npj Precision Oncology

https://doi.org/10.1038/s41698-026-01328-x

Article in Press

Charting cell-type-specific positive genetic interaction at single-cell resolution for lung adenocarcinoma

Received: 4 April 2025

Accepted: 4 February 2026

Cite this article as: Chen, B., Liu, M., Dong, Q. *et al.* Charting cell-type-specific positive genetic interaction at single-cell resolution for lung adenocarcinoma. *npj Precis. Onc.* (2026). https://doi.org/10.1038/s41698-026-01328-x

Bo Chen, Mingyue Liu, Qi Dong, Chen Lv, Kaidong Liu, Huiming Han, Linzhu Wang, Nan Zhang, Wenyuan Zhao, Junjie Lv & Yunyan Gu

Charting cell-type-specific positive genetic interaction at single-cell resolution for lung adenocarcinoma

Bo Chen[1#], Mingyue Liu[1#], Qi Dong[1#], Chen Lv[1], Kaidong Liu[1], Huiming Han[1], Linzhu Wang[1], Nan Zhang[1], Wenyuan Zhao[1], Junjie Lv[2*], Yunyan Gu[1, 3*]

[1] Department of Systems Biology, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China.

[2] Department of Biological Physics, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China.

[3] State Key Laboratory of Frigid Zone Cardiovascular Diseases (SKLFZCD), Harbin Medical University, Harbin, 150081, China.

[#] Contributed equally to this work.

[*] Corresponding authors: Yunyan Gu (guyunyan@ems.hrbmu.edu.cn), or Junjie Lv (lvjunjie525@126.com).

Abstract

Genetic interactions (GIs) drive carcinogenesis and treatment resistance via non-additive phenotypic effects between genes. Traditional bulk-based methods fail to capture cell-type-specific interactions in heterogeneous tumors like lung adenocarcinoma (LUAD), limiting precision oncology. Resolving cell-type-specific GIs at single-cell resolution persists as a major hurdle, hindered by limited analytical methodologies. Here we develop scPGI-finder, a computational framework that identifies gene-pairs whose coordinated high expression is associated with higher proliferation-related fitness at single-cell resolution, which we refer to operationally as single-cell positive genetic interactions (scPGIs). Using scPGI-finder, we identify 49,808 and 15,896 scPGIs spanning epithelial cells and T cells in LUAD, respectively. The predicted scPGIs display tighter junctions in the protein interaction network compared to non-scPGIs. Furthermore, we demonstrate the predictive power of scPGIs for malignancy and immunotherapy response through multi-omics validation across diverse cohorts. Notably, with a mean area under the ROC curve (AUROC) of 0.974 in bulk tissue validation, the epithelial-derived scPGI classifier enables concordant malignancy identification across scales ranging from epithelial single cells and lung cancer cell lines, through spatial transcriptomic maps, to bulk LUAD tissue profiles. Additionally, a six scPGI T cell signature reliably forecasts immunotherapy efficacy, with AUROC values exceeding 0.80 across multiple datasets. Together, our research advances the understanding of underlying cancer positive GIs at the single-cell level. scPGIs of epithelial and T cells serve as robust biomarkers for malignancy evaluation and treatment response, offering a translational framework for precision oncology.

Keywords: Genetic interactions; Single-cell RNA sequencing; Cell-specific network; Tumor malignancy; Immunotherapy.

Introduction

Cellular functions and phenotype deviation are mediated by the functional relationships between genes. Genetic interactions (GIs) are functional relationships between genes, where the phenotypic effect of perturbing one gene is modified by the activity of another[1,2]. GIs, including positive GIs and negative GIs, are crucial for understanding complex diseases and provide novel insight into the personalized treatment[3]. Positive (buffering/alleviating) GIs describe scenarios where the combined phenotype exhibits higher fitness than expected based on the individual phenotypes of the two genes[4]. Conversely, negative (synthetic/amplifying) GIs describe scenarios where the combined phenotype exhibits lower fitness than expected[4,5]. GIs have a critical role in understanding cancer mechanisms and predicting resistance to immunotherapy[6,7]. Contemporary investigations have primarily focused on the GIs within cancer cells using traditional bulk RNA sequencing (RNA-seq) data. However, these studies overlook the interactions that are grounded on specific cell types within the complex tumor microenvironment. The advent of single-cell RNA sequencing (scRNA-seq) facilitates dissecting complex multicellular niches, providing opportunities to discover cell type-specific GIs.

To date, genetic screens and computational approaches have been employed to systematically establish causal links between genotypes and phenotypes[1,8]. Genetic perturbations, such as RNA interference and CRISPR/Cas9 technology, have led to well-established methods for

systematically mapping genetic interactomes across processes affecting cell phenotypes[9-11]. Nevertheless, it is impractical to screen all potential GIs by laboratory-based methods. Numerous computational approaches have been developed for genome-wide prediction of GIs using tumor transcriptomic data, thereby reducing the search space and time required[6,12]. For instance, Jerby-Arnon *et al.* developed a data mining synthetic lethality identification pipeline for predicting negative GIs based on the assumption that negative GIs tend to have co-expression but seldom co-inactivation[13]. Additionally, network-based methods were used to identify GIs by constructing multi-molecular network, such as protein-protein interaction (PPI) network and metabolism network[1]. Moreover, artificial intelligence models have recently emerged as useful methods to infer GIs[1]. Overall, existing methods for GIs inference have primarily relied on certain hypotheses or recognized GI features via traditional RNA-seq data and molecular networks. A significant limitation of these methodologies is their dependence on the aggregate gene expression data derived from heterogeneous cell populations[14]. The inferred GIs are confounded by varying cell-type compositions across samples and are limited to an aggregated view of GIs that may differ considerably across cell types.

Conversely, screening for GIs in different cell types facilitates uncovering cell-type-specific components of genetic networks and elucidating the clinical outcomes underlying these interactions, which remain to be fully explored in cancer research. With scRNA-seq technology development, various computational methods have been employed to infer gene regulatory or co-expression networks, offering a great opportunity for identifying GIs in single cells[15-17]. A computational approach is needed to identify cell-type-specific GIs at single-cell resolution, which

helps to clarify the contribution of tumor heterogeneity, particularly in epithelial cells and T cells. Given that damaged and dead cells are removed during sample preparation in scRNA-seq[18], we focused on dissecting potential positive interactions in **s**pecific **c**ell types. In classical genetics, GIs are defined by non-additive phenotypic effects of combined perturbations. In this work, we use the term "single-cell positive genetic interactions (scPGIs)" operationally to denote gene pairs whose coordinated high expression at single-cell resolution is associated with higher proliferation-related fitness and cell viability, which are correlation-based expression patterns rather than experimentally validated GIs. In this study, cell viability denotes a more favorable cellular state in terms of adaptation and proliferative capacity, reflected by higher proliferation signature scores.

Here, we present a single-cell RNA expression data-driven approach to find scPGIs (scPGI-finder) (Fig. 1A). scPGI-finder delineated the cell-type-specific positive genetic interactions in epithelial cells (E-scPGIs) and T cells (T-scPGIs) from scRNA-seq of lung adenocarcinoma (LUAD). Notably, we show that predicted scGPIs of higher confidence offer novel avenues for predicting malignant epithelial cells, clinical prognosis, and immunotherapy responses (Fig. S1A-B). Our framework extends beyond standard scRNA-seq gene network methods by specifically identifying cell-type-specific positive interactions informed by coordinated high expression enrichment and proliferation-associated signals.

Results

Overview of scPGI-finder at single-cell resolution

scPGI-finder is a computational pipeline for statistically inferring candidate scPGIs from scRNA-

seq data (Fig. 1A). We identified cell-type-specific scPGIs in epithelial and T-cell compartments, representing gene-pairs whose coordinated high expression is associated with increased proliferation. This computational pipeline includes five steps: (i) *Identification of over-expressed genes for the desired cell type*, (ii) *cell ratio test*, (iii) *differential cell proliferation analysis*, (iv) *cell-type-specific positive co-expression inference*, and (v) *functional similarity analysis*. A brief overview is described as follows.

Identification of over-expressed genes in per cell type: Given the sparse data at single-cell resolution, scPGI-finder identifies differentially over-expressed genes for the desired cell types to predict scPGI.

The Salcher *et al.* dataset was used as the discovery dataset to construct E-scPGI and T-scPGI networks via scPGI-finder (Table S1)[19]. After quality control and data preprocessing, 161,041 cells were retained and annotated into eight major cell types: T cells, fibroblasts, endothelial cells, B cells, natural killer (NK) cells, myeloid cells, epithelial cells, and mast cells (Fig. S2A-D; Table S2). The LUAD dataset by Salcher *et al.* includes 31,413 epithelial cells and 61,480 T cells. 2,426 and 1,317 over-expressed genes were identified in epithelial cells and T cells, respectively, for further analysis.

Cell ratio test: The strategy further mines over-represented candidate gene pairs, whose co-activation is significantly more frequent than expected based on their individual activation frequencies, indicating that co-activation of gene pair is under positive selection[20,21].

Differential cell proliferation analysis: scPGI-finder selects a gene pair when the cell group with co-activation of the given pair exhibits higher cell proliferation levels than the cell group

where it is not, for the desired cell types, indicating that the gene pair is likely to increase cell fitness when co-active[22,23].

Cell-type-specific positive co-expression inference: Pairwise gene co-expression analysis is based on functional similarity, where GI pairs tend to have similar functions and hence likely to be co-expressed in specific cell types[6,13]. Gene pairs with a significant positive correlation were considered as potential scPGIs in the specific cell type.

To compare alternative implementations of co-expression inference, we evaluated four models (Pearson, CS-CORE, MAST and Dozer) using 220 experimentally validated SV gene pairs[16,24,25]. Although SV pairs derive from bulk or cell-line perturbations and are not a direct gold standard for single-cell co-activation, they provide a consistent reference for comparing model behavior. Across epithelial and T cell subsets, CS-CORE and MAST achieved higher area under the ROC curve (AUROC) and recall relative to other methods (Fig. S2E-F). Computational benchmarking showed lower runtime for CS-CORE, although more computational memory was needed (Fig. S2G-H). Based on this balance of performance and efficiency, CS-CORE was selected as the default implementation in scPGI-finder.

Functional similarity analysis: Two genes with GIs typically participate in close biological processes. Consequently, the positions of two genes with GIs in the GO topological structure should be close[13]. We defined FS score based on similarities in BP, MF and CC, and identified gene pairs with high functional similarity.

Gene pairs that passed all the procedures described above were taken as candidate scPGIs and utilized for following analysis.

The landscape of scPGIs in epithelial cells and T cells for LUAD

In epithelial cells, we identified 49,808 E-scPGIs among 2,068 genes in the LUAD dataset via scPGI-finder (Fig. S3A, Table S3). To assess sensitivity to sampling depth, we re-ran scPGI-finder on random subsets of 500-25,000 epithelial cells. The recovered scPGIs showed increasing overlap with the full network as cell numbers increased (Table S4), indicating that stronger interactions are progressively captured with larger sampling.

The E-scPGIs network exhibited a scale-free property (Fig. S3B). Notably, 392 genes were identified as central due to consistent overlap across five distinct network metrics (Fig. S3C). The central genes quantified distinct types of biological importance in E-scPGIs network, significantly enriched in pathways related to epithelial cell development and protein processing, including protein processing in the endoplasmic reticulum and p53 signaling pathway, which plays a paramount role in tumor suppression ($P < 0.05$, Hypergeometric test; Fig. S3D). In coordinated high expression based scPGI networks, edges do not represent directional information flow, so shortest-path-based metrics such as betweenness and closeness have no clear mechanistic interpretation, whereas degree provides a more straightforward measure of a gene's involvement in scPGI programs. *MLF2* showed the top degree in E-scPGIs network (Fig. 1B). The partner genes of *MLF2* were enriched in various transcription pathways, including the spliceosome, and transcription machinery pathways ($P < 0.05$, Hypergeometric test; Fig. 1C).

In T cells, scPGI-finder obtained a network of 15,896 T-scPGIs among 1,117 genes (Fig. S3E), which exhibited scale-free characteristics (Fig. S3F). The 181 central genes were significantly

enriched in multiple immunological pathways, such as Th1 and Th2 cell differentiation, Th17 cell differentiation, T cell receptor signaling pathway, and NK cell mediated cytotoxicity (Fig. S3G-H). The interactions formed by immune genes may play a critical role in immune regulation within the tumor microenvironment and in maintaining T cell viability. In T-scPGIs network, the top 20 genes with the highest degree included *PSMD8*, *UBE2V1,* and *SUMO1* (Fig. 1D). Most notably, the partners of *CTLA4*, a target of immunotherapy and essential for T cell activation, were enriched in immune-related pathways, such as CD molecules and TCR pathway ($P < 0.05$, Hypergeometric test; Fig. 1E). These results may indicate that the complex positive GIs between immune genes were crucial for T cell viability.

