SCISSOR™: a single-cell inferred site-specific omics resource for tumor microenvironment association study

Xiang Cui [1], **Fei Qin** [2], **Xuanxuan Yu**[2], **Feifei Xiao** [2] **and Guoshuai Cai**[1,*]

[1]Department of Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA and [2]Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC 29208, USA

ABSTRACT

Tumor tissues are heterogeneous with different cell types in tumor microenvironment, which play an important role in tumorigenesis and tumor progression. Several computational algorithms and tools have been developed to infer the cell composition from bulk transcriptome profiles. However, they ignore the tissue specificity and thus a new resource for tissue-specific cell transcriptomic reference is needed for inferring cell composition in tumor microenvironment and exploring their association with clinical outcomes and tumor omics. In this study, we developed SCISSOR™ (https://thecailab.com/scissor/), an online open resource to fulfill that demand by integrating five orthogonal omics data of >6031 large-scale bulk samples, patient clinical outcomes and 451 917 high-granularity tissue-specific single-cell transcriptomic profiles of 16 cancer types. SCISSOR™ provides five major analysis modules that enable flexible modeling with adjustable parameters and dynamic visualization approaches. SCISSOR™ is valuable as a new resource for promoting tumor heterogeneity and tumor–tumor microenvironment cell interaction research, by delineating cells in the tissue-specific tumor microenvironment and characterizing their associations with tumor omics and clinical outcomes.

GRAPHICAL ABSTRACT
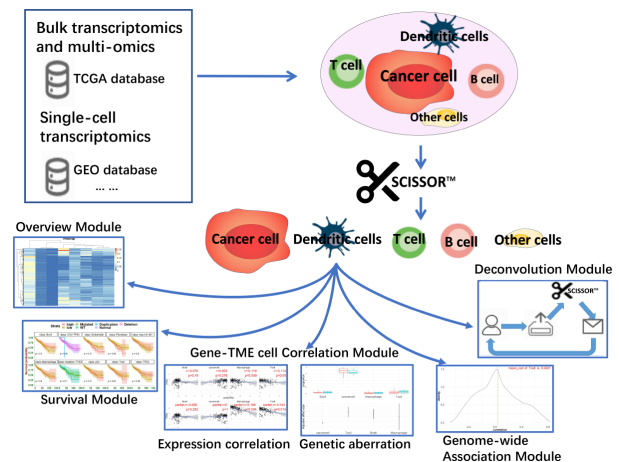
INTRODUCTION

With the advancement of transcriptome analysis, tumor research has made breakthrough progress. RNA sequencing (RNA-seq) has been widely applied to profile transcriptome in tumor biopsies to investigate transcriptomic dysregulation, detect new biomarkers and guide therapeutic treatment (1). However, this traditional bulk tissue analysis only captures the average gene expression in a tissue and masks the variation between cells. Besides tumor cells, various types of cells exist and function in tumors, including infiltrating inflammatory cells, tumor stromal cells, blood vessel cells and other associated tissue cells. Together with extracellular factors, these cells create a unique tumor microenvironment (TME) and play important roles in tumorigenesis and tumor progression (2). Evidence has suggested that these important roles of TME cells are tissue-specific (3). For example, resident myofibroblast-like stellate cells specifically contribute to carcinogenic mechanisms in pancreas and liver tumors (4). Moreover, tumor metastasis fits the 'seed and soil' theory that the metastasis of tu-

*To whom correspondence should be addressed. Tel: +1 803 777 4120; Email: GCAI@mailbox.sc.edu

mor cells is directed by the interaction between the cancer cells ('seed') and the host organ ('soil') (5). For accommodating metastatic tumor cells, the microenvironment of the distant organ can be induced and transformed into a tissue-specific niche (6). Therefore, it is critical to delineate the cells in the tissue-specific TME to understand the cancer mechanisms. Currently, single-cell RNA-seq (scRNA-seq) provides an unprecedented opportunity by profiling transcriptome in individual single cells. Using scRNA-seq, landscapes of immune cells and their phenotypic mapping have been studied in various tissues, including liver, breast and acute myeloid leukemia (7–9). However, the high economic and labor cost of scRNA-seq and its requirement of living cells obstruct the application in clinical settings (10).

Theoretically, the observed gene expression in a bulk tissue is a linear proportional summation of that in its cell subpopulations. Therefore, the proportions of tumor and TME cells can be estimated from bulk tumor transcriptomic data by deconvolution referencing to cell-type specific expression profiles. This approach overcomes the above limitations in bulk RNA-seq or scRNA-seq and enables cost- and effort-efficient sorting of TME cells. Several deconvolution methods have been proposed, including the non-negative least-squares (NNLS) method (11) that was originally applied to deconvolve the blood microarray data (12). Based on NNLS, post-modified methods such as quadprog (13) and limSolve (14) were developed and used in different tissues, including brain and heart (15–17). Alternatively, CIBER-SORT used support vector regression to infer 22 immune cell abundances from microarray and RNA-seq transcriptomic data, which has been widely used in tumor research (18). Later, MuSiC enabled the deconvolution with multi-subject single-cell expression reference using weighted non-negative least-squares regression (19). With cell compositions inferred from bulk tissue measurement, valuable online resources, including TCIA (20), PRECOG (21) and TIMER (22), were further developed to characterize infiltrating immune cells in tumor tissues. However, the deconvolution and cell proportion inference in these resources are based on one set of cell transcriptome profile references such as the 22 immune cell types in CIBERSORT and 6 immune cell types in TIMER. This strategy could introduce bias and lead to erroneous conclusions (23), because the transcriptome profiles of immune cells vary in different tissues (24). Also, their analyses are limited to tumor-infiltrating immune cells, without the capacity to study the full spectrum of TME cells, including immune cells, epithelial cells, fibroblasts, blood vessels and others.

In this study, we developed a new publicly available web server resource, Single Cell Inferred Site Specific Omics Resource for Tumor Microenvironments (SCISSOR™), for inferring tissue-specific TME cell composition from bulk transcriptomic data and investigating interactions between TME cell composition and tumor omics and clinical outcomes. By integrating The Cancer Genome Atlas (TCGA) large-scale bulk multimodal data and high-granularity single-cell transcriptomic data of a broad range of cancer types, SCISSOR™ provides a new opportunity to characterize tissue-specific TME cells and brings new insight into mechanisms of tumor development systematically and comprehensively (25).

MATERIALS AND METHODS

Single-cell transcriptomic data

We searched for scRNA-seq data for 33 TCGA cancer types in PubMed and GEO (Gene Expression Omnibus) databases and found 46 previously published studies. We filtered out 22 datasets with data of <1000 cells [except for the colon cancer dataset GSE81861 (26), which is of interest for our colon cancer research] or unavailable online. After removing studies without reported cell types, we finally obtained publicly available large-scale scRNA-seq datasets for 16 cancer types, including GSE81861 (26) (colon cancer), GSE103322 (27) (head and neck squamous cell carcinoma), GSE114725 (28) (breast cancer), GSE84465 (29) (glioblastoma), GSE125449 (30) (cholangiocarcinoma), GSE116256 (31) (acute myeloid leukemia), GSE72056 (32) (melanoma), GSE131907 (33) (lung adenocarcinoma), GSE154763 (34) (with data for seven cancer types: pancreatic adenocarcinoma, uterine corpus endometrial carcinoma, esophageal carcinoma, ovarian serous cystadenocarcinoma, kidney renal papillary cell carcinoma, lymphoid neoplasm diffuse large B-cell lymphoma and thyroid carcinoma), GSE140228 (35) (liver hepatocellular carcinoma), GSE123139 (36) (two datasets for melanoma), CRA001160 (37) (pancreatic ductal adenocarcinoma), E-MTAB-6149 and E-MTAB-6653 (38) (lung adenocarcinoma), datasets from Young *et al.* (39) (kidney cancer) and heiDATA (40) (nodal B-cell lymphomas).

