EBioMedicine 41 (2019) 120–133

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Research paper

The prognostic landscape of interactive biological processes presents treatment responses in cancer

Bin He [a,1], Rui Gao [a,b,c,1], Dekang Lv [b], Yalu Wen [d], Luyao Song [b], Xi Wang [a], Suxia Lin [a], Qitao Huang [a], Ziqian Deng [b], Zifeng Wang [a], Min Yan [a], Feimeng Zheng [e], Eric W.-F. Lam [f], Keith W. Kelley [g], Zhiguang Li [b,*], Quentin Liu [a,b,**,2]

[a] State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, PR China
[b] Institute of Cancer Stem Cell, Dalian Medical University, Dalian 116044, PR China
[c] Department of Medical Oncology, The Seventh Affiliated Hospital, Sun Yat-sen University, Shenzhen 510275, PR China
[d] Department of Statistics, University of Auckland, New Zealand
[e] Department of Medical Oncology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510060, PR China
[f] Department of Surgery and Cancer, Imperial College London, Hammersmith Hospital Campus, London, W12 ONN, UK
[g] Laboratory of Immunophysiology, Department of Animal Sciences, College of ACES and Department of Pathology, College of Medicine, University of Illinois at Urbana-Champaign, Urbana, USA

ARTICLE INFO

Article history:
Received 1 October 2018
Received in revised form 15 January 2019
Accepted 31 January 2019
Available online 22 February 2019

Keywords:
Prognosis
Biological processes
Treatment response
Cell-proliferation
Immune processes

ABSTRACT

Background: Differential gene expression patterns are commonly used as biomarkers to predict treatment responses among heterogeneous tumors. However, the link between response biomarkers and treatment-targeting biological processes remain poorly understood. Here, we develop a prognosis-guided approach to establish the determinants of treatment response.
Methods: The prognoses of biological processes were evaluated by integrating the transcriptomes and clinical outcomes of ~26,000 cases across 39 malignancies. Gene-prognosis scores of 39 malignancies (GEO datasets) were used for examining the prognoses, and TCGA datasets were selected for validation. The Oncomine and GEO datasets were used to establish and validate transcriptional signatures for treatment responses.
Findings: The prognostic landscape of biological processes was established across 39 malignancies. Notably, the prognoses of biological processes varied among cancer types, and transcriptional features underlying these prognostic patterns distinguished response to treatment targeting specific biological process. Applying this metric, we found that low tumor proliferation rates predicted favorable prognosis, whereas elevated cellular stress response signatures signified resistance to anti-proliferation treatment. Moreover, while high immune activities were associated with favorable prognosis, enhanced lipid metabolism signatures distinguished immunotherapy resistant patients.
Interpretation: These findings between prognosis and treatment response provide further insights into patient stratification for precision treatments, providing opportunities for further experimental and clinical validations.
Fund: National Natural Science Foundation, Innovative Research Team in University of Ministry of Education of China, National Key Research and Development Program, Natural Science Foundation of Guangdong, Science and Technology Planning Project of Guangzhou, MRC, CRUK, Breast Cancer Now, Imperial ECMC, NIHR Imperial BRC and NIH.

1. Introduction

Heterogeneity among tumors, which leads to differential treatment responses, remains the leading challenge for effective treatment [1]. Extensive efforts have been devoted to distinguish treatment responders from non-responders. Individual genes or gene signatures are commonly used as the determinants of treatment response [2]. Yet, each gene may be involved in many distinct biological processes, rendering its biological function highly context-dependent [3,4]. Gene signatures

* Correspondence to: Zhiguang Li, Institute of Cancer Stem Cell, Dalian Medical University, Dalian 116044, PR China
** Correspondence to: Quentin Liu, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, PR China
E-mail addresses: zhiguangli2013@126.com (Z. Li), liuq9@mail.sysu.edu.cn (Q. Liu).
[1] Co-first author.
[2] First Corresponding Author.

https://doi.org/10.1016/j.ebiom.2019.01.064

Research in context

Evidence before this study

Differential gene expression patterns, defined by either experimental or data-mining efforts, are commonly used as biomarkers to predict responses in cancer patients. However, how these biomarkers are linked to treatment targeting biological processes remains poorly understood, limiting the identification of determinants for effective treatment. Treatment response is currently evaluated by the clinical outcomes of patients, a parameter that is also the primary readout for evaluating the potency of biological processes in cancer. Here we integrate the transcriptomic and clinical profiles from a variety of databases to link biomarkers and treatment targeting biological processes, thereby defining the determinants of response.

Added value of this study

Based on the hypothesis that a molecular context whereby a treatment targeting biological process correlate with an adverse clinical outcome informs features of treatment responders, we have developed an integrative strategy to establish the determinants of treatment responses. The prognoses of biological processes are defined based upon the transcriptome and clinical outcomes of distinct human malignancies. Based on the prognosis-treatment response link, we have identified principles of responses to treatments targeting both cell-proliferation and immune processes. Although high cell-proliferation rate of tumors predicted a poor prognosis, we found a favorable cell-proliferation prognosis within the context of a stress response signature, which indicates resistance to anti-proliferation treatment. Moreover, while high immune activity generally led to a favored prognosis, we showed that a lipid metabolism signature for adverse immune prognoses indicates immunotherapeutic resistance.

Implications of all the available evidence

Prognostic landscape of biological processes among cancer types provides a unique opportunity to predict vulnerabilities of distinct malignancies and to identify the molecular context responsible for treatment responses. The finding that response to treatment is dependent on the prognosis of the treatment targeting biological process provides a link for identifying treatment biomarkers. Our study proves the concept of using a prognoses guided approach to define the contextual determinants of treatment responses in cancer patients.

that contain a group of genes are more informative since they represent biological functions in defined contexts. Indeed, as genes in a signature are not functionally equal, the biological processes underlying these signatures make a major contribution to determining treatment responses [2]. Therefore, establishing gene signatures and the underlying biological processes is key to determining treatment response. However, how these response biomarkers are linked to treatment targeting biological processes remains poorly understood, which impedes the process in identifying determinants of effective treatment.

To link response biomarkers and treatment targeting biological processes, gene set-based approaches can be used [3]. Gene sets according to cellular response to chemical/genetic perturbations have previously been developed in the Molecular Signatures Database (MSigDB) [5]. As treatments act by targeting specific biological process, biological process gene sets (BPGSs) associated with particular clinical outcomes may reveal molecular features of treatment response. The prognoses of BPGSs among distinct cancer cohorts provide an opportunity to identify the underlying biological processes for treatment response. Prognostic evaluation in cancer datasets has hitherto been achieved by single-gene or signature-based approaches [2]. Recently, the PREdiction of Clinical Outcomes from Genomic (PRECOG) [6] and the Human Pathology Atlas [7] have provided a systematic approach to define the prognosis of individual genes, yet it is difficult to mine out the biological insights. Gene set analysis of gene-prognosis profiles could be an effective way to systematically evaluate the prognoses of BPGSs. The prognostic patterns of BPGSs among heterogeneous cancer types thus provide an approach to learn the determinant of prognosis as well as treatment response.

Here we report a gene-prognosis ranked Gene Set Enrichment Analysis (GSEA) [3] approach to evaluate the prognostic values of BPGSs in a variety of cancers. Based on the gene-prognosis information established from the tumor transcriptomes and clinical outcomes of ~26,000 patients, we systematically evaluated the prognostic values of BPGSs across 39 distinct cancer types. We found the prognoses of BPGSs vary among cancer types, and that the transcriptional contexts are responsible for the intertumor heterogeneity caused prognostic variations. Notably, we demonstrate that the transcriptional signatures and underlying biological processes associated with prognoses of BPGSs can distinguish treatment response for both anti-proliferation treatments and immunotherapy. The prognostic landscape of BPGSs reveals interactive biological processes for treatment responses, providing a novel approach for developing precision treatment strategies through available cancer datasets.

2. Materials and methods

2.1. TCGA, PRECOG z-scores, pathology atlas and gene expression data

TCGA datasets for 36 malignancies were downloaded from the BROAD GDAC Firehouse (https://gdac.broadinstitute.org/). Genome-wide gene-prognosis data were organized according to the meta-*z* score method [6]. Specifically, for each dataset, RNA-seq and clinical data were downloaded and matched. The association of each gene with survival outcomes was assessed via Cox proportional hazards regression using the 'coxph' function of the R 'survival' package. *P* values, *z*-score cox coefficients and hazard ratios with 95% confidence intervals were obtained for each gene. *P* values for each gene were transformed into meta-*z* scores. Weighted meta-*z*-scores were collapsed into a global meta-*z*-score for each gene as described (Table S1). In this manner, each gene is assigned a specific meta-*z*-score for each type of cancer, as well as a global meta-*z*-score across all cancer types. The *z*-score represents the capability of a gene to differentiate two subsets of patients with distinct survival outcomes. The 'meta-*z*-scores' for prognostic outcome of each gene in 39 malignancies were downloaded from the PRECOG project (http://precog.stanford.edu) [6]. Survival analysis results for all protein-coding genes in 17 TCGA major cancer types were downloaded from the pathology atlas database (http://www.proteinatlas.org/pathology) [7]. Gene expression values for patients with distinct sensitivity to therapies were downloaded from the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo). Gene expression levels (mRNA Expression *z*-Scores (RNA Seq V2 RSEM)) in the TCGA datasets were downloaded from the cBioPortal database (http://www.cbioportal.org/) [8,9].

2.2. Gene set enrichment analysis and single sample GSEA

Genes were ordered in a decreasing fashion according to pan-cancer (unweighted) or individual cancer prognostic *z*-scores. Pre-defined gene sets (eight major collections, H and C1-C7) were obtained from the Molecular Signatures Database (MSigDB) [5]. Chemical and genetic perturbations (CGP) gene sets containing "target" or "response" as a

keyword were selected to generate a ".gmt" gene set file for analysis. A customized gene set was defined as a set of genes in specific gene signature. Ranked lists were submitted to GSEA using the 'PreRanked' tool of GSEA software [3]. Interpretation of enrichment score (*ES*) to prognostic value is similar to the gene expression based GSEA. For a randomly distributed gene set, *ES* will be relatively small, but if it is concentrated at the top (adverse prognosis) or bottom (favorable prognosis) of the list, or otherwise non-randomly distributed, then the *ES* will be correspondingly high. For GSEA on CCLE, cell lines were grouped as sensitive or resistant according to their sensitivity to cell-proliferation targeting compounds. Enrichment of gene sets in both groups was determined. For GSEA of GEO datasets, patients were grouped as sensitive or resistant according to the authors' instructions, and then analyzed with candidate gene sets. Significantly enriched gene sets were defined using a False Discovery Rate (FDR) q-value <0.25 and a nominal P value <.05. All analyses were performed using GSEA v2.2.1 software with the preranked list and 1000 data permutations. Leading edge genes were defined by GSEA as genes in the gene set that appear in the ranked list at, or before the point where the running sum reaches its maximum deviation from zero, interpreted as the core of a gene set that accounts for the enrichment signal.

To perform single-sample gene set enrichment (ssGSEA), normalized gene expression data (downloaded from the CCLE portal) were submitted to the GenePattern platform. The ssGSEA Projection program was used to calculate separate enrichment scores for each pairing of a sample and gene set. Samples were normalized by rank, and the weighting exponent was set as 0.75. Enrichment scores for c5.bp.v6.0 (MSigDB) gene sets were subjected to Cluster 3.0 software and both gene sets and cell lines were clustered by average linkage. A clustered heat map was analyzed and visualized by TreeView.

2.3. Biomarker validation by PROGgene and SurvExpress

Candidate gene sets were submitted to the PROGgeneV2 [10] and SurvExpress online database [11]. Distinct types of cancer, including glioblastoma multiforme (TCGA), breast cancer (TCGA), colon cancer (GSE41258), lung adenocarcinoma (TCGA), and lung squamous cell carcinoma (TCGA) were analyzed using the SurvExpress. For the Cox Survival Analysis in the SurvExpress, two risk groups (high/low risk group) were defined by the median of submitted gene set expression, with patients categorized by survival time.

2.4. Hierarchical clustering

Normalized enrichment scores (NES) of each hallmark gene set for individual cancers (Table S3) were subjected to Cluster 3.0 software and both gene set and cancer type were clustered by average linkage. A clustered heat map was analyzed and visualized by TreeView. For hierarchical clustering of z-scores or gene expression values, the z-scores or gene expression values of candidate genes were subjected to the Cluster 3.0 software, clustered by average linkage according to their correlation or the Euclidean distance, and visualized by TreeView. Clustering of gene expression values was conducted by centering genes according to mean expression values (microarrays values or mRNA Expression z-Scores), clustered by average linkage and visualized by TreeView.

2.5. Validation in the TCPA dataset

Markers for enriched processes, such as cell proliferation and the immune system, were examined by the webpage survival analysis of The Cancer Proteome Atlas project [12] (TCPA, http://app1.bioinformatics.mdanderson.org/tcpa/_design/basic/index.html). Datasets of different cancer types were examined as follows: bladder urothelial carcinoma (TCGA), breast invasive carcinoma (TCGA), colon adenocarcinoma (TCGA), glioblastoma multiforme (TCGA), head and heck squamous cell carcinoma (TCGA), kidney renal clear cell carcinoma (TCGA), brain lower grade glioma (TCGA), lung adenocarcinoma (TCGA), lung squamous cell carcinoma (TCGA), endometrial carcinoma (MDACC), ovarian carcinoma (Japan), stomach adenocarcinoma (TCGA), uterine corpus endometrioid carcinoma (TCGA) and TCGA womens cancer (combined breast, uterine and ovarian cancer). Survival plots and log-rank P-values were downloaded for further analysis.

