Single-cell transcriptome analysis revealed a role of the transcription factor TOX in promoting CD8$^+$ T-cell exhaustion in cancer

Kyungsoo Kim[1#], Seyeon Park[2#], Ga Min Kim[3], Su Myeong Park[3], Seong Yong Park[4], Da Hee Kim[5], Young Min Park[5], Yoon Woo Koh[5], Hye Ryun Kim[3], Sang-Jun Ha[2*], and Insuk Lee[1,6*]

[1]Department of Biotechnology, College of Life Science & Biotechnology, Yonsei University, Seoul 03722, Korea

[2]Department of Biochemistry, College of Life Science & Biotechnology, Yonsei University, Seoul 03722, Korea

[3]Division of Medical Oncology, Department of Internal Medicine, Yonsei Cancer Center, Yonsei University College of Medicine, Seoul 03722, Korea

[4]Department of Thoracic and Cardiovascular Surgery, Yonsei University College of Medicine, Seoul 03722, Korea

[5]Department of Otorhinolaryngology, Yonsei University College of Medicine, Seoul 03722, Korea

[6]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul 03722, Korea

[#]These authors contributed equally.

* Correspondence to:

Insuk Lee, Department of Biotechnology, College of Life Science & Biotechnology, Yonsei University, Seoul 03722, Korea. Tel: +82-2-2123-5559 insuklee@yonsei.ac.kr

Sang-Jun Ha, Department of Biochemistry, College of Life Science & Biotechnology, Yonsei University, Seoul 03722, Korea. Tel: +82-2-2123-2696 sjha@yonsei.ac.kr

ABSTRACT

Background: Functional states of T cells are heterogeneous in tumor microenvironment. Immune checkpoint inhibitor (ICI) can reinvigorate only stem-like exhausted T cells, suggesting that impeding progress of exhaustion will improve immunotherapy efficacy. Transcription factors promoting T-cell exhaustion could be potential targets for delaying the process, improving efficacy of ICI.

Methods: Analyzing single-cell transcriptome data derived from melanoma and non-small cell lung cancer (NSCLC), we divided tumor-infiltrating $CD8^+$ T-cell population by PDCD1 (also known as PD-1) expression level into PDCD1-high cells and PDCD1-low cells and identified differentially expressed genes as candidate factors promoting intratumoral T-cell exhaustion. Co-expression of candidate genes with immune checkpoint molecules among tumor-infiltrating $CD8^+$ T cells was confirmed by single-cell trajectory analysis and flow-cytometry analysis. Loss-of-function effect of the candidate regulator was examined by cell-based knockdown assay. Clinical effect of the candidate regulator was evaluated based on overall survival and anti-PD-1 responses.

Results: We identified TOX among the most differentially expressed transcription factors between PDCD1-high subset and PDCD1-low subset of tumor-infiltrating $CD8^+$ T cells in melanoma and NSCLC. TOX expression level tend to increase as $CD8^+$ T cells become more exhausted. Flow-cytometry analysis of tumor-infiltrating T cells showed correlation of TOX expression with severity of intratumoral T-cell exhaustion. TOX knockdown resulted in downregulation of PD-1 and TIM-3, suggesting that TOX promotes intratumoral T-cell exhaustion by inducing immune checkpoint molecules. Finally, we found that TOX expression in tumor-infiltrating T cells was predictive for overall survival and anti-PD-1 efficacy in melanoma and NSCLC.

Conclusions: We found TOX to promote intratumoral $CD8^+$ T-cell exhaustion via positive regulation of PD-1—TIM-3 cooperation, suggesting that inhibition of TOX potentially impede T-cell exhaustion and improve efficacy of ICI. In addition, TOX expression in tumor-infiltrating T cells would be a useful feature for patient stratification in anti-tumor treatments including anti-PD-1 immunotherapy.

Keywords: single-cell RNA sequencing, intratumoral T-cell exhaustion, immune checkpoint molecules, anti-PD-1 immunotherapy

Running Head: TOX promotes $CD8^+$ T-cell exhaustion in tumor

BACKGROUND

Although it is generally accepted that T-cell dysfunction is a hallmark of cancers (1), the underlying molecular processes, especially alterations in transcriptional regulatory networks leading to the exhaustion of tumor-infiltrating $CD8^+$ T cells, are not completely understood. T-cell exhaustion develops progressively during chronic antigen stimulation, resulting in a heterogeneous population of exhausted T cells (2). Recent studies revealed that subsets of exhausted T cells called 'progenitor-exhausted' and 'terminally-exhausted' $CD8^+$ T cells, differ in both intrinsic effector function and reinvigoration potential by programmed cell death 1 (PD-1) blockade (reviewed in (2)). The same scenario is valid for tumor-infiltrating $CD8^+$ T cells, as both 'progenitor-exhausted' and 'terminally-exhausted' subsets with distinct epigenetic and transcriptional characteristics were also detected in the tumor microenvironment (3, 4).

Although TCF7 (also known as TCF1) was identified as a key transcription factor (TF) essential for progenitor-exhausted $CD8^+$ T cells (2), a master regulator inducing differentiation into terminally-exhausted $CD8^+$ T cells was not known. As PD-1 expression level is closely related to the severity of T-cell exhaustion, many studies have focused on PD-1 expression and its associated regulatory factors to identify a master regulator triggering exhaustion of $CD8^+$ T cells. Although the expression of certain regulators such as eomesodermin (Eomes) and Smad4 have been reported to define terminally-exhausted $CD8^+$ T cells (2, 5), whether these regulators can directly regulate the exhaustion program remains unclear. Therefore, better understanding of the mechanisms underlying establishment and maintenance of terminally-exhausted $CD8^+$ T cells will aid development of novel strategies for anti-cancer treatment.

In this study, we present a strategy for predicting genes involved in cellular differentiation based on analysis of single-cell transcriptome data, which was used to identify TFs promoting $CD8^+$ T-cell exhaustion in tumors. Using expression kinetics analysis along the pseudo-

temporally ordered $CD8^+$ T cells in tumors, we identified thymocyte selection-associated high mobility group box gene (TOX) as a key TF promoting intratumoral T-cell exhaustion. We also observed that TOX positively regulates PD-1 and TIM-3, known to cooperate for $CD8^+$ T-cell exhaustion in chronic viral infections and cancer (6, 7), and that TOX expression level in tumor-infiltrating $CD8^+$ T cells is predictive for overall survival and response to anti-PD-1 therapy. Together, these results suggest that TOX can be used for patient stratification during anti-cancer treatment, including immunotherapy, and can be targeted for combination therapy with immune checkpoint inhibitors.

METHODS

Preprocessing of single-cell transcriptome data and differential expression analysis

We analyzed single-cell transcriptome data of tumor samples derived from melanoma and non-small cell lung cancer (NSCLC), both of which were based on full-length single-cell RNA sequencing (scRNA-seq) (8, 9). These data provide gene expression profile based on transcript per million (TPM) values. For both scRNA-seq datasets, we transformed TPM values into the log normalized expression value using the formula $E_{i,j} = \log_2(\text{TPM} + 1)$ (for the i-th gene of the j-th cell). To investigate the transition of transcriptional states of $CD8^+$ T cells in the tumor microenvironment, we first isolated only T cells by considering cells annotated as "T cell" in melanoma dataset. Then, using the E value, we selected only $CD4^-$ $CD8^+$ T cells, which express at least either *CD8A* or *CD8B*, but not *CD4*. For the NSCLC dataset, we used only cells annotated as "TTC cells" (tumor cytotoxic T-cells), which are subsequently filtered for $CD4^-CD8^+$ T cells as for the melanoma dataset. We divided the $CD4^-$ CD8+ T cells into two subsets on the basis of the expression level of PDCD1 (also known as PD-1) into PDCD1-low (cells with below median E_{PDCD1}) and PDCD1-high (cells with above median E_{PDCD1}). Then, we analyzed differential expression of each gene between PDCD1-low subset and PDCD1-high subset using Wilcoxon rank sum test. Adjusted P values were calculated using the p.adjust function of the R package. For both tumor scRNA-seq datasets, we selected differentially expressed genes (DEGs) for adjusted $P < 0.05$ and then further filtered for the mean of $E > 2$ in either subset. This resulted in 179 and 92 DEGs for melanoma and NSCLC datasets, respectively (**Supplementary Table 1**).