Bulk transcriptomic data

RNA-seq raw counts of TCGA bulk tumor samples were collected from Broad GDAC Firehose (https://gdac.broadinstitute.org/). We downloaded and preprocessed the data of cancers that have scRNA-seq data available, which cover 6031 samples across 16 different cancer types. The raw counts were used for deconvolution to estimate the proportions of cell types in bulk samples. Counts per million on the $\log_2$ scale were used in association analysis.

Deconvolution

In SCISSOR™, MuSiC and CIBERSORTx were utilized to infer cell proportions of tumor and TME cells in TCGA tumor samples from RNA-seq data by referencing to tissue-specific single-cell transcriptome profiles obtained from the aforementioned scRNA-seq studies. A detailed introduction to these two methods is provided in the Supplementary Data. As different methods may provide different results, we recommended users report both results and use them for cross-validation. To accommodate the limitation of cell numbers for CIBERSORTx's web server (https://cibersortx.stanford.edu/), we applied a downsampling strategy on datasets with over 3000 cells, by randomly selecting 3000 cells for analysis. The robustness of this cell sampling strategy was validated by its high correlation ($r > 0.99$ for most tests) in cell proportion estimation with bootstrap resampling of 50 times (Supplementary Figures S1 and S2), in two independent datasets. SCISSOR™ calculated the proportion of each cell type within TME by letting the summation of proportions of TME cells equal to 1. To better com-

ply with the assumption of linear model fitting, we also provided two transformations to cell proportions: the arcsine square root transformation and the logit transformation. In the GSE81861 dataset of colon cancer, we observed a cell type undefined in its original study, which showed strong similarities to cancer cells based on our analysis. We believe that the 'undefined' cell type is also cancer cells and did not include it in the analysis.

Tumor purity estimation

In SCISSOR™, the tumor purity can be obtained from the proportion of cancer cells estimated by deconvolution based on scRNA-seq data. Additionally, SCISSOR™ implemented other methods, including ABSOLUTE (41), ESTIMATE (42), IHC (43), CPE (43) and LUMP (43), which were available in R package TCGAbiolinks (44). A detailed introduction to each method is provided in the Supplementary Data.

TME cell composition, multi-omics and survival association testing

SCISSOR™ investigates the association between cancer prognosis and multi-omics data in tumors. Copy number aberration, mutation, miRNA-seq and protein expression measurements for each tumor sample were downloaded from Firehose. Also, clinical records for each patient were downloaded, including common variables such as age, sex and vital status, as well as variables that were known to be associated with specific cancers, such as smoking and alcohol history in pancreatic adenocarcinoma and esophageal carcinoma. The mutation status of the gene has been identified by TCGA. As for the CNAs (copy number alterations), we downloaded level 3 data and used R packages GenomicRanges (45) and Homo.sapiens (46) to locate CNA detection to genes. CNAs were categorized into deletion, normal and duplication by comparing segment mean to -0.2 and 0.2. For miRNAs, we downloaded level 3 miRNA expression data and performed log transformation. miRDB (47) was employed to identify miRNAs interacted with mRNAs of interest, and miRNAs with prediction scores over 50 were considered as significant candidates. Level 3 protein expression data were mapped to genes using database UniProtKB (48) and quantile normalization was applied.

SCISSOR™ provides statistical testing on associations between cell type proportion in TME with cancer prognosis and multi-omics. Appropriate models were used for different data types, respectively. For a specific gene, we used the Cox regression to evaluate the prognostic value of TME cell proportion, multi-omics and other clinical covariates by

$$\ln\left(\frac{h\left(t\right)}{h_0\left(t\right)}\right) = \beta p + \gamma x + \delta z,$$

where the independent variable is the hazard function with time and vital status. $h(t)$ is the conditional hazard function at time t. $h_0(t)$ is the baseline hazard, which represents the hazard when all covariates are equal to zero. p, x and z are matrices of covariates in all samples, where p represents proportions of interested TME cell types, x represents omics data including mRNA, miRNA, protein, mu-

tation or CNAs, and z represents covariates of clinical features and tumor purity. The coefficients, hazard ratio (HR), confidence interval (CI) and P-value can be calculated in SCISSOR™ modules. Also, Kaplan–Meier (KM) plots with CI and log-rank P-value were used to visualize the survival difference between groups of interest.

To test the associations between TME cell proportions and qualitative omics measurements adjusting for tumor purity, we used multiple linear regression and calculated the adjusted effect size, CI and P-value for qualitative omics. We modeled

$$p = \gamma x + \delta z,$$

where p is the proportion of a TME cell type as the response variable and x indicates a vector of omics data from mRNA, miRNA, protein, mutation or CNAs. We recoded mutation to 0 and 1 for wild-type and mutated status, and CNA status to -1, 0 and 1 for deletion, normal and duplication. z stands for a vector of tumor purity as a covariate to adjust for. Box plots were used to visualize the distribution of genomic aberrations (mutations and CNAs) in different TME cells. Dot plots were also used to visualize the effect size of mutation or CNA and its CI, which were adjusted for tumor purity in multiple linear regression. The associations between cell type proportions and continuous mRNA, miRNA and protein expression were shown by scatter plots. We also calculated the correlations between cell type proportions and omics measurements. To adjust for tumor purity, partial Pearson correlation coefficients controlling for tumor purity were calculated by R package ppcor (49).

Gene set enrichment analysis

To identify molecular functions associated with TME cell composition, we performed gene set enrichment analysis on identified TME associated genes. PANTHER (50) and GO (Gene Ontology) (51) databases were used for this purpose.

Web server and application construction

SCISSOR™ framework was constructed using R 3.6.1 and R shiny package. The web server was built on a CentOS Linux 7 server, with six CPUs and 16 GB memory in the environment of Jetstream (52,53). We will maintain and further develop the web server for at least 5 years.

RESULTS

TME cells play important roles in tumorigenesis and tumor progression tissue-specifically (3) and SCISSOR™ fulfills the need for a user-friendly and efficient scientific gateway for comprehensively studying the tissue-specific associations among tissue-specific tumor omics, TME cell composition and cancer prognosis. Our resource covers 16 different cancer types and integrated multi-omics data of 6031 tumor bulk tissues and scRNA-seq data from over 451 917 cells in tumors. It provides analyses on multiple omics, including mRNA expression, miRNA expression, protein expression, mutation and copy number aberration, and clinical features, including age, gender, race, histological type,

pathological stage and others collected by TCGA. SCIS-SOR™ provides five major modules, including Overview, Survival, Gene–TME cell correlation with two submodules for gene expression and genetic aberration, Genome-wide TME–omics association and Deconvolution (see Graphical abstract). Each module enables flexible modeling with adjustable parameters and dynamic visualization approaches. Figures in PDF format are also available for download. In the Survival, Gene–TME cell correlation and Genome-wide TME–omics association modules, the processed data can be downloaded in TXT, CSV and JSON formats. Detailed descriptions for each module are provided below. In this study, MuSiC deconvolution and tumor purity estimated from scRNA-seq data were mainly applied. The results of other deconvolution and tumor purity estimation methods are also summarized in tables.

Overview module

The Overview module profiles the distribution of TME cell composition in each tumor sample. With the cancer type of user's interest, SCISSOR™ will automatically load preprocessed bulk transcriptomics and other omics data as well as clinical data. Also, matched and processed single-cell data, deconvolution methods and tumor purity estimation methods will be provided to users for customized inference of TME cell composition and tumor purity in their analyses. According to user-defined parameters, a heatmap will be produced for visualizing the variation and similarity of the TME cell type proportion in tumor samples, which is useful for the identification of potential subtypes of tumors and clusters of TME cell types. Also, the distribution of cell type proportion among samples will be provided in box plots. In the application study of COAD, we inferred TME cell proportions using MuSiC and observed enterocyte-like cells, T cells, macrophages and fibroblasts were correlated by clustering analysis (Figure 1A). The correlation between enterocyte-like cells and immune cells was aligned with the finding of Stettner *et al*. (54) that nitric oxide produced by enterocytes was protective against colitis by acting as part of the innate immune response.