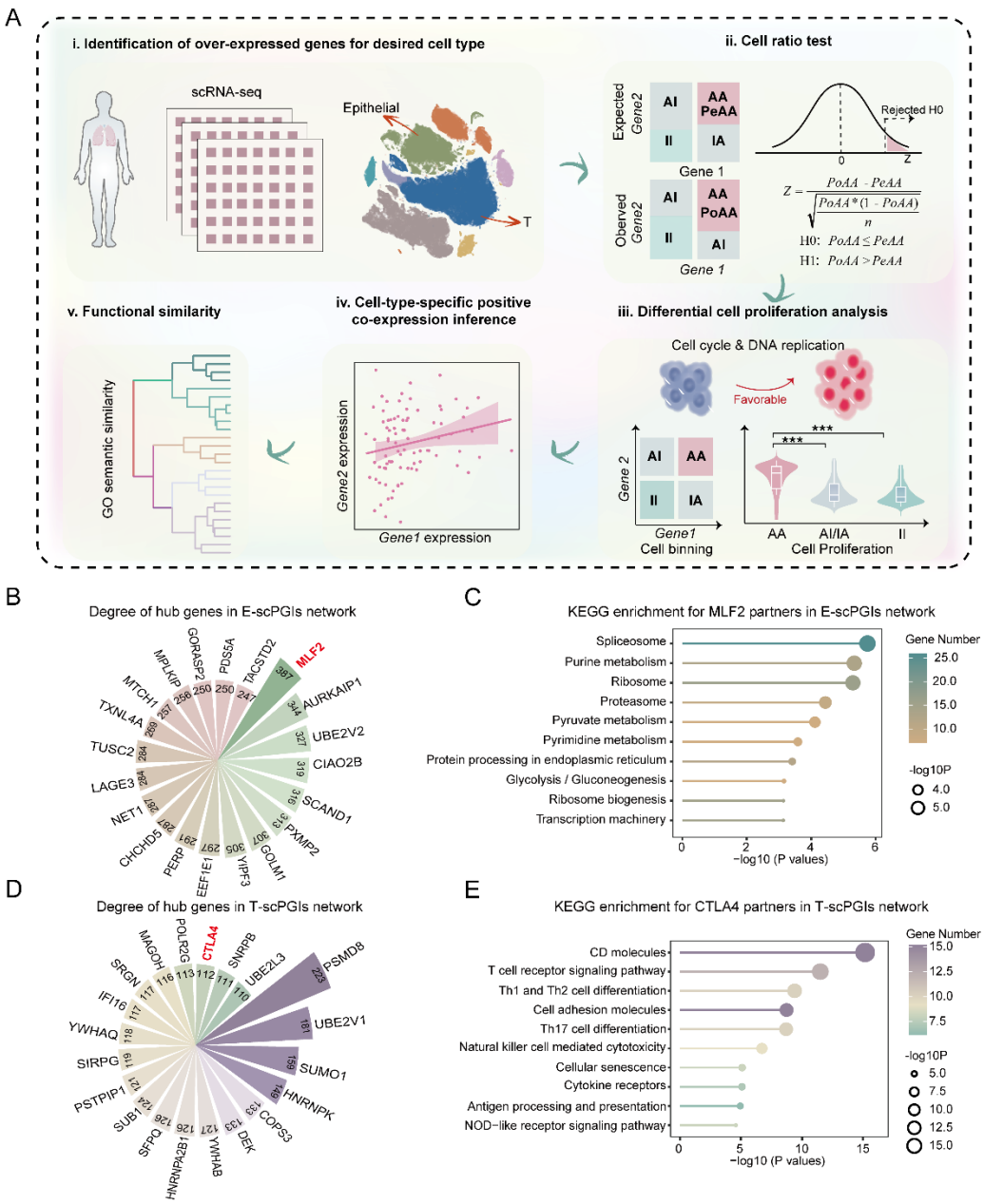

Fig. 1. Workflow of scPGI-finder and hub-centered scPGI networks in LUAD epithelial cells and T cells.

(A) Workflow of scPGI-finder. (B) Degree distribution of hub genes in the E-scPGIs network. (C) KEGG enrichments of the *MLF2* partners in the E-scPGIs network. (D) Degree distribution of hub genes in the T-scPGIs network. (E) KEGG enrichments of the *CTLA4* partners in the T-scPGIs

network. $P < 0.05$ was considered as statistically significant in C and E.

Validation of scPGIs via well-recognized GI features and cell-type-specific functions

To assess the reliability of scPGIs, we collected multiple well-recognized features for GIs, including comparative, and boolean (true/false) properties (Fig. S1A; Table S5). We randomly selected an equal number of gene pairs as non-E-scPGIs (non-T-scPGIs), composed of genes from E-scPGIs (T-scPGIs). Compared to non-E-scPGIs, E-scPGIs had more shared and union neighbor genes in the PPI network, as well as more shared and union pathways ($P < 2.2e-16$, Two-sided Wilcoxon rank-sum test; Fig. 2A). E-scPGI pairs also displayed higher essentiality (lower gene effect score) of shared PPI, union PPI, and the sum of two genes in E-scPGIs, along with higher subcellular colocalization scores and co-expression coefficients than non-E-scPGIs ($P < 2.2e-16$, Two-sided Wilcoxon rank-sum test; Fig. 2A). In addition, the genes in E-scPGIs tend to have a lower average shortest distance in the PPI network ($P < 2.2e-16$, Pearson's Chi-squared test; Fig. 2B).

To further evaluate the correlation between scPGIs and cell characteristics, we extracted the hub genes with the top 20 highest degrees and its partners in the scPGIs network, forming the hub subnet. The hub subnet score, defined as an aggregate score of the subnet for each hub gene and their partners, was used to evaluate the cell viability. We then calculated the correlation between the scores of these 20 subnets and cell-type-specific function scores (see Methods). In epithelial cells, the subnet scores showed a significantly positive correlation with the activity of specific biological processes relevant to epithelial cells and cancer hallmarks ($P < 0.05$, Spearman's rank

correlation test; Fig. 2C). Furthermore, epithelial cells with high subnet scores exhibited higher CytoTRACE scores ($P < 2.2e-16$, Two-sided Wilcoxon rank-sum test; Fig. 2D), indicating lower differentiation levels and higher stemness. Functional reproducibility was performed using the independent Zhou *et al.* dataset. In agreement with Salcher *et al.*, the subnet scores were positively correlated with cancer hallmark activity and showed similar changes in differentiation levels in an independent LUAD dataset by Zhou *et al.* (Fig. S4A-F)[26]. Additionally, the subnet scores in epithelial cells derived from advanced-stage samples were higher than those from early-stage samples in the Salcher *et al.* dataset ($P < 0.05$, Two-sided Wilcoxon rank-sum test; Fig. 2E).

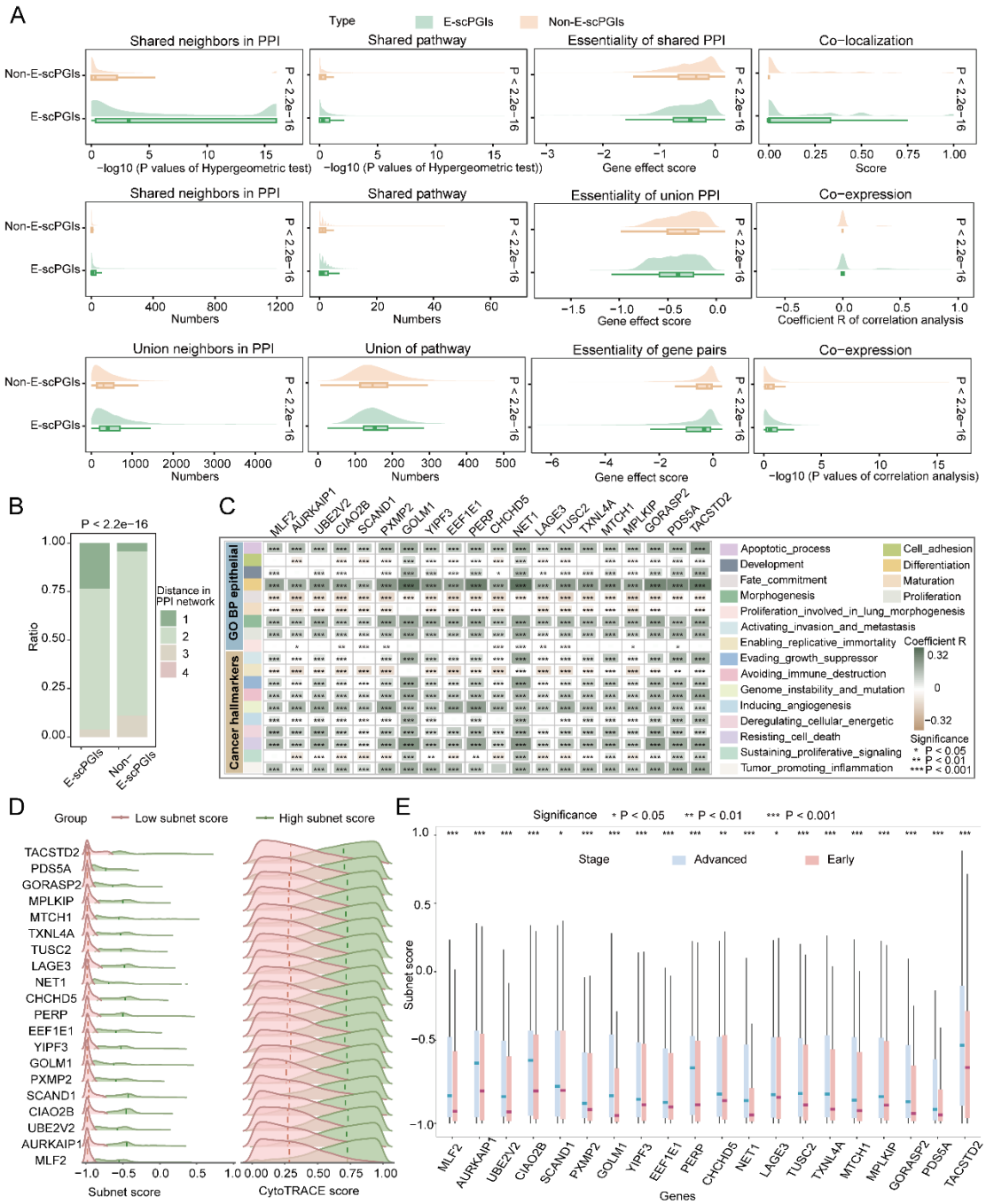

Fig. 2. Assessing the predictive power of E-scPGIs using various recognized GIs properties.

(A-B) Significant differences in multiple features between E-scPGIs and non-E-scPGIs gene pairs. Gene effect score represents CRISPR-based gene essentiality from DepMap (lower values indicate stronger dependency). (C) Association between the subnet score of the top 20 hub genes in E-

scPGIs and the epithelial cell specific functions or tumor hallmarks in Salcher *et al.*, calculated by Spearman's rank correlation test. (D) Distribution of CytoTRACE scores between cells with high and low subnet scores for top 20 hub genes in Salcher *et al.*. (E) Distribution of the subnet score for the top 20 hub genes between the epithelial cells in early and advanced stages in Salcher *et al.*. *P* values were calculated by Two-sided Wilcoxon rank-sum test in (A, D and E). Pearson's chi-squared test was performed in B. *P* values were calculated by Spearman's rank correlation test in C. $P < 0.05$ was considered as statistically significant. For *P* value, * denotes $P < 0.05$, ** denotes $P < 0.01$, and *** denotes $P < 0.001$.

In accordance with E-scPGIs, T-scPGIs exhibited a greater number of shared and union neighbors in PPI network, shared and union pathways, sub-cellular co-localization scores, and co-expression coefficients than non-T-scPGIs ($P < 0.001$, Two-sided Wilcoxon rank-sum test; Fig. 3A). Correspondingly, the genes in T-scPGIs tend to have a lower average shortest distance in the PPI network compared to non-T-scPGIs ($P < 2.2e-16$, Pearson's Chi-squared test; Fig. 3B).

Overall, the candidate E-scPGIs and T-scPGIs typically displayed well-known GIs characteristics. Particularly, compared with matched non-scPGI gene pairs, scPGIs showed modest but consistent increases in shared/union PPI neighbors, shared/union pathway membership, co-expression, essentiality correlations, and shorter PPI distances. Although effect sizes are small, the convergent enrichments across multiple independent functional dimensions indicate that scPGIs capture subtle but reproducible functional coherence.

In addition, the top 20 subnet scores were significantly positively correlated with T cell-related functions in T-scPGIs in the discovery (Salcher *et al.*) and independent validation dataset

(Zhou *et al.*), such as the T cell biological processes of activation, activation involved in immune response and activation via TCR contact with antigen bound to MHC molecule on antigen presenting cell ($P < 0.05$, Spearman's rank correlation test; Fig. 3C and S4G). Notably, T cells with higher subnet scores exhibited higher CytoTRACE scores ($P < 0.001$, Two-sided Wilcoxon rank-sum test; Fig. 3D and S4H) and displayed higher TCR clones ($P < 0.05$, Two-sided Wilcoxon rank-sum test; Fig. 3E). The above results reveal the association between T-scPGIs and T cell functions, suggesting the potential of T-scPGIs in assessing T cell function and predicting immunotherapy response.

We further compared scPGI hub subnet scores with cell-type-specific functional modules using matched non-scPGI hubs as controls. For each scPGI hub in two cell types, we randomly sampled the same number of genes from the same gene pool to construct these control hubs. In both Salcher *et al.* and Zhou *et al.* datasets, scPGI hubs showed consistently higher correlations with cell-type-specific functional modules than matched non-scPGI hubs, supporting a closer alignment of scPGIs with epithelial- and T-cell-specific biological programs (Fig. S5).

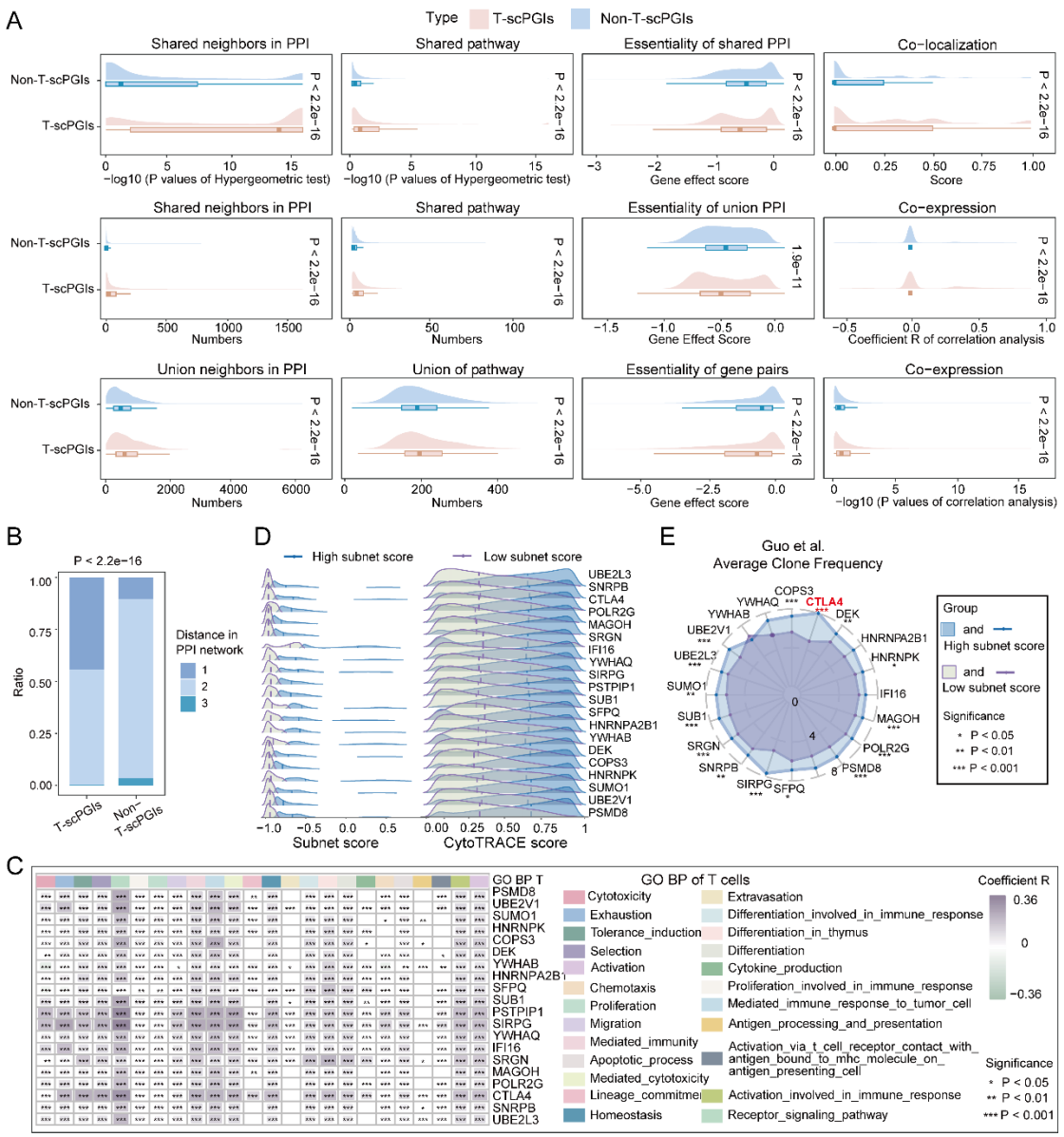

Fig. 3. Assessing the predictive power of T-scPGIs.

