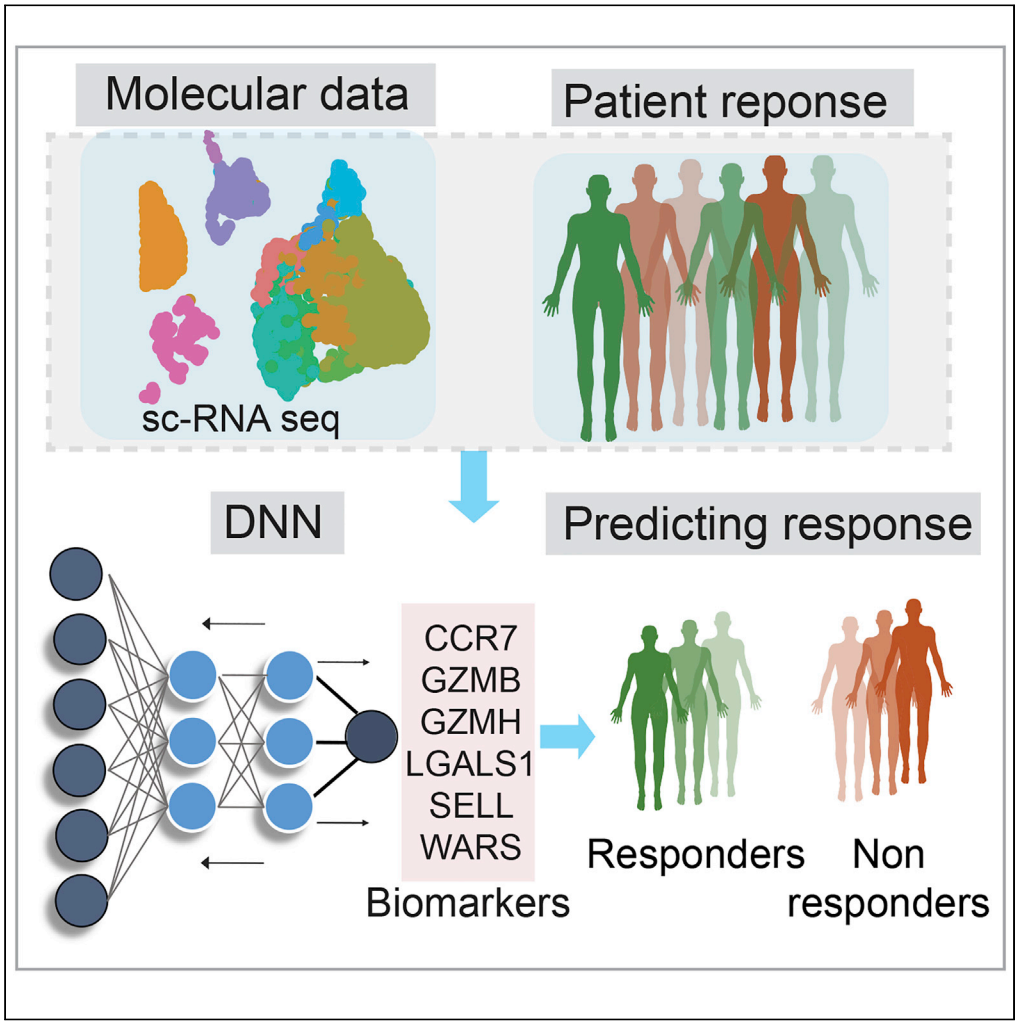
iScience

CellPress
OPEN ACCESS

Article

Deep neural network modeling identifies biomarkers of response to immune-checkpoint therapy



Yuqi Kang, Siddharth Vijay, Taranjit S. Gujral

tgujral@fredhutch.org

Highlights

Predicting biomarkers for immunotherapy response remains a challenge

DeepGeneX combines neural networks with single-cell RNAseq to predict responders

LGALS1 and WARS-expressing macrophages in nonresponders impact T cell activation

DeepGeneX enables biomarker discovery and elucidates underlying molecular mechanism

Kang et al., iScience 25, 104228
May 20, 2022
https://doi.org/10.1016/j.isci.2022.104228

iScience

CellPress
OPEN ACCESS

Article

Deep neural network modeling identifies biomarkers of response to immune-checkpoint therapy

Yuqi Kang,[1,3] Siddharth Vijay,[1,3] and Taranjit S. Gujral[1,2,4,*]

SUMMARY

Immunotherapy has shown significant promise as a treatment for cancer, such as lung cancer and melanoma. However, only 10%–30% of the patients respond to treatment with immune checkpoint blockers (ICBs), underscoring the need for biomarkers to predict response to immunotherapy. Here, we developed DeepGeneX, a computational framework that uses advanced deep neural network modeling and feature elimination to reduce single-cell RNA-seq data on ~26,000 genes to six of the most important genes (*CCR7*, *SELL*, *GZMB*, *WARS*, *GZMH*, and *LGALS1*), that accurately predict response to immunotherapy. We also discovered that the high LGALS1 and WARS-expressing macrophage population represent a biomarker for ICB therapy nonresponders, suggesting that these macrophages may be a target for improving ICB response. Taken together, DeepGeneX enables biomarker discovery and provides an understanding of the molecular basis for the model's predictions.

INTRODUCTION

In order to prevent the immune system from destroying cells indiscriminately, T cells use immune checkpoints mediated by immune checkpoint molecules. These molecules are membrane receptors classified as either stimulatory or inhibitory, depending on whether they activate or inhibit T cell response. Many tumors have strategies to evade immunosurveillance by expressing ligands for inhibitory immune checkpoint molecules, such as programmed cell death 1 (PD1) and cytotoxic T-lymphocyte-associated protein 4 (CTLA4), which prevent T cells from recognizing and destroying tumor cells. Recently, drugs that block these interactions, called immune checkpoint blockers (ICBs), have emerged as effective therapies for cancer, especially melanoma and lung cancer (Waldman et al., 2020). In comparison to conventional cancer treatments, such as chemotherapy and radiotherapy, which harm the immune system due to their untargeted (systemic) effects, ICBs were shown to be more specific and restrained, with a significant enhancement in the patients' survival (Esfahani et al., 2020; Dwary et al., 2017; Vera Aguilera et al., 2020). However, despite these benefits, clinical data highlight that ICBs are not universally effective, as only 10%–30% of patients who receive ICBs respond to treatment (Ventola, 2017). Furthermore, as these agents activate the immune response, they pose a risk for triggering a severe autoimmune response (Staff, 2019). These deficiencies highlight the need for strategies to identify patients who will respond to the ICB therapy and unravel physiological and mechanistic differences between the responders and nonresponders.

To address these challenges, numerous research efforts have been directed toward the discovery of predictive biomarkers of the positive therapeutic effect (Bai et al., 2020). Currently, parameters such as tumor mutational burden, the status of DNA damage response pathways, and tumor immune microenvironment, as well as liquid biopsy biomarkers, including circulating tumor DNA, have been investigated and/or adopted as a part of clinical practice (Bai et al., 2020). An emerging strategy for predicting ICB response and investigating the molecular basis for clinically observed differences between responders and nonresponders is single-cell RNA sequencing (scRNA-seq), i.e., single-cell transcriptomics. Previously, scRNA-seq analysis of CD8 T cell population from a large cohort of melanoma patients treated with ICBs led to the identification of *TCF7*, a gene coding for a transcription factor involved in T cell differentiation, as a biomarker that correlates with ICB therapy (Sade-Feldman et al., 2018). Although this study presented promising findings, they relied on simplistic linear correlational models. Furthermore, they focused on CD8 T cells, thus disregarding the impact of the tumor microenvironment (TME), which drives multiple

[1]Division of Human Biology, Fred Hutchinson Cancer Center, Seattle, WA, USA

[2]Department of Pharmacology, University of Washington, Seattle, WA, USA

[3]These authors contributed equally

[4]Lead contact

*Correspondence: tgujral@fredhutch.org

https://doi.org/10.1016/j.isci.2022.104228

aspects of tumorigenesis, including response to therapy (Binnewies et al., 2018). More recently, bulk RNA-seq data across 18 solid cancers from more than 7,500 patients, combined with a range of prior biological knowledge, was used to develop a machine learning model to construct systems-level signatures predictive of ICB response (Lapuente-Santana et al., 2021). However, given their complexity, systems biomarkers may be challenging to interpret and act upon in routine clinical practice.

To address the need for easy-to-use, clinically actionable biomarkers of ICB response that take into account features of the TME, we developed a new analytical framework, DeepGeneX. DeepGeneX uses sc-RNA-seq data, advanced deep neural network modeling, and features elimination steps to identify a smaller set of genes that could predict a patient's immune response to ICB therapy. DeepGeneX models outperformed linear models and identified a set of six genes that could predict the response to ICB in melanoma with 100% accuracy. We further examined the expression of these marker genes in different types of immune cells in the TME and identified two genes, *LGALS1* and *WARS*, that expressed significantly higher in macrophages of nonresponders compared with those of responders. Gene set enrichment and cell-cell interaction analysis suggest the macrophages with high expression of *LGALS1* and *WARS* ($M\varphi^{LW}$-high) are immunosuppressive and could directly promote T cell exhaustion and suppress T cell function. Thus, the $M\varphi^{LW}$-high macrophage population not only represents a biomarker for ICB therapy but targeting this population could enhance ICB therapy in nonresponders.

