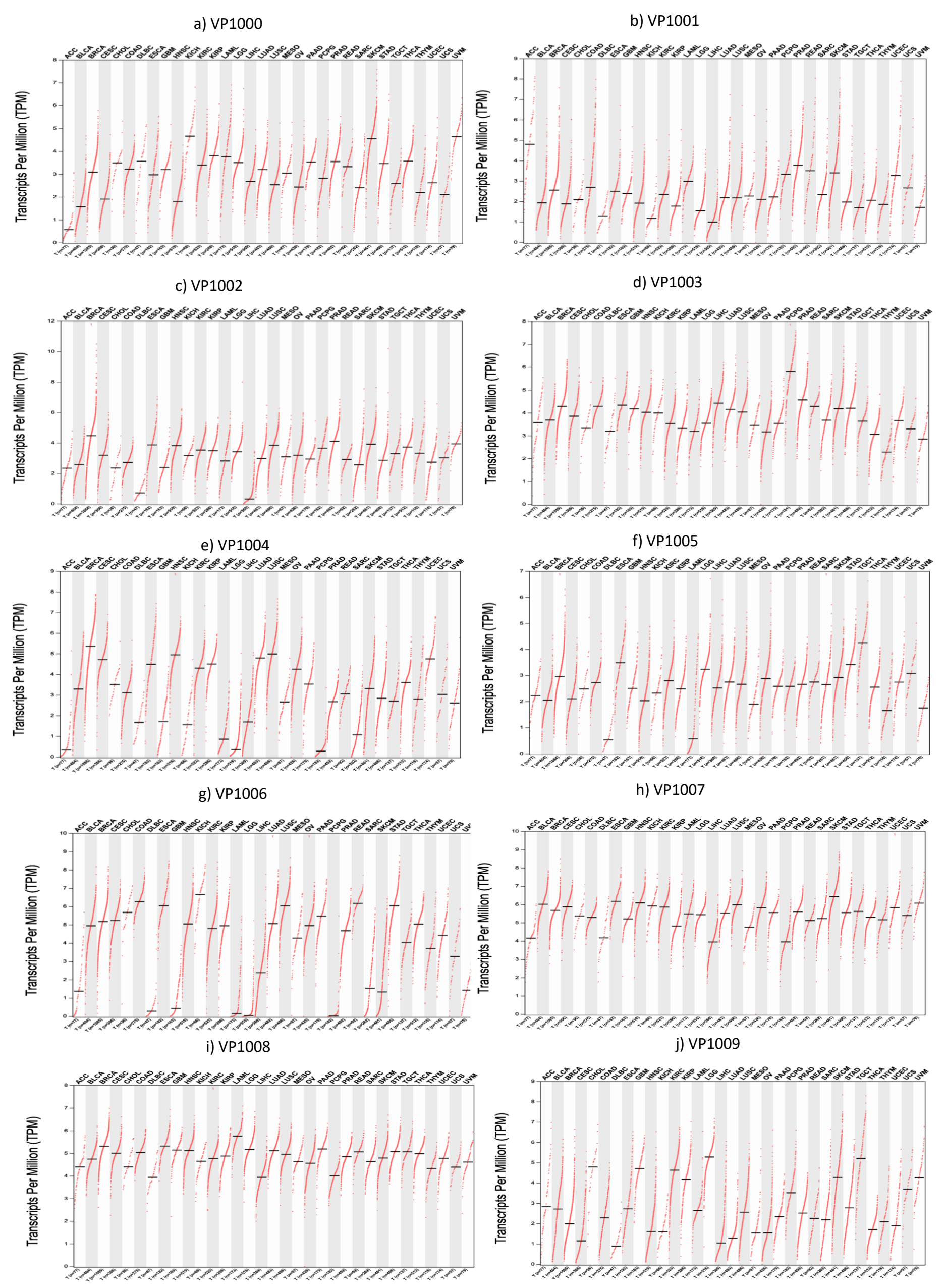

Fig 2. Dot plots showing expression of gene in TCGA cohorts. Each TCGA cohort and the number of subjects enrolled is shown on the x-axis and expression values in Log2(Transcripts per million + 1) are shown on the y-axis. Cohort abbreviations shown in full in the legend to the left. These graphs were generated on http://gepia.cancer-pku.cn/index.html[2]