(A-B) Significant differences in multiple features between T-scPGIs and non-T-scPGIs gene pairs. (C) Association between the subnet score of the top 20 hub genes in T-scPGIs and T cell-specific functions in the study by Salcher *et al.*, calculated by Spearman's rank correlation test. (D) Distribution of CytoTRACE scores between cells with high and low subnet scores for the top 20 hub genes in the study by Salcher *et al.*. (E) Distribution of the average clone frequency between

T cells with high and low subnet scores for top 20 hub genes in Guo *et al.*. *P* values were calculated by Two-sided Wilcoxon rank-sum test in (A, D and E). Pearson's chi-squared test was performed in B. *P* values were calculated by Spearman's rank correlation test in C. $P < 0.05$ was considered as statistically significant. For *P* value, * denotes $P < 0.05$, ** denotes $P < 0.01$, and *** denotes $P < 0.001$.

Prediction of malignant epithelial cells using an E-scPGI-based classifier

Considering that the subnet scores of hub genes in E-scPGIs were significantly positively correlated with cancer hallmarks, we investigated whether E-scPGIs could provide an alternative methodology for evaluating tumor malignancy (Fig. 2C and S4E). Given the association between cancer and large-scale chromosomal alterations, we first utilized CNV inferred from RNA expression to classify epithelial cells into malignant and non-malignant epithelial cells in the LUAD dataset. Compared to control cells (fibroblasts and endothelial cells) and non-malignant epithelial cells (n = 8,872), malignant epithelial cells (n = 22,541) displayed larger changes in relative expression intensities across the genome (Fig. S6A-C). Next, we identified 423 ME-scPGIs that serve as a hallmark to gauge the malignancy of epithelial cells. ME-scPGIs exhibited a more highly co-active expression pattern in the malignant cells compared to non-malignant cells ($P < 0.05$, Pearson's Chi-squared test; Fig. 4A; See Methods). Classifier performance in predicting malignant cells was assessed by scoring epithelial cells using the ME-scPGIs. Remarkably, malignant cells showed significantly higher ME-scPGI scores than non-malignant cells ($P < 2.2e-16$, Two-sided Wilcoxon rank-sum test; Fig. 4B-C). We further assessed the predictive

performance of the ME-scPGI scores by analyzing the ROC and computing the AUROC. The ME-scPGI scores demonstrated predictive power in both the training LUAD dataset and an independent dataset (AUROC in Salcher *et al.*: 0.707; AUROC in Zhou *et al.*: 0.802; Fig. 4D-E, and S6D-F). Additionally, epithelial cells with higher ME-scPGI scores (epithelial-H) had more and stronger interactions with other cell types than those with lower ME-scPGI scores (epithelial-L), in both sender and receiver roles (Fig. 4F-I). Moreover, we investigated specific communications in epithelial-H cells. In the study by Salcher *et al.*, interactions involving MIF signaling, including *CD74-CXCR4* and *CD74-CD44*, were specifically present in epithelial-H cells, indicating distinct communications in the highly malignant cells (Fig. S7A-B).

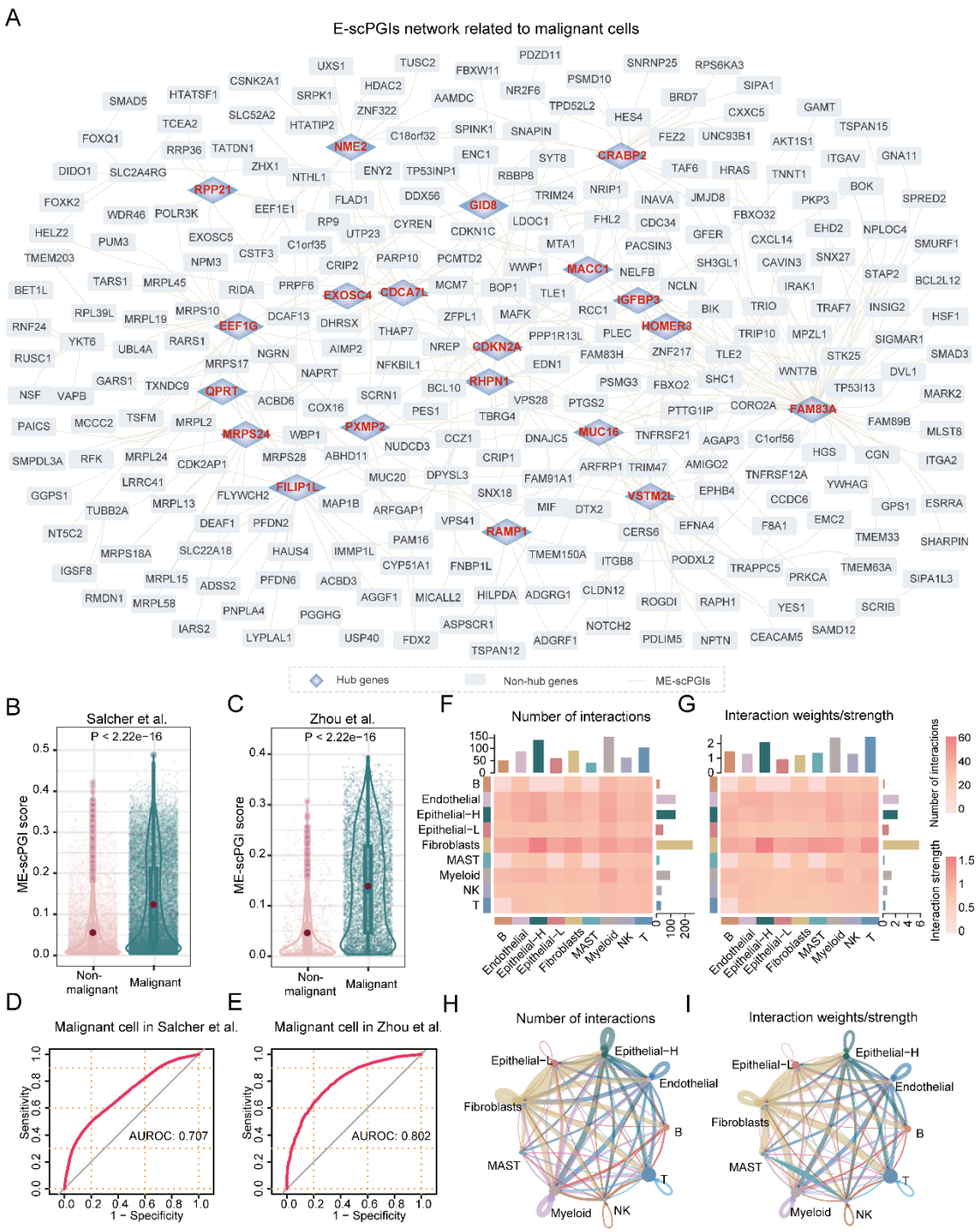

Fig. 4. Malignant epithelial cells prediction by an E-scPGIs-based classifier.

(A) Malignant epithelial cell-specific GIs network derived from E-scPGIs. (B-C) Comparison of the scores of malignant epithelial cells-specific GIs between cells with low and high copy-number variation levels in the discovery (Salcher *et al.*) and independent validation (Zhou *et al.*) datasets.

(D-E) ROC curve for ensemble classifiers assessing the malignancy of epithelial cells in Salcher *et al.* and Zhou *et al.*datasets. (F-I) Communications among epithelial-H, epithelial-L, and other cell subtypes in Salcher *et al.*. *P* values were calculated by Two-sided Wilcoxon rank-sum test in D and E.

Benchmarking ME-scPGI against expression- and CNV-based classifiers

Across both the discovery (Salcher *et al.*) and independent validation (Zhou *et al.*) datasets, the ME-scPGI classifier outperformed hallmark-based expression signatures in distinguishing malignant from non-malignant epithelial cells (Fig. S8A-D). ME-scPGI scores were higher in aneuploid epithelial cells than in diploid cells (Fig. S8E-F). These results suggest that ME-scPGI captures a malignant epithelial program that provides additional discriminatory value beyond expression-based signatures. Although performance differences are moderate, the ME-scPGI classifier adds discriminatory information beyond expression-based signatures and serves as a more computationally efficient alternative to CNV-based approaches.

To prioritize ME-scPGI associated surface markers for malignant epithelial cells, we performed differential expression restricted to GO_CELL_SURFACE genes in both the discovery and validation scRNA-seq datasets. We identified a consensus set of 36 surface-associated genes consistently upregulated in epithelial cells with higher ME-scPGI scores (adjusted $P < 0.05$, avg_log2FC > 0.1 and pct.1 > 0.2; Fig.S9), including *EPCAM*, *SDC1*, *CD151*, *ITGB4*, *TPBG* and *ADAM9*. Their expression patterns define an in-silico panel of candidate surface markers associated with the ME-scPGI program.

Validating the ME-scPGI classifier in lung cancer patients

Subsequently, we investigated the association between ME-scPGIs and tumor malignancy in bulk datasets. To characterize the pathways represented by the ME-scPGI program, we examined ME-scPGI network genes in eight independent LUAD bulk transcriptomic datasets with paired tumor and normal samples (Table S6)[27-36]. Nineteen ME-scPGI genes were consistently downregulated and enriched for C-type lectin receptors, NF-κB and TNF signaling and endocytosis (Fig. S10). 154 ME-scPGI consistently upregulated genes and enriched for pathways related to ribosome biogenesis, cell cycle and ECM–receptor interaction, consistent with enhanced proliferative and biosynthetic activity in tumors (Fig. S10). The remaining 68 ME-scPGI genes without differential expression formed a distinct module enriched for Wnt, Notch, and pluripotency-related pathways (Fig. S10), indicating that part of the ME-scPGI program resides within broadly active epithelial regulatory circuits rather than being solely driven by differential expression.

Most of the top 20 hubs in the ME-scPGI network exhibited elevated expression in tumor tissues compared to adjacent tissues (Fig. 5A). We calculated the ME-scPGI scores in bulk RNA-seq datasets to determine the malignancy of tissues. We found augmented levels of ME-scPGI scores in tumor tissues than in adjacent tissues ($P < 0.001$, Two-sided Wilcoxon rank-sum test; Fig. 5B). Additionally, the unsupervised ME-scPGI-based classifier exhibited robust discriminative ability, with a mean AUROC of 0.974 in distinguishing tumor tissue from adjacent tissue across all eight lung cancer datasets (Fig. 5C). By assessing the association of the ME-scPGI scores and immune cell infiltration by CIBERSORT, ESTIMATE, and xCell, we revealed that ME-scPGI

scores were significantly positively correlated with tumor purity in estimate, and significantly negatively correlated with the infiltration of several activated immune cells, such as $CD8^+$ T-cells in xCell ($P < 0.05$, Spearman's rank correlation test; Fig. 5D-E). Next, we investigated whether the ME-scPGI scores could provide an alternative methodology for prognosis stratification[37-40]. Patients were divided into two groups based on the ME-scPGI scores. As expected, Kaplan-Meier curves showed patients with high ME-scPGI scores had a poorer prognosis compared to low ME-scPGI scores (Tomida *et al.*: $P = 0.010$; Tang *et al.*: $P = 0.130$; Rousseaux *et al.*: $P = 0.013$; Schabath *et al.*: $P < 0.001$; TCGA_LUAD: $P = 0.001$, Log-rank test; Fig. 5F; Table S6). Patients with high ME-scPGI scores had poorer overall survival (OS). Overall, the ME-scPGI scores served as an indicator of tumor malignancy and a worse prognosis.