Survival module

Emerging evidence showed that heterogeneous TME cells, like infiltrating immune cells and cancer-associated fibroblastic cells, correlated with cancer clinical prognosis (55). We developed a 'Survival' module in SCISSOR™ to enable a dynamic platform for evaluating the prognostic association of omics measurements and TME cell composition. With the gene of interest, a flexible framework of multivariate Cox proportion hazard (CoxPH) model will be applied for assessing the hazard effects of any single or combination of cell type proportions, omics data and clinical features. It is worth noting that tumor purity, also known as the proportion of cancer cells in tumor samples, is a significant confounding factor in tumor research and has been recognized as a challenge in omics studies (56). It could confound the associations between TME cell composition and interested omics and clinical outcomes; thus, SCISSOR™ considered tumor purity as a covariate in the association model. Tu-

mor purity estimation by multiple existing methods, including ABSOLUTE, ESTIMATE, IHC, CPE and LUMP, was provided in SCISSOR™, and alternatively, it can be inferred from the proportion of cancer cells by bulk tissue deconvolution based on tissue-specific single-cell transcriptomics. Results including coefficients, the HR, CI and the *P*-value will be output. The survival difference between high-risk and low-risk groups will be visualized by KM plots. By default, high-risk and low-risk groups will be identified by the median of expression value, mutation and CNA status, or the median of TME cell type proportion. We also provided a slider for users to easily identify high-risk and low-risk groups by the upper and lower *X* percentiles. Log-rank *P*-value and CI will also be calculated and displayed in KM plots.

To provide a demo application, we utilized SCISSOR™ to investigate the prognostic value of TME cell composition in COAD. We adjusted the association testing for age, gender, race and tumor purity in multivariate Cox regression. SCISSOR™ found that lower tumor purity was associated with a worse survival outcome (CoxPH model, HR = 0.071, *P*-value = 0.042), which is consistent with previous reports (57,58). Also, KM curves and the log-rank test indicated a significant association of enterocyte-like cells' proportion with a favored survival outcome (*P*-value = 0.017, Figure 1B). This implies consistency with the finding of Sadanandam *et al*. (59) that enterocytes were enriched in a subtype of colorectal cancer, and such enterocyte subtype showed a significantly better prognosis than the stem-like subtype.

To validate the analysis of SCISSOR™, we study a gene *EPHB2*, which was found to be positively associated with the survival outcome in COAD [CoxPH model, HR = 0.45, *P*-value = 0.035 from Jubb *et al*. (60); HR = 0.43, *P*-value < 0.001 from Martinez-Romero *et al*. (61)]. The Gene–TME cell correlation module of SCISSOR™ (shown below) found a negative correlation between TME T-cell proportion and *EPHB2* gene expression (Pearson coefficient $r = -0.37$, *P*-value < 0.001, Table 1), which is aligned well with the previous finding (62,63) that EphB/ephrin-B could suppress T-cell activation (64). Consistent associations were found in analyses adjusted for tumor purities estimated from five out of six methods (Table 1). Further, SCISSOR™ validated the prognostic value of *EPHB2* (multivariate CoxPH model, HR = 0.696, *P*-value = 0.016) in COAD, with the effect from age, gender, race, TME cell proportion and tumor purity adjusted. The prognostic effect of *EPHB2* was consistently detected using different deconvolution and tumor purity estimation methods (Table 2), and the survival curves were significantly differentiated by the median of *EPHB2* expression (log-rank test, *P*-value = 0.029, Figure 1C). By analyzing omics data, we found the miRNA *hsa-mir-424*, which was predicted to target at *EPHB2* in miRDB (47), was significantly associated with COAD prognosis (CoxPH model, HR = 1.765, *P*-value = 0.010, Table 2). This identification supports the previous report of Kandhavelu *et al*. (65) that *hsa-mir-424* potentially targets 10 colon cancer hallmark genes. Moreover, Oba *et al*. (66) found that the chromosome deletion in *EPHB2* could lead to loss of function. This result was also validated by the association of *EPHB2* expression with its CNA and COAD prognosis in SCISSOR™ (CoxPH model, HR = 1.985,
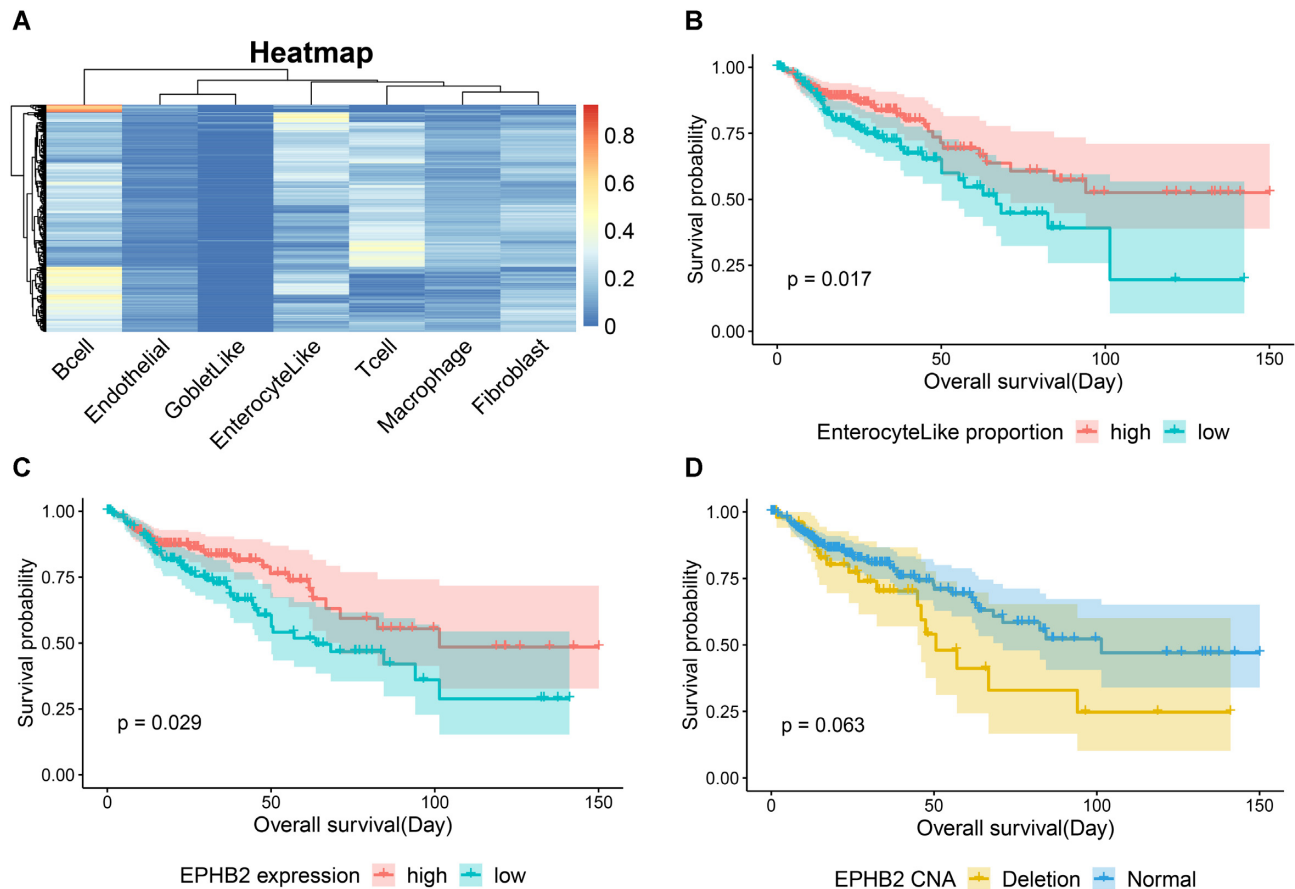
Figure 1. Application of Overview module and Survival module on COAD study. (**A**) The heatmap of a two-way hierarchical clustering analysis consisting of the cell proportions (column) deconvolved from bulk transcriptomic data of tumor samples (row). Red color represents a relatively high proportion of cell type, while blue color indicates a relatively lower proportion. (**B–D**) The KM plots of the gene *EPHB2*, miRNA *mir-155* and CNA status of *EPHB2*, with the *P*-value of the log-rank test and CI shown by shade.