2.6. Oncomine based signature analysis

Differential expression analyses (Cancer vs Cancer Analysis) were performed in the Oncomine database [13]. Datasets in this study included: Astrocytoma (**ASTR**: *Bredel Brain 2, Cancer Res, 2005; Freije Brain, Cancer Res, 2004; Shai Brain, Oncogene, 2003; Sun Brain, Cancer Cell, 2006; vandenBoom Brain, Am J Pathol, 2003; Yamanaka Brain, Oncogene, 2006*), Glioblastoma multiforme (**GBM**: *Bredel Brain 2, Cancer Res, 2005; Freije Brain, Cancer Res, 2004; Liang Brain, Proc Natl Acad Sci, 2005; Nutt Brain, Cancer Res, 2003; Ramaswamy Multicancer, Proc Natl Acad Sci, 2001; Shai Brain, Oncogene, 2003; Sun Brain, Cancer Cell, 2006; vandenBoom Brain, Am J Pathol, 2003; Yamanaka Brain, Oncogene, 2006*), Glioma (**GLIO**: *Bredel Brain 2, Cancer Res, 2005; Freije Brain, Cancer Res, 2004; French Brain, Cancer Res, 2006; Liang Brain, Proc Natl Acad Sci, 2005; vandenBoom Brain, Am J Pathol, 2003)*, Medulloblastoma (**MEDU**: *Fattet Brain, J Pathol, 2009; Kool Brain, Plos One, 2008; Northcott Brain 3, Nature, 2012; Pomeroy Brain, Nature, 2002; Ramaswamy Multi-cancer, Proc Natl Acad Sci, 2001*), Neuroblastoma (**NEUB**: *Ramcrswomy Muiticancer, PNAS, 2001; Albino Brain, Cancer, 2003),* lung adenocarcinoma (**LUAD**, *Bittner Lung, Not Published, 2005; Chen Lung 3, N Engl J Med, 2007; Ding Lung, Nature, 2008; Garber Lung, Proc Natl Acad Sci, 2001; Hou Lung, Plos One 2010; Tomida Lung, Oncogene, 2004; Wigle Lung, Cancer Res, 2002; Yamagata Lung, Clin Cancer Res, 2003; Zhu Lung, J Clin Oncol, 2010*), lung squamous cell carcinoma (**LUSC,** *Bhattacharjee Lung, Proc Nat Acad Sci, 2001; Bild Lung, Nature, 2006; Bittner Lung, Not Published, 2005; Chen Lung 3, N Engl J Med, 2007; Garber Lung, Proc Natl Acad Sci, 2001; Hou Lung, Plos One, 2010; Kuner Lung, Lung Cancer, 2009; Lee Lung Clin Cancer Res, 2006; Rohrbeck Lung, J Transl Med, 2008; Tomida Lung, Oncogene, 2004; Wigle Lung, Cancer Res, 2002; Yamagata Lung, Clin Cancer Res, 2003; Zhu Lung, J Clin Oncol, 2010*) and small cell lung carcinoma (**SCLC**, *Chen Lung 3, N Engl J Med, 2007; Rohrbeck Lung, J Transl Med, 2008; Garber Lung, Proc Natl Acad Sci, 2001; Bhattacharjee Lung, Proc Natl Acad Sci,2001*).

Eight signatures for distinct types of cancers defined as the top high and low expressed genes for each cancer (compared with the other cancers of the same dataset) were downloaded for signature analysis. *Cell-cycle Con* signature (CycleC) was defined by overlapping up-regulated genes in **GLIO**, **ASTR**, **MEDU** and down-regulated genes in **GBM**; and up-regulated genes in **LUSC** overlapping down-regulated genes in **SCLC**, respectively. *Cell-cycle Rev* signature (CycleR) was defined by overlapping down-regulated genes in **GLIO**, **ASTR**, **MEDU** and up-regulated genes in **GBM**; and down-regulated genes in **LUSC** overlapping up-regulated genes in **SCLC**, respectively (illustrated in Fig. S3e, gene lists in Table S4). *Immune Con* signature (ImmuC) was defined by overlapping up-regulated genes in **NEUB**, **LUSC** and down-regulated genes in **MEDU**, **LUAD**, respectively. *Immune Rev* signature (ImmuR) was defined by overlapping down-regulated genes in **NEUB**, **LUSC** and up-regulated genes in **MEDU** and **LUAD**, respectively (illustrated in Fig. S5f, gene lists in Table S6).

2.7. Gene ontology annotation

Gene lists of interest were submitted to the MSigDB database [5]. The overlaps between gene list and pre-defined gene sets (BP: GO biological process) were computed. P value from the hypergeometric distribution for (k-1, K, N - K, n), where k is the number of genes in the intersection of the query set with a set from MSigDB, K is the number of genes in the set from MSigDB, N is the total number of gene universe

(all known human gene symbols) and n is the number of genes in the query set. FDR *q*-value is the false discovery rate analog of hypergeometric *P*-value after correction for multiple hypothesis testing according to Benjamini and Hochberg, and those gene sets with FDR $q < .05$ were selected.

2.8. Prognostic analysis of gene signatures

To assess the prognostic values of signatures, the normalized counts were scaled to a to *z*-score for each signature-gene, using the following function: *z*-score = ((normalized counts − mean counts across all samples)/(standard deviation of counts across all samples)). Subsequently, we used the average *z*-scores of all signature-genes to define a signature score for each single sample. Finally, Kaplan-Meier analysis was performed in all cases with stratification of risk sub-groups based on the median value of signature scores.

2.9. Immunet analysis

For gene set analysis, the top 200 favorable prognostic genes were analyzed using Immunet [14] gene set analysis (http://immunet.princeton.edu/geneset/). The immune networks of ImmuNet were used as features in a SVM classifier to generate gene predictions for the genes that were investigated. The resulting gene predictions and their probabilities are provided as Table S5. The proportion of genes in the submitted gene set (GS = 1) those not in the set (GS = 0) and the randomly selected negative set (GS = −1) with SVM-Prob >0.9 were analyzed.

2.10. Expression profile analysis in CCLE

Candidate genes were analyzed using the GENE-E tool of the CCLE database [15] (http://www.broadinstitute.org/ccle/data/browse-Analyses). To identify lymphocyte specific genes, expression data in the CCLE were downloaded from the cBioPortal [8,9] database. Unsupervised consensus clustering was performed with nonnegative matrix factorization [16] (NMF, v 0.20.6, default Brunet algorithm, 50 and 100 iterations for the rank survey and clustering runs, respectively). A preferred cluster result was selected by considering profiles of the cophenetic score for clustering solutions between 2 and 5 clusters. Profiles of genes or samples were generated by reordering with silhouette widths in the consensus clusters.

2.11. Functional immune genes

Genes in functional immune processes were identified using the Gene Expression Commons database (https://gexc.stanford.edu). Gene sets of the mouse hematopoiesis model were selected by functional immune key words (e.g., dendritic, T cell, NK cell, etc.) inquiries.

2.12. Vector constructions, lentivirus production and transfection

To knockdown PLAT, PLAU and SERPINE1 in human cancer cell lines, we constructed shRNA or scramble (shGFP) in pLKO-Tet-On vector (addgene #21915, RRID: Addgene_21915). Briefly, vector was digested by *Age*I and *Eco*RI (Thermo Scientific) for 15 min at 37 °C and recovered from agarose gel by DNA recovery kit (Tiangen). Annealed oligoes were ligated with digested vector with T4 DNA ligase (Thermo Scientific) for 2 h at room temperature. Competent Stbl3 *E. coli* cells were transformed with the constructs, selected and identified by Sanger sequencing. The lentiviral vector (10 μg) and the two packaging viral vectors, pMD.2G (3.3 μg) and pspPax2 (6.7 μg), were co-transfected into 293 T (RRID: CVCL_0045) cells using Lipo 2000 (ThermoFisher, #11668019, 1 μg/μL). The virus supernatant was collected at 48 h. MDA-MB-231 (bought from ATCC, RRID: CVCL_0062) cells were transfected with virus and selected with puromycin (Sigma-Aldrich, #P8833, 2 μg/ml) for six days.

2.13. Reverse transcription PCR (RT-PCR)

Cells were treated with doxycycline (Sigma-Aldrich, #D9891, 200 ng/ml) for 72 h. Total RNA was isolated using SuperPrep Cell Lysis kit (TOYOBO, #SCQ-101) as described by manufacturer and reverse transcribed with SuperPrep RT kit for qPCR (TOYOBO, #SCQ-101) for cDNA synthesis. Target genes were detected by real time PCR in CFX96 Touch™ according to manufacturer's protocol. Transcripts were quantified relative to the housekeeping gene, *GAPDH*. The probes used for this study were listed in Table S7.

2.14. Cell viability assay

Cells were maintained with doxycycline (200 ng/ml) for 72 h before vehicle (DMSO), MLN8237 (Selleckchem, #S1133, 200 nM) or Palbociclib (Sigma-Aldrich, #PZ0199-PD0332991, 500 nM) treatment. Cells were seeded in 96-well plates at a density of 3×10^3 cells/well, allowed to adhere overnight, and treated with vehicle (DMSO) or drug for 72 h, after which they were processed for cell viability using Cell Counting Kit-8 reagent (Dojindo Molecular Technologies, #CK04). Each condition was performed in replicates of eight wells.

2.15. Statistical analysis

Statistical analyses were performed using SPSS version 13.0 (SPSS Inc.). Kaplan–Meier statistic and log-rank tests were performed to estimate the relevance of candidate markers in overall and disease-free survival of patients. Multivariate analysis was performed using the Cox proportional hazards model. Other *P* values were determined using the two-tailed Student's *t*-test. Significant results were defined as $P < .05$.

3. Results

3.1. A prognosis-ranked GSEA for evaluating biological processes

To identify the interactive pattern of treatment targeting biological processes, we used cancer datasets to develop a stepwise bioinformatic approach (Fig. 1a). In this model, we proposed that response to a treatment in heterogeneous patients depends on two factors: 1) inactivation of the target and 2) prognostic contribution of the targeting biological process. An alternative treatment that successfully inactivate the target rescues type II, but not type I resistance. The mechanisms for type I resistance is key to stratifying patients and effective treatment. To this end, the prognostic values of BPGSs were used to define patient cohorts where BPGSs indicate favorable (BPGS-Fav) or adverse prognosis (BPGS-Adv). A comparative analysis between BPGS-Fav and BPGS-Adv patients was followed to establish gene signatures that distinguish BPGS-Fav from BPGS-Adv patients. Gene signatures and interacting biological processes were tested in treatment response datasets to determine their capacity to distinguish responding patients (Fig. 1b). The prognoses for individual genes (gene-prognosis scores) were examined by integrating gene expression patterns and clinical outcomes of cancer patients [6]. The 39 cancer types whose gene-prognosis scores generated by PRECOG from microarray gene expression data (Fig. 1b, **Methods**) were used for examining the prognoses of gene sets. In addition, 19 TCGA datasets (PRECOG matched cancer types, in-house generated gene-prognosis scores) were selected for validation (Table S1). To evaluate the prognostic values of gene sets, we performed an enrichment analysis of genes sets in the gene list decreasingly ranked by gene-prognosis scores, termed prognosis-ranked GSEA here. This prognosis-ranked GSEA defined two distinct prognostic classes indicating favorable and adverse prognosis, respectively (Fig. 1b) [3,4]. Pan-cancer gene-prognosis scores showed a high concordance between the TCGA and PRECOG datasets (Figs. S1a and S1b). Consistently, normalized enrichment score (NES) of the gene sets (hallmark in MSigDB

[5]) were highly correlated between matched TCGA and PRECOG datasets (Pearson $r > 0.35$, $P < .05$ in 13/19 cancer types, Fig. S1c). These results indicate the reproducibility between PRECOG and TCGA cancer datasets and the consistency of our evaluation approach.

To examine whether prognosis-ranked GSEA is capable of identifying biologically relevant gene sets, a pan-cancer prognosis-ranked GSEA was carried out in gene sets that were up- or down-regulated by specific chemical and genetic perturbations (CGP, Table S2). Notably, we found the opposite enrichment scores of -up/−down gene set pairs within either favorable or adverse prognoses (Fig. 1c). For instance, the prognostic NES for gene sets that were up- and down-regulated by serum in "CHANG_SERUM_RESP" were 2.798 and −2.331, respectively. The opposite NES values that define distinct prognostic association were prevalent among CGP gene sets. Among the 148 gene set pairs enriched in either the adverse phenotype (AdvP) or favorable phenotype (FavP), 83 (56%) exhibited the opposite NES for up- and down-regulated genes (gene sets with $P < .05$, FDR $q < 0.25$ in prognosis-ranked GSEA, Fig. 1d and Table S2). This finding is consistent with the notion that genes with opposing biological roles exhibit distinct prognostic contributions. Specifically, gene sets that were up-regulated by oncogenes such as *MYC* and *TERT* were enriched in AdvP, whereas the tumor suppressor *TP53* and *TFRC* up-regulated gene sets were enriched in FavP (Fig. 1d). Based upon these findings, we next explored the opposite enrichment of CGP gene set pairs in cancer type specific gene-prognosis scores by prognosis-ranked GSEA. Consistently, clear examples of opposite enrichments included Aplidin (−1.860 for up genes, 2.536 for down genes), CDK4i (−1.643 for up genes, 2.674 for down genes) and Cisplatin (2.004 for up genes, −1.456 for down genes) in Ewing sarcoma, as well as MYC (2.595 for up genes, −3.316 for down genes) and KDM1A (−1.660 for up genes, 1.407 for down genes) in melanoma (Fig. S1d). Enrichment patterns of paired gene sets that were up- or down-regulated by specific perturbations further confirmed the consistency and reproducibility of our approach.

