Comprehensive pan-cancer characterization of cancer-testis genes at single-cell resolution

Authors

Chunyang Fu[1], Shumin Yin[1], Linjin Li[2], Xueying Liu[1], Xiuyuan Jin[1], Ke Liu[1*]

Affiliations

1. [1]Department of Medical Dataology, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China.
2. [2]Administration Office, Shenzhen People's Hospital, Shenzhen, China.

First author: Chunyang Fu

Corresponding author: Ke Liu (keliu.iluke@email.sdu.edu.cn)

Abstract

Cancer-testis genes (CTGs) represent promising therapeutic targets due to their restricted expression in the testis and aberrant activation in cancers. However, their expression, regulation, and clinical relevance have not been systematically characterized at single-cell resolution. Here, by integrating large-scale single-cell transcriptomic datasets across 12 cancer types, we identified 407 CTGs. Autosomal CTGs exhibited broader expression across samples than those on the X chromosome. CTG activation in malignant cells was closely linked to epigenetic regulation, with chromatin remodelers *ARID4B* and *YEATS2* playing pivotal roles. Malignant cells consistently expressed a greater number of CTGs across cancer types, which led us to develop a computational pipeline for rapid malignancy annotation. Clinically, several CTGs showed potential as targets for T-cell receptor-engineered T-cell (TCR-T) therapy. Furthermore, CTG activation appeared to foster an immune-evasive tumor microenvironment. Collectively, we established a single-cell atlas of CTGs, deepening mechanistic understanding and facilitating their translational application in cancer treatment.

Introduction

Cancer-testis genes (CTGs) represent a distinct class of genes characterized by high expression in the testis, aberrant activation in various cancer types, and minimal or absent expression in most normal tissues (*1, 2*). Owing to the immune-privileged status of the testis, many CTG-encoded proteins—termed cancer-testis antigens (CTAs)—are immunogenic and serve as promising candidates for cancer vaccine development (*3, 4*). Beyond their immunogenicity, emerging evidence indicates that CTGs contribute to diverse intrinsic oncogenic programs, including malignant proliferation (*5*), metastatic dissemination (*6*), and therapeutic resistance (*7*). In addition, CTG expression in

malignant cells can remodel the tumor immune microenvironment, thereby promoting tumor progression through extrinsic mechanisms. For example, CTGs such as *KDM5B (8)* and *ITPRIPL1* (*9*) have been shown to regulate T cell infiltration and activation, ultimately enabling immune evasion.

Given the broad biological and clinical significance of CTGs, a comprehensive catalog of these genes would substantially advance research in this field. The CTdatabase (*10*) compiles over 200 CTGs reported in the literature, providing a valuable foundation for CTG research. More recently, Wang et al. (*11*) identified CTGs across 19 cancer types using transcriptomic data from The Cancer Genome Atlas (TCGA) (*12*), while Carter et al. (*13*) performed a similar analysis based on sequencing data from the GTEx (*14*) and Genomic Data Commons (GDC) (*15*) databases. These studies have greatly advanced the field and markedly expanded the CTG repository.

However, these omics-based studies rely primarily on bulk RNA sequencing (bulk RNA-seq) data, which present several inherent limitations. First, bulk RNA-seq measures average gene expression across heterogeneous cellular populations, potentially confounding CTG detection due to non-malignant cell contamination. Second, elucidating the regulatory mechanisms of CTGs requires accurate transcriptional profiles of malignant cells—something bulk RNA-seq cannot readily provide. Third, bulk RNA-seq fails to capture intratumoral transcriptional heterogeneity, thereby obscuring cell type-specific or subclonal CTG activation patterns that could have important biological and clinical implications.

The advent of single-cell transcriptomic technologies provides a powerful means to overcome these challenges, enabling precise quantification of gene expression at cellular resolution. Publicly available single-cell resources—such as 3CA (*16*), TISCH2 (*17*), and HTAN (*18*)—now encompass hundreds of cancer studies with detailed cell-type annotations, providing an unprecedented opportunity for comprehensive pan-cancer discovery and refined characterization of CTGs at single-cell resolution. In this study, we systematically identified 407 CTGs across 12 cancer types through large-scale analysis of curated single-cell and single-nucleus RNA sequencing (scRNA-seq and snRNA-seq) datasets (**Fig. 1A**). We comprehensively characterized their expression patterns and regulatory features, and demonstrated their utility for rapid malignancy annotation across sc/snRNA-seq, single-nucleus ATAC-seq (snATAC-seq), and spatial transcriptomic (ST) datasets. We further examined their clinical relevance and identified several CTGs as potential targets for T-cell receptor-engineered T-cell (TCR-T) therapy. Notably, CTG activation in malignant cells appears to promote an immune-evasive tumor microenvironment, particularly in hepatocellular carcinoma, where it correlates with increased infiltration of regulatory T cells (Tregs).

Results

Definition of cancer-testis genes

We developed a systematic pipeline to identify CTGs across multiple cancer types based on single-cell transcriptomic data (**Fig. 1B**). We began by compiling a list of 1,294 unique CTG candidate genes from two sources: 242 previously curated CTGs from the CTdatabase, and 1,212 testis-enriched protein-coding genes from the Human Protein Atlas (*19*) (**Table S1**; **Fig. S1A**). As expected, the majority of these genes exhibited peak expression in testicular germ cells among 81 normal human cell types (**Fig. S1B**). Next, we curated 62 single-cell and single-nucleus RNA-seq datasets generated using the 10x Genomics platform to construct a discovery cohort comprising 828 tumor samples, spanning 12 cancer types (**Table S2**; **Fig. S1C–F**). These included glioblastoma (GBM); breast cancer (BRCA), subtyped into triple-negative (TNBC) and non-TNBC; colorectal cancer (CRC); head and neck squamous cell carcinoma (HNSCC); clear cell renal cell carcinoma (ccRCC); hepatocellular carcinoma (HCC); lung adenocarcinoma (LUAD); neuroblastoma (NB); high-grade serous ovarian cancer (HGSOC); pancreatic ductal adenocarcinoma (PDAC); prostate cancer (PRAD); and melanoma (SKCM). For each cancer type, at least 16 tumor samples were included. We defined a gene *g* as a CTG for cancer type *C* if its expression was detected in more than 1% of malignant cells in at least half of the samples of *C*. Applying this criterion, we identified 407 unique CTGs (**Table S4**), with a median of 173 CTGs per cancer. Among them, SKCM and ccRCC exhibited the highest and lowest CTG numbers, respectively (**Fig. 1C**). Notably, CTG count was not significantly correlated with sample size or the total number of expressed genes per cancer type (**Fig. 1C**), suggesting that CTG variability reflects intrinsic tumor biology rather than technical artifacts.

To validate the quality of the identified CTGs, we performed two independent analyses. First, we assessed CTG expression in malignant cells using an assembled Smart-seq2 cohort (validation cohort 1; 172 tumor samples across nine cancer types; **Table S2**; **Fig. S1C–F**). In seven cancer types, more than 65% of CTGs exhibited expression levels exceeding 1 Transcripts Per Million (TPM) (**Fig. 1D**). Although LUAD and PRAD exhibited relatively lower expression, median TPM values still approximated to 1. Moreover, the proportion of malignant cells expressing each CTG was highly correlated between the 10x Genomics and Smart-seq2 platforms (**Fig. S2A**), supporting the robustness and reproducibility of CTG detection using 10x-based data.

Second, we assessed the ability of the identified CTGs to distinguish tumor from normal samples using bulk RNA-seq data from the TCGA and GTEx projects. When aggregated CTG expression was applied as a classification feature, the median area under the ROC curve (ROC-AUC) achieved 0.89 across cancer types in the discovery cohort (**Fig. 1E**). Additionally, CTG expression was positively correlated with tumor purity in most cancers (median Spearman's $\rho = 0.48$; **Fig. 1F**). By contrast, ccRCC displayed both minimal tumor-normal expression difference and weak purity correlation, potentially attributable to the smaller number of CTGs identified in this tumor type, consistent with its previously reported CTG-poor phenotype (*20*).

Interestingly, similar results were obtained even in cancer types not included in the discovery cohort, further underscoring the generalizability of our CTG set (**Fig. 1E–F**). We further compared our CTG set against the initial candidate genes and previously defined CTGs (*10, 11, 13*). While our CTG set demonstrated comparable tumor-normal discrimination to established CTG sets, it markedly outperformed the original candidate genes (**Fig. S2B**). Moreover, it demonstrated significantly higher tumor purity correlation than prior CTG sets (**Fig. S2C**).

In summary, by integrating large-scale single-cell transcriptomic datasets, we systematically identified and validated 407 CTGs across diverse human cancers.

Cancer-type specificity of CTGs

We performed Uniform Manifold Approximation and Projection (UMAP) analysis (*21*) on TCGA tumor samples using the CTGs defined above. The analysis revealed clear segregation of tumor samples based on cancer type (**Fig. 2A**). Specifically, breast invasive carcinoma samples were partitioned into two distinct clusters corresponding to TNBC and non-TNBC. Squamous cell carcinomas—including CESC, LUSC, ESCA, BLCA, and HNSC—clustered together, consistent with prior evidence of conserved molecular programs across squamous cancers arising from divergent anatomical sites (*22*). These findings indicated that CTG expression robustly captures lineage-related molecular programs across various cancer types.

To further characterize CTG expression patterns, we quantified the number of cancer types in which each CTG was detected and examined the resulting distribution. Interestingly, the distribution appears to be bimodal, suggesting that most CTGs were either highly specific to a particular cancer type, or broadly expressed across multiple cancer types (**Fig. 2B**). We therefore stratified the CTGs into three categories: cancer-specific ($n \leq 3$), pan-cancer ($n \geq 11$), and others. The proportions of CTGs in each category were 48.6%, 23.6%, and 27.8%, respectively.

Given the predominance of cancer-specific CTGs, we next examined their expression profiles in malignant cells. As anticipated, the majority of these CTGs exhibited markedly confined expression to individual cancer types (**Fig. 2C**). For instance, *ROPN1* was predominantly expressed in TNBC, while *IL13RA2* showed marked expression in GBM. Within breast cancer subtypes, *ELL3* and *TEKT5* were primarily expressed in non-TNBC, whereas *ROPN1* showed stronger expression in TNBC (**Fig. 2C**). This subtype-specific pattern was recapitulated in luminal/HER2-amplified and basal-A breast cancer cell lines (**Fig. 2D**). The observed expression specificities were further validated using bulk RNA-seq data from TCGA and TARGET (*23*) (**Fig. S3A**).

X-CTGs exhibit pronounced sample-restricted expression patterns

CTGs are commonly stratified by chromosomal location into autosomal CTGs and X-chromosome CTGs (X-CTGs) (*1*), and the latter group has attracted particular interest since the initial characterization of the cancer-testis antigen MAGEA1 (*24*). Among the initial 1,294 CTG candidates, 14.5% were located on the X chromosome; however, in the final CTG list, the proportion declined to 6.1%. Surprisingly, multiple canonical X-CTGs were not detected by our pipeline, such as *CTAG1B* (also known as NY-ESO-1), and MAGEA family members (e.g., *MAGEA2*, *MAGEA4*). Fisher's exact test demonstrated that the CTGs we identified were predominantly enriched on the autosomes (OR = 3.55; **Fig. 2E**).