P-value = 0.022, log-rank *P*-value = 0.063, Figure 1D). Similar results were also observed with other deconvolution and tumor purity estimation methods (Table 2). Collectively, these results showed the value of SCISSOR™ and also validated the prognostic association of *EPHB2* expression and potential related regulatory programs.

Gene–TME cell correlation module

Genes interact with TME cell composition at different omic dimensions (67–70). To investigate the associations between the proportion of tissue-specific cell types and omics measurements, we designed the Gene–TME composition correlation module. This module enables the identification of TME cell markers as well as genetic factors involved in tumor–TME interactions. It provides two submodules: the expression correlation module for quantitative omics measurements (mRNA, miRNA and protein expression) and the gene aberration module for qualitative omics measurements (mutation and copy number aberration).

Expression correlation submodule

Evidence has shown that mRNA, miRNA and protein could be valuable biomarkers for cancer (71); thus, we developed three components for studying them respectively and comprehensively in this submodule. With omics data of a specific tissue or cancer type, TME cell proportion estimation and inferred tumor purity, users can investigate the interaction of a particular gene, miRNA or protein with the proportion of cell types of their interest in TME. Scatter plots with the fitted curves, Pearson correlation coefficients and *P*-values will be provided for association evaluation. SCISSOR™ will also calculate partial Pearson correlation coefficients and *P*-value to adjust the confounding effect from tumor purity.

We tested whether SCISSOR™ can detect the association between T cells and *CD4*, which codes a glycoprotein on the surface of immune cells such as T cells and macrophages. SCISSOR™ indeed found a significant and strong correlation between *CD4* gene expression with T-cell proportion without ($r = 0.81$, *P*-value < 0.001) or with adjustment for tumor purity estimated from single-cell transcriptomic data ($r = 0.49$, *P*-value < 0.001, Figure 2A). Consistent results were also found by different tumor purity estimation methods (Table 1). Also, we studied *mir-155*, which is required for optimal T-cell activation and reinforcement of T-cell response (72). With tumor purity adjusted, SCISSOR™ found *mir-155* expression positively correlated with T-cell proportion (adjusted partial $r = 0.52$, *P*-value < 0.001, Table 1,

Table 1. Associations between *EPHB2*, *CD4* and *hsa-mir-155* expression with T-cell proportion in COAD using different tumor purity estimation and deconvolution methods

Deconvolution	Tumor purity	EPHB2			CD4			hsa-mir-155		
		r	*r'*	*P*	*r*	*r**	*P*	*r*	*r'*	*P*
CIBERSORTx	Absolute	−0.07	−0.06	0.39	0.22	0.19	<0.01*	0.33	0.34	<0.01*
	Estimate	0.08	0.11	0.05*	0.08	0.06	0.31	0.31	0.30	<0.01*
	CPE	−0.07	−0.03	0.61	0.22	0.16	0.01*	0.33	0.28	<0.01*
	IHC	−0.07	−0.06	0.29	0.22	0.23	<0.01*	0.33	0.33	<0.01*
	LUMP	−0.07	−0.05	0.44	0.22	0.11	0.07	0.33	0.26	<0.01*
	Sc-Est	0.08	0.09	0.10	0.08	0.11	0.06	0.31	0.32	<0.01*
MuSiC	Absolute	−0.30	−0.22	<0.01*	0.81	0.74	<0.01*	0.44	0.42	<0.01*
	Estimate	−0.37	−0.13	0.02*	0.81	0.36	<0.01*	0.40	0.58	<0.01*
	CPE	−0.30	−0.19	<0.01*	0.81	0.68	<0.01*	0.44	0.37	<0.01*
	IHC	−0.30	−0.32	<0.01*	0.81	0.81	<0.01*	0.44	0.43	<0.01*
	LUMP	−0.30	−0.25	<0.01*	0.81	0.77	<0.01*	0.44	0.36	<0.01*
	Sc-Est	−0.37	0.05	<0.01*	0.81	0.49	<0.01*	0.40	0.52	<0.01*

Note: HR and *P*-value were calculated by Cox regression adjusting for age, gender, race, cell type proportions and tumor purity. *r'*: partial correlation adjusting for tumor purity. * indicates significance (*P*<0.05). CNA: copy number aberration, normal versus deletion. Sc-Est: single-cell estimation.

Table 2. Associations between *EPHB2* expression, *hsa-mir-424* expression and *EPHB2* CNA with survival outcome in COAD using different tumor purity estimation and deconvolution methods

Deconvolution	Tumor purity	EPHB2		hsa-mir-424		EPHB2 CNA	
		HR	*P*	HR	*P*	HR	*P*
CIBERSORTx	Absolute	0.914	0.603	1.692	0.027*	1.521	0.251
	Estimate	0.749	0.044*	1.605	0.031*	1.892	0.041*
	CPE	0.808	0.174	1.543	0.052	1.963	0.028*
	IHC	0.815	0.192	1.528	0.056	1.794	0.046*
	LUMP	0.851	0.317	1.489	0.073	1.689	0.101
	Sc-Est	0.752	0.055	1.597	0.031*	1.820	0.045*
MuSiC	Absolute	0.809	0.236	1.729	0.015*	1.353	0.393
	Estimate	0.680	0.010*	1.687	0.017*	1.971	0.023*
	CPE	0.721	0.044*	1.665	0.019*	1.921	0.032*
	IHC	0.727	0.050	1.631	0.024*	1.904	0.031*
	LUMP	0.778	0.129	1.600	0.028*	1.711	0.086
	Sc-Est	0.696	0.016*	1.765	0.010*	1.985	0.022*

Note: HR and *P*-value were calculated by Cox regression adjusting for age, gender, race, cell type proportions and tumor purity. * indicates significance (*P*<0.05). CNA: copy number aberration, normal versus deletion. Sc-Est: single-cell estimation.

Supplementary Figure S3). Given the important role of T cells in immune surveillance of tumors, this observation was aligned with the reported association of *mir-155* expression with immune response and favorable prognosis in melanoma patients (72).

Genetic aberration submodule

To explore the association between TME cell composition and genetic aberrations, we developed two components for mutation and CNA separately. In the 'Mutation' component, SCISSOR™ sorts genes according to frequencies of their somatic mutation so that users can easily identify potential tumor driver genes. Box plots will be produced for visualizing the distribution of mutation status (mutation and wild type) or CNA status (deletion, normal and duplication) in each cell type. An adjusted effect size will be calculated using multiple linear regression with tumor purity as a covariate (see the 'Materials and Methods' section).

In application, we found *APC* was the most frequently somatically mutated gene in COAD. In higher purity tumors, it is more likely to observe mutation of *APC* (*P*-value = 0.020, Figure 2D). This result is supported by the findings that the mutation of *APC* can lead to colon cancer by inducing familial adenomatous polyposis, a significant hereditary predisposition indicator for colon adenocarcinoma (73), and by activating the Wnt signal transduction pathway (74,75). Interestingly, SCISSOR™ detected more mutations of *APC* in tumors with lower T-cell proportions than those with higher T-cell proportions (*P*-value = 0.016, Figure 2D). This could be an outcome of previous findings that the loss of function of *APC* is associated with reduced T-cell differentiation and lower cytokine production like IL-10 (76), which could lead to the suppression of T-cell function and the promotion of inflammation, and further cause adenocarcinoma progression (77).