To confirm the validity of prognostically enriched gene sets, we performed gene signature-based survival analyses in the PROGgeneV2 cancer datasets as previously reported [10]. In agreement with our findings that MYC up-regulated genes were enriched in AdvP (Fig.1d), MYC up-regulated genes were associated with poor survival outcomes in multiple tumor cohorts (Fig. S1e). Similar findings were obtained for EZH2 up-regulated genes and TFRC down-regulated genes (Figs. S1f and S1g). As shown in Fig.1d, genes down-regulated by Rapamycin, APLIDIN and GSK3 inhibitor SB216763 (GSK3i) were enriched in AdvP. Consistently, "Rapamycin response DN", "Response to APLIDIN DN" and "Response to GSK3 inhibitor SB216763 DN" were associated with adverse prognoses in the majority of tumor cohorts (log-rank $P < .05$, 41/42 datasets for Rapamycin, 43/46 datasets for GSK3i, Figs. 1e, 1f and S1h). In summary, prognostic enrichment of gene sets was consistent with their biological functions and prognostic associations, further confirming the capacity of our approach to evaluate the prognoses of biological processes.

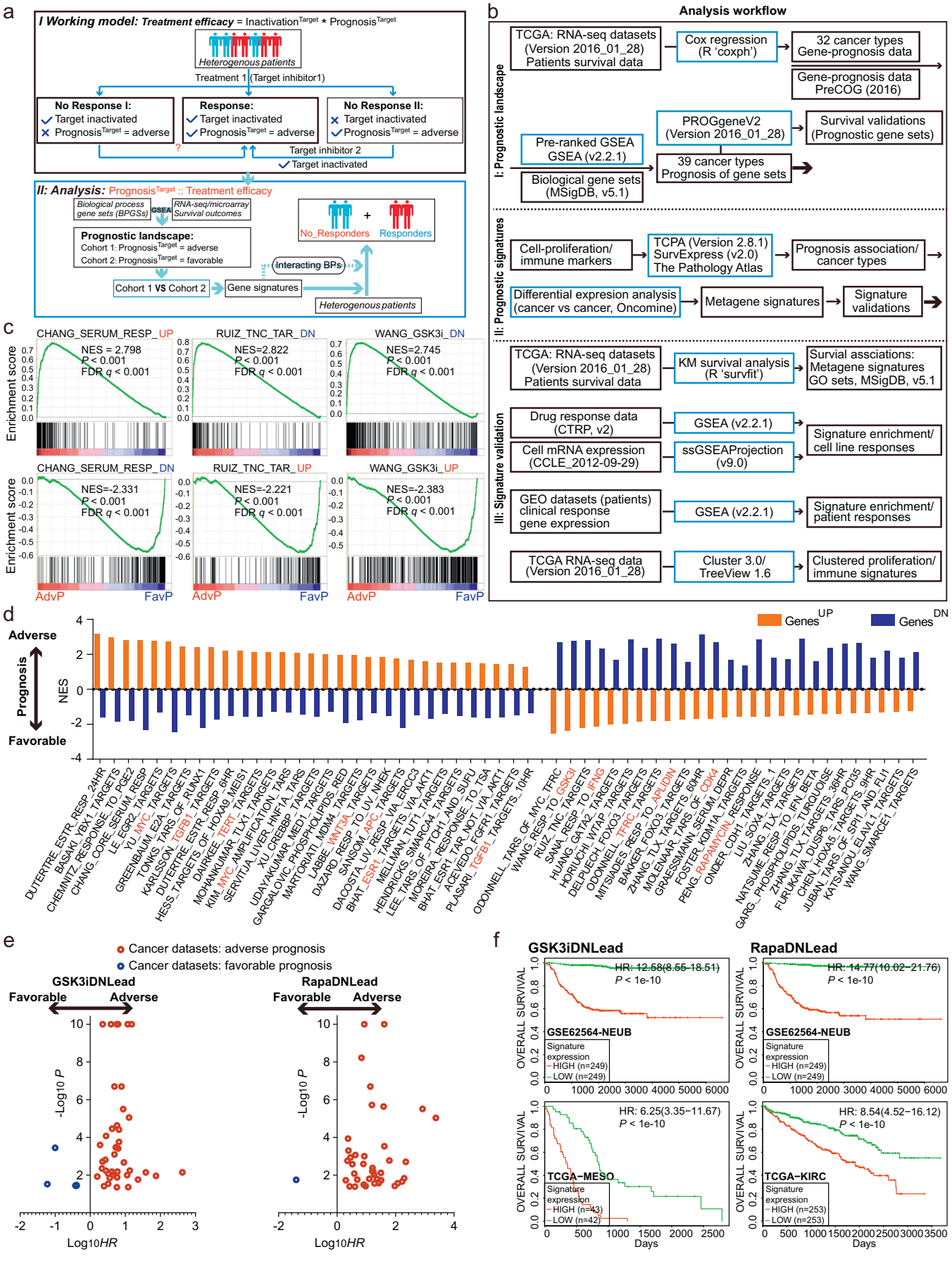
3.2. Prognoses of biological processes vary across cancer types

To assess the prognoses of key biological processes in cancer, we performed pan-cancer prognosis-ranked GSEA in 50 hallmark gene sets (MSigDB). Consistent with previous reports [6,7], cell-proliferation programs represented by cell-cycle related gene sets were found to be associated with adverse prognoses and enriched in AdvP, as 6 of the top 10 AdvP-enriched gene sets were cell-proliferation programs (gene sets with $P < .05$, FDR $q < 0.25$, Figs. 2a, left, S2a). In contrast, immune processes represented by immune-related gene sets were associated with favorable prognoses and enriched in FavP, as 6 of the top 10 FavP-enriched gene sets were immune processes (gene sets with $P < .05$, FDR $q < 0.25$, Figs. 2a, right, S2a). Next, the prognostic association of these processes was validated by a gene signature-based survival analysis in the PROGgeneV2 cancer datasets. Consistent with prognosis-ranked GSEA, leading edge genes of cell-proliferation programs (49 genes combining leading edge genes of E2F_TARGETS and MYC_TARGETS_V1 gene sets) were associated with adverse prognoses in 43 of 47 datasets (log-rank $P < .05$, Figs. 2b and 2d). In contrast, leading edge genes of immune processes (46 leading edge genes from INTERFERON_GAMMA_RESPONSE and INTERFERON_ALPHA_RESPONSE gene sets) forecasted favorable prognoses in 13 of 15 datasets (log-rank $P < .05$, Figs. 2c and 2d). In addition, "E2F targets" associated gene signature was coupled with adverse prognoses (Fig. S2b), while "IFNG response" and "IFNA response" gene signatures defined favorable prognoses (Fig. S2c). Collectively, these data demonstrate that cell-proliferation programs and immune processes are core processes of prognostic significance in cancer.

To characterize cancer type specific prognoses of key biological processes, we assessed the hallmark biological processes in 39 distinct PRECOG malignancies. Hierarchical clustering of prognostic NES revealed a prognostic landscape of 50 core biological processes in 39 malignancies (Fig. 2e and Table S3). The prognoses of individual biological processes exhibited considerable cancer type-dependent variations (Fig. S2d). Notably, in contrast to their AdvP enrichment in major cancer types, cell-proliferation programs were enriched in FavP in colon cancer, small cell lung carcinoma and glioblastoma (cancer types labeled in blue, Fig. 2e). Similarly, immune processes were classified as favorable processes in most cancer types, but were enriched in AdvP in bladder, pancreatic and brain cancers (cancer types labeled in red, Fig. 2e). Prognostic enrichments of the top biological processes were validated in matched TCGA datasets (lower panel, Fig. 2e). To confirm these findings, we assessed the prognoses of biological processes from the Gene Ontology Biological Process (GOBP gene sets from MSigDB) in 39 PRECOG malignancies (Table S3). We found that for the majority of cancer types, immune-associated GOBP were enriched in FavP, while those cell-proliferation related GOBP were enriched in AdvP (Figs. 2f and 2g). Consistently, cancer type dependent prognostic variations were observed in these gene sets (Figs. 2f and 2g). Indeed, ≤ 3 of 39 malignancies were significantly enriched in AdvP for immune-associated GOBPs (gene sets with $P < .05$, FDR $q < 0.25$, Fig. S2e) and ≤ 1 of 39 malignancies was significantly enriched in FavP for cell-proliferation associated GOBPs ($P < .05$, FDR $q < 0.25$, Fig. S2f). In addition, cancer type dependent prognostic variations were confirmed in the combined prognostic scores of core CGP gene sets (combined scores of CGP_UP and CGP_DN gene sets, Fig. S2g). Collectively, these findings establish the prognostic landscape of key biological processes and a distinct cancer type dependent property of prognostic association across a variety of malignancies (Fig. S2h).

3.3. Cell-proliferation prognoses point to treatment response

The cancer-type dependent prognostic variations provide an opportunity to examine whether transcriptional context of BPGS prognosis could distinguish treatment response. To determine the representative cancer types, we defined malignancies that displayed the consistent prognoses with pan-cancer NES as *Con* cancers, and those that showed reversing trends as *Rev* cancers. The analysis was focused on the prognostic pattern of the cell-proliferation program, which is a principal treatment target in a variety of cancers. Using data from the Human Pathology Atlas [7], we found that cancer-type dependent prognostic variations in the cell-proliferation program were consistent with prognostic NES (Fig. 2e). Specifically, the prognoses of individual cell-proliferation leading edge genes were adverse in 11/17 cancer types, while favorable prognoses were observed in malignancies such as glioma and colon cancer (Fig. 3a). These prognostic patterns were also validated in TCPA datasets [12], where cell-cycle markers (cyclin B1/E) showed adverse prognoses in several cancers, but favorable in colon cancer (Figs. S3a and S3b). Similarly, cell-proliferation leading edge genes displayed elevated expression in "high risk" in breast cancer (*Cell-cycle Con* cancer, Fig. S3c), but displayed lower expression in "high


a
I Working model: Treatment efficacy = Inactivation^Target * Prognosis^Target
Heterogenous patients
Treatment 1 (Target inhibitor1)
No Response I:
Target inactivated
Prognosis^Target = adverse
Response:
Target inactivated
Prognosis^Target = adverse
No Response II:
Target inactivated
Prognosis^Target = adverse
Target inhibitor 2
Target inactivated
II: Analysis: Prognosis^Target :: Treatment efficacy
Biological process gene sets (BPGSs)
GSEA
RNA-seq/microarray Survival outcomes
Prognostic landscape:
Cohort 1: Prognosis^Target = adverse
Cohort 2: Prognosis^Target = favorable
No_Responders
Responders
Interacting BPs
Cohort 1 VS Cohort 2
Gene signatures
Heterogenous patients
b
Analysis workflow
I: Prognostic landscape
TCGA: RNA-seq datasets (Version 2016_01_28) Patients survival data
Cox regression (R 'coxph')
32 cancer types Gene-prognosis data
Gene-prognosis data PreCOG (2016)
PROGgeneV2 (Version 2016_01_28)
Survival validations (Prognostic gene sets)
Pre-ranked GSEA GSEA (v2.2.1)
Biological gene sets (MSigDB, v5.1)
39 cancer types Prognosis of gene sets
II: Prognostic signatures
Cell-proliferation/ immune markers
TCPA (Version 2.8.1) SurvExpress (v2.0) The Pathology Atlas
Prognosis association/ cancer types
Differential expresion analysis (cancer vs cancer, Oncomine)
Metagene signatures
Signature validations
III: Signature validation
TCGA: RNA-seq datasets (Version 2016_01_28) Patients survival data
KM survival analysis (R 'survfit')
Survial asscations: Metagene signatures GO sets, MSigDB, v5.1
Drug response data (CTRP, v2)
Cell mRNA expression (CCLE_2012-09-29)
GSEA (v2.2.1)
ssGSEAProjection (v9.0)
Signature enrichment/ cell line responses
GEO datasets (patients) clinical response gene expression
GSEA (v2.2.1)
Signature enrichment/ patient responses
TCGA RNA-seq data (Version 2016_01_28)
Cluster 3.0/ TreeView 1.6
Clustered proliferation/ immune signatures
c
CHANG_SERUM_RESP_UP
NES = 2.798
P < 0.001
FDR q < 0.001
RUIZ_TNC_TAR_DN
NES=2.822
P < 0.001
FDR q < 0.001
WANG_GSK3i_DN
NES=2.745
P < 0.001
FDR q < 0.001
CHANG_SERUM_RESP_DN
NES=-2.331
P < 0.001
FDR q < 0.001
RUIZ_TNC_TAR_UP
NES=-2.221
P < 0.001
FDR q < 0.001
WANG_GSK3i_UP
NES=-2.383
P < 0.001
FDR q < 0.001
Enrichment score
AdvP
FavP
d
Adverse
Prognosis
Favorable
NES
Genes^UP
Genes^DN
DUTERTRE_ESTR_RESP_24HR
BASAKI_YBX1_TARGETS
CHEMMITZ_RESPONSE_TO_PGE2
CHANG_CORE_SERUM_RESP
LE_EGR2_TARGETS
YU_MYC_TARGETS
GREENBAUM_E2A_TARGETS
KARLSSON_TGFB1_TARGETS_UP
TONKS_TARGETS_OF_RUNX1
DUTERTRE_ESTR_RESP_6HR
HESS_TARGETS_OF_HOXA9_MEIS1
DAIRKEE_TERT_TARGETS
MOHANKUMAR_TLX1_TARGETS
KIM_MYC_AMPLIFICATION_TARS
SERVITJA_LIVER_HNF1A_TARS
XU_CREBBP_TARGETS
UDAYAKUMAR_MED1_TARGETS
GARGALOVIC_PHOSPHOLIPIDS_RED
MARTORIATI_MDM4_TARGETS
LABBE_WNT2A_TARGETS
DAZARD_RESP_TO_UV_NHEK
SANSOM_APC_TARGETS
DACOSTA_UV_RESP_VIA_ERCC3
BHAT_ESR1_TARGETS_VIA_AKT1
MELLMAN_TUT1_TARGETS
HENDRICKS_SMARCA4_TARGETS
LEE_TARS_OF_PTCH1_AND_SUFU
MOREIRA_RESPONSE_TO_TSA
BHAT_ESR1_TAR_NOT_VIA_AKT1
ACEVEDO_FGFR1_TARGETS
PLASARI_TGFB1_TARGETS_10HR
ODONNELL_TARS_OF_MYC_TFRC
WANG_RESP_TO_GSK3i
RUIZ_TNC_TARGETS
SANA_RESP_TO_IFNG
HUANG_GATA2_TARGETS
HORIUCHI_WTAP_TARGETS
DELPUECH_FOXO3_TARGETS
ODONNELL_TFRC_TARGETS
MITSIADES_RESP_TO_APLIDIN
BAKKER_FOXO3_TARGETS
ZHANG_TLX_TARGETS_60HR
MOLENAAR_TARS_OF_CDK4
GRAESSMANN_SERUM_DEPR
FOSTER_KDM1A_TARGETS
PENG_RAPAMYCIN_RESPONSE
ONDER_CDH1_TARGETS_1
LIU_SOX4_TARGETS
ZHANG_TLX_TARGETS
NATSUME_RESP_TO_IFN_BETA
GARG_PHOSPHOLIPIDS_TURQUOISE
ZHANG_TLX_TARGETS_36HR
FURUKAWA_DUSP6_TARGETS_PCI35
CHEN_HOXA5_TARGETS_9HR
JUBAN_TARS_OF_SPI1_AND_FLI1
KATSANOU_ELAVL1_TARGETS
WANG_SMARCE1_TARGETS
e
Cancer datasets: adverse prognosis
Cancer datasets: favorable prognosis
GSK3iDNLead
RapaDNLead
Favorable
Adverse
-Log10 P
Log10HR
f
GSK3iDNLead
RapaDNLead
HR: 12.58(8.55-18.51)
P < 1e-10
GSE62564-NEUB
Signature expression
HIGH (n=249)
LOW (n=249)
HR: 14.77(10.02-21.76)
P < 1e-10
GSE62564-NEUB
Signature expression
HIGH (n=249)
LOW (n=249)
HR: 6.25(3.35-11.67)
P < 1e-10
TCGA-MESO
Signature expression
HIGH (n=43)
LOW (n=42)
HR: 8.54(4.52-16.12)
P < 1e-10
TCGA-KIRC
Signature expression
HIGH (n=253)
LOW (n=253)
OVERALL SURVIVAL
Days