The unexpected exclusion of these canonical X-CTGs prompted us to further investigate their expression profiles in malignant cells. Although some X-CTGs were undetectable at the single-cell level, others—particularly MAGEA family genes—showed high expression only in a minority of tumor samples, with negligible or absent expression in the remainder (**Fig. S4A**). This sample-restricted expression pattern contrasts starkly with the broadly expressed CTGs in our list. To quantify sample specificity, we calculated the tau (*25*) index for each identified CTG. This index ranges from 0 to 1, with values approaching 1 indicating greater expression restriction to specific samples. Remarkably, we observed that X-CTGs exhibited significantly higher tau values than autosomal CTGs across most cancers in the discovery cohort (**Fig. 2F**). An exception was HCC, which displayed the lowest median tau for X-CTGs, slightly below that of autosomal CTGs, although the difference was not statistically significant. Analysis of TCGA and TARGET bulk RNA-seq data also confirmed elevated expression of multiple X-CTGs in HCC compared to other cancers (**Fig. S4B**), suggesting an aberrant activation of X-CTGs in HCC.

Independent validation using RNA-seq data from CCLE (*26*) further corroborated the heightened sample specificity of X-CTGs (**Fig. S4C**), indicating that this pattern is not attributable to the sparsity of scRNA-seq data. Moreover, CTGs excluded by our pipeline—whether X-linked or autosomal—consistently displayed more restricted expression patterns than those retained, reinforcing the validity of our pipeline in identifying CTGs with broad expression across tumor samples.

Delineation of regulatory mechanisms of CTGs

Currently, the regulatory mechanisms underlying CTG activation remain incompletely characterized. To investigate this, we performed co-expression analysis using pseudo-bulk profiles of malignant cells and found CTGs exhibited significantly stronger correlations with each other than with non-CTGs (**Fig. 3A**), suggesting the transcription of CTGs is highly coordinated. Notably, similar patterns were also observed in TCGA and TARGET bulk RNA-seq datasets (**Fig. S5A**). We next employed CS-CORE (*27*) to reconstruct co-expression networks of CTGs in malignant cells. Remarkably, a CTG

cluster enriched in canonical cell cycle-associated genes (e.g., *ATAD2* and *CENPF*) emerged across multiple cancer types (**Fig. 3B and S5B**), indicating a tight coupling between neoplastic proliferation and CTG transcriptional activation.

Given the extensive co-expression observed among CTGs, we hypothesized that their upstream transcriptional regulators would exhibit broad positive correlations with them. In each cancer type of the discovery cohort, we ranked non-CTG genes by the number of CTGs with which they exhibited strong positive co-expression and performed enrichment analysis on the top 2,000 genes (**Fig. 3C**). As anticipated, cell-cycle related terms were significantly enriched in multiple cancer types; in addition, we found a surprising enrichment of genes involved in cilium assembly, echoing our earlier observation that numerous germ cell-specific genes are also expressed in ciliated cells (**Fig. S1B**). Notably, pathways associated with epigenetic regulation (including "chromatin modifying enzymes" and "chromatin organization") were likewise significantly enriched across cancer types (**Fig. 3C**), indicating epigenetics may play important roles in the regulation of CTG activation. Since cell cycle is also under tight regulation of epigenetic mechanisms (*28, 29*), we wondered whether the enrichment could arise from CTGs simultaneously influenced by both processes. We partitioned CTGs into cell cycle-associated and non-cell cycle groups, and repeated the above enrichment analysis. The former displayed exclusive enrichment for mitotic pathways, whereas the latter were enriched for epigenetic pathways (**Fig. S6A–B**). Similar findings were obtained in TCGA and TARGET bulk RNA-seq data (**Fig. S6C–E**). These results underscored the critical influence of epigenetics on CTG regulation, particularly for non-cell cycle-associated CTGs.

To further identify potential epigenetic regulators involved in CTG activation, we extracted genes annotated in the "chromatin modifying enzymes" pathway and ranked them by the number of cancer types in which they were detected (**Fig. 3D**). Notably, most of these genes targeted histones, as annotated in the EpiFactors database (*30*), underscoring the central role of histone modifications in CTG regulation. Six genes—*ARID4B*, *KDM3B*, *PBRM1*, *SETD2*, *SUZ12*, and *YEATS2*—were consistently identified among the top-ranked epigenetic regulators across both sc/snRNA-seq and bulk RNA-seq datasets (**Fig. 3D**). To confirm the regulatory roles of *YEATS2* and *ARID4B*, we analyzed bulk RNA-seq data from genetic perturbation experiments (*31, 32*). Gene Set Enrichment Analysis (GSEA) (*33*) revealed that CTGs were significantly enriched among the downregulated genes following knockdown (**Fig. 3E and S6F**), confirming their roles in CTG activation. Given the high mutation frequency of *PBRM1* (*34*) and "CTG-poor phenotype" of ccRCC, we further examined whether its loss was associated with CTG inactivation. Differential expression analysis showed reduced *PBRM1* expression in mutated samples compared with wild-type tumors (**Fig. S6G**), and GSEA demonstrated that CTGs absent from ccRCC were significantly enriched among the downregulated genes (**Fig. S6H**), suggesting that *PBRM1* may act as a key modulator of CTG expression.

Finally, we leveraged a snATAC-seq dataset (*35*) to evaluate chromatin accessibility correlations between CTGs and the six candidate epigenetic regulators identified above. Genes were ranked by their chromatin accessibility correlation with each regulator, and subsequent GSEA revealed strong positive correlations between CTGs and these regulators across multiple cancer types (**Fig. 3F**). Notably, in CRC, HNSCC, PDAC, and SKCM, at least three regulators exhibited such positive correlations, further emphasizing the coordinated epigenetic control of CTG activation.

Ultra-fast annotation of malignant cells using CTGs

Annotation of malignant cells in cancer single-cell transcriptomic data is usually the foundation of downstream analyses. Building on our previous observation that CTGs effectively distinguish tumor from normal samples in bulk RNA-seq datasets, we investigated their capacity to differentiate malignant from non-malignant cells at single-cell resolution. Interestingly, we found that malignant cells consistently exhibited higher number of unique CTGs detected per cell (abbreviated as n_CTGs) than normal cell types across multiple cancer types in both the discovery cohort and validation cohort 1 (**Fig. 4A–B and S7A**). Since the n_CTGs value derived from the full 407 CTGs correlated strongly with those based on CTGs identified in corresponding individual cancers (**Fig. S7B**), we employed the complete CTG set for all n_CTGs associated analyses to ensure consistency.

To validate the generalizability of the finding, we assessed n_CTGs differences in 21 independent sc/snRNA-seq datasets (validation cohort 2; **Table S2**; **Fig. S1C–F**) encompassing 17 cancer types, 12 of which were not even included in the discovery cohort. Consistent with our initial finding, malignant cells exhibited significantly higher n_CTGs values (**Fig. 4B and S8A**). To further confirm this elevation was attributable to malignancy rather than lineage-intrinsic expression patterns, we defined normal reference cells for each cancer type based on its cellular origin (e.g., epithelial cells for carcinoma, glial cells for GBM). In the majority of datasets, malignant cells displayed significantly higher n_CTGs than their corresponding normal reference cells (**Fig. S8B**), thereby substantiating high n_CTGs as a malignancy-specific feature. In addition, analysis of the aforementioned snATAC-seq dataset further confirmed the finding (**Fig. S8C**), suggesting the utility of n_CTGs as a robust and scalable metric for malignant cell annotation across multi-omics data.

We next evaluated the performance of n_CTGs for annotation of malignant cells within individual samples. Notably, in the discovery cohort, 84.6% (11/13) of cancer types achieved a median ROC-AUC ≥ 0.75 (**Fig. 4C**), although ccRCC exhibited a relatively lower performance (median ROC-AUC = 0.55), consistent with its CTG-poor characteristic. As expected, such ideal performance could also be achieved in the two validation cohorts (**Fig. S9A**). Importantly, the n_CTGs index outperformed both aggregated CTG expression and n_Genes (the total number of genes detected per cell) in discriminating malignant from non-malignant cells in both discovery and validation

cohorts (**Fig. S9A–C**). Furthermore, we did not observe significant performance drop of a specific platform, suggesting the robustness of the index (**Fig. S9D–E**).

To further enhance the discriminatory power of n_CTGs in identifying malignant cells, we further developed a classification pipeline based on SingleR (*36*) (termed SingleR-CTG) (**Fig. 4D**). Given an sc/sn-RNAseq dataset, individual cells were first ranked by n_CTGs, with the top and bottom 5% designated as malignant-like and normal-like, respectively. Following pseudo-bulk aggregation, these malignant-like and normal-like cells were used to create a two-column reference expression matrix. SingleR was then employed to compute the transcriptomic correlation (which we termed as SingleR-CTG score) of each single cell with the malignant-like reference. As expected, the SingleR-CTG score outperformed the raw n_CTGs index in the three cohorts (**Fig. 4F**). For example, in a representative TNBC sample from the discovery cohort, this score substantially improved the separation between malignant and normal cells relative to raw n_CTGs (**Fig. 4E**).

When benchmarking against established malignant cell annotation tools, SingleR-CTG demonstrated comparable classification accuracy to inferCNV (*37*) and slightly outperformed ikarus (*38*) and scCancer2 (*39*) in both discovery and validation cohorts (**Fig. 4F**). When the analysis was restricted to malignant and normal reference cells, similar results were observed (**Fig. S9F**). Notably, although inferCNV (the most widely used gold-standard tool for malignancy annotation) and SingleR-CTG yielded similar performances, our method offered substantial computational accelerations (**Fig. 4G**). For instance, in datasets with ~5,000 cells, inferCNV required about 11 minutes for completion, whereas SingleR-CTG only took 1.3 seconds. This characteristic makes our method particularly suitable for large-scale malignancy annotation.

The n_CTGs acts as an effective surrogate of tumor purity in spatial transcriptomics data

Spatial transcriptomics technologies enable the interrogation of tumor subclonal architecture and spatial interactions between cancer cells and the tumor microenvironment (*40*). Efficient and generalizable identification of malignant spots is a prerequisite for exploiting expanding tumor ST datasets.

To investigate the ability of n_CTGs in malignant spot identification, we first analyzed 12 oral squamous cell carcinoma (OSCC) samples with expert pathological annotations (*41*). As anticipated, n_CTGs were significantly higher in malignant spots in all cases (**Fig. 5A**). When using n_CTGs to predict malignant spots, the median ROC-AUC reached 0.74 (**Fig. S10A**). Similarly, in an HCC ST dataset with definitive cell annotations (*42*), n_CTGs were markedly elevated in malignant cells relative to all other cell types (**Fig. 5B**), reinforcing the malignancy-specific nature of n_CTGs.

We next collected 245 ST slides (encompassing 15 cancer types) to comprehensively assess the utility of n_CTGs in malignant spot identification (**Table S2**). Given that the spots typically contain multiple cells (ranging from 1 to 10), ESTIMATE (*43*) was used to infer tumor purity for each spot. Notably, n_CTGs demonstrated a strong positive correlation with estimated tumor purity (Spearman's $\rho > 0.5$) in 10 of 15 cancer types (**Fig. 5C–D and S10B**), confirming its utility as a purity surrogate. As observed in sc/sn-RNAseq data, n_CTGs showed superior correlation with tumor purity than either aggregate CTG expression or n_Genes (**Fig. S10C**). Recognizing that aneuploidy is a hallmark of malignancy and a widely accepted indicator of malignant cell identity (*44*), we further assessed the relationship between n_CTGs and copy number variation (CNV). As expected, n_CTGs were significantly higher in CNV-high clusters compared to CNV-low counterparts (**Fig. 5E**) and demonstrated a positive correlation with CNV scores across multiple cancer types (**Fig. S10D**). Collectively, these findings suggest that n_CTGs serves as a robust proxy for tumor purity in ST data and provides a practical index for the rapid annotation of malignant spots.

Identification of CTGs as potential TCR-T therapy targets

In recent years, TCR-T cell therapy targeting CTGs has shown promising efficacy in treating solid tumors. Tecelra (*45*), which targets the canonical cancer-testis gene *MAGEA4*, was the first FDA-approved TCR-T therapy and demonstrated encouraging effects in the treatment of advanced synovial sarcoma; in addition, TCR-T therapies targeting *ROPN1* (*46*) and *PRAME* (*47*) (the two CTGs identified in our study) have exhibited promising efficacy in TNBC and SKCM.