RESULTS

Modeling sc-RNA-seq data to identify molecular drivers of response to immunotherapy

We reasoned that the interactions between T cells and other cells in the TME could affect immune checkpoint therapy response. To model and identify molecular signatures from the broader tumor immune microenvironment (TIME) that could predict responses to immunotherapy, we used a recently published sc-RNA-seq dataset from melanoma patients treated with various immune checkpoint therapy (Sade-Feldman et al., 2018). This dataset includes 48 tissue samples (31 from nonresponders and 17 from responders) from 32 melanoma patients treated with checkpoint immunotherapy. In addition, 19 out of these 32 patients have single-cell RNA sequencing data available, including sequencing data of 5,920 immune cells collected before immunotherapy treatment. First, we analyzed the distribution of different immune cells in the stroma from responders and nonresponders and found a 2-fold higher number of CD8 T cells and a 4-fold higher number of macrophages in nonresponders than the responders (Figures 1A and 1B). In addition, most CD4 T cells, which are known to correlate with poor clinical outcomes, were also observed in higher frequency in nonresponders (Pan et al., 2020) (Figures 1A and 1B). These observations are consistent with the previous study (Sade-Feldman et al., 2018) and suggest that increase in the myeloid/macrophage population may suppress or cause exhaustion of CD8 T cells in nonresponders.

To identify molecular markers of immune checkpoint therapy response, we applied naive predictive modeling to the data from all cells in the tumor or macrophages or CDT cells. Specifically, we applied the support vector machine (SVM) and XGBoost, to distinguish the responder and nonresponder population using the immune cell gene expression data. The SVM classifies patients as responders or nonresponders based on drawing a plane to separate patients into two classes, whereas XGBoost adapts a decision-tree algorithm that separates patients with each branching and assigns a label (response or not) at the final leaf node. Our data show that SVM required the expression data from around 20 genes to accurately predict the outcome from all immune cell populations, whereas the optimal performance of the XGBoost model was approximately 0.8 (Figure 1C), suggesting a more complex and nonlinear modeling approach is warranted to predict patient outcome accurately.

Deep neural networks identify genesets that can predict patient response

Another shortcoming of XGBoost models is that they may not perform well on large datasets. Given that our data measure the activity of more than 26,000 genes, we hypothesized that a deep neural network architecture might model the large dataset better. We explored deep neural network (DNN) modeling to identify biomarkers of immune checkpoint therapy response using data from all immune cells. DNNs are nonlinear models that are analogous to neurons in the human brain (Zupan, 1994). DNNs have an input layer, output layer, and hidden layers in between connected by weighted links that capture complex relations in data. DNNs have previously been applied to biological modeling, including proteomic, genomic, and other high-throughput data (Grapov et al., 2018). We build the DNN through several stages, as conceptualized in Figure 2A.

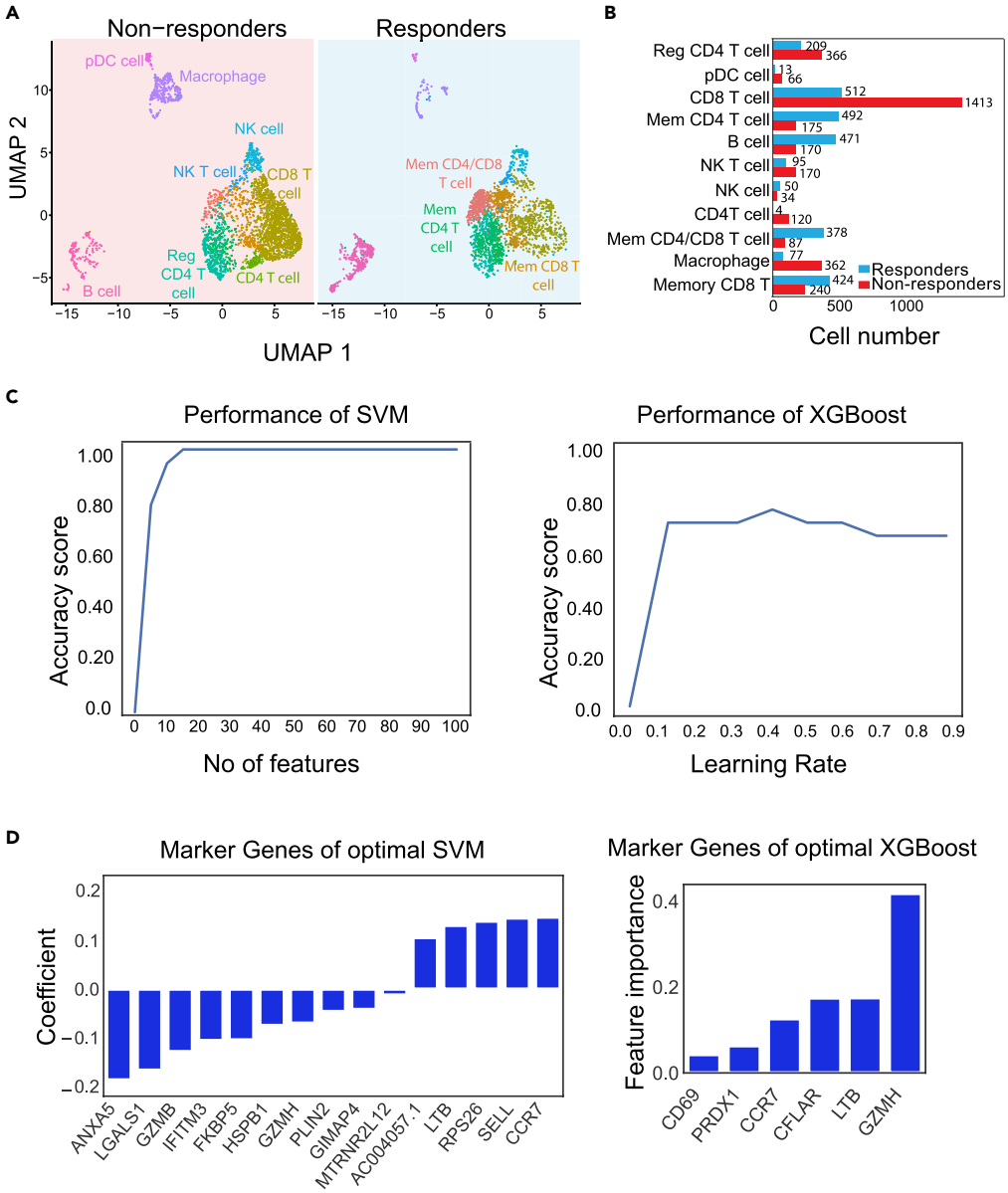


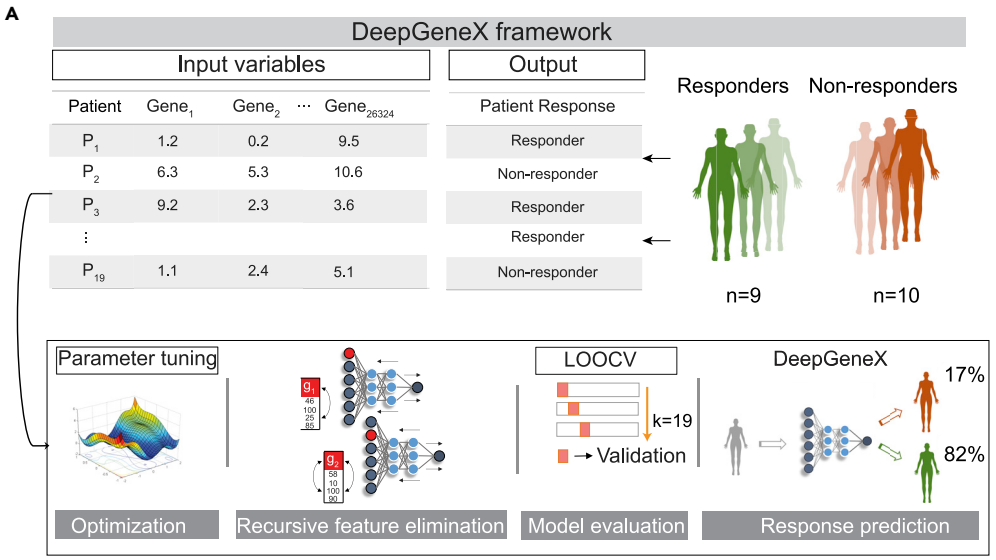
Figure 1. Modeling single-cell RNA seq data from melanoma patients for immunotherapy response

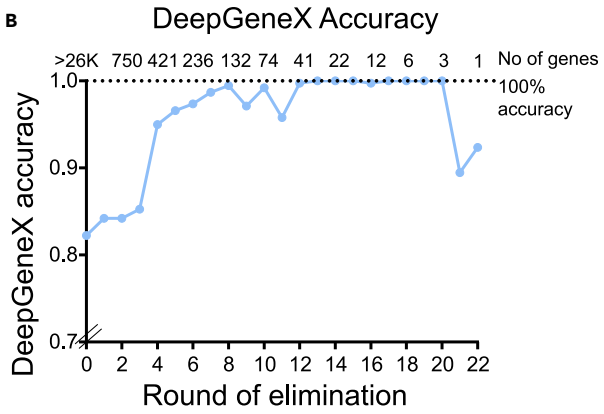
(A) The UMAP (Uniform Manifold Approximation and Projection) and bar plot (B) showing the immune cell distribution between nonresponders and responders of the checkpoint immunotherapy.