risk" colon cancer compared with the "low risk" subgroups (*Cell-cycle Rev* cancer, Fig. S3d). Therefore, colon and brain cancers are representative *Rev* cancers for the cell-proliferation program.

To characterize contextual determinants of this variation, we established the Prognosis Variation Signature (PVS) according to the six representative cancer types (GLIO, GBM, ASTR, MEDU, SCLC and LUSC). These four pairs of cell-proliferation representative cancer types (GLIO-GBM, ASTR-GBM, MEDU-GBM and SCLC-LUSC) shared similar prognostic patterns for 50 hallmark gene sets except cell-proliferation programs (Fig. 2e). The top 200 up-regulated and down-regulated genes within each cancer type were analyzed in the Oncomine database [13] to establish the PVS (see **Methods**, Fig. S3e and Table S4). Genes that were up-regulated in cancer types with favorable cell-proliferation prognoses were defined as *Cell-cycle Rev* signature (CycleR), and genes up-regulated in cancer types with adverse cell-proliferation prognoses were defined as *Cell-cycle Con* signature (CycleC). GO annotation indicated a significant enrichment of stress response, angiogenesis and regeneration processes in CycleR, whereas synapse development was prominent in CycleC (Fig. S3f). To test whether CycleR stratified patients displayed distinct cell-proliferation prognoses, survival analyses of cell cycle genes (GO) were performed in lung adenocarcinoma subsets classified by CycleR. In agreement with our previous results, the cell-proliferation programs were associated with a favorable prognosis in $CycleR^{high}$, but not in the $CycleR^{low}$ patients (log-rank $P = .000935$, Fig. 3b).

Next, we tested whether the PVS and underlying biological processes for the cell-proliferation program indicate response to cell-proliferation targeting perturbations. To this end, cancer cell lines that were sensitive (15/823 cell lines) or resistant (15/823 cell lines) to cell-proliferation targeted reagents were defined according to treatment sensitivity data in the Cancer Therapeutics Response Portal [17,18]. GSEA in the transcriptomic profiles (Cancer Cell Line Encyclopedia, CCLE) of representative cell lines uncovered a significant enrichment of CycleR and epithelial mesenchymal transition (EMT) gene sets in the treatment resistant phenotype (Fig. S3g). Using single sample GSEA (ssGSEA) [19] of GOBP gene sets (MsigDB) in these cell lines, we found that enrichment scores of GOBP gene sets distinguished sensitive from resistant cells. Consistently, stress response, angiogenesis and regeneration processes were significantly enriched in the resistant phenotype (Fig. 3c). To confirm the functional roles of stress response genes in treatment resistance, we validated three of these genes (*PLAT*, *PLAU* and *SERPINE1*) in MDA-MB-231 cells in response to MLN8237 (AURKA inhibitor) and Palbociclib (CDK4/6 inhibitor). Consistently, cells were more sensitive to both inhibitors after knocking-down *PLAT*, *PLAU* or *SERPINE1* by conditional shRNA (Figs. 3d and S3h). These findings suggest that PVS for the cell-proliferation program indicate response to cell-proliferation targeting treatments.

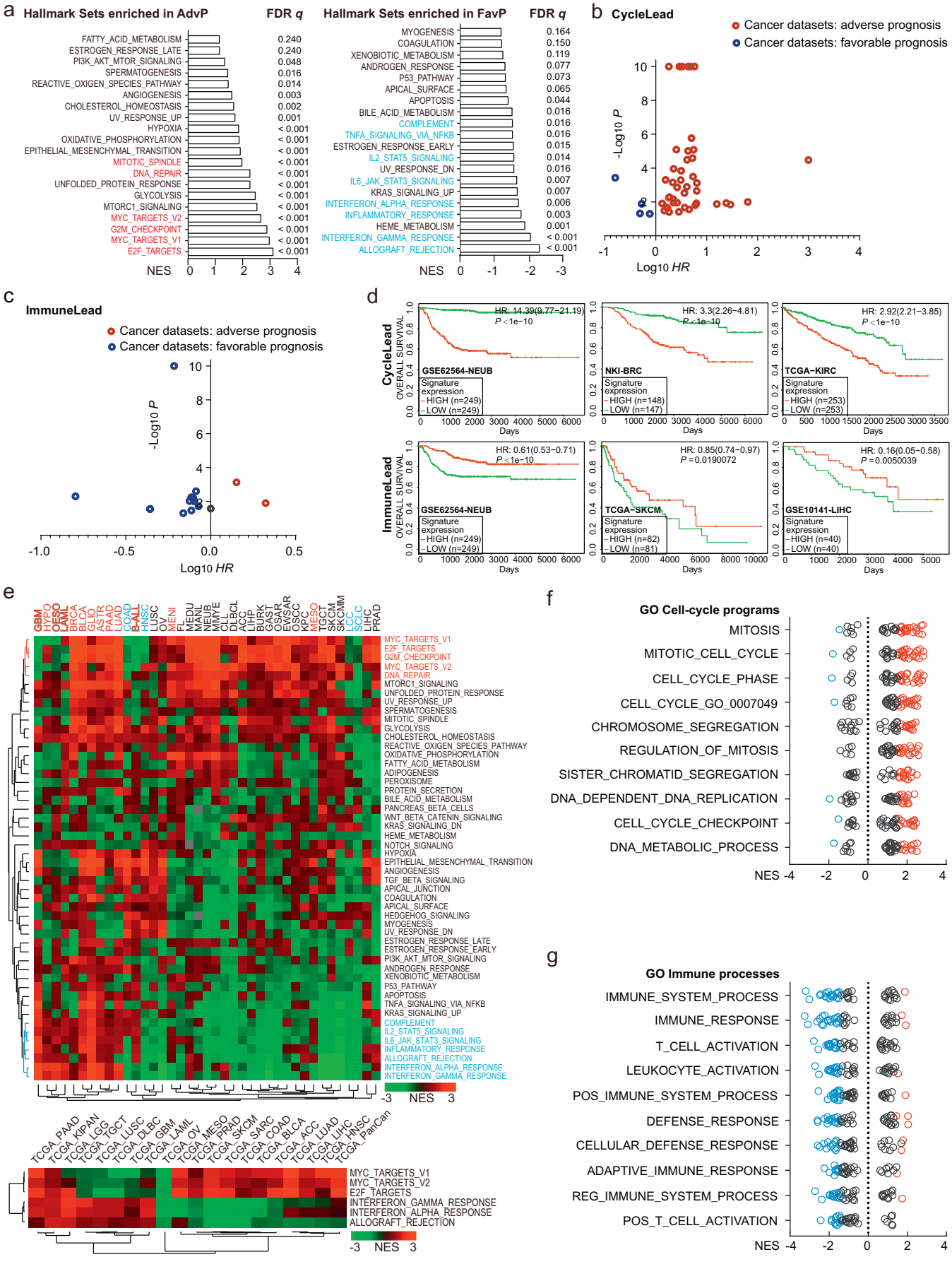
To determine whether PVS distinguish responders from non-responders in patients, we inspected the transcriptional profiles of pretreated patients in eight anti-proliferation treatment-response cancer datasets (GSE106977, GSE19697, GSE41998, GSE50948, GSE5820, GSE6667, GSE66999, GSE76360) [20] [21]. GSEA in the transcriptional profiles confirmed that gene sets including CycleR, EMT, hypoxia and angiogenesis were enriched in the non-responders (Fig. 3e). Indeed, CycleR was upregulated in non-responders (NES < 0, Fig. 3f), and was significantly enriched in non-responders in three of them (GSEA nominal $P < .05$ in GSE5820, GSE76360 and GSE106977, Fig. 3f). In addition, stress response CycleR genes distinguished resistant patients from sensitive ones in the treatment-response dataset [22] (GSE5820 and GSE32962, Fig. S3i). Therefore, transcriptional signatures that were defined by prognostic variation could distinguish response to anti-proliferation treatments.

Co-enriched transcriptional signatures of tumors that distinguish tumor subsets have been proposed to facilitate patient stratification [23]. To determine whether the cell-proliferation PVS is co-enriched in tumors, we explored the gene expression patterns in other cancer datasets [8,9] and found that PVS were co-enriched in tumor subsets in multiple malignancies (Figs. 3g and 3h). Consistent with the prognostic NES patterns, the majority of glioblastoma (*Cell-cycle Rev* cancer) patients expressed low levels of CycleC genes, and a larger proportion of pancreatic cancer (*Cell-cycle Con* cancer) patients were found to be in the $CycleC^{high}$ group (Fig. 3g). Subsets of CycleR co-enriched tumors were observed in glioblastoma, pancreatic cancer and additional tumor cohorts (Figs. 3h and S3j). To determine whether CycleC and CycleR are associated with established molecular subtyping, we examined the correlation of CycleC and CycleR with molecular subtypes and *IDH1* mutation in the TCGA glioblastoma dataset [24]. Interestingly, CycleC and CycleR showed a reverse pattern of correlation with glioblastoma subtypes. CycleC was expressed at higher levels in neural and proneural and lower in mesenchymal subtype, whereas CycleR was enriched in mesenchymal and lower in neural and proneural subtypes (Figs. 3i and S3k). These results are consistent with previous findings that mesenchymal subtype is associated with treatment resistance and relapse in glioblastoma [25]. In addition, CycleC was enriched *IDH1* mutant subset, while CycleR was enriched in *IDH1* wild-type subtype (Figs. 3j and S3l), consistent with previous report that *IDH1* mutant glioblastoma patients were more sensitive to radiochemotherapy [26]. Thus, transcriptional signatures that were associated with both cell-proliferation prognosis and treatment response stratified patient subsets in multiple tumors.