Dimension reduction and visualization of single-cell transcriptome data

To visualize relationship among individual cells based on high-dimensional gene expression data, we used t-stochastic neighbor embedding (tSNE) (10), which is one of the most popular methods for dimension reduction. We conducted tSNE analysis using function "Rtsne" of R package with parameters of 30, 100, and 1000 for perplexity, number of PCA dimension, and the number of iterations, respectively. We projected individual cells on the first two tSNE dimensions. In addition, we used violin plots to present density distribution of cells with specific expression level of a gene of interest for the PDCD1-low and PDCD1-high subsets.

Single-cell trajectory analysis

To investigate kinetics of gene expression during $CD8^+$ T-cell differentiation in the tumor microenvironment, we performed single-cell trajectory analysis using the Monocle2 software (11). Using scRNA-seq profiles for T cells derived from melanoma (8), we reconstructed single-cell trajectories for effector, memory, and exhaustion states. To isolate only $CD8^+$ T cells, we first classified T cells based on the CD4, CD8A, and CD8B expression. Owing to the presence of cells that were not properly classified into certain cell types, we also used unsupervised clustering by Monocle2 for cell type assignment. We clustered cells using the function "clusterCells" with cell-type specific marker genes selected using the function "markerDiffTable". Then, we assigned cell types for each cluster with the function "clusterCells" based on the frequency of $CD8^+$ T cells in the cluster. We defined the three T-cell states of stable endpoint based on the expression of three marker genes (12-14). Using the Monocle2 function "classifyCells", cells were classified into cell types based on the following criteria: effector ($CD62L^-$, $CD127^-$, $PDCD1^-$), exhausted ($PDCD1^+$), memory (either $CD62L^+$ or $CD127^+$), ambiguous (classified into multiple cell types), and unknown (classified into none of the cell types). After selecting the group-specific marker genes using the function "markerDiffTable", we could pseudo-temporally order the cells using the function "reduceDimension" and "orderCells". For the visualization of expression dynamics along the trajectories, we used the BEAM analysis tools of Monocle2.

Clinical sample collection

Fresh tumor specimens were provided by the Department of Internal Medicine at the Severance Hospital with the consent to conduct the following study. We enrolled the 20 patients with NSCLC and 15 patients with head and neck squamous cell carcinoma (HNSCC) who were treated between 2017 and 2019 in Korea. Detailed information on human subjects is listed in **Table 1**. This study was approved by the Institutional Review Board (IRB) of Severance Hospital (No. 4-2016-0788). All patients signed a written informed consent for genetic analysis.

Isolation of tumor-infiltrating lymphocytes (TILs) from a primary tumor

Primary tumor tissues obtained from surgical resection of patient tumors and from tumor burdens of mice were separated, minced into 1 mm^3 pieces, and digested with a solution containing 1 mg/mL collagenase type IV (Worthington Biochemical corp., Lakewood, NJ, USA) and 0.01 mg/mL DNase I (Millipore Sigma Corp., St. Louis, MO, USA) at 37°C for 20 min. The dissociated tissues were filtered using a 40-μm cell strainer (Falcon, Corning, NY, USA) into Roswell Park Memorial Institute (RPMI)-1640 medium (Corning Inc., Corning, NY, USA). TILs were separated using a Percoll gradient (Millipore Sigma Corp.) and suspended in PBS with 2% fetal bovine serum (FBS; Biowest, Riverside, MO, USA). Thereafter, single-cell suspensions were stained with the indicated fluorescent-conjugated antibodies.

Flow cytometry analysis

Single-cell suspensions were analyzed using FACS Cytoplex (Beckman Coulter, Brea, CA, USA) after staining the following antibodies: for murine tissues, the antibodies from Biolegend were CD4-BV785 (RM4-5, Cat# 100552), CD8-Alexa Fluor 700 (53-6.7, Cat# 100730), CD25-Percp-Cy5.5 (PC61, Cat# 102030), PD-1-BV605 (29F.1A12, Cat# 135220), PD-1-BV421 (29F.1A12, Cat#135218), TIM-3-BV605 (RMT3-23, Cat#119721), and TIGIT-APC (1G9, Cat#142105); antibodies from Miltenyi Biotec: TOX-PE (REA473, Cat#130-107-

837), and REA control (I)-PE (REA293, Cat#130-104-613). Dead cells were stained using the Live/Dead Fixable Near-IR Dead cell stain kit (Invitrogen, Cat# L10119). For transcription factor staining, TILs were fixed and permeabilized using Foxp3 fixation/permeabilization solution (eBioscience, 00-5523-00). TOX and isotype control were stained after permeabilization.

Antibodies used for human sample staining were as follows: Biolegend antibodies: CD3-BV785 (OKT3, Cat#317329), CD8-BV605 (RPA-T8, Cat#301040), and PD-1-BV421 (EH12.2H7, Cat#329920); R & D systems antibodies: TIM-3-Alexa Fluor 488 (344823, Cat#FAB2365G) and TIGIT-APC (741182, Cat#FAB7898A); BD Biosciences antibody: CD4-APC-H7 (L200, Cat#560837); Miltenyi Biotec antibodies: TOX-PE (REA473, Cat#130-107-837) and REA control (I)-PE (REA293, Cat#130-104-613); Invitrogen antibodies: TOX (TXRX10, Cat#12-6502-82) and Rat IgG2a kappa isotype control (eBR2a, Cat#12-4321-80). Dead cells were excluded by staining with Live/Dead™ Fixable Red Dead cell stain kit (Invitrogen, Cat# L23102). For transcription factor staining, the cells were fixed and permeabilized using Foxp3 fixation/permeabilization solution (eBioscience, 00-5523-00), followed by TOX and isotype control staining. The cells were analyzed using the FlowJo software (Tree Star Inc., Ashland, OR, USA).

Mice

Five- to six-week-old female C57BL/6 mice and Balb/c mice were purchased from Charles River Laboratories and The Jackson Laboratory, respectively. Mice were maintained in accordance with the Laboratory Animals Welfare Act, the Guide for the Care and Use of Laboratory Animals, and the Guidelines and Policies for Rodent experiment provided by the IACUC in the Yonsei University Health System.

***In vivo* tumor models**

MC38 colon cancer cells, TC-1 lung cancer cells, or LLC1 lung cancer cells were inoculated subcutaneously (10^6 cells) in C57BL/6 mice. CT26 colon cancer cells were inoculated subcutaneously (10^6 cells) in Balb/c mice. Mice were sacrificed on day 21 after tumor

inoculation.

TOX knockdown assay

Healthy human peripheral blood mononuclear cells (PBMCs) were activated using anti-CD3 antibody (Miltenyi Biotec) in RPMI-1640 medium containing 10% fetal bovine serum (FBS; Young In Frontier Co. Ltd., Korea) for 24 h. After incubation, TOX-targeting siRNA and negative control siRNA (Invitrogen) were added to the activated cells and transfected using the Neon transfection system (Invitrogen). Following the transfection, the transfected cells were re-stimulated with the anti-CD3 antibody for 72 h. Gene knockdown was confirmed using flow cytometry analysis.

Statistical test for experimental data

Statistical significance using two-tailed unpaired Student's *t*-tests was analyzed using the Prism 5.02 software (GraphPad). Data are expressed as means ± SEM. Differences of $P < 0.05$ (*), $P < 0.01$ (**), and $P < 001$(***) were considered significant.

Survival analysis and anti-PD-1 response analysis

We conducted survival analysis with transcriptome and clinical data compiled from The Cancer Genome Atlas (TCGA) for melanoma (SKCM, skin cutaneous melanoma) and NSCLC (LUAD, lung adenocarcinoma and LUSC, lung squamous cell carcinoma). Bulk RNA-seq data for tumor samples were downloaded from the Xena database (https://xena.ucsc.edu/) and clinical data were downloaded from TCGA-CDR (15). For the survival analysis of patients with NSCLC, we used those with only top 25% tumor mutation burden. To evaluate clinical effect based on the TOX expression in only tumor-infiltrating T cells, we normalized TOX expression to the level of tumor-infiltrating T cells using the geometric mean of CD3D, CD3E, and CD3G expression. Then, we compared the overall survival rates of patients with top 30% TOX level (TOX-high) with those of all other patients (TOX-low).

