RESOURCE

nature medicine

The prognostic landscape of genes and infiltrating immune cells across human cancers

Andrew J Gentles[1,2,12], Aaron M Newman[3,4,12], Chih Long Liu[3,4], Scott V Bratman[3,5,11], Weiguo Feng[3,5], Dongkyoon Kim[3], Viswam S Nair[6], Yue Xu[7], Amanda Khuong[7], Chuong D Hoang[7,11], Maximilian Diehn[3,5,8], Robert B West[9], Sylvia K Plevritis[1,2,13] & Ash A Alizadeh[1,3,4,8,10,13]

Molecular profiles of tumors and tumor-associated cells hold great promise as biomarkers of clinical outcomes. However, existing data sets are fragmented and difficult to analyze systematically. Here we present a pan-cancer resource and meta-analysis of expression signatures from ~18,000 human tumors with overall survival outcomes across 39 malignancies. By using this resource, we identified a forkhead box MI (*FOXM1*) regulatory network as a major predictor of adverse outcomes, and we found that expression of favorably prognostic genes, including *KLRB1* (encoding CD161), largely reflect tumor-associated leukocytes. By applying CIBERSORT, a computational approach for inferring leukocyte representation in bulk tumor transcriptomes, we identified complex associations between 22 distinct leukocyte subsets and cancer survival. For example, tumor-associated neutrophil and plasma cell signatures emerged as significant but opposite predictors of survival for diverse solid tumors, including breast and lung adenocarcinomas. This resource and associated analytical tools (http://precog.stanford.edu) may help delineate prognostic genes and leukocyte subsets within and across cancers, shed light on the impact of tumor heterogeneity on cancer outcomes, and facilitate the discovery of biomarkers and therapeutic targets.

Genomic features of tumors and their microenvironments represent promising candidates for predictive and prognostic biomarkers[1–3]. Despite intensive efforts over the past decade to leverage emerging high-throughput genomic technologies[4,5], heterogeneity in patient cohorts, treatment regimens and technological platforms, among other factors, has led to apparently inconsistent results and modest translational impact[6–11]. To address these issues, and to gain new insights into the molecular features of tumors with prognostic associations, we integrated tumor gene expression profiles (GEPs) and overall survival data from nearly 18,000 patients within a meta-analytical framework that enhances statistical power and improves reproducibility[12]. This study differs from previous efforts, which were not amenable to meta-analysis[13] or were limited to either single cancer types[14] or single data sets per cancer type[15].

We applied the recently described CIBERSORT (Cell type Identification By Estimating Relative Subsets Of known RNA Transcripts) method[16] to this data resource to analyze associations between clinical outcomes and abundance of diverse tumor-associated leukocyte (TAL) subsets, several of which have been linked to tumor growth[17,18], cancer progression and outcome[19]. This approach quantifies the relative expression of cell type specific signatures in bulk tumors, and it is largely independent of methods that rely on cell isolation, preservation, and reagent quality, all of which are difficult to standardize across large numbers of tumors.

Using this resource and these analytical tools, we constructed a global pan-cancer map of the landscape of both genes and TALs predicting clinical outcomes, integrating with existing resources such as ENCODE[20]. Our findings reveal genome-wide molecular portraits of human tumors and identify candidate genes and TALs for prognostic stratification and targeted therapy.

RESULTS

A cancer-wide atlas of prognostic genes

We assembled, curated, and integrated cancer gene expression and clinical outcome data from the public domain into a new resource for PREdiction of Clinical Outcomes from Genomic profiles (PRECOG, http://precog.stanford.edu; **Fig. 1a** and **Supplementary Data 1**). PRECOG encompasses nearly 30,000 expression profiles from 166 cancer expression data sets covering 39 distinct malignant histologies (**Fig. 1b**). Importantly, we manually curated all data with respect to

[1]Center for Cancer Systems Biology (CCSB), Stanford University, Stanford, California, USA. [2]Department of Radiology, Stanford University, Stanford, California, USA. [3]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, California, USA. [4]Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University, Stanford, California, USA. [5]Department of Radiation Oncology, Stanford University, Stanford, California, USA. [6]Department of Medicine, Division of Pulmonary and Critical Care Medicine, Stanford University, Stanford, California, USA. [7]Department of Cardiothoracic Surgery, Division of Thoracic Surgery, Stanford University, Stanford, California, USA. [8]Stanford Cancer Institute, Stanford University, Stanford, California, USA. [9]Department of Pathology, Stanford University, Stanford, California, USA. [10]Department of Medicine, Division of Hematology, Stanford Cancer Institute, Stanford University, Stanford, California, USA. [11]Present addresses: Department of Radiation Oncology, Princess Margaret Cancer Centre, University of Toronto, Toronto, Ontario, Canada (S.V.B.); Thoracic and Gastrointestinal Oncology Branch, National Cancer Institute, Bethesda, Maryland, USA (C.D.H.). [12]These authors contributed equally to this work. [13]These authors jointly directed this work. Correspondence should be addressed to A.A.A. (arasha@stanford.edu).

Received 19 January; accepted 19 June; published online 20 July 2015; doi:10.1038/nm.3909

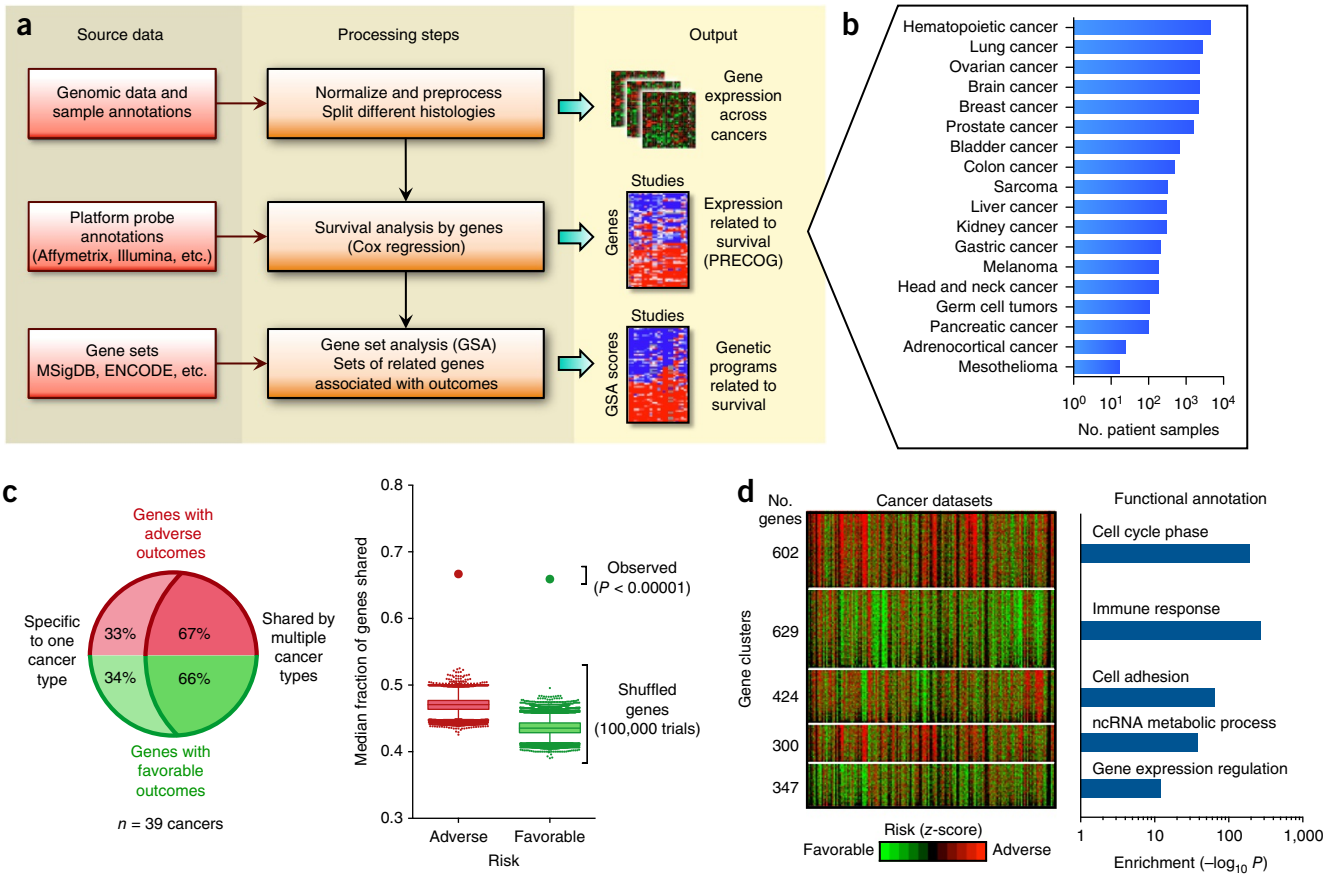

Figure 1 Prognostic landscape of gene expression across human cancers. (**a**) Schematic depicting PRECOG data pre-processing and analysis steps. (**b**) Number of patient samples with survival data included in PRECOG, organized by cancer type. Thirty-nine distinct histologies (for example, adenocarcinoma and squamous cell carcinoma in lung cancer, different types of blood cancer) have been grouped into 18 clusters for concise display. (**c**) Left, approximately two-thirds of prognostic genes (filtered for $|\text{meta-}z| > 3.09$, or nominal one-sided $P < 0.001$) are prognostic in more than one of the 39 distinct cancer histologies for which meta-*z*-scores were computed, while the remaining one-third are prognostic in only a single histology; the latter are cancer-specific. Right, same analysis shown at left but applied to randomly shuffled gene labels for each cancer in PRECOG. On the basis of 100,000 trials, the empirical *P* value for the observed enrichment of shared genes is $P < 10^{-5}$ (Monte Carlo simulation). (**d**) Left, heat map showing genes (rows) clustered by association between expression levels and survival outcomes across 166 individual cancer studies (columns). *Z*-scores represent the statistical significance of each gene's association with survival, with poor prognosis genes in red, and favorable prognosis genes in green. All identified clusters were ranked by compound scores that integrate cluster size with the prognostic significance of genes within each cluster; the five top-ranking clusters are shown (left). Right, representative functional enrichments for each of the five clusters, determined by analyzing annotated gene sets with a Bonferroni-corrected hypergeometric test. All clusters, including associated data sets and compound scores, are provided in **Supplementary Data 3**.

relevant clinical parameters, including disease status, stage and histology, and we verified key results from selected studies (for example, survival plots; see Online Methods). To avoid ambiguities in outcome annotation, we restricted the results presented herein to the ~18,000 cases with overall survival data. We further confirmed the consistency and quality of integrated data sets and curated annotations by assessing the concordance between molecularly inferred versus clinically reported patient gender. In all, 98% of tumors were gender-concordant, with less than 2% of tumors affected by probable annotation errors (such as 'off-by-one' mismatches) (**Supplementary Fig. 1a–c**). All microarray studies in PRECOG were consistently normalized and pre-processed (see Online Methods) and associations between gene expression and survival were assessed by univariate Cox regression.