3.4. Immune prognoses exhibit intrinsic intertumor variation

We next examined this prognosis-guided strategy in immunotherapy, which is at present hampered by their low response rates and a

Fig. 1. GSEA defines the prognoses of biological processes (a) Overview of the working model. Response to a treatment in heterogeneous patients depends on two factors according to the treatment-response model: 1) inactivation of the target and 2) prognostic contribution of the target. An alternative target inactivating treatment rescues type II, but not type I resistance, and the mechanisms for type I resistance is key to stratifying patients and effective treatment. The prognoses of treatment targeting biological process in distinct cancer types were evaluated by integrating transcriptomic and clinical outcomes in datasets. Gene signatures and interacting biological processes that determine prognostic variation and treatment response were established to distinguish treatment responses. PrognosisTarget, the prognostic score of treatment targeting processes. InactivationTarget, the ratio of inactivated target by specific treatment. (b) A detailed analysis workflow. Black boxes indicate input or output data and blue boxes analysis method. Based on the gene sets defined by biological processes and treatment interventions available in the MSigDB database, a method was developed that relied on the genome-wide gene-prognosis metadata (PRECOG (GEO microarray datasets) or TCGA (in-house generated RNA-seq datasets) gene-prognosis data) and GSEA algorithm. This prognosis-ranked GSEA provided an estimate of the prognostic values of biological gene sets. Metagene signatures were developed according to cancer type dependent prognostic and transcriptional features, and validated by a variety of bioinformatic resources. (c) Representative enrichment plots for chemical and genetic perturbations (CGP) gene set pairs (_UP: upregulated by perturbation, _DN: downregulated by perturbation) in prognosis-ranked GSEA (pan-cancer prognostic z-score). NES, normalized enrichment score. AdvP, Adverse Prognostic Phenotype; FavP, Favorable Prognostic Phenotype. FDR, false discovery rate. FDR and Nominal P values were defined by GSEA software. (d) Prognostic NES of 57 paired CGP gene sets in PRECOG prognosis-ranked GSEA (pan-cancer prognostic z-score). Gene sets of known oncogenes or treatment perturbations are labeled in red. Orange bars indicate "UP" (upregulated by perturbation) gene sets and blue bars indicate "DN" (downregulated by perturbation) gene sets. (e) Plots indicating –Log10HR (Hazard Ratio) and –Log10P of GSK3iDNLead and RapaDNLead genes in the TCGA and GEO datasets. Leading edge genes were defined by GSEA, which was determine as genes in the gene set that appear in the ranked list at, or before the point where the running sum reaches its maximum deviation from zero, interpreted as the core of a gene set that accounts for the enrichment signal. P values (log-rank test) and HR were determined between the two patient groups stratified by median level of expression of leading edge genes using Kaplan Meier survival analysis. Each point stands for an independent dataset. Red and blue dots indicate adverse and favorable prognosis, respectively. (f) Representative curves for Kaplan Meier analyses of patients with low (green) or high (red) GSK3iDNLead (left) RapaDNLead (right) expression (stratified by median expression, n = 249/249 for GSE62564-NEUB, n = 43/42 for TCGA-MESO and n = 253/253 for TCGA-KIRC) in the cancer datasets. P values were determined using the log-rank test.


a
Hallmark Sets enriched in AdvP
FDR q
FATTY_ACID_METABOLISM 0.240
ESTROGEN_RESPONSE_LATE 0.240
PI3K_AKT_MTOR_SIGNALING 0.048
SPERMATOGENESIS 0.016
REACTIVE_OXIGEN_SPECIES_PATHWAY 0.014
ANGIOGENESIS 0.003
CHOLESTEROL_HOMEOSTASIS 0.002
UV_RESPONSE_UP 0.001
HYPOXIA < 0.001
OXIDATIVE_PHOSPHORYLATION < 0.001
EPITHELIAL_MESENCHYMAL_TRANSITION < 0.001
MITOTIC_SPINDLE < 0.001
DNA_REPAIR < 0.001
UNFOLDED_PROTEIN_RESPONSE < 0.001
GLYCOLYSIS < 0.001
MTORC1_SIGNALING < 0.001
MYC_TARGETS_V2 < 0.001
G2M_CHECKPOINT < 0.001
MYC_TARGETS_V1 < 0.001
E2F_TARGETS < 0.001
NES 0 1 2 3 4
Hallmark Sets enriched in FavP
FDR q
MYOGENESIS 0.164
COAGULATION 0.150
XENOBIOTIC_METABOLISM 0.119
ANDROGEN_RESPONSE 0.077
P53_PATHWAY 0.073
APICAL_SURFACE 0.065
APOPTOSIS 0.044
BILE_ACID_METABOLISM 0.016
COMPLEMENT 0.016
TNFA_SIGNALING_VIA_NFKB 0.016
ESTROGEN_RESPONSE_EARLY 0.015
IL2_STAT5_SIGNALING 0.014
UV_RESPONSE_DN 0.016
IL6_JAK_STAT3_SIGNALING 0.007
KRAS_SIGNALING_UP 0.007
INTERFERON_ALPHA_RESPONSE 0.006
INFLAMMATORY_RESPONSE 0.003
HEME_METABOLISM 0.001
INTERFERON_GAMMA_RESPONSE < 0.001
ALLOGRAFT_REJECTION < 0.001
NES 0 -1 -2 -3
b
CycleLead
Cancer datasets: adverse prognosis
Cancer datasets: favorable prognosis
-Log10 P
Log10 HR
c
ImmuneLead
Cancer datasets: adverse prognosis
Cancer datasets: favorable prognosis
-Log10 P
Log10 HR
d
CycleLead
ImmuneLead
OVERALL SURVIVAL
HR: 14.39(9.77−21.19) P <1e−10
GSE62564-NEUB
HR: 3.3(2.26−4.81) P <1e−10
NKI-BRC
HR: 2.92(2.21−3.85) P <1e−10
TCGA−KIRC
HR: 0.61(0.53−0.71) P <1e−10
GSE62564-NEUB
HR: 0.85(0.74−0.97) P =0.0190072
TCGA−SKCM
HR: 0.16(0.05−0.58) P =0.0050039
GSE10141-LIHC
Signature expression
HIGH
LOW
Days
e
MYC_TARGETS_V1
E2F_TARGETS
G2M_CHECKPOINT
MYC_TARGETS_V2
DNA_REPAIR
MTORC1_SIGNALING
UNFOLDED_PROTEIN_RESPONSE
UV_RESPONSE_UP
SPERMATOGENESIS
MITOTIC_SPINDLE
GLYCOLYSIS
CHOLESTEROL_HOMEOSTASIS
REACTIVE_OXIGEN_SPECIES_PATHWAY
OXIDATIVE_PHOSPHORYLATION
FATTY_ACID_METABOLISM
ADIPOGENESIS
PEROXISOME
PROTEIN_SECRETION
BILE_ACID_METABOLISM
PANCREAS_BETA_CELLS
WNT_BETA_CATENIN_SIGNALING
KRAS_SIGNALING_DN
HEME_METABOLISM
NOTCH_SIGNALING
HYPOXIA
EPITHELIAL_MESENCHYMAL_TRANSITION
ANGIOGENESIS
TGF_BETA_SIGNALING
APICAL_JUNCTION
COAGULATION
APICAL_SURFACE
HEDGEHOG_SIGNALING
MYOGENESIS
UV_RESPONSE_DN
ESTROGEN_RESPONSE_LATE
ESTROGEN_RESPONSE_EARLY
PI3K_AKT_MTOR_SIGNALING
ANDROGEN_RESPONSE
XENOBIOTIC_METABOLISM
P53_PATHWAY
APOPTOSIS
TNFA_SIGNALING_VIA_NFKB
KRAS_SIGNALING_UP
COMPLEMENT
IL2_STAT5_SIGNALING
IL6_JAK_STAT3_SIGNALING
INFLAMMATORY_RESPONSE
ALLOGRAFT_REJECTION
INTERFERON_ALPHA_RESPONSE
INTERFERON_GAMMA_RESPONSE
-3 NES 3
TCGA_PAAD TCGA_KIPAN TCGA_LGG TCGA_TGCT TCGA_LUSC TCGA_DLBC TCGA_GBM TCGA_LAML TCGA_OV TCGA_MESO TCGA_PRAD TCGA_SKCM TCGA_SARC TCGA_COAD TCGA_BLCA TCGA_ACC TCGA_LUAD TCGA_LIHC TCGA_HNSC TCGA_PanCan
MYC_TARGETS_V1
MYC_TARGETS_V2
E2F_TARGETS
INTERFERON_GAMMA_RESPONSE
INTERFERON_ALPHA_RESPONSE
ALLOGRAFT_REJECTION
-3 NES 3
f
GO Cell-cycle programs
MITOSIS
MITOTIC_CELL_CYCLE
CELL_CYCLE_PHASE
CELL_CYCLE_GO_0007049
CHROMOSOME_SEGREGATION
REGULATION_OF_MITOSIS
SISTER_CHROMATID_SEGREGATION
DNA_DEPENDENT_DNA_REPLICATION
CELL_CYCLE_CHECKPOINT
DNA_METABOLIC_PROCESS
NES -4 -2 0 2 4
g
GO Immune processes
IMMUNE_SYSTEM_PROCESS
IMMUNE_RESPONSE
T_CELL_ACTIVATION
LEUKOCYTE_ACTIVATION
POS_IMMUNE_SYSTEM_PROCESS
DEFENSE_RESPONSE
CELLULAR_DEFENSE_RESPONSE
ADAPTIVE_IMMUNE_RESPONSE
REG_IMMUNE_SYSTEM_PROCESS
POS_T_CELL_ACTIVATION
NES -4 -2 0 2 4

lack of treatment biomarkers. We found that immune processes were the highest ranked biological processes associated with favorable prognoses (higher gene expression in patients with better survival outcomes). Core immune gene sets, such as antigen presentation, lymphocyte development and activation, T cell response and natural killer (NK) cell-mediated cytotoxicity were enriched in the FavP ($P < .05$ in GSEA, Fig. S4a). In addition, Gene Ontology (GO) annotation of the top 200 favorable genes (PRECOG pan-cancer) demonstrated that immune processes were highly enriched (Fig. S4b). These top favorable genes were found to be involved in well-defined immune networks (determined by ImmuNet [14]), such as those involved in antigen presentation, lymphocyte activation and NK cell mediated killing pathways (Fig. S4c and Table S5). Pan-cancer prognosis-ranked GSEA for gene sets representing 28 immune cell subpopulations and 10 immune cell processes indicated that the prognostic NES between TCGA and PRECOG were positively correlated (Pearson $r = 0.7295$, $P < .0001$, Figs. S4d and S4e). Specifically, dendritic cells, CD8 T cells, B cells and NK cells were associated with favorable prognoses, whereas macrophage and regulatory T cells indicated adverse prognoses (Fig. S4d).

We then assessed the cancer type specific prognoses of these immune processes by prognosis-ranked GSEA in cancer type specific gene-prognosis scores. Hierarchical clustering of prognostic NES for individual cancer types confirmed that dendritic cells, CD8 T cells, B cells and NK cells were associated with favorable prognoses in the majority of the 39 malignancies (Fig. S4f). However, adverse prognoses were observed in a subset of cancers, including pancreatic cancer, glioma and glioblastoma (red labels, Fig. S4f). TCGA datasets confirmed the favorable prognosis of immune processes in the majority of cancers, as well as the variations across different types of cancers (Fig. S4g). These results reveal the overall favorable prognoses of core immune processes and cancer-type dependent variations.

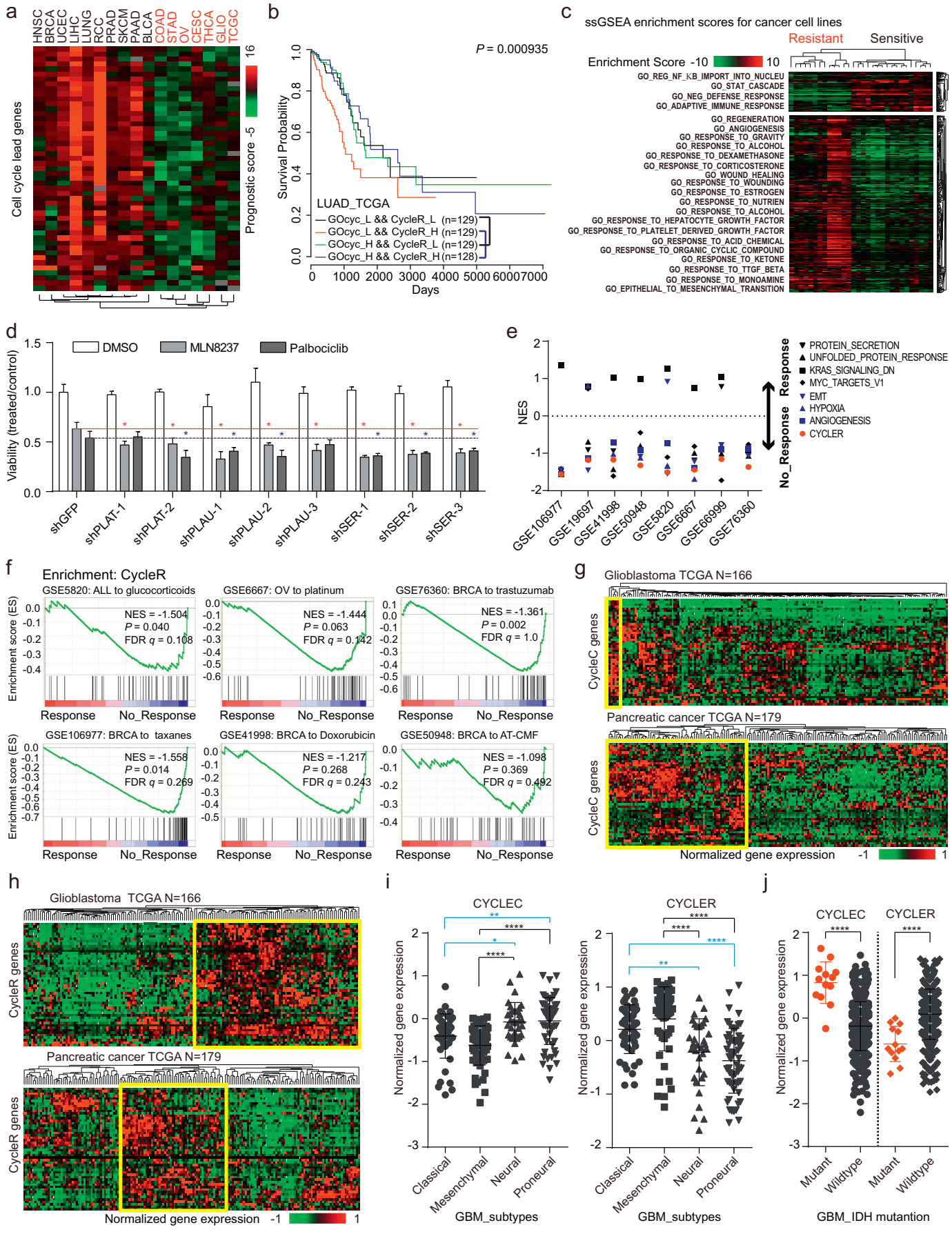
As inflammatory signaling in cancer cells promotes cancer progression [27], inflammatory genes from immune gene sets that are expressed in cancer cells might confer an adverse immune prognosis. To inspect this possibility, we defined lymphocyte specific genes by expression-based non-negative matrix factorization (NMF) clustering of immune leading edge genes in the CCLE dataset (Cluster 2, left panel, Fig. S5a). A notable portion of lymphocyte-specific leading edge genes exhibited favorable prognoses in malignancies with adverse immune prognostic NES, including breast cancer and lung adenocarcinoma, but not in cancers such as glioma (right panel, Fig. S5a). We examined the Gene Enrichment Profiler database, performed NMF clustering and collected a subset of 32 Lymphocyte Specific Functional genes (LSF genes, Fig. S5b). Similar to the prognoses of immune cell gene sets (Figs. S4f and S4g), the prognoses of favorable LSF gene sets again varied among cancer types (Fig. 4a). The prognostic patterns of immune process were substantiated in representative cancer datasets from the SurvExpress [11], with immune leading edge genes indicating "low risk" in lung squamous cell carcinoma patients (*Immune Con* cancer, Fig. S5c), but "high risk" in glioblastoma (*Immune Rev* cancer, Fig. S5d). Moreover, survival outcomes of individual immune leading edge genes were favorable in 13/17 cancer types whereas adverse prognoses were observed in malignancies such as glioma and pancreatic cancer according to datasets from the Human Pathology Atlas [7] (Fig. S5e). Therefore, non-hematopoietic specific gene sets partially contribute to adverse immune prognoses in certain situations, but not in malignancies such as brain cancers.