For analyzing the correlation of anti-PD-1 response with different levels of TOX expression in tumor infiltrating T cells, we used bulk RNA-seq data along with clinical information for two independent cohorts of patients treated with anti-PD-1 immunotherapy (16, 17). Similar to the survival analysis, we normalized TOX expression and drew waterfall plots after dividing patients on the basis of mean TOX expression values across patients. For the Hugo et al. (16) dataset, we assigned patients annotated by "Partial Response" and "Complete Response" into 'responder' group and those by "Progressive Disease" into 'non-responder' group. For the Riaz et al. (17) dataset, we excluded patients annotated by "Stable Disease" and assigned the rest of patients into two groups similar to the Hugo et al. dataset.

RESULTS

scRNA-seq data analysis suggested association of TOX with intratumoral $CD8^+$ T-cell exhaustion

Distinct cellular states can be often depicted using single marker gene expression. Thus, we may identify key genes involved in the progression of T-cell exhaustion by analyzing differentially expressed genes (DEGs) between 'progenitor-exhausted' T cells and 'terminally-exhausted' T cells. However, exhausted $CD8^+$ T cells in the tumor microenvironment show a continuous spectrum of transcriptional states for different levels of exhaustion severity (18). Therefore, we developed a strategy for identifying genes involving T-cell exhaustion using single-cell transcriptome data (**Figure 1A**). Exhausted $CD8^+$ T cells with intermediate level of PD-1 (also known as PDCD1) expression can be reinvigorated by PD-1 blockade, whereas those with high level of PD-1 expression are refractory to this effect (19). Therefore, we divided $CD8^+$ T cells into two subsets, the PDCD1-high and PDCD1-low subtypes, using the median expression value of PDCD1. Observation of localized distribution of PDCD1-high cells in the 2-dimensional latent space of the tSNE plot confirms the capability of the PD-1 marker to segregate terminally-exhausted cells from progenitor-exhausted cells. The DEGs between the two subsets can be potential candidate genes associated with T-cell exhaustion, which might also be confirmed from the similar distribution of DEG-high cells in the same 2-dimensional latent space.

We applied the strategy to scRNA-seq profiles of $CD8^+$ T cells derived from melanoma (8) (**Figure 1B**). We observed localized distribution of PDCD1-high cells in the tSNE plot. We identified 179 DEGs between the PDCD1-high subset and PDCD1-low subset using Wilcoxon test (adjusted $P < 0.05$) (**Supplementary Table S1A**). Notably, immune checkpoint genes such as *HAVCR2* (also known as *TIM-3*), *CTLA4*, and *TIGIT* were upregulated in the PDCD1-high subset compared to in the PDCD1-low subset, and showed similar distribution of DEG-high cells to that of PDCD1-high cells. Similarly, we performed subset analysis with scRNA-seq profiles of $CD8^+$ T cells derived from NSCLC (9) (**Figure 1C**) and identified 92 DEGs (**Supplementary Table S1B**). We also observed upregulation of immune checkpoint genes in the PDCD1-high subset using the single-cell transcriptome profiles of NSCLC samples. The correlation between the expression of PDCD1 and other immune checkpoint genes validated the effectiveness of subset analysis with single-cell transcriptome data in identifying genes involved in T-cell exhaustion.

For the identification of key trans-acting regulators regulating intratumoral T-cell exhaustion, we focused on 11 and 5 TFs among the identified DEGs from melanoma and NSCLC samples, respectively. We successfully retrieved TFs previously known for T-cell exhaustion, such as BATF (20), NFATC1 (20), PRDM1 (21), and TCF7 (2), which further highlight the effectiveness of predicting regulatory factors using single-cell transcriptome data. Interestingly, among uncharacterized candidates, only TOX was identified in both melanoma and NSCLC samples, and hence we included it for follow-up functional analysis.

TOX transcript level tended to increase during progression of $CD8^+$ T-cell exhaustion in tumors

Tumor-infiltrating T cells that are initially in the effector state (T_{eff}) soon start becoming dysfunctional and are converted to exhausted T cells (T_{exh}) due to the highly immunosuppressive tumor microenvironment. A subset of persisting T_{eff} cells differentiate into long-lived and self-renewable memory T cells (T_{mem}). We hypothesized that if TOX promotes T-cell exhaustion, its expression dynamics during transition from T_{eff} to T_{exh} should differ from that during the transition from T_{eff} to T_{mem}. To test this hypothesis, we reconstructed single-cell trajectories composed of pseudo-temporally ordered $CD8^+$ T cells

across the three distinct T-cell states using the software Monocle2 (11). We observed that each of the three single-cell trajectories was enriched for the corresponding cell type assigned by Monocle2 (**Figure 2A**). Thus, we classified tumor-infiltrating $CD8^+$ T cells into three T-cell states based on their locations in the single-cell trajectory model (**Figure 2B**).

As expected, we observed upward trend in the expression of immune checkpoint genes such as *CTLA4*, *HAVCR2*, *PDCD1*, and *TIGIT* along the pseudotime of $CD8^+$ T-cell exhaustion, whereas their expression tended to decrease along the pseudotime trajectory to memory state (**Figure 2C-E**). Notably, we observed the same trend of TOX expression changes along the pseudotime trajectories (**Figure 2C-E**), which indicated correlation in the expression of TOX and immune checkpoint genes. Considering that expression level of immune checkpoint molecules correlate with severity of $CD8^+$ T-cell exhaustion, these results also suggest association of TOX expression with severity of $CD8^+$ T-cell exhaustion in tumors.

TOX protein level correlated with severity of intratumoral $CD8^+$ T-cell exhaustion

To test whether expression of TOX protein is associated with severity of intratumoral T-cell exhaustion, we conducted flow cytometry analysis for tumor-infiltrating lymphocytes (TILs) isolated from human primary tumor specimens from 20 patients with NSCLC and 15 patients with head and neck squamous cell carcinoma (HNSCC) who underwent surgical resection at the Severance Hospital (**Table 1**). We observed correlation between TOX expression and PD-1, TIM-3, and TIGIT expression at the protein level (**Figure 3A**). In addition, the proportion of TOX^+ TILs was significantly associated with expression of the immune checkpoint molecules in both NSCLC and HNSCC (**Figure 3B**). Considering that PD-1^+TIM-3^+ TILs exhibit the most severely exhausted phenotype in both chronic viral infection and tumor (6, 22), we sub-gated TILs into PD-1^-TIM-3^-, PD-1^+TIM-3^-, and PD-1^+TIM-3^+ cells and compared their distribution of TOX levels. We observed that TOX level was significantly higher in more severely exhausted TILs, which can be arranged in the following order of decreasing TOX expression: PD-1^+TIM-3^+ > PD-1^+TIM-3^- > PD-1^-TIM-3^- (**Figure 3C**). Flow cytometry analysis for mouse $CD8^+$ TILs isolated from various cancer models, including MC38 colon cancer, CT26 colon cancer, TC-1 lung cancer, and LLC1 lung cancer, showed similar relationship between TOX and exhaustion severity of TILs (**Figure 4**). Taken

together, these results strongly suggest that TOX expression is associated with the severity of intratumoral $CD8^+$ T-cell exhaustion.