Robust tumor expression coupled with negligible expression in normal tissues forms the foundation for TCR-T target identification, ensuring selective tumor cytotoxicity while minimizing off-tumor toxicity (*48*). Motivated by these results, we applied a filtering strategy to the 407 identified CTGs to prioritize potential TCR-T therapy targets (**Materials and Methods**) and ultimately derived 37 genes (**Fig. 6A**). Surprisingly, only *CT83* and *MAGEA1* among these have ongoing TCR-T clinical trials, underscoring that the therapeutic potential of CTGs remains largely underexplored. For example, *CSAG1* is broadly expressed in malignant cells of HCC while displaying minimal expression in normal tissues, making it an attractive candidate TCR-T target in HCC; similarly, *HMX1* appears to be a promising target for neuroblastoma, the most common pediatric solid tumor. Cell-lineage expression analysis (*49*) further confirmed that both genes are selectively expressed in malignant cells (**Fig. 6B**).

Clinical relevance of CTGs

To evaluate the clinical relevance of CTGs, we systematically analyzed associations between individual CTG expression levels and patient overall survival across cancer types (**Table S5**). Among the 407 CTGs assessed, 129 (31.2%) demonstrated significant associations with survival in at least one cancer type. Notably, CTGs were significantly

more enriched for survival-associated genes than non-CTGs in GBM, HCC, LUAD, and PDAC (**Fig. 7A**). A similar trend was also observed in ccRCC, though not reaching statistical significance. Interestingly, several CTGs showed context-dependent survival associations. For example, elevated *SPATA24* expression correlated with poorer survival in HCC (HR = 1.43, p = 0.00098) and NB (HR = 1.73, p = 0.00023), while being associated with improved outcomes in PDAC (HR = 0.60, p = 0.00236) and ccRCC (HR = 0.67, p = 0.00004).

Given the established role of CTGs in modulating the tumor immune microenvironment (*8, 9*), we subsequently investigated the relationship between CTG expression levels in malignant cells and immune cell infiltration using sc/snRNA-seq data from the discovery cohort (**Fig. 7B**). To minimize confounding from immune remodeling in metastatic lesions, this analysis was restricted to primary tumors. In half of the tumor types evaluated, CTG expression level in malignant cells was significantly correlated with the abundance of at least three immune cell types, and most frequently demonstrating a negative correlation with $CD8^+$ T cells and a positive correlation with M2 macrophages. Interestingly, HCC exhibited a distinct pattern, with elevated CTG expression being associated with increased infiltration of Tregs (**Fig. 7B and S11A**). These findings suggest that CTG activation in malignant cells may promote an immunosuppressive microenvironment. Consistent with this, elevated CTG expression was associated with downregulation of immune activation pathways (e.g., interferon gamma response, inflammatory response) in malignant cells across multiple cancer types (**Fig. 7B**).

We next conducted a deeper investigation into HCC to elucidate the potential mechanisms, as elevated CTG expression in this tumor type was associated with poorer overall survival and more advanced clinical stage (**Fig. 7C–D**). We used two independent HCC scRNA-seq datasets: one from the discovery cohort (Lu et al.) (*50*), and another from Sun et al. (*51*). In Sun's dataset, samples with higher CTG expression in malignant cells also exhibited significantly elevated levels of Treg infiltration (**Fig. 7E**), consistent with the findings from the discovery cohort. We performed differential gene expression analysis between CTG-high and CTG-low malignant cells in both datasets, identifying 62 consistently upregulated and 96 consistently downregulated genes (**Fig. 7F**). As expected, we observed significant upregulation of cell cycle regulators (e.g., *BIRC5*, *TOP2A*) and downregulation of genes involved in antigen presentation (e.g., *CD74*, *HLA-DRA*). Surprisingly, we noticed that cholesterol biosynthesis-related genes (e.g., *INSIG1*), cytochrome P450 enzymes (e.g., *CYP27A1*, *CYP2A6*), and apolipoproteins (e.g., *APOB*, *APOC1*) were down-regulated. Hallmark (*52*) pathway analysis corroborated these observations, showing increased activity of proliferation-related programs (e.g., *E2F* and *MYC* targets), repression of immune response (e.g., interferon gamma response, complement cascade) and metabolic pathways (e.g., cholesterol homeostasis, bile acid metabolism) (**Fig. S11B–C**). Considering the liver's central role in lipid metabolism, the profound suppression of hepatic metabolic programs in CTG-high malignant cells suggests a dedifferentiated,

less hepatocyte-like state, which is further confirmed by CytoTRACE2 analysis (*53*) (**Fig. 7G**). We next performed complementary analysis using TCGA HCC bulk RNA-seq data and derived consistent conclusions (**Fig. 7H and S11D–E**). Taken together, these findings suggest that elevated CTG expression characterizes a stem-like status of HCC, which may explain the immune-evasive tumor microenvironment and unfavorable clinical outcomes.

Discussion

CTGs have garnered considerable attention in cancer research (*1-4*). In contrast to previous CTG catalogs constructed using bulk RNA-seq data from tumor tissues (*11, 13*), we refined a comprehensive compendium of 407 CTGs across 12 prevalent cancer types by leveraging single-cell transcriptomic data. This strategy effectively mitigates the confounding influence of non-malignant cells within the tumor microenvironment, providing a more accurate representation of CTG expression in cancer cells.

Interestingly, nearly half of the identified CTGs displayed cancer-type-specific expression. This observation raises an important question: what mechanisms regulate the cancer-type-specific activation of CTGs? Most current knowledge of CTG regulation derives from canonical X-CTGs, where DNA methylation plays a central role (*54-56*). However, DNA methylation alone appears insufficient to fully determine context-specific gene regulation. As demonstrated by our study, chromatin remodelers also play a crucial role in CTG activation. In the future, we intend to utilize multi-omics technologies to further elucidate these regulatory mechanisms. Notably, such research holds substantial translational potential. If we can use some additional treatments to enhance the expression of CTGs in malignant cells, it could potentially boost the efficacy of immunotherapies targeting CTGs.

Historically, X-CTGs have been the focus of cancer immunotherapy research, particularly in the development of cancer vaccines and adoptive T-cell therapies (ACT) (*3, 4*). However, clinical outcomes from trials evaluating these therapies have been limited and inconsistent (*57, 58*), possibly due to the pronounced sample-restricted expression of X-CTGs. For example, for the canonical target NY-ESO-1 (*59, 60*), we scarcely detected its expression in malignant cells of the 828 tumor samples in our discovery cohort. In contrast, CTGs located on autosomes exhibit broader expression across tumor samples. We propose that prioritizing CTGs with broad expression could enhance the therapeutic applicability of CTG-based strategies, benefiting a larger patient population. Encouragingly, several ACT therapies targeting autosomal CTGs (e.g., *ROPN1*, *PRAME*, and *IL13RA2*) have shown promising results in preclinical and clinical trials (*46, 47, 61*), further underscoring their translational potential.

We found that malignant cells consistently exhibited higher n_CTGs than normal cell types across more than 20 cancer types, as revealed by scRNA-seq, snATAC-seq, and

ST data. Although prior studies have reported that tumor cells generally display heightened transcriptional activities (*62, 63*) (**Fig. S9B**), such elevated n_CTGs in malignant cells appear not to be merely a byproduct of global transcriptional increases. Instead, malignant cells seem to "purposely" up-regulate the expression of CTGs. This hypothesis is supported by two key observations: first, n_CTGs exhibited significantly greater discriminatory power for malignant cell identification compared to n_random_genes (the number of unique non-CTG genes detected per cell) (**Fig. S9A**). Second, even after adjusting for global transcriptional activity, n_CTGs remained significantly elevated in malignant cells relative to normal cells, whereas n_random_genes did not (**Fig. S9C**). These findings highlight the unique and selective nature of CTG activation in malignant cells.

Motivated by the consistently higher n_CTGs in malignant cells, we developed the SingleR-CTG pipeline, which enables robust malignant cell annotation across multiple cancer types. SingleR-CTG performed comparably to the widely used inferCNV, while achieving substantially faster runtimes. In addition, SingleR-CTG does not require the manual selection of reference normal cells, a step that often influences inferCNV results. These two features make SingleR-CTG an outstanding tool for large-scale integration analyses of cancer scRNA-seq data.

Clinically, CTGs were significantly enriched for genes associated with patient survival relative to non-CTGs, suggesting their obligate participation in multiple hallmarks of cancer (*64*). Particularly in HCC, CTG-high malignant cells exhibited an immune-evasive, stem-like phenotype with enhanced proliferative activity and increased Treg infiltration, which may explain the negative association between CTG activation and overall survival. Future studies focusing on the perturbation of CTGs to remodel tumor-immune interactions would be worth further exploring.

In summary, our comprehensive pan-cancer analysis presents a systematic CTG catalog with broad tumor applicability. The insights gained from this study not only enhance our understanding of CTG biology but also pave the way for developing novel therapeutic strategies targeting CTGs. Despite these advances, there are several limitations to consider. First, tumor evolution is a dynamic, multi-stage process. Our study focused solely on primary and metastatic specimens, which may overlook CTGs specifically activated in circulating tumor cells or precancerous lesions. Especially, identifying CTGs that emerge during premalignant stages may provide sensitive biomarkers for early detection of cancer. Second, most of the cancer types included in this study were adult cancers, and it remains unclear whether our conclusions are applicable to pediatric cancers. As multistage, longitudinal single-cell omics datasets accumulate across more cancer types, iterative refinement of the CTG compendium is expected to enhance its precision and clinical utility.

Materials and Methods

Assembly of CTG candidate genes

A total of 276 genes were retrieved from the CTdatabase (http://www.cta.lncc.br/). After removal of genes lacking approved HGNC symbols and corresponding Ensembl IDs, 242 genes were retained. In parallel, testis-enriched protein-coding genes were extracted from the Human Protein Atlas (HPA; https://www.proteinatlas.org/). Genes were retained if their expression level in testis or a tissue group (two to five tissues including testis) was at least fourfold higher than in any other tissue. A total of 1,212 genes with approved HGNC symbols met this criterion. The union of these two sources yielded 1,294 unique CTG candidates (**Table S1**).

Datasets collection

A total of 96 well-annotated cancer sc/snRNA-seq datasets were collected, primarily from the Single Cell Portal (https://singlecell.broadinstitute.org/single_cell), 3CA, HTAN, and TISCH2. For each dataset *d*, the following processing steps were applied: (1) if *d* encompassed samples from distinct cancer subtypes or sequencing platforms, then it was stratified into separate datasets; (2) only tumor samples containing more than 10 malignant cells and adjacent normal samples were retained; and (3) specimens obtained from the same patient at multiple timepoints (e.g., pre- and post-treatment) or anatomical sites (e.g., tumor core and edge) were treated as independent samples when sufficient metadata were available. These datasets constituted the discovery and two validation cohorts (**Table S2**).

ST data comprising 245 tumor sections across 15 cancer types were collected from publicly available repositories, including GEO, SCAR (*65*), HTAN, Mendeley Data, and CROST (*66*). The snATAC-seq dataset from Terekhanova et al. was retrieved from HTAN. Comprehensive metadata for these datasets are provided in **Table S2**.