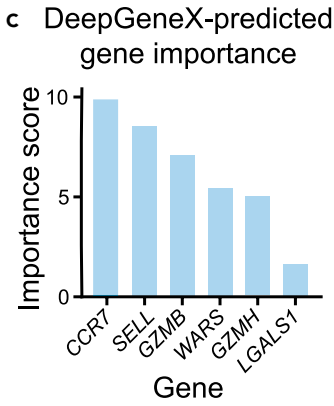
(C) The performance of SVM (support vector machine)-based model (left) and that of XGBoost model (right) using all immune cells.

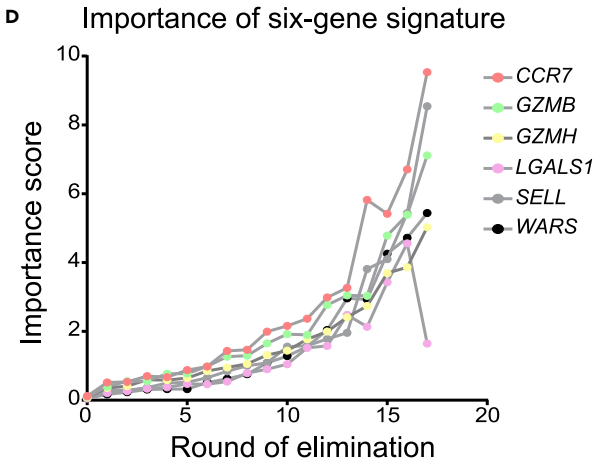
(D) The marker genes and corresponding coefficient or feature importance scores from the optimal SVM and XGBoost models.

To build DNN modeling of the sc-RNA-seq data, we first used a multi-stage Grid Search method to optimize the model hyperparameters. Next, model performance was evaluated using leave-one-out cross-validation (LOOCV) binary cross-entropy between predicted and observed responses to the immune checkpoint inhibitor. In LOOCV, we used a dataset that contains information on 19 patients in total, and at each step, data from 18 patients were used to train the model to predict the remaining patient's responses. The Grid Search method was used to search through hundreds of combinations of hyperparameters to identify a single combination with the lowest LOOCV binary cross-entropy cost. In each round of optimization, two hyperparameters were tuned: epoch and batch size, weight initializer and optimizer, and









E Performance Comparison

Model	Accuracy	Genes Used
SVM	100%	15
XGBoost	79%	6
DeepGeneX	100%	6

Figure 2. DeepGeneX identifies genesets that can predict patient response to immunotherapy

(A) A schematic illustration of DeepGeneX framework.

(B) A plot showing the LOOCV accuracy of DeepGeneX after each round of feature elimination. The number of genes used to build the model in each round is also indicated.

(C) The importance score of the top six genes predicted by DeepGeneX.

(D) A plot showing the importance score of the top six genes predicted by DeepGeneX in each round of recursive gene elimination.

(E) A table comparing the performance of three predictive models.

hidden layers and nodes per hidden layer in rounds 1, 2, and 3, respectively (Figure S1). The resulting optimized network involved two hidden layers with 100 nodes per layer, the normal weight initialization, exponential linear unit (elu) activation function, and the Adam optimizer. The model was trained for 45 epochs with a batch size of 4. The average accuracy of the model was 0.82 in LOOCV.

Next, we sought to improve the model's predictive accuracy while also identifying which of the ~26,000 genes in the model were indicative of ICB response. We employed a method we call "permutation gene importance" (PGI). First, we assigned an initial baseline error score to the DNN model based on the model's binary cross-entropy error of predicted response on the original, unchanged 19 patient training matrix. Subsequently, we shuffled each gene's activity across all 19 patients while keeping the remaining matrix of features unchanged and inputted the data into the DNN model, tracking the binary cross-entropy error after each shuffle. Each gene was assigned a "gene importance" score, which was calculated by subtracting the baseline binary cross-entropy error from the error after permuting the feature. By measuring the model's error changes, we estimated the importance of different gene's activity in contributing to a positive or negative response of the patient, with higher error changes (i.e., gene importance scores) indicating greater reliance of the model on that specific gene's activity. From this, we obtained a ranked list of the most important genes. After ranking the genes by importance score, we used the top 1,000 genes to build a new model. Having filtered the genes with only the top 1,000 genes, we then iteratively used the process of recursive gene elimination—(1) ranking genes by importance score, (2) removing the bottom 25% of genes, and (3) assessing LOOCV binary cross-entropy of the DNN model built using the remaining 75% of genes (Figure 2B).

Applying the above strategy led to significant improvement in the model's accuracy from 0.82 to 1.0 over several rounds of gene elimination (Figure 2B). We identified seven gene sets (all <10 genes) that resulted in a 100% LOOCV accuracy. Of these, a six gene set signature was chosen for further exploration. This set includes *CCR7*, *SELL*, *GZMB*, *WARS*, *GZMH*, and *LGALS1*, in order of predicted importance (Figure 2C). Overall, the process of permutation gene importance reduced the model's matrix of features from ~26,000 genes to six of the most important genes. The importance scores of these six genes in each round of elimination are shown in Figure 2D. These six genes were used to build the final DNN model we call DeepGeneX, and we compared the performance of DeepGeneX with our two baseline models (Figure 2E). Among the six genes, five of them were also among the15 marker genes identified by the SVM model (*GZMB*, *GZMH*, *SELL*, *CCR7*, *LGALS1*) and two of them were used in XGBoost (*GZMH*, *CCR7*) (Figure 1D). We further rebuilt the SVM and XGBoost models with these six genes as input and observed a 100% accuracy for SVM and an enhancement in accuracy for XGBoost, from 79% to 89%, showing the robustness of the marker gene set. The coefficients of the SVM model using the six marker genes assigned negative coefficients for *GZMB*, *GZMH*, *LGALS1*, and *WARS* and the positive coefficient for *CCR7* and *SELL*, whereas XGBoost with these six marker genes put importance on *GZMH*, *LGALS1*, *SELL*, and *CCR7*. These model performance indicators suggested that the DeepGeneX model effectively predicted whether a patient is a responder or nonresponder and that recursive gene elimination could reduce the dimensionality of the input space to identify gene signatures that were indicative of immune checkpoint response.

DeepGeneX-identified marker genes are differentially expressed in responders and nonresponders

To gain biological insights into the marker genes identified by DeepGeneX, we next analyzed the expression pattern of six marker genes in the sc-RNAseq data from responders and nonresponders. Our data show that all six genes were differentially expressed between responders and nonresponders (Figure 3A). *SELL* and *CCR7* were expressed at significantly higher levels in responders, whereas *GZMB*, *GZMH*, *LGALS1*, and *WARS* expression in responders was significantly lower (Figure 3A). Further, we also observed differential expression of these marker genes in specific immune cell types. Consistent with previous studies (Martin and Badovinac, 2018; Sade-Feldman et al., 2018), we observed the predominant expression