TOX knockdown disrupts PD-1 and TIM-3 expression in $CD8^+$ T cells

Considering the positive correlation of TOX expression with severity of $CD8^+$ T-cell exhaustion, we hypothesized TOX as a positive regulator of the exhaustion process. Therefore, we evaluated loss-of-function effect of TOX on immune checkpoint molecules in $CD8^+$ T cells. We first activated healthy human peripheral blood mononuclear cells (PBMCs) with an antibody against CD3 for 24 h, and then transfected them with siRNA targeting *TOX*. Interestingly, when *TOX* was knocked down in activated $CD8^+$ T cells, PD-1 and TIM-3, but not TIGIT, were significantly downregulated (**Figure 5**), indicating that TOX positively regulates the expression of both PD-1 and TIM-3 to promote $CD8^+$ T-cell exhaustion. PD-1 and TIM-3 cooperate for $CD8^+$ T-cell exhaustion in chronic viral infection and cancer (6, 7). Overall, these observations may suggest that TOX is a key regulator of differentiation into terminally exhausted $CD8^+$ T cells, which acts by inducing both PD-1 and TIM-3.

TOX expression in tumor-infiltrating T cells is predictive for overall survival and anti-PD-1 efficacy

As T cells are major players in eliminating cancer cells in tumor, their functional state affects prognosis and therapeutic efficacy. TOX exhibited a highly specific expression pattern in T cells (**Figure 6A**). Considering the association of TOX expression with severity of intratumoral $CD8^+$ T-cell exhaustion, we hypothesized that TOX expression in tumor-infiltrating T cells may be used as a clinical indicator during anti-cancer treatment. To test this hypothesis, we interrogated TCGA survival data with respect to TOX expression after normalizing the level of tumor-infiltrating T cells by geometric mean of CD3D, CD3E, and CD3G expression as described in a previous study (23). We observed that lower expression of TOX in T cells was associated with enhanced overall survival rate (P = 0.0022, log-rank test) for TCGA melanoma cohort (SKCM, skin cutaneous melanoma) (**Figure 6B**). Similarly, survival analysis for TCGA NSCLC cohort (LUAD, lung adenocarcinoma and LUSC, lung

squamous cell carcinoma) showed enhanced overall survival rate with lower expression of TOX in T cells (P = 0.0393, log-rank test) (**Figure 6C**). These results suggest that TOX expression of tumor-infiltrating T cells can be used to predict the overall survival of cancer patients.

Next, we assessed whether TOX expression of tumor-infiltrating T cells can predict response to anti-PD-1 immunotherapy by interrogating two previously published independent cohorts treated with anti-PD-1 therapy with available bulk RNA-seq data and clinical information (16, 17). Interestingly, we observed that higher proportion of patients with lower TOX expression in tumor-infiltrating T cells respond to anti-PD-1 immunotherapy in both cohorts (**Figure 6D**), indicating that TOX expression of tumor-infiltrating T cells is predictive for anti-PD-1 efficacy. Taken together, these results suggest that TOX expression in tumor-infiltrating T cells can be used for patient stratification in anti-cancer treatment, including anti-PD-1 immunotherapy.

DISCUSSION

In this study, we identified TOX as a regulatory factor promoting intratumoral $CD8^+$ T-cell exhaustion by analyzing scRNA-seq data with a method mimicking subset analysis of flow or mass cytometry data. Contrary to conventional cytometry that can quantify maximum ~50 pre-selected proteins, scRNA-seq enables genome-wide expression analysis at the transcript level. Our unbiased search successfully retrieved not only immune checkpoint genes, but also TFs known for T-cell exhaustion such as BATF, NFATC1, PRDM1, and TCF7. These results demonstrated the effectiveness of subset analysis for scRNA-seq data in identifying regulatory molecules mediating cellular transitions across a continuous spectrum of transcriptional states. The proposed method may be applied for studying cellular transition during the progression of various diseases with appropriate marker genes.

We observed that TOX positively regulates expression of PD-1 and TIM-3, but not that of TIGIT, and that TOX expression is associated with severity of intratumoral $CD8^+$ T-cell exhaustion in cancer. PD-1 and TIM-3 are known to cooperate for $CD8^+$ T-cell exhaustion in chronic viral infection (6) and cancer (7), and correlation of their co-expression in $CD8^+$ T cells with poor clinical outcome was recently reported in renal cell carcinoma (24). Taken

together, we hypothesized that TOX positively regulates co-expression of PD-1 and TIM-3, which subsequently results in transition into terminally-exhausted $CD8^+$ T cells and non-responsiveness of tumors to immune checkpoint inhibitors. Therefore, inhibition of TOX may potentially impede the cooperative process toward generation of terminally-exhausted T cells, consequently enhancing chances of reinvigorating progenitor-exhausted T cells.

ACKNOWLEDGMENTS

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2018M3C9A5064709, NRF-2018R1A5A2025079 to I.L.; 2017R1A5A1014560, 2018M3A9H3024850, 2018R1A2A1A05076997 to S.-J.H.).

REFERENCES

1. Wherry EJ, Ha S-J, Kaech SM, Haining WN, Sarkar S, Kalia V, et al. Molecular signature of CD8+ T cell exhaustion during chronic viral infection. Immunity. 2007;27(4):670-84.

2. Pauken KE, Wherry EJ. Overcoming T cell exhaustion in infection and cancer. Trends in immunology. 2015;36(4):265-76.

3. Miller BC, Sen DR, Al Abosy R, Bi K, Virkud YV, LaFleur MW, et al. Subsets of exhausted CD8(+) T cells differentially mediate tumor control and respond to checkpoint blockade. Nature immunology. 2019;20(3):326-36.

4. Kurtulus S, Madi A, Escobar G, Klapholz M, Nyman J, Christian E, et al. Checkpoint Blockade Immunotherapy Induces Dynamic Changes in PD-1(-)CD8(+) Tumor-Infiltrating T Cells. Immunity. 2019;50(1):181-94 e6.

5. Huang AC, Postow MA, Orlowski RJ, Mick R, Bengsch B, Manne S, et al. T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. Nature. 2017;545(7652):60-5.

6. Jin H-T, Anderson AC, Tan WG, West EE, Ha S-J, Araki K, et al. Cooperation of Tim-3 and PD-1 in CD8 T-cell exhaustion during chronic viral infection. Proceedings of the National Academy of Sciences. 2010;107(33):14733-8.

7. Zhou Q, Munger ME, Veenstra RG, Weigel BJ, Hirashima M, Munn DH, et al. Coexpression of Tim-3 and PD-1 identifies a CD8+ T-cell exhaustion phenotype in mice with disseminated acute myelogenous leukemia. Blood. 2011;117(17):4501-10.

8. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016;352(6282):189-96.

9. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nature Medicine. 2018;24(7):978.

10. van der Maaten H, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;9:2579–605.

11. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner H, et al. Reversed graph embedding resolves complex single-cell developmental trajectories. bioRxiv. 2017.

12. Wherry EJ. T cell exhaustion. Nature immunology. 2011;12(6):492.

13. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nature Reviews Immunology. 2015;15(8):486.

14. Fuertes Marraco SA, Neubert NJ, Verdeil G, Speiser DE. Inhibitory receptors beyond T cell exhaustion. Frontiers in immunology. 2015;6:310.

15. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell. 2018;173(2):400-16. e11.

16. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell. 2016;165(1):35-44.

17. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. Cell. 2017;171(4):934-49. e15.

18. Hashimoto M, Kamphorst AO, Im SJ, Kissick HT, Pillai RN, Ramalingam SS, et al. CD8 T Cell Exhaustion in Chronic Infection and Cancer: Opportunities for Interventions. Annual review of medicine. 2018;69:301-18.

19. Blackburn SD, Shin H, Freeman GJ, Wherry EJ. Selective expansion of a subset of exhausted CD8 T cells by alphaPD-L1 blockade. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(39):15016-21.

20. Man K, Gabriel SS, Liao Y, Gloury R, Preston S, Henstridge DC, et al. Transcription Factor IRF4 Promotes CD8(+) T Cell Exhaustion and Limits the Development of Memory-like T Cells during Chronic Infection. Immunity. 2017;47(6):1129-41 e5.

21. Chihara N, Madi A, Kondo T, Zhang H, Acharya N, Singer M, et al. Induction and transcriptional regulation of the co-inhibitory gene module in T cells. Nature. 2018;558(7710):454-9.

22. Sakuishi K, Apetoh L, Sullivan JM, Blazar BR, Kuchroo VK, Anderson AC. Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. J Exp Med. 2010;207(10):2187-94.