Bulk RNA-seq data from TCGA, TARGET, and GTEx projects were downloaded from UCSC Xena (https://xenabrowser.net/datapages/). Bulk RNA-seq data for CCLE was retrieved from the DepMap portal (https://depmap.org/portal/). Differential expression data from gene knockdown experiments were accessed through GEO (GSE90781, GSE116668). All data sources for bulk RNA-seq analyses are detailed in **Table S3**.

sc/snRNA-seq data processing

Given that all the datasets have already gone through rigorous quality control in their corresponding original analysis, no additional quality control procedure was performed in this analysis. For each dataset, the raw UMI count or TPM expression matrix were normalized using the LogNormalize method in the NormalizeData function to mitigate the influence of variable sequencing depth across cells. Subsequently, the top 2,000 highly variable genes were identified using the vst method in the FindVariableFeatures function. The expression matrix was scaled using the ScaleData function, and principal

565 component analysis (PCA) was performed for dimensionality reduction via the
566 RunPCA function. A two-dimensional embedding was generated using the RunUMAP
567 function on the top 30 principal components (PCs) for visualization. All the procedures
568 were conducted using the Seurat R package (v4.4.0) (*67*).
569
570 When the analysis was restricted to normal cells, after PCA, harmony (v1.2.0) (*68*) was
571 run with default parameters to correct for potential batch effects among samples within
572 each dataset. UMAP was then performed using the top 30 batch-corrected PCs.
573
574
575 **Cell identity validation and immune subtype annotation**
576 To harmonize the original cell annotations provided in each dataset, all cells were
577 classified into eight major categories: Normal Epithelial, Stromal, Immune, Brain Glial,
578 Skin Melanocyte, Malignant, Other (cell types not falling into the preceding categories,
579 e.g., adipocytes and myocytes), and Unknown (e.g., low-quality or unclassified cells).
580
581 To validate the identity of the five major normal cell types (Normal Epithelial, Stromal,
582 Immune, Brain Glial, and Skin Melanocyte), two complementary analyses were
583 conducted for each dataset. First, UMAP (restricted to normal cells, following batch
584 correction) was inspected to verify that each cluster predominantly comprised cells of
585 a single annotated type. Second, differential expression analysis was performed using
586 the FindAllMarkers function in Seurat to obtain the top 20 upregulated genes per cell
587 type. These marker genes were subjected to enrichment analysis against CellMarker
588 (*69*) gene sets using compareCluster function in the clusterProfiler R package (v4.6.2)
589 (*70*) to check whether the corresponding gene sets were up-regulated in each cell type.
590
591 Malignant cell annotations were validated using two orthogonal criteria. First,
592 malignant cells from distinct patients were expected to form separate clusters on the
593 UMAP (constructed using all cells without batch correction), reflecting high inter-
594 patient transcriptional heterogeneity, whereas normal cells of the same type typically
595 clustered together. Second, malignant cells generally exhibit extensive copy number
596 alterations (CNAs); accordingly, their inferred CNV scores (see **Inference of Copy**
597 **Number Variations in sc/snRNA-seq Data**) were expected to be significantly higher
598 than those of normal cells.
599
600 Immune cell subtype annotations were refined using SingleR (v2.0.0). Initial annotation
601 was performed against the reference constructed from Blueprint and ENCODE data to
602 confirm immune lineage identity. One dataset ("Data_Kidney_Obradovic2021") was
603 excluded from downstream immune infiltration analysis, since only 17.5% of its
604 annotated immune cells mapped to immune references. High-resolution subtype
605 annotation was then conducted using the LM22 (*71*) signature matrix as the reference.
606
607
608 **Inference of copy number variations in sc/snRNA-seq data**

Large-scale copy number variations (CNVs) in individual cells were inferred at the sample level using the infercnv R package (v1.17.0). For each sample, the raw count matrix was used as input. Reference cells were defined as the Stromal or Immune population with the largest number of cells; if only malignant annotations were available, all non-malignant cells were designated as reference. Gene-level CNV values were calculated by subtracting 1 from the residual expression values. The CNV score for each cell was defined as the mean of the squared gene-level CNV values, reflecting the deviation from the inferred diploid baseline and serving as a proxy for the magnitude of large-scale copy number alterations.

Calculation of aggregated CTG expression

For each sample or pseudo-bulk sample, CTG set expression was quantified using single-sample Gene Set Enrichment Analysis (ssGSEA) (*72*) via the gsva function in the GSVA package (v1.46.0) (*73*). For individual cells, aggregated CTG expression was computed using the AddModuleScore function in Seurat.

Calculation of tau index

For a given gene *g* in cancer type *C*, let x_i denotes its expression in sample *i* ($i = 1, 2, ..., N$), where *N* is the total number of samples in *C*. The tau (τ) index for gene *g* was calculated as:

$$\tau = \frac{\sum_{i=1}^{N}(1-x_i/x_{\max})}{N-1} \quad (1)$$

where x_{max} represents the maximum expression observed across all samples. In the discovery cohort, x_i corresponded to the percentage of malignant cells in sample *i* with detectable expression of gene *g*. In the CCLE dataset, x_i was defined as the $\log_2$-transformed TPM value for gene *g*, and only solid tumor types represented by more than 30 cell lines were included.

CTG co-expression analysis

CTG co-expression was inferred using the CS-CORE R package (v0.0.0.9000). As this method requires sequencing depth information, only 10x Genomics datasets with raw UMI counts were included. Within each dataset, the CS-CORE estimate quantified the co-expression strength between CTG pairs. CTG pairs with adjusted *P*-values ≥ 0.01 were considered not significant, and their CS-CORE estimates were set to zero. Only CTG pairs showing consistent directionality (i.e., CS-CORE estimates across datasets within a cancer type were either all ≥ 0 or all ≤ 0) were retained. To summarize co-expression patterns across datasets for a given cancer type, the median CS-CORE estimate for each CTG pair was calculated and used as its representative co-expression score.

Co-expression networks were visualized using the ggraph R package (v2.1.0), retaining only CTG pairs with a co-expression score > 0.25. CTG clusters were identified using the cluster_louvain function (default parameters) implemented in the igraph R package (v1.5.1).

Definition of cell cycle-associated CTGs

For each dataset, AUCell (v1.20.2) (*74*) was employed to quantify the activity of G2/M and S phase gene sets in malignant cells. Let x and y denote the AUCell scores for G2/M and S phase gene sets, respectively. As the majority of malignant cells were non-cycling, a malignant cell i was classified as cycling if:

$$x_i > \text{median}(x) + 5 \times \text{MAD}(x) \text{ or } y_i > \text{median}(y) + 5 \times \text{MAD}(y) \quad (2)$$

where MAD refers to median absolute deviation.

For each dataset d within cancer type c, differential expression analysis was conducted between cycling and non-cycling malignant cells, restricted to CTGs identified in cancer type c. The Wilcoxon rank-sum test was performed using the FindMarkers function in Seurat, with pseudocount.use = 1e-9 to account for the generally low expression levels of CTGs. A CTG was deemed significantly upregulated in dataset d if it met the following criteria: adjusted P-value < 0.01, average $\log_2$ fold-change > 1, and a difference in the fraction of expressing cells > 0.1. Within cancer type c, CTGs upregulated in ≥ 50% of datasets were designated as cell cycle-associated; for cancer types with fewer than three datasets, the threshold was adjusted to ≥ 2 datasets.

Enrichment analysis of genes correlated with multiple CTGs

Non-CTG genes were defined as genes not classified as CTGs in any cancer type. For each non-CTG gene g in cancer type c, Spearman's rank correlation was computed between g and all CTGs identified in c using sample level expression profiles. A strong positive co-expression was defined as $\rho > 0.5$ and adjusted P-value < 0.01. Non-CTG genes were ranked by the number of CTGs with which they exhibited strong positive co-expression. The top-ranked non-CTG genes per cancer type were selected for enrichment analysis using the compareCluster function in clusterProfiler (v4.6.2) against the REACTOME pathway database (*75*). The top 2,000 non-CTG genes were used for pseudo-bulk samples derived from sc/snRNA-seq data, and the top 500 were used for bulk RNA-seq data from TCGA and TARGET cohorts.

Differential gene expression and gene set enrichment analysis

DESeq2 (v1.38.3) (*76*) was used to identify differentially expressed genes (DEGs) between *PBRM1* non-silent mutated and wild-type ccRCC tumors. DEGs were defined by $|\log_2(\text{fold change})| \geq 0.25$ and adjusted P-value ≤ 0.05. Gene set enrichment analysis was conducted using the GSEA function in the clusterProfiler package.

Calculation of n_CTGs, n_Genes and n_random_genes

For each cell *i* in a given sc/snRNA-seq dataset, n_CTGs (or n_Genes) was defined as the number of CTGs (or all genes) with detectable expression (counts > 0). To derive n_random_genes, all genes were first stratified into 24 bins based on their mean expression levels across all cells. For each CTG, a non-CTG gene was randomly selected from the same bin corresponding to the CTG, resulting in a background gene set that recapitulated the global expression distribution of the CTG set. This procedure was repeated 100 times to generate 100 independent gene sets, and the number of genes with non-zero expression was computed for each random gene set. Finally, the median value across the 100 iterations was used as n_random_genes.

For ST data, these metrics were computed for each spot. For snATAC-seq data, gene activity was first inferred by aggregating fragment counts within gene bodies and promoter regions using the GeneActivity function in the Signac package (v1.10.0) (*77*). n_CTGs was then defined as the number of CTGs with non-zero activity per cell.

Benchmarking of malignant cell identification methods

CNV scores were derived as described in the **Inference of copy number variations in sc/snRNA-seq data** section. The ikarus score, representing predicted malignancy probability, was obtained using the default ikarus pipeline. For scCancer2, the author-provided XGBoost classifier was applied to a scaled expression matrix composed of the top 5,000 highly variable genes, and the predicted probability of malignancy for each cell was recorded as the scCancer2 score. All benchmarking analyses were conducted at the sample level to minimize inter-sample batch effects.

ST data processing

For each ST section, the raw count matrix was normalized and scaled using the SCTransform function. Dimensionality reduction was performed by principal component analysis via the RunPCA function. A shared nearest-neighbor graph was constructed based on the top 30 principal components using the FindNeighbors function, followed by clustering with the original Louvain algorithm at a resolution of 1.5. All analyses were conducted using Seurat.

ESTIMATE scores were computed for each spatial spot using the ESTIMATE R package (v1.0.13), and tumor purity was inferred according to the following formula:

$$\textit{Tumor purity} = \cos(0.6049872018 + 0.0001467884 \times \textit{ESTIMATE score}) \quad (3)$$

Inference of copy number variations in ST data

CNV inference for ST data followed the method of Xun et al. (*78*). To define appropriate reference (i.e., normal) spots, an immune score was calculated per spot

using AddModuleScore based on canonical immune cell markers (*PTPRC*, *CD2*, *CD3D*, *CD3E*, *CD3G*, *CD5*, *CD7*, *CD79A*, *MS4A1*, *CD19*). For each section, the cluster with the highest average immune score, determined from prior clustering results, was selected as the candidate reference cluster. To ensure immune cell enrichment in the selected reference, a Wilcoxon rank-sum test was performed to compare immune scores between the candidate cluster and all other spots. Only clusters with significantly higher immune scores (ROC-AUC > 0.6 and adjusted *P*-value < 0.01) were accepted as references for CNV analysis.

The infercnv package was then applied at the sample level using the raw count matrix as input, with the following parameters: cutoff = 0.1, analysis_mode = "subclusters", cluster_by_groups = FALSE, denoise = TRUE, HMM = TRUE, HMM_type = "i6", BayesMaxPNormal = 0, and tumor_subcluster_partition_method = "random_trees". The random tree partitioning approach divides observation spots into eight CNV-defined subclusters, enhancing discrimination between malignant and non-malignant regions. Given a spot s and a gene $i \in \{1, 2, ..., N\}$, where x_i denotes the inferred CNV state of gene i, the CNV score of spot s was computed as:

$$CNV\ score(s) = \sum_{i=1}^{N} |x_i - 3| \quad (4)$$

All CNV clusters, including the normal reference, were ranked by average CNV scores. The two clusters with the highest average scores were designated as CNV-high, whereas the two with the lowest were defined as CNV-low.