To compare prognostic associations across independent data sets, and to minimize the confounding influence of batch effects, the statistical associations between genes and clinical outcomes were assessed by *z*-scores. *Z*-scores are directly related to *P* values, but they conveniently encode both the directionality and robustness of statistical associations (**Supplementary Fig. 1d**). Moreover, *z*-scores have a straightforward interpretation; they represent the number of s.d. from the mean of a normal distribution. For example, $|z| > 1.96$ is equivalent to a two-sided $P < 0.05$ (**Supplementary Fig. 1d**). Unlike parameters such as hazard ratios, *z*-scores are independent of different timescales measuring survival follow-up times, and of the range/scale of predictor variables, permitting direct comparison across studies and platforms. To facilitate cross-cancer analyses, *z*-scores for individual studies were combined to yield 'meta-*z*-scores' for the prognostic significance of each gene in each cancer type (**Supplementary Data 1**). We observed high concordance between meta-*z*-scores and *z*-scores, where the latter were obtained by first merging expression data from multiple studies of the same cancer (for example, lung adenocarcinoma, Spearman's $R = 0.9$, $P < 2.2 \times 10^{-16}$; see Online Methods). To further evaluate the robustness of the meta-*z* metric, we calculated a global meta-*z*-score for each gene across all cancers,

and compared these to meta-*z*-scores computed across an external validation set of nine independent studies (**Supplementary Data 1**). *Z*-scores of globally prognostic genes were significantly correlated between PRECOG and the validation set ($R = 0.55$, $P < 2.2 \times 10^{-16}$; **Supplementary Fig. 2a**,**b**). In addition, pan-cancer prognostic genes were significantly concordant between PRECOG and another validation set comprised of studies profiled by RNA-seq from The Cancer Genome Atlas (TCGA) ($R = 0.52$, $P < 2.2 \times 10^{-16}$; **Supplementary Fig. 2a**,**b**). We also evaluated the influence of batch effects[21] on *z*-score values. Notably, we observed only modest differences in *z*-scores after batch effect removal (for example, for samples run on different dates) (**Supplementary Fig. 2c**–**e**).

Pan-cancer prognostic genes

The PRECOG system provides an unprecedented opportunity to quantify commonalities in prognostic genes across a large number of human malignancies. We found that prognostic genes (filtered at |meta-*z*| > 3.09, or nominal one-sided $P < 0.001$) are significantly more likely to be shared by distinct tumor types than would be expected by random chance (**Fig. 1c** and **Supplementary Data 2**). This result was reproducible across a broad range of statistical thresholds (**Supplementary Fig. 3a**,**b**), and it is reminiscent of the high cancer-wide concordance reported among somatic aberrations influencing genome-wide copy number alterations[22]. Conversely, cancer-specific prognostic genes are less frequent than would be expected by random chance (**Fig. 1c** and **Supplementary Fig. 3a**,**b**), and they predominantly reflect their tissues of origin (**Supplementary Fig. 3c** and **Supplementary Data 2**).

To obtain a global map of prognostic patterns, we clustered survival-associated *z*-scores across all 166 PRECOG data sets using AutoSOME, an unsupervised method that is not sensitive to outliers and does not require pre-specification of the number of clusters[23] (**Fig. 1d** and **Supplementary Data 3**). Prognostic clusters include genes involved in cell adhesion and epithelial–mesenchymal transitions, vascularization, and immunological and proliferative processes (**Supplementary Data 3**). When clusters were ordered by a metric that integrates gene-level meta-*z*-scores and cluster size, the two largest clusters were most highly ranked (**Fig. 1d**, left). One of these two clusters is broadly associated with inferior outcomes, and is functionally linked to cell proliferation and cell cycle phase (**Fig. 1d**, right). Although this cluster is prognostic in many solid tumors, such as breast and lung adenocarcinoma, we found that proliferation genes were not adversely prognostic in some cancers, including colon cancer and acute myeloid leukemia (AML) (**Supplementary Data 1**), two malignancies for which the clinical relevance of generally quiescent cancer stem cells has been demonstrated[24,25]. The other large cluster is associated with favorable survival and is highly enriched in immunological processes and immune-response genes (**Fig. 1d** and **Supplementary Data 3**), suggesting that the immune microenvironment is a key factor that contributes to favorable survival across cancers.

To further explore cancer-wide prognostic signatures, we used PRECOG to define robust pan-cancer survival models. First, we determined the number of histologies needed to identify genes with maximal prognostic power. Using a cross-validation approach to avoid outliers, we observed quantitative improvements in the significance of pan-cancer prognostic genes until ~30 distinct histologies were sampled, after which additional gains were marginal (**Fig. 2a**, left). With this framework, we identified the top ten adverse and top ten favorable prognostic genes, largely consisting of genes that regulate or oscillate with the mitotic cell cycle, as well as markers of distinct lymphocyte subsets (**Fig. 2a**, right). Expression of the proto-oncogene *FOXM1* (ref. 26) was most frequently associated with adverse risk, and it outperformed *MKI67*, which encodes a protein (Ki-67) that is used clinically as a marker of proliferation[27] (**Supplementary Fig. 4a**). Expression of *KLRB1* (encoding CD161, a surface marker on several T cell subsets[28] and NK cells; **Supplementary Fig. 4b**) was most frequently associated with favorable outcomes. Notably, *FOXM1* and *KLRB1* were among the top pan-cancer genes in validation sets, including the TCGA RNA-seq data (**Fig. 2b** and **Supplementary Fig. 2b**). Bivariate models based on these two genes outperformed either gene alone (**Supplementary Data 4**). Moreover, when weighted by their Cox regression coefficients in a training set (**Supplementary Data 4**), a *FOXM1-KLRB1* composite score stratified patient survival in diverse internal validation data sets, including carcinomas, brain tumors, and hematopoietic neoplasms (**Fig. 2c**), and was prognostic in multivariate models integrating common clinical indices (**Supplementary Data 4**). Furthermore, the *FOXM1-KLRB1* composite score remained prognostic in external validation data sets, including TCGA RNA-seq data sets (**Supplementary Fig. 4c**,**d**).

We next asked whether integrating cancers in a meta-analysis might illuminate additional functional elements of biological programs that affect survival. To address this, we evaluated connectivity among top prognostic genes within protein-protein association (PPA) networks. For adversely prognostic genes, a considerably higher connectivity was achieved when considering all cancer histologies together as opposed to separately (**Fig. 2d** and **Supplementary Data 5**), suggesting that individual malignancies in PRECOG contribute complementary information to a pervasive proliferation program (**Fig. 2d**, middle). We further characterized this network by integrating PRECOG data with data from ENCODE[20], ChEA[29], and mSigDB[30] for 1,006 gene sets defined by shared transcription factor binding sites. Adversely prognostic genes in this network and across PRECOG were most significantly enriched in *FOXM1* ChIP-seq binding targets[31] ($P < 2.2 \times 10^{-22}$; **Supplementary Fig. 4e** and **Supplementary Data 5**), which together with *FOXM1* itself, may be drivers of inferior survival. In contrast, we observed no difference in protein-protein connectivity among favorably prognostic genes, whether they were identified from pooled cancers or individual cancers (**Fig. 2d** and **Supplementary Data 5**). We hypothesized that transcriptome heterogeneity among distinct TALs, which accounts for the majority of favorable pan-cancer prognostic associations, may not be well captured by aggregated PPA studies. In addition, several highly prognostic genes, including *KLRB1*, are expressed by multiple immune subsets (**Supplementary Fig. 4b**). Therefore we performed an '*in silico* dissection' of PRECOG to infer the specific leukocyte subsets associated with survival across cancers.

Leukocyte composition in bulk tumors

Infiltration of tumors by specific leukocyte cell subsets such as $CD8^+$ and $CD45RO^+$ memory T-lymphocytes has been linked with favorable outcomes in different cancers[2,32], whereas other subsets, such as regulatory T cells and macrophages, can confer either good or poor prognosis depending on context[33–36]. To systematically and comprehensively map compositional differences in TALs and their relationships to survival, we applied a new machine-learning tool known as CIBERSORT[16]. CIBERSORT outperforms previous deconvolution methods with respect to noise, unknown mixture content, and closely related cell types, in statistically estimating relative proportions of cell subsets from expression profiles of complex tissues (for example, bulk tumors)[16]. As input, we used purified