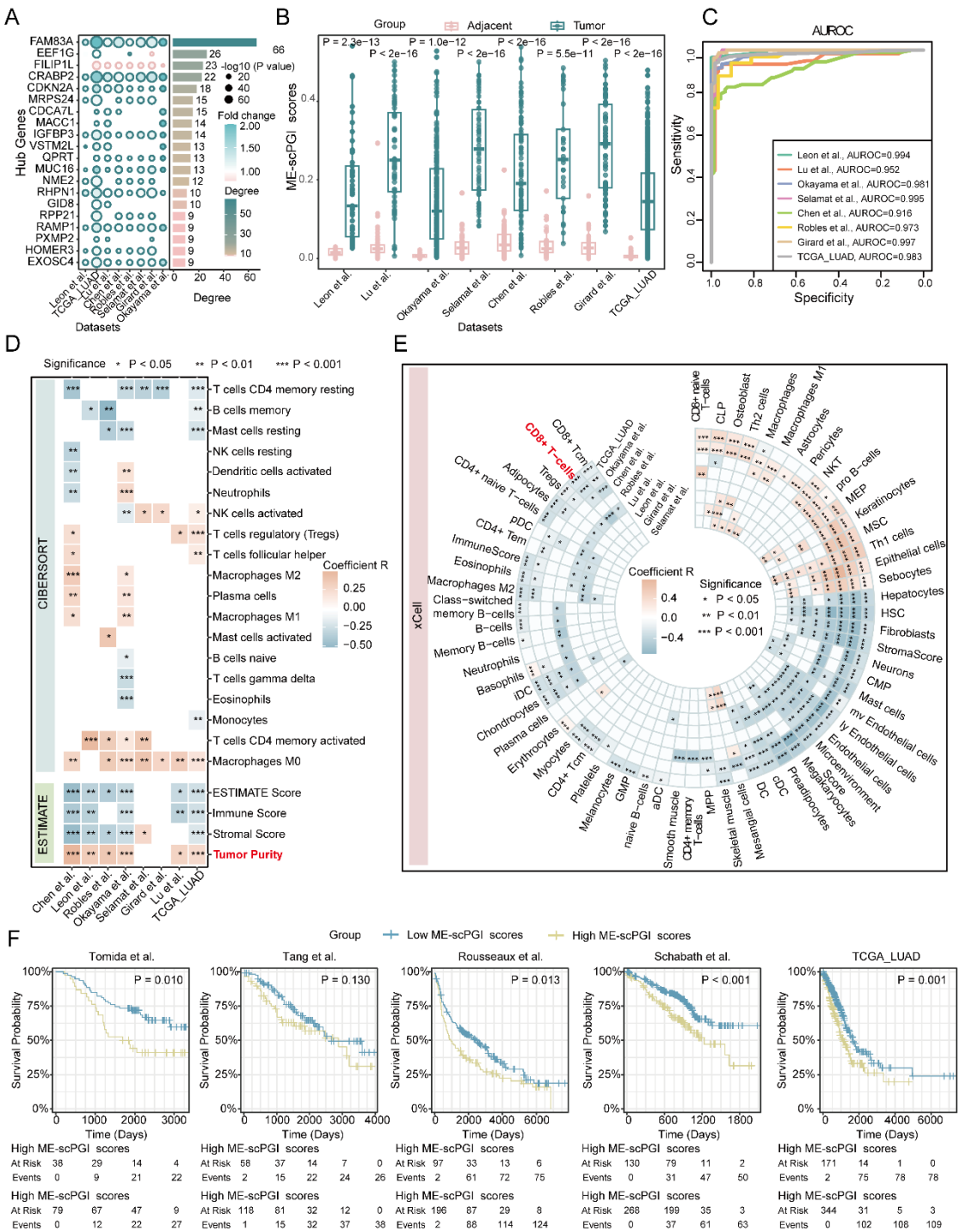

Fig. 5. ME-scPGIs stratifies patients across multiple lung cancer datasets.

(A) Significant differential expression of the top 20 hub genes in ME-scPGIs between tumor and adjacent normal tissues. (B) Comparison of ME-scPGI scores in ME-scPGIs between tumor and

adjacent normal tissues. (C) ROC curve for the ensemble classifier assessing the malignancy of tumor tissues across multiple lung cancer datasets. (D) Association between ME-scPGI scores and tumor immune microenvironment using CIBERSORT and ESTIMATE. (E) Association between ME-scPGI scores and tumor immune microenvironment in xCell. (F) Kaplan Meier plots for lung cancer patients with low and high ME-scPGI scores (threshold was upper terciles of scores). The *P* values were calculated by Log-rank test. *P* values were calculated by Two-sided Wilcoxon rank-sum test in A and B. *P* values were calculated by Spearman's rank correlation test in D and E. For *P* value, * denotes $P < 0.05$, ** denotes $P < 0.01$, and *** denotes $P < 0.001$.

To evaluate whether the bulk-level prognostic association of ME-scPGI could be explained solely by tumor purity or immune composition, we performed multivariate Cox regression in TCGA-LUAD and Schabath *et al.* cohorts. We fitted three types of models: (i) ME-scPGI plus tumor purity; (ii) ME-scPGI, tumor purity, and major CIBERSORT immune populations; and (iii) ME-scPGI, tumor purity, and immune populations that were significant in univariate Cox analysis. Across all models and in both cohorts, ME-scPGI remained significantly associated with overall survival ($P < 0.05$) with stable hazard ratios, whereas tumor purity lost significance once ME-scPGI was included, and adding immune covariates did not attenuate its effect (Fig. S11). These findings indicate that ME-scPGI provides prognostic information beyond epithelial abundance or immune depletion.

Spatial transcriptomic support for ME-scPGI patterns

To explore whether ME-scPGI captures spatially structured malignant epithelial programs, we

analyzed LUAD spatial transcriptomic datasets from the Ad-SpatialAnalysis repository[41]. Across sections, ME-scPGI scores with well-differentiated, proliferative and invasive signatures exhibited non-random spatial organization, and in several samples ME-scPGI high regions coincided with poorly differentiated, more proliferative/invasive tumor areas (Fig. 6A). We further examined the spatial distribution of *FAM83A*, the highest-degree gene in the ME-scPGI network, and its top partners. *FAM83A* showed strong spatial correlation with the hub partners (Fig. 6B).

In both two representative sections (FFPE_LUAD_No_2C and FFPE_LUAD_No_2D), ME-scPGI-H regions overlap with low well-differentiated scores and locally elevated proliferative/invasive signal. *FAM83A*, the top-degree ME-scPGI gene, displayed strong spatial co-variation with *MACC1* and *MUC16*, whereas *CRABP2* localized to a distinct region, indicating heterogeneous but structured spatial coordination among ME-scPGI partners (Fig. 6C).

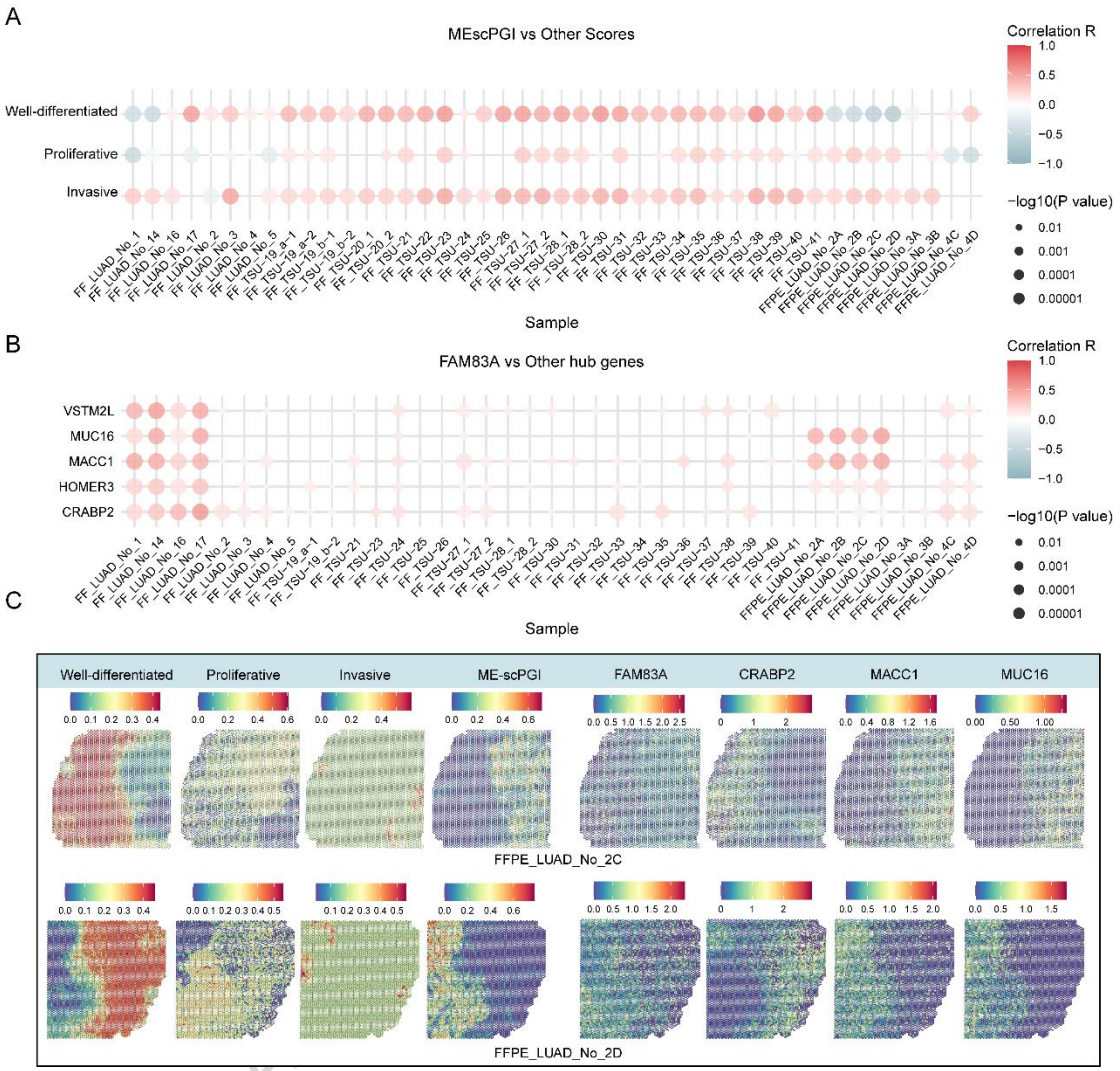

Fig. 6. Spatial validation of ME-scPGI patterns and ME-scPGI hub gene relationships in LUAD spatial transcriptomics.

(A) Correlation between ME-scPGI scores and published differentiation programs across LUAD spatial transcriptomic samples. **(B)** Correlation between FAM83A and its ME-scPGI hub partners across LUAD spatial sections. **(C)** Representative spatial maps showing ME-scPGI and related programs in two LUAD sections (FFPE_LUAD_No_2C and FFPE_LUAD_No_2D). Spearman correlations were computed in A and B. Only significant correlations ($P < 0.05$ and $|r| > 0.3$) are shown. Dot size reflects –log10(P value); dot color encodes correlation coefficients (red = positive,

blue = negative).

Identification of resistant and sensitive drugs for high malignancy tumor cells

Furthermore, we examined the relationship between ME-scPGI scores and drug efficacy in lung cancer cells based on four pharmacological screening datasets ($P < 0.05$, Spearman's rank correlation test; Fig. 7A; Table S7). Eight drugs demonstrated significant correlations with ME-scPGI scores in at least two of the four datasets (Fig. 7B). The response to six drugs was negatively correlated (Afatinib, Gefitinib, Pelitinib, Wz8040, P22077, and Tosedostat; Fig. 7C-H) and two drugs were positively correlated with ME-scPGI scores (Methotrexate and Oxaliplatin; Fig. 7I-J). Notably, the response to three *EGFR* inhibitors, including afatinib, gefitinib, and pelitinib, demonstrated a negative correlation with ME-scPGI scores (Fig. 7C-E). These findings suggested that cancer cell lines with higher ME-scPGI scores, corresponding to higher malignancy, may benefit from EGFR inhibitors.

To clarify the molecular context of the observed association between ME-scPGI scores and EGFR-inhibitor response, we compared genomic alterations between ME-scPGI-high and ME-scPGI-low cell lines in GDSC1 and GDSC2. *EGFR* mutations were significantly enriched in the ME-scPGI-high group, and copy-number gains of *MYC* and *POU5F1B*, characteristic of proliferative LUAD subtypes, were also more frequent in this group (Fisher's exact test, $P < 0.05$; Fig. S12). Importantly, ME-scPGI-high cell lines included both *EGFR*-mutant and *EGFR* wild-type models, indicating that the signature is broader than EGFR mutation status alone. *KRAS* and *SMARCA4* mutations tended to be more frequent in the ME-scPGI-low group, although without

statistical significance. Taken together, these patterns suggest that the ME-scPGI signature reflects an epithelial, proliferation-associated LUAD subtype that only partly overlaps with EGFR-driven disease.

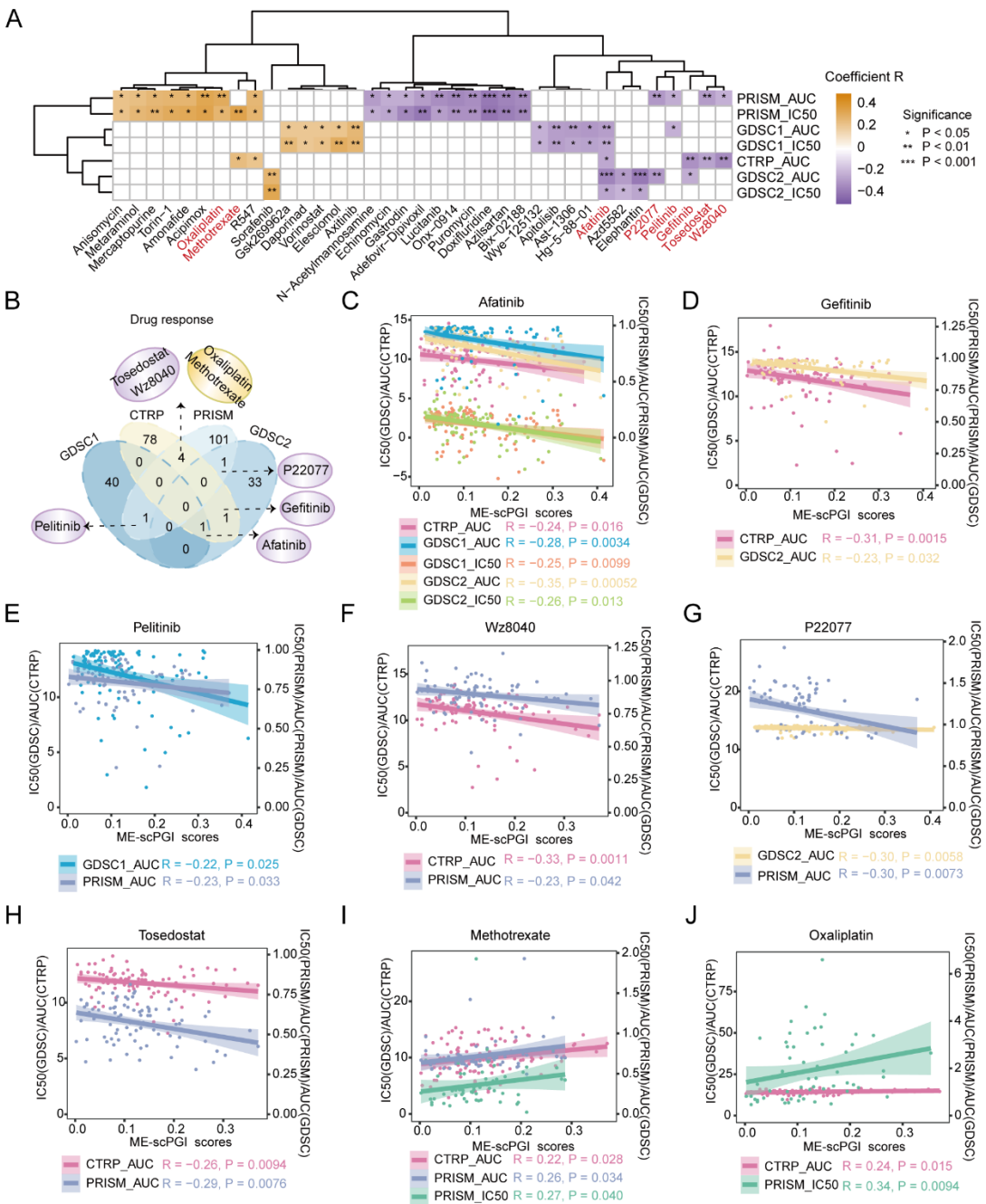

Fig. 7. ME-scPGI score indicates EGFR inhibitors response in lung cancer cells.

