

Integrative single-cell and pan-cancer analyses identify LINC01871 and TBX21 as candidate regulators of antitumor T-cell immunity in breast cancer

Xi Chen

Third Affiliated Hospital of Soochow University

Tao Tang

Third Affiliated Hospital of Soochow University

Wenyue Da

Third Affiliated Hospital of Soochow University

Tongbing Chen

Third Affiliated Hospital of Soochow University

Hui Wang

Third Affiliated Hospital of Soochow University

Pengyi Yu

75898389@qq.com

Third Affiliated Hospital of Soochow University

Research Article

Keywords: breast cancer, tumour immune microenvironment, prognostic signature, immunotherapy, long non-coding RNA

Posted Date: August 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10586247/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Immune checkpoint blockade benefits only a subset of patients with breast cancer, and the molecular determinants of an effective CD8⁺ T-cell response remain incompletely understood. We integrated single-cell transcriptomes from primary breast tumors, breast cancer brain metastases and normal tissues with bulk RNA-sequencing data from The Cancer Genome Atlas (TCGA). Genes co-expressed with cytotoxic CD8⁺ T cells converged on a coordinated network encompassing cytotoxic effector function, T-cell receptor signalling and immune-checkpoint regulation. From this network, we derived an eight-gene CD8⁺ T-cell activity signature comprising CCL5, TBX21, LINC01871, PLA2G2D, GFI1, KLRB1, CCR7 and PTPN7. The signature stratified patient survival independently of tumour stage and identified tumors exhibiting concurrent cytotoxic activation and checkpoint enrichment. Pan-cancer analysis across 33 cancer types revealed that LINC01871 and TBX21 were downregulated in tumors, whereas higher expression of either gene was associated with longer survival and inflamed molecular subtypes. In silico knockout analysis further predicted that loss of KLRB1, LINC01871, PTPN7 or TBX21 would disrupt the cytotoxic CD8⁺ T-cell programme. Together, these findings define a clinically relevant CD8⁺ T-cell activity programme and nominate LINC01871 and TBX21 as candidate regulators of antitumor immunity and potential immunotherapeutic targets in breast cancer.

1. Introduction

Breast cancer is the most commonly diagnosed malignancy among women and remains a leading cause of cancer-related mortality worldwide[1]. Immune checkpoint inhibitors targeting PD-1 or PD-L1 have improved clinical outcomes in triple-negative breast cancer, particularly when combined with chemotherapy. However, only a subset of patients derives durable benefit, and currently used biomarkers, including PD-L1 immunohistochemistry and tumour mutational burden capture only part of the biological complexity that determines therapeutic response[2–4]. The efficacy of checkpoint blockade ultimately depends on both the abundance and functional competence of tumour-infiltrating lymphocytes. Among these populations, cytotoxic CD8⁺ T cells are the principal effectors responsible for direct tumour-cell killing[5]. Defining the molecular programmes that sustain CD8⁺ T-cell activity in the breast tumour microenvironment, as well as the mechanisms that constrain or extinguish this response, is therefore essential for extending the benefits of immunotherapy to a broader population of patients.

CD8⁺ tumour-infiltrating lymphocytes are functionally heterogeneous, occupying a spectrum of states ranging from naive-like and effector populations to terminally exhausted cells. Tumour control and response to immune checkpoint blockade are therefore more closely associated with the composition and functional balance of these states than with the absolute abundance of T cells alone[6, 7]. This cellular heterogeneity is largely obscured by bulk transcriptomic profiling, in which signals from immune cells are diluted by those from malignant and stromal compartments. Consequently, bulk measurements cannot readily distinguish a genuinely inflamed, cytotoxic tumour from one that is merely infiltrated by dysfunctional lymphocytes. Single-cell RNA sequencing can resolve these immune states directly; however, single-cell cohorts are generally limited in size and are rarely accompanied by sufficiently

mature survival data[8, 9]. By contrast, TCGA and other large bulk-transcriptomic resources provide the extensive clinical annotation and outcome information that single-cell studies often lack[10]. Integrating a single-cell definition of CD8⁺ T-cell activity with validation in a large, clinically annotated cohort therefore offers a rational framework for identifying immune programmes with prognostic relevance and for prioritising candidate therapeutic targets.

We applied this integrative framework to breast cancer by constructing a single-cell atlas spanning normal breast tissue, primary tumours and brain metastases. We delineated the constituent T-cell states and identified genes whose expression was associated with cytotoxic CD8⁺ T-cell activity. From this gene set, we derived a prognostic signature that was developed and evaluated in the TCGA breast cancer cohort, characterised in relation to the cellular composition and functional state of the tumour immune microenvironment, and subsequently examined across 33 cancer types. To move beyond association and explore potential functional dependencies, we performed network-based *in silico* knockout analyses to identify signature genes predicted to be required for maintaining the CD8⁺ T-cell activity programme.

Two genes within the signature were particularly notable and provide a conceptual focus for the remainder of this study. TBX21 encodes T-bet, a master transcriptional regulator of type 1 immunity that promotes T helper 1 differentiation, drives CD8⁺ T-cell effector commitment and sustains cytotoxic function; accordingly, loss of TBX21 impairs antitumor T-cell responses[11, 12]. Its recovery within the activity signature was therefore mechanistically plausible and provided biological validation of the analytical framework. By contrast, LINC01871 is a poorly characterised long non-coding RNA with no defined immune function. Long non-coding RNAs are increasingly recognised as widespread regulators of immunity, controlling lymphocyte differentiation and activation and modulating transcriptional programmes in both innate and adaptive immune cells through chromatin-dependent, transcriptional and post-transcriptional mechanisms[13, 14]. Several exert context-specific functions during T-cell development and effector differentiation[15]. In cancer, long non-coding RNAs can either reinforce or alleviate immune suppression, making them an expanding source of candidate immunotherapeutic targets. Although LINC01871 has repeatedly appeared in immune-related lncRNA prognostic models of breast cancer, its underlying mechanism and association with a specific immune programme remain undefined. Its consistent concordance with TBX21 across our analyses, together with the predicted disruption of the cytotoxic CD8⁺ T-cell programme following its *in silico* loss, led us to prioritise LINC01871 as a candidate regulator rather than a passive prognostic marker. We therefore nominate LINC01871 and TBX21 for further mechanistic evaluation as potential modulators and therapeutic targets of antitumor T-cell immunity.

2. Materials and methods

2.1 Single-cell data acquisition and processing

Single-cell RNA-sequencing data for breast cancer primary tumors, brain metastases and normal tissue were obtained from the Gene Expression Omnibus under accessions GSE281490 and GSE281580. Raw

count matrices were processed in R with Seurat. Cells were filtered on detected-gene count, unique-molecular-identifier count and mitochondrial content; libraries were merged and integrated to correct batch effects. Counts were log-normalised, highly variable genes were selected, and these were used for principal-component analysis, shared-nearest-neighbour graph construction, Louvain clustering and Uniform Manifold Approximation and Projection (UMAP) embedding.

2.2 Cell-type annotation

Clusters were annotated by reference-based classification with SingleR against a human primary-cell reference and confirmed against canonical lineage markers displayed as dot plots and row-Z-score heatmaps of average expression. Twenty-two clusters were assigned to nine lineages. The T-cell compartment was re-clustered and annotated into regulatory T cells, naive CD8⁺, effector CD8⁺, exhausted CD8⁺, $\gamma\delta$ T and NKT states. Cell-type proportions were compared across tissue contexts.

2.3 CD8⁺ T-cell co-expression network and enrichment

Genes co-expressed with the cytotoxic CD8⁺ T-cell state were identified by correlation analysis and rendered as a correlation network. Network genes were tested for GO enrichment and KEGG pathway enrichment, with a q-value threshold defining significance.