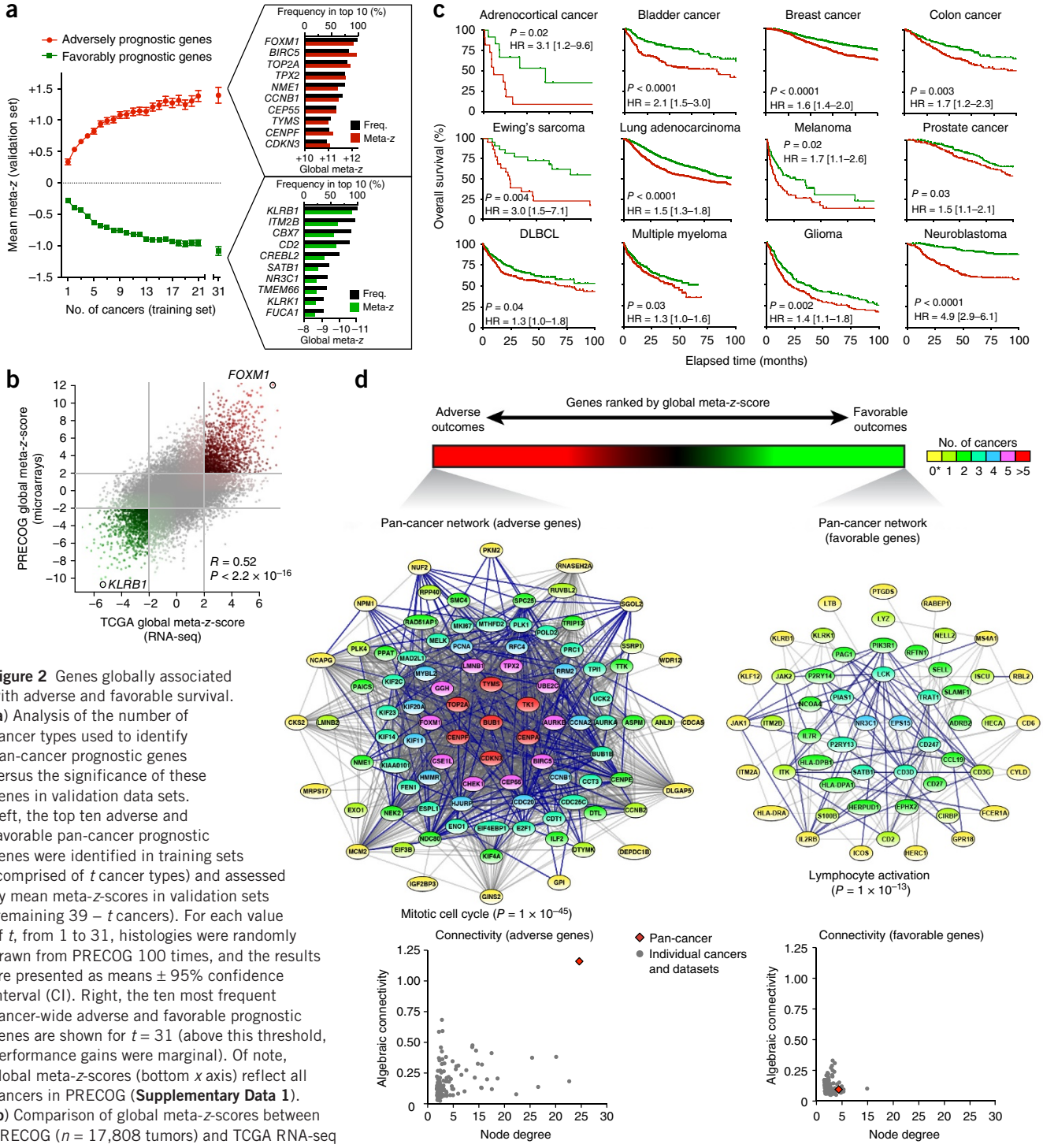

Figure 2 Genes globally associated with adverse and favorable survival. (**a**) Analysis of the number of cancer types used to identify pan-cancer prognostic genes versus the significance of these genes in validation data sets. Left, the top ten adverse and favorable pan-cancer prognostic genes were identified in training sets (comprised of *t* cancer types) and assessed by mean meta-*z*-scores in validation sets (remaining 39 – *t* cancers). For each value of *t*, from 1 to 31, histologies were randomly drawn from PRECOG 100 times, and the results are presented as means ± 95% confidence interval (CI). Right, the ten most frequent cancer-wide adverse and favorable prognostic genes are shown for *t* = 31 (above this threshold, performance gains were marginal). Of note, global meta-*z*-scores (bottom *x* axis) reflect all cancers in PRECOG (**Supplementary Data 1**). (**b**) Comparison of global meta-*z*-scores between PRECOG (*n* = 17,808 tumors) and TCGA RNA-seq data (*n* = 6,663 tumors), with *FOXM1* and *KLRB1* indicated. Points lying between the parallel gray lines represent insignificant genes in PRECOG, TCGA, or both (nominal two-sided $P > 0.05$). (**c**) Kaplan-Meier curves showing differences in overall survival for patients in validation sets stratified by a *FOXM1* and *KLRB1* expression score (see Online Methods). For each cancer, a median split was used and curve separation was assessed by a log-rank test. Survival units from different studies were standardized to months. Lung cancers were primarily stage I (approximately two-thirds), and the melanoma data consisted primarily of metastatic samples (see Online Methods). 95% confidence intervals are presented in brackets. HR, hazard ratio. (**d**) Top, genes ranked by mean meta-*z*-scores across all data sets in PRECOG (*n* = 23,288 genes). Center, PPA networks for the top 100 genes determined by mean pan-cancer meta-*z*-scores. Edges are colored to denote experimentally confirmed interactions and/or associations in curated databases (blue edges), and other sources of evidence (gray edges) (see Online Methods and **Supplementary Data 5**). Functional annotation *P* values were determined using a Benjamini-Hochberg–corrected hypergeometric test. Genes in the pan-cancer prognostic networks are colored according to the number of cancer-specific PPA networks in which they are also found. 0* indicates genes only found in PPA networks derived from all cancers. Bottom, two metrics of network connectivity are compared among PPA networks for the top 100 prognostic genes derived from all cancers (red diamonds) versus individual cancers and studies in the PRECOG data (gray circles): *x* axis = node degree, the average number of edges *e* (i.e., PPAs) per node *n* (i.e., protein); *y* axis = algebraic connectivity, a graph theoretic measure of overall network connectedness.

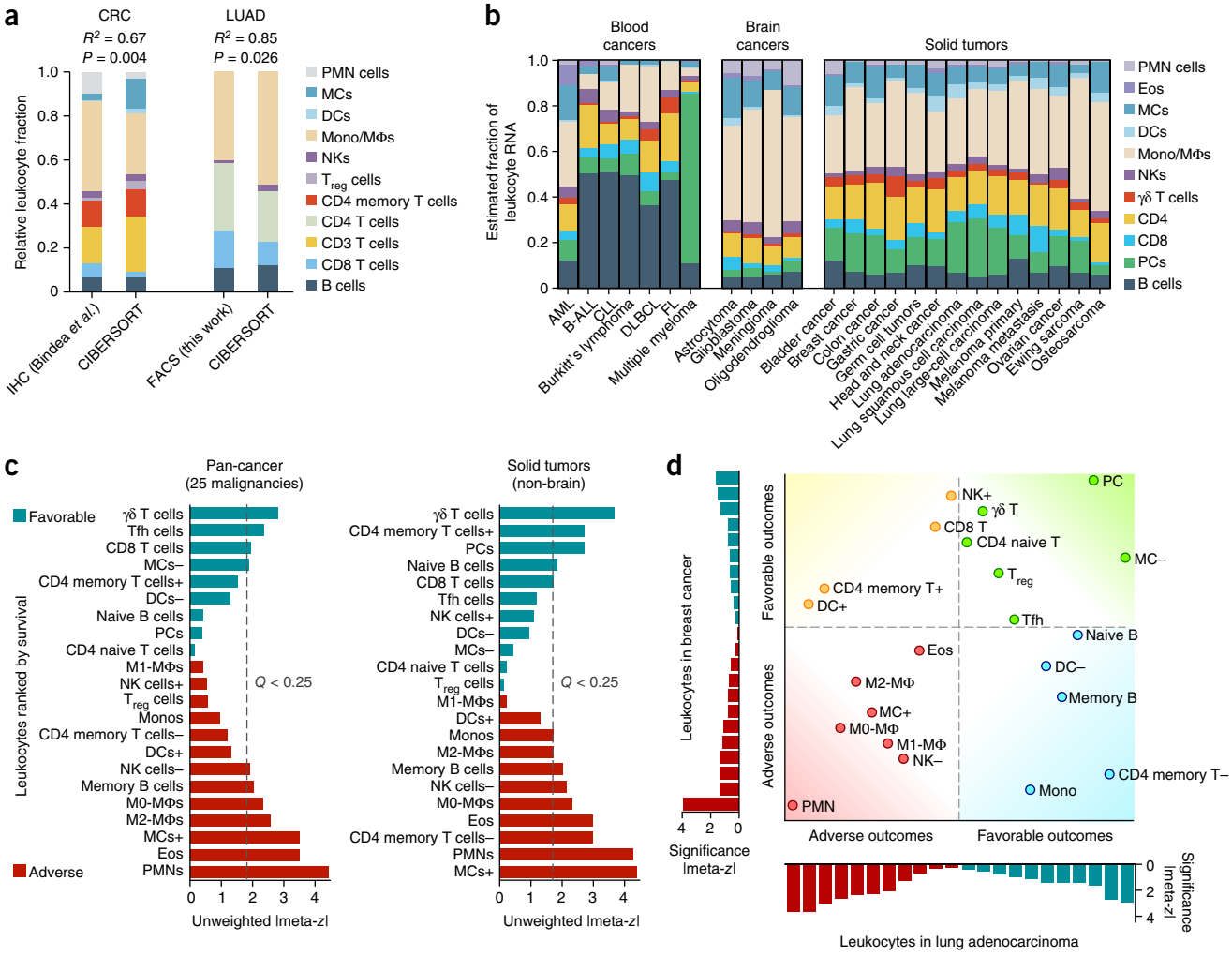

Figure 3 Inferred leukocyte frequencies and prognostic associations in 25 human cancers. (**a**) Relative leukocyte fractions enumerated in solid tumors by CIBERSORT versus immunohistochemical analysis (IHC) or flow cytometry (FACS) on independent samples. CRC, colorectal cancer; LUAD, lung adenocarcinoma. To approximate ground truth proportions in CRC biopsies, levels were inferred by averaging previously reported leukocyte counts from the tumor center and invasive margin of 107 patients (Bindea *et al.*[51]). Baseline leukocyte fractions in LUAD biopsies were enumerated by FACS (n = 13 tumors; data represented as medians; details in Online Methods). CIBERSORT results are represented as mean leukocyte fractions for the corresponding histologies (**Supplementary Data 6**). (**b**) Estimated mRNA fractions of 22 leukocyte subsets across 25 cancers (Affymetrix platforms only; see Online Methods), pooled into 11 immune populations here for clarity (for full details, see **Supplementary Data 6**). (**c**) Global prognostic associations for 22 leukocyte types across 25 cancers (n = 5,782 tumors; left) and 14 solid non-brain tumors (n = 3,238 tumors; right), ranked by unweighted meta-z-score, with a false discovery rate (FDR) threshold of 25% indicated for each plot. Additional FDR thresholds are provided in **Supplementary Figure 6d**. For individual cancers, see **Supplementary Figure 6a**,**c**. (**d**) Concordance and differences in TAL prognostic associations between breast cancers and lung adenocarcinoma (for FDRs, see **Supplementary Fig. 6c**). Resting and activated subsets in **c**,**d** are indicated by – and +, respectively. All leukocyte subset abbreviations are defined in **Supplementary Data 6**. Red and blue bars in **c**,**d** indicate adverse and favorable prognostic associations, respectively.

expression profiles for 22 distinct leukocyte subsets, and defined 'barcodes' of gene expression signatures that robustly distinguish these cell types without requiring cell type–specific marker genes[16]. At a |meta-z-score| > 3.3 (corresponding to two-sided $P < 0.001$), 28% of these barcode genes (152 of 547) are individually significant in PRECOG, out of 2,851 total pan-cancer prognostic genes at the same significance threshold. This is higher than would be expected by random chance ($P < 0.001$, chi-squared test). Whether directly or indirectly compared against flow cytometry and immunohistochemistry data, CIBERSORT exhibited robust performance when applied to solid tumors, accurately estimating relative fractions of leukocyte subsets in colorectal cancer and lung adenocarcinoma (**Fig. 3a**), as well as follicular lymphoma[16].