Genome-wide association module

The 'genome-wide association' module was developed to enable a systematic investigation of tumor immunology and tumor–TME interaction by identifying transcriptome-wide TME-associated genes. For a particular TME cell type, SCISSOR™ can identify and summarize the top 100 positively and negatively genes correlated with the cell proportion, and visualize these correlations in a heatmap.
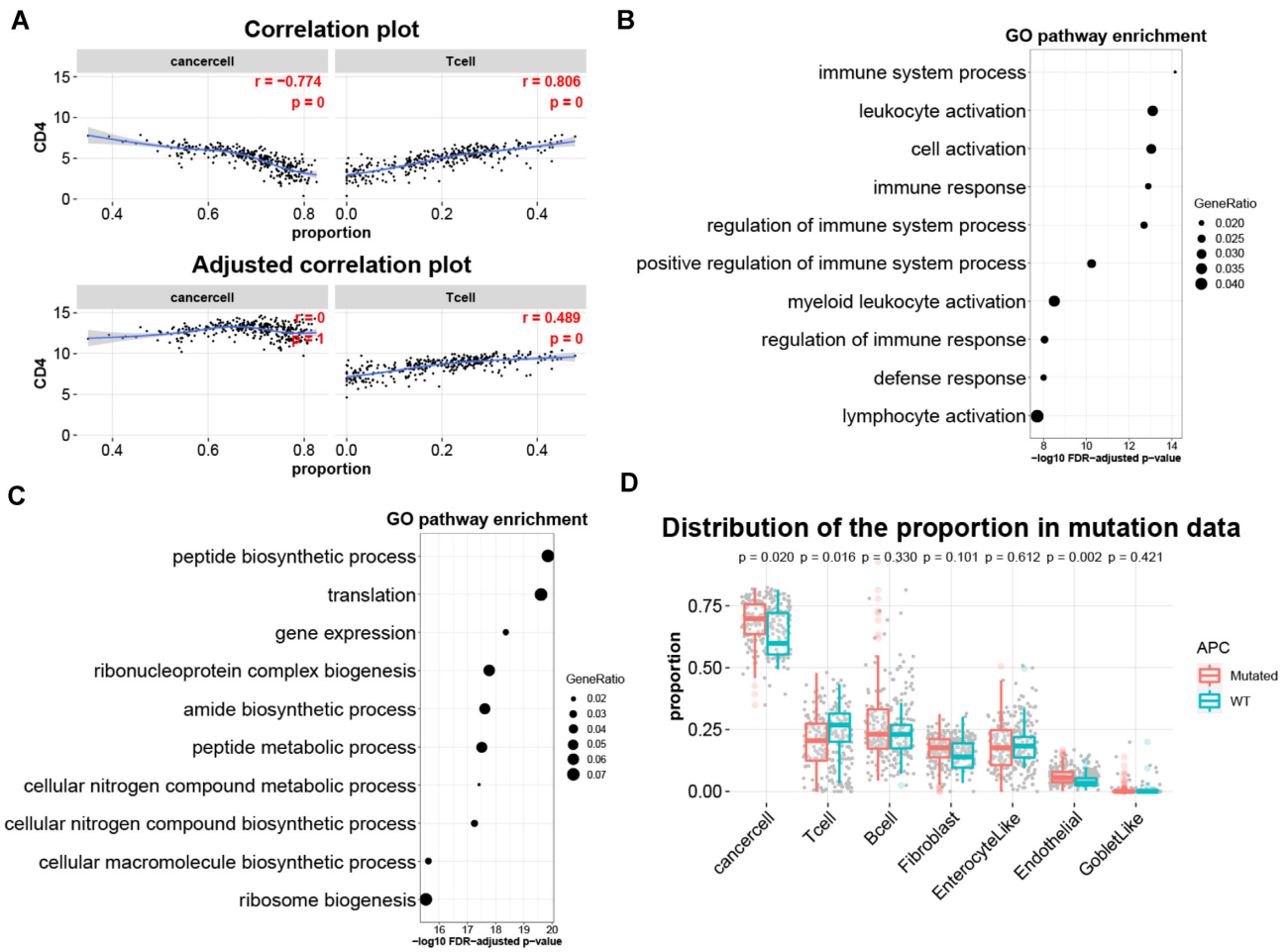
Figure 2. Application of Gene–TME cell correlation module, genetic aberration module and expression correlated module on COAD study. (**A**) Correlation plots of the *CD4* expression with TME T-cell proportion in the Gene–TME cell correlation module. Pearson correlation coefficients and *P*-values are shown in plots. The result adjusting for tumor purity is also shown in the bottom panel. (**B**, **C**) The GO biological processes enriched by the top 100 positively and negatively correlated genes, respectively, which were recognized through the expression correlated module. (**D**) The distribution of T-cell proportion in tumors with mutated or wild-type *APC* in the genetic aberration module.

In this study, we applied SCISSOR™ and identified genes that were most associated with TME cells of COAD. Hierarchical clustering and heatmap of detected associations showed that macrophages, T cells and fibroblasts were closely related (Supplementary Figure S4). Gene set enrichment analysis found that the top 100 genes that were positively correlated with TME T cells were enriched in the GO biological process of the immune system (Figure 2B, Table 3), while the negatively correlated genes were interestingly enriched in the biological process of cell metabolism (Figure 2C, Table 4).

Deconvolution module

SCISSOR™ also provides a platform for inferring tissue-specific TME cell proportions from users' private bulk transcriptome datasets with raw counts. The deconvolution method MuSiC and cell type transcriptome profile reference from single-cell transcriptomics specific to cancers of colon, breast, brain, bile duct, kidney, liver, lung, pancreas, head and neck, skin and blood are currently available for use. Other deconvolution methods and more tissue-specific

transcriptome profile references will be continuously added to our web server. Moreover, SCISSOR™ provides a user-friendly interface and will automatically send results to users by email.

DISCUSSION

Large-scale association studies of omics data, TME cell composition and tumor prognosis are highly needed for understanding the important role of TME in tumor progression. Here, we introduced a new web server resource for tissue-specific deconvolution of TME cell proportion from bulk transcriptomic data and characterization of the association of TME cell composition with tumor prognosis and omics, with five comprehensive modules, a flexible modeling framework, and processed data of large-scale omics and high-granularity single-cell transcriptomics. SCISSOR™ fills the gap of large-scale TME cell sorting by integrating TCGA bulk RNA-seq data of various types of tumors with tissue-specific scRNA-seq data and establishes the first online open-access resource for inferring the proportion of tissue-specific cell types from bulk expres-

Table 3. Top 20 enriched biological processes of the top 100 positively correlated genes with T-cell proportion in TME of COAD

Pathway	GeneRatio	FDR adjusted *P*-value
Immune system process (GO:0002376)	1.83%	6.94E−15
Leukocyte activation (GO:0045321)	3.32%	7.58E−14
Cell activation (GO:0001775)	3.04%	9.12E−14
Immune response (GO:0006955)	2.13%	1.28E−13
Regulation of immune system process (GO:0002682)	2.29%	2.01E−13
Positive regulation of immune system process (GO:0002684)	2.75%	5.73E−11
Myeloid leukocyte activation (GO:0002274)	3.58%	3.10E−09
Regulation of immune response (GO:0050776)	2.39%	9.33E−09
Defense response (GO:0006952)	2.12%	1.02E−08
Lymphocyte activation (GO:0046649)	4.35%	2.05E−08
Regulation of response to stimulus (GO:0048583)	1.20%	4.25E−08
Response to biotic stimulus (GO:0009607)	1.99%	7.23E−08
Positive regulation of immune response (GO:0050778)	2.85%	1.06E−07
Immune effector process (GO:0002252)	2.28%	1.15E−07
Response to external biotic stimulus (GO:0043207)	1.97%	1.53E−07
Activation of immune response (GO:0002253)	3.40%	1.54E−07
Response to other organism (GO:0051707)	1.97%	1.58E−07
Leukocyte activation involved in immune response (GO:0002366)	3.02%	2.69E−07
Cell activation involved in immune response (GO:0002263)	3.00%	2.82E−07
Positive regulation of response to stimulus (GO:0048584)	1.54%	2.94E−07

GeneRatio: Gene ratio is calculated as the number of selected genes in the pathway, divided by the total number of genes in the reference dataset that make up the pathway.