2.4 TCGA data and construction of the prognostic signature

Bulk RNA-sequencing and matched clinical data for breast cancer were downloaded from TCGA. The 240 network genes present in the TCGA transcriptome were screened for association with overall survival by univariate Cox proportional-hazards regression. Prognostic genes entered LASSO-penalised Cox regression, and the penalty parameter was chosen by ten-fold cross-validation at the value minimising partial-likelihood deviance. The eight retained genes and their coefficients defined the risk score as the sum of expression values weighted by coefficient. Patients were dichotomised at the median score. Overall survival was compared by Kaplan–Meier analysis and the log-rank test, and discrimination was assessed by time-dependent receiver-operating-characteristic analysis at one, three and five years.

2.5 Nomogram and independent prognostic analysis

A nomogram integrating risk score, age, sex and tumour stage was constructed to predict one-, three- and five-year survival. Independence from clinical covariates was tested by multivariable Cox regression, and the distribution of clinical features across risk groups was shown as a stratified heatmap. Individual signature genes were evaluated by Kaplan–Meier analysis.

2.6 Immune infiltration and functional analysis

Immune-cell composition was estimated with CIBERSORT. Correlations among signature genes, the risk score, immune-cell fractions and immune-checkpoint genes were computed by Spearman correlation. Immune-cell fractions and single-sample gene set enrichment analysis (ssGSEA) immune-function

scores were compared between risk groups. Stromal, immune and ESTIMATE scores were computed with the ESTIMATE algorithm.

2.7 Pan-cancer analysis

Expression, survival and molecular-subtype data across 33 TCGA cancer types were used to evaluate the signature genes. Tumour-versus-normal differential expression, univariate Cox survival, expression across immune subtypes C1–C6 and per-cancer Kaplan–Meier survival were assessed, with emphasis on *LINC01871* and *TBX21*.

2.8 In-silico gene knockout

Network-based virtual knockout was performed with scTenifoldKnk on the T-cell single-cell data. For each of *KLRB1*, *LINC01871*, *PTPN7* and *TBX21*, a single-cell gene-regulatory network was constructed and the target gene deleted; genes were ranked by predicted change, and the top perturbed genes were subjected to GO and KEGG enrichment analysis.

2.9 Statistical analysis

Analyses were performed in R. Group comparisons used the Wilcoxon rank-sum or *t*-test as appropriate, correlations used Spearman's coefficient, and survival used the log-rank test and Cox regression. Two-sided $P < 0.05$ was considered significant, with thresholds denoted $P < 0.05$ (*), $P < 0.01$ (**) and $P < 0.001$ (***).

3. Results

3.1 A single-cell atlas resolves the breast cancer T-cell compartment across tissue contexts

We assembled single-cell transcriptomes from two Gene Expression Omnibus series (GSE281490 and GSE281580)[16] spanning normal tissue, primary breast tumour and breast cancer brain metastasis. After quality control and integration to remove batch effects, unsupervised clustering produced 22 clusters, which reference-based classification with SingleR[17, 18], collapsed into nine lineages: epithelial cells, tissue stem cells, macrophages/microglia, monocytes, T cells, B cells, endothelial cells, astrocytes and neurons (Fig. 1a). The assignments matched canonical markers – *CD3D*, *IL7R*, *CCL5* and *NKG7* for T cells, *HLA-DRA* and *LYZ* for the myeloid compartment, *IGHA1* and *IGKC* for B/plasma cells, and *VWF* and *CLDN5* for endothelium (Fig. 1b,c).

Cellular composition differed across tissue compartments. Normal samples were characterised by abundant resident macrophage/microglial and endothelial populations, whereas primary tumours and brain-metastatic lesions showed greater representation of epithelial and tissue stem-like cells, together with variable levels of T-cell infiltration (Fig. 1d). Re-clustering the T cells resolved six states: regulatory T cells, naive CD8⁺, effector CD8⁺ (CD8_Teff), exhausted CD8⁺ (CD8_Tex), $\gamma\delta$ T cells and NKT cells (Fig.

1e). Exhausted CD8⁺ T cells were enriched in a subset of tumour samples, whereas naive and effector populations predominated in normal tissue. These findings indicate that the intratumoural CD8⁺ T-cell compartment is skewed toward dysfunctional states, with the magnitude of this shift varying markedly among patients.

3.2 Genes co-expressed with cytotoxic CD8⁺ T cells define a T-cell activity programme

To identify potential regulators of cytotoxic antitumor immunity, we analysed bulk RNA-sequencing data from the TCGA-BRCA cohort. Immune-cell infiltration was inferred using CIBERSORT, and genes associated with cytotoxic CD8⁺ T-cell infiltration were identified by correlation analysis. These genes were subsequently used to construct a co-expression network anchored to the cytotoxic CD8⁺ T-cell programme. The network was dense and dominated by the machinery of effector T cells (*GZMK*, *GZMB*, *PRF1*, *NKG7*, *CD8A*, *IFNG*, *CCL5*, *EOMES* and *CXCR3*), alongside numerous T-cell receptor variable segments and the inhibitory receptors (*PDCD1*, *CTLA4*, *BTLA* and *TIGIT*) (Fig. 2a). KEGG pathway enrichment analysis of the network genes revealed a coherent immune-related functional landscape, with prominent enrichment of T-cell receptor signaling, cytokine–cytokine receptor interaction, natural killer cell-mediated cytotoxicity, Th1 and Th2 cell differentiation, Th17 cell differentiation, and the PD-L1 expression and PD-1 checkpoint pathway in cancer. These pathways collectively connect antigen recognition and receptor-proximal signaling with cytokine-mediated intercellular communication, cytotoxic effector function, helper T-cell polarisation, and immune-checkpoint regulation. This functional convergence indicates that the co-expression network captures a coordinated antitumor immune programme rather than merely reflecting the overall abundance of infiltrating lymphocytes (Fig. 2b). Together, these findings indicate that the programme captures both sides of the antitumor immune response: the cytotoxic effector machinery that mediates tumour-cell killing and the checkpoint pathways that restrain its activity.

3.3 An eight-gene CD8⁺ T-cell activity signature independently stratifies survival in breast cancer

We next tested whether this programme carried prognostic information in bulk disease. Of the network genes, 240 were measured in the TCGA breast cancer transcriptome. Screening each against overall survival by univariate Cox regression retained a prognostic subset, and LASSO-penalised Cox regression[19] reduced this to a parsimonious eight-gene signature, including *CCL5*, *TBX21*, *LINC01871*, *PLA2G2D*, *GFI1*, *KLRB1*, *CCR7* and *PTPN7*, with the penalty chosen by ten-fold cross-validation at the minimum partial-likelihood deviance (Fig. 2c,d). Every gene carried a protective hazard ratio, matching the expectation that the programme reads productive rather than exhausted immunity.

For each patient, a risk score was derived from the coefficient-weighted expression of the eight signature genes. Using the median risk score as the cutoff, patients were dichotomised into high- and low-risk groups. Kaplan–Meier survival analysis showed that high-risk patients had significantly shorter overall survival than low-risk patients, indicating that the eight-gene signature effectively discriminates patients

with distinct clinical outcomes(log-rank $P < 0.001$; Fig. 2e). The score discriminated outcome with time-dependent areas under the curve of 0.65, 0.68 and 0.63 at one, three and five years (Fig. 2f). In univariable Cox regression analyses, each of the eight signature genes was individually associated with overall survival, and higher expression consistently corresponded to a reduced risk of death. Kaplan–Meier analyses further confirmed that patients with high expression of each gene experienced longer survival than those with low expression. Together, these findings suggest that the prognostic value of the signature reflects the coordinated contribution of multiple protective genes rather than dependence on a single component(Fig. 3c).