23. Zheng C, Zheng L, Yoo J-K, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell. 2017;169(7):1342-56. e16.

24. Granier C, Dariane C, Combe P, Verkarre V, Urien S, Badoual C, et al. Tim-3 Expression on Tumor-Infiltrating PD-1(+)CD8(+) T Cells Correlates with Poor Clinical Outcome in Renal Cell Carcinoma. Cancer Res. 2017;77(5):1075-82.

Table 1. Baseline characteristics of clinical samples

Characteristics	Cancer type	
	Head and neck cancer	Non-small cell lung cancer
Number of patients	15	20
Age in years (median, range)	60.4 (37–76)	62.4 (44–75)
Sex		
Male	12 (80.0%)	13 (65.0%)
Female	3 (20.0%)	7 (35.0%)
Primary site		
Oropharyngeal cancer	7 (46.7%)	NA
Hypopharyngeal cancer	2 (13.3%)	NA
Oral cavity cancer	6 (40.0%)	NA
Lung cancer	NA	20 (100%)
Histology		
Adenocarcinoma	0 (0%)	15 (75.0%)
Squamous cell carcinoma	15 (100%)	5 (25.0%)
Mutational status		
EGFR mutation	NA	7 (35.0%)
HPV status		
Positive	6 (40.0%)	NA
Negative	9 (60.0%)	NA
Stage		
I	1 (6.7%)	11 (55.0%)
II	5 (33.3%)	8 (40.0%)
III	3 (20.0%)	1 (5.0%)
IV	6 (40.0%)	0 (0%)

FIGURE LEGENDS

Figure 1. (A) Overview of identifying candidate genes associated with T-cell exhaustion using single-cell transcriptome profiles of tumor-infiltrating $CD8^+$ T cells. (B-C) Correlation of the expression levels of immune checkpoint genes and TOX with PDCD1, which is a marker for exhaustion state in (B) melanoma and (C) NSCLC. Gene expression values are based on $\log_2$(TPM+1). Individual cells that express a gene of interest at values higher than the median value are indicated in red in the tSNE plots. Distributions for single-cell expression of a gene of interest for PDCD1-low subset and PDCD1-high subset are summarized as violin plots, and their difference was tested using Wilcoxon rank sum test (**, adjusted $P < 0.05$).

Figure 2. (A) Single-cell trajectories across three distinct states of $CD8^+$ T cells. Cells were classified into cell types using Monocle2 based on the following criteria: effector ($CD62L^-$, $CD127^-$, $PDCD1^-$), exhausted ($PDCD1^+$), memory (either $CD62L^+$ or $CD127^+$), ambiguous (classified into multiple cell types), and unknown (classified into none of the cell types) (B) Based on enriched cell type, cells were classified into three states of the $CD8^+$ T cell: effector, exhausted, and memory state. (C-E) Expression dynamics of immune checkpoint genes and *TOX* along the pseudotime of $CD8^+$ T cells in two alternative trajectories from the effector state to the memory state or to the exhausted state were summarized using BEAM analysis (C) and scatter plots with regression curves (D for the trajectory of exhausted state and E for the trajectory of memory state).

Figure 3. Correlation of TOX expression with severity of human TIL $CD8^+$ T-cell exhaustion

(A-C) Flow cytometry analysis for TIL $CD8^+$ T cells isolated from human NSCLC (n = 20) and HNSCC (n = 15). (A) Representative flow cytometry plots showing co-expression of TOX and immune checkpoint molecules, namely, PD-1, TIM-3, and TIGIT, in TIL $CD8^+$ T cells. (B) Percentage of TOX^+ cells in TIL $CD8^+$ T cells for each pair of subpopulations derived from the same tumor but either expressing/not expressing a specific immune checkpoint molecule. (C) TOX protein levels in three subsets of TIL $CD8^+$ T cells for different severity of exhaustion: $PD\text{-}1^-TIM\text{-}3^-$ (orange), $PD\text{-}1^+TIM\text{-}3^-$ (blue), and $PD\text{-}1^+TIM\text{-}3^+$ (red). Histogram represents percentage distribution of cells that express a specific level of

TOX. The mean fluorescence intensity (MFI) for TOX in each subset is indicated in parenthesis. A dashed line represents the boundary separating the expression of TOX protein. Distribution of TOX-expressing TIL $CD8^+$ T cells across patients are summarized in grouped scattered plots. ns, not significant, *P < 0.05, **P < 0.01, ***P < 0.001. All statistical analysis was performed with unpaired Student's *t*-test.

Figure 4. Correlation of TOX expression with severity of mouse TIL $CD8^+$ T-cell exhaustion

(A-C) Flow cytometry analysis for TIL $CD8^+$ T cells isolated from tumor tissue of mice bearing MC38, CT26, TC1, or LLC1 tumor. (A) Representative flow cytometry plot showing co-expression of TOX and PD-1 (left) and percentage of TOX^+ cells in TIL $CD8^+$ T cells from MC38 tumor samples (n = 5) for each pair of subpopulations derived from same tumor samples with or without PD-1 expression (right). (B) Percentage of TOX^+ cells in TIL $CD8^+$ T cells from mouse tumors for each pair of subpopulations derived from same tumor samples with or without PD-1 expression in CT26 (n = 3), TC1 (n = 3), and LLC1 (n = 3). (C) TOX protein levels in three subsets of TIL $CD8^+$ T cells from MC38 tumor for different severity of exhaustion: PD-1^-TIM-3^- (orange), PD-1^+TIM-3^- (blue), and PD-1^+TIM-3^+ (red). Histogram represents percentage distribution of cells that express specific level of TOX. The mean fluorescence intensity (MFI) for TOX in each subset is indicated in parenthesis. A dashed line represents the boundary separating the expression of TOX protein. The distribution of TOX-expressing TIL $CD8^+$ T cells across MC38 tumor samples is shown in grouped scattered plots. (D) The distribution of TOX-expressing TIL $CD8^+$ T cells across tumor samples from CT26 (n = 3), TC1 (n = 3), and LLC1 (n = 3) is summarized in grouped scattered plots. ns, not significant; *, P < 0.05; **, P < 0.01; ***, P < 0.001. All statistical analysis was performed with unpaired Student's *t*-test.

Figure 5. TOX-dependent upregulation of PD-1 and TIM-3 during human $CD8^+$ T-cell activation

Expression level of immune checkpoint molecules after *TOX* mRNA knockdown. Human PBMCs were treated with control siRNA or TOX siRNA during activation and TOX expression, along with those of immune checkpoint molecules, are represented in the plot. Percentages of cells positive for the indicated proteins are shown as mean + SEM for triplicates in the bar graph. The data are representative of three to four independent

experiments. ns, not significant, *$P < 0.05$, **$P < 0.01$. All statistical analysis was performed with unpaired Student's *t*-test.

Figure 6. (A) Violin plots to depict distribution of TOX expression levels for three groups of cells derived from melanoma: T cells, other immune cells, cancer cells. (B) Overall survival analysis for TCGA patients with subcutaneous melanoma (SKCM). (C) Overall survival analysis for TCGA patients with non-small cell lung cancer (NSCLC) (with only top 25% tumor mutation burden). Patients were classified into high-TOX for those with top 30% TOX level and low-TOX for the rest. (D) Waterfall plot to depict response to anti-PD-1 immunotherapy based on two independent cohorts. Baseline represents median level of TOX expression normalized to the level of tumor-infiltrating T cells.