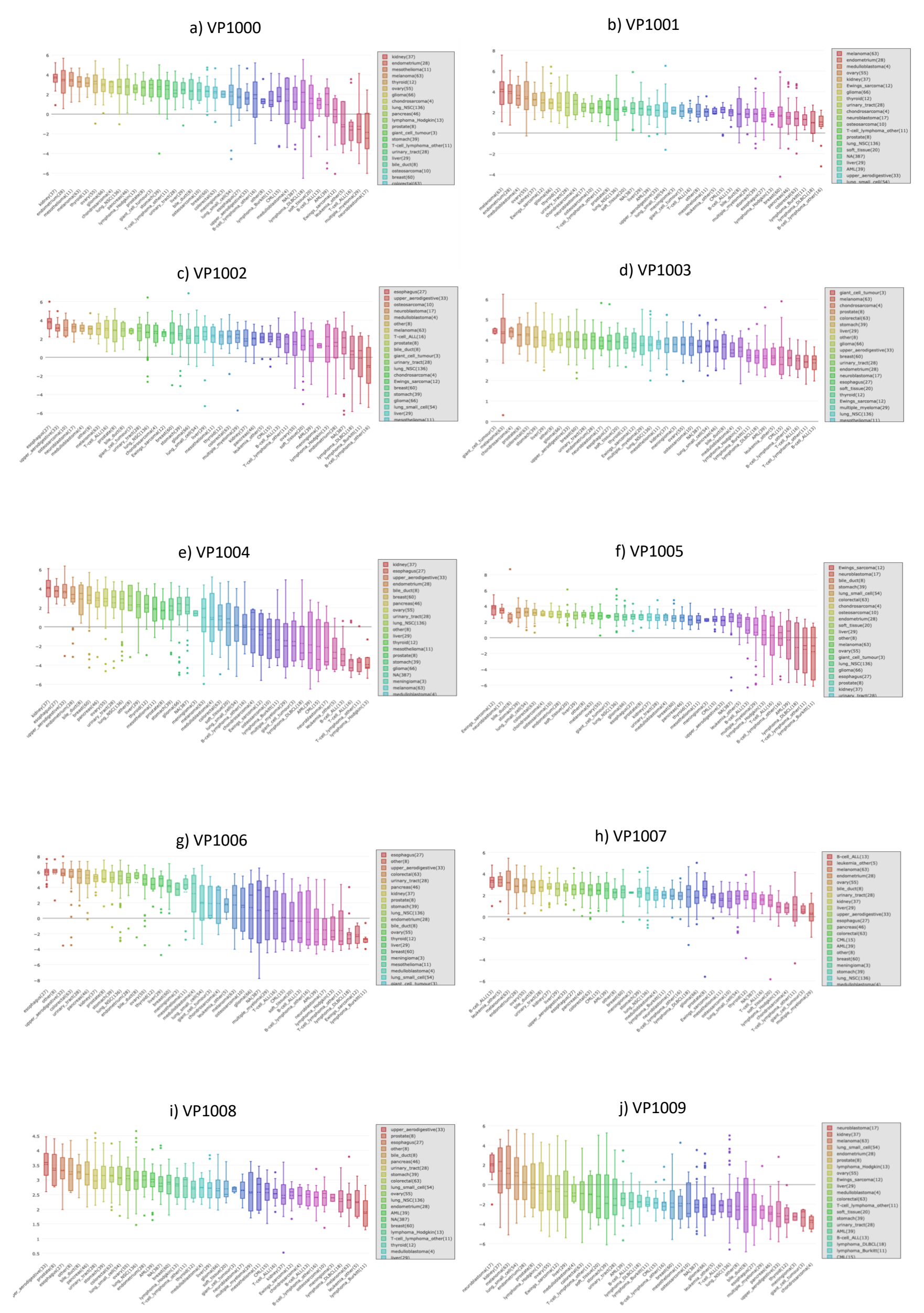

Fig 3. Box plots displaying the distribution of the genes expression data in cancer cell line lineages. The box plot is sorted and colored by average distribution of a gene's expression in a lineage, and lineages are composed of a number of cell lines from the same area or system of the body. The number next to the lineage name indicates how many cell lines are in the lineage. The highest average distribution is on the left and is colored red, and the dashed line within a box is the mean. When looking for cell lines to test our CAR-T therapies on we can look within the high expression cell line lineages or cell lines with comparable expression to cancer patients. These graphs are from https://portals.broadinstitute.org/ccle[4]