3.4 The eight-gene signature provides stage-independent prognostic stratification in breast cancer

To place the score alongside routine clinical variables, we built a nomogram combining risk score with age, sex and tumour stage (Fig. 3a). Risk score, age and stage each contributed to predicted one-, three- and five-year survival, and the risk score carried weight comparable to stage. A stratified heatmap showed that high-risk status accumulated with older age and higher stage yet remained separable from them (Fig. 3b), and in multivariable Cox regression the signature stayed significant after adjustment for these covariates. The score adds prognostic information that staging alone does not provide.

3.5 The eight-gene signature marks an inflamed, cytotoxic tumour immune microenvironment

Because the signature was built from T-cell activity genes, we asked what immune context a favourable score reflects. The signature genes showed broadly positive correlations with the estimated abundance of immune effector and pro-inflammatory cell populations, including CD8⁺ T cells, activated and resting memory CD4⁺ T cells, follicular helper T cells and M1 macrophages. By contrast, their expression was inversely associated with M0 and M2 macrophages. This coordinated pattern links the eight-gene signature to an immune-infiltrated, cytotoxic and pro-inflammatory tumour microenvironment, while suggesting a relative reduction in undifferentiated or immunoregulatory macrophage states(Fig. 4a). The risk score itself ran opposite to the effector molecules *GZMB* and *PRF1* and to the checkpoint genes *PDCD1*, *CD274*, *CTLA4*, *HAVCR2*, *IDO1* and *PDCD1LG2* (Fig. 4b). Low-risk tumors were simultaneously enriched for cytotoxic immune activity and checkpoint expression, defining an inflamed, checkpoint-active tumour microenvironment. This phenotype suggests the presence of pre-existing antitumor immunity under adaptive immune restraint and may therefore identify tumors that are more amenable to immune checkpoint blockade.

Cell-fraction estimates were consistent with the functional immune profiles. Low-risk tumors were enriched for CD8⁺ T cells, activated memory CD4⁺ T cells, follicular helper T cells and plasma cells, whereas high-risk tumors contained higher proportions of M0 and M2 macrophages. This contrast supports an effector-rich, immune-inflamed phenotype in low-risk tumors and a more immunosuppressive, myeloid-dominated phenotype in high-risk tumors(Fig. 4c). Single-sample gene set enrichment analysis placed the low-risk group higher across antigen presentation, co-stimulation,

cytolytic activity, checkpoint, HLA, inflammation-promoting and type I/II interferon-response modules (Fig. 4d). The signature reads a coordinated immune axis, not the abundance of any single cell type.

3.6 Pan-cancer analysis prioritises *LINC01871* and *TBX21* as candidate regulators of antitumour immunity

We then asked whether the signature genes matter beyond breast cancer by profiling them across 33 TCGA cancer types. *LINC01871* and *TBX21* were differentially expressed between tumour and normal tissue in most cancers and were generally lower in tumors (Fig. 5a,b). Despite this repression, higher expression of the signature genes was protective across cancers in univariate Cox analysis (Fig. 5c), and the genes reached their highest levels in the immunologically active molecular subtypes C2 (IFN- γ dominant) and C3 (inflammatory) and their lowest in C4 (lymphocyte-depleted) and C5 (immunologically quiet) (Fig. 5d). Per-cancer Kaplan–Meier analysis confirmed that high *LINC01871* and *TBX21* predicted longer survival in breast cancer, low-grade glioma, melanoma, renal and ovarian cancers (Fig. 5e). Their tumour-associated repression, favourable prognostic associations and preferential expression in inflamed molecular subtypes collectively nominate *LINC01871* and *TBX21* as candidate regulators of antitumor immunity. These observations suggest that reduced expression of the two genes may accompany immune escape, whereas restoring their activity could potentially strengthen cytotoxic T-cell responses, a hypothesis that warrants direct experimental validation.

3.7 *LINC01871* and *TBX21* partition the tumor microenvironment

Turning to these two genes, we split breast tumors by *LINC01871* or by *TBX21* expression and characterised the resulting microenvironments. In tumors high for either gene, the co-regulated genes were enriched for cytokine–cytokine receptor interaction, chemokine signaling, antigen processing and presentation and T-cell receptor signaling (Fig. 6a,b). High-expression tumors had greater stromal, immune and ESTIMATE scores, indicating higher immune and stromal content and lower tumor purity (Fig. 6c,d), and were enriched for CD8⁺ T cells, plasma cells and follicular helper T cells while depleted of M0 and M2 macrophages and resting mast cells (Fig. 6e). Across immune populations, *LINC01871* correlated most strongly and positively with CD8⁺ T cells, activated CD4⁺ memory T cells, follicular helper T cells and M1 macrophages, and negatively with M2 macrophages (Fig. 6f). This single gene therefore recapitulated the principal immunological axis captured by the full eight-gene signature, showing concordant associations with cytotoxic lymphocyte infiltration, interferon-related activity and immune-checkpoint expression. Its expression thus reflected not only the abundance of tumour-infiltrating immune cells, but also the functional state of an inflamed and cytotoxic tumour microenvironment.

3.8 Network-based in silico knockout identifies candidate dependencies of the cytotoxic CD8⁺ T-cell programme

Correlation cannot separate the genes that drive the activity programme from those that merely ride along with it. To rank candidates by dependency, we performed network-based in-silico knockout with

scTenifoldKnk[20], which fits a single-cell gene-regulatory network and predicts the transcriptome-wide effect of deleting a gene. Virtual knockout of *KLRB1*, *LINC01871*, *PTPN7* and *TBX21* each produced strong, coordinated dysregulation of the cytotoxic and identity modules of the T-cell network. The most strongly affected genes were largely shared across the four knockout models, including *CCL5*, *CD8A*, *CD3D*, *GZMA*, *GZMB*, *NKG7*, *CTSW* and *CD52*. These genes form a coherent module encompassing T-cell receptor signaling, lymphocyte activation and cytotoxic effector function. The extensive overlap among the perturbation profiles suggests that *KLRB1*, *LINC01871*, *PTPN7* and *TBX21* converge on a common CD8⁺ T-cell effector programme and represent complementary candidate dependencies within the same regulatory network(Fig. 7a–d).

Enrichment of the perturbed genes reinforced this reading. Losing these genes disrupted biological processes governing T-cell activation and differentiation, leukocyte-mediated cytotoxicity and cell killing; cellular components including the cytolytic granule and the external side of the plasma membrane; molecular functions built on MHC protein binding and immune-receptor activity; and KEGG pathways for cytokine–cytokine receptor interaction, cell adhesion molecules, antigen processing and presentation and Th1/Th2 differentiation (Fig. 7e–h). *KLRB1*, *LINC01871*, *PTPN7* and *TBX21* therefore emerge as candidate network dependencies that may help sustain the cytotoxic CD8⁺ T-cell programme, rather than merely serving as correlates of immune activation. Their predicted perturbation of a shared downstream effector module suggests that altering any of these genes could reconfigure the broader cytotoxic network.

4. Discussion

In this study, we sought to define the gene-level architecture of an effective antitumor CD8⁺ T-cell response in breast cancer and to translate this definition into a prioritised set of candidate therapeutic targets. By deriving the programme from single-cell transcriptomic data and evaluating it in a large, outcome-annotated bulk cohort, we established an eight-gene activity signature that stratified survival independently of clinical stage and captured the immune composition and functional state of the tumor microenvironment. Notably, the signature identified tumors characterised by concurrent cytotoxic activity and checkpoint expression, a biologically relevant state in which pre-existing antitumor immunity may be amenable to reinvigoration by immune checkpoint blockade. The signature may therefore help identify patients most likely to benefit from such treatment, although its predictive value will require prospective validation in independent, immunotherapy-treated cohorts.