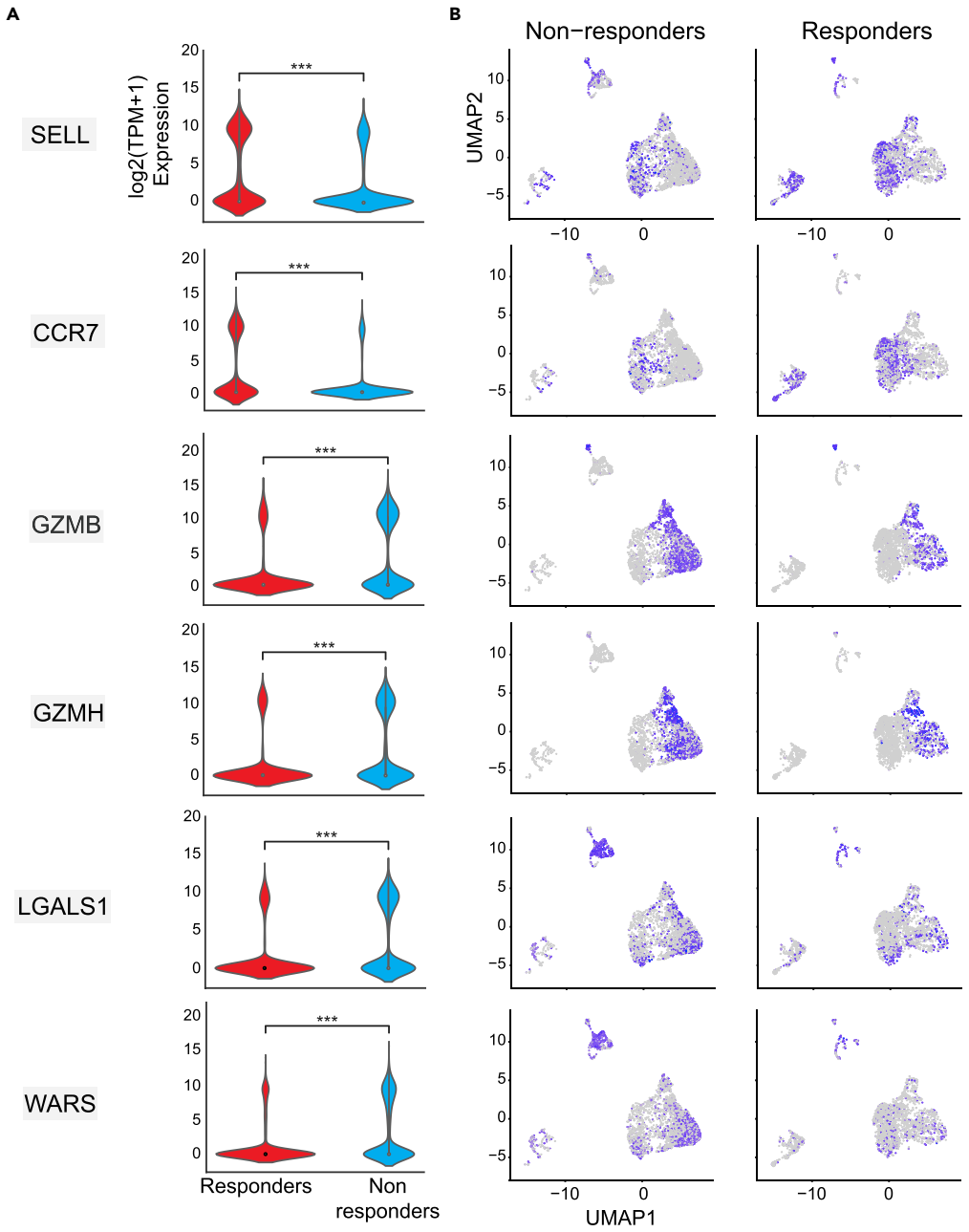


Figure 3. Expression and distribution of DeepGeneX-predicted marker genes in responders and nonresponders population

(A) Violin plots comparing the overall expression distribution across all immune cells between responders and nonresponders. ***p < 0.0005, Mann Whitney U test.

(B) The UMAPs comparing the difference in the expression of the six marker genes: *SELL*, *CCR7*, *GZMB*, *GZMH*, *LGALS1*, and *WARS* in different immune cell populations.

of *SELL* and *CCR7* in memory T cells. These genes were also expressed in a more significant proportion of memory T cells in responders compared with nonresponders. In contrast, *GZMB* and *GZMH*, known to be expressed in cytotoxic cells (Hashimoto et al., 2019), were mainly expressed in the natural killer (NK) cells and CD8 T cells of nonresponders (Figure 3A). Lastly, *LGALS1* and *WARS* were expressed primarily in the macrophages of nonresponders and showed a significant correlation (coefficient = 2.75, p value = 0.00014). Previous studies have shown that *LGALS1* plays an essential role in promoting the differentiation of M2-like

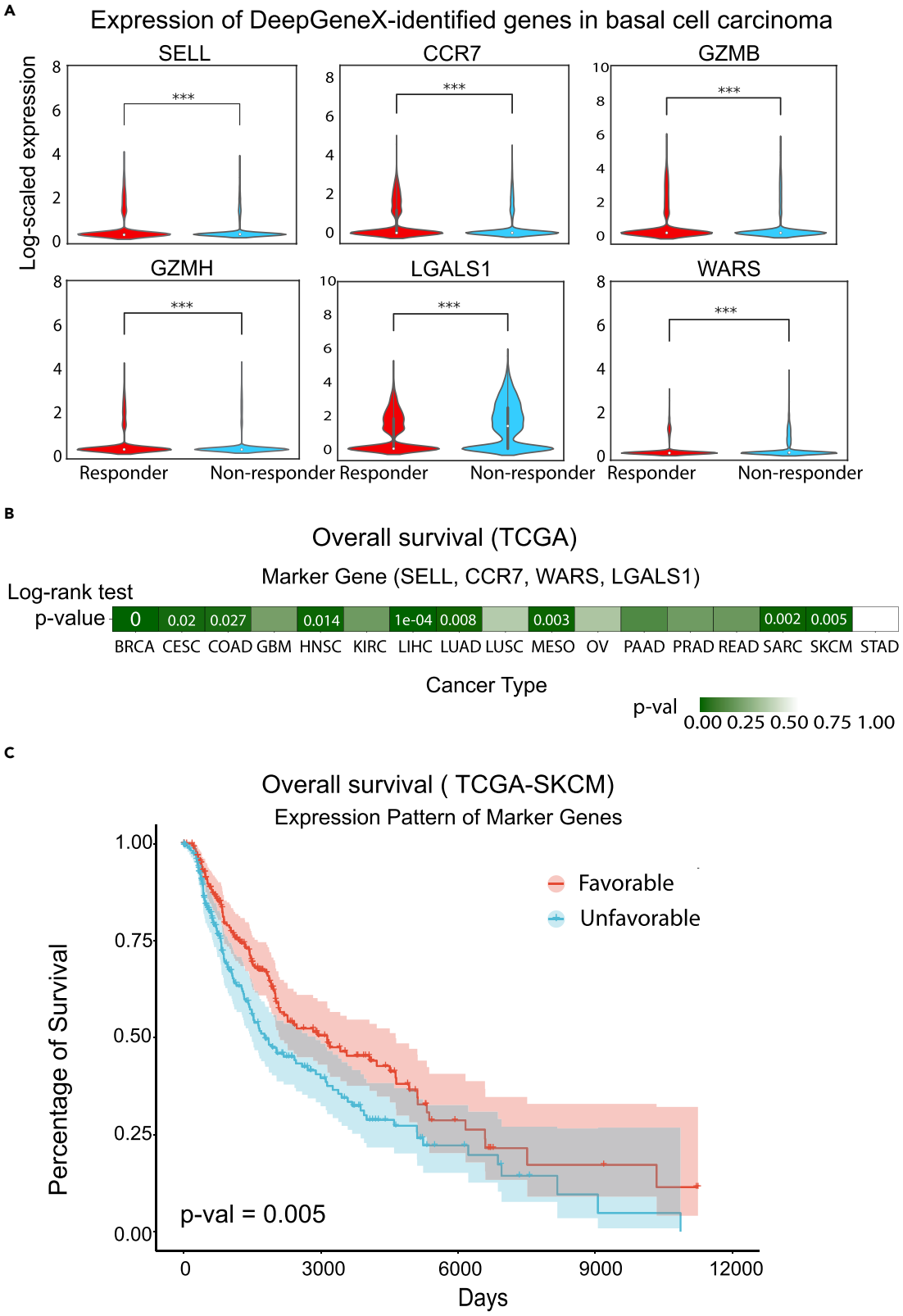

A
Expression of DeepGeneX-identified genes in basal cell carcinoma
SELL
CCR7
GZMB
GZMH
LGALS1
WARS
Log-scaled expression
***
Responder
Non-responder
B
Overall survival (TCGA)
Marker Gene (SELL, CCR7, WARS, LGALS1)
Log-rank test
p-value
0
0.02
0.027
0.014
1e-04
0.008
0.003
0.002
0.005
BRCA CESC COAD GBM HNSC KIRC LIHC LUAD LUSC MESO OV PAAD PRAD READ SARC SKCM STAD
Cancer Type
p-val
0.00 0.25 0.50 0.75 1.00
C
Overall survival (TCGA-SKCM)
Expression Pattern of Marker Genes
Favorable
Unfavorable
Percentage of Survival
p-val = 0.005
Days

Figure 4. Validation of DeepGeneX-identified marker genes in other cancers

(A) Violin plots showing the difference in expression of six marker genes between responders and nonresponders in patients with basal cell carcinoma. *p < 0.05, **p < 0.005, ***p < 0.0005, Mann Whitney U test.

(B) Log rank test results comparing the overall survival difference between patients with the favorable expression pattern of marker genes (high *SELL/CCR7* and low *LGALS1/WARS*) and patients with unfavorable expression patterns across all cancer types in the TCGA database.

(C) Kaplan-Meier survival curve comparing patients with favorable/unfavorable marker genes' expression pattern within TCGA-SKCM (melanoma) dataset, p-val < 0.005, log rank test.

macrophage and therefore driving an immunosuppressive TME (Abebayehu et al., 2017; Chen et al., 2019). Similarly, *WARS* was shown to induce the secretion of a panel of Interferons (IFN) and cytokines, which contributed to the activation of tumor-associated macrophages (TAMs) (Brown et al., 2010; Nie et al., 2019). Thus, based on the higher expression of *LGALS1* and *WARS* in the macrophages of the nonresponder population and the previously documented role of these genes, we posit that high *LGALS1*- and *WARS*-expressing TAM population may negatively impact response to immune checkpoint therapy.