(A) Association between ME-scPGI scores and drug efficiency in lung cancer cells across GDSC1, GDSC2, CTRP and PRISM datasets. Rows (datasets) and columns (drugs) were hierarchically clustered to reveal groups of compounds with similar correlation patterns with ME-scPGI activity; colors indicate Spearman correlation coefficient. (B) Overlap of resistant and sensitive drugs for cells with higher ME-scPGI scores in different datasets. (C-H) Negative association between ME-scPGI scores and AUC/IC50 of various drugs, including Afatinib, Gefitinib, Pelitinib, Wz8040, P22077 and Tosedostat. (I-J) Positive association between ME-scPGI scores and AUC/IC50 of Methotrexate and Oxaliplatin. *P* values were calculated by Spearman's rank correlation test in A and C-J. $P < 0.05$ was considered as statistically significant. For *P* value, * denotes $P < 0.05$, ** denotes $P < 0.01$, and *** denotes $P < 0.001$.

Prediction of immunotherapy efficacy using T-scPGI

The notable correlation between T-scPGIs and T cell functions may prove to be a valuable methodology for assessing T cell viability and predicting immunotherapy responses (Fig. 3C-E). Next, we explored the ability of T-scPGIs to predict clinical response to immune checkpoint blockade (Fig. S1B; Table S8). To identify the T-scPGIs that are predictive of immunotherapy efficacy, we analyzed scRNA-seq data from three pre-treatment and twelve post-treatment patients with NSCLC who received neoadjuvant PD-1 blockade combined with chemotherapy. As expected, the ratio of T cells changed across distinct groups, including pre-treatment or post-treatment samples with PR and SD groups (Fig. 8A, and S13A-C). A comparison of pre-treatment samples with PR to post-treatment surgery samples with PR revealed an increase in the ratio of T cells,

indicating that T cells had been activated by immunotherapy (Fig. 8A). We identified 516 T-scPGIs related to immunotherapy response and 330 T-scPGIs related to immunotherapy resistance by analyzing the scRNA-seq and RNA-seq data of T cells from pre-treatment biopsy samples (See Methods). The hub genes with the top 50 degrees involved in SD-related T-scPGIs were enriched in transcription machinery and the spliceosome pathway (Fig. 8B). Interestingly, the hub genes involved in PR-related T-scPGIs were enriched in multiple immune-related pathways, particularly in CD molecules and the TCR signaling pathway (Fig. 8C).

Two prediction models for immunotherapy response were constructed using T-scPGIs involved in CD and TCR pathways related to PR, based on bulk data from Hu *et al.*[42]. These models were named CD-scPGI and TCR-scPGI, respectively. Importantly, the TCR-scPGI scores of responders were significantly higher than non-responders[42-45] (P = 4.83e-03, P = 1.65e-03, P = 5.51e-03, P = 9.45e-04, Two-sided Wilcoxon rank-sum test; Fig; 8D-G; Table S8). Correspondingly, patients with low TCR-scPGI scores were significantly associated with unfavorable responses to immunotherapy treatment in the training cohort and three validation cohorts (P = 4.58e−03, P = 4.81e−03, P = 2.48e−02, P = 3.69e−03, Fisher's exact test; Fig. 8H-K). We found that the TCR-scPGI results for the therapies were successfully predicted, with AUROC values exceeding 0.80 in three out of four datasets (Fig. 8L). Moreover, the prediction performance of TCR-scPGIs was overall superior to a variety of expression-based signatures across multiple datasets[46-54] (Fig. 8M-P; Table S9), including T-effector and IFN-γ signaling marker, T cell inflamed marker, cytolytic markers, and the CD-scPGI signature. Finally, the TCR-scPGI scores were also predictive of overall patient outcomes. Patients with high TCR-scPGI

scores had longer OS and progression-free survival (PFS) times than patients with low scores across multiple validation datasets (Fig. 8Q-S). We revealed that the TCR-scPGI scores were significantly negatively correlated with tumor purity and significantly positively correlated with the infiltration of several activated immune cells and immune-related genes, such as $CD8^+$ T-cells and neutrophils both in CIBERSORT and ESTIMATE result ($P < 0.05$, Spearman's rank correlation test; Fig. S14A-C). Notably, patients with elevated TCR-scPGI scores may exhibit a "hot" immune microenvironment, which is associated with enhanced responsiveness to immunotherapeutic interventions.

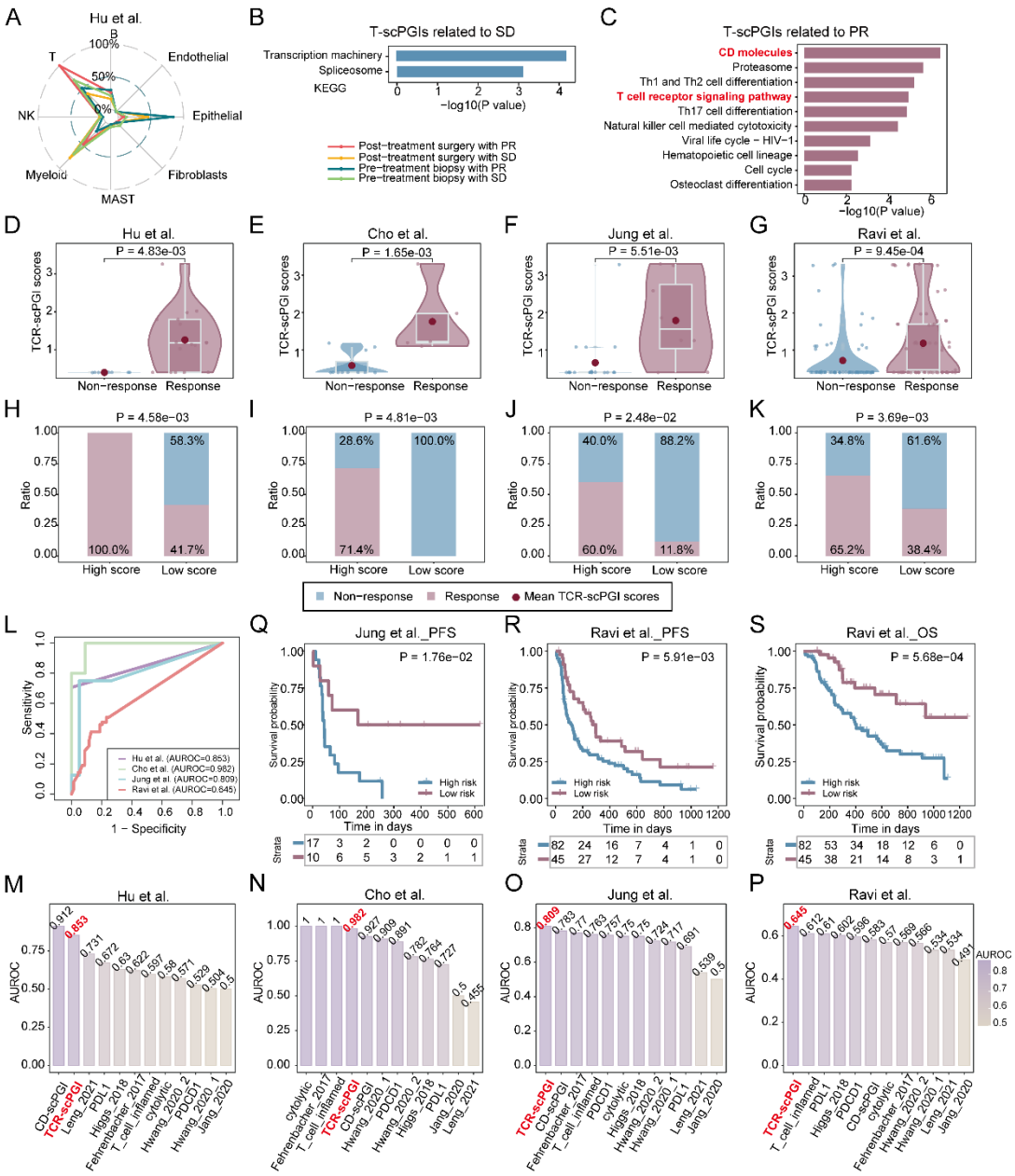

Fig. 8. TCR-scPGIs signature stratifies patients for immunotherapy across multiple datasets.

(A) Distribution of cell ratios in patients with different immunotherapy responses in Hu *et al.*. (B) KEGG Enrichment of genes involved in SD-related T genetic interactions. (C) KEGG Enrichment of genes involved in PR-related T genetic interactions. (D-G) Comparison of the TCR-scPGI scores between responders and non-responders in Hu *et al.*, Cho *et al.*, Jung *et al.* and Ravi *et al.*

datasets. (H-K) Proportion of lung cancer patients who responded to immunotherapy treatment with high TCR-scPGI scores and low TCR-scPGI scores. (L) ROC curve showing the prediction performance of the TCR-scPGIs signature in Hu *et al.*, Cho *et al.*, Jung *et al.* and Ravi *et al.* datasets. (M-P) Bar graphs showing the AUROC-based predictive performance of TCRGI-based predictors versus published prediction signatures across multiple datasets. (Q-S) Kaplan-Meier plots depicting the survival of patients with low versus high TCR-scPGI scores in the Jung *et al.* and Ravi *et al.* datasets. *P* values were calculated by Log-rank test.

Discussion

By analyzing scRNA-seq data across epithelial cells and T cells, we comprehensively sculpt the landscape of scPGIs within the broader context of cancer research. Our work extends previous studies on GIs in cancer research, which have been almost entirely focused on tissue-specific or cancer-specific GIs[55-58], to investigate cell-specific types of interactions. The discovered scPGIs demonstrated promise and viability in depicting cell activity, as evidenced by their favorable association with cell-specific functional activity. Although the effect sizes are modest, two comparisons point in the same direction. Hub subnet scores show higher correlations with cell-type-specific functional modules for scPGI hubs than for matched non-scPGI hubs, and scPGI gene pairs exhibit modest but reproducible increases in shared PPI neighbors, pathway co-membership, essentiality correlations and shorter PPI distances compared with matched controls. Together, these patterns suggest that scPGIs capture subtle but biologically meaningful cell-type-specific cooperative programs, while the evidence should be viewed as supportive rather than

definitive.

More importantly, the E-scPGIs were predictive of patient survival and drug response, elucidated tumor malignancy specificity, and revealed clinically relevant lung cancer subtypes. In addition, the set of T-scPGIs related to immunotherapy response was linked to an immune activity. Overall, these results illustrated the pivotal role of molecular biology and cellular processes represented by scPGIs. The distinct cell characteristics between groups with different subnet scores may be partly attributed to the scPGIs, which were inferred based on their unique expression patterns. However, the subgroups of patients and cell lines determined by the detected scPGIs, whether in epithelial cells or T cells, were also characterized by distinct tumor progression, survival trends, treatment responses and tumor microenvironments, which were not taken into consideration when determining the scPGIs. Overall, these results underscored the significance of molecular context represented by E-scPGIs and T-scPGIs, both at the resolution of single cells and cancer patients.

Interestingly, the ME-scPGI genes without differential expression still formed a coherent functional module enriched for multiple signaling pathways (e.g., Wnt, TGF-β, Hedgehog, VEGF, mTOR, ErbB), all of which are known to regulate epithelial plasticity, stemness, or microenvironmental adaptation. In contrast, consistently upregulated ME-scPGI genes were enriched for proliferation and biosynthesis pathways, whereas consistently downregulated genes were associated with immune- and stress-response programs. These results indicate that the biological relevance of ME-scPGIs is not solely driven by differential expression; rather, ME-scPGIs capture network-level coordination involving both DE and non-DE genes that participate

in malignant epithelial states through distinct functional modules.

In this study, the fundamental unit of inference is the pairwise scPGI, rather than the single gene. While we occasionally highlight individual hub genes for interpretability, their biological relevance arises from their interaction neighborhoods: scPGIs identify gene pairs whose coordinated high expression is enriched beyond expectation within a cell type and associated with higher proliferation scores. In the malignant epithelial network, many ME-scPGI pairs involve genes that are already upregulated in tumors, but their joint activation provides a much finer-grained readout of malignant-state heterogeneity than either gene alone. For example, in the spatial transcriptomic analysis, the co-variation patterns of *FAM83A-MACC1* and *FAM83A-MUC16* pairs were far more spatially coherent than the patterns of either gene alone, forming tightly overlapping hotspots within ME-scPGI high regions that corresponded to poorly differentiated, proliferative and invasive areas. Aggregating these pairwise interactions into the ME-scPGI program improves tumour-normal classification and patient stratification compared with expression-based hallmark signatures and remains prognostic after adjusting for tumour purity and immune infiltration. Thus, scPGI-based networks provide information that is complementary to single-gene analyses by resolving cooperative, cell-type-specific programs in LUAD.

scPGIs are fundamentally different from previous efforts for GIs inference and therapy response prediction in one important way: the interactions underlying the prediction are inferred from specific cell types via scRNA-seq data, making them more context-sensitive and more effective at predicting immunotherapy responses when T-scPGIs are applied. Additionally, the predictive performance of TCR-scPGI is superior to existing predictors for NSCLC

immunotherapy response. Reassuringly, TCR-scPGI maintains its prediction performance with AUROC exceeding 0.80 in three datasets.