Two observations support the biological relevance of the prioritised targets. First, all eight signature genes were associated with favourable outcomes and correlated with cytotoxic lymphocyte infiltration and interferon-related activity, indicating that the signature primarily captures productive antitumor immunity rather than the dysfunctional or exhausted states that may also accompany dense immune infiltration. Second, *LINC01871* and *TBX21* showed a consistent pan-cancer pattern: both were downregulated in tumors, whereas higher expression was associated with improved survival and inflamed molecular subtypes. Although these associations do not establish causality, they are

compatible with the possibility that reduced expression of these genes contributes to immune escape. TBX21 encodes T-bet, a master regulator of type 1 immunity and CD8⁺ T-cell effector differentiation^{13,14}; its identification therefore provides mechanistic support for the biological validity of the signature. The striking concordance of LINC01871, a poorly characterised long non-coding RNA, with TBX21 across multiple independent analyses raises the hypothesis that the two may participate in a shared regulatory axis. This consistency, together with the predicted disruption of the cytotoxic programme following LINC01871 loss, positions the lncRNA as a compelling but largely unexplored candidate regulator of antitumor CD8⁺ T-cell activity.

The in silico knockout analysis provided a first step beyond correlative prioritisation by probing predicted network dependencies. scTenifoldKnk indicated that deletion of KLRB1, LINC01871, PTPN7 or TBX21 would disrupt a shared module enriched for cytotoxic effector function and antigen presentation, consistent with these genes occupying functionally influential positions within the CD8⁺ T-cell network rather than acting as passive markers. This distinction is therapeutically relevant: a gene that merely correlates with favourable outcome may offer limited value as an intervention point, whereas one whose perturbation is predicted to destabilise the effector programme represents a higher-priority candidate for mechanistic validation. The convergence of all four genes on a common downstream programme further suggests that they may constitute complementary or alternative points of therapeutic intervention. Whether they can be targeted in combination, or selected according to the tractability of a given therapeutic modality, will require direct experimental evaluation.

Several limitations temper these conclusions. First, the single-cell cohort was relatively small and included brain-metastatic specimens; consequently, the T-cell states identified here may not fully capture the heterogeneity of primary breast tumors and will require validation in larger, site-specific datasets. Second, the eight-gene signature was derived and evaluated within TCGA and has not yet undergone locked-model validation in an independent external cohort. Its ability to predict benefit from immune checkpoint blockade also remains untested in an immunotherapy-treated population. Third, the scTenifoldKnk results represent computationally inferred dependencies rather than experimentally established causal relationships. The molecular function of LINC01871, including its regulatory relationship with TBX21, likewise remains unresolved. Priority next steps therefore include gain- and loss-of-function perturbation of LINC01871 and TBX21 in primary human CD8⁺ T cells, together with mechanistic and therapeutic evaluation in immunocompetent breast cancer models. Such studies will be necessary to determine whether either gene directly sustains cytotoxic T-cell function and can be therapeutically manipulated to enhance antitumor immunity.

By distilling CD8⁺ T-cell activity into a compact, biologically grounded transcriptional readout and prioritising candidate regulators according to predicted network dependency rather than correlation alone, this study provides both a potential framework for patient stratification and a focused set of candidate targets for breast cancer immunotherapy, with LINC01871 and TBX21 emerging as the leading candidates for further mechanistic and therapeutic evaluation.

Declarations

Ethics approval and consent to participate

Ethical approval was not applicable because the data were obtained from the public database.

Conflict of interest

The authors declare no competing interests.

Funding

The study was supported by two Changzhou Science and Technology Projects (project numbers: CJ20259007 and QN202516).

Author Contribution

Xi Chen✉Tao Tang: Conceptualization, established the analytical framework for single-cell transcriptome data mining, performed quality control and dimensionality reduction analysis on the GSE281490 and GSE281580 datasets, constructed the CD8⁺ T cell trajectory map, wrote the core results section of the original manuscript. Wenyue Da: Conducted the pan-cancer survival integration analysis based on TCGA database, screened and verified the 8-gene CD8⁺ T cell activity signature, completed the statistical test of prognostic correlation across multiple clinical endpoints. Tongbing Chen: Designed and implemented the network-guided in silico perturbation analysis, identified LINC01871 and TBX21 as key structural regulators of the cytotoxic T cell effector program, interpreted the regulatory mechanism of the two candidate molecules. Hui Wang: Curated the clinical metadata of breast cancer patients, analyzed the correlation between the 8-gene signature and immunotherapy response, organized all supplementary tables and figure legends. Pengyi Yu : Supervised the entire research project, provided critical academic guidance for the study design, reviewed and revised the manuscript for key intellectual content, supported the funding for data acquisition. All authors have read and approved the final submitted version of the manuscript. Every listed author has made substantive, verifiable intellectual contributions to this work, and no honorary or ghost authors are included. All authors agree to be fully accountable for the accuracy and integrity of the entire research content, and consent to the exclusive submission of this manuscript to Discovery Oncology.

Data availability

Single-cell data are available from the Gene Expression Omnibus under accessions GSE281490 and GSE281580. TCGA data were obtained from the Genomic Data Commons

References

1. Bray, F., et al., *Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA Cancer J Clin, 2024. **74**(3): p. 229–263.
2. Cortes, J., et al., *Pembrolizumab plus Chemotherapy in Advanced Triple-Negative Breast Cancer*. N Engl J Med, 2022. **387**(3): p. 217–226.
3. Chisaki, Y., et al., *Cost-Effectiveness Analysis of Pembrolizumab Plus Chemotherapy Compared with Chemotherapy as First-Line Treatment for Advanced PD-L1-Positive Triple-Negative Breast Cancer from a Japanese Healthcare Perspective*. Clin Drug Investig, 2025. **45**(6): p. 317–326.
4. Schmid, P., et al., *Atezolizumab and Nab-Paclitaxel in Advanced Triple-Negative Breast Cancer*. N Engl J Med, 2018. **379**(22): p. 2108–2121.
5. Savas, P., et al., *Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis*. Nat Med, 2018. **24**(7): p. 986–993.
6. Wherry, E.J. and M. Kurachi, *Molecular and cellular insights into T cell exhaustion*. Nat Rev Immunol, 2015. **15**(8): p. 486–99.
7. van der Leun, A.M., D.S. Thommen, and T.N. Schumacher, *CD8(+) T cell states in human cancer: insights from single-cell analysis*. Nat Rev Cancer, 2020. **20**(4): p. 218–232.
8. Wu, S.Z., et al., *A single-cell and spatially resolved atlas of human breast cancers*. Nat Genet, 2021. **53**(9): p. 1334–1347.
9. Azizi, E., et al., *Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment*. Cell, 2018. **174**(5): p. 1293–1308 e36.
10. Cancer Genome Atlas Research, N., et al., *The Cancer Genome Atlas Pan-Cancer analysis project*. Nat Genet, 2013. **45**(10): p. 1113–20.
11. Kallies, A. and K.L. Good-Jacobson, *Transcription Factor T-bet Orchestrates Lineage Development and Function in the Immune System*. Trends Immunol, 2017. **38**(4): p. 287–297.
12. Szabo, S.J., et al., *A novel transcription factor, T-bet, directs Th1 lineage commitment*. Cell, 2000. **100**(6): p. 655–69.
13. Atianand, M.K., D.R. Caffrey, and K.A. Fitzgerald, *Immunobiology of Long Noncoding RNAs*. Annu Rev Immunol, 2017. **35**: p. 177–198.
14. Chen, Y.G., A.T. Satpathy, and H.Y. Chang, *Gene regulation in the immune system by long noncoding RNAs*. Nat Immunol, 2017. **18**(9): p. 962–972.
15. Hu, G., et al., *Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation*. Nat Immunol, 2013. **14**(11): p. 1190–8.
16. Caire, R., et al., *A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer*. Cell, 2026. **189**(12): p. 3701–3718 e30.