Table 4. Top 20 enriched biological processes of the top 100 negatively correlated genes with T-cell proportion in TME of COAD

Pathway	GeneRatio	FDR adjusted *P*-value
Peptide biosynthetic process (GO:0043043)	7.09%	1.39E−20
Translation (GO:0006412)	7.29%	2.48E−20
Gene expression (GO:0010467)	2.33%	4.38E−19
Ribonucleoprotein complex biogenesis (GO:0022613)	6.02%	1.66E−18
Amide biosynthetic process (GO:0043604)	5.49%	2.39E−18
Peptide metabolic process (GO:0006518)	5.40%	3.11E−18
Cellular nitrogen compound metabolic process (GO:0034641)	1.76%	3.93E−18
Cellular nitrogen compound biosynthetic process (GO:0044271)	2.73%	5.58E−18
Cellular macromolecule biosynthetic process (GO:0034645)	2.54%	2.52E−16
Ribosome biogenesis (GO:0042254)	6.93%	2.92E−16
SRP-dependent cotranslational protein targeting to membrane (GO:0006614)	16.67%	3.88E−16
Macromolecule biosynthetic process (GO:0009059)	2.48%	4.58E−16
Cotranslational protein targeting to membrane (GO:0006613)	15.84%	6.81E−16
Nuclear-transcribed mRNA catabolic process (GO:0000956)	9.69%	1.18E−15
Protein targeting to ER (GO:0045047)	14.41%	2.30E−15
Translational initiation (GO:0006413)	11.89%	3.38E−15
Establishment of protein localization to endoplasmic reticulum (GO:0072599)	13.91%	3.39E−15
Viral transcription (GO:0019083)	13.79%	3.63E−15
mRNA catabolic process (GO:0006402)	8.80%	4.70E−15
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (GO:0000184)	13.33%	5.33E−15

GeneRatio: Gene ratio is calculated as the number of selected genes in the pathway, divided by the total number of genes in the reference dataset that make up the pathway.

sion data. We demonstrated the value of SCISSOR™ by application studies on the prognostic effect of *EPHB2* and its mapped miRNA *hsa-mir-424* and CNA in COAD, and the association between *CD4* and *hsa-mir-155* with T-cell proportion, etc.

SCISSOR™ detected a negative association between metabolism-related genes in tumor with T-cell proportion in TME, which brings new insight and hypothesis of cancer metabolism and immunity interactions. The negative correlation between T-cell proportion in TME and the metabolism process in tumors may be related to nutrient competition and hypoxia within the TME. The expansion of both tumor cells and immune cells requires aerobic glycolysis (78), which may cause competition for metabolic re-

sources. Indeed, studies indicated that tumor cells compete for glucose resources from neighboring cells to maintain their metabolism (79). This might affect the proliferation and activation of T cells as glucose is a critical substrate for T cells to play an active anti-tumor role (80). Moreover, the hypoxia in tumors induces glucose utilization and lactate release in cancer cells (81), and continuous exposure to extracellular lactic acid can strongly inhibit the expansion of T cells, consequently decreasing the immune activity (82). Further investigation was required to characterize this potential negative association between T-cell composition and metabolism in tumors.

SCISSOR™ currently provides scRNA-seq datasets for 16 cancer types, and will maintain and expand the database

with newly generated datasets. One of the limitations of our studies could be the incomplete identification of cell subtypes in each single-cell transcriptomic study. To address this challenge, we will analyze the latest scRNA-seq studies and update the hosted database frequently. The current study is based on datasets from different studies. The variations in study design, experiment and analysis platform brought difficulties in cross-study comparisons. In future studies, we will collaborate with experts in the research of each cancer type and re-analyze the data in the same analysis framework. Also, several scRNA-seq datasets were filtered out by the current exclusion criteria due to the missing cell type information. We will re-annotate those studies and include them in SCISSOR™. In addition, the downsampling strategy used in SCISSOR™ may lead to random selection bias; more efficient methods will be searched and implemented for utilizing all information. Despite these limitations, SCISSOR™ is valuable as a new resource for promoting tumor heterogeneity and tumor–TME interaction research.

DATA AVAILABILITY

Bulk data were downloaded from Broad GDAC Firehose. scRNA-seq data were available in multiple databases, including GEO, EMBL, heiDATA, GSA (Genome Sequence Archive) and EGA (European Genome-phenome Archive). Detailed information of each dataset is described in the 'Materials and Methods' section.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Cancer Online.

REFERENCES

1. Wang,Y., Mashock,M., Tong,Z., Mu,X., Chen,H., Zhou,X., Zhang,H., Zhao,G., Liu,B. and Li,X. (2020) Changing technologies of RNA sequencing and their applications in clinical oncology. *Front. Oncol.*, **10**, 447.
2. Runa,F., Hamalian,S., Meade,K., Shisgal,P., Gray,P.C. and Kelber,J.A. (2017) Tumor microenvironment heterogeneity: challenges and opportunities. *Curr. Mol. Biol. Rep.*, **3**, 218–229.
3. Schneider,G., Schmidt-Supprian,M., Rad,R. and Saur,D. (2017) Tissue-specific tumorigenesis: context matters. *Nat. Rev. Cancer*, **17**, 239–253.
4. Omary,M.B., Lugea,A., Lowe,A.W. and Pandol,S.J. (2007) The pancreatic stellate cell: a star on the rise in pancreatic diseases. *J. Clin. Invest.*, **117**, 50–59.
5. Fidler,I.J. (2003) The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat. Rev. Cancer*, **3**, 453–458.
6. Peinado,H., Zhang,H., Matei,I.R., Costa-Silva,B., Hoshino,A., Rodrigues,G., Psaila,B., Kaplan,R.N., Bromberg,J.F., Kang,Y. *et al.* (2017) Pre-metastatic niches: organ-specific homes for metastases. *Nat. Rev. Cancer*, **17**, 302–317.
7. van Galen,P., Hovestadt,V., Wadsworth,I.M., Hughes,T.K., Griffin,G.K., Battaglia,S., Verga,J.A., Stephansky,J., Pastika,T.J., Lombardi,S.J. *et al.* (2019) Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. *Cell*, **176**, 1265–1281.
8. Azizi,E., Carr,A.J., Plitas,G., Cornish,A.E., Konopacki,C., Prabhakaran,S., Nainys,J., Wu,K., Kiseliovas,V., Setty,M. *et al.* (2018) Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell*, **174**, 1293–1308.
9. Zheng,C., Zheng,L., Yoo,J.K., Guo,H., Zhang,Y., Guo,X., Kang,B., Hu,R., Huang,J.Y., Zhang,Q. *et al.* (2017) Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell*, **169**, 1342–1356.
10. Avila Cobos,F., Vandesompele,J., Mestdagh,P. and De Preter,K. (2018) Computational deconvolution of transcriptomics data from mixed cell populations. *Bioinformatics*, **34**, 1969–1979.
11. Venet,D., Pecasse,F., Maenhaut,C. and Bersini,H. (2001) Separation of samples into their constituents using gene expression data. *Bioinformatics*, **17**, S279–S287.
12. Abbas,A.R., Wolslegel,K., Seshasayee,D., Modrusan,Z. and Clark,H.F. (2009) Deconvolution of blood microarray data identifies cellular activation patterns in systemic lupus erythematosus. *PLoS One*, **4**, e6098.
13. Turlach,B. and Weingessel,A. (2013) In: Quadprog: functions to solve quadratic programming problems.
14. Soetaert,K., Van den Meersche,K. and Oevelen,D. (2009) In: Package limSolve, solving linear inverse models in R.
15. Repsilber,D., Kern,S., Telaar,A., Walzl,G., Black,G.F., Selbig,J., Parida,S.K., Kaufmann,S.H. and Jacobsen,M. (2010) Biomarker discovery in heterogeneous tissue samples—taking the *in-silico* deconfounding approach. *BMC Bioinformatics*, **11**, 27.
16. Venet,D., Pecasse,F., Maenhaut,C. and Bersini,H. (2001) Separation of samples into their constituents using gene expression data. *Bioinformatics*, **17**, S279–S287.
17. Zuckerman,N.S., Noam,Y., Goldsmith,A.J. and Lee,P.P. (2013) A self-directed method for cell-type identification and separation of gene expression microarrays. *PLoS Comput. Biol.*, **9**, e1003189.
18. Newman,A.M., Liu,C.L., Green,M.R., Gentles,A.J., Feng,W., Xu,Y., Hoang,C.D., Diehn,M. and Alizadeh,A.A. (2015) Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*, **12**, 453–457.
19. Wang,X., Park,J., Susztak,K., Zhang,N.R. and Li,M. (2019) Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.*, **10**, 380.
20. Charoentong,P., Finotello,F., Angelova,M., Mayer,C., Efremova,M., Rieder,D., Hackl,H. and Trajanoski,Z. (2017) Pan-cancer immunogenomic analyses reveal genotype–immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.*, **18**, 248–262.
21. Gentles,A.J., Newman,A.M., Liu,C.L., Bratman,S.V., Feng,W., Kim,D., Nair,V.S., Xu,Y., Khuong,A., Hoang,C.D. *et al.* (2015) The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.*, **21**, 938–945.
22. Li,T., Fan,J., Wang,B., Traugh,N., Chen,Q., Liu,J.S., Li,B. and Liu,X.S. (2017) TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res.*, **77**, e108–e110.
23. Chen,Z., Ji,C., Shen,Q., Liu,W., Qin,F.X. and Wu,A. (2020) Tissue-specific deconvolution of immune cell composition by integrating bulk and single-cell transcriptomes. *Bioinformatics*, **36**, 819–827.
24. Mass,E., Ballesteros,I., Farlik,M., Halbritter,F., Günther,P., Crozet,L., Jacome-Galarza,C.E., Händler,K., Klughammer,J., Kobayashi,Y. *et al.* (2016) Specification of tissue-resident macrophages during organogenesis. *Science*, **353**, aaf4238.
25. Yan,J., Risacher,S.L., Shen,L. and Saykin,A.J. (2018) Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief. Bioinform.*, **19**, 1370–1381.