Identifying scPGIs is only a first step toward capturing the complexity of cellular networks in epithelial cells and T cells. Future endeavors will leverage a greater variety of cell types to infer scPGIs at cell-type-specific expression levels and help elucidate the contribution of tumor composition and tumor microenvironment. In this study, we focused on epithelial and T-cell compartments, which represent the principal axes of LUAD biology and are the best-represented and phenotypically well-defined populations in current scRNA-seq datasets. Nonetheless, the scPGI-finder framework can be extended to other immune and stromal populations (e.g., macrophages, NK cells, dendritic cells), which may harbor additional clinically relevant interaction programs and further refine our understanding of the LUAD tumor microenvironment. As more data and investigations accumulate, we may further learn how to combine interactions arising from distinct cell types to further boost prediction performance for therapeutic responses in the foreseeable future.

Despite its focus on LUAD, which enables the joint analysis of single-cell, bulk, spatial, and pharmacogenomic cohorts within a single disease context, the scPGI-finder framework is universally applicable to other tumor types. A systematic pan-cancer extension is currently being pursued as a separate line of research.

In this work, we introduce the first systematic scRNA-seq-based computational pipeline to establish cell-type-specific positive GI networks in epithelial cells and T cells: scPGI-finder. Moreover, this study provides epithelial cell-specific and T cell-specific gene pair-based classifiers,

which can be used to assess cancer malignancy and predict immunotherapy response. This work lays a conceptual and computational foundation for future studies of scPGIs, with potential translational applications.

Like other genome-wide computational methods for GIs prediction in bulk, scPGI-finder has several limitations that should be acknowledged and improved upon in the future. First, due to biological and technical limitations in scRNA-seq, damaged and dead cells were removed[18]. Consequently, we concentrated exclusively on gene pairs with specific joint high expression levels that may contribute to better cell fitness, rather than those associated with decreased fitness. Future work can go beyond the expression types studied here to investigate more complex expression patterns with distinct cell viability. For instance, a given interacting gene pair with co-inactive expression could confer detrimental effects on cell fitness in specific cell types. Additionally, the current version of scPGI-finder considers multiple types of GIs features to enhance the reliability of scPGIs. Further features could be considered in the future as more research accumulates.

In addition, CS-CORE has reported limitations when expression is extremely sparse or spatially restricted. In our application, most genes entering scPGI-finder were sufficiently expressed in epithelial and T-cell populations, and CS-CORE was used only as a secondary co-expression filter, so this limitation is unlikely to materially affect the reported results. Consistently, our down-sampling analysis in epithelial cells suggests that several thousand cells per cell type are generally required for stable co-activation estimation, whereas very small subsets provide limited power; applications to rare or sparsely sampled populations may therefore require additional sampling or alternative modelling strategies.

Notably particularly, this study is observational and based on transcriptomic correlations. By construction, scPGIs are defined as gene pairs whose coordinated high expression at single-cell resolution is associated with proliferation-related scores, and therefore they do not by themselves establish causal, non-additive genetic interactions in the classical perturbation sense. Experimental perturbations (e.g., combinatorial CRISPR screens) will be required to test whether specific scPGIs correspond to bona fide genetic interactions.

Methods

Key resources table

See details in Supplementary Tables.

Identification of scPGIs via scPGI-finder pipeline

Identification of over-expressed genes in per cell type: To infer the scPGIs, we first identified differentially over-expressed genes for the desired cell subset within a given scRNA-seq dataset, using the default Wilcoxon rank sum test with only. pos = TRUE via R package "Seurat" (v 4.2.1).

Cell ratio test: We identified candidate over-expressed gene pairs by analyzing expression data in the desired cell type. This step searches for gene pairs whose co-activations are under positive selection. We mined gene expression data to identify gene pairs whose co-activation (defined below) was significantly more frequent than expected by One proportion Z-test. The range of each gene's expression values across cells was divided into two bins according to the median expression value, without regard to non-zero expression. Cells with expression values greater than the threshold are considered in the active (A) state bins. Conversely, cells with expression values

less than the threshold are considered in the inactive (I) state bins. PoAA denotes the observed probability of Gene1 and Gene2 with co-activation. PeAA is the expected probability of Gene1 and Gene2 with co-activation, which equals to the probability of Gene1 in the active state multiplied the probability of Gene2 in the active state. The observed probability of cells with co-activation in both Gene1 and Gene2 being greater than random chance was calculated according to Equation (1). The *P* value was subjected to the Benjamini and Hochberg (BH) correction for multiple tests. The gene pairs with false discovery rate (FDR) < 0.01 were selected for further analysis. The test statistic was given as:

$$Z = \frac{PoAA - PeAA}{\sqrt{\frac{PoAA*(1-PoAA)}{n}}} \quad \text{Equation (1)},$$

$$H0: PoAA \leq PeAA; \quad H1: PoAA > PeAA,$$

where n represents the number of cells in the desired cell type.

Differential cell proliferation analysis: In this step, the desired cells were divided into three groups according to the threshold in step ii: both Gene1 and Gene2 have active expression (AA), one gene is active while the other is inactive (AI or IA) and both Gene1 and Gene2 are inactive (II). The cell proliferation score was evaluated through the cell cycle and DNA replication enrichment scores of each cell using the "AUCell" (v 1.22.0) R package for counts data[15,59]. For each gene pair, One-sided Wilcoxon rank-sum test was used to examine whether the cell proliferation levels of AA cells were greater than the levels of AI/IA cell bins and II cell bins, respectively. The *P* value was subjected to BH correction for multiple tests. The gene pairs with FDR < 0.01 were selected for downstream analysis.

Cell-type-specific positive co-expression inference: Due to the limitations of data sparsity, we utilized "CS-CORE" (v0.0.0.9000) with seq_depth = NULL to infer cell-type-specific positive co-expression gene pairs from scRNA-seq data[16]. The co-expressed *P* value was subjected to the BH correction for multiple tests. Gene pairs with FDR < 0.01 and R > 0.1 were selected for further analysis.

Functional similarity analysis: Functional similarity (FS) scores between gene pairs were calculated based on the semantic similarities in biological process (SimBP), molecular function (SimMF), and cellular component (SimCC) aspects of the Gene ontology (GO) terms, which account for both the function and location of genes[6,13,60]. The FS score for a gene pair was given as:

$$\text{FS score} = \sqrt[3]{\text{SimBP} * \text{SimMF} * \text{SimCC}} \quad \text{Equation (2),}$$

Semantic similarities were measured based on the GO topological biological process (BP), molecular function (MF), and cellular component (CC) through the "GOSemSim" (v 2.26.0) package with semData = bp, cc or mf[61]. Gene pairs with FS scores >0.5 were considered to have high functional interactions and were used for further analysis.

GO semantic similarity was used as a downstream refinement step after all expression-based criteria had been applied, to modestly narrow the candidate gene-pair set. The core identification of scPGIs is driven by co-activation enrichment within a cell type, cell-type-specific co-expression, and association with proliferation-related programs.

Evaluation of co-expression models

Several statistical models were considered for the co-expression steps, including Pearson

correlation, CS-CORE, MAST and Dozer. To compare their relative performance, we used 220 experimentally validated synthetic-viability gene pairs reported in previous studies. SV pairs reflect cooperative viability effects in bulk or cell-line perturbations and therefore do not constitute a strict single-cell benchmark, but they provide a useful reference set for assessing the relative discrimination ability of each model.

For CS-CORE and Pearson we applied commonly used significance/effect-size thresholds (CS-CORE: $P < 0.05$ and est > 0.1; Pearson: $P < 0.05$ and r > 0.1), and required propR est > 0.1. MAST was used to regress out technical covariates (e.g. CDR, batch) via the hurdle model, and Pearson correlations were then computed on the resulting residuals. For each model we recorded AUROC, recall, runtime and memory usage to compare discrimination performance and computational cost.

scRNA-seq data processing and analysis

We first obtained a LUAD atlas by extracting scRNA-seq data from a mixed non-small cell lung cancer (NSCLC) atlas to infer scPGIs for epithelial cells and T cells (Table S1)[19]. Only primary LUAD patients who were treatment-naive were used for downstream analysis. The R package "Seurat" (v 4.2.1) was used to preprocess raw data of single cell[62]. Cell types were annotated based on canonical marker genes derived from the literature (Table S2)[63,64].

We applied three quality measures on raw gene-cell-barcode matrix for each cell. All cells were removed that had either over 50,000 or below 100 unique molecular identifiers (UMIs), over 6,000 or below 200 expressed genes, or over 20% UMIs derived from mitochondrial genome. For each batch, we used the filtered cells to remove genes that are expressed at low levels by counting

the number of cells (min.cells) having expression of each Genei, and excluded genes with min.cells < 0.1% cells. For the remaining cells and genes, R package "harmony" (v 1.1.0) (https://cran.r-project.org/web/packages/harmony/index.html) were used to integrate cells across individuals from different datasets. Subsequently, count data was normalized and scaled using the NormalizeData function and ScaleData function. FindVariableFeatures function was used to screen out the top 2,000 highly variable genes from the normalized expression matrix, then we performed the principal component analysis using these genes. We performed FindNeighbors function to get nearest neighbors for graph clustering based on top 30 PCs and performed FindCluster function to determine the optimal resolution. t-SNE algorithm was used to visualize identified clusters on the 2D map. Cell types were annotated based on the canonical marker genes derived from the literature (Supplementary Tables S2). Ultimately, we defined a set of cell-type notations considering the 8 major cell types: T, fibroblast, endothelial, B, NK, myeloid, epithelial and mast cells. Over-expressed genes for the desired cell subset within a given scRNA-seq dataset were generated by Seurat FindMarkers function using the default Wilcoxon rank sum test with the parameters: only.pos = TRUE.

Salcher *et al.* was used as the discovery dataset to construct cell-type-specific scPGI networks. Zhou et al. served as an independent validation dataset for evaluating the reproducibility and functional characteristics of the identified scPGIs. Guo *et al.* was used as a biological validation dataset, in which the T-scPGI-derived hub subnet was applied to stratify T-cell states. A concise summary of all scRNA-seq datasets and their roles is provided in Table S1.

Pathway activity evaluation

Genes related to cell-specific biological processes were retrieved by searching the keywords "gobp_epithelial_cell" and "gobp_t_cell" in GO[65]. Cancer hallmarks were obtained from COSMIC[66]. Cell-specific biological process enrichment scores for each cell were calculated by "AUCell" (v 1.22.0) R package[15]. In addition, we used the "CytoTRACE" (v 0.3.3) R package to assess the differentiation (stemness) status of cells[67]. Cells were assigned a CytoTRACE score representing their differentiation potential, with higher scores indicating a lower level of differentiation and higher stemness. The desired cells were grouped by the median of subnet scores into high score and low score groups. Cell-cell communications were analyzed using the R package "CellChat" (v 1.6.1)[68].

Benchmarking of scPGIs using external functional feature databases

To assess the functional coherence of scPGIs[6], we used gene-pair features from several external resources. Protein-protein interaction (PPI) neighbors were obtained from PCNet[69]; pathway co-membership from MSigDB; gene essentiality correlations of lung cancer cell lines from DepMap; and subcellular colocalization annotations from the Human Protein Atlas (HPA, https://www.proteinatlas.org/). Gene essentiality (gene effect score) was obtained from the DepMap Achilles CRISPR knockout screens (https://depmap.org/portal/), where lower gene effect values indicate stronger dependency and reduced cell viability upon gene perturbation. In addition, the "igraph" (v 2.0.3) R package was used to calculate the average shortest distance for each gene pair.

For each cell type, we constructed a matched control set (“non-scPGIs”) by randomly sampling an equal number of gene pairs from the same gene pool used to derive scPGIs. All

functional features (shared and union PPI neighbors, shared and union pathways, co-expression, essentiality correlations, and colocalization scores) were computed in parallel for scPGIs and non-scPGIs. Statistical comparisons were performed using two-sided Wilcoxon rank-sum tests or chi-squared tests. Detailed descriptions of the features are provided in Table S5.

Copy-number variation detection and epithelial subset analysis

The R package "infercnv" (v 1.10.1) was used to estimate the copy-number variation (CNV) via count data in epithelial cells (n=31,413), with fibroblast cells and endothelial cells as baselines. We scored each cell for the extent of CNV signal. The CNV score was used to be re-clustered by k-means clustering (k = 8). Epithelial cells that clustered together with spiked in controls (Fibroblast cells and Endothelial cells) were labeled "non-malignant" (n = 8,872) in cluster 7, whereas the remaining seven clusters were labeled as "malignant" (n = 22,541).

The same methodology was applied to infer malignant cell subsets in the dataset from Zhou *et al.*. Ultimately, 927 epithelial cells were grouped with fibroblast and endothelial cells into clusters 3 and 4, and were classified as "non-malignant." In contrast, 5,275 epithelial cells were classified as "malignant".

Measuring node centrality in scPGIs network

The centrality of genes is used to quantify the importance in network, such as the E-scPGI network and T-scPGI network in our case. There are various metrics to measure node centrality. Here, we employed degree, betweenness, closeness, pagerank, and eigenvalue. All these centralities were calculated with the R package “igraph” (v 2.0.3) (https://cloud.r-project.org/web/packages/igraph/citation.html). Genes were considered as central if they were

ranked in the top 30% across all five given centrality metrics.

Hub gene identification in scPGI networks

For initial characterization of epithelial and T-cell scPGI networks, we computed several standard centrality measures (including degree-based and distance-based metrics) and defined a consensus "core" by selecting genes that ranked highly across multiple measures for functional enrichment. For downstream analyses, hubs were defined by degree centrality, because scPGI edges represent undirected co-activation rather than directional signal propagation, and degree directly reflects the number of partners with which a gene repeatedly co-activates.