Validating DeepGeneX-predicted marker genes across cancer types

To determine whether DeepGeneX-identified marker genes could predict response to immunotherapy in other cancers, we compared the differential expression between responders and nonresponders using a scRNA-seq dataset from basal cell carcinoma patients (Yost et al., 2019). This dataset includes pretreatment scRNA sequencing data from eleven basal carcinoma patients treated with anti-PD1 therapy. After removing tumor cells, we achieved a final dataset of 18,799 immune cells. We found that four out of six genes (*SELL*, *CCR7*, *LGALS1*, and *WARS*) showed a consistent trend with immunotherapy response. As predicted by DeepGeneX, we found significantly higher expression of *SELL* and *CCR7* in responders and substantially higher expression of *LGALS1* and *WARS* in nonresponders in basal cell carcinoma patients (Figure 4A). To further extend the validity and generality of DeepGeneX-defined marker genes, we then assessed the expression pattern of four genes and patients' overall survival across seventeen cancer types using the bulk RNA-seq data and clinical data from The Cancer Genome Atlas (TCGA) (Figure 4B). Our data show that patients with the favorable expression pattern of marker genes (high expression in *SELL/CCR7* and low expression in *LGALS1/WARS*) showed significantly better survival rates compared with those with unfavorable gene expression patterns (Log rank test, p value < 0.05) in nine out of seventeen cancer types, including SKCM (Skin Cutaneous Melanoma) (Figures 4B and 4C). Together, these orthogonal analyses suggest that the applicability of our DeepGeneX-identified marker gene set is not limited to skin melanoma cancer type or sc-RNA-seq data.

High *LGALS1*- and *WARS*-expressing macrophages are immunosuppressive

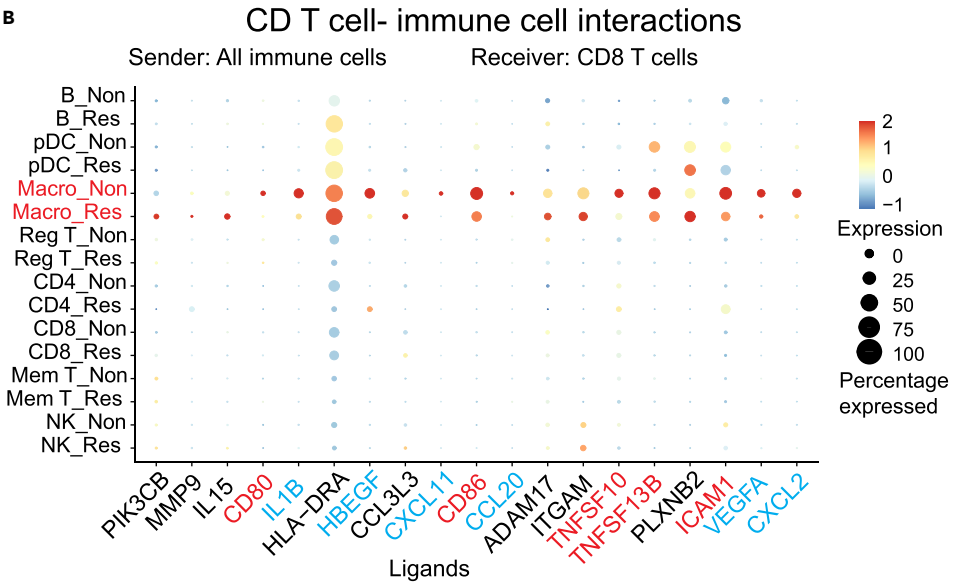
Our analysis of sc-RNA-seq from melanoma and basal cell carcinoma and bulk RNA-seq from melanoma suggests that the expression of *LGALS1* and *WARS* correlates with nonresponsiveness to immunotherapy (Figure 3) and poor overall survival (Figure 4). To determine whether high *LGALS1*- and *WARS*-expressing macrophages ($M\varphi^{LW}$-high) played a role in immunotherapy resistance, we explored the biological processes and pathways enriched in macrophages from nonresponders compared with those from responders using Gene Set Enrichment Analysis (GSEA). We then identified pathways enriched in $M\varphi^{LW}$-high population but not in $M\varphi^{LW}$-low population of nonresponders. Our data showed that $M\varphi^{LW}$-high population exhibited a significant overlap in the pathways related to the activation and polarization of macrophages and the interaction between macrophages and CD8 T cells (NES>2, FDR <0.05, Figure 5A). For example, we noted enrichment in interferon-gamma (IFN-y) signaling and response, which may indicate the polarization toward proinflammatory type macrophages (Muller et al., 2018), and upregulation of major histocompatibility complex class I (MHC-I), which could result in cytotoxic CD8 T cell activation and proliferation by processing and presenting the antigen to naive T cells (Castro et al., 2018). In addition, we found enrichment of the pattern recognition receptor signaling pathways, which include signaling receptors in *TLR* (Toll-like receptor) and *NLR* (node-like receptor) families (Figure 5A), indicating activation and differentiation of macrophages (Cen et al., 2018; Franchi et al., 2009; Singh et al., 2014). Together, these enriched pathways suggest that $M\varphi^{LW}$-high in nonresponders are fully activated and polarized. Thus, we propose that these mature macrophages could infiltrate into the TME as TAMs and induce differentiation and proliferation of naive CD8 T cells to cytotoxic CD8 T cells via MHC-I antigen presentation. Further, these data suggest that CD8 T cells in nonresponders could have already been activated by macrophages and exhausted prior anti-PD1 and anti-CTLA4 therapy. Therefore, we posit that although TAMs themselves contribute to an immunosuppressive environment, their impact on the continuous activation of CD8 T cells could partially explain the resistance to immune checkpoint therapy.

A

Pathway enrichment in MΦ LW^{-high} macrophages from non-responders

Pathways enriched in MΦ LW^{-high} macrophages	NES - LW^{-high}	NES -Nonres
Response to interferon gamma	2.62	2.59
Interferon gamma mediated signaling pathway	2.50	2.38
Antigen processing and presentation	2.39	2.31
Interleukin 1 mediated singaling pathway	2.48	2.34
Nik NFkb signaling	2.44	2.35
Pattern recognition receptor signaling pathway	2.45	2.45
Antigen processing and presentation of peptide antigen via MHC1	2.38	2.29
Fc receptor signaling pathway	2.38	2.29

B



C

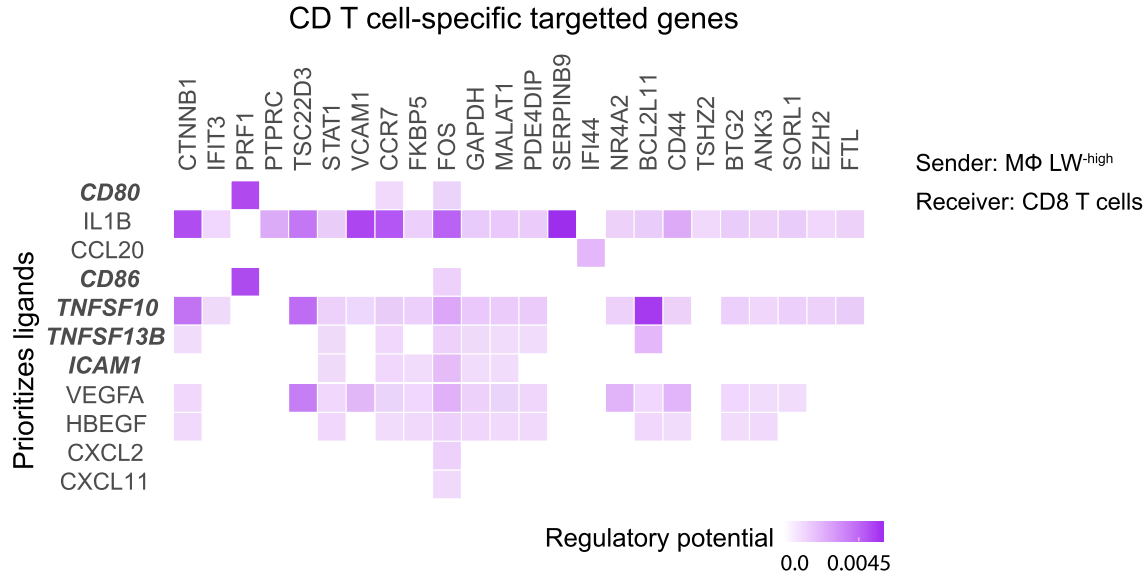


Figure 5. Pathway enrichment and cell-cell interactions of Mφ ^{LW}-high macrophages

(A) GO pathways enriched in Mφ ^{LW}-high from nonresponders compared with macrophages from responders' population. NES; normalized enrichment score.

(B) A dot plot showing the expression and distribution of ligands predominantly secreted by macrophages (in red) that could contribute to the CD8 T cell difference between responders and nonresponders.

(C) A heatmap showing the potential targeted genes in CD8 T cells in response to ligands expressed in nonresponders' macrophages. Ligands with bold italic font are differentially expressed in Mφ ^{LW}-high only.