26. Li,H., Courtois,E.T., Sengupta,D., Tan,Y., Chen,K.H., Goh,J.J.L., Kong,S.L., Chua,C., Hon,L.K., Tan,W.S. *et al.* (2017) Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat. Genet.*, **49**, 708–718.

27. Puram,S.V., Tirosh,I., Parikh,A.S., Patel,A.P., Yizhak,K., Gillespie,S., Rodman,C., Luo,C.L., Mroz,E.A., Emerick,K.S. *et al.* (2017) Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell*, **171**, 1611–1624.

28. Azizi,E., Carr,A.J., Plitas,G., Cornish,A.E., Konopacki,C., Prabhakaran,S., Nainys,J., Wu,K., Kiseliovas,V., Setty,M. *et al.* (2018) Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell*, **174**, 1293–1308.

29. Darmanis,S., Sloan,S.A., Croote,D., Mignardi,M., Chernikova,S., Samghababi,P., Zhang,Y., Neff,N., Kowarsky,M., Caneda,C. *et al.* (2017) Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.*, **21**, 1399–1410.

30. Ma,L., Hernandez,M.O., Zhao,Y., Mehta,M., Tran,B., Kelly,M., Rae,Z., Hernandez,J.M., Davis,J.L., Martin,S.P. *et al.* (2019) Tumor cell biodiversity drives microenvironmental reprogramming in liver cancer. *Cancer Cell*, **36**, 418–430.

31. van Galen,P., Hovestadt,V., Wadsworth,I.M., Hughes,T.K., Griffin,G.K., Battaglia,S., Verga,J.A., Stephansky,J., Pastika,T.J., Lombardi,S.J. *et al.* (2019) Single-cell RNA-seq reveals AML hierarchies relevant to disease progression and immunity. *Cell*, **176**, 1265–1281.

32. Tirosh,I., Izar,B., Prakadan,S.M., Wadsworth,M.N., Treacy,D., Trombetta,J.J., Rotem,A., Rodman,C., Lian,C., Murphy,G. *et al.* (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, **352**, 189–196.

33. Kim,N., Kim,H.K., Lee,K., Hong,Y., Cho,J.H., Choi,J.W., Lee,J.I., Suh,Y.L., Ku,B.M., Eum,H.H. *et al.* (2020) Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.*, **11**, 2285.

34. Cheng,S., Li,Z., Gao,R., Xing,B., Gao,Y., Yang,Y., Qin,S., Zhang,L., Ouyang,H., Du,P *et al.* (2021) A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. *Cell*, **184**, 792–809.

35. Zhang,Q., He,Y., Luo,N., Patel,S.J., Han,Y., Gao,R., Modak,M., Carotta,S., Haslinger,C., Kind,D. *et al.* (2019) Landscape and dynamics of single immune cells in hepatocellular carcinoma. *Cell*, **179**, 829–845.

36. Li,H., van der Leun,A.M., Yofe,I., Lubling,Y., Gelbard-Solodkin,D., van Akkooi,A., van den Braber,M., Rozeman,E.A., Haanen,J., Blank,C.U. *et al.* (2019) Dysfunctional CD8 T cells form a proliferative, dynamically regulated compartment within human melanoma. *Cell*, **176**, 775–789.

37. Peng,J., Sun,B.F., Chen,C.Y., Zhou,J.Y., Chen,Y.S., Chen,H., Liu,L., Huang,D., Jiang,J., Cui,G.S. *et al.* (2019) Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res.*, **29**, 725–738.

38. Lambrechts,D., Wauters,E., Boeckx,B., Aibar,S., Nittner,D., Burton,O., Bassez,A., Decaluwé,H., Pircher,A., Van den Eynde,K. *et al.* (2018) Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.*, **24**, 1277–1289.

39. Young,M.D., Mitchell,T.J., Vieira,B.F., Tran,M., Stewart,B.J., Ferdinand,J.R., Collord,G., Botting,R.A., Popescu,D.M., Loudon,K.W. *et al.* (2018) Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science*, **361**, 594–599.

40. Roider,T., Seufert,J., Uvarovskii,A., Frauhammer,F., Bordas,M., Abedpour,N., Stolarczyk,M., Mallm,J.P., Herbst,S.A., Bruch,P.M. *et al.* (2020) Dissecting intratumour heterogeneity of nodal B-cell lymphomas at the transcriptional, genetic and drug-response levels. *Nat. Cell Biol.*, **22**, 896–906.

41. Carter,S.L., Cibulskis,K., Helman,E., McKenna,A., Shen,H., Zack,T., Laird,P.W., Onofrio,R.C., Winckler,W., Weir,B.A. *et al.* (2012) Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.*, **30**, 413–421.

42. Yoshihara,K., Shahmoradgoli,M., Martínez,E., Vegesna,R., Kim,H., Torres-Garcia,W., Treviño,V., Shen,H., Laird,P.W., Levine,D.A. *et al.* (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.*, **4**, 2612.

43. Aran,D., Sirota,M. and Butte,A.J. (2015) Systematic pan-cancer analysis of tumour purity. *Nat. Commun.*, **6**, 8971.

44. Colaprico,A., Silva,T.C., Olsen,C., Garofano,L., Cava,C., Garolini,D., Sabedot,T.S., Malta,T.M., Pagnotta,S.M., Castiglioni,I. *et al.* (2016) TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.*, **44**, e71.

45. Lawrence,M., Huber,W., Pagès,H., Aboyoun,P., Carlson,M., Gentleman,R., Morgan,M.T. and Carey,V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.*, **9**, e1003118.