Figure 1

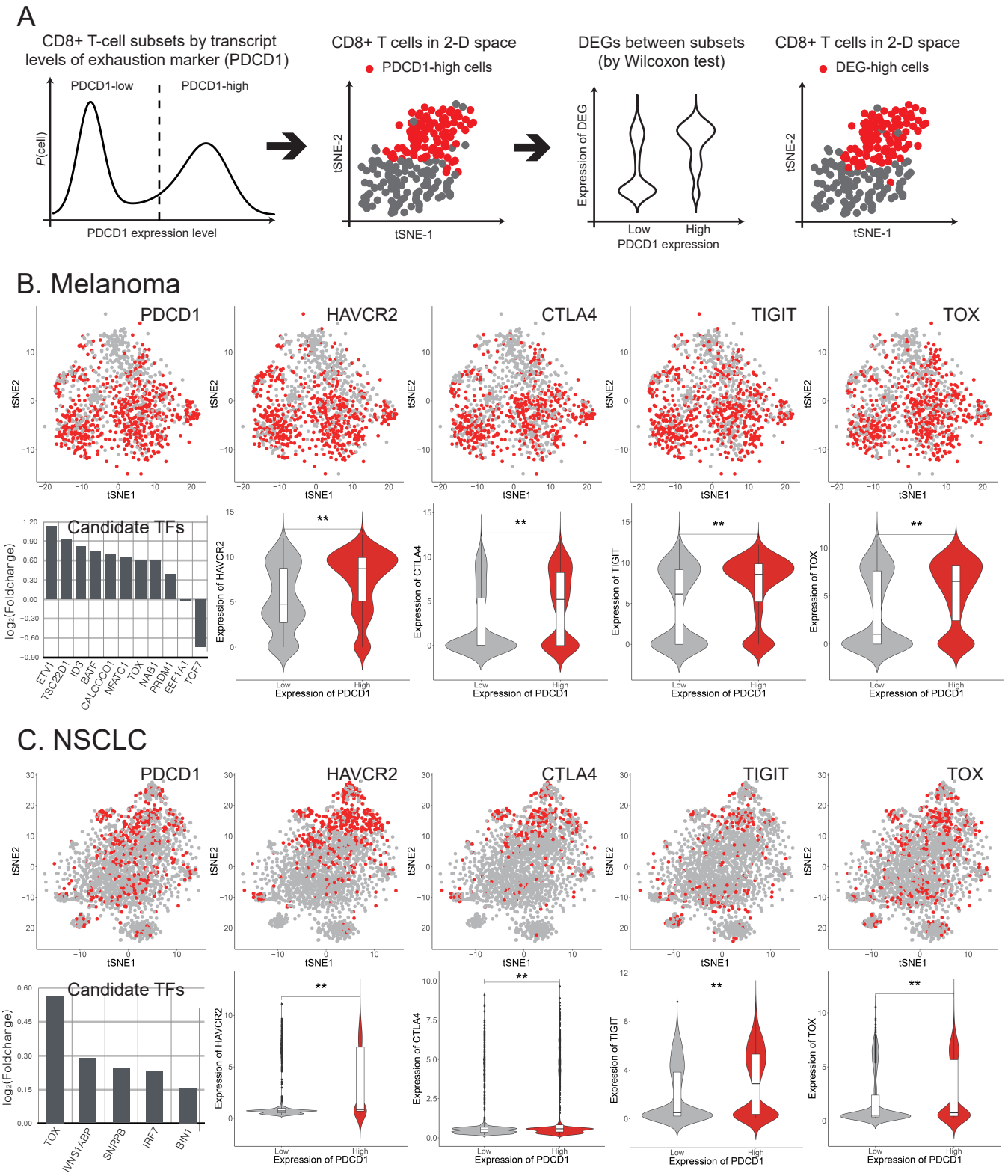

A
CD8+ T-cell subsets by transcript levels of exhaustion marker (PDCD1)
PDCD1-low
PDCD1-high
P(cell)
PDCD1 expression level
CD8+ T cells in 2-D space
PDCD1-high cells
tSNE-2
tSNE-1
DEGs between subsets (by Wilcoxon test)
Expression of DEG
Low
High
PDCD1 expression
CD8+ T cells in 2-D space
DEG-high cells
tSNE-2
tSNE-1
B. Melanoma
PDCD1
HAVCR2
CTLA4
TIGIT
TOX
tSNE2
tSNE1
Candidate TFs
log2(Foldchange)
ETV1
TSC22D1
ID3
BATF
CALCOCO1
NFATC1
TOX
NAB1
PRDM1
EEF1A1
TCF7
Expression of HAVCR2
Expression of CTLA4
Expression of TIGIT
Expression of TOX
Low
High
Expression of PDCD1
C. NSCLC
PDCD1
HAVCR2
CTLA4
TIGIT
TOX
tSNE2
tSNE1
Candidate TFs
log2(Foldchange)
TOX
IVNS1ABP
SNRPB
IRF7
BIN1
Expression of HAVCR2
Expression of CTLA4
Expression of TIGIT
Expression of TOX
Low
High
Expression of PDCD1

Figure 2

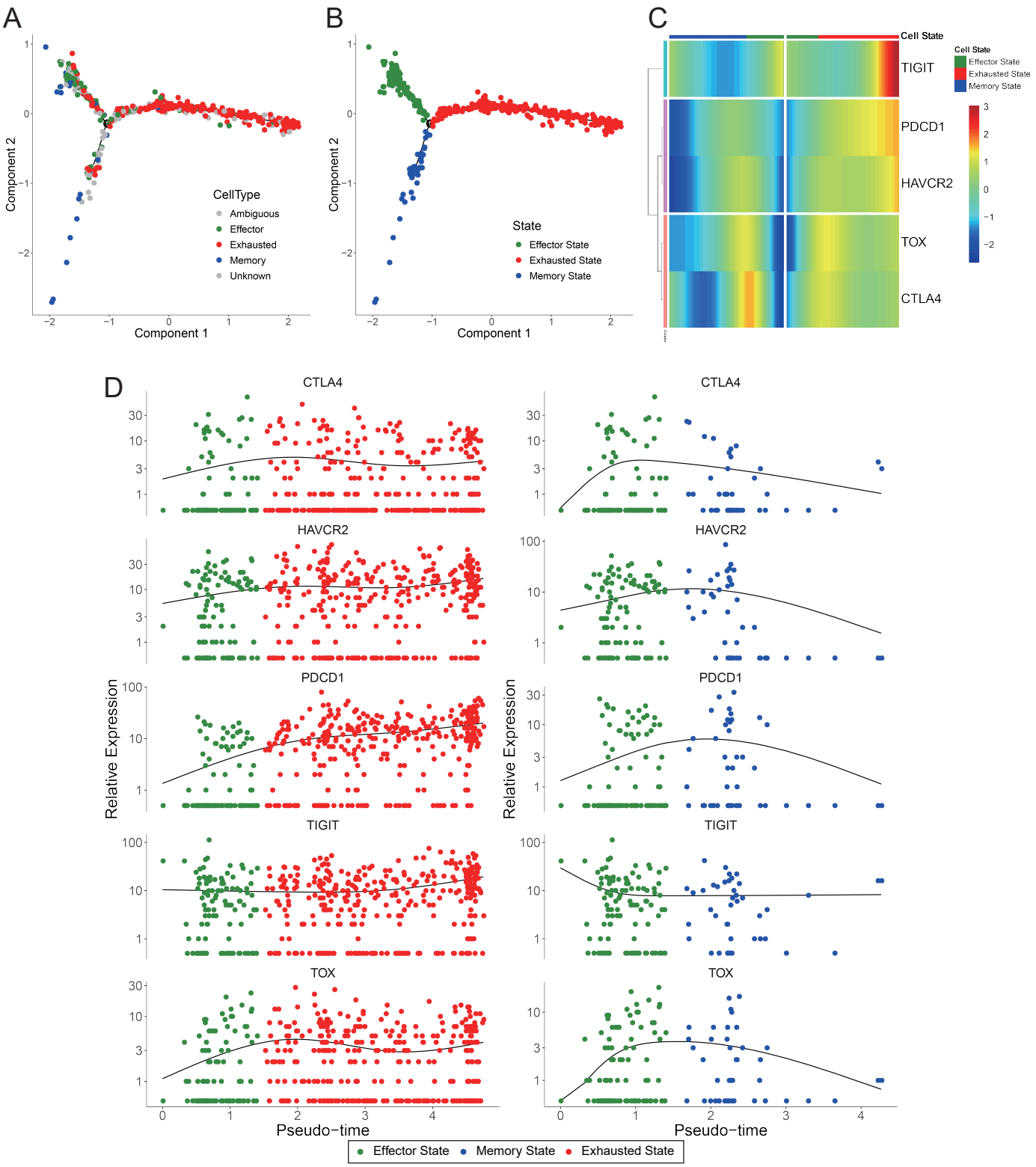

A
Component 2
Component 1
CellType
Ambiguous
Effector
Exhausted
Memory
Unknown
B
Component 2
Component 1
State
Effector State
Exhausted State
Memory State
C
Cell State
TIGIT
PDCD1
HAVCR2
TOX
CTLA4
Cell State
Effector State
Exhausted State
Memory State
D
CTLA4
HAVCR2
PDCD1
TIGIT
TOX
Relative Expression
Pseudo−time
Effector State
Memory State
Exhausted State