17. Hao, Y., et al., *Integrated analysis of multimodal single-cell data*. Cell, 2021. **184**(13): p. 3573–3587 e29.
18. Aran, D., et al., *Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage*. Nat Immunol, 2019. **20**(2): p. 163–172.
19. Simon, N., et al., *Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent*. J Stat Softw, 2011. **39**(5): p. 1–13.
20. Osorio, D., et al., *scTenifoldKnk: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation*. Patterns (N Y), 2022. **3**(3): p. 100434.

Figures

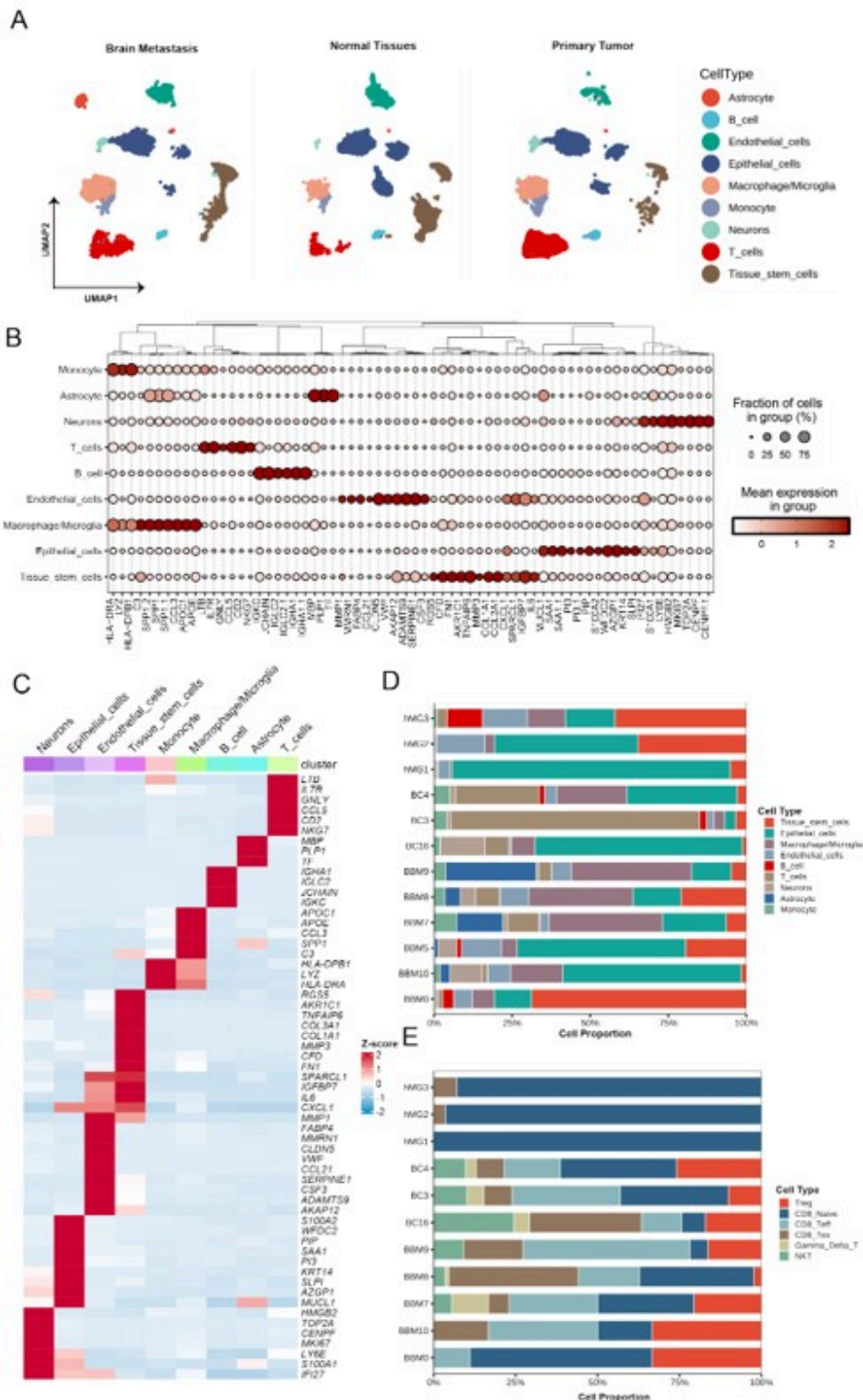


Figure 1

A single-cell atlas resolves the breast cancer T-cell compartment across tissue contexts.

a, UMAP of integrated single-cell transcriptomes coloured by cell type and split by tissue context (brain metastasis, normal tissue, primary tumour). **b**, Dot plot of lineage-marker expression across cell types; dot size, fraction of expressing cells; colour, mean expression. **c**, Heatmap of top marker genes per

lineage (row Z-score). **d**, Cell-type composition of each library. **e**, Composition of the re-clustered T-cell compartment (Treg, CD8_Naive, CD8_Teff, CD8_Tex, $\gamma\delta$ T, NKT).