Identification of TCR-T therapy targets

To discover novel TCR-T therapy targets beyond established candidates, CTGs were filtered according to two criteria: (1) significant overexpression in tumors relative to normal tissues (ROC-AUC > 0.8; adjusted *P*-value < 0.01). For each CTG in a given cancer type, statistical significance was assessed using two-sided Wilcoxon rank-sum tests with false discovery rate correction, and effect size was quantified by ROC-AUC, where values closer to 1 indicate greater upregulation of the CTG in tumor samples; and (2) minimal expression in non-testicular normal tissues (TPM < 1).

Existing clinical trials targeting each CTG were identified by querying the ClinicalTrials.gov database (accessed on September 27, 2025).

Survival analysis

For each gene, univariate Cox proportional hazards regression was used to evaluate the association between expression and patient survival within each cancer type using the coxph function in the survival R package (v3.5.7). Patients were stratified into “High” and “Low” groups according to the median aggregated CTG expression. Kaplan–Meier curves were generated using the survfit function, and survival differences were assessed using the log-rank test via the survdiff function.

Statistical analysis

If not specified, the two-sided Wilcoxon rank-sum test was used to evaluate the differences between the two groups, and Spearman's correlation test was used to assess the correlations between two variables. Multiple hypothesis testing was corrected using the Benjamini–Hochberg method. All analyses and visualizations were performed using R (v4.2.3).

References

1. Z. A. Gibbs, A. W. Whitehurst, Emerging Contributions of Cancer/Testis Antigens to Neoplastic Behaviors. *Trends Cancer* **4**, 701-712 (2018).
2. O. Gordeeva, Cancer-testis antigens: Unique cancer stem cell biomarkers and targets for cancer therapy. *Semin. Cancer Biol.* **53**, 75-87 (2018).
3. C. M. Fan, H. K. Qu, X. Wang, N. Sobhani, L. M. Wang, S. L. Liu, W. Xiong, Z. Y. Zeng, Y. Li, Cancer/testis antigens: from serology to mRNA cancer vaccine. *Semin. Cancer Biol.* **76**, 218-231 (2021).
4. P. Yang, M. Meng, Q. S. Zhou, Oncogenic cancer/testis antigens are a hallmarker of cancer and a sensible target for cancer immunotherapy. *Bba-Rev Cancer* **1876**, 188558 (2021).
5. Q. Deng, K. Y. Li, H. Chen, J. H. Dai, Y. Y. Zhai, Q. Wang, N. Li, Y. P. Wang, Z. G. Han, RNA Interference Against Cancer/Testis Genes Identifies Dual Specificity Phosphatase 21 as a Potential Therapeutic Target in Human Hepatocellular Carcinoma. *Hepatology* **59**, 518-530 (2014).
6. Z. D. Xie, H. B. Lin, Y. C. Wu, Y. A. Yu, X. T. Liu, Y. T. Zheng, X. J. Wang, J. S. Wu, M. F. Xu, Y. T. Han, Q. Y. Zhang, Y. Deng, L. Lin, Y. Linzhu, L. Qingyun, X. J. Lin, Y. Huang, P. Chi, USP4-mediated CENPF deubiquitylation regulated tumor metastasis in colorectal cancer. *Cell Death Dis.* **16**, 81 (2025).
7. M. AlHossiny, L. Luo, W. R. Frazier, N. Steiner, Y. Gusev, B. Kallakury, E. Glasgow, K. Creswell, S. Madhavan, R. Kumar, G. Upadhyay, Ly6E/K Signaling to TGFβ Promotes Breast Cancer Progression, Immune Escape, and Drug Resistance. *Cancer Res.* **76**, 3376-3386 (2016).
8. S. M. Zhang, W. L. Cai, X. N. Liu, D. Thakral, J. S. Luo, L. H. Chan, M. K. McGeary, E. Song, K. R. M. Blenman, G. Micevic, S. Jessel, Y. Y. Zhang, M. Z. Yin, C. J. Booth, L. B. Jilaveanu, W. Damsky, M. Sznol, H. M. Kluger, A. Iwasaki, M. W. Bosenberg, Q. Yan, KDM5B promotes immune evasion by recruiting SETDB1 to silence retroelements. *Nature* **598**, 682-+ (2021).
9. S. Deng, Y. Zhang, H. Wang, W. Liang, L. Xie, N. Li, Y. Fang, Y. Wang, J. Liu, H. Chi, Y. Sun, R. Ye, L. Shan, J. Shi, Z. Shen, Y. Wang, S. Wang, J.-P. Brosseau, F. Wang, G. Liu, Y. Quan, J. Xu, ITPRIPL1 binds CD3ε to impede T cell activation and enable tumor immune evasion. *Cell* **187**, 2305-2323.e2333 (2024).
10. L. G. Almeida, N. J. Sakabe, A. R. Deoliveira, M. C. C. Silva, A. S. Mundstein, T. Cohen, Y. T. Chen, R. Chua, S. Gurung, S. Gnjatic, A. A. Jungbluth, O. L. Caballero, A. Bairoch, E. Kiesler, S. L. White, A. J. G. Simpson, L. J. Old, A. A. Camargo, A. T. R. Vasconcelos, CTdatabase: a

knowledge-base of high-throughput and curated data on cancer-testis antigens. *Nucleic Acids Res.* **37**, D816-D819 (2009).

11. C. Wang, Y. Gu, K. Zhang, K. Xie, M. Zhu, N. Dai, Y. Jiang, X. Guo, M. Liu, J. Dai, L. Wu, G. Jin, H. Ma, T. Jiang, R. Yin, Y. Xia, L. Liu, S. Wang, B. Shen, R. Huo, Q. Wang, L. Xu, L. Yang, X. Huang, H. Shen, J. Sha, Z. Hu, Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types. *Nat Commun* **7**, 10499 (2016).

12. J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. M. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J. M. Stuart, C. G. A. R. Network, The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113-1120 (2013).

13. J. A. Carter, B. Matta, J. Battaglia, C. Somerville, B. D. Harris, M. LaPan, G. S. Atwal, B. J. Barnes, Identification of pan-cancer/testis genes and validation of therapeutic targeting in triple-negative breast cancer: Lin28a-based and Siglece-based vaccination induces antitumor immunity and inhibits metastasis. *J Immunother Cancer* **11**, e007935 (2023).

14. J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, B. Foster, M. Moser, E. Karasik, B. Gillard, K. Ramsey, S. Sullivan, J. Bridge, H. Magazine, J. Syron, J. Fleming, L. Siminoff, H. Traino, M. Mosavel, L. Barker, S. Jewell, D. Rohrer, D. Maxim, D. Filkins, P. Harbach, E. Cortadillo, B. Berghuis, L. Turner, E. Hudson, K. Feenstra, L. Sobin, J. Robb, P. Branton, G. Korzeniewski, C. Shive, D. Tabor, L. Q. Qi, K. Groch, S. Nampally, S. Buia, A. Zimmerman, A. Smith, R. Burges, K. Robinson, K. Valentino, D. Bradbury, M. Cosentino, N. Diaz-Mayoral, M. Kennedy, T. Engel, P. Williams, K. Erickson, K. Ardlie, W. Winckler, G. Getz, D. DeLuca, D. MacArthur, M. Kellis, A. Thomson, T. Young, E. Gelfand, M. Donovan, Y. Meng, G. Grant, D. Mash, Y. Marcus, M. Basile, J. Liu, J. Zhu, Z. D. Tu, N. J. Cox, D. L. Nicolae, E. R. Gamazon, H. K. Im, A. Konkashbaev, J. Pritchard, M. Stevens, T. Flutre, X. Q. Wen, E. T. Dermitzakis, T. Lappalainen, R. Guigo, J. Monlong, M. Sammeth, D. Koller, A. Battle, S. Mostafavi, M. McCarthy, M. Rivas, J. Maller, I. Rusyn, A. Nobel, F. Wright, A. Shabalin, M. Feolo, N. Sharopova, A. Sturcke, J. Paschal, J. M. Anderson, E. L. Wilder, L. K. Derr, E. D. Green, J. P. Struewing, G. Temple, S. Volpi, J. T. Boyer, E. J. Thomson, M. S. Guyer, C. Ng, A. Abdallah, D. Colantuoni, T. R. Insel, S. E. Koester, A. R. Little, P. K. Bender, T. Lehner, Y. Yao, C. C. Compton, J. B. Vaught, S. Sawyer, N. C. Lockhart, J. Demchok, H. F. Moore, The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580-585 (2013).

15. R. L. Grossman, A. P. Heath, V. Ferretti, H. E. Varmus, D. R. Lowy, W. A. Kibbe, L. M. Staudt, Toward a Shared Vision for Cancer Genomic Data. *N. Engl. J. Med.* **375**, 1109-1112 (2016).

16. A. Gavish, M. Mints, A. C. Greenwald, R. Hoefflin, D. Simkin, R. Tschernichovsky, N. G. Darnell, E. Somech, C. Barbolin, T. Antman, D. Kovarsky, T. Barrett, L. N. G. Castro, D. Halder, R. Chanoch-Myers, J. Laffy, M. Mints, A. Wider, R. Tal, A. Spitzer, T. Hara, M. Raitses-Gurevich, C. Stossel, T. Golan, A. Tirosh, M. L. Suvà, S. V. Puram, I. Tirosh, Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598-+ (2023).

17. Y. Han, Y. Wang, X. Dong, D. Sun, Z. Liu, J. Yue, H. Wang, T. Li, C. Wang, TISCH2: expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. *Nucleic Acids Res.* **51**, D1425-D1431 (2023).

18. O. Rozenblatt-Rosen, A. Regev, P. Oberdoerffer, T. Nawy, A. Hupalowska, J. E. Rood, O. Ashenberg, E. Cerami, R. J. Coffey, E. Demir, L. Ding, E. D. Esplin, J. M. Ford, J. Goecks, S.