Figure 3

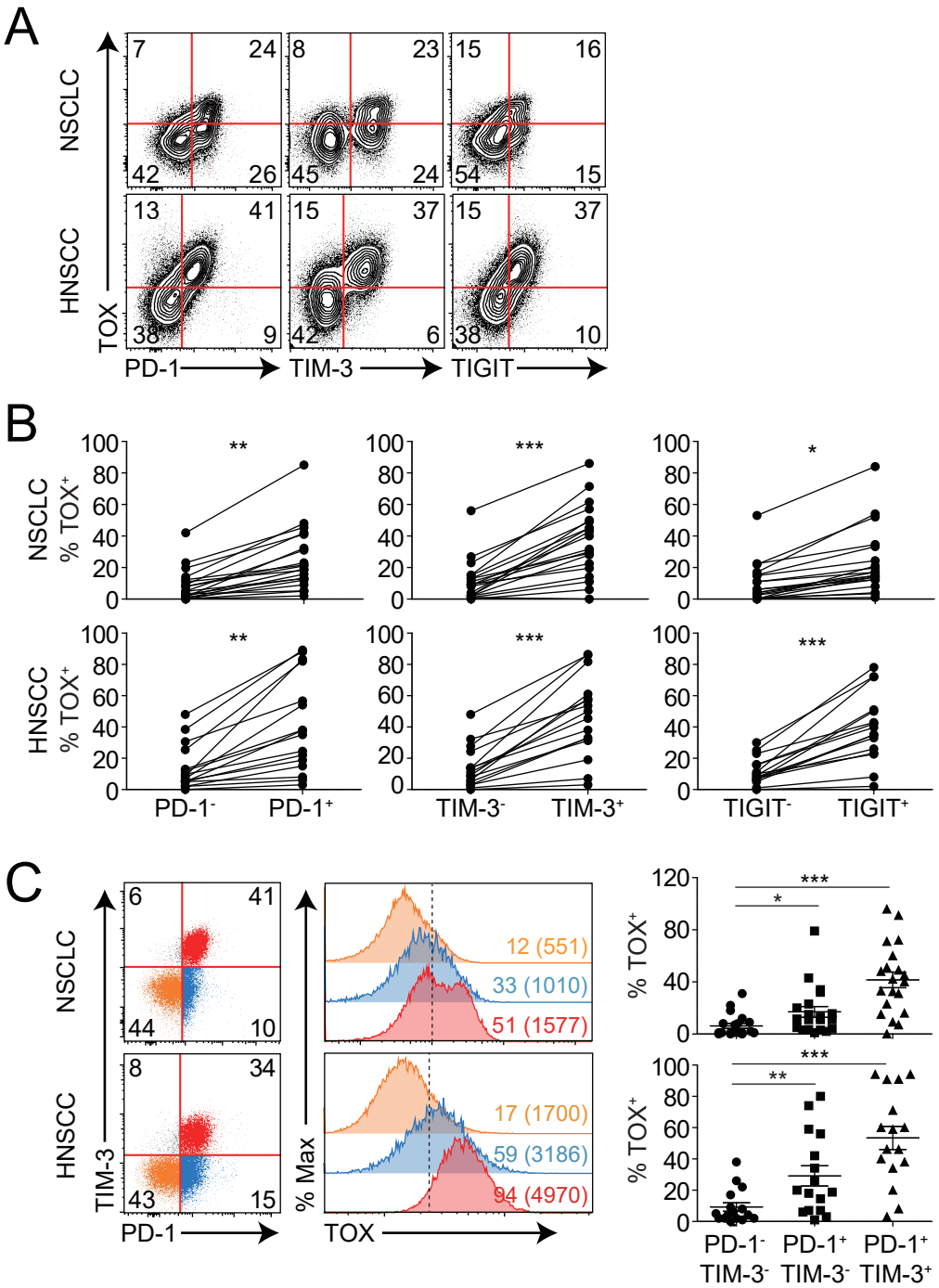

A
NSCLC
HNSCC
TOX
PD-1
TIM-3
TIGIT
7
24
42
26
8
23
45
24
15
16
54
15
13
41
38
9
15
37
42
6
15
37
38
10
B
NSCLC % TOX+
HNSCC % TOX+
PD-1-
PD-1+
TIM-3-
TIM-3+
TIGIT-
TIGIT+
C
NSCLC
HNSCC
TIM-3
PD-1
% Max
TOX
6
41
44
10
8
34
43
15
12 (551)
33 (1010)
51 (1577)
17 (1700)
59 (3186)
94 (4970)
% TOX+
PD-1- TIM-3-
PD-1+ TIM-3-
PD-1+ TIM-3+