Applied to PRECOG data, CIBERSORT revealed striking differences in relative leukocyte composition between hematopoietic neoplasms, brain cancers, and non-brain solid tumors (**Fig. 3b** and **Supplementary Data 6**). Variation in TAL content was also consistent and reproducible across independent studies of the same cancer type, including solid tumors (**Supplementary Fig. 5a**). Of note, whereas the majority of tumors profiled within PRECOG were unpurified and uncontrolled with respect to tumor content (**Supplementary Data 1**), CIBERSORT correctly inferred high fractions of plasma cells in multiple myeloma–enriched specimens (**Fig. 3b**). Furthermore, as expected, B cell signatures were found to predominate in B cell malignancies (**Fig. 3b**), suggesting that CIBERSORT has general utility for discerning the cell of origin in diverse cancers.

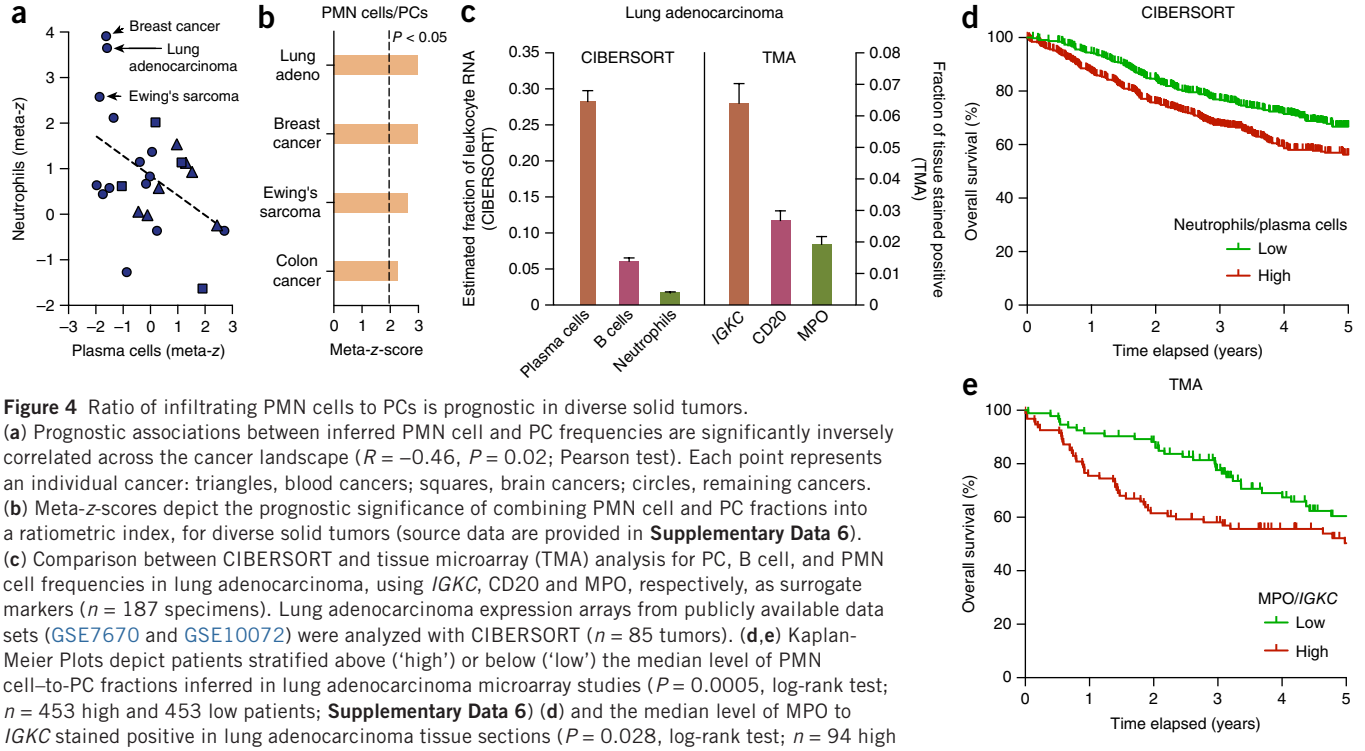

Figure 4 Ratio of infiltrating PMN cells to PCs is prognostic in diverse solid tumors. (**a**) Prognostic associations between inferred PMN cell and PC frequencies are significantly inversely correlated across the cancer landscape ($R = -0.46$, $P = 0.02$; Pearson test). Each point represents an individual cancer: triangles, blood cancers; squares, brain cancers; circles, remaining cancers. (**b**) Meta-*z*-scores depict the prognostic significance of combining PMN cell and PC fractions into a ratiometric index, for diverse solid tumors (source data are provided in **Supplementary Data 6**). (**c**) Comparison between CIBERSORT and tissue microarray (TMA) analysis for PC, B cell, and PMN cell frequencies in lung adenocarcinoma, using *IGKC*, CD20 and MPO, respectively, as surrogate markers ($n = 187$ specimens). Lung adenocarcinoma expression arrays from publicly available data sets (GSE7670 and GSE10072) were analyzed with CIBERSORT ($n = 85$ tumors). (**d**,**e**) Kaplan-Meier Plots depict patients stratified above ('high') or below ('low') the median level of PMN cell–to-PC fractions inferred in lung adenocarcinoma microarray studies ($P = 0.0005$, log-rank test; $n = 453$ high and 453 low patients; **Supplementary Data 6**) (**d**) and the median level of MPO to *IGKC* stained positive in lung adenocarcinoma tissue sections ($P = 0.028$, log-rank test; $n = 94$ high and 93 low patients). (**e**) Hazard ratios were 1.5 (1.2–1.9, 95% CI) for **d** and 1.7 (1.1–2.6, 95% CI) for **e**. Inferred levels of PMN cells to PCs were also significantly prognostic in continuous models assessed by univariate Cox regression in **d** ($P = 0.003$, $z = 2.98$) and **e** ($P = 0.0005$, $z = 3.46$). Data in **c** are presented as means ± s.e.m. All patients were right censored after 5 years in **d** and **e**.

Prognostic associations of TALs

To complement our gene-centric survival analysis, we assembled a global map of prognostic associations for 22 immune populations across human malignancies (**Supplementary Fig. 6a**). We observed considerable variation between cell subsets and cancer-specific outcomes, and many of these associations are statistically significant (**Supplementary Fig. 6b–d**). Pooling cancer types yielded global leukocyte prognostic patterns, in which higher levels of estimated T cell fractions were found to generally correlate with superior survival while increasing levels of myeloid populations primarily correlated with poorer survival. Intra-tumoral γδ T cell[37,38] and polymorphonuclear (PMN) cell[39,40] signatures emerged as the most significant favorable and adverse cancer-wide prognostic populations, respectively (**Fig. 3c**, left). Moreover, when inferred leukocyte fractions were compared with *KLRB1* expression across cancers, γδ T cell and CD8 T cell signatures were most highly correlated (**Supplementary Fig. 5b**), suggesting a link to the prognostic significance of this gene. We found no relationship between estimated PMN cell fractions estimated by CIBERSORT and known necrotic tissue content, suggesting that intra-tumoral PMN cells are not simply a correlate of tissue necrosis[41]. Furthermore, consistent with previous reports[33,42], signatures of tumor-associated M2 macrophages were found to predict worse outcomes than pro-inflammatory M1 macrophages, and anti-CD3/anti-CD28-costimulated, but not resting, CD45RO$^+$ memory helper T cells were correlated with superior outcomes.

Prognostic TALs in solid tumors

By comparing leukocyte survival signatures in breast and lung cancer—two of the most highly profiled cancers in PRECOG—we identified two populations, PMN cells and plasma cells (PCs), with unexpectedly strong yet reciprocal relationships to survival (**Fig. 3d**). PC signatures are significant predictors of favorable survival across solid tumors in general (**Fig. 3c**, right), and we found that they were the most inversely correlated prognostic population to PMN cells (**Fig. 4a**) when assessed globally in a cross-correlation analysis of human cancers (**Supplementary Fig. 5c**). Estimated PC levels were not correlated with tumor stage (**Supplementary Fig. 7a**). Because PC signatures were found to be higher in tumors than in adjacent normal tissues (**Supplementary Fig. 7b**), the prognostic value of tumor-infiltrating PCs is unlikely to be a proxy for general immunological health, thus supporting a role for the antigen-driven processes required for their clonal expansion and emergent humoral immune responses[43]. Furthermore, a simple ratio of estimated PMN cells to PCs was found to be significantly prognostic in diverse solid tumors (**Fig. 4b**).

To experimentally evaluate the reciprocal survival associations of PMN cell and PC signatures, we assessed their infiltration in 187 lung adenocarcinomas using tissue microarray (TMA) analysis (**Supplementary Data 7**). Characteristics of both cell types were observed by H&E staining of tissue sections (**Supplementary Fig. 7c**,**d**), and the presence of tumor-infiltrating plasmacytic cells (i.e., plasmablasts or plasma cells) was confirmed in fresh tumor specimens using both flow cytometry (**Supplementary Fig. 7e**) and morphological assessment (**Supplementary Fig. 7f**). Moreover, we confirmed an elevated presence of plasmacytic cells in non-small-cell lung cancer (NSCLC) tumors, as compared to normal adjacent tissues (**Supplementary Fig. 7g**,**h**). In serial lung adenocarcinoma tissue sections, we stained for the presence of MPO (myeloperoxidase protein) and *IGKC* (immunoglobulin kappa constant RNA), markers of PMN cells and PCs, respectively (**Supplementary Fig. 8a**). Because B cells express varying levels of *IGKC*, we also tested for CD20, a surface protein of mature B cells but not PCs (**Supplementary Fig. 7e**). We found <10% overlap with CD20, indicating the high specificity of

IGKC for PCs (**Supplementary Fig. 8b**). Next, we quantified the staining area for each marker in the tissue array (**Supplementary Fig. 8c**,**d**). Despite operating on differing scales and being measured on independent tumor specimens, fractional levels of these three markers measured *in situ* on TMAs were comparable to the relative infiltrate levels inferred by CIBERSORT (**Fig. 4c**). Moreover, in both continuous and binary models, we found a strong relationship between inferior survival and a higher ratio of PMN cells to PCs in lung adenocarcinoma, whether measured in PRECOG (**Fig. 4d**), in microarray data sets not included in PRECOG (**Supplementary Fig. 8e**), or by surrogate markers in TMA specimens (**Fig. 4e**). Furthermore, TMA results remained significant in multivariate models incorporating relevant clinical parameters (P = 0.002; **Supplementary Data 7**). Together, these data validate our computational approach, and they demonstrate that tumor-associated PMN cells and PCs exhibit opposite associations with overall survival.