Mφ^{LW}-high populations from nonresponders produce a set of ligands affecting CD8 T cells

We hypothesized that the Mφ^{LW}-high population is immunosuppressive and may directly inhibit the function of CD8 T cells. Specifically, ligands or secreted factors from macrophages could contribute to the difference in the function and amount of CD8 T cells between responders and nonresponders. To verify this, we applied NichNet (Browaeys et al., 2020), a method that identifies ligands secreted by sender cells that could contribute to the differential gene expression in the receiver cells. First, we designated all immune cells as sender cells and CD8 T cells as receiver cells to identify ligands expressed in other immune cells that could affect CD8 T cell function between responders and nonresponders. To exclude the impact of other immune cells from the result, we identified a list of ligands that are uniquely or dominantly expressed by macrophages (Figure 5B). To examine the possible relationship between immunotherapy response and expression of *LGALS1* and *WARS* in macrophages, we separated the macrophages from nonresponders into two subpopulations as defined previously: Mφ^{LW}-high and Mφ^{LW}-low. Next, the Mann Whitney U test was applied to identify a subset of ligands differentially expressed between Mφ ^{LW}-high and macrophages from responders. Among these ligands, *CD80*, *CD86*, *TNFSF10 (TRAIL)*, *TNFSF13B (TACI)*, and *ICAM1* were found to be upregulated in Mφ^{LW}-high, whereas *CXCL2*, *VEGFA*, *CCL20*, *CXCL11*, *HBGEF*, and *IL1B* were overexpressed in both Mφ^{LW}-high and Mφ ^{LW}-low compared with responders' macrophages (Figure 5B, Figure S2). Next, we determined macrophage-specific target genes in CD8 T cells affected by the ligands (Figure 5C). We found that the CD8 T cells from nonresponders had higher expression of *GAPDH*, *EZH2*, *VCAM1*, *PRF1*, *TSCCD3 (GILZ)*, *STAT1*, *FKBP5*, *IFIT3*, *CTNNB1*, and *BCL2L11*, whereas CD8 T cells from responders expressed higher levels of *BTG2*, *CD44*, *FOS*, *MALAT1*, and *NR4A2* (Figure 5C, Figure S2). Gratifyingly, many of the genes identified as overexpressed in CD8 T cells from nonresponders have known roles in T cell exhaustion and impaired function (Dimeloe et al., 2017; Kornberg et al., 2018; Levine et al., 2021), T cell differentiation (Kakaradov et al., 2017; Stairiker et al., 2020), and enhanced infiltration and recruitment of immune cells, including M2-macrophages. Together, these data suggest that secreted factors from Mφ^{LW}-high macrophages promote an overactivated or exhausted state in T cells. Further studies are warranted to determine the mechanistic understanding of how ligands from Mφ^{LW}-high activate markers of T cell exhaustion.

DISCUSSION

Although artificial intelligence (AI)-based approaches such as DNNs represent a promising opportunity to identify disease biomarkers, their lack of interpretability impedes their application in biological sciences. Commonly referred to as "black-box" models, it is increasingly difficult to comprehend the underlying biological mechanisms through which DNNs reach their predictions as their complexity increases. Due to this shortcoming, it is often more useful to employ simpler interpretable models irrespective of lower predictive accuracy over more complex models. In this manuscript, we developed an integrated approach by combining DNNs and recursive feature elimination, called DeepGeneX, that allows accurate prediction of outcome and reveals the underlying features important for the predictions.

We applied DeepGeneX to sc-RNA-seq data from melanoma patients and identified a set of six genes, *GZMB*, *GZMH*, *SELL*, *CCR7*, *LAGLS1*, and WARS, that could predict a patient's response to ICB therapy. This finding was validated on a sc-RNA-seq dataset from basal cell carcinoma (Yost et al., 2019). Among the six genes, we further investigated the biological impact of LGALS1 and WARS in macrophages on other cell types in the microenvironment and the effectiveness of immunotherapy. GSEA of high *LGALS1*- and *WARS*-expressing macrophages indicated a heightened activation and polarization of the macrophage population. We then applied NicheNet to examine the impact of macrophages with high expression of LGALS1 and WARS on CD8 T cells. We found ligands that mainly were or uniquely secreted by macrophages, such as VEGFA, ICAM1, PLXNB2, targeted genes in CD8 T cells, and modulated activation, differentiation, and infiltration of naive T cells. Further, our analyses of Mφ^{LW}-high/CD8 T cells revealed differentially expressed genes in CD8 T cells. For example, we found higher expression of *CD44*, *EZH2*, and

BTG2, which are known to suppress T cell function in CD8 T cells from patients with $M\varphi^{LW}$-high macrophages. CD8 T cells from patients with high expression of LGALS1 and WARS seemed to be fully activated and differentiated into effector T cells. In contrast, the CD8 T cells from the responders of ICB therapy or patients with low expression of *LGALS1* and *WARS* population overrepresented markers of quiescent T cell population and memory T cells. Because immune checkpoint therapy such as anti-PD1 and anti-CTLA4 aims to boost the immune system's potency and activate quiescent T cells, its effect could be reduced or diminished on already activated and exhausted T cells found in nonresponders. Thus, the $M\varphi^{LW}$-high macrophages-driven shift in T cell state could partially explain the differential response to ICB therapy.

The clinical response to ICB therapy is an elaborate consequence combining the interplay of several complex and multifaceted molecular mechanisms and signaling pathways in the TME, within and between cells. Current ICB therapy response prediction methods sacrifice the required complexity to develop computational models that can be interpreted. Compared with existing methods, DeepGeneX can simultaneously model highly complex relations in data driven by DNNs (known for their ability to model complex data) to predict patient outcomes and produce a set of descriptive genes that characterize nonresponders and responders. The recursive gene elimination algorithm improves neural network prediction while concurrently reducing the number of genes into a set of smaller gene signatures. Consequently, these smaller gene signatures (<10) can easily be measured in clinical or preclinical settings to predict response to ICB therapy. In addition, the ease of sc-RNA data collection allows for the rapid and straightforward collection of data required for accurate DeepGeneX prediction of response.

Overall, we present a broadly applicable, DNN-based approach called DeepGeneX that uses recursive feature elimination and significance scoring to reduce a complex dataset (~26,000 genes) into a clinically actionable biomarker gene set (6 genes) with 100% accuracy. We also demonstrate that DeepGeneX accurately models complex biological data and elucidates the underlying molecular mechanisms behind the predictions. Furthermore, DeepGeneX is a generally applicable approach that can predict the effects from any dataset and in any disease context, given a training set of measurements. In conclusion, DeepGeneX is a significant step toward a more robust machine-based strategy for predicting phenotypic and clinical response to therapeutics with a complex mechanism of action, and as such, an essential addition to the current set of methodologies in this area.

Limitations of the study

Despite the exceptional performance of the DeepGeneX models, there are a few limitations of this study. First, the power of deep neural networks can be fully realized when working with large and complex datasets. However, the melanoma dataset we used to generate the DeepGeneX model consisted of only 19 patients, a relatively smaller dataset. Secondly, the sc-RNAseq data did not contain information on cancer cells. Thus, we could not consider the impact of tumor cells on the performance of immune checkpoint therapy. Overall, the methods and approaches described in this study could be readily applied to molecular data, including sc-RNAseq data from other cancers, to identify clinically actionable gene sets for predicting response to therapy.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Single-cell (sc) RNA sequencing data analysis
  - Baseline model construction
  - DNN model construction
  - Recursive gene elimination
  - Cell interaction analysis
  - GSEA analysis
  - Survival analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.104228.

ACKNOWLEDGMENTS

This work was supported in part by the NSF under Grant No. 2047289 and Research Scholar Grant 133870-RSG-19-197-01-CDD from the American Cancer Society. We thank Drs. Marina Chan and Milka Kostic for helpful comments on the manuscript.

AUTHOR CONTRIBUTIONS

T.S.G. conceived the study. Y.K. performed the SVM and XGBoost modeling. SV performed the neural network modeling. Y.K. analyzed cell-cell interactions data. Y.K. and S.V. wrote the initial draft of the manuscript. All authors revised the manuscript.

DECLARATION OF INTERESTS

TSG and SV have filed for patent protection on the DeepGeneX approach. No other authors have any competing interests. TSG serves on the editorial board of iScience.

Received: December 28, 2021
Revised: March 18, 2022
Accepted: April 5, 2022
Published: May 20, 2022

REFERENCES

Abebayehu, D., Spence, A., Boyan, B.D., Schwartz, Z., Ryan, J.J., and McClure, M.J. (2017). Galectin-1 promotes an M2 macrophage response to polydioxanone scaffolds. J. Biomed. Mater. Res. A *105*, 2562–2571.

Bai, R., Lv, Z., Xu, D., and Cui, J. (2020). Predictive biomarkers for cancer immunotherapy with immune checkpoint inhibitors. Biomark. Res. *8*, 1–17.

Binnewies, M., Roberts, E.W., Kersten, K., Chan, V., Fearon, D.F., Merad, M., Coussens, L.M., Gabrilovich, D.I., Ostrand-Rosenberg, S., and Hedrick, C.C. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. Nat. Med. *24*, 541–550.

Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162.

Brown, J., Wallet, M.A., Krastins, B., Sarracino, D., and Goodenow, M.M. (2010). Proteome bioprofiles distinguish between M1 priming and activation states in human macrophages. J. Leukoc. Biol. *87*, 655–662.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Castro, F., Cardoso, A.P., Goncalves, R.M., Serre, K., and Oliveira, M.J. (2018). Interferon-gamma at the crossroads of tumor immune surveillance or evasion. Front. Immunol. *9*, 847.

Cen, X., Liu, S., and Cheng, K. (2018). The role of toll-like receptor in inflammation and tumor immunity. Front. Pharmacol. *9*, 878.

Chan, M., Vijay, S., McNevin, J., McElrath, M.J., Holland, E.C., and Gujral, T.S. (2021). Machine learning identifies molecular regulators and therapeutics for targeting SARS-CoV2-induced cytokine release. Mol. Syst. Biol. *17*, e10426.

Chen, T., and Guestrin, C. (2016). Xgboost: a scalable tree boosting system. In Paper Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM).