3.5. Immune interactive processes for treatment response

To uncover key factors that were responsible for immune prognostic variations, we established the immune PVS by combining transcriptional signatures of tumors from representative cancer types (MEDU, NEUB, LUSC and LUAD) in the Oncomine database [13]. These two pairs of immune representative cancer types (MEDU-NEUB and LUSC-LUAD) exhibited similar prognostic patterns for the 50 hallmark gene sets, with the exception of immune processes (Fig. 2e). The top 300 up-regulated and down-regulated genes of representative cancer types were combined to generate the gene signatures associated with prognoses (Fig. S5f and Table S6). Genes up-regulated in malignancies with adverse immune prognoses were defined as *Immune Rev* signature (ImmuR), and genes up-regulated in malignancies with favorable immune prognoses as *Immune Con* signature (ImmuC). GO annotation of these signatures indicated a significant enrichment of lipid metabolism, regulatory T cell (Treg) and macrophage processes in the ImmuR, while T-cell activation process was in the ImmuC (Fig. S5g). To test whether patients stratified by PVS displayed distinct prognoses of immune processes, we performed survival analysis of immune genes (GO) in patient subsets classified by the ImmuC and ImmuR signatures, respectively. Consistently, the immune processes were associated with adverse prognoses in the $ImmuR^{high}$ patients (Fig. 4b), whereas the immune processes were correlated with favorable prognoses in the $ImmuC^{high}$ patients (Fig. 4c).

Encouraged by results from the cell-proliferation program, we speculated that immune PVS may inform treatment response in immunotherapies. To test this possibility, we performed ssGSEA on transcriptional profiles of individual patients treated with anti-PD-1 immunotherapy (GSE78220 [23], GSE67501 [28], GSE91061 and Aa5951 [29]). In agreement with our hypothesis, enrichment scores for ImmuC/ImmuR (combined ssGSEA score for ImmuC and ImmuR) were significantly higher in responders than non-responders (Student's *t*-test, $P = .097$, 0.040, 0.008 and 0.0201, respectively, Fig. 4d). We carried out GSEA of ImmuC, ImmuR and hallmark gene sets in grouped patients (responders vs non-responders) in each dataset. Consistent with the report that EMT is associated with immunotherapy resistance, EMT was found to be enriched in the immunotherapy non-responder phenotype (GSE78220, $P = .004$ in GSEA, Fig. S5h) [23]. The ImmuR leading edge genes, especially *MGLL* that is involved in membrane lipid metabolism processes, were up-regulated in immunotherapy-resistant patients (Fig. 4e). Notably, ImmuR was enriched in the immunotherapy non-responder phenotype ($P = .040$, 0.046 and 0.081 in GSEA, upper panel, Fig. 4f), whereas ImmuC was enriched in the responder population ($P = .108$, 0.053 and 0.155 in GSEA, lower panel, Fig. 4f). Indeed, among hallmark gene sets, ImmuC, ImmuR and lipid metabolic

Fig. 2. The prognostic landscape of biological processes in cancer (a) The NES of hallmark biological processes in prognosis-ranked GSEA (pan-cancer prognostic z-score). NES values were generated for all of the 50 GSEA hallmark processes using the decreasingly ranked pan-cancer prognostic z-scores. Gene sets enriched in adverse (NES > 0, left) and favorable (NES < 0, right) prognoses were ranked by NES. Red: cell-proliferation related gene sets. Blue: immune gene sets. FDR and Nominal P values were defined by GSEA software. (b) and (c) Plots indicating –Log10HR and –Log10P of the cell-proliferation programs (b) and immune processes (c) in the TCGA and GEO datasets. CycleLead here includes the 49 leading edge genes from the E2F_TARGETS and MYC_TARGET_V1 whereas the ImmuneLead here includes 46 leading edge genes from the INTERFERON_GAMMA and INTERFERON_ALPHA gene sets. P values and hazard ratios were determined by Kaplan Meier survival analyses of leading edge genes (patients were stratified by median gene expression). Each point stands for an independent dataset. Red and blue dots indicate adverse and favorable prognoses, respectively. (d) Representative curves for Kaplan Meier analyses of patients with low (green) or high (red) CycleLead (upper) or ImmuneLead (lower) expression in the GEO datasets (patients were stratified by median gene expression, n = 249/249 for GSE62564-NEUB, n = 148/147 for NKI-BRC, n = 253/253 for TCGA-KIRC, n = 82/81 for TCGA-SKCM and n = 40/40 for GSE10141-LIHC). P values were determined using the log-rank test. (e) Upper panel: Hierarchical clustering of prognostic NES for 50 hallmark gene sets in 39 malignancies. Abbreviations for cancer types are listed in Table S1. NES for the 50 hallmark gene sets were calculated in the GSEA software according to the prognostic z-scores of each cancer type that were download from the PRECOG website. Cancers with cell-proliferation programs conferring a favorable prognosis were marked in blue whereas those with immune processes indicating poor prognosis were marked in red. Lower panel: Hierarchical clustering of NES for 6 representative gene sets in 19 TCGA datasets whose prognostic z-scores were generated in this study (See Methods). Sidebar, NES index for adverse (red) and favorable (green) prognoses. (f) and (g) Plots depicting prognostic NES of immune (f) and cell-proliferation (g) related GO processes assessed in 39 individual cancer types. Significant enrichments (P <.05, FDR q <0.25) for each gene set is labeled in red (AdvP) and blue (FavP), respectively.


a
HNSC BRCA UCEC LIHC LUNG RCC PRAD SKCM PAAD BLCA COAD STAD OV CESC THCA GLIO TCGC
Cell cycle lead genes
Prognostic score -5 16
b
P = 0.000935
Survival Probability
LUAD_TCGA
GOcyc_L && CycleR_L (n=129)
GOcyc_L && CycleR_H (n=129)
GOcyc_H && CycleR_L (n=129)
GOcyc_H && CycleR_H (n=128)
Days
c
ssGSEA enrichment scores for cancer cell lines
Resistant
Sensitive
Enrichment Score -10 10
GO_REG_NF_KB_IMPORT_INTO_NUCLEU
GO_STAT_CASCADE
GO_NEG_DEFENSE_RESPONSE
GO_ADAPTIVE_IMMUNE_RESPONSE
GO_REGENERATION
GO_ANGIOGENESIS
GO_RESPONSE_TO_GRAVITY
GO_RESPONSE_TO_ALCOHOL
GO_RESPONSE_TO_DEXAMETHASONE
GO_RESPONSE_TO_CORTICOSTERONE
GO_WOUND_HEALING
GO_RESPONSE_TO_WOUNDING
GO_RESPONSE_TO_ESTROGEN
GO_RESPONSE_TO_NUTRIEN
GO_RESPONSE_TO_ALCOHOL
GO_RESPONSE_TO_HEPATOCYTE_GROWTH_FACTOR
GO_RESPONSE_TO_PLATELET_DERIVED_GROWTH_FACTOR
GO_RESPONSE_TO_ACID_CHEMICAL
GO_RESPONSE_TO_ORGANIC_CYCLIC_COMPOUND
GO_RESPONSE_TO_KETONE
GO_RESPONSE_TO_TTGF_BETA
GO_RESPONSE_TO_MONOAMINE
GO_EPITHELIAL_TO_MESENCHYMAL_TRANSITION
d
DMSO
MLN8237
Palbociclib
Viability (treated/control)
shGFP
shPLAT-1
shPLAT-2
shPLAU-1
shPLAU-2
shPLAU-3
shSER-1
shSER-2
shSER-3
e
NES
Response
No_Response
PROTEIN_SECRETION
UNFOLDED_PROTEIN_RESPONSE
KRAS_SIGNALING_DN
MYC_TARGETS_V1
EMT
HYPOXIA
ANGIOGENESIS
CYCLER
GSE10697
GSE19697
GSE41998
GSE50948
GSE5820
GSE6667
GSE66999
GSE76360
f
Enrichment: CycleR
GSE5820: ALL to glucocorticoids
NES = -1.504
P = 0.040
FDR q = 0.108
GSE6667: OV to platinum
NES = -1.444
P = 0.063
FDR q = 0.142
GSE76360: BRCA to trastuzumab
NES = -1.361
P = 0.002
FDR q = 1.0
GSE106977: BRCA to taxanes
NES = -1.558
P = 0.014
FDR q = 0.269
GSE41998: BRCA to Doxorubicin
NES = -1.217
P = 0.268
FDR q = 0.243
GSE50948: BRCA to AT-CMF
NES = -1.098
P = 0.369
FDR q = 0.492
Enrichment score (ES)
Response
No_Response
g
Glioblastoma TCGA N=166
Pancreatic cancer TCGA N=179
CycleC genes
Normalized gene expression -1 1
h
Glioblastoma TCGA N=166
Pancreatic cancer TCGA N=179
CycleR genes
Normalized gene expression -1 1
i
CYCLEC
CYCLER
Normalized gene expression
Classical
Mesenchymal
Neural
Proneural
GBM_subtypes
j
CYCLEC
CYCLER
Normalized gene expression
Mutant
Wildtype
GBM_IDH mutantion

signatures were consistently enriched in these datasets (Fig. 4g). These results led us to propose that transcriptional signatures and the underlying biological processes that define immune prognostic variations are associated with clinical response to immunotherapies.

To test whether these immune PVS distinguish patient subsets in a general fashion, we explored expression of ImmuC and ImmuR signatures in a panel of TCGA datasets. Indeed, ImmuC genes were co-expressed and classified into distinct transcriptional subsets in multiple types of cancer (Figs. 4h and S5i). The ImmuR signature, which was expressed at low levels in melanoma patients (Fig. S5j), defined an ImmuR^{high} (high expression of ImmuR signature) subset in lung cancers (Fig. 4i). Notably, the proportions of ImmuC^{high}, ImmuC^{low}, ImmuR^{high}, and ImmuR^{low} patient subsets were consistent with the prognoses of immune processes in those cancer types. For instance, the *Immune Rev* cancers (e.g., glioblastoma) consisted of a small ImmuC^{high} subset and a large ImmuR^{high} subset, whereas the majority of *Immune Con* cancers (e.g., lung squamous cell carcinoma) were composed of the ImmuR^{low} subset (Figs. 4h and 4i). These findings established that the prognoses of immune processes are dependent on specific transcriptional contexts of tumors that are applicable to distinguish immunotherapy response. In summary, these conglomerate analyses reinforce the concept that prognosis of treatment targeting biological process is an indicator for treatment response. By integrating transcriptomic profiles and clinical parameters, this study presents an approach to establish transcriptional signatures and underlying biological processes as determinants that distinguish treatment responders across multiple cancers (Fig. 4j).