Circulating leukocytes, including PMN cells and B lymphocytes, contribute to the tumor microenvironment[44–46], and leukocyte frequencies of innate and adaptive effectors in peripheral blood can have prognostic value[43,47]. Therefore, we examined a subset of NSCLC patients from the TMA with available peri-operative complete blood counts to assess the concordance between levels of circulating leukocytes and TALs. Although intra-tumoral PMN cell–to-PC ratios remained significantly prognostic within this subset (P = 0.035), we found no significant correlation between circulating and infiltrating compartments, and no prognostic value from circulating leukocyte levels (**Supplementary Data 7**).

DISCUSSION

Because expression signatures of small numbers of genes[48] and cells[47] can have utility for predicting response to therapy, including anti-tumor agents, we envision that these resources, available at http://precog.stanford.edu, will be useful for the future discovery of predictive biomarkers, including immunotherapies.

PRECOG has unique advantages over related resources[49]. First, multiple data sets are included for most human cancers, and the use of a robust survival meta-*z* approach to integrate studies reduces the potential for erroneous conclusions drawn from single data sets. For example, in contrast to one recent study[15], we found that molecular markers can indeed add significant prognostic information to clinical variables (see integrated discrimination improvement and net reclassification improvement analyses for *FOXM1-KLRB1* in **Supplementary Data 4**), which is consistent with our previous work[50]. Nevertheless, we acknowledge that prognostic associations remain modest in some cancers (for example, lung squamous cell carcinoma), and the development of clinically applicable molecular models in such cases may remain challenging. Second, although recent studies have inferred TAL activities in relation to patient outcomes[32,51,52], most have focused on one or two cell types in multiple cancer histologies[32,52], have analyzed a relatively small number of markers that are not uniquely expressed by single immune cell types (for example, granzyme A (*GZMA*) and perforin 1 (*PRF1*); **Supplementary Fig. 4b**)[52], or have studied multiple infiltrating immune cell types in a single cancer histology[51].

By applying CIBERSORT, a computational method for inferring leukocyte representation in bulk tumors[16], we identified complex relationships between 22 immune subset signatures and overall survival across 25 cancer histologies. One potential limitation of our approach is the fidelity of input reference profiles, which could deviate from the expression programs and functional states of tumor-associated leukocytes. However, CIBERSORT was developed to be robust with respect to noise, and thus far we have observed strong agreement with ground truth assessments in bulk tumors (see **Fig. 3a** and ref. 16). We identified and validated an intriguing reciprocal prognostic association of PC and PMN cell signatures in diverse and common solid tumors, although the immunological basis for this observation remains unclear. For example, it is unknown whether infiltrating PCs passively reflect nonspecific host immunity or actively contribute to anti-tumor humoral immune responses. Future studies could help define novel tumor-associated antigens targeted by these PCs, with potential implications for new monoclonal antibody therapies. We also speculate that tumor-associated PMN cells may be functionally related to myeloid-derived suppressor cells, which are known to inhibit active T cell anti-tumor immune responses[53]. Indeed, our analysis revealed several broadly favorable prognostic T cell signatures, including γδ and CD8 T cells, whose corresponding cell subsets express *KLRB1* (CD161), the top favorable pan-cancer prognostic gene across PRECOG and a marker of enhanced innate immune characteristics in diverse T cell subsets[28].

The favorable prognostic association of specific TALs across cancers, including effector T cell subsets, is relevant to emerging cancer immunotherapies. Recently, therapeutic antibodies targeting T cell checkpoints, including PD-1 and PD-L1, have yielded unprecedented successes for immunotherapy in some cancers but not others[54]. The immunological diversity underlying the presence or absence of such responses in different cancers and in individual patients remains poorly understood[55]. Our approach provides a framework for characterizing diverse immune effectors as candidate biomarkers for predicting clinical response to these and other novel immunotherapies. It is also amenable to pharmacodynamic measurements that assess the recruitment and/or activation of targeted immune effectors by specific immunotherapies as a means of better understanding treatment responses or failures.

Our results illuminate the prognostic landscape of genes and TALs across human cancers, and they suggest numerous hypotheses for further investigation and clinical translation. Moreover, the data and tools presented here should facilitate a variety of future studies, including the investigation of prognostic biomarkers within molecular subtypes and the assessment of multivariate interactions between genes, TALs, and other clinical indices.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to thank S. Galli, I. Weissman, P. Brown, R. Levy and H. Kohrt for critically reading the manuscript, and members of the Center for Cancer Systems Biology and the Plevritis, Diehn, Levy, and Alizadeh laboratories for valuable guidance and suggestions. This work was supported by grants from the Doris Duke Charitable Foundation (A.A.A.), Damon Runyon Cancer Research Foundation (A.A.A.), V-Foundation (A.A.A.); by the US Public Health Service/National Institutes of Health U01 CA194389 (A.A.A.), R01 CA188298 (M.D. and A.A.A.), U54 CA149145 (S.K.P.), U01CA154969 (S.K.P.), and 5T32 CA09302-35 (A.M.N.); by the Bent & Janet Cardan Oncology Research Fund (A.A.A.); by the Ludwig Institute for Cancer Research (A.A.A.); by a Department of Defense grant W81XWH-12-1-0498 (A.M.N.); and by a grant from the Siebel Stem Cell Institute and the Thomas and Stacey Siebel Foundation (A.M.N.).

AUTHOR CONTRIBUTIONS

A.J.G., S.K.P. and A.A.A. conceived PRECOG, and A.M.N. and A.A.A. conceived immune-PRECOG. A.J.G., A.M.N. and A.A.A. designed the framework, collected and curated the primary data, and developed strategies for implementation and optimizations in related experiments, analyzed the data, and wrote the paper. A.M.N. and A.J.G. wrote all bioinformatics software for PRECOG and related analyses. A.J.G. and C.L.L. implemented web infrastructure for hosting PRECOG. S.V.B., V.S.N., R.B.W. and M.D. curated the NSCLC tumor GEP and TMA data, including clinical annotations. Y.X., A.K. and C.D.H. identified and provided viable NSCLC patient specimens. D.K. and W.F. assisted with flow cytometry characterizations of primary NSCLC tumor specimens and enumeration of corresponding TALs. V.S.N. and R.B.W. constructed the NSCLC TMA and R.B.W. performed *in situ* hybridizations and immunohistochemical characterizations for TALs. A.A.A. and S.K.P. contributed equally as senior authors to supervising and funding the project. All authors discussed the results and their implications, and commented on the manuscript at all stages.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Coussens, L.M. & Werb, Z. Inflammation and cancer. *Nature* **420**, 860–867 (2002).
2. Zhang, L. *et al.* Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N. Engl. J. Med.* **348**, 203–213 (2003).
3. Topalian, S.L. *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* **366**, 2443–2454 (2012).
4. Fan, J.B., Chee, M. & Gunderson, K. Highly parallel genomic assays. *Nat. Rev. Genet.* **7**, 632–644 (2006).
5. Koscielny, S. Why most gene expression signatures of tumors have not been useful in the clinic. *Sci. Transl. Med.* **2**, ps2 (2010).
6. Dupuy, A. & Simon, R.M. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. *J. Natl. Cancer Inst.* **99**, 147–157 (2007).
7. Subramanian, J. & Simon, R. Gene expression–based prognostic signatures in lung cancer: ready for clinical use? *J. Natl. Cancer Inst.* **102**, 464–474 (2010).
8. Dalton, W.S. & Friend, S.H. Cancer biomarkers: an invitation to the table. *Science* **312**, 1165–1168 (2006).
9. Ein-Dor, L., Zuk, O. & Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc. Natl. Acad. Sci. USA* **103**, 5923–5928 (2006).
10. Ransohoff, D.F. Rules of evidence for cancer molecular-marker discovery and validation. *Nat. Rev. Cancer* **4**, 309–314 (2004).
11. Varmus, H. Ten years on: the human genome and medicine. *N. Engl. J. Med.* **362**, 2028–2029 (2010).
12. Lee, H.K., Hsu, A.K., Sajdak, J., Qin, J. & Pavlidis, P. Coexpression analysis of human genes across many microarray data sets. *Genome Res.* **14**, 1085–1094 (2004).
13. Mizuno, H., Kitada, K., Nakai, K. & Sarai, A. PrognoScan: a new database for meta-analysis of the prognostic value of genes. *BMC Med. Genomics* **2**, 18 (2009).
14. Hebestreit, K. *et al.* Leukemia Gene Atlas: a public platform for integrative exploration of genome-wide molecular data. *PLoS ONE* **7**, e39148 (2012).
15. Yuan, Y. *et al.* Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nat. Biotechnol.* **32**, 644–652 (2014).
16. Newman, A.M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
17. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).
18. Mantovani, A. *et al.* Chemokines in the recruitment and shaping of the leukocyte infiltrate of tumors. *Semin. Cancer Biol.* **14**, 155–160 (2004).
19. Coussens, L.M., Zitvogel, L. & Palucka, A.K. Neutralizing tumor-promoting chronic inflammation: a magic bullet? *Science* **339**, 286–291 (2013).
20. Gerstein, M.B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
21. Leek, J.T. *et al.* Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11**, 733–739 (2010).
22. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
23. Newman, A.M. & Cooper, J.B. AutoSOME: a clustering method for identifying gene expression modules without prior knowledge of cluster number. *BMC Bioinformatics* **11**, 117 (2010).
24. Gentles, A.J., Plevritis, S.K., Majeti, R. & Alizadeh, A.A. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *J. Am. Med. Assoc.* **304**, 2706–2715 (2010).
25. Zeuner, A., Todaro, M., Stassi, G. & De Maria, R. Colorectal cancer stem cells: from the crypt to the clinic. *Cell Stem Cell* **15**, 692–705 (2014).
26. Myatt, S.S. & Lam, E.W.-F. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat. Rev. Cancer* **7**, 847–859 (2007).
27. Scholzen, T. & Gerdes, J. The Ki-67 protein: from the known and the unknown. *J. Cell. Physiol.* **182**, 311–322 (2000).
28. Fergusson, J.R. *et al.* CD161 defines a transcriptional and functional phenotype across distinct human T cell lineages. *Cell Rep.* **9**, 1075–1088 (2014).
29. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438–2444 (2010).
30. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
31. Chen, X. *et al.* The forkhead transcription factor FOXM1 controls cell cycle-dependent gene expression through an atypical chromatin binding mechanism. *Mol. Cell. Biol.* **33**, 227–236 (2013).
32. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960–1964 (2006).
33. Fridman, W.H., Pagès, F., Sautès-Fridman, C. & Galon, J. The immune contexture in human tumours: impact on clinical outcome. *Nat. Rev. Cancer* **12**, 298–306 (2012).
34. Lewis, C.E. & Pollard, J.W. Distinct role of macrophages in different tumor microenvironments. *Cancer Res.* **66**, 605–612 (2006).
35. de Visser, K.E., Eichten, A. & Coussens, L.M. Paradoxical roles of the immune system during cancer development. *Nat. Rev. Cancer* **6**, 24–37 (2006).
36. Beyer, M. & Schultze, J.L. Regulatory T cells in cancer. *Blood* **108**, 804–811 (2006).
37. Girardi, M. *et al.* Regulation of cutaneous malignancy by γδ T cells. *Science* **294**, 605 (2001).
38. Haas, W., Pereira, P. & Tonegawa, S. Gamma/delta cells. *Annu. Rev. Immunol.* **11**, 637–685 (1993).
39. Fridlender, Z.G. & Albelda, S.M. Tumor-associated neutrophils: friend or foe? *Carcinogenesis* **33**, 949–955 (2012).
40. Di Carlo, E. *et al.* The intriguing role of polymorphonuclear neutrophils in antitumor reactions. *Blood* **97**, 339–345 (2001).
41. Vakkila, J. & Lotze, M.T. Inflammation and necrosis promote tumour growth. *Nat. Rev. Immunol.* **4**, 641–648 (2004).
42. Sica, A., Schioppa, T., Mantovani, A. & Allavena, P. Tumour-associated macrophages are a distinct M2 polarised population promoting tumour progression: potential targets of anti-cancer therapy. *Eur. J. Cancer* **42**, 717–727 (2006).
43. Teramukai, S. *et al.* Pretreatment neutrophil count as an independent prognostic factor in advanced non-small-cell lung cancer: an analysis of Japan Multinational Trial Organisation LC00–03. *Eur. J. Cancer* **45**, 1950–1958 (2009).
44. de Visser, K.E., Korets, L.V. & Coussens, L.M. *De novo* carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. *Cancer Cell* **7**, 411–423 (2005).
45. Liyanage, U.K. *et al.* Prevalence of regulatory T cells is increased in peripheral blood and tumor microenvironment of patients with pancreas or breast adenocarcinoma. *J. Immunol.* **169**, 2756–2761 (2002).
46. Minárik, I. *et al.* Regulatory T cells, dendritic cells and neutrophils in patients with renal cell carcinoma. *Immunol. Lett.* **152**, 144–150 (2013).
47. Yamanaka, T. *et al.* The baseline ratio of neutrophils to lymphocytes is associated with patient prognosis in advanced gastric cancer. *Oncology* **73**, 215–220 (2007).
48. Paik, S. *et al.* A multigene assay to predict recurrence of Tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**, 2817–2826 (2004).
49. Chen, X., Sun, X. & Hoshida, Y. Survival analysis tools in genomics research. *Hum. Genomics* **8**, 21 (2014).
50. Alizadeh, A.A. *et al.* Prediction of survival in diffuse large B-cell lymphoma based on the expression of 2 genes reflecting tumor and microenvironment. *Blood* **118**, 1350–1358 (2011).
51. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39**, 782–795 (2013).
52. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
53. Gabrilovich, D.I. & Nagaraj, S. Myeloid-derived suppressor cells as regulators of the immune system. *Nat. Rev. Immunol.* **9**, 162–174 (2009).
54. Pardoll, D.M. The blockade of immune checkpoints in cancer immunotherapy. *Nat. Rev. Cancer* **12**, 252–264 (2012).
55. Ribas, A. Tumor immunotherapy directed at PD-1. *N. Engl. J. Med.* **366**, 2517–2519 (2012).