Chen, Q., Han, B., Meng, X., Duan, C., Yang, C., Wu, Z., Magafurov, D., Zhao, S., Safin, S., Jiang, C., et al. (2019). Immunogenomic analysis reveals LGALS1 contributes to the immune heterogeneity and immunosuppression in glioma. Int. J. Cancer *145*, 517–530.

Colaprico, A. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71. https://doi.org/10.1093/nar/gkv1507.

Davidson-Pilon, C. (2019). lifelines: survival analysis in Python. J. Open Source Softw. *4*, 1317.

Dimeloe, S., Burgener, A.V., Grahlert, J., and Hess, C. (2017). T-cell metabolism governing activation, proliferation and differentiation; a modular view. Immunology *150*, 35–44.

Dwary, A.D., Master, S., Patel, A., Cole, C., Mansour, R., Mills, G., Koshy, N., Peddi, P., Burton, G., Hammoud, D., et al. (2017). Excellent response to chemotherapy post immunotherapy. Oncotarget *8*, 91795–91802.

Esfahani, K., Roudaia, L., Buhlaiga, N., Del Rincon, S.V., Papneja, N., and Miller, W.H., Jr. (2020). A review of cancer immunotherapy: from the past, to the present, to the future. Curr. Oncol. *27*, S87–S97.

Franchi, L., Warner, N., Viani, K., and Nunez, G. (2009). Function of Nod-like receptors in microbial recognition and host defense. Immunol. Rev. *227*, 106–128.

Grapov, D., Fahrmann, J., Wanichthanarak, K., and Khoomrung, S. (2018). Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. OMICS *22*, 630–636.

Grossman, R.L., Heath, A.P., Ferretti, V., Varmus, H.E., Lowy, D.R., Kibbe, W.A., and Staudt, L.M. (2016). Toward a shared vision for cancer genomic data. N. Engl. J. Med. *375*, 1109–1112.

Hashimoto, K., Kouno, T., Ikawa, T., Hayatsu, N., Miyajima, Y., Yabukami, H., Terooatea, T., Sasaki, T., Suzuki, T., Valentine, M., et al. (2019). Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. Proc. Natl. Acad. Sci. U S A *116*, 24242–24251.

Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., and Scholkopf, B. (1998). Support vector machines. IEEE Intell. Syst. Appl. *13*, 18–28.

Kakaradov, B., Arsenio, J., Widjaja, C.E., He, Z., Aigner, S., Metz, P.J., Yu, B., Wehrens, E.J., Lopez, J., Kim, S.H., et al. (2017). Early transcriptional and epigenetic regulation of CD8(+) T cell differentiation revealed by single-cell RNA sequencing. Nat. Immunol. *18*, 422–432.

Kornberg, M.D., Bhargava, P., Kim, P.M., Putluri, V., Snowman, A.M., Putluri, N., Calabresi, P.A., and Snyder, S.H. (2018). Dimethyl fumarate targets GAPDH and aerobic glycolysis to modulate immunity. Science *360*, 449–453.

Lapuente-Santana, Ó., van Genderen, M., Hilbers, P.A., Finotello, F., and Eduati, F. (2021). Interpretable systems biomarkers predict

response to immune-checkpoint inhibitors. Patterns *2*, 100293.

Levine, L.S., Hiam-Galvez, K.J., Marquez, D.M., Tenvooren, I., Madden, M.Z., Contreras, D.C., Dahunsi, D.O., Irish, J.M., Oluwole, O.O., Rathmell, J.C., et al. (2021). Single-cell analysis by mass cytometry reveals metabolic states of early-activated CD8(+) T cells during the primary immune response. Immunity *54*, 829–844.e5.

Martin, M.D., and Badovinac, V.P. (2018). Defining memory CD8 T cell. Front. Immunol. *9*, 2692.

Muller, E., Speth, M., Christopoulos, P.F., Lunde, A., Avdagic, A., Oynebraten, I., and Corthay, A. (2018). Both type I and type II interferons can activate antitumor M1 macrophages when combined with TLR stimulation. Front. Immunol. *9*, 2520.

Nie, A., Sun, B., Fu, Z., and Yu, D. (2019). Roles of aminoacyl-tRNA synthetases in immune regulation and immune diseases. Cell Death Dis. *10*, 901.

Pan, H., Wang, X., Huang, W., Dai, Y., Yang, M., Liang, H., Wu, X., Zhang, L., Huang, W., Yuan, L., et al. (2020). Interferon-Induced protein 44 correlated with immune infiltration serves as a potential prognostic indicator in head and neck squamous cell carcinoma. Front. Oncol. *10*, 557157.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

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.

Sade-Feldman, M., Yizhak, K., Bjorgaard, S.L., Ray, J.P., de Boer, C.G., Jenkins, R.W., Lieb, D.J., Chen, J.H., Frederick, D.T., Barzily-Rokni, M., et al. (2018). Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Cell *175*, 998–1013.e20.

Singh, M., Khong, H., Dai, Z., Huang, X.F., Wargo, J.A., Cooper, Z.A., Vasilakos, J.P., Hwu, P., and Overwijk, W.W. (2014). Effective innate and adaptive antimelanoma immunity through localized TLR7/8 activation. J. Immunol. *193*, 4722–4731.

Staff, N. (2019). New Drugs, New Side Effects: Complications of Cancer Immunotherapy (National Cancer Institute).

Stairiker, C.J., Thomas, G.D., and Salek-Ardakani, S. (2020). EZH2 as a regulator of CD8+ T cell fate and function. Front. Immunol. *11*, 593203.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U S A *102*, 15545–15550.

Van Der Walt, S., Colbert, S.C., and Varoquaux, G. (2011). The NumPy array: a structure for efficient numerical computation. Comput. Sci. Eng. *13*, 22–30.

Ventola, C.L. (2017). Cancer immunotherapy, Part 3: challenges and future trends. P T *42*, 514–521.

Vera Aguilera, J., Paludo, J., McWilliams, R.R., Zhang, H., Li, Y., Kumar, A.B., Failing, J., Kottschade, L.A., Block, M.S., Markovic, S.N., et al. (2020). Chemo-immunotherapy combination after PD-1 inhibitor failure improves clinical outcomes in metastatic melanoma patients. Melanoma Res. *30*, 364–375.

Vijay, S., and Gujral, T.S. (2020). Non-linear deep neural network for rapid and accurate prediction of phenotypic responses to kinase inhibitors. iScience *23*, 101129.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

Waldman, A.D., Fritz, J.M., and Lenardo, M.J. (2020). A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat. Rev. Immunol. *20*, 651–668.

Waskom, M. (2021). Seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021.

Yost, K.E., Satpathy, A.T., Wells, D.K., Qi, Y., Wang, C., Kageyama, R., McNamara, K.L., Granja, J.M., Sarin, K.Y., Brown, R.A., et al. (2019). Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat. Med. *25*, 1251–1259.

Zupan, J. (1994). Introduction to artificial neural network (ANN) methods: what they are and how to use them. Acta Chim. Slov. *41*, 327.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python	Python Software Foundation	https://www.python.org/
Numpy	(Van Der Walt et al., 2011)	https://numpy.org/
Pandas	pandas – Python Data Analysis Library	https://pandas.pydata.org/
Seurat package version	R package	https://cran.r-project.org/web/packages/Seurat/index.html
SciPy	(Virtanen et al., 2020)	https://www.scipy.org/scipylib/download.html
XGBoost python package	(Chen and Guestrin, 2016)	https://github.com/dmlc/xgboost/tree/master/python-package
Python package Scikit-learn	(Pedregosa et al., 2011)	https://scikit-learn.org/stable/
Seaborn	(Waskom, 2021)	https://seaborn.pydata.org/index.html
TCGABioLink	(Colaprico, 2016)	https://bioconductor.org/packages/release/bioc/html/TCGAbiolinks.html
DeepGeneX	This paper	https://github.com/gujrallab/biomarkers

RESOURCE AVAILABILITY

Lead contact

Further information about the methods and requests for data or scripts should be directed to and will be fulfilled by the lead contact, Taranjit S Gujral (tgujral@fredhutch.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All data reported in this paper will be shared by the lead contact upon request. The source code for the SVM and XGBoost is available in the supplemental information. The source code for the DeepGeneX is available at: https://github.com/gujrallab/biomarkers. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

METHOD DETAILS

Single-cell (sc) RNA sequencing data analysis

The sc-RNA sequencing data and the corresponding patients' immunotherapy response and treatment record were achieved from the published paper (Sade-Feldman et al., 2018). The gene expression values of single cells were normalized as log2(TPM+1). Then, we applied Seurat to plot the immune cells of pre-treatment samples based on the normalized values of gene expression for each cell (Butler et al., 2018). The cell types were labeled according to the marker genes from the paper (Sade-Feldman et al., 2018). UMAPs from Seurat were plotted to show the different distribution of immune cell populations of responders and non-responders and show the differential expression of identified marker genes for predicting immune response. The Mann Whitney U test was applied to examine the statistical difference in expression of marker genes between responders and non-responders. Fisher Exact test was used to correlate the expression of two genes, where the threshold of high or low expression was defined as 2 of log2(TPM+1) value (Sade-Feldman et al., 2018). We also obtained a dataset for basal cell carcinoma and processed the data with the above workflow to validate and generalize our findings (Yost et al., 2019). All statistical tests were performed using the python package SciPy (Virtanen et al., 2020).