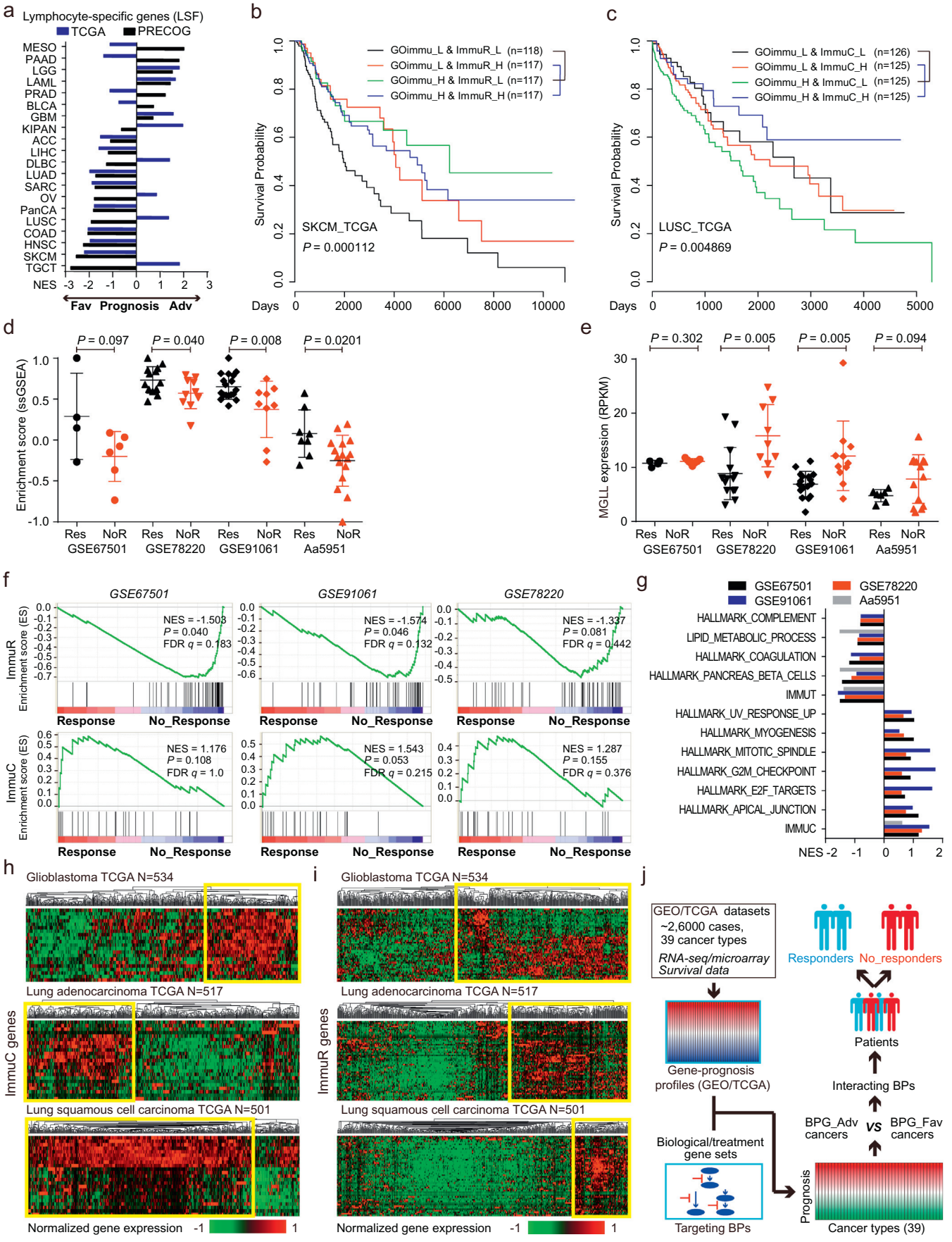
4. Discussion

Distinguishing treatment responders from non-responders is the leading challenge to effective treatment of cancer patients. Intensive experimental and data-mining efforts has been devoted to establish differentially expressed gene patterns to distinguish responsive patients. However, how this response biomarkers are linked to treatment targeting biological processes remains poorly understood. Inspired by the fact that clinical outcomes are the primary readout for the evaluation treatment efficacy, we have developed a prognosis based strategy to define treatment biomarkers. Current approaches for prognoses of gene sets can be affected by the absolute level of gene expression, leading to biased results that are dominated by genes that are abundantly expressed in the gene set [4]. Here we linked gene-prognosis profiles to biological processes through prognosis-ranked GSEA. The prognosis-ranked GSEA approach is validated by the finding that biologically paired gene sets are consistently enriched, both in pan-cancer and cancer specific gene-prognosis profiles. This systematic approach establishes a prognostic landscape of biological processes that uncovers both general and cancer type specific prognostic patterns across 39 distinct malignancies. Based on the prognostic patterns, a comparative analysis of patient cohorts with distinct prognostic patterns establishes transcriptional signatures that distinguish patients cohorts. Notably, transcriptional signatures that define intertumor prognostic variations of treatment targeting biological processes distinguish treatment response, a concept that is proved in both cell-proliferation and immune processes.

The prognoses of biological processes showed both general and cancer-type specific patterns that are consistent with the oncogenic or tumor suppressive roles of these processes. Biological processes, such as cell-proliferation programs, mTORC1 signaling, the unfolded protein response and glycolysis, are frequently associated with adverse prognoses. By contrast, immune processes, TP53 pathway and apoptosis process are linked to favorable prognoses. Notably, the prognoses of all gene sets are context dependent among different cancer types. Cell-proliferation programs, which are the primary targets for current anti-cancer therapeutics, generally confer poor outcomes but signify favorable prognoses in glioblastoma and colon cancer. Similarly, immune processes indicate favorable outcomes in cancers, such as melanoma and lung cancer, yet the same immune processes suggest poor prognoses in pancreatic and a number of brain cancers. These prognostic variations may result from differences in cell-of-origin, oncogenic drivers and microenvironment among different cancer types [1]. For instance, enriched cancer stem-like cells may contribute to favorable prognoses of cell-proliferation markers in brain cancer types, as stem-like cells are slow-cycling (low cell-proliferating activity) and responsible for treatment resistance and relapse [30]. Both tumor intrinsic and microenvironmental heterogeneity may contribute to the prognostic variations of immune processes. Pancreatic cancer and glioblastoma are characterized by infiltrating immune suppressive cells, which may contribute to adverse prognoses of immune processes in these cancer types [31–33]. The prognostic variations not only strengthen the importance of our cancer type-specific prognostic resource, but also provide an opportunity to establish the determinants for prognosis and treatment response.

Tumor heterogeneity presents daunting challenges to the development of biomarkers and treatments for cancer [1]. To examine the potential of our approach in identifying treatment biomarkers, we focused on the cell-proliferation program, the primary target of cancer treatment [34]. Enrichment of stress response processes in CycleR is in accordance with the concept that cell-cycle activation in the context of stress impairs cell survival [35,36]. The enrichment of CycleR and stress response processes in the anti-proliferation non-responder phenotype in multiple datasets demonstrates the consistency of our approach. A consistent correlation with molecular subtypes of glioblastoma further confirms the clinical significance of these signatures [24]. In addition, co-enrichment of PVS in patients indicates that transcriptional programs distinguishing prognostic patterns and treatment response exist across multiple cancers. As anti-proliferative treatments are the primary anti-cancer strategies [37–39], these transcriptional signatures

Fig. 3. Cell-proliferation PVS links prognosis to treatment response (a) Clustering of prognostic values for individual CycleLead genes (Fig. 2b) determined by the Human Pathology Atlas. –Log10PLogRank values were calculated (negative values for favorable prognosis, and positive values for unfavorable prognoses) and clustered by average linkage clustering. Cancer types with favorable prognoses in cell-proliferation associated leading edge genes are labeled in red. Sidebar, prognostic score index. (b) Kaplan Meier analysis of "GO cell cycle" gene set in patients classified by their CycleR signature expression in the TCGA lung adenocarcinoma. Expression levels (low, L; high, H) were defined according to the median value of the gene set expression of the cohort (n = 129/129/129/128 for four groups in TCGA-LUAD). P values were determined using the log-rank test. (c) Hierarchical clustering of ssGSEA enrichment scores of GOBP gene sets (MSigDB) that were differentially enriched between sensitive and resistant cell lines. Sensitivity to 22 compounds targeting cell-proliferation program were defined by the AUC values in the Cancer Therapeutics Response Portal database. Transcriptomic profiles for cell lines were downloaded from the CCLE database. The enrichment of GOBP gene sets were analyzed by ssGSEAProjection in the GenePattern. (d) Independent shRNA targeting PLAT, PLAU, SERPINE1 control (shGFP) were delivered to MDA-MB-231 cells by lentivirus (pLKOTeton). Cells were selected for six days by puromycin (2 mg/ml) two days after transfection. After four days of doxycycline (Dox, 200 ng/ml) treatment, cells were subjected to cell proliferation assay. For proliferation assay, cells were maintained in Dox and treated with DMSO, MLN8237 (AURKA inhibitor, 200 nM) or Palbociclib (CDK4/6 inhibitor, 500 nM). Cell viability was determined by CCK-8 kit, and quantified compared to control cells (shGFP, Dox + DMSO). *, P <.01 (two-tailed Student's t-test, n = 4). (e) The NES of CycleR and top enriched Hallmark gene sets in eight treatment response GEO datasets. GSEA were performed in pre-treated transcriptomes of patients classified as responders and non-responders according to their response to treatments (including glucocorticoids, platinum, taxanes and doxorubicin). (f) Enrichment plots for CycleR in six GEO datasets. GSEA was performed in transcriptomes of responders and non-responders in each dataset. (g) and (h) Hierarchical clustering of CycleC (g) and CycleR (h) genes in TCGA glioblastoma (upper) and pancreatic cancer (lower) cohorts. Yellow boxes indicate transcriptomic subsets with high gene expression. Sidebars in (g) and (h) show the gene expression index. (i) Expression of CycleC (left) and CycleR (right) in TCGA glioblastoma subtypes. Glioblastoma were pre-classified according to their transcriptional features. Signature expression is defined as the mean value of all genes in the signature. *, P <.05; **, P <.01; ****, P <.0001 (Dunnett multiple comparisons following One-Way ANOVA). (j) Expression of CycleC (left) and CycleR (right) in TCGA glioblastoma. Glioblastoma were pre-classified according to their IDH mutation status. Signature expression is defined as the mean value of all genes in the signature. ****, P <.0001 (two-tailed Student's t-test).


a
Lymphocyte-specific genes (LSF)
TCGA
PRECOG
MESO
PAAD
LGG
LAML
PRAD
BLCA
GBM
KIPAN
ACC
LIHC
DLBC
LUAD
SARC
OV
PanCA
LUSC
COAD
HNSC
SKCM
TGCT
NES -3 -2 -1 0 1 2 3
Fav Prognosis Adv
b
GOimmu_L & ImmuR_L (n=118)
GOimmu_L & ImmuR_H (n=117)
GOimmu_H & ImmuR_L (n=117)
GOimmu_H & ImmuR_H (n=117)
Survival Probability
SKCM_TCGA
P = 0.000112
Days 0 2000 4000 6000 8000 10000
c
GOimmu_L & ImmuC_L (n=126)
GOimmu_L & ImmuC_H (n=125)
GOimmu_H & ImmuC_L (n=125)
GOimmu_H & ImmuC_H (n=125)
Survival Probability
LUSC_TCGA
P = 0.004869
Days 0 1000 2000 3000 4000 5000
d
P = 0.097
P = 0.040
P = 0.008
P = 0.0201
Enrichment score (ssGSEA)
Res NoR
GSE67501
GSE78220
GSE91061
Aa5951
e
P = 0.302
P = 0.005
P = 0.005
P = 0.094
MGLL expression (RPKM)
f
GSE67501
GSE91061
GSE78220
ImmuR Enrichment score (ES)
NES = -1.503
P = 0.040
FDR q = 0.183
NES = -1.574
P = 0.046
FDR q = 0.132
NES = -1.337
P = 0.081
FDR q = 0.442
ImmuC Enrichment score (ES)
NES = 1.176
P = 0.108
FDR q = 1.0
NES = 1.543
P = 0.053
FDR q = 0.215
NES = 1.287
P = 0.155
FDR q = 0.376
Response
No_Response
g
GSE67501
GSE91061
GSE78220
Aa5951
HALLMARK_COMPLEMENT
LIPID_METABOLIC_PROCESS
HALLMARK_COAGULATION
HALLMARK_PANCREAS_BETA_CELLS
IMMUT
HALLMARK_UV_RESPONSE_UP
HALLMARK_MYOGENESIS
HALLMARK_MITOTIC_SPINDLE
HALLMARK_G2M_CHECKPOINT
HALLMARK_E2F_TARGETS
HALLMARK_APICAL_JUNCTION
IMMUC
NES -2 -1 0 1 2
h
Glioblastoma TCGA N=534
Lung adenocarcinoma TCGA N=517
ImmuC genes
Lung squamous cell carcinoma TCGA N=501
Normalized gene expression -1 1
i
Glioblastoma TCGA N=534
Lung adenocarcinoma TCGA N=517
ImmuR genes
Lung squamous cell carcinoma TCGA N=501
Normalized gene expression -1 1
j
GEO/TCGA datasets
~2,6000 cases,
39 cancer types
RNA-seq/microarray
Survival data
Gene-prognosis profiles (GEO/TCGA)
Biological/treatment gene sets
Targeting BPs
Responders
No_responders
Patients
Interacting BPs
BPG_Adv cancers
VS
BPG_Fav cancers
Prognosis
Cancer types (39)

and underlying biological processes have significant implications for patient stratification and effective treatment. The concept that PVS distinguishes treatment response has now been proven, and this is not limited to cell-proliferation program. This approach may also be applicable to treatments targeting other biological processes, especially treatments that are newly approved or under preclinical development, thereby providing a strategy for patient stratification based on existing cancer datasets.

Immunotherapy is now being evaluated in a wide range of cancers [38,39], yet the predictors and mechanisms of response remains poorly understood. For example, mutational loads have been demonstrated to benefit anti-CTLA4 responses in melanoma [40] and anti-PD1 response in lung cancer [41,42], while another study reports that mutational loads benefit survival, but not anti-PD-1 response [23]. The limited patient numbers may contribute to the substantial inconsistencies between these studies [43–52]. In contrast, our approach establishes transcriptional signatures and the underlying biological processes based on the molecular profiles of ~26,000 cases, and are applicable in multiple independent datasets. This high levels of consistency indicates a general mechanism of immunotherapy responses. Specifically, the T cell activation processes that are critical for successful immunotherapy [53], are enriched in ImmuC. Similarly, enrichment of Treg and lipid metabolism processes in ImmuR is consistent with the immunosuppressive function of Treg [54] and requirement for lipid uptake and metabolism for survival of tissue-resident memory T cells [55]. Indeed, lipid metabolism controls the functional switch between immune protection and disease state in T cells [56]. Interestingly, cancer types with favorable immune prognoses, such as melanoma and lung cancer, are responsive to immunotherapies [38,39,57], whereas cancer types with adverse immune prognoses, such as pancreatic and several brain cancers, show a poor response to immunotherapy [58]. These findings further confirm a link between prognosis of treatment targeting process and treatment response.