ONLINE METHODS

PRECOG assembly and quality control. To identify cancer gene expression data sets with corresponding patient outcome data we queried the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), European Bioinformatics Institute (EBI) ArrayExpress, the National Cancer Institute (NCI) caArray, and the Stanford Microarray Database for the terms 'survival', 'prognosis', 'prognostic', or 'outcome'. Perl scripts were implemented to download processed and raw data, and associated annotation. For NCBI data, the array platform was determined from the SOFT format file, and the corresponding annotation file was retrieved from GEO. From these, the Probe ID, GenBank accession number, HUGO gene symbol and gene description were extracted based on the internal headers of the SOFT annotation file. The desired fields were specified manually if this automated procedure failed. For older platforms, such as cDNA microarrays, for which annotations had not been recently updated, we re-mapped the probe sequences to HUGO gene symbols via the GenBank or RefSeq accession number through the NCBI Entrez gene identifier. In cases without available accessions, but with the DNA sequence of the probe, we performed the mapping using BLAST-like alignment tool (BLAT) to compare probes to a RefSeq reference and look for unique highest-scoring hits. Scripts were written to extract sample annotation information from GEO SOFT format files and parse them into tables. Since the contents of annotation fields are not semantically enforced, sample data can be contained in various fields, including 'Sample_title', 'Sample_characteristics', 'Sample_description' and 'Sample_source'. Moreover, not all fields are specified for every sample. To parse this information into a tabular format, we attempted to estimate the correct variable name (column header) by searching for common substrings across samples. In some cases, a data set clearly had survival information, but was not deposited with the genomic data. In such cases, we first searched the supplementary information of the corresponding literature for the missing information. Failing this, we contacted corresponding and first authors, of which roughly half supplied the requested data. All tabulations of clinical annotations were further checked and manually curated. This process included verification of results in selected studies by direct comparison of Kaplan-Meier plots and timescales with those in the corresponding primary publications, as well as consistency of prognostic genes across studies. Separately, errors due to technical issues or the curation process were estimated by comparing annotated gender to the ratio of *RPS4Y1* to *XIST* (male:female) expression levels[56] after microarray normalization, as detailed below (**Supplementary Fig. 1a–c**). Furthermore, identical samples present in more than one data set were identified using MD5 checksums for Affymetrix data, and by cross-correlation analysis of expression vectors, and redundant samples were accordingly eliminated.

We applied the following gene expression normalization strategy to allow unification of data from diverse microarray platforms within PRECOG. For Affymetrix GeneChip data, raw CEL files were obtained when possible, and were normalized with the MAS5 algorithm ('affy' package v. 1.26 of Bioconductor v. 1.8 in R 2.15.1), using a custom CDF (Chip Definition File) for probeset summarization, which updates and maps array oligonucleotides to Entrez gene identifiers[57–59] (http://brainarray.mbni.med.umich.edu/Brainarray/). Each data set, regardless of platform, was quantile normalized separately. Moreover, each gene was $\log_2$ transformed if not already in log space, and was then unit mean/variance standardized across samples within a given data set. Although alternative microarray normalization methods have been proposed (for example, RMA[60], gcRMA[61], fRMA[62], SCAN-UPC[63]), for survival analysis we did not observe any significant benefit in comparing Affymetrix data normalized as described above to alternate normalization strategies (data not shown). TCGA RNA-seq and clinical data were downloaded from the TCGA Data Coordinating Center using TCGA-assembler[64]. The gene-level RNA-seq data were preprocessed using TCGA-assembler's 'ProcessRNASeqData' function. RNA-seq and clinical data were matched via the patient barcode provided by TCGA.

For each study, the association of each probe on an array platform with survival outcomes was assessed via Cox proportional hazards regression using the 'coxph' function of the R 'survival' package (v. 2.37). Cox coefficients, hazard ratios with 95% confidence intervals, *P* values, and *z*-scores were obtained for each array probe. For data sets that had not been processed with Custom CDF, which yields a unique per-gene expression value, survival *z*-scores for probes were collapsed to the gene level by averaging the *z*-scores of probes that matched to the same HUGO gene symbol. *Z*-scores for each gene were summarized across all data sets in each malignancy using Lipták's weighted meta-*z* test[65,66], with weights set to the square roots of sample sizes[67]. To identify genes with cancer-wide prognostic significance, and avoid bias due to cancers with different sample sizes, we further combined weighted meta-*z*-scores into a single global meta-*z*-score for each gene using Stouffer's method (unweighted)[66].

Validation of *z*-statistics in PRECOG. Using lung adenocarcinoma as a test case, we assessed the relationship between the weighted meta-*z*-score metric and standard *z*-scores, the latter of which were derived from a merged expression matrix consisting of GEPs from lung adenocarcinoma studies in PRECOG (**Supplementary Data 1**). For this purpose, we selected data sets that had at least 40 stage I samples (indicated in **Supplementary Data 1**). To mitigate batch effects, we standardized each gene in each data set such that it had unit mean and variance across stage I samples. Sample annotations were manually reviewed to ensure that staging corresponded to American Joint Committee on Cancer (AJCC) version 6 (2002), based on TNM (tumor-nodes-metastasis) information. Many data sets pre-dated version 7 of AJCC, and did not contain the required detail for annotating to that standard. These refinements and standardizations permitted merging of samples from different data sets comprising different array platforms and different distributions of tumor stage across the cohort. In all, we compared lung adenocarcinoma GEPs from $n = 1{,}106$ patients, and found that weighted meta-z scores are significantly correlated with merged *z*-scores (Spearman's $R = 0.9$, $P < 2.2 \times 10^{-16}$). We observed similar results when comparing the meta- and merged *z*-statistics for a compendium of five AML studies, thus validating our use of the meta-*z*-statistic. Of note, while we applied batch-correction procedures to merge expression data sets before calculating cross-study *z*-scores, these steps were not necessary with the meta-*z* metric, as *z*-scores from individual studies were directly integrated. This suggests that the meta-*z* approach effectively overcomes batch differences across data sets.

We further evaluated the influence of batch effects[21] within individual data sets using Combat[68] (**Supplementary Fig. 2c–e**). Applied to microarray processing dates in four AML studies, we observed only a modest effect on prognostic *z*-scores, as pre- and post-batch-corrected data were all highly correlated ($R \geq 0.92$, $P < 2.2 \times 10^{-16}$; **Supplementary Fig. 2c**). To test whether batch correction of samples profiled by different study sites would improve data quality, we compared pre- and post-batch-corrected expression data from the NCI director's challenge lung adenocarcinoma data set (ca00182) with a control data set consisting of prognostic meta-*z* scores from a pooled set of all remaining 19 lung adenocarcinoma studies in PRECOG. Little difference in performance was observed for the most prognostic genes, with changes primarily affecting genes whose association with survival outcomes was subtle. (**Supplementary Fig. 2d,e**).