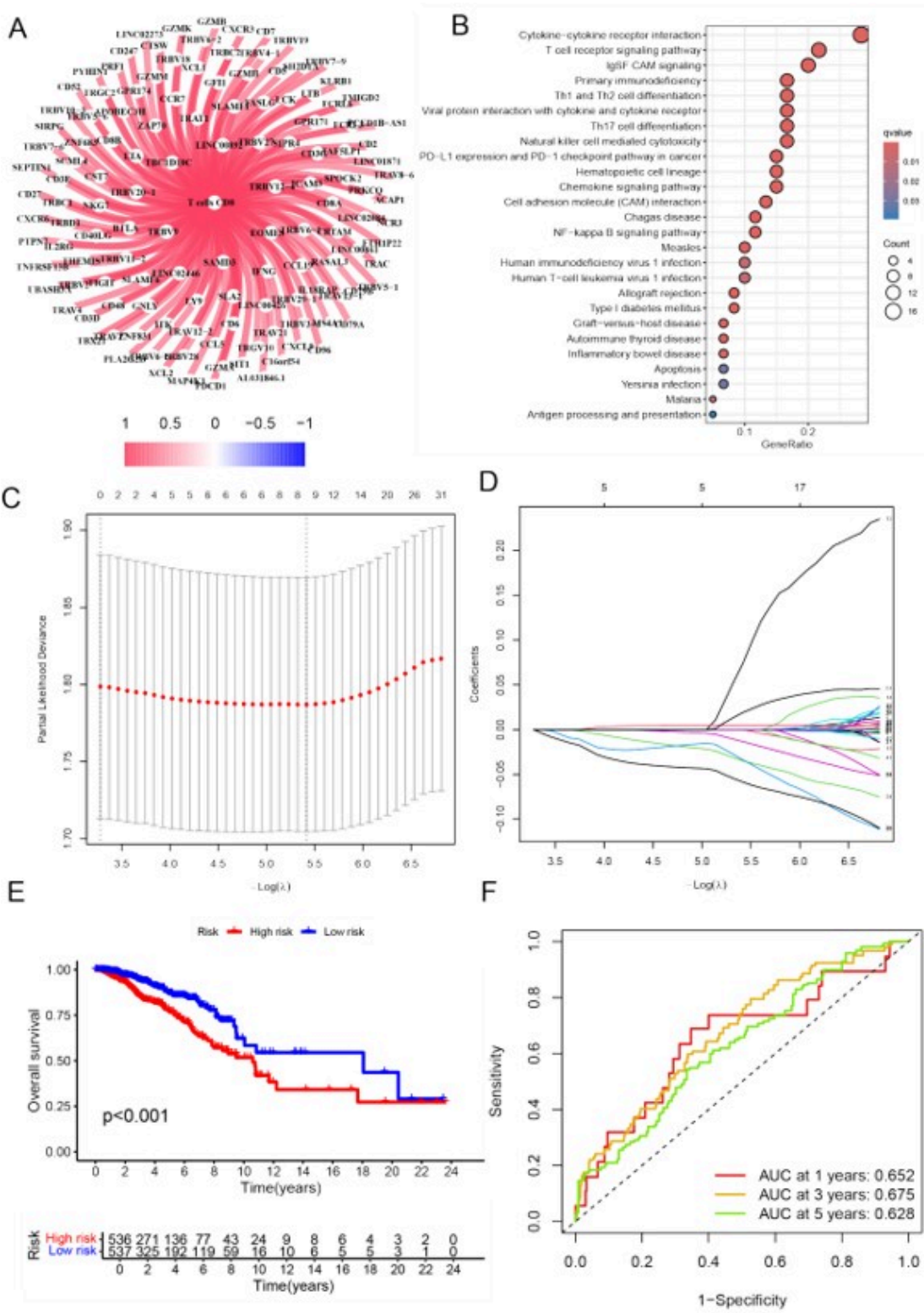


Figure 2

A CD8⁺ T-cell activity programme yields a prognostic signature.

a, Correlation network of genes co-expressed with cytotoxic CD8⁺ T cells. **b**, KEGG enrichment of network genes. **c**, LASSO cross-validation partial-likelihood deviance versus log(λ). **d**, LASSO coefficient paths. **e**, Kaplan–Meier overall survival for high- versus low-risk patients (log-rank $P < 0.001$). **f**, Time-dependent ROC at one, three and five years.

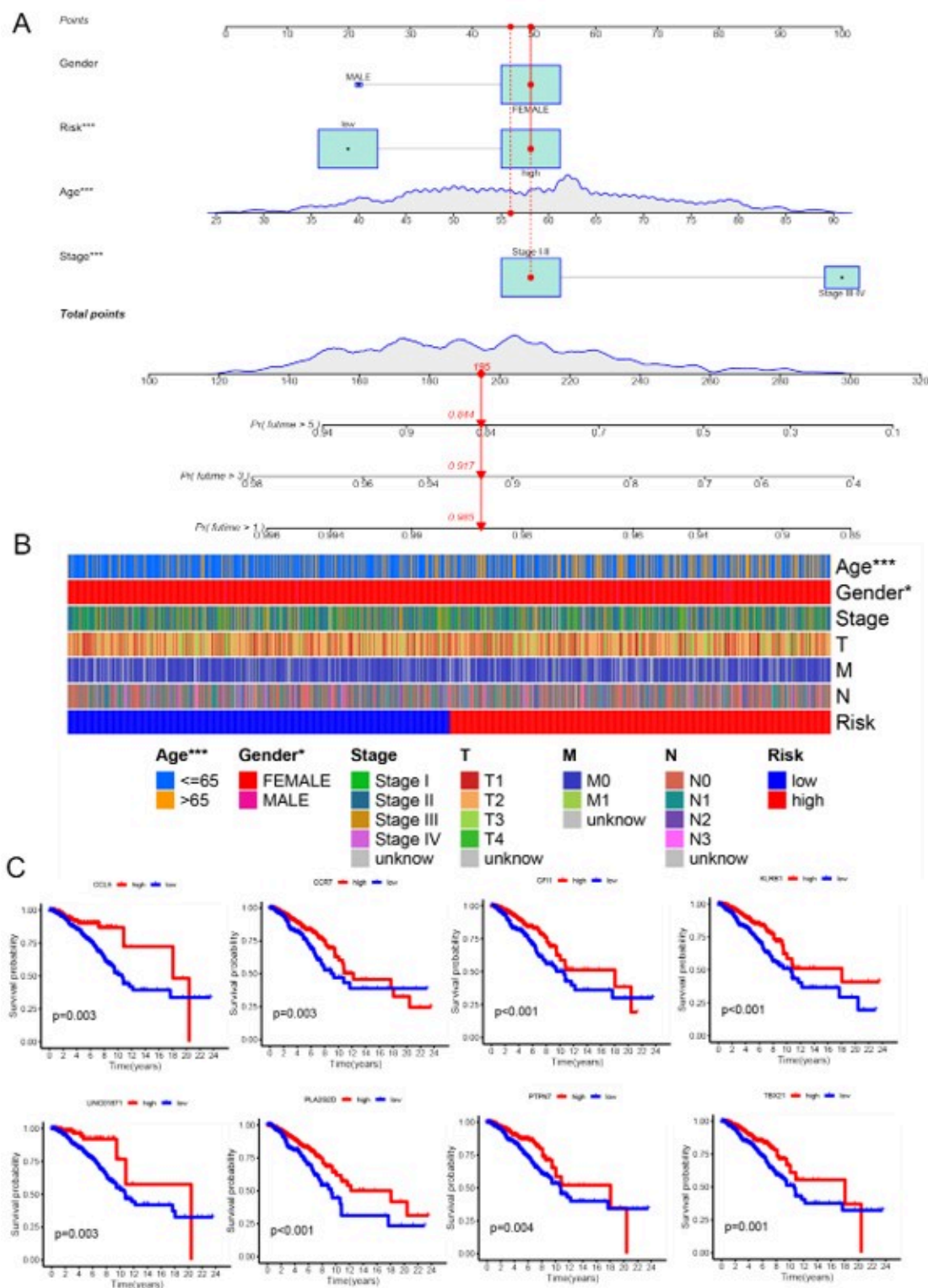


Figure 3

The signature is an independent prognostic factor.

a, Nomogram integrating risk score, age, sex and stage. **b**, Stratified heatmap of clinical features by risk group. **c**, Kaplan–Meier survival for individual signature genes (*CCL5*, *CCR7*, *GFI1*, *KLRB1*, *LINC01871*, *PLA2G2D*, *PTPN7*, *TBX21*).