871 Ghosh, J. W. Gray, J. Guinney, S. E. Hanlon, S. K. Hughes, E. S. Hwang, C. A. Iacobuzio-
872 Donahue, J. Jané-Valbuena, B. E. Johnson, K. S. Lau, T. Lively, S. A. Mazzilli, D. Pe'er, S.
873 Santagata, A. K. Shalek, D. Schapiro, M. P. Snyder, P. K. Sorger, A. E. Spira, S. Srivastava, K.
874 Tan, R. B. West, E. H. Williams, H. T. A. Network, The Human Tumor Atlas Network:
875 Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell* **181**,
876 236-249 (2020).
877 19. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å.
878 Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani,
879 C. A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist,
880 H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von
881 Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F.
882 Pontén, Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
883 20. P. Kulkarni, T. Shiraishi, K. Rajagopalan, R. Kim, S. M. Mooney, R. H. Getzenberg,
884 Cancer/testis antigens and urological malignancies. *Nature Reviews Urology* **9**, 386-396
885 (2012).
886 21. L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection
887 for dimension reduction. *The Journal of Open Source Software* **3**, 861 (2018).
888 22. J. D. Campbell, C. Yau, R. Bowlby, Y. X. Liu, K. Brennan, H. H. Fan, A. M. Taylor, C. Wang,
889 V. Walter, R. Akbani, L. A. Byers, C. J. Creighton, C. Coarfa, J. Shih, A. D. Cherniack, O.
890 Gevaert, M. Prunello, H. Shen, P. Anur, J. H. Chen, H. Cheng, D. N. Hayes, S. Bullman, C. S.
891 Pedamallu, A. I. Ojesina, S. Sadeghi, K. L. Mungall, A. G. Robertson, C. Benz, A. Schultz, R.
892 S. Kanchi, C. M. Gay, A. Hegde, L. X. Diao, J. Wang, W. C. Ma, P. Sumazin, H. S. Chiu, T. W.
893 Chen, P. Gunaratne, L. Donehower, J. S. Rader, R. Zuna, H. Al-Ahmadie, A. J. Lazar, E. R.
894 Flores, K. Y. Tsai, J. H. Zhou, A. K. Rustgi, E. Drill, R. L. Shen, C. K. Wong, J. M. Stuart, P. W.
895 Laird, K. A. Hoadley, J. N. Weinstein, M. Peto, C. R. Pickering, Z. Chen, C. Waes, C. G. A. R.
896 Network, Genomic, Pathway Network, and Immunologic Features Distinguishing
897 Squamous Carcinomas. *Cell Rep.* **23**, 194-+ (2018).
898 23. T. J. Pugh, O. Morozova, E. F. Attiyeh, S. Asgharzadeh, J. S. Wei, D. Auclair, S. L. Carter, K.
899 Cibulskis, M. Hanna, A. Kiezun, J. Kim, M. S. Lawrence, L. Lichenstein, A. McKenna, C. S.
900 Pedamallu, A. H. Ramos, E. Shefler, A. Sivachenko, C. Sougnez, C. Stewart, A. Ally, I. Birol,
901 R. Chiu, R. D. Corbett, M. Hirst, S. D. Jackman, B. Kamoh, A. H. Khodabakshi, M. Krzywinski,
902 A. Lo, R. A. Moore, K. L. Mungall, J. Qian, A. Tam, N. Thiessen, Y. J. Zhao, K. A. Cole, M.
903 Diamond, S. J. Diskin, Y. P. Mosse, A. C. Wood, L. Y. Ji, R. Sposto, T. Badgett, W. B. London,
904 Y. Moyer, J. M. Gastier-Foster, M. A. Smith, J. M. G. Auvil, D. S. Gerhard, M. D. Hogarty, S.
905 J. M. Jones, E. S. Lander, S. B. Gabriel, G. Getz, R. C. Seeger, J. Khan, M. A. Marra, M.
906 Meyerson, J. M. Maris, The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**,
907 279-284 (2013).
908 24. A. V. Pel, P. Van Der Bruggen, P. G. Coulie, V. G. Brichard, B. Lethéa, B. Van Den Eynde, C.
909 Uyttenhove, J. C. Renauld, T. Boon, Genes Coding for Tumor Antigens Recognized by
910 Cytolytic T Lymphocytes. *Immunol. Rev.* **145**, 229-250 (2006).
911 25. I. Yanai, H. Benjamin, M. Shmoish, V. Chalifa-Caspi, M. Shklar, R. Ophir, A. Bar-Even, S.
912 Horn-Saban, M. Safran, E. Domany, D. Lancet, O. Shmueli, Genome-wide midrange
913 transcription profiles reveal expression level relationships in human tissue specification.
914 *Bioinformatics* **21**, 650-659 (2005).

26. J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, L. A. Garraway, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012).

27. C. Su, Z. C. Xu, X. N. Shan, B. Cai, H. Y. Zhao, J. F. Zhang, Cell-type-specific co-expression inference from single cell RNA-sequencing data. *Nat Commun* **14**, 4846 (2023).

28. C. Alabert, A. Groth, Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Bio* **13**, 153-167 (2012).

29. C. Van Rechem, F. Ji, D. Chakraborty, J. C. Black, R. I. Sadreyev, J. R. Whetstine, Collective regulation of chromatin modifications predicts replication timing during cell cycle. *Cell Rep.* **37**, 109799 (2021).

30. D. Marakulina, I. E. Vorontsov, I. V. Kulakovskiy, A. Lennartsson, F. Drabløs, Yulia A. Medvedeva, EpiFactors 2022: expansion and enhancement of a curated database of human epigenetic factors and complexes. *Nucleic Acids Res.* **51**, D564-D570 (2023).

31. W. Y. Mi, H. P. Guan, J. Lyu, D. Zhao, Y. X. Xi, S. M. Jiang, F. H. Andrews, X. L. Wang, M. Gagea, H. Wen, L. Tora, S. Y. R. Dent, T. G. Kutateladze, W. Li, H. T. Li, X. B. Shi, YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. *Nat Commun* **8**, 1088 (2017).

32. R. C. Wu, I. C. Young, Y. F. Chen, S. T. Chuang, A. Toubaji, M. Y. Wu, Identification of the PTEN-ARID4B-PI3K pathway reveals the dependency on ARID4B by PTEN-deficient prostate cancer. *Nat Commun* **10**, 4332 (2019).

33. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550 (2005).

34. C. J. Ricketts, A. A. De Cubas, H. Fan, C. C. Smith, M. Lang, E. Reznik, R. Bowlby, E. A. Gibb, R. Akbani, R. Beroukhim, D. P. Bottaro, T. K. Choueiri, R. A. Gibbs, A. K. Godwin, S. Haake, A. A. Hakimi, E. P. Henske, J. J. Hsieh, T. H. Ho, R. S. Kanchi, B. Krishnan, D. J. Kwiatkowski, W. Liu, M. J. Merino, G. B. Mills, J. Myers, M. L. Nickerson, V. E. Reuter, L. S. Schmidt, C. S. Shelley, H. Shen, B. Shuch, S. Signoretti, R. Srinivasan, P. Tamboli, G. Thomas, B. G. Vincent, C. D. Vocke, D. A. Wheeler, L. Yang, W. Y. Kim, A. G. Robertson, P. T. Spellman, W. K. Rathmell, W. M. Linehan, S. J. Caesar-Johnson, J. A. Demchok, I. Felau, M. Kasapi, M. L. Ferguson, C. M. Hutter, H. J. Sofia, R. Tarnuzzer, Z. Wang, L. Yang, J. C. Zenklusen, J. Zhang, S. Chudamani, J. Liu, L. Lolla, R. Naresh, T. Pihl, Q. Sun, Y. Wan, Y. Wu, J. Cho, T. DeFreitas, S. Frazer, N. Gehlenborg, G. Getz, D. I. Heiman, J. Kim, M. S. Lawrence, P. Lin, S. Meier, M. S. Noble, G. Saksena, D. Voet, H. Zhang, B. Bernard, N. Chambwe, V. Dhankani, T. Knijnenburg, R. Kramer, K. Leinonen, Y. Liu, M. Miller, S. Reynolds, I. Shmulevich, V. Thorsson, W. Zhang, R. Akbani, B. M. Broom, A. M. Hegde, Z. Ju, R. S. Kanchi, A. Korkut, J.

Li, H. Liang, S. Ling, W. Liu, Y. Lu, G. B. Mills, K.-S. Ng, A. Rao, M. Ryan, J. Wang, J. N. Weinstein, J. Zhang, A. Abeshouse, J. Armenia, D. Chakravarty, W. K. Chatila, I. de Bruijn, J. Gao, B. E. Gross, Z. J. Heins, R. Kundra, K. La, M. Ladanyi, A. Luna, M. G. Nissan, A. Ochoa, S. M. Phillips, E. Reznik, F. Sanchez-Vega, C. Sander, N. Schultz, R. Sheridan, S. O. Sumer, Y. Sun, B. S. Taylor, J. Wang, H. Zhang, P. Anur, M. Peto, P. Spellman, C. Benz, J. M. Stuart, C. K. Wong, C. Yau, D. N. Hayes, J. S. Parker, M. D. Wilkerson, A. Ally, M. Balasundaram, R. Bowlby, D. Brooks, R. Carlsen, E. Chuah, N. Dhalla, R. Holt, S. J. M. Jones, K. Kasaian, D. Lee, Y. Ma, M. A. Marra, M. Mayo, R. A. Moore, A. J. Mungall, K. Mungall, A. G. Robertson, S. Sadeghi, J. E. Schein, P. Sipahimalani, A. Tam, N. Thiessen, K. Tse, T. Wong, A. C. Berger, R. Beroukhim, A. D. Cherniack, C. Cibulskis, S. B. Gabriel, G. F. Gao, G. Ha, M. Meyerson, S. E. Schumacher, J. Shih, M. H. Kucherlapati, R. S. Kucherlapati, S. Baylin, L. Cope, L. Danilova, M. S. Bootwalla, P. H. Lai, D. T. Maglinte, D. J. Van Den Berg, D. J. Weisenberger, J. T. Auman, S. Balu, T. Bodenheimer, C. Fan, K. A. Hoadley, A. P. Hoyle, S. R. Jefferys, C. D. Jones, S. Meng, P. A. Mieczkowski, L. E. Mose, A. H. Perou, C. M. Perou, J. Roach, Y. Shi, J. V. Simons, T. Skelly, M. G. Soloway, D. Tan, U. Veluvolu, H. Fan, T. Hinoue, P. W. Laird, H. Shen, W. Zhou, M. Bellair, K. Chang, K. Covington, C. J. Creighton, H. Dinh, H. Doddapaneni, L. A. Donehower, J. Drummond, R. A. Gibbs, R. Glenn, W. Hale, Y. Han, J. Hu, V. Korchina, S. Lee, L. Lewis, W. Li, X. Liu, M. Morgan, D. Morton, D. Muzny, J. Santibanez, M. Sheth, E. Shinbrot, L. Wang, M. Wang, D. A. Wheeler, L. Xi, F. Zhao, J. Hess, E. L. Appelbaum, M. Bailey, M. G. Cordes, L. Ding, C. C. Fronick, L. A. Fulton, R. S. Fulton, C. Kandoth, E. R. Mardis, M. D. McLellan, C. A. Miller, H. K. Schmidt, R. K. Wilson, D. Crain, E. Curley, J. Gardner, K. Lau, D. Mallery, S. Morris, J. Paulauskis, R. Penny, C. Shelton, T. Shelton, M. Sherman, E. Thompson, P. Yena, J. Bowen, J. M. Gastier-Foster, M. Gerken, K. M. Leraas, T. M. Lichtenberg, N. C. Ramirez, L. Wise, E. Zmuda, N. Corcoran, T. Costello, C. Hovens, A. L. Carvalho, A. C. de Carvalho, J. H. Fregnani, A. Longatto-Filho, R. M. Reis, C. Scapulatempo-Neto, H. C. S. Silveira, D. O. Vidal, A. Burnette, J. Eschbacher, B. Hermes, A. Noss, R. Singh, M. L. Anderson, P. D. Castro, M. Ittmann, D. Huntsman, B. Kohl, X. Le, R. Thorp, C. Andry, E. R. Duffy, V. Lyadov, O. Paklina, G. Setdikova, A. Shabunin, M. Tavobilov, C. McPherson, R. Warnick, R. Berkowitz, D. Cramer, C. Feltmate, N. Horowitz, A. Kibel, M. Muto, C. P. Raut, A. Malykh, J. S. Barnholtz-Sloan, W. Barrett, K. Devine, J. Fulop, Q. T. Ostrom, K. Shimmel, Y. Wolinsky, A. E. Sloan, A. De Rose, F. Giuliante, M. Goodman, B. Y. Karlan, C. H. Hagedorn, J. Eckman, J. Harr, J. Myers, K. Tucker, L. A. Zach, B. Deyarmin, H. Hu, L. Kvecher, C. Larson, R. J. Mural, S. Somiari, A. Vicha, T. Zelinka, J. Bennett, M. Iacocca, B. Rabeno, P. Swanson, M. Latour, L. Lacombe, B. Têtu, A. Bergeron, M. McGraw, S. M. Staugaitis, J. Chabot, H. Hibshoosh, A. Sepulveda, T. Su, T. Wang, O. Potapova, O. Voronina, L. Desjardins, O. Mariani, S. Roman-Roman, X. Sastre, M.-H. Stern, F. Cheng, S. Signoretti, A. Berchuck, D. Bigner, E. Lipp, J. Marks, S. McCall, R. McLendon, A. Secord, A. Sharp, M. Behera, D. J. Brat, A. Chen, K. Delman, S. Force, F. Khuri, K. Magliocca, S. Maithel, J. J. Olson, T. Owonikoko, A. Pickens, S. Ramalingam, D. M. Shin, G. Sica, E. G. Van Meir, H. Zhang, W. Eijckenboom, A. Gillis, E. Korpershoek, L. Looijenga, W. Oosterhuis, H. Stoop, K. E. van Kessel, E. C. Zwarthoff, C. Calatozzolo, L. Cuppini, S. Cuzzubbo, F. DiMeco, G. Finocchiaro, L. Mattei, A. Perin, B. Pollo, C. Chen, J. Houck, P. Lohavanichbutr, A. Hartmann, C. Stoehr, R. Stoehr, H. Taubert, S. Wach, B. Wullich, W. Kycler, D. Murawa, M. Wiznerowicz, K. Chung, W. J. Edenfield, J. Martin, E. Baudin, G. Bubley, R. Bueno, A. De Rienzo, W. G.