PRECOG false discovery rate. Whereas *z*-scores and meta-*z*-scores were analyzed in this work, *Q* values for global unweighted meta-*z* and weighted cancer-specific meta-*z*-scores were estimated using the FDR method of Storey and Tibshirani[69], and are available for all analyzed *z*-score matrices online (http://precog.stanford.edu). Notably, of 23,288 HUGO gene symbols in PRECOG, 4,385 (19%) have a global meta-*z* significant at $Q < 0.05$ ($|\text{meta-}z| > 2.6$), and 2,986 (13%) are significant at a $Q < 0.01$ ($|\text{meta-}z| > 3.22$) (**Supplementary Data 1**).

Blinding and sample selection criteria. No blinding was used in this work. Duplicate and non-diagnostic (relapse) samples were excluded from analysis.

Statistical analysis of shared prognostic genes. To determine an empirical *P* value for the fraction of overlapping genes in **Figure 1c**, we randomized gene labels for every cancer, and for each cancer, calculated the fraction of prognostic genes shared with at least one other cancer (again, using $|\text{meta-}z| > 3.09$). We then calculated the median fraction of shared genes across all cancers and repeated the analysis 100,000 times. We performed a similar analysis in **Supplementary Figure 3a,b**, but used 1,000 Monte Carlo iterations to test a broad range of *P* value and *Q* value thresholds, considering prognostic genes shared by at least two, three or four tumor types in PRECOG.

doi:10.1038/nm.3909

Clustering of the PRECOG *z*-score matrix. To identify groups of genes with similar prognostic patterns across the entire PRECOG database (166 cancer data sets and 22,461 gene symbols; 827 genes represented in only a small subset of cancers were excluded), we performed unsupervised cluster analysis using AutoSOME[23]. Prior to clustering, columns (i.e., cancer data sets) were pre-processed to unit variance, rows (i.e., genes) were median-centered, and sum of squares = 1 normalization was applied to rows and then columns[23]. AutoSOME was run with a *P* value threshold of 0.01, 100 ensemble iterations, and otherwise default settings. In all, 665 clusters were identified, and ~50% genes were assigned to the 55 largest clusters (**Supplementary Data 3**). Functional annotation for each cluster was assessed using a variety of published gene sets and significance of overlap was determined by a hypergeometric test (for PubMed IDs, see **Supplementary Data 3**). Gene sets with a Bonferroni-corrected $P < 0.01$ were considered significant. To further evaluate each cluster, we calculated a global meta-*z*-score for each gene (as in **Supplementary Data 1**), but considered only absolute meta-*z*-scores to avoid biasing clusters with pan-cancer prognostic genes with the same survival orientation. We then calculated a compound score for each cluster by applying the same formula used for the unweighted meta-*z*-score, which weights cluster-wide meta-*z*-scores by cluster size. Importantly, this formula is not intended to yield a prognostic score in this context, since intra-cluster meta-*z*-scores are not independent variables. Moreover, our approach is not simply a proxy for cluster size, as the five top-ranking clusters were originally ranked 2, 1, 4,13, and 7 with respect to size. The top five clusters ranked by decreasing compound scores are shown in **Figure 1d**.

Cross-validation of top prognostic genes. The utility of PRECOG as a cancer-wide meta-analysis tool was assessed by cross-validation. Cancer histologies (from **Supplementary Data 1**) were randomly selected and pooled in a training set, with the number of cancers *t* ranging from 1 to 31 (of 39) total cancers. For each *t*, the top ten adverse and favorable prognostic genes were identified using the mean meta-*z*-score across the training set, and the mean meta-*z* for the same genes was determined in the remaining $(39 - t)$ cancers (validation set). This process was repeated 100 times for each *t*, and the results are plotted in **Figure 2a**. To determine the top prognostic genes across PRECOG, the top ten adverse and favorable prognostic genes in each validation set were recorded at $t = 31$, and the most frequently occurring genes are shown in **Figure 2a**.

Construction and assessment of a *FOXM1-KLRB1* prognostic model. In brief, we used the coxph function in the R 'survival' package to test the prognostic significance of *FOXM1* and *KLRB1* in bivariate models across PRECOG (each data set was normalized as described for PRECOG construction). Bivariate models with a $P < 0.05$ (Wald test) were considered significant. Derivation of the *FOXM1-KLRB1* composite score is described in detail below.

To integrate *FOXM1* and *KLRB1* into a composite score, we analyzed cancer types with at least two independent data sets having a *FOXM1*, *KLRB1* bivariate $P < 0.05$. Of these, we extracted bivariate coefficients from a randomly selected group of 20% of the corresponding data sets, which served as the training set ($n = 8$; also see **Supplementary Data 4**). The median value of each coefficient in the training set was used as an independent weight in the following formula to define a composite score, $(FOXM1 \times 0.243) + (KLRB1 \times -0.169)$. The median value of the *FOXM1-KLRB1* composite score was determined for each left-out data set (validation set), and used to stratify corresponding patients into high and low risk groups, which were then aggregated by cancer type, and subjected to Kaplan-Meier analysis (**Fig. 2c**). Importantly, each cancer type in **Figure 2c** only includes left-out data sets (i.e., validation data) for which *FOXM1* and *KLRB1* expression values were both available (see **Supplementary Data 1** for data sets used). Moreover, the composite score was not re-optimized for RNA-seq when applied to TCGA data.

PRECOG network connectivity analysis. We evaluated the functional coherence of top prognostic genes in PRECOG by interrogation of human PPAs in STRING version 9.0 (ref. 70). STRING integrates evidence from multiple sources, including curated databases, experimentally confirmed physical interactions, co-expression data, and associations inferred via text mining[70]. We compared the connectivity of STRING networks (for edges with a combined confidence level of at least 0.4) for the top 100 adverse and favorable prognostic genes identified from all cancers (by mean meta-*z*-score), individual cancers (meta-*z*-score), and individual data sets (*z*-score). For each set of proteins, the largest connected component was assessed by two metrics: (i) the average number of associations (i.e., edges) per protein (i.e., node), and (ii) algebraic connectivity, a graphed theoretic measure of overall network connectedness[71]. Only networks with at least ten nodes were considered. Functional enrichment of global networks shown in **Figure 2d** was performed with ToppFun[72], and *P* values were determined using a Benjamini-Hochberg–corrected hypergeometric test.

Gene set analysis of PRECOG clusters and ranked *z*-scores. Enrichment of biological processes with respect to survival *z*-scores was assessed in two ways. First, cluster memberships defined by AutoSOME were compared to pre-defined gene sets by hypergeometric test. Comparison gene sets were obtained from the Molecular Signatures Database (mSigDB)[73], and additional sets of genes that are targets of specific transcription factors were extracted from CHEA[29] and ENCODE[20]. Ranked lists of gene survival *z*-scores were also analyzed with respect to these gene sets using the 'PreRanked' tool of Gene Set Enrichment Analysis[74].

Inferring TAL levels in bulk tumor GEPs. The samples profiled within PRECOG primarily represent bulk diagnostic pre-therapy tumor specimens, which often contain a variety of cell types, including diverse TALs. Given the enrichment of lymphocyte markers in favorably prognostic genes across PRECOG (**Figs. 1d** and **2d**), a method to systematically 'unmix' or deconvolve bulk tumor GEPs in PRECOG may reveal new insights into tumor immunobiology. We recently developed a new approach for CIBERSORT, a machine-learning method that outperformed other approaches in benchmarking experiments[16]. CIBERSORT produces an empirical *P* value for the deconvolution using Monte Carlo sampling. Like other linear deconvolution methods, CIBERSORT only operates on expression values in non-log linear space[75].

TAL heterogeneity and prognostic associations. CIBERSORT was applied to all normalized PRECOG GEPs from Affymetrix HGU133 platforms (57 studies and 25 cancers; **Supplementary Data 6**). In all, 5,782 tumor GEPs were successfully deconvolved (CIBERSORT $P < 0.005$). For each data set, estimated mRNA fractions of each leukocyte subset were related to survival using univariate Cox regression. Weighted meta-*z*-scores were determined using the same approach described for PRECOG in order to build an immune-centric version of PRECOG (iPRECOG, **Supplementary Fig. 6a**), and unweighted global meta-*z*-scores were used to summarize pan-cancer leukocyte associations in **Figure 3c**.

Immune-PRECOG false discovery rate. To differentiate real from stochastic variation in inferred leukocyte prognostic associations, we first compared *P* values and meta-*z*-scores in immune-PRECOG (**Supplementary Fig. 6b**), as any deviation from a standard normal distribution must be considered when drawing statistical conclusions. We generated 1,000 null meta-*z*-score matrices by (i) shuffling the cell type fractions inferred for each data set in **Supplementary Data 6**, and (ii) computing *z*-scores and corresponding meta-*z*-scores to capture relationships to overall survival. We found a tight correspondence between the distribution of null meta-*z*-scores and a standard normal distribution (**Supplementary Figure 6b**). Having validated the normality of the meta-*z*-score, we then filtered **Supplementary Figure 6a** using a range of statistical significance thresholds, and at each cutoff, compared the observed versus expected fractions for all leukocyte prognostic associations (**Supplementary Fig. 6c**). At a two-sided *P* value threshold of 0.05 ($|z| > 1.96$), we found nearly three times more prognostic associations than would be expected by random chance; at $P < 0.01$, there is a fivefold enrichment, which continues to increase with lower *P* value cutoffs (**Supplementary Fig. 6c**).

Separately, we performed a similar analysis on the global meta-*z*-scores shown in **Figure 3c**. Here, we integrated the null meta-*z* scores from **Supplementary Figure 6c** into null global meta-*z*-scores and recomputed the analysis shown for pan-cancer leukocyte prognostic associations (plotted as the fraction of leukocyte subsets retained at different significance thresholds; **Supplementary Fig. 6d**). Taken together, these results explicitly quantify significant versus stochastic variation in leukocyte prognostic associations at different statistical cutoffs, and they allow others to tune the nominal statistical threshold to achieve a desired FDR.

Relative PMN cell fractions versus necrotic tissue content. Relative RNA fractions of PMN cells inferred by CIBERSORT were not correlated with annotated necrotic content in lung squamous cell carcinoma (TCGA; $R^2 = 0.01$; P = nonsignificant) or melanoma (microarray data set GSE8401 (ref. 76); $R^2 \sim 0$; P = nonsignificant).