Figure 4

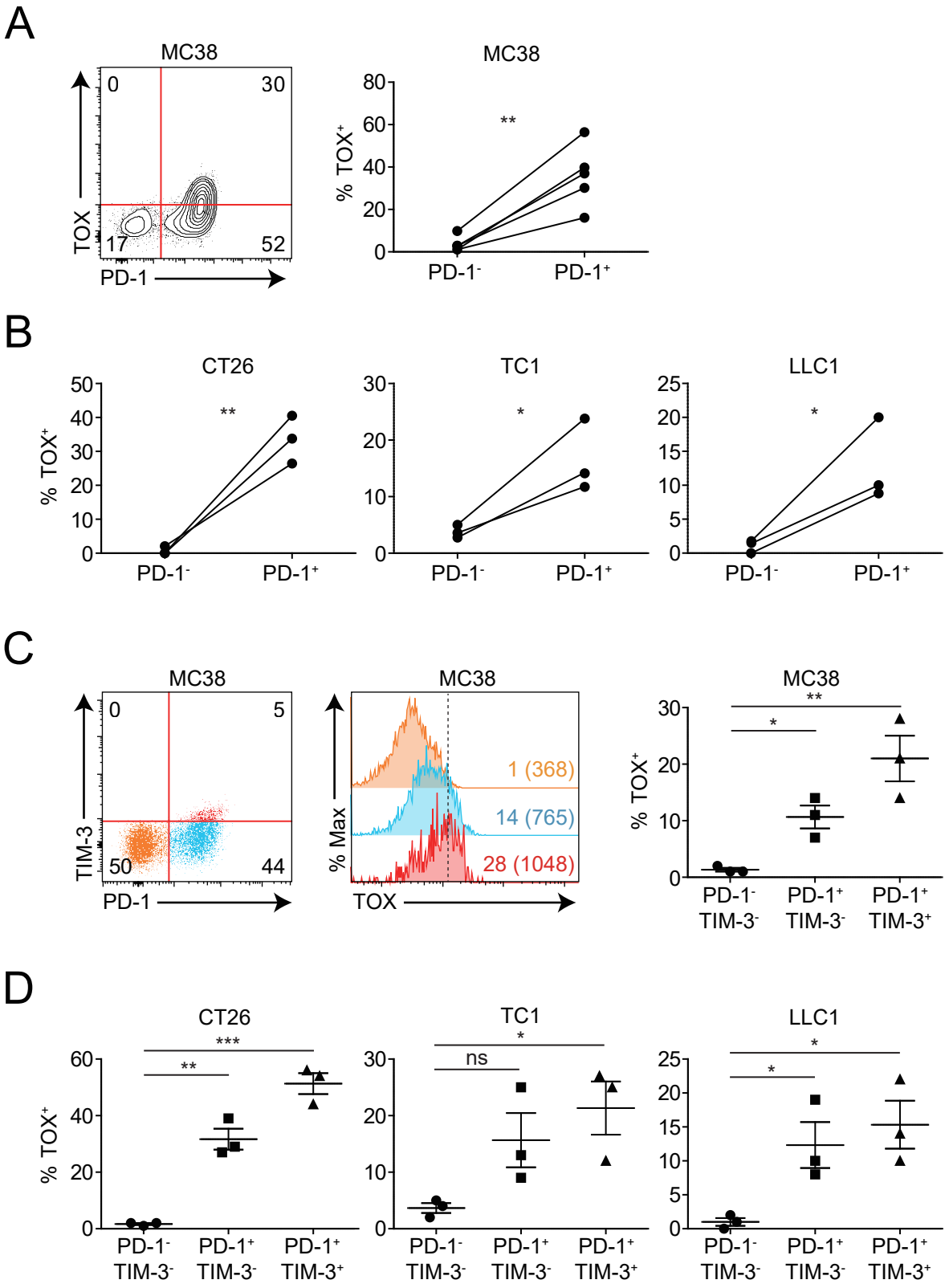

A
MC38
0
30
17
52
TOX
PD-1
MC38
% TOX+
80
60
40
20
0
**
PD-1-
PD-1+
B
CT26
TC1
LLC1
% TOX+
50
40
30
20
10
0
**
*
*
PD-1-
PD-1+
C
MC38
0
5
50
44
TIM-3
PD-1
MC38
% Max
1 (368)
14 (765)
28 (1048)
TOX
MC38
% TOX+
30
20
10
0
**
*
PD-1-
TIM-3-
PD-1+
TIM-3-
PD-1+
TIM-3+
D
CT26
% TOX+
60
40
20
0
***
**
TC1
30
20
10
0
*
ns
LLC1
25
20
15
10
5
0
*
*
PD-1-
TIM-3-
PD-1+
TIM-3-
PD-1+
TIM-3+

Figure 5

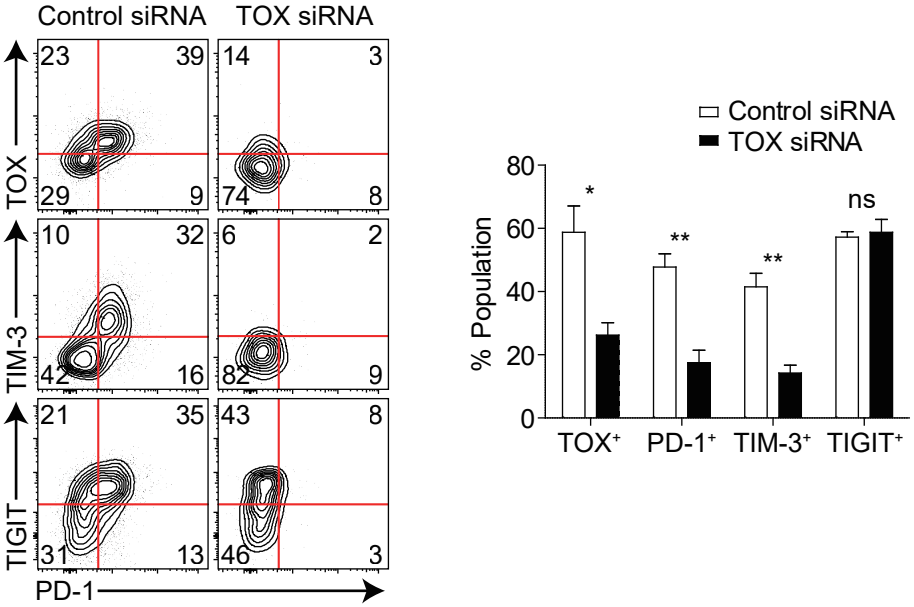
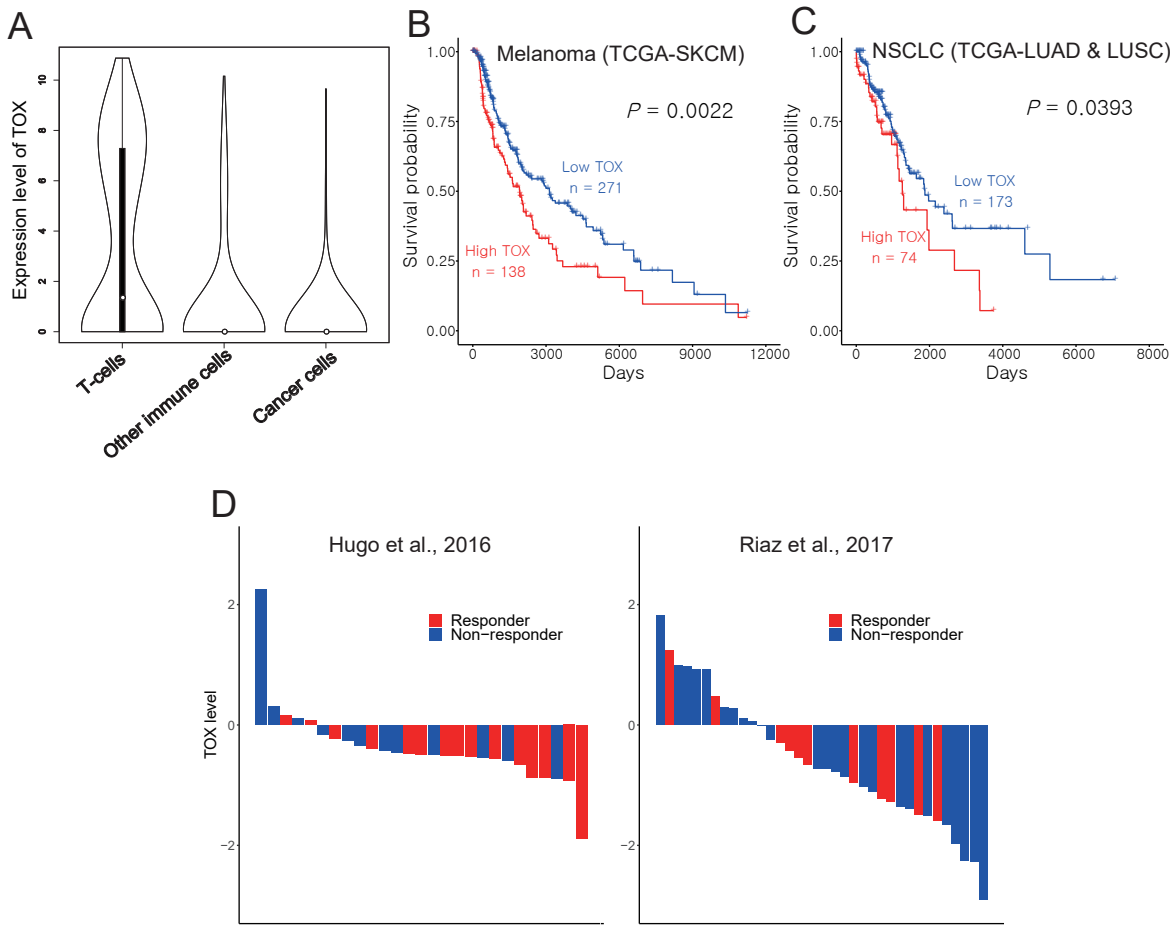

Control siRNA
TOX siRNA
TOX
TIM-3
TIGIT
PD-1
23
39
29
9
14
3
74
8
10
32
42
16
6
2
82
9
21
35
31
13
43
8
46
3
Control siRNA
TOX siRNA
% Population
80
60
40
20
0
*
**
**
ns
TOX+
PD-1+
TIM-3+
TIGIT+

Figure 6


A
Expression level of TOX
T-cells
Other immune cells
Cancer cells
B
Melanoma (TCGA-SKCM)
P = 0.0022
Survival probability
Low TOX
n = 271
High TOX
n = 138
Days
C
NSCLC (TCGA-LUAD & LUSC)
P = 0.0393
Survival probability
Low TOX
n = 173
High TOX
n = 74
Days
D
Hugo et al., 2016
Riaz et al., 2017
Responder
Non−responder
TOX level