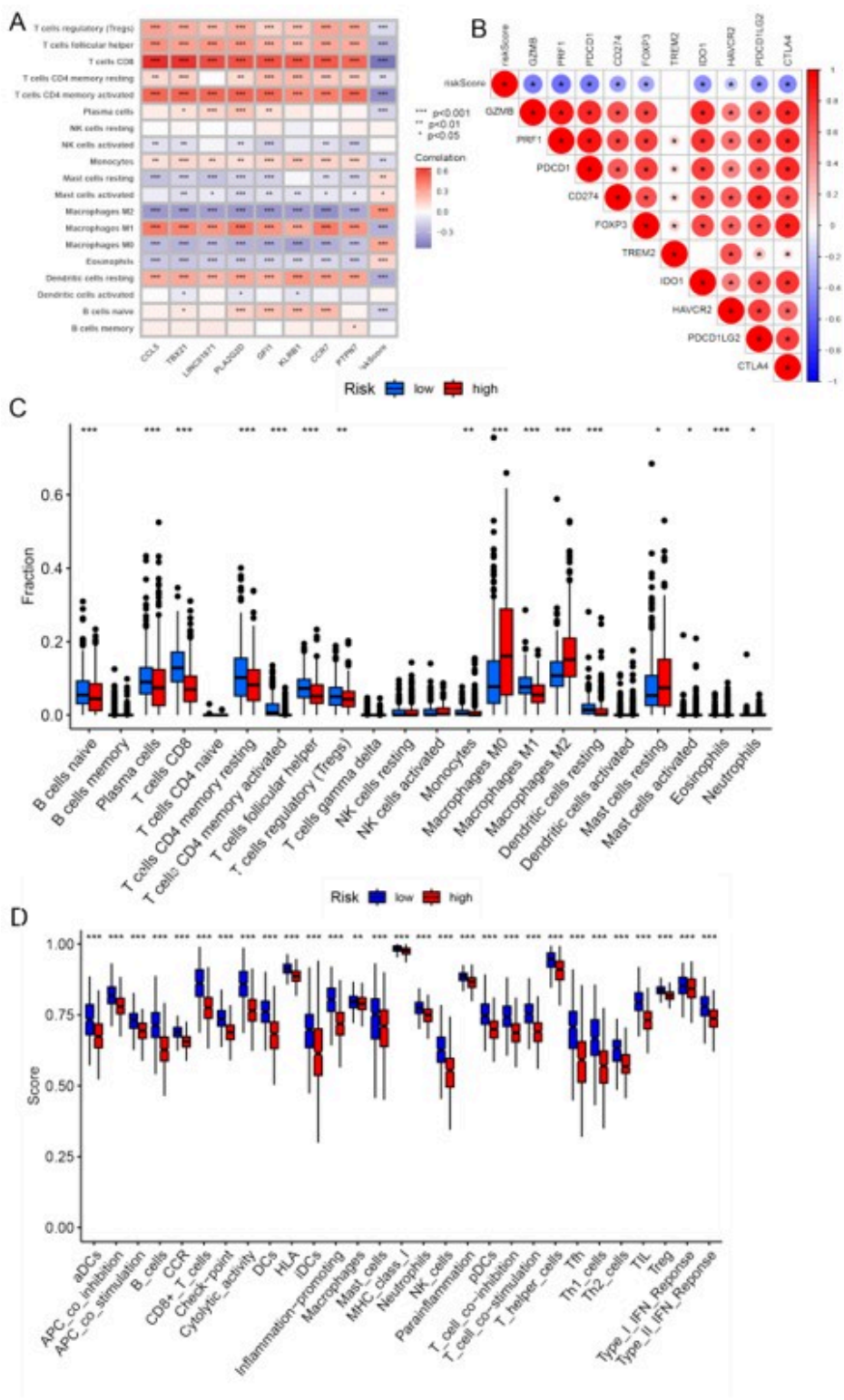


Figure 4

The signature marks an inflamed, cytotoxic microenvironment.

a, Correlation of signature genes and risk score with CIBERSORT immune-cell fractions. **b**, Correlation of risk score with effector and checkpoint genes. **c**, Immune-cell fractions by risk group. **d**, ssGSEA immune-function scores by risk group.

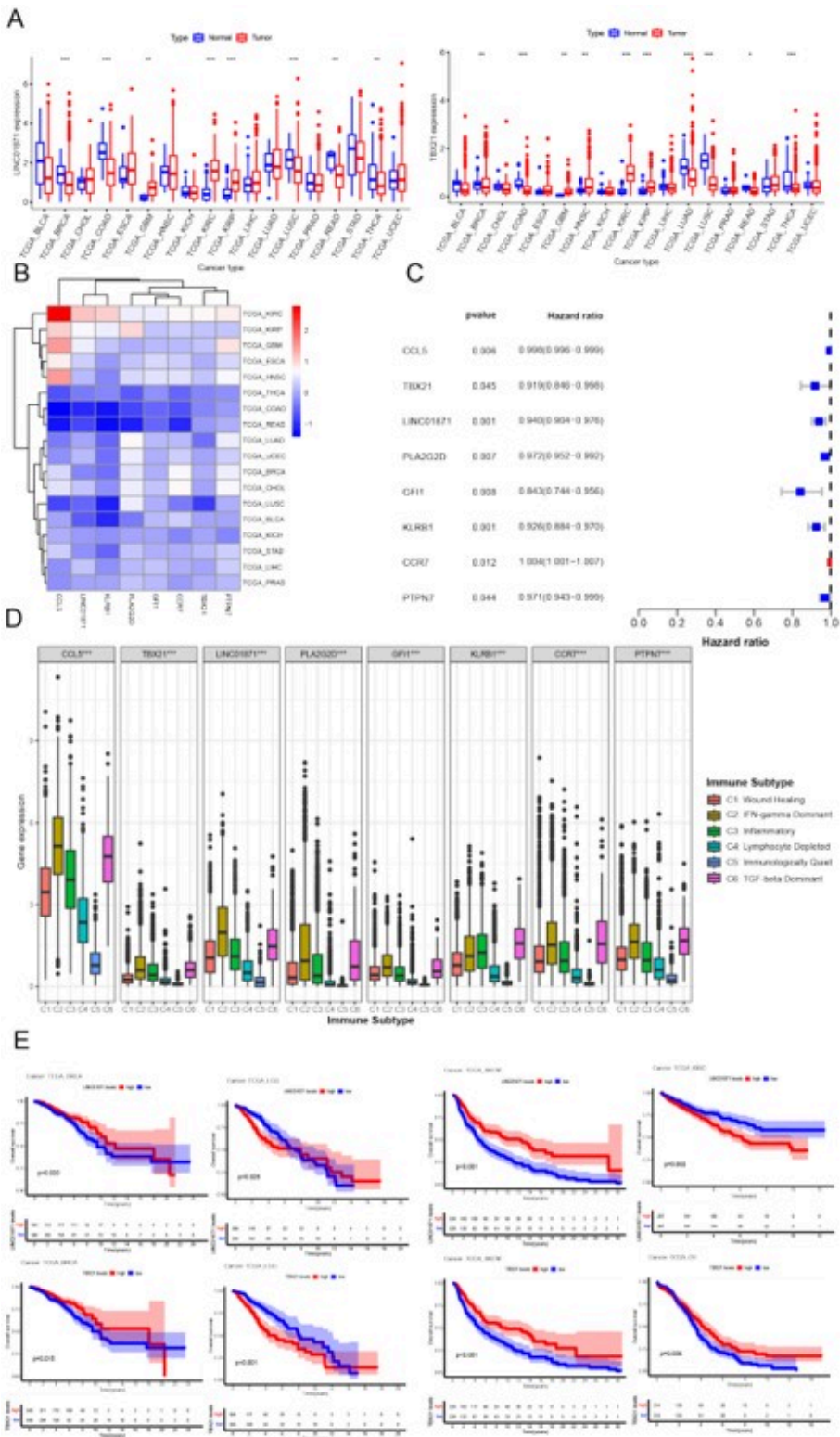


Figure 5

Pan-cancer analysis of the signature genes.

a,b, *LINC01871* (**a**) and *TBX21* (**b**) expression in tumour versus normal tissue across TCGA cancers, with a heatmap of expression by cancer type. **c**, Univariate Cox hazard ratios. **d**, Expression across immune subtypes C1–C6. **e**, Per-cancer Kaplan–Meier survival for *LINC01871* and *TBX21*.