Flow cytometry versus CIBERSORT. Flow cytometry analysis of non-small cell lung cancer tumor ($n = 13$) specimens was performed as described below, and median fractions of $CD4^+$, $CD8^+$, $CD19^+$, $CD56^+$, and $CD14^+$ populations were normalized by overall $CD45^+$ content (**Fig. 3a**). For comparison with CIBERSORT, leukocyte signature matrix populations were grouped into the same cluster of differentiation categories: $CD14^+$, monocytes, macrophages, and dendritic cells; $CD4^+$, all T cell subsets except CD8 and γδ T cells; $CD8^+$, CD8 T cells; $CD19^+$, naïve and memory B cells, $CD56^+$, resting and activated NK cells. Median CIBERSORT-inferred fractions for lung adenocarcinoma GEPs, shown in **Figure 3a**, were determined from two publicly available microarray data sets, GSE7670 (ref. 77) and GSE10072 (ref. 78),

Patient samples. All aspects of this study were approved by the Stanford Institutional Review Board in accordance with the Declaration of Helsinki guidelines for the ethical conduct of research, and all patients involved provided informed consent. For **Figure 3a**, fresh human lung tumor samples were obtained from the Stanford Tissue Bank. For tissue microarray analyses (**Fig. 4c**,**e** and **Supplementary Fig. 7c–h**), patient samples were retrieved from the surgical pathology archives at the Stanford Department of Pathology and linked to a clinical database using the Cancer Center Database and STRIDE Database tools from Stanford.

Human lung dissociation and flow cytometry. Fresh human lung tumor samples were cut into small pieces and dissociated into single cell suspensions by 45 min of Collagenase I (STEMCELL Technologies) digestion. Dissociated single cells were suspended at 1×10^7 per ml in staining buffer (HBSS with 2% heat-inactivated calf serum). After 10 min of blocking with 10 μg μl^{-1} rat IgG, the cells were stained for at least 10 min with the antibodies listed below. After washing, stained cells were re-suspended in staining buffer with 1μg/ml DAPI, analyzed, and sorted with a FACS Aria II cell sorter (BD Biosciences). Antibodies used for experiments related to **Figure 3a**: CD45-A700 (cat. no. cat. no. 304024), CD14-PE (cat. no. 325606), CD8-APC (cat. no. 300912), CD4-FITC (cat. no. 357406), CD56-PE-cy7 (cat. no. 318318), and CD19-PerCP-cy5.5 (cat. no. 302230). Antibodies used for enumeration of plasmacytic cells: CD45-PE-cy7 (cat. no. 304016), CD20-PerCP-cy5.5 (cat. no. RUO-340508), CD138-PE (cat. no. 325305), CD38-APC (cat. no. 303509), CD19-A700 (cat. no. 302225), and CD27-FITC (cat. no. 302806). All antibodies were obtained from BioLegend.

Tissue microarray (TMA) cohort. We reviewed patients with lung cancer who had surgically treated disease and paraffin-embedded tumor biopsies from 1995 through June, 2010 for inclusion. Patients with recurrent or metastatic disease samples only were excluded. Medical charts were reviewed to clinically annotate the tumor specimens with demographic, operative procedures, imaging data, and follow-up. Pathology reports were reviewed to confirm specimen type, site, pathology, stage, histology, invasion status and operative procedure. Treated samples (neoadjuvant therapy) were excluded, resulting in a final analysis cohort of 187 pre-treated lung adenocarcinoma tumor specimens with follow-up data.

TMA cohort follow-up. Recurrence was defined by imaging or biopsy and patients with advanced disease or who did not have at least 6 months of follow-up were censored for further analyses. The National Death Index (NDI) was used to define vital status through October 30, 2010. Patients not indicated as dead according to the NDI were assumed to be alive except for those who had left the country or were from other countries (who were censored), as the NDI relies on a social security number for vital status assessment. Synchronous tumors resected over time were eligible for prognostic assessment in patients with two primaries.

TMA construction. The Stanford Lung Cancer TMA was developed from surgical specimens that contained viable tumor from duplicate slides that were reviewed by a board-certified pathologist (R.B.W.). The pathologist was not blinded to sample identity. The area of highest tumor content was marked for coring blocks corresponding to the slides. We used 2-mm cores to build the tissue microarray. These cores were aligned by histology and stage and negative controls were taken from the West Lab and included a variety of benign and malignant tissues (65 cores) that included normal non-lung tissue (12 cores), abnormal non-lung tissue (13 cores), placental markers (23 cores) and normal lung (17 cores). Normal lung consisted of a specimen adjacent to but distinct from tumor over the years 1995–2010 in order to assess the variability of staining by year. OligoDT analysis was performed on the finished array to assess the architecture of selected cores and adequacy of tissue content before target IHC analysis. A co-registered H&E slide was used as well to verify tumor location for cases in which this was unclear on initial inspection.

TMA immunohistochemistry. MPO (DAKO) and CD20 (clone L26, DAKO) immunohistochemistry performed on 4 mm sections using the Ventana BenchMark XT automated immunostaining platform (Ventana Medical Systems/Roche, Tucson, AZ).

TMA RNA *in situ* hybridization. The RNA *in situ* hybridization probe for *IGKC* was designed against chr2: 88,937,790–88,938,290 (hg18) using primer 5′-CTG TTG TGT GCC TGC TGA AT-3′ and the T7 promoter-tagged primer 5′-CTA ATA CGA CTC ACT ATA GGG TTA AAG CCA AGG AGG AGG AG-3′. RNA *in situ* hybridizations were performed on TA369, as described previously[79].

TMA microscopy. Eleven slides were scanned at 20× on an Ariol imaging analysis system (originally built by Applied Imaging).

TMA staining quantification and analysis. To facilitate consistency and reproducibility in quantitating TMA staining patterns, we evaluated the performance of GemIdent[80], a supervised *in silico* image segmentation system. As an initial exercise, we trained GemIdent on a single lung adenocarcinoma specimen to recognize both *IGKC* stains and non-tissue background (white space). GemIdent was then applied to 10 TMA specimens to generate separate image masks of both *IGKC* localization and non-tissue background (i.e., 'empty space'). A custom Perl script was used to process each image mask and quantify the staining area of *IGKC* for each specimen (by first removing non-tissue white space to calculate the surface area of each tissue). To test the utility of this approach, a board-certified pathologist (R.B.W.) scored *IGKC* for the same 10 specimens. The pathologist had no knowledge of the results from automated staining, but was not blinded to sample identity. Both assessments were highly correlated ($R^2 = 0.98$; **Supplementary Fig. 8c**). In a separate exercise, two independent operators trained GemIdent on distinct CD20-stained specimens. CD20-stained fractions were then quantified across the entire TMA ($n = 187$ lung adenocarcinomas) and results were processed as described above. The concordance between independent operators was very high ($R^2 \sim 1$; **Supplementary Fig. 8d**). These data support the utility of GemIdent coupled with image post-processing for automated scoring of TMA specimens. We applied this approach to quantitatively score *IGKC*, CD20, and MPO for all lung adenocarcinoma TMA specimens (for example, see **Supplementary Fig. 8a**).

Comparison between TALs and circulating leukocytes. Among patients with available perioperative circulating leukocyte (lymphocyte and PMN cells) counts, we analyzed the sample closest to the date of procedure within −120 to +28 d, where precedence was given to preoperative samples (total $n = 46$ lung adenocarcinoma patients). As shown in **Supplementary Data 7**, no relationships were found between circulating leukocyte levels and TALs quantified on the TMA. Moreover, while the ratio of MPO to *IGKC* levels remained significantly prognostic within this patient subset ($P = 0.035$), circulating leukocyte levels had no significant relationship to survival.

56. Day, A., Carlson, M.R., Dong, J., O'Connor, B.D. & Nelson, S.F. Celsius: a community resource for Affymetrix microarray data. *Genome Biol.* **8**, R112 (2007).
57. Dai, M. *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **33**, e175 (2005).

doi:10.1038/nm.3909

58. Gautier, L., Cope, L., Bolstad, B.M. & Irizarry, R.A. affy—analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
59. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
60. Irizarry, R.A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
61. Wu, Z., Irizarry, R.A., Gentleman, R., Martinez-Murillo, F. & Spencer, F. A model-based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.* **99**, 909–917 (2004).
62. McCall, M.N., Bolstad, B.M. & Irizarry, R.A. Frozen robust multiarray analysis (fRMA). *Biostatistics* **11**, 242–253 (2010).
63. Piccolo, S.R. *et al.* A single-sample microarray normalization method to facilitate personalized-medicine workflows. *Genomics* **100**, 337–344 (2012).
64. Zhu, Y., Qiu, P. & Ji, Y. TCGA-Assembler: open-source software for retrieving and processing TCGA data. *Nat. Methods* **11**, 599–600 (2014).
65. Lipták, T. On the combination of independent tests. *Magyar Tud. Akad. Mat. Kutato Int. Közl* **3**, 171–196 (1958).
66. Stouffer, S., DeVinney, L. & Suchmen, E. *The American Soldier: Adjustment During Army Life* (Princeton University Press, 1949).
67. Zaykin, D.V. Optimally weighted *Z*-test is a powerful method for combining probabilities in meta-analysis. *J. Evol. Biol.* **24**, 1836–1841 (2011).
68. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
69. Storey, J.D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440–9445 (2003).
70. Szklarczyk, D. *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* **39**, D561–D568 (2011).
71. Fiedler, M. Algebraic connectivity of graphs. *Czech. Math. J.* **23**, 298–305 (1973).
72. Chen, J., Bardes, E.E., Aronow, B.J. & Jegga, A.G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
73. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
74. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
75. Zhong, Y. & Liu, Z. Gene expression deconvolution in linear space. *Nat. Methods* **9**, 8–9 (2012).
76. Xu, L. *et al.* Gene expression changes in an animal melanoma model correlate with Aggressiveness of Human Melanoma Metastases. *Mol. Cancer Res.* **6**, 760–769 (2008).
77. Su, L.-J. *et al.* Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics* **8**, 140 (2007).
78. Landi, M.T. *et al.* Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE* **3**, e1651 (2008).
79. Brunner, A.L. *et al.* Transcriptional profiling of long non-coding RNAs and novel transcribed regions across a diverse panel of archived human cancers. *Genome Biol.* **13**, R75 (2012).
80. Holmes, S., Kapelner, A. & Lee, P.P. An interactive java statistical image segmentation system: Gemident. *J. Stat. Softw.* **30**, i10 (2009).