The findings that prognoses are linked to treatment responses for specific biological processes may be useful for repurposing currently available targets and treatments based on the prognoses of related biological processes. The prognosis-repurposed vulnerabilities provide a resource of both therapeutic targets (e.g., MYC, EZH2, CDK4) and interventions (e.g., Rapamycin, Aplidin, Darapladib) that will facilitate development of novel therapeutic approaches across 39 human malignancies. This cancer type-dependent property of treatment response establishes resources that repurpose established treatments in specific cancer types. Identifying therapeutic vulnerabilities for cancers, especially those malignancies for which there are currently no effective treatments, is urgently needed. Therefore, our analysis also has the important advantage of presenting valuable resources indicative of potential treatment approaches for distinct malignancies. Although we have tried our best to optimize these analyses by keeping the versions of data consistent among different online tools and databases, there're still several limitations in this study. Firstly, the prognostic analyses here are based on overall survival because the lack of recurrence-free survival information, which is more relevant to treatment response in cancer datasets. Secondly, the prognostic landscape built by GSEA is more exploratory and less accurate as predictive biomarkers here, and more efforts are needed before their clinical application [4]. Thirdly, although several findings from our analyses are consistent with previous experimental publications, we provided only cell-based experimental validations of CycleR signature. Since both tumor-specific and immune cell-specific factors should be considered in responses to immunotherapy [43–52], immune signatures are not experimentally validated because responses to immunotherapy are not well recapitulated in cell based experiments in vitro. Future in vivo studies will be considered for cell-proliferation, immune processes and other biological processes of potential interest.

Collectively, by integrating transcriptomic profiles, clinical outcomes and biological processes, we have now characterized the prognostic landscape of biological processes for a variety of cancers. These prognoses show intertumor variation patterns, which is a valuable property needed to define representative cancer types and transcriptional determinants of prognostic variation. This link between prognosis and treatment response establishes a strategy to stratify sensitive and resistant patients using transcriptional signatures that define prognostic variations. The concept that prognosis of target distinguishes treatment response may be applicable to a variaty of treatments, providing potential biomarkers for further experimental and clinical validations. These findings highlight the utility of bioinformatic resources to develop biomarkers and treatment strategies that will allow more precisely targeted and immune therapies for cancer patients.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.01.064.

Acknowledgements

We thank all members of Liu laboratory and Li laboratory for their critical comments and technical support.

Funding sources

This work was supported by the National Natural Science Foundation of China (81630005, 81130040, and 81573025 to Q.L., 81773166 to Z.W.), the Innovative Research Team in University of Ministry of Education of China (IRT13049 to Q.L.), the National Key Research and Development Program of China (2017YFA0505600-04 to Q.L.), the Natural Science Foundation of Guangdong (2016A030311038 and 2017A030313608 to Q.L., 2017A020215098 to Z.W.), the Science and Technology Planning Project of Guangzhou (201604020163 to Q.L.), and MRC (MR/N012097/1), CRUK (A12011), Breast Cancer Now (2012MayPR070; 2012NovPhD016), the Cancer Research UK Imperial Centre, Imperial ECMC and NIHR Imperial BRC to E. W.-F. Lam and the NIH grant (RO1 SUB UT00000712 to K.W. Kelley).

Fig. 4. Immune interactive processes indicate immunotherapy response (a) The prognostic NES of favorable LSF gene set (32 genes that are specifically expressed in lymphocytes that exhibited favorable prognostic z-scores in > 60% (24/39) PRECOG cancer datasets) in 19 matched TCGA (blue bars) and PRECOG (black bars) cancer types. (b) and (c) Kaplan Meier analysis of the "GO immune response" gene set in patient subsets classified by ImmuR (b) or ImmuC (c) signature expression in the TCGA melanoma (SKCM, b) and lung squamous cell carcinoma (LUSC, c) datasets. Expression level (low, L; high, H) was defined according to the median value of gene set expression of the cohort (n = 118/117/117/117 for four groups in SKCM, and n = 125/125/125/125 for four groups in LUSC). P values were determined by log-rank test. (d) The combined ssGSEA enrichment scores of ImmuC and ImmuR gene sets (score ImmuC-score ImmuR) in individual patients from four independent datasets (GSE67501, GSE78220, GSE91061 and Aa5951). Patients were classified as responders (Res, black dots) and non-responders (NoR, red dots) to immunotherapy according to their clinical outcomes in each dataset. P values were calculated using a two-tailed Student's t-test. (e) Expression of MGLL in individual patients from four independent datasets (GSE67501, GSE78220, GSE91061 and Aa5951). Patients were classified as responders (Res, labeled in black) or non-responders (NoR, labeled in red) to immunotherapy according to their clinical outcomes. P values were calculated using a two-tailed Student's t-test. (f) The NES of ImmuR (upper panel) and ImmuC (lower panel) in three independent GEO datasets. GSEA were performed in transcriptomes of patients classified as responders and non-responders according to their response to immunotherapy. (g) The NES of ImmuR, ImmuC and Hallmark gene sets in three GEO datasets (black, GSE67501; blue, GSE91061; red, GSE78220; grey, Aa5951). GSEA were performed in transcriptomes of patients classified as responders and non-responders according to their response to treatment. (h) and (i) Hierarchical clustering of ImmuC (h) and ImmuR (i) genes in three TCGA datasets, including glioblastoma (upper panel), lung adenocarcinoma (middle panel) and lung squamous cell carcinoma (lower panel). Yellow boxes indicate distinct clustered subsets with high gene expression. Sidebars, gene expression index. (j) A summary of the strategy to identify interactive biological processes for treatment responses and to stratify treatment responders from non-responders. The prognoses of BPGSs were systematically evaluated among cancer datasets. Differentially expressed genes between BPGS-favorable and BPGS-adverse patients were defined to identify the underlying biological processes and to distinguish responder patients. The concept has now been proven in treatments targeting both tumor (anti-proliferation treatment) and non-tumor cells (immunotherapy).

Declaration of interests

The authors declare that no conflict of interest exists.

Author contributions

Bin He and Rui Gao participated in design, acquisition, analysis and interpretation of data and manuscript preparation for the entire project. Dekang Lv, Yalu Wen, Luyao Song and Ziqian Deng performed the assembly of TCGA datasets, NMF analysis and statistical analysis of validation datasets. Xi Wang, Suxia Lin, Qitao Huang, Min Yan, Zifeng Wang, Feimeng Zheng, Eric W.-F. Lam, Keith W. Kelley and Zhiguang Li participated in manuscript editing and making critical revisions. Quentin Liu participated in conception, design, analysis and interpretation of data in the entire project and made critical revisions in the manuscript.

References

[1] Bedard PL, Hansen AR, Ratain MJ, et al. Tumour heterogeneity in the clinic. Nature 2013;501(7467):355–64.
[2] Nevins JR, Potti A. Mining gene expression profiles: expression signatures as cancer phenotypes. Nat Rev Genet 2007;8(8):601–9.
[3] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545–50.
[4] de Leeuw CA, Neale BM, Heskes T, et al. The statistical properties of gene-set analysis. Nat Rev Genet 2016;17(6):353–64.
[5] Liberzon A, Birger C, Thorvaldsdottir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst 2015;1(6):417–25.
[6] Gentles AJ, Newman AM, Liu CL, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med 2015;21(8):938–45.
[7] Uhlen M, Zhang C, Lee S, et al. A pathology atlas of the human cancer transcriptome. Science 2017;357(6352).
[8] Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6(269):pl1.
[9] Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2(5):401–4.
[10] Goswami CP, Nakshatri H. PROGgene: gene expression based survival analysis web application for multiple cancers. J Clin Bioinforma 2013;3(1):22.
[11] Aguirre-Gamboa R, Gomez-Rueda H, Martinez-Ledesma E, et al. SurvExpress: an online biomarker validation tool and database for cancer gene expression data using survival analysis. PLoS One 2013;8(9):e74250.
[12] Li J, Lu Y, Akbani R, et al. TCPA: a resource for cancer functional proteomics data. Nat Methods 2013;10(11):1046–7.
[13] Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;6(1):1–6.
[14] Gorenshteyn D, Zaslavsky E, Fribourg M, et al. Interactive big data resource to elucidate human immune pathways and diseases. Immunity 2015;43(3):605–14.
[15] Barretina J, Caponigro G, Stransky N, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483(7391):603–7.
[16] Brunet JP, Tamayo P, Golub TR, et al. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 2004;101(12):4164–9.
[17] Basu A, Bodycombe NE, Cheah JH, et al. An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell 2013;154(5):1151–61.
[18] Seashore-Ludlow B, Rees MG, Cheah JH, et al. Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer Discov 2015;5(11):1210–23.
[19] Barbie DA, Tamayo P, Boehm JS, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462(7269):108–12.
[20] Wei G, Twomey D, Lamb J, et al. Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. Cancer Cell 2006;10(4):331–42.
[21] Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30(1):207–10.
[22] Spijkers-Hagelstein JA, Schneider P, Hulleman E, et al. Elevated S100A8/S100A9 expression causes glucocorticoid resistance in MLL-rearranged infant acute lymphoblastic leukemia. Leukemia 2012;26(6):1255–65.
[23] Hugo W, Zaretsky JM, Sun L, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell 2016;165(1):35–44.
[24] Verhaak RG, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 2010;17(1):98–110.
[25] Phillips HS, Kharbanda S, Chen R, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell 2006;9(3):157–73.
[26] Tran AN, Lai A, Li S, et al. Increased sensitivity to radiochemotherapy in IDH1 mutant glioblastoma as demonstrated by serial quantitative MR volumetry. Neuro Oncol 2014;16(3):414–20.
[27] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144(5):646–74.
[28] Ascierto ML, McMiller TL, Berger AE, et al. The intratumoral balance between metabolic and immunologic gene expression is associated with anti-PD-1 response in patients with renal cell carcinoma. Cancer Immunol Res 2016;4(9):726–33.
[29] Miao D, Margolis CA, Gao W, et al. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science 2018;359(6377):801–6.
[30] Saygin C, Matei D, Majeti R, et al. Targeting cancer stemness in the clinic: from hype to hope. Cell Stem Cell 2019;24(1):25–40.
[31] Preusser M, Lim M, Hafler DA, et al. Prospects of immune checkpoint modulators in the treatment of glioblastoma. Nat Rev Neurol 2015;11(9):504–14.
[32] Lim M, Xia Y, Bettegowda C, et al. Current state of immunotherapy for glioblastoma. Nat Rev Clin Oncol 2018;15(7):422–42.
[33] Bayne LJ, Beatty GL, Jhala N, et al. Tumor-derived granulocyte-macrophage colony-stimulating factor regulates myeloid inflammation and T cell immunity in pancreatic cancer. Cancer Cell 2012;21(6):822–35.
[34] Sherr CJ. Principles of tumor suppression. Cell 2004;116(2):235–46.
[35] Kruman II, Wersto RP, Cardozo-Pelaez F, et al. Cell cycle activation linked to neuronal cell death initiated by DNA damage. Neuron 2004;41(4):549–61.
[36] Nguyen MD, Boudreau M, Kriz J, et al. Cell cycle regulators in the neuronal death pathway of amyotrophic lateral sclerosis caused by mutant superoxide dismutase 1. J Neurosci 2003;23(6):2131–40.
[37] Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. J Clin Oncol 2015;33(17):1974–82.
[38] Topalian SL, Drake CG, Pardoll DM. Immune checkpoint blockade: a common denominator approach to cancer therapy. Cancer Cell 2015;27(4):450–61.
[39] Kirkwood JM, Butterfield LH, Tarhini AA, et al. Immunotherapy of cancer in 2012. CA Cancer J Clin 2012;62(5):309–35.
[40] Van Allen EM, Miao D, Schilling B, et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 2015;350(6257):207–11.
[41] Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015;348(6230):124–8.
[42] Hellmann MD, Callahan MK, Awad MM, et al. Tumor mutational burden and efficacy of nivolumab monotherapy and in combination with ipilimumab in small-cell lung cancer. Cancer Cell 2018;33(5):853–61 [e854].
[43] Ascierto ML, Makohon-Moore A, Lipson EJ, et al. Transcriptional mechanisms of resistance to anti-PD-1 therapy. Clin Cancer Res 2017;23(12):3168–80.
[44] Patel SJ, Sanjana NE, Kishton RJ, et al. Identification of essential genes for cancer immunotherapy. Nature 2017;548(7669):537–42.
[45] Le DT, Durham JN, Smith KN, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 2017;357(6349):409–13.
[46] Topalian SL, Taube JM, Anders RA, et al. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. Nat Rev Cancer 2016;16(5):275–87.
[47] De Henau O, Rausch M, Winkler D, et al. Overcoming resistance to checkpoint blockade therapy by targeting PI3Kgamma in myeloid cells. Nature 2016;539(7629):443–7.
[48] Gopalakrishnan V, Spencer CN, Nezi L, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science 2018;359(6371):97–103.
[49] Krieg C, Nowicka M, Guglietta S, et al. High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. Nat Med 2018;24(2):144–53.
[50] Pan D, Kobayashi A, Jiang P, et al. A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. Science 2018;359(6377):770–5.
[51] Chowell D, Morris LGT, Grigg CM, et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. Science 2018;359(6375):582–7.
[52] Shukla SA, Bachireddy P, Schilling B, et al. Cancer-germline antigen expression discriminates clinical outcome to CTLA-4 blockade. Cell 2018;173(3):624–633 e628.
[53] Tumeh PC, Harview CL, Yearley JH, et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. Nature 2014;515(7528):568–71.
[54] Sakaguchi S, Miyara M, Costantino CM, et al. FOXP3+ regulatory T cells in the human immune system. Nat Rev Immunol 2010;10(7):490–500.
[55] Pan Y, Tian T, Park CO, et al. Survival of tissue-resident memory T cells requires exogenous lipid uptake and metabolism. Nature 2017;543(7644):252–6.
[56] Wang C, Yosef N, Gaublomme J, et al. CD5L/AIM regulates lipid biosynthesis and restrains Th17 cell pathogenicity. Cell 2015;163(6):1413–27.
[57] Hamid O, Robert C, Daud A, et al. Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. N Engl J Med 2013;369(2):134–44.
[58] Royal RE, Levy C, Turner K, et al. Phase 2 trial of single agent Ipilimumab (anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. J Immunother 2010;33(8):828–33.