Baseline model construction

We constructed predictive models using gene expression values as inputs and immunotherapy responses as labels. Using the expression values of individual cells would result in overwhelming zero values. Therefore, we use

the mean values of the expression values for genes within all immune cell populations for each patient. Two baseline models were selected for the prediction of immunotherapy response: support vector machine (SVM) (Hearst et al., 1998) and XGBoost (Chen and Guestrin, 2016). SVM is a supervised predictive model that is used to classify the classes of points (vectors of numeric values, vectors of gene expression values per patients here specifically) by finding a hyperplane to separate these points in space. XGBoost, as a decision-tree based algorithm, works differently from SVM. Instead of identifying a plane, decision-tree like models construct a tree-like model that separates samples with each branching. More than a traditional decision tree model, XGB is able to adjust the existing tree models using the new input (gene expression data of patients and their response to immunotherapy) and minimize the prediction error via gradient boosting. To avoid overfitting due to the unbalance of the sample space (19 patients) and the number of features (25,000 + genes), we first excluded genes with low variance (<1). For SVM, we applied sklearn build-in recursive feature elimination to select the top 100 genes as final input features for model building (estimator as SVM, linear kernel) (Pedregosa et al., 2011). The performance of SVM using increasing percentages of the top 100 genes was evaluated by the average accuracy of leave-one-out cross validation (LOOCV). In LOOCV, each time $n - 1$ patients' scRNA data and response are used to train the model to predict the remaining patient's response. The process is repeated n times, excluding and predicting every patient. For XGBoost, due to its bagging property, we used all gene expression values as input features for self-selecting the features of importance. The parameters of XGBoost were tuned via grid search to evaluate a large number of hyperparameter combinations from a user-defined range and get the one with optimal performance, considering the following parameters: learning rate, number of estimators, and the maximum depth of the built tree. We used the following setting: learning rate = 0.41, number of estimators = 150, and maximum tree depth = 5. The performance was evaluated by the average model accuracy of LOOCV. We built the SVM model using the Python package Scikit-learn and XGBoost model with XGBoost package (Chen and Guestrin, 2016; Pedregosa et al., 2011). In the feature elimination step of SVM and parameter optimization step of XGBoost, to avoid overfitting, we only include 80% of datapoints randomly shuffled from the dataset. All performance evaluation, validation and random shuffling steps were also done via the Scikit-learn package.

DNN model construction

We developed DNN models using gene expression values as inputs and immunotherapy responses as output. As was done with XGBoost and SVM models, the mean expression values for genes for each patient were used, which eliminated hundreds of genes with 0 values. The implementation of the DNN model was achieved using the Keras and TensorFlow Deep Learning libraries as described previously (Chan et al., 2021; Vijay and Gujral, 2020). A multi-phase Grid Search method was used to optimize the DNN hyperparameters (epochs, batch size, optimizer, weight initializer, hidden layer quantity, and nodes per hidden layer). Grid Search evaluates several hundred hyperparameter combinations in order to identify the model hyperparameters that result in the lowest binary cross-entropy error between observed and predicted responses to the drug. The error function that was used to compare numerous models was LOOCV (Leave-One-Out-Cross-Validation) Binary Cross-Entropy, as it is a commonly used error function for binary classification. In LOOCV, each time $n - 1$ patients' scRNA data and response are used to train the model to predict the remaining patient's response. The process is repeated n times, excluding and predicting every patient. Binary cross entropy between predicted and observed responses is used to assign an error score to each model built with various combinations of hyperparameter values. In each phase of Grid Search, combinations of hyperparameters were evaluated, and the combination with the lowest LOOCV Binary cross-entropy was used in the subsequent phase of optimization until the final phase was reached. After the optimal hyperparameters were identified, the architecture was used to build the preliminary DeepGeneX network.

Recursive gene elimination

After the development of a trained preliminary DeepGeneX DNN model defined by f and a dataset defined by $[y, X_1, X_2, X_3, \ldots]$, the baseline error ($e_{baseline}$) can be computed, assuming a pre-defined cost function. The cost function was defined as follows:

$$C(Binary\ \ Cross\ \ Entropy) = -\frac{1}{n}\sum_{i=1}^{n} y_i * \log\hat{y}_i + (1 - y_i) * \log(1 - \hat{y}_i) \qquad \text{(Equation 1)}$$

The baseline error was calculated as defined below:

$$e_{baseline} = \frac{1}{n}\sum_{i=1}^{n} C(f(y_i, X_{1_i}, X_{2_i}, X_{3_i}, \ldots)) \qquad \text{(Equation 2)}$$

To calculate the post-permutation error ($e_{permutation}$), each feature is shuffled one-by-one for a total of 200 random shuffles. The matrix of features with a single feature permuted once can defined by $X^{permutation}$. Accordingly, the post-permutation error for an individual feature is computed as follows:

$$e_{permutation} = \frac{1}{200}\sum_{j=1}^{200}\frac{1}{n}\sum_{i=1}^{n} C\left(f\left(y, X_i^{permutation}\right)\right) \quad \text{(Equation 3)}$$

Lastly, the relative gene importance (*RGI*) or error difference for an individual feature is computed:

$$RGI = e_{permutation} - e_{baseline} \quad \text{(Equation 4)}$$

For each gene, an RGI score was computed, and the genes are ranked from highest to lowest importance. In the first round of elimination, the top 1000 genes based on RGI were selected to build the new model. Subsequently, for every following round of recursive gene elimination, the bottom 25% of genes were eliminated. Using just the top 75% of genes, a new DeepGeneX model was built, and LOOCV accuracy was used to track the model's overall relative performance across several rounds. This three-step process— (1) ranking genes by importance score, (2) removing the bottom 25% of genes, and (3) assessing LOOCV accuracy of the DeepGeneX model built using only the remaining genes is repeated until the LOOCV Accuracy of the model achieves an inflection point where the accuracy starts to decrease as the number of inputs decrease.

Cell interaction analysis

NicheNet was adopted to examine the difference in cell-cell interaction in the tumor microenvironment between responders and non-responders, especially from the aspect of how macrophages would affect CD8 T cells (Browaeys et al., 2020). By specifying the cell types of sender and receiver cells and the condition to compare with, NicheNet identified ligands of the sender cells that were likely to cause the differential gene expression in the receiver cells between two conditions: responder to immunotherapy or not in our case. Here, we first choose CD8 T cells as receivers and macrophages as senders to obtain ligands produced by macrophages that could contribute to the difference in CD8 T cells between responders and non-responders. We also examine the situation where all immune cells were considered as senders and CD8 T cells as receivers. The overlapping ligands from these two analyses identified the ligands that were uniquely or mostly secreted by macrophages that would not be masked by background – the interaction between CD8 T cell and other immune cells. It also provided the information of the possible targeted genes in CD8 T cells of these identified ligands, allowing us to link this information and the following-up pathway enrichment analysis. To examine how *LGALS1* and *WARS* expression in macrophages would contribute to the resistance to immunotherapy, we considered Mφ ^{LW}-high as macrophages with high expression of both *LGALS1* and *WARS*, while the rest as Mφ ^{LW}-low, using the expression threshold defined above. We compared the expression of identified ligands from the Mφ ^{LW}-high/Mφ ^{LW}-low population of non-responders to those from responders. Mann Whitney U test was applied to show the significant difference in ligand expression (Virtanen et al., 2020).

GSEA analysis

We applied GSEA analysis on the gene expression data of specific immune cell populations to investigate the distinction in pathway regulation between patients with different immune responses or marker gene expressions (Subramanian et al., 2005), using the GO biological process pathway dataset. We focused on differentially regulated pathways that are enriched macrophages from non-responder compared to those from responders. Pathways with a false discovery rate less than 0.05 and a normalized enrichment score of more than two were kept. To identify pathway enrichments that are dominantly impacted by Mφ ^{LW}-high, we separate the macrophages from non-responders by their *LGALS1* and *WARS* expression and compared the enriched pathways compared to macrophages from responders accordingly. We then intersected the pathways enriched in non-responders with those upregulated in Mφ ^{LW}-high, but not in Mφ ^{LW}-low to achieve a final list of pathways that are uniquely enriched in Mφ ^{LW}-high from non-responders and could contribute to the distinct immunotherapy response.

Survival analysis

The clinical data (overall survival data) and the expression data (htseq-count) of seventeen cancer types were achieved from the TCGA database, GDC portal (Grossman et al., 2016). The expression data were normalized to CPM (counts per million) value using edgeR (Robinson et al., 2010). To examine the impact

of the expression of marker genes (*SELL*, *CCR7*, *WARS and LGALS1*) on patients' survival, for each cancer types, patients were split into two groups: one with favorable gene expression pattern and the other with the unfavorable pattern. To determine the expression pattern, the expression values for four marker genes were first ranked across patients, *SELL*/*CCR7* in descending order, since DeepGeneX indicates that the higher expression of these two genes linked with immunotherapy response, while *LGALS1*/*WARS* in ascending order. The rank value of these four genes were then summed for each patient. For each cancer type, if the sum value of a patient is greater than the median, that patient is classified to have an unfavorable expression pattern, otherwise, favorable. Kaplan-Meier analysis was used to compare the survival difference between these two groups of patients and generate corresponding survival curve (Davidson-Pilon, 2019). Log rank test was used to examine the statistical significance of such difference (Virtanen et al., 2020).