Calculation the subnet scores for epithelial cells and T cells

Genes with the top 20 degrees were considered as hubs in the scPGIs network. The subnet score for each hub gene with its partner genes was calculated based on the co-active scPGIs. For single-cell RNA-seq data, within each cell type, each gene's non-zero expression values were dichotomized at their median; cells with expression ≥ this median were assigned to the active (A) state and those below to the inactive (I) state. For a gene pair (g_i, g_j), a cell c was said to be co-activated (AA) if both genes were in the A state, and a pairwise indicator was defined as

$$I_{ij}(c) = \begin{cases} 1, if\ g_i\ and\ g_j\ \ are\ both\ active\ (AA)\ in\ cell\ c, \\ 0, otherwise, \end{cases}$$

For a given hub gene h with K_h scPGI partners, the subnet score in cell c was calculated as

$$Subnet\ score\ (c,h) = \frac{1}{K_h}\sum_{k=1}^{K_h} I_{hk}(c)\ \ Equation\ (3),$$

which summarizes the fraction of its scPGI partners that are co-activated with the hub in that cell.

Cell-type-specific function scoring

For predefined biological processes relevant to each cell type (e.g., epithelial differentiation, epithelial-mesenchymal transition, T-cell activation, cytotoxicity), we computed single-cell activity scores using AUCell, which quantifies the enrichment of a gene set in the top-ranked expressed genes of each cell. The resulting AUCell scores were used as cell-type-specific functional scores and correlated with scPGI hub module scores to assess functional relevance.

Identification of the E-scPGIs related to malignant cells

In the epithelial cells arising from the scRNA-seq dataset, we screened the E-scPGIs related to malignant cells (ME-scPGIs) for two procedures: (1) The E-scPGIs showed a higher AA expression pattern in the malignant than the non-malignant epithelial cells by Pearson's chi-squared test with $P < 0.01$. (2) The fraction of AA expression in malignant cells was at least twice than in non-malignant epithelial cells for each E-scPGI.

ME-scPGI score in epithelial cells

For each epithelial cell c, the ME-scPGI score was defined by aggregating the activity of all M ME-scPGI pairs (g_i & g_j):

$$\text{ME-scPGI}(c) = \frac{1}{M}\sum_{m=1}^{M} I_{ij}\,(c) \quad \text{Equation (4)},$$

where $I_{ij}(c) = 1$ if both genes in ME-scPGI pair (g_i & g_j) are in the A state (AA) in cell c, and 0 otherwise, as defined above. A data-driven threshold of 0.0827 (estimated using the *cutoff* R package) was used to distinguish malignant from non-malignant epithelial cells in the discovery scRNA-seq dataset and independent dataset.

Comparison of the ME-scPGI classifier with other methods

The ME-scPGI score for each epithelial cell was defined as the proportion of ME-scPGI pairs exhibiting co-activation in that cell. For comparison, AUCell scores of canonical tumor hallmark programs were used as expression-based classifiers. inferCNV and CopyKAT weak labels were used as reference annotations for malignant versus non-malignant epithelial cells.

ME-scPGI score in bulk tumors and cancer cell lines

The ME-scPGI scores were calculated across multiple bulk RNA-seq datasets (Table S6) and cell lines (Table S7). For bulk RNA-seq datasets and cancer cell lines, each gene's expression across samples was divided into tertiles, and the upper tertile was used to define a high/active state. For an ME-scPGI pair M $(g_i \,\&\, g_j)$ in sample s, we defined

$$J_{ij}(s) = \begin{cases} 1, if \text{ both } g_i \text{ and } g_j \text{ fall into the upper tertile in } s, \\ 0, otherwise, \end{cases}$$

and the bulk or cell-line ME-scPGI score was computed as

$$M\text{E-scPGI}(s) = \frac{1}{M}\sum_{m=1}^{M} J_{ij}\,(c) \quad \text{Equation (5)}\,.$$

The same definition was used across all bulk LUAD cohorts and pharmacogenomic cell-line panels.

Pathway enrichment of ME-scPGI genes in bulk LUAD datasets

To assess the pathway context of ME-scPGI genes in bulk tumors, we analyzed eight independent lung cancer transcriptomic datasets with paired tumor and normal samples. For each cohort, differential expression of ME-scPGI genes between tumor and normal tissues was computed using the Wilcoxon rank-sum test with FDR correction (adjusted $P < 0.05$), and genes were classified as upregulated, unchanged, or downregulated. We classified each ME-scPGI gene as downregulated, upregulated, or unchanged in tumors, and selected genes with a concordant direction of change in

⩾4 cohorts for pathway enrichment analysis. Genes showing the same direction of change in at least four cohorts were then subjected to KEGG pathway enrichment using clusterProfiler, with $P < 0.05$ considered significant.

Evaluation of immune infiltration in bulk RNA-seq

The proportions of infiltrating immune cells were estimated via the "CIBERSORT" (v 0.1.0), "estimate" (v 1.0.13), and "xCell" (v 1.1.0) methods based on gene expression data. Immune checkpoint genes were obtained from the study of Hu *et al.*[70].

Survival analysis in multiple lung cancer dataset

The patients were grouped by upper terciles of the ME-scPGI scores into high ME-scPGI scores and low ME-scPGI scores. The OS time of the two groups was tested using Log-rank test, $P < 0.05$ was considered statistically significant and the results were represented by Kaplan-Meier plots.

Spatial transcriptomic analysis

LUAD spatial transcriptomic dataset from the Ad-SpatialAnalysis repository were processed to compute well-differentiated, proliferative and invasive scores per spatial spot using AUCell and ME-scPGI scores. AUCell scores for ME-scPGI and for well-differentiated, proliferative and invasive signatures were computed for each spatial spot. Spatial co-variation between different scores, *FAM83A* and its ME-scPGI partner genes (*CRABP2*, *MACC1*, *MUC16*, *VSTM2L*, *HOMER3*) was evaluated using Spearman correlation and visualized with spatial heatmaps.

Assessment the association of ME-scPGI scores and drug response in lung cancer cells

The pharmacological screening data were downloaded from (1) Genomics of Drug Sensitivity in Cancer (GDSC) (GDSC1 and GDSC2, v8.5); (2) Cancer Therapeutics Response Portal (CTRP)

(v2.0); (3) DepMap portal (PRISM Repurposing v19Q4). Cell lines in CTRP and PRISM Repurposing were referenced from the Cancer Cell Line Encyclopedia (CCLE). Cell line information and normalized gene expression values were obtained from GDSC and DepMap. AUC and IC50 values of drugs were collected as measures of cancer cell line sensitivity to small molecule treatments. Lower IC50/AUC values typically indicate greater sensitivity of the cell line to the drug. We identified drugs that IC50/AUC were significant positive or negative correlation with ME-scPGI score by Spearman's rank correlation analysis in NSCLC cell lines.

Identification of the T-scPGIs related to immunotherapy response

In the T cells arising from the scRNA data and matched bulk data with immunotherapy information, we screened the T-scPGIs related to immunotherapy response with two procedures: (1) In the scRNA data of Hu et al., the T-scPGIs showed a higher AA expression pattern in the T cells from pre-treatment biopsy with partial response (PR) than in the T cells from pre-treatment biopsy with stable disease (SD) by One proportion Z-test with $P < 0.05$. (2) In the bulk RNA-seq data of Hu et al., the T-scPGIs showed a higher AA expression pattern in the response patients than in the non-response patients of pre-treatment biopsy by One proportion Z-test with $P < 0.05$.

The T-scPGIs related to immunotherapy resistance with two procedures: (1) In the scRNA data of Hu *et al.*, the T-scPGIs showed a higher AA expression pattern in the T cells from Pre-treatment biopsy with SD than in the T cells from Pre-treatment biopsy with PR by One proportion Z-test with $P < 0.05$. (2) In the bulk RNA-seq data of Hu *et al.*, the T-scPGIs showed a higher AA expression pattern in the non-response patients than in the response patients of Pre-treatment biopsy by One proportion Z-test with $P < 0.05$.

The responses of NSCLC patients to immunotherapy were assessed in accordance with the response evaluation criteria in solid tumors (RECIST) clinical practice guidelines[71]. Patients with complete response (CR) or PR or SD > six months were classified as responders, while patients with progressive disease (PD) or SD < six months were classified as nonresponders based on RECIST when the stable disease times can be obtained. Otherwise, patients with CR or PR were classified as responders, while patients with PD or SD were classified as non-responders based on RECIST when the stable disease times cannot be obtained.

Prediction of the immunotherapy response by T-scPGI

We implemented the least absolute shrinkage and selection operator (LASSO) model to further optimize the T-scPGIs involved in T cell receptor signaling (TCR) genes or the cluster of differentiation (CD) genes related to immunotherapy response. The prediction model for immunotherapy response using T-scPGIs involved in TCR pathways were named as TCR-scPGIs. The prediction model for immunotherapy response using T-scPGIs involved in CD pathways were named as CD-scPGIs. The "glmnet" (v 4.1.7) package was used to perform the LASSO logistic regression analysis to predict immunotherapy response by minimizing the model error. The parameters of "glmnet" were set as family = "binomial", alpha = 1, and nfolds = 10, and the coefficients at lambda.min were selected to construct the predictive model. The receiver-operating characteristic (ROC) curve was plotted using the "pROC" (v 1.18.2) package. The best threshold value for ROC curve was determined by "cutoff" (v 1.3) R package to predict immunotherapy response[72].

Construction of the TCR-scPGIs signature

For T-scPGIs involving TCR-related genes, we derived a TCR-scPGI signature by selecting six T-scPGI pairs via LASSO logistic regression in the discovery cohort. For each TCR scPGI pair, expression across patients was divided into tertiles, and the pair score was defined analogously to the bulk ME-scPGI definition:

$$Score_{g_i\_g_j}(s) = \begin{cases} 1, & if \text{ both } g_i \text{ and } g_j \text{ fall into the upper tertile in } s, \\ 0, & otherwise. \end{cases}$$

The TCR-scPGI score for sample s was then calculated as:

$$\begin{aligned} \text{TCR-scPGI score} = {} & 0.0254 * \text{Score}_{\text{CD3D\_TNFRSF18}} + 0.0392 * \text{Score}_{\text{CD3D\_CKLF}} + 0.7874 \\ & * \text{Score}_{\text{CD3E\_HAVCR2}} + 0.6008 * \text{Score}_{\text{IFNG\_LAG3}} + 0.7963 * \text{Score}_{\text{IFNG\_HAVCR2}} \\ & + 0.6809 * \text{Score}_{\text{PDCD1\_TAP1}} + 0.3961, \text{Equation (6).} \end{aligned}$$

A cutoff of 1.02 (estimated using the *cutoff* package) was used to stratify patients in the discovery cohort, and the same formula and cutoff were applied to all validation cohorts. Finally, the same formula and best threshold were applied to the other validation cohorts.

Construction of the CD-scPGIs signature

For the T-scPGIs involved in CD genes, we quantified a CD-scPGI score for each patient based on the eight T-scPGIs pairs signature through LASSO logistic regression analysis. The formula of the CD-scPGI score was as follows:

$$\begin{aligned} \text{CD-scPGI score} = {} & \\ & 0.4755 * \text{Score}_{\text{CD3D\_LAG3}} + 1.2820 * \text{Score}_{\text{HCST\_HAVCR2}} + \\ & 0.5779 * \text{Score}_{\text{IFNG\_LAG3}} + 0.6265 * \text{Score}_{\text{IFNG\_HAVCR2}} + \\ & 0.7362 * \text{Score}_{\text{PDCD1\_TAP1}} + 0.2370 * \text{Score}_{\text{PRF1\_HAVCR2}} + \\ & 0.0677 * \text{Score}_{\text{SIRPG\_HAVCR2}} + 1.2498 * \text{Score}_{\text{TNFRSF18\_TNFRSF25}} + \\ & 0.0883 \qquad \text{Equation (7),} \end{aligned}$$

and the pair score was defined analogously to the bulk TCR-scPGI definition. The best threshold

value for ROC curve was determined as 0.952 by cutoff package. Finally, the same formula and threshold were applied to the other validation cohorts.

Survival analysis in NSCLC treated with immunotherapy

The patients were grouped by upper terciles of the TCR-scPGI scores into TCR-scPGI high scores and TCR-scPGI low scores. The OS time or PFS time of the two groups was tested using Log-rank test, $P < 0.05$ was considered statistically significant and the results were represented by Kaplan-Meier plots.

Functional enrichment analysis

Functional enrichment analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway were conducted via the "clusterProfiler" (v 4.8.1) R package[73].

Statistics

Two-sided Wilcoxon rank-sum test was used to assess the discrepancy between scPGIs and non-scPGIs, the CytoTRACE scores between cells with high subnet scores and low subnet scores, the subnet scores of hub genes between epithelia cells from patients with advanced stage and early stage, the average clone frequency between cells with high subnet scores and low subnet scores, Spearman's rank correlation test was used to estimate the association between scores of two cell groups. All statistical analyses and presentations were performed using R software (v 4.3.0).

Abbreviations

See details in Table S10.

Supporting information

Supplementary file 1: The file contains the accompanying Supplementary Figures S1-14 and Supplementary Tables S1-10.

Funding

This work was supported by grants from the National Key Research and Development Program of China (No. 2023YFF1204600), the National Natural Science Foundation of China (No. 32470702 and 32270710), the Scientific Research Project of Provincial Scientific Research Institutes of Heilongjiang Province (No. CZKYF2024-1-A010), National Multidisciplinary Innovation Team Project in Traditional Chinese Medicine (No. ZYYCXTD-D-202407), and Chunyan Team Program of Heilongjiang Province (No. CYQN24043).

Acknowledgements

The authors acknowledge the efforts of all the researchers who have contributed the data to the public databases of TCGA, GEO, DepMap, CTRP, GDSC, COSMIC, and MSigDB. The interpretation and reporting of these data are the sole responsibility of the authors.

Data availability

All data analyzed during this study can be downloaded from public databases or retrieved from associated files of papers, such as The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/), Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/), CCLE, DepMap (https://depmap.org/portal/) and the

molecular signatures database (MSigDB, https://www.gsea-msigdb.org/gsea/msigdb), CTRP (https://portals.broadinstitute.org/ctrp/), GDSC (https://www.cancerrxgene.org/). Detailed information was provided in the Supplementary Tables.