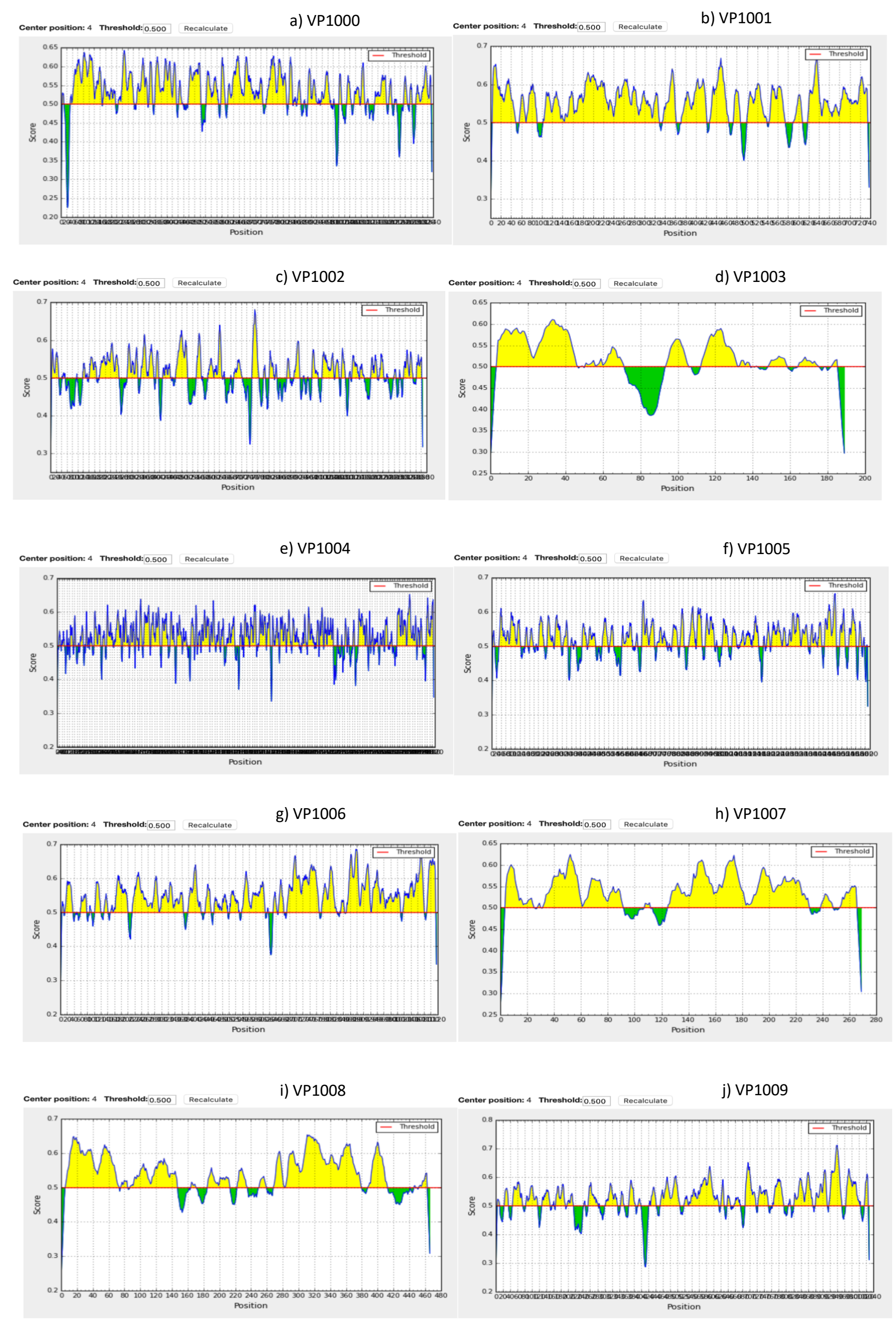

Fig 4. Prediction of continuous antibody epitope from protein sequence. Tools.Immuneepitope.org/bcell[3] was used to predict the possible external sites in a protein from the primary sequence that could function as docking sites for antibodies. The algorithm used to generate these graphs is the BepiPred-2.0: Sequential B-Cell Epitope Predictor with a threshold of 0.5 and a center position of 4. Yellow peaks are predicted external sites. Each graph represents one of the top 10 genes in order.