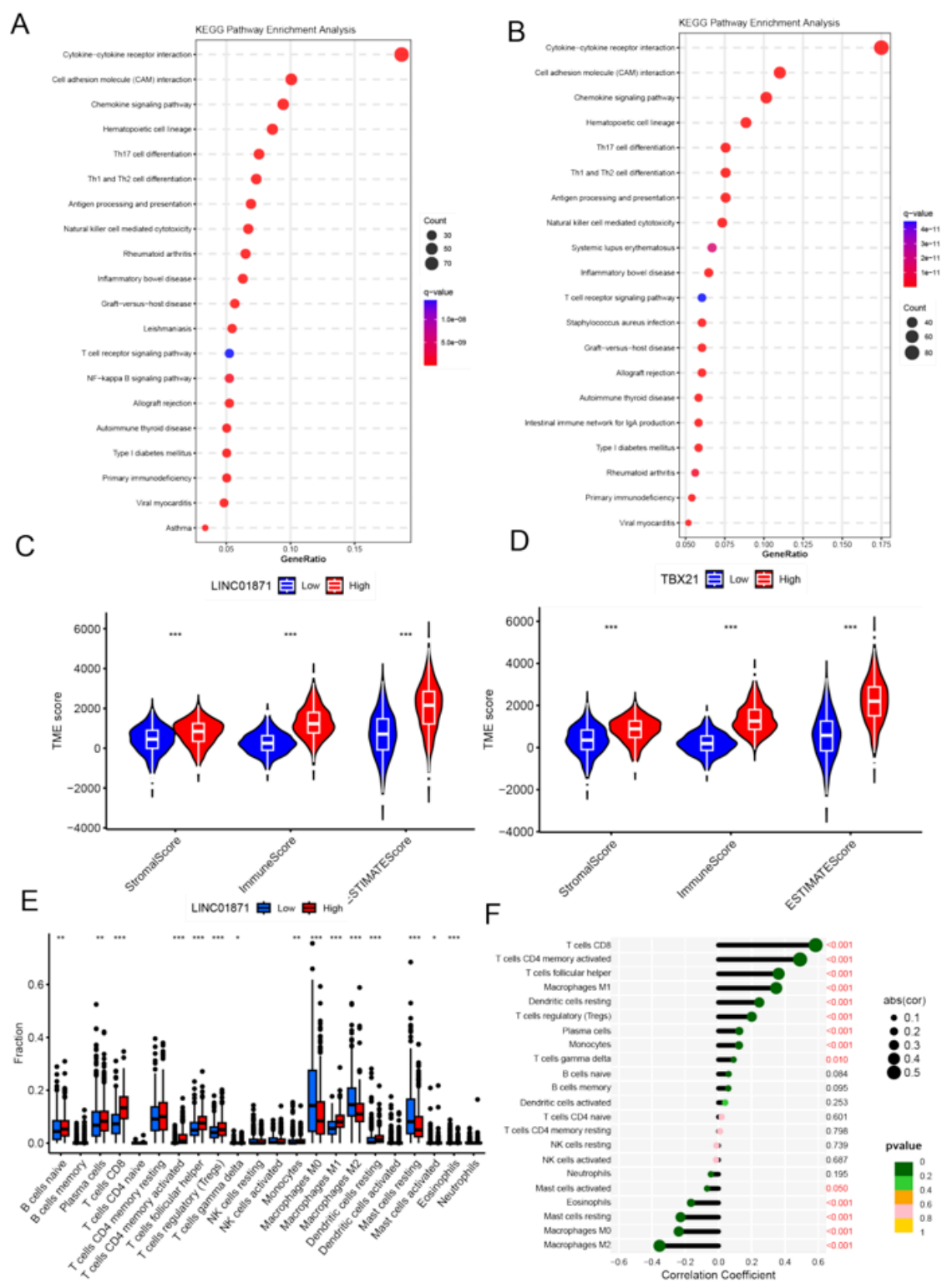


Figure 6

LINC01871 and TBX21 partition the tumour microenvironment.

a,b, KEGG enrichment of genes co-regulated with *LINC01871* and *TBX21*. **c,d**, Stromal, immune and ESTIMATE scores by high/low expression. **e**, Immune-cell fractions by *LINC01871* level. **f**, Correlation of *LINC01871* with immune-cell populations.

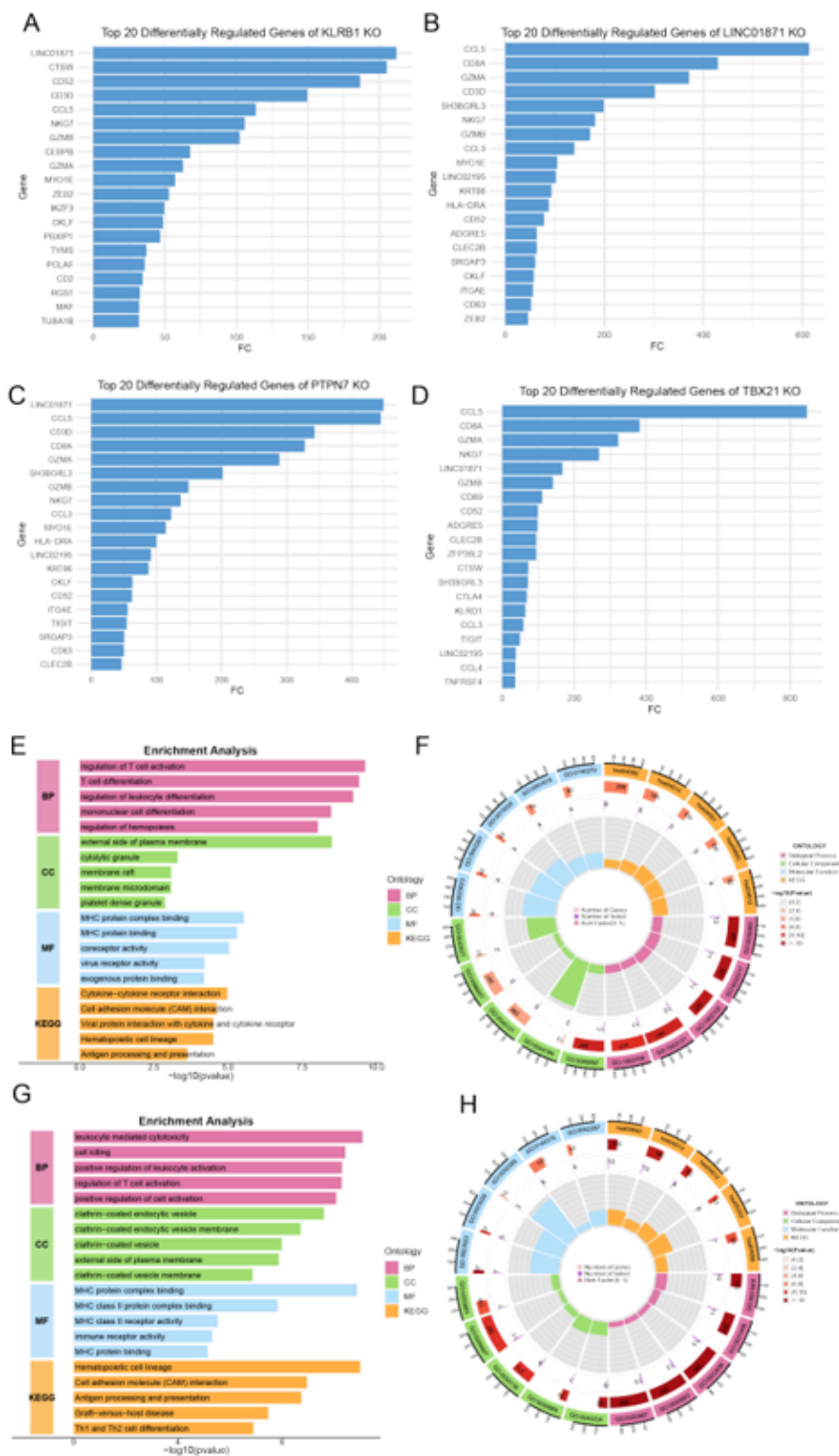


Figure 7

In-silico knockout identifies structural regulators of the T-cell activity network.

a–d, Top 20 differentially regulated genes after virtual knockout of *KLRB1*, *LINC01871*, *PTPN7* and *TBX21* (scTenifoldKnk). **e–h**, GO and KEGG enrichment of the perturbed genes.