Code availability

The scPGI-finder pipeline is available at GitHub (https://github.com/chencccchen/scPGI-finder-pipeline) with an archived version on Zenodo (https://zenodo.org/records/17817076). Because scPGI-finder is implemented as an analysis workflow rather than a standalone software package, the scripts can be executed directly in a standard R environment without installation. The repository provides step-by-step instructions to reproduce the analyses reported in this study. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Author contributions

YYG, BC, and JJL designed the project. BC, MYL, and QD wrote the manuscript. BC, KDL and CL collected data and performed bioinformatic analysis. HMH, LZW, NZ, and WYZ provided valuable suggestions for the manuscript. All authors read and approved of the final manuscript.

Competing interests

The authors declare no competing financial or non-financial interests

References

1 Wang, J. *et al.* Computational methods, databases and tools for synthetic lethality prediction. *Brief Bioinform* **23**, doi:10.1093/bib/bbac106 (2022).

2 Setton, J. *et al.* Synthetic Lethality in Cancer Therapeutics: The Next Generation. *Cancer Discov* **11**, 1626-1635, doi:10.1158/2159-8290.CD-20-1503 (2021).

3 Fong, S. H. *et al.* A multilineage screen identifies actionable synthetic lethal interactions in human cancers. *Nat Genet* **57**, 154-164, doi:10.1038/s41588-024-01971-9 (2025).

4 van de Haar, J. *et al.* Identifying Epistasis in Cancer Genomes: A Delicate Affair. *Cell* **177**, 1375-1383, doi:10.1016/j.cell.2019.05.005 (2019).

5 O'Meara, C. P. *et al.* Genetic landscape of T cells identifies synthetic lethality for T-ALL. *Commun Biol* **4**, 1201, doi:10.1038/s42003-021-02694-x (2021).

6 Liu, M. *et al.* Synthetic viability induces resistance to immune checkpoint inhibitors in cancer cells. *Br J Cancer* **129**, 1339-1349, doi:10.1038/s41416-023-02404-w (2023).

7 Sahu, A. D. *et al.* Genome-wide prediction of synthetic rescue mediators of resistance to targeted and immunotherapy. *Mol Syst Biol* **15**, e8323, doi:10.15252/msb.20188323 (2019).

8 Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *Am J Hum Genet* **102**, 717-730, doi:10.1016/j.ajhg.2018.04.002 (2018).

9 Zhou, P. *et al.* Single-cell CRISPR screens in vivo map T cell fate regulomes in cancer. *Nature* **624**, 154-163, doi:10.1038/s41586-023-06733-x (2023).

10 Kuzmin, E. *et al.* Systematic analysis of complex genetic interactions. *Science* **360**, doi:10.1126/science.aao1729 (2018).

11 Costanzo, M. *et al.* A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, doi:10.1126/science.aaf1420 (2016).

12 Lee, J. S. *et al.* Synthetic lethality-mediated precision oncology via the tumor transcriptome. *Cell* **184**, 2487-2502 e2413, doi:10.1016/j.cell.2021.03.030 (2021).

13 Jerby-Arnon, L. *et al.* Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell* **158**, 1199-1209, doi:10.1016/j.cell.2014.07.027 (2014).

14 Avila Cobos, F., Vandesompele, J., Mestdagh, P. & De Preter, K. Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics* **34**, 1969-1979, doi:10.1093/bioinformatics/bty019 (2018).

15 Bravo Gonzalez-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* **20**, 1355-1367, doi:10.1038/s41592-023-01938-4 (2023).

16 Su, C. *et al.* Cell-type-specific co-expression inference from single cell RNA-sequencing data. *Nat Commun* **14**, 4846, doi:10.1038/s41467-023-40503-7 (2023).

17 Wilk, A. J., Shalek, A. K., Holmes, S. & Blish, C. A. Comparative analysis of cell-cell communication at single-cell resolution. *Nat Biotechnol* **42**, 470-483, doi:10.1038/s41587-023-01782-z (2024).

18 Lafzi, A., Moutinho, C., Picelli, S. & Heyn, H. Tutorial: guidelines for the experimental design of single-cell RNA sequencing studies. *Nat Protoc* **13**, 2742-2757, doi:10.1038/s41596-018-0073-y (2018).

19 Salcher, S. *et al.* High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. *Cancer Cell* **40**, 1503-1520 e1508, doi:10.1016/j.ccell.2022.10.008 (2022).

20 Gu, Y. *et al.* A landscape of synthetic viable interactions in cancer. *Brief Bioinform* **19**, 644-655, doi:10.1093/bib/bbw142 (2018).

21 Magen, A. *et al.* Beyond Synthetic Lethality: Charting the Landscape of Pairwise Gene Expression States Associated with Survival in Cancer. *Cell Rep* **28**, 938-948 e936, doi:10.1016/j.celrep.2019.06.067 (2019).

22 Zhang, Y. *et al.* A T cell resilience model associated with response to immunotherapy in multiple tumor types. *Nat Med* **28**, 1421-1431, doi:10.1038/s41591-022-01799-y (2022).

23 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457-462, doi:10.1093/nar/gkv1070 (2016).

24 Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, 278, doi:10.1186/s13059-015-0844-5 (2015).

25 Lu, S. & Keles, S. Dozer: Debiased personalized gene co-expression networks for population-scale scRNA-seq data. *bioRxiv*, doi:10.1101/2023.04.25.538290 (2023).

26 Yang, L. *et al.* Single-cell transcriptome analysis revealed a suppressive tumor immune microenvironment in EGFR mutant lung adenocarcinoma. *J Immunother Cancer* **10**, doi:10.1136/jitc-2021-003534 (2022).

27 Moreno Leon, L. *et al.* The nuclear hypoxia-regulated NLUCAT1 long non-coding RNA contributes to an aggressive phenotype in lung adenocarcinoma through regulation of oxidative stress. *Oncogene* **38**, 7146-7165, doi:10.1038/s41388-019-0935-y (2019).

28 Lu, T. P. *et al.* Identification of regulatory SNPs associated with genetic modifications in lung adenocarcinoma. *BMC Res Notes* **8**, 92, doi:10.1186/s13104-015-1053-8 (2015).

29 Lu, T. P. *et al.* Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev* **19**, 2590-2597, doi:10.1158/1055-9965.EPI-10-0332 (2010).

30 Selamat, S. A. *et al.* Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res* **22**, 1197-1211, doi:10.1101/gr.132662.111 (2012).

31 Chang, I. S. *et al.* Genetic Modifiers of Progression-Free Survival in Never-Smoking Lung Adenocarcinoma Patients Treated with First-Line Tyrosine Kinase Inhibitors. *Am J Respir Crit Care Med* **195**, 663-673, doi:10.1164/rccm.201602-0300OC (2017).

32 Chen, K. Y. *et al.* Estrogen Receptor Gene Polymorphisms and Lung Adenocarcinoma Risk in Never-Smoking Women. *J Thorac Oncol* **10**, 1413-1420, doi:10.1097/JTO.0000000000000646 (2015).

33 Okayama, H. *et al.* Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* **72**, 100-111,

doi:10.1158/0008-5472.CAN-11-1403 (2012).

34 Yamauchi, M. *et al.* Epidermal growth factor receptor tyrosine kinase defines critical prognostic genes of stage I lung adenocarcinoma. *PLoS One* **7**, e43923, doi:10.1371/journal.pone.0043923 (2012).

35 Robles, A. I. *et al.* An Integrated Prognostic Classifier for Stage I Lung Adenocarcinoma Based on mRNA, microRNA, and DNA Methylation Biomarkers. *J Thorac Oncol* **10**, 1037-1048, doi:10.1097/JTO.0000000000000560 (2015).

36 Girard, L. *et al.* An Expression Signature as an Aid to the Histologic Classification of Non-Small Cell Lung Cancer. *Clin Cancer Res* **22**, 4880-4889, doi:10.1158/1078-0432.CCR-15-2900 (2016).

37 Tomida, S. *et al.* Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *J Clin Oncol* **27**, 2793-2799, doi:10.1200/JCO.2008.19.7053 (2009).

38 Tang, H. *et al.* A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res* **19**, 1577-1586, doi:10.1158/1078-0432.CCR-12-2321 (2013).

39 Rousseaux, S. *et al.* Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med* **5**, 186ra166, doi:10.1126/scitranslmed.3005723 (2013).

40 Schabath, M. B. *et al.* Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* **35**, 3209-3216, doi:10.1038/onc.2015.375 (2016).

41 Takano, Y. *et al.* Spatially resolved gene expression profiling of tumor microenvironment reveals key steps of lung adenocarcinoma development. *Nat Commun* **15**, 10637, doi:10.1038/s41467-024-54671-7 (2024).

42 Hu, J. *et al.* Tumor microenvironment remodeling after neoadjuvant immunotherapy in non-small cell lung cancer revealed by single-cell RNA sequencing. *Genome Med* **15**, 14,

doi:10.1186/s13073-023-01164-9 (2023).

43 Jung, H. *et al.* DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nat Commun* **10**, 4278, doi:10.1038/s41467-019-12159-9 (2019).

44 Kim, J. Y., Choi, J. K. & Jung, H. Genome-wide methylation patterns predict clinical benefit of immunotherapy in lung cancer. *Clin Epigenetics* **12**, 119, doi:10.1186/s13148-020-00907-4 (2020).

45 Ravi, A. *et al.* Genomic and transcriptomic analysis of checkpoint blockade response in advanced non-small cell lung cancer. *Nat Genet* **55**, 807-819, doi:10.1038/s41588-023-01355-5 (2023).

46 Fehrenbacher, L. *et al.* Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* **387**, 1837-1846, doi:10.1016/S0140-6736(16)00587-0 (2016).

47 Higgs, B. W. *et al.* Interferon Gamma Messenger RNA Signature in Tumor Biopsies Predicts Outcomes in Patients with Non-Small Cell Lung Carcinoma or Urothelial Cancer Treated with Durvalumab. *Clin Cancer Res* **24**, 3857-3866, doi:10.1158/1078-0432.CCR-17-3451 (2018).

48 Hwang, S. *et al.* Immune gene signatures for predicting durable clinical benefit of anti-PD-1 immunotherapy in patients with non-small cell lung cancer. *Sci Rep* **10**, 643, doi:10.1038/s41598-019-57218-9 (2020).

49 Leng, Y. *et al.* GDPLichi: a DNA Damage Repair-Related Gene Classifier for Predicting Lung Adenocarcinoma Immune Checkpoint Inhibitors Response. *Front Oncol* **11**, 733533, doi:10.3389/fonc.2021.733533 (2021).

50 Jang, H. J. *et al.* Transcriptome-based molecular subtyping of non-small cell lung cancer may predict response to immune checkpoint inhibitors. *J Thorac Cardiovasc Surg* **159**, 1598-1610 e1593, doi:10.1016/j.jtcvs.2019.10.123 (2020).

51 Ayers, M. *et al.* IFN-gamma-related mRNA profile predicts clinical response to PD-1

blockade. *J Clin Invest* **127**, 2930-2940, doi:10.1172/JCI91190 (2017).

52 Doroshow, D. B. *et al.* PD-L1 as a biomarker of response to immune-checkpoint inhibitors. *Nat Rev Clin Oncol* **18**, 345-362, doi:10.1038/s41571-021-00473-5 (2021).

53 Herbst, R. S. *et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563-567, doi:10.1038/nature14011 (2014).

54 Agioti, S. & Zaravinos, A. Immune Cytolytic Activity and Strategies for Therapeutic Treatment. *Int J Mol Sci* **25**, doi:10.3390/ijms25073624 (2024).

55 Lin, C. J. *et al.* Genetic interactions reveal distinct biological and therapeutic implications in breast cancer. *Cancer Cell* **42**, 701-719 e712, doi:10.1016/j.ccell.2024.03.006 (2024).

56 Birkealv, S. *et al.* Mutually exclusive genetic interactions and gene essentiality shape the genomic landscape of primary melanoma. *J Pathol* **259**, 56-68, doi:10.1002/path.6019 (2023).

57 Liu, F. *et al.* Cryo-shocked tumor cells deliver CRISPR-Cas9 for lung cancer regression by synthetic lethality. *Sci Adv* **10**, eadk8264, doi:10.1126/sciadv.adk8264 (2024).

58 Pacini, C. *et al.* A comprehensive clinically informed map of dependencies in cancer cells and framework for target prioritization. *Cancer Cell* **42**, 301-316 e309, doi:10.1016/j.ccell.2023.12.016 (2024).

59 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

60 Yang, C. *et al.* Mapping the landscape of synthetic lethal interactions in liver cancer. *Theranostics* **11**, 9038-9053, doi:10.7150/thno.63416 (2021).

61 Yu, G. *et al.* GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976-978, doi:10.1093/bioinformatics/btq064 (2010).

62 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).

63 Kim, N. *et al.* Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun* **11**, 2285,

doi:10.1038/s41467-020-16164-1 (2020).

64 Wu, F. *et al.* Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat Commun* **12**, 2540, doi:10.1038/s41467-021-22801-0 (2021).

65 Gene Ontology, C. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, doi:10.1093/genetics/iyad031 (2023).

66 Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941-D947, doi:10.1093/nar/gky1015 (2019).

67 Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411, doi:10.1126/science.aax0249 (2020).

68 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nat Commun* **12**, 1088, doi:10.1038/s41467-021-21246-9 (2021).

69 Huang, J. K. *et al.* Systematic Evaluation of Molecular Networks for Discovery of Disease Genes. *Cell Syst* **6**, 484-495 e485, doi:10.1016/j.cels.2018.03.001 (2018).

70 Hu, F. F., Liu, C. J., Liu, L. L., Zhang, Q. & Guo, A. Y. Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. *Brief Bioinform* **22**, doi:10.1093/bib/bbaa176 (2021).

71 Schwartz, L. H. *et al.* RECIST 1.1-Update and clarification: From the RECIST committee. *Eur J Cancer* **62**, 132-137, doi:10.1016/j.ejca.2016.03.081 (2016).

72 Altman, D. G., Lausen, B., Sauerbrei, W. & Schumacher, M. Dangers of using "optimal" cutpoints in the evaluation of prognostic factors. *J Natl Cancer Inst* **86**, 829-835, doi:10.1093/jnci/86.11.829 (1994).

73 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).