Richards, S. Kalkanis, T. Mikkelsen, H. Noushmehr, L. Scarpace, N. Girard, M. Aymerich, E. Campo, E. Giné, A. L. Guillermo, N. Van Bang, P. T. Hanh, B. D. Phu, Y. Tang, H. Colman, K. Evason, P. R. Dottino, J. A. Martignetti, H. Gabra, H. Juhl, T. Akeredolu, S. Stepa, D. Hoon, K. Ahn, K. J. Kang, F. Beuschlein, A. Breggia, M. Birrer, D. Bell, M. Borad, A. H. Bryce, E. Castle, V. Chandan, J. Cheville, J. A. Copland, M. Farnell, T. Flotte, N. Giama, T. Ho, M. Kendrick, J.-P. Kocher, K. Kopp, C. Moser, D. Nagorney, D. O'Brien, B. P. O'Neill, T. Patel, G. Petersen, F. Que, M. Rivera, L. Roberts, R. Smallridge, T. Smyrk, M. Stanton, R. H. Thompson, M. Torbenson, J. D. Yang, L. Zhang, F. Brimo, J. A. Ajani, A. M. A. Gonzalez, C. Behrens, J. Bondaruk, R. Broaddus, B. Czerniak, B. Esmaeli, J. Fujimoto, J. Gershenwald, C. Guo, A. J. Lazar, C. Logothetis, F. Meric-Bernstam, C. Moran, L. Ramondetta, D. Rice, A. Sood, P. Tamboli, T. Thompson, P. Troncoso, A. Tsao, I. Wistuba, C. Carter, L. Haydu, P. Hersey, V. Jakrot, H. Kakavand, R. Kefford, K. Lee, G. Long, G. Mann, M. Quinn, R. Saw, R. Scolyer, K. Shannon, A. Spillane, o. Stretch, M. Synott, J. Thompson, J. Wilmott, H. Al-Ahmadie, T. A. Chan, R. Ghossein, A. Gopalan, D. A. Levine, V. Reuter, S. Singer, B. Singh, N. V. Tien, T. Broudy, C. Mirsaidi, P. Nair, P. Drwiega, J. Miller, J. Smith, H. Zaren, J.-W. Park, N. P. Hung, E. Kebebew, W. M. Linehan, A. R. Metwalli, K. Pacak, P. A. Pinto, M. Schiffman, L. S. Schmidt, C. D. Vocke, N. Wentzensen, R. Worrell, H. Yang, M. Moncrieff, C. Goparaju, J. Melamed, H. Pass, N. Botnariuc, I. Caraman, M. Cernat, I. Chemencedji, A. Clipca, S. Doruc, G. Gorincioi, S. Mura, M. Pirtac, I. Stancul, D. Tcaciuc, M. Albert, I. Alexopoulou, A. Arnaout, J. Bartlett, J. Engel, S. Gilbert, J. Parfitt, H. Sekhon, G. Thomas, D. M. Rassl, R. C. Rintoul, C. Bifulco, R. Tamakawa, W. Urba, N. Hayward, H. Timmers, A. Antenucci, F. Facciolo, G. Grazi, M. Marino, R. Merola, R. de Krijger, A.-P. Gimenez-Roqueplo, A. Piché, S. Chevalier, G. McKercher, K. Birsoy, G. Barnett, C. Brewer, C. Farver, T. Naska, N. A. Pennell, D. Raymond, C. Schilero, K. Smolenski, F. Williams, C. Morrison, J. A. Borgia, M. J. Liptay, M. Pool, C. W. Seder, K. Junker, L. Omberg, M. Dinkin, G. Manikhas, D. Alvaro, M. C. Bragazzi, V. Cardinale, G. Carpino, E. Gaudio, D. Chesla, S. Cottingham, M. Dubina, F. Moiseenko, R. Dhanasekaran, K.-F. Becker, K.-P. Janssen, J. Slotta-Huspenina, M. H. Abdel-Rahman, D. Aziz, S. Bell, C. M. Cebulla, A. Davis, R. Duell, J. B. Elder, J. Hilty, B. Kumar, J. Lang, N. L. Lehman, R. Mandt, P. Nguyen, R. Pilarski, K. Rai, L. Schoenfield, K. Senecal, P. Wakely, P. Hansen, R. Lechan, J. Powers, A. Tischler, W. E. Grizzle, K. C. Sexton, A. Kastl, J. Henderson, S. Porten, J. Waldmann, M. Fassnacht, S. L. Asa, D. Schadendorf, M. Couce, M. Graefen, H. Huland, G. Sauter, T. Schlomm, R. Simon, P. Tennstedt, O. Olabode, M. Nelson, O. Bathe, P. R. Carroll, J. M. Chan, P. Disaia, P. Glenn, R. K. Kelley, C. N. Landen, J. Phillips, M. Prados, J. Simko, K. Smith-McCune, S. VandenBerg, K. Roggin, A. Fehrenbach, A. Kendler, S. Sifri, R. Steele, A. Jimeno, F. Carey, I. Forgie, M. Mannelli, M. Carney, B. Hernandez, B. Campos, C. Herold-Mende, C. Jungk, A. Unterberg, A. von Deimling, A. Bossler, J. Galbraith, L. Jacobus, M. Knudson, T. Knutson, D. Ma, M. Milhem, R. Sigmund, A. K. Godwin, R. Madan, H. G. Rosenthal, C. Adebamowo, S. N. Adebamowo, A. Boussioutas, D. Beer, T. Giordano, A.-M. Mes-Masson, F. Saad, T. Bocklage, L. Landrum, R. Mannel, K. Moore, K. Moxley, R. Postier, J. Walker, R. Zuna, M. Feldman, F. Valdivieso, R. Dhir, J. Luketich, E. M. M. Pinero, M. Quintero-Aguilo, C. G. Carlotti, J. S. Dos Santos, R. Kemp, A. Sankarankuty, D. Tirapelli, J. Catto, K. Agnew, E. Swisher, J. Creaney, B. Robinson, C. S. Shelley, E. M. Godwin, S. Kendall, C. Shipman, C. Bradford, T. Carey, A. Haddad, J. Moyer, L. Peterson, M. Prince, L. Rozek, G. Wolf, R. Bowman, K. M. Fong, I. Yang, R. Korst,

W. K. Rathmell, J. L. Fantacone-Campbell, J. A. Hooke, A. J. Kovatich, C. D. Shriver, J. DiPersio, B. Drake, R. Govindan, S. Heath, T. Ley, B. Van Tine, P. Westervelt, M. A. Rubin, J. I. Lee, N. D. Aredes, A. Mariamidze, The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep.* **23**, 313-326.e315 (2018).

35. N. V. Terekhanova, A. Karpova, W. W. Liang, A. Strzalkowski, S. Q. Chen, Y. Z. Li, A. N. Southard-Smith, M. D. Iglesia, M. C. Wendl, R. G. Jayasinghe, J. X. Liu, Y. Z. Song, S. Cao, A. Houston, X. T. Liu, M. A. Wyczalkowski, R. J. H. Lu, W. Caravan, A. Shinkle, N. N. Al Deen, J. M. Herndon, J. Mudd, C. Ma, H. Sarkar, K. Sato, O. M. Ibrahim, C. K. Mo, S. E. Chasnoff, E. Porta-Pardo, J. M. Held, R. Pachynski, J. K. Schwarz, W. E. Gillanders, A. H. Kim, R. Vij, J. F. Dipersio, S. V. Puram, M. G. Chheda, K. C. Fuh, D. G. Denardo, R. C. Fields, F. Chen, B. J. Raphael, L. Ding, Epigenetic regulation during cancer transitions across 11 tumour types. *Nature* **623**, 432-+ (2023).

36. D. Aran, A. P. Looney, L. Q. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R. P. Naikawadi, P. J. Wolters, A. R. Abate, A. J. Butte, M. Bhattacharya, Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163-+ (2019).

37. inferCNV of the Trinity CTAT Project. https://github.com/broadinstitute/inferCNV.

38. J. Dohmen, A. Baranovskii, J. Ronen, B. Uyar, V. Franke, A. Akalin, Identifying tumor cells at the single-cell level using machine learning. *Genome Biol.* **23**, 123 (2022).

39. Z. Y. Chen, Y. X. Miao, Z. Y. Tan, Q. F. Hu, Y. H. Wu, X. Q. Li, W. B. Guo, J. Gu, scCancer2: data-driven in-depth annotations of the tumor microenvironment at single-level resolution. *Bioinformatics* **40**, btae028 (2024).

40. G. S. Gulati, J. P. D'Silva, Y. H. Liu, L. H. Wang, A. M. Newman, Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. *Nat Rev Mol Cell Bio* **26**, 11-31 (2025).

41. R. Arora, C. Cao, M. Kumar, S. Sinha, A. Chanda, R. McNeil, D. Samuel, R. K. Arora, T. W. Matthews, S. Chandarana, R. Hart, J. C. Dort, J. Biernaskie, P. Neri, M. D. Hyrcza, P. Bose, Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat Commun* **14**, 5029 (2023).

42. Y. Liu, Z. Z. Xun, K. Ma, S. H. Liang, X. Y. Li, S. Zhou, L. M. Sun, Y. F. Liu, Y. H. Du, X. Y. Guo, T. M. Cui, H. R. Zhou, J. Z. Wang, D. L. Yin, R. P. Song, S. G. Zhang, W. Cai, F. Z. Meng, H. R. Guo, B. Zhang, D. Yang, R. J. Bao, Q. S. Hu, J. B. Wang, Y. Q. Ye, L. X. Liu, Identification of a tumour immune barrier in the HCC microenvironment that determines the efficacy of immunotherapy. *J. Hepatol.* **78**, 770-782 (2023).

43. K. Yoshihara, M. Shahmoradgoli, E. Martínez, R. Vegesna, H. Kim, W. Torres-Garcia, V. Treviño, H. Shen, P. W. Laird, D. A. Levine, S. L. Carter, G. Getz, K. Stemke-Hale, G. B. Mills, R. G. W. Verhaak, Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* **4**, 2612 (2013).

44. M. Andreatta, J. Garnica, S. J. Carmona, Identification of malignant cells in single-cell transcriptomics data. *Commun Biol* **8**, 1264 (2025).

45. S. J. Keam, Afamitresgene Autoleucel: First Approval. *Mol. Diagn. Ther.* **28**, 861-866 (2024).

46. D. Kortleve, D. Hammerl, M. van Brakel, R. Wijers, D. Roelofs, K. Kroese, M. M. Timmermans, C. Y. Liao, S. Z. Huang, A. Trapman-Jansen, R. Foekens, J. Michaux, M. T. A. de Beijer, S. I. Buschow, J. A. A. Demmers, M. Kok, E. H. J. Danen, M. Bassani-Sternberg, J. W. M. Martens,

R. J. M. Abbott, R. Debets, TCR-Engineered T Cells Directed against Ropporin-1 Constitute a Safe and Effective Treatment for Triple-Negative Breast Cancer. *Cancer Discov.* **14**, 2450-2470 (2024).