46. Bioconductor Core Team (2015) Homo.sapiens: Annotation package for the Homo.sapiens object.

47. Chen,Y. and Wang,X. (2020) miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.*, **48**, D127–D131.

48. UnitProt Consotium. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.

49. Kim,S. (2015) ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Commun. Stat. Appl. Methods*, **22**, 665–674.

50. Mi,H., Muruganujan,A., Ebert,D., Huang,X. and Thomas,P.D. (2019) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.*, **47**, D419–D426.

51. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

52. Towns,J., Cockerill,T., Dahan,M., Foster,I., Gaither,K., Grimshaw,A., Hazlewood,V., Lathrop,S., Lifka,D., Peterson,G.D. *et al.* (2014) XSEDE: accelerating scientific discovery. *Comput. Sci. Eng.*, **16**, 62–74.

53. Stewart,C.A., Cockerill,T.M., Foster,I., Hancock,D., Merchant,N., Skidmore,E., Stanzione,D., Taylor,J., Tuecke,S., Turner,G. *et al.* (2015) In: *XSEDE '15*. NY.

54. Stettner,N., Rosen,C., Bernshtein,B., Gur-Cohen,S., Frug,J., Silberman,A., Sarver,A., Carmel-Neiderman,N.N., Eilam,R., Biton,I. *et al.* (2018) Induction of nitric-oxide metabolism in enterocytes alleviates colitis and inflammation-associated colon cancer. *Cell Rep.*, **23**, 1962–1976.

55. Hanahan,D. and Coussens,L.M. (2012) Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell*, **21**, 309–322.

56. Petralia,F., Wang,L., Peng,J., Yan,A., Zhu,J. and Wang,P. (2018) A new method for constructing tumor specific gene co-expression networks based on samples with tumor purity heterogeneity. *Bioinformatics*, **34**, i528–i536.

57. Gong,Z., Zhang,J. and Guo,W. (2020) Tumor purity as a prognosis and immunotherapy relevant feature in gastric cancer. *Cancer Med.*, **9**, 9052–9063.

58. Zhang,C., Cheng,W., Ren,X., Wang,Z., Liu,X., Li,G., Han,S., Jiang,T. and Wu,A. (2017) Tumor purity as an underlying key factor in glioma. *Clin. Cancer Res.*, **23**, 6279–6291.

59. Sadanandam,A., Lyssiotis,C.A., Homicsko,K., Collisson,E.A., Gibb,W.J., Wullschleger,S., Ostos,L.C., Lannon,W.A., Grotzinger,C., Del,R.M. *et al.* (2013) A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.*, **19**, 619–625.

60. Jubb,A.M., Zhong,F., Bheddah,S., Grabsch,H.I., Frantz,G.D., Mueller,W., Kavi,V., Quirke,P., Polakis,P. and Koeppen,H. (2005) EphB2 is a prognostic factor in colorectal cancer. *Clin. Cancer Res.*, **11**, 5181.

61. Martinez-Romero,J., Bueno-Fortes,S., Martín-Merino,M., Ramirez,D.M.A. and De Las,R.J. (2018) Survival marker genes of colorectal cancer derived from consistent transcriptomic profiling. *BMC Genomics*, **19**, 857.

62. Kawano,H., Katayama,Y., Minagawa,K., Shimoyama,M., Henkemeyer,M. and Matsui,T. (2012) A novel feedback mechanism by Ephrin-B1/B2 in T-cell activation involves a concentration-dependent switch from costimulation to inhibition. *Eur. J. Immunol.*, **42**, 1562–1572.

63. Nguyen,T.M., Arthur,A., Hayball,J.D. and Gronthos,S. (2013) EphB and Ephrin-B interactions mediate human mesenchymal stem cell suppression of activated T-cells. *Stem Cells Dev.*, **22**, 2751–2764.

64. Shiuan,E. and Chen,J. (2016) Eph receptor tyrosine kinases in tumor immunity. *Cancer Res.*, **76**, 6452–6457.

65. Kandhavelu,J., Subramanian,K., Khan,A., Omar,A., Ruff,P. and Penny,C. (2019) Computational analysis of miRNA and their gene targets significantly involved in colorectal cancer progression. *MicroRNA*, **8**, 68–75.

66. Oba,S.M., Wang,Y.J., Song,J.P., Li,Z.Y., Kobayashi,K., Tsugane,S., Hamada,G.S., Tanaka,M. and Sugimura,H. (2001) Genomic structure and loss of heterozygosity of EPHB2 in colorectal cancer. *Cancer Lett.*, **164**, 97–104.

67. Rupaimoole,R., Calin,G.A., Lopez-Berestein,G. and Sood,A.K. (2016) miRNA deregulation in cancer cells and the tumor microenvironment. *Cancer Discov.*, **6**, 235–246.

68. Bindra,R.S. and Glazer,P.M. (2005) Genetic instability and the tumor microenvironment: towards the concept of microenvironment-induced mutagenesis. *Mutat. Res.*, **569**, 75–85.

69. Joyce,J.A. and Fearon,D.T. (2015) T cell exclusion, immune privilege, and the tumor microenvironment. *Science*, **348**, 74–80.

70. Hasin,Y., Seldin,M. and Lusis,A. (2017) Multi-omics approaches to disease. *Genome Biol.*, **18**, 83.

71. Barbieri,I. and Kouzarides,T. (2020) Role of RNA modifications in cancer. *Nat. Rev. Cancer*, **20**, 303–322.

72. Ekiz,H.A., Huffaker,T.B., Grossmann,A.H., Stephens,W.Z., Williams,M.A., Round,J.L. and O'Connell,R.M. (2019) MicroRNA-155 coordinates the immunological landscape within murine melanoma and correlates with immunity in human cancers. *JCI Insight*, **4**, e126543.

73. Fearon,E.R. (2011) Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.*, **6**, 479–507.

74. Zhang,L. and Shay,J.W. (2017) Multiple roles of APC and its therapeutic implications in colorectal cancer. *J. Natl Cancer Inst.*, **109**, djw332.

75. Fodde,R., Smits,R. and Clevers,H. (2001) APC, signal transduction and genetic instability in colorectal cancer. *Nat. Rev. Cancer*, **1**, 55–67.

76. Gounaris,E., Blatner,N.R., Dennis,K., Magnusson,F., Gurish,M.F., Strom,T.B., Beckhove,P., Gounari,F. and Khazaie,K. (2009) T-regulatory cells shift from a protective anti-inflammatory to a cancer-promoting proinflammatory phenotype in polyposis. *Cancer Res.*, **69**, 5490–5497.

77. Agüera-González,S., Burton,O.T., Vázquez-Chávez,E., Cuche,C., Herit,F., Bouchet,J., Lasserre,R., Del Río-Iñiguez,I., Di Bartolo,V. and Alcover,A. (2017) Adenomatous polyposis coli defines Treg differentiation and anti-inflammatory function through microtubule-mediated NFAT localization. *Cell Rep.*, **21**, 181–194.

78. Pearce,E.L., Poffenberger,M.C., Chang,C.H. and Jones,R.G. (2013) Fueling immunity: insights into metabolism and lymphocyte function. *Science*, **342**, 1242454.

79. Buck,M.D., Sowell,R.T., Kaech,S.M. and Pearce,E.L. (2017) Metabolic instruction of immunity. *Cell*, **169**, 570–586.

80. Chang,C.H., Qiu,J., O'Sullivan,D., Buck,M.D., Noguchi,T., Curtis,J.D., Chen,Q., Gindin,M., Gubin,M.M., van der Windt,G.J. *et al.* (2015) Metabolic competition in the tumor microenvironment is a driver of cancer progression. *Cell*, **162**, 1229–1241.

81. Eales,K.L., Hollinshead,K.E. and Tennant,D.A. (2016) Hypoxia and metabolic adaptation of cancer cells. *Oncogenesis*, **5**, e190.

82. Fischer,K., Hoffmann,P., Voelkl,S., Meidenbauer,N., Ammer,J., Edinger,M., Gottfried,E., Schwarz,S., Rothe,G., Hoves,S. *et al.* (2007) Inhibitory effect of tumor cell-derived lactic acid on human T cells. *Blood*, **109**, 3812–3819.