Results – continued

ACC - Adrenocortical carcinoma
BLCA - Bladder Urothelial Carcinoma
BRCA - Breast invasive carcinoma
CESC - Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL - Cholangiocarcinoma
COAD - Colon adenocarcinoma
DLBC - Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
ESCA - Esophageal carcinoma
GBM - Glioblastoma multiforme
HNSC - Head and Neck squamous cell carcinoma
KICH - Kidney Chromophobe
KIRC - Kidney renal clear cell carcinoma
KIRP - Kidney renal papillary cell carcinoma
LAML - Acute Myeloid Leukemia
LGG - Brain Lower Grade Glioma
LIHC - Liver hepatocellular carcinoma
LUAD - Lung adenocarcinoma
LUSC - Lung squamous cell carcinoma
MESO - Mesothelioma
OV - Ovarian serous cystadenocarcinoma
PAAD - Pancreatic adenocarcinoma
PCPG - Pheochromocytoma and Paraganglioma
PRAD - Prostate adenocarcinoma
READ - Rectum adenocarcinoma
SARC - Sarcoma
SKCM - Skin Cutaneous Melanoma
STAD - Stomach adenocarcinoma
TGCT - Testicular Germ Cell Tumors
THCA - Thyroid carcinoma
THYM - Thymoma
UCEC - Uterine Corpus Endometrial Carcinoma
USC - Uterine Carcinosarcoma
UVM - Uveal Melanoma

TCGA cancer cohort abbreviations[1]

Gene	Expression Cutoff (TPM) mean	Expression Cutoff (TPM) mean + 1stdev	Sum of percentage patient coverage	Sum of TCGA expressions	Critical Organs sum of means	Stomach sum of means	Muscle sum of means	CCLE sum of means
VP1000	6.7	7.6	612.3	215.5	70.3	14.8	1.2	5455.8
VP1001	6.2	7.0	596.8	198.0	101.8	29.0	3.5	8871.8
VP1002	7.1	8.2	350.2	226.1	228.1	52.9	4.7	10861.1
VP1003	6.6	7.6	257.6	212.0	190.6	50.3	2.8	9587.7
VP1004	6.0	7.8	216.0	191.7	45.8	34.9	0.1	9063.6
VP1005	6.6	7.6	199.7	212.1	160.8	36.2	5.5	8480.4
VP1006	6.9	8.9	191.0	220.5	48.3	36.1	0.7	22083.2
VP1007	7.2	8.5	107.7	229.9	149.9	58.0	3.8	12244.1
VP1008	6.8	7.7	18.0	218.5	250.0	50.2	6.0	8689.6
VP1009	5.9	7.1	992.2	188.9	209.4	22.9	1.6	2518.8

Table 3. Sorting criteria used to determine which surface genes would be good targets. 1 and 0 were assigned to expression values depending on cutoff chosen, then the genes were sorted accordingly. Critical Organs include heart, lung, stomach, skeletal muscle, liver and brain tissues.

Conclusions

By modeling various properties of cell surface proteins to establish future development of CAB-CAR-T products it is possible to achieve 90% patient coverage with 10 distinct targets. While antigen escape is anticipated in a fraction of successfully targeted patients combinatorial algorithms may provide optimal treatment strategies. Utilizing the algorithm with different sorting criteria can help us discern percent patient coverage with various combination therapies. In the future more complex modeling will be performed to examine combination therapies where tumor heterogeneity is an important criteria for the ranked list to achieve efficacy, with a goal of maximizing complete responses (CRs) and minimize the chance of relapse.

References

1. Cerami et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery. May 2012 2; 401
2. Tang, Z. et al. (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic Acids Res, 10.1093/nar/gkx247.
3. Jespersen MC, Peters B, Nielsen M, Marcatili P. 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res (Web Server issue). 2:2.
4. Jordi Barretina, Giordano Caponigro, Nicolas Stransky, Kavitha Venkatesan, William R. Sellers, Robert Schlegel, Levi A. Garraway, et. al. 2012. The Cancer Cell Line Encyclopedia Enables Predictive Modelling of Anticancer Drug Sensitivity.Nature 483 (7391):603–7.
5. The expansion of targetable biomarkers for CAR T cell therapy Townsend et al. Journal of Experimental & Clinical Cancer Research (2018) 37:163
6. RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.
7. The Genotype-Tissue Expression (GTEx) Project
8. Robert Petryszak, et al. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants, Nucleic Acids Research, Volume 44, Issue D1, 04 January 2016, Pages D746–D752