47. M. Wermke, W. Alsdorf, D. Araujo, M. Chatterjee, N. Hilf, T. A. W. Holderried, A. A. Jazaeri, M. Kalra, A. Mayer-Mokler, R. Mendrzyk, A. Mohamed, S. P. Patel, R. Reshef, A. Satelli, H. Singh, A.-M. Tsimberidou, S. Walter, T. Weinschenk, C. M. Britten, J. Luke, Abstract PR018: IMA203 TCR-T targeting PRAME demonstrates potent anti-tumor activity in patients with different types of metastatic solid tumors. *Mol. Cancer Ther.* **22**, PR018-PR018 (2023).

48. E. Baulu, C. Gardet, N. Chuvin, S. Depil, TCR-engineered T cell therapy in solid tumors: State of the art and perspectives. *Sci Adv* **9**, eadf3700 (2023).

49. L. Gong, J. Luo, E. Yang, B. Ru, Z. Qi, Y. Yang, A. Rani, A. Purohit, Y. Zhang, G. Guan, R. Paul, T. Vu, Z. Chen, R. Ji, C.-P. Day, C. Wu, G. Merlino, D. Fitzgerald, G. Altan-Bonnet, K. Aldape, J. Wu, X. Guan, P. Jiang, Cancer immunology data engine reveals secreted AOAH as a potential immunotherapy. *Cell* **188**, 5062-5080.e5032 (2025).

50. Y. M. Lu, A. Q. Yang, C. Quan, Y. W. Pan, H. Y. Zhang, Y. F. Li, C. M. Gao, H. Lu, X. T. Wang, P. B. Cao, H. X. Chen, S. C. Lu, G. Q. Zhou, A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. *Nat Commun* **13**, 4594 (2022).

51. Y. F. Sun, L. Wu, Y. Zhong, K. Q. Zhou, Y. Hou, Z. F. Wang, Z. F. Zhang, J. R. Xie, C. Q. Wang, D. D. Chen, Y. L. Huang, X. C. Wei, Y. H. Shi, Z. K. Zhao, Y. H. Li, Z. W. Guo, Q. C. Yu, L. Q. Xu, G. Volpe, S. J. Qiu, J. Zhou, C. Ward, H. C. Sun, Y. Yin, X. Xu, X. D. Wang, M. A. Esteban, H. M. Yang, J. Wang, M. Dean, Y. G. Zhang, S. P. Liu, X. R. Yang, J. Fan, Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. *Cell* **184**, 404-+ (2021).

52. A. Liberzon, C. Birger, H. Thorvaldsdottir, M. Ghandi, J. P. Mesirov, P. Tamayo, The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**, 417-425 (2015).

53. M. Kang, G. S. Gulati, E. L. Brown, Z. Qi, S. Avagyan, J. J. A. Armenteros, R. Gleyzer, W. Zhang, C. B. Steen, J. P. D'Silva, J. Schwab, M. F. Clarke, A. A. Chaudhuri, A. M. Newman, Improved reconstruction of single-cell developmental potential with CytoTRACE 2. *Nat Methods*, 10.1038/s41592-41025-02857-41592 (2025).

54. A. J. G. Simpson, O. L. Caballero, A. Jungbluth, Y. T. Chen, L. J. Old, Cancer/testis antigens, gametogenesis and cancer. *Nature Reviews Cancer* **5**, 615-625 (2005).

55. A. R. Karpf, S. X. Bai, S. R. James, J. L. Mohler, E. M. Wilson, Increased Expression of Androgen Receptor Coregulator MAGE-11 in Prostate Cancer by DNA Hypomethylation and Cyclic AMP. *Mol. Cancer Res.* **7**, 523-535 (2009).

56. A. Van Tongelen, A. Loriot, C. De Smet, Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes. *Cancer Lett.* **396**, 130-137 (2017).

57. J. Vansteenkiste, M. Zielinski, A. Linder, J. Dahabreh, E. E. Gonzalez, W. Malinowski, M. Lopez-Brea, T. Vanakesa, J. Jassem, H. Kalofonos, J. Perdeus, R. Bonnet, J. Basko, R. Janilionis, B. Passlick, T. Treasure, M. Gillet, F. F. Lehmann, V. G. Brichard, Adjuvant MAGE-A3 Immunotherapy in Resected Non-Small-Cell Lung Cancer: Phase II Randomized Study Results. *J. Clin. Oncol.* **31**, 2396-+ (2013).

58. J. F. Vansteenkiste, B. C. Cho, T. Vanakesa, T. De Pas, M. Zielinski, M. S. Kim, J. Jassem, M. Yoshimura, J. Dahabreh, H. Nakayama, L. Havel, H. Kondo, T. Mitsudomi, K. Zarogoulidis, O. A. Gladkov, K. Udud, H. Tada, H. Hoffman, A. Bugge, P. Taylor, E. E. Gonzalez, M. L. Liao, J. X. He, J. L. Pujol, J. Louahed, M. Debois, V. Brichard, C. Debruyne, P. Therasse, N. Altorki,

Efficacy of the MAGE-A3 cancer immunotherapeutic as adjuvant therapy in patients with resected MAGE-A3-positive non-small-cell lung cancer (MAGRIT): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* **17**, 822-835 (2016).

59. R. Thomas, G. Al-Khadairi, J. Roelands, W. Hendrickx, S. Dermime, D. Bedognetti, J. Decock, NY-ESO-1 Based Immunotherapy of Cancer: Current Perspectives. *Front. Immunol.* **9**, 947 (2018).

60. A. Alsalloum, J. A. Shevchenko, S. Sennikov, NY-ESO-1 antigen: A promising frontier in cancer immunotherapy. *Clin Transl Med* **14**, e70020 (2024).

61. C. E. Brown, J. C. Hibbard, D. Alizadeh, M. S. Blanchard, H. M. Natri, D. Wang, J. R. Ostberg, B. Aguilar, J. R. Wagner, J. A. Paul, R. Starr, R. A. Wong, W. Chen, N. Shulkin, M. Aftabizadeh, A. Filippov, A. Chaudhry, J. A. Ressler, J. Kilpatrick, P. Myers-McNamara, M. Chen, L. D. Wang, R. C. Rockne, J. Georges, J. Portnow, M. E. Barish, M. D'Apuzzo, N. E. Banovich, S. J. Forman, B. Badie, Locoregional delivery of IL-13Rα2-targeting CAR-T cells in recurrent high-grade glioma: a phase 1 trial. *Nat. Med.* **30**, 1001-1012 (2024).

62. D. D. Bhuva, C. W. Tan, A. Salim, C. Marceaux, M. A. Pickering, J. J. Chen, M. Kharbanda, X. Y. Jin, N. Liu, K. Feher, G. Putri, W. D. Tilley, T. E. Hickey, M. L. Asselin-Labat, B. Phipson, M. J. Davis, Library size confounds biology in spatial transcriptomics data. *Genome Biol.* **25**, 99 (2024).

63. D. Lambrechts, E. Wauters, B. Boeckx, S. Aibar, D. Nittner, O. Burton, A. Bassez, H. Decaluwé, A. Pircher, K. Van den Eynde, B. Weynand, E. Verbeken, P. De Leyn, A. Liston, J. Vansteenkiste, P. Carmeliet, S. Aerts, B. Thienpont, Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277-+ (2018).

64. K. E. Maxfield, P. J. Taus, K. Corcoran, J. Wooten, J. Macion, Y. Zhou, M. Borromeo, R. K. Kollipara, J. Yan, Y. Xie, X.-J. Xie, A. W. Whitehurst, Comprehensive functional characterization of cancer–testis antigens defines obligate participation in multiple hallmarks of cancer. *Nat Commun* **6**, 8840 (2015).

65. Y. Deng, P. Chen, J. Xiao, M. Li, J. Shen, S. Qin, T. Jia, C. Li, A. Chang, W. Zhang, H. Liu, R. Xue, N. Zhang, X. Wang, L. Huang, D. Chen, SCAR: Single-cell and Spatially-resolved Cancer Resources. *Nucleic Acids Res.* **52**, D1407-D1417 (2024).

66. G. Wang, S. Wu, Z. Xiong, H. Qu, X. Fang, Y. Bao, CROST: a comprehensive repository of spatial transcriptomics. *Nucleic Acids Res.* **52**, D882-D890 (2024).

67. Y. H. Hao, S. Hao, E. Andersen-Nissen, W. I. I. I. Mauck, S. W. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-+ (2021).

68. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289-+ (2019).

69. C. Hu, T. Li, Y. Xu, X. Zhang, F. Li, J. Bai, J. Chen, W. Jiang, K. Yang, Q. Ou, X. Li, P. Wang, Y. Zhang, CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res.* **51**, D870-D876 (2023).

70. T. Z. Wu, E. Q. Hu, S. B. Xu, M. J. Chen, P. F. Guo, Z. H. Dai, T. Z. Feng, L. Zhou, W. L. Tang, L. Zhan, X. C. Fu, S. S. Liu, X. C. Bo, G. C. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation-Amsterdam* **2**, 100141 (2021).
71. A. M. Newman, C. L. Liu, M. R. Green, A. J. Gentles, W. G. Feng, Y. Xu, C. D. Hoang, M. Diehn, A. A. Alizadeh, Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-+ (2015).
72. D. A. Barbie, P. Tamayo, J. S. Boehm, S. Y. Kim, S. E. Moody, I. F. Dunn, A. C. Schinzel, P. Sandy, E. Meylan, C. Scholl, S. Fröhling, E. M. Chan, M. L. Sos, K. Michel, C. Mermel, S. J. Silver, B. A. Weir, J. H. Reiling, Q. Sheng, P. B. Gupta, R. C. Wadlow, H. Le, S. Hoersch, B. S. Wittner, S. Ramaswamy, D. M. Livingston, D. M. Sabatini, M. Meyerson, R. K. Thomas, E. S. Lander, J. P. Mesirov, D. E. Root, D. G. Gilliland, T. Jacks, W. C. Hahn, Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108-112 (2009).
73. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 7 (2013).
74. S. Aibar, C. B. González-Blas, T. Moerman, A. H. T. Van, H. Imrichova, G. Hulselmans, F. Rambow, J. C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-+ (2017).
75. M. Milacic, D. Beavers, P. Conley, C. Q. Gong, M. Gillespie, J. Griss, R. Haw, B. Jassal, L. Matthews, B. May, R. Petryszak, E. Ragueneau, K. Rothfels, C. Sevilla, V. Shamovsky, R. Stephan, K. Tiwari, T. Varusai, J. Weiser, A. Wright, G. M. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res.* **52**, D672-D678 (2023).
76. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
77. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat Methods* **18**, 1333-+ (2021).
78. Z. Z. Xun, X. Y. Ding, Y. Zhang, B. Y. Zhang, S. J. Lai, D. W. Zou, J. K. Zheng, G. Q. Chen, B. Su, L. Han, Y. Q. Ye, Reconstruction of the tumor spatial microenvironment along the malignant-boundary-nonmalignant axis. *Nat Commun* **14**, 933 (2023).

Acknowledgments

Funding: The research is supported by National Natural Science Foundation of China (Fund 32370715), Science Fund Program for Distinguished Overseas Young Scholars of Shandong Province (2023HWYQ-015), Taishan Young Scholar of Shandong Province (tsqn202312020), and the Cheeloo Young Scholar Program of Shandong University. The content is solely the responsibility of the authors and does not necessarily represent the official views of sponsors.

Author contributions: C.F. and K.L. conceived and designed the study. C.F. performed the formal analyses. C.F. and K.L. drafted the manuscript. S.Y., L.L., X.L., and X.J.

provided critical feedback during manuscript preparation and revision. All authors reviewed and approved the final version of the manuscript.

Competing interests: The researchers declare no competing interests.

Code availability: The R scripts used to analyze the data and generate figures are available on GitHub (https://github.com/chunyangyyyyyyy/Pan.Cancer.CTGs).