The vascular landscape of human cancer

Benjamin M. Kahn, … , Robert B. Faryabi, Ben Z. Stanger

Research Article **Angiogenesis** **Oncology**

Graphical abstract
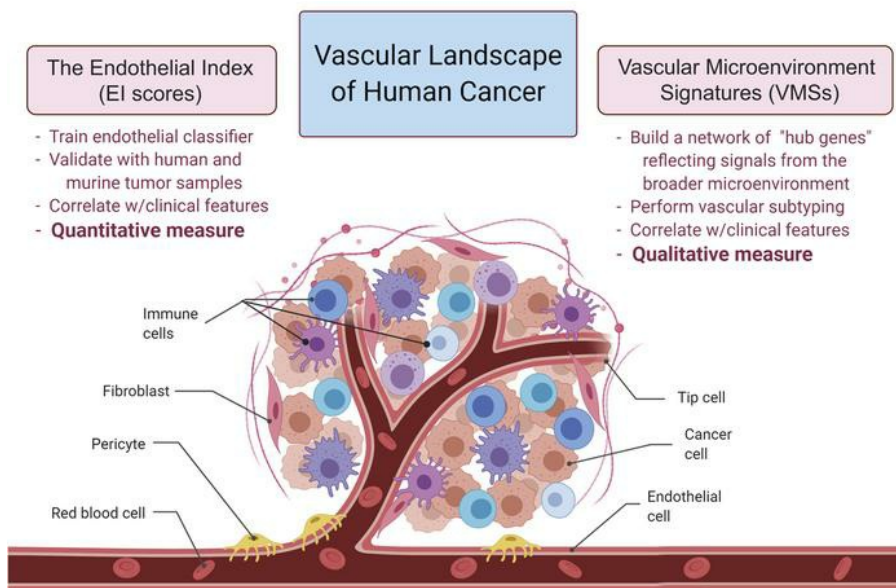
The vascular landscape of human cancer

Benjamin M. Kahn,[1,2,3,4] Alfredo Lucas,[1,2,3,4] Rohan G. Alur,[1,2,3,4] Maximillian D. Wengyn,[1,2,3,4] Gregory W. Schwartz,[3,4,5,6,7] Jinyang Li,[1,2,3,4] Kathryn Sun,[1,2,3,4] H. Carlo Maurer,[8] Kenneth P. Olive,[8] Robert B. Faryabi,[3,4,5,6,7] and Ben Z. Stanger[1,2,3,4]

[1]Department of Medicine, [2]Department of Cell and Developmental Biology, [3]Abramson Family Cancer Research Institute, [4]Abramson Cancer Center, [5]Department of Pathology and Laboratory Medicine, [6]Penn Epigenetics Institute, and [7]Department of Cancer Biology Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA, [8]Department of Medicine, Division of Digestive Liver Diseases and Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, New York, USA.

Tumors depend on a blood supply to deliver oxygen and nutrients, making tumor vasculature an attractive anticancer target. However, only a fraction of patients with cancer benefit from angiogenesis inhibitors. Whether antiangiogenic therapy would be more effective if targeted to individuals with specific tumor characteristics is unknown. To better characterize the tumor vascular environment both within and between cancer types, we developed a standardized metric — the endothelial index (EI) — to estimate vascular density in over 10,000 human tumors, corresponding to 31 solid tumor types, from transcriptome data. We then used this index to compare hyper- and hypovascular tumors, enabling the classification of human tumors into 6 vascular microenvironment signatures (VMSs) based on the expression of a panel of 24 vascular "hub" genes. The EI and VMS correlated with known tumor vascular features and were independently associated with prognosis in certain cancer types. Retrospective testing of clinical trial data identified VMS2 classification as a powerful biomarker for response to bevacizumab. Thus, we believe our studies provide an unbiased picture of human tumor vasculature that may enable more precise deployment of antiangiogenesis therapy.

Introduction

The pathways by which tumors derive a blood supply are attractive anticancer targets (1–3). However, only a fraction of patients derive benefit from antiangiogenesis therapies, and drugs that inhibit angiogenesis — particularly those that block *VEGFA-VEGF* receptor interactions — are approved for a select handful of cancer types (4). Heterogeneity poses a major challenge to the development of effective therapeutics, as diverse (and likely redundant) drivers of vessel outgrowth may operate across different tumor types and individual tumors (4–6). Thus, a better understanding of the features that distinguish the vascular properties of tumors across cancer is needed to inform precision-based antiangiogenesis strategies.

Clinical measurements of tumor vascularity have traditionally relied on immunohistochemical estimates of microvessel density (MVD). However, this approach can be confounded by methodological variability and limited tissue availability and quality; thus, the literature regarding vascular properties of human tumors contains inconsistent and sometimes contradictory findings (7–10). Moreover, these methods fail to leverage the vast amounts of transcriptional data now available through consortia such as The Cancer Genome Atlas (TCGA), which can be used to correlate a tumor's phenotype with its molecular features. To systematical-

ly evaluate tumor vasculature at the level of individual human tumors, we undertook a comprehensive computational approach to predict vascular density and extract further clinical and molecular correlates across human cancer. We believe our studies put forth an unbiased approach for molecular classification of tumor vascular environments that may guide targeted antiangiogenic therapies and ultimately improve their efficacy.

Results

A classifier for estimating tumor vascularity. To better define the relationship between a tumor's vascularity and its molecular features, we developed an in silico algorithm for predicting vascular density from bulk tumor mRNA expression data. Blood and lymphatic vessels are lined by endothelial cells, a highly specialized cell type that can be identified by the expression of cell-specific genes. Thus, a quantitative measure for endothelial cell content is predicted to reflect vascularity in a tissue or tumor.

To generate such an "endothelial classifier," we first curated a training set of 176 normal human tissue microarray samples that mimicked the cellular complexity and content of the tumor microenvironment (Figure 1A and Supplemental Data File 1; supplemental material available online with this article; https://doi.org/10.1172/JCI136655DS1). The training set was divided into an endothelial group (*n* = 45), consisting of endothelial cells sourced from various dissected microvessels and arteries, and a nonendothelial group (*n* = 131), consisting of various immune, stromal, epithelial, and other cell types commonly present in the tumor microenvironment. We then trained an ElasticNet classifier to discriminate the 2 groups (Figure 1A and Supplemental Figure 1A). ElasticNet is a penalized logistic regression classifier and was used because it selects a small and accessible set of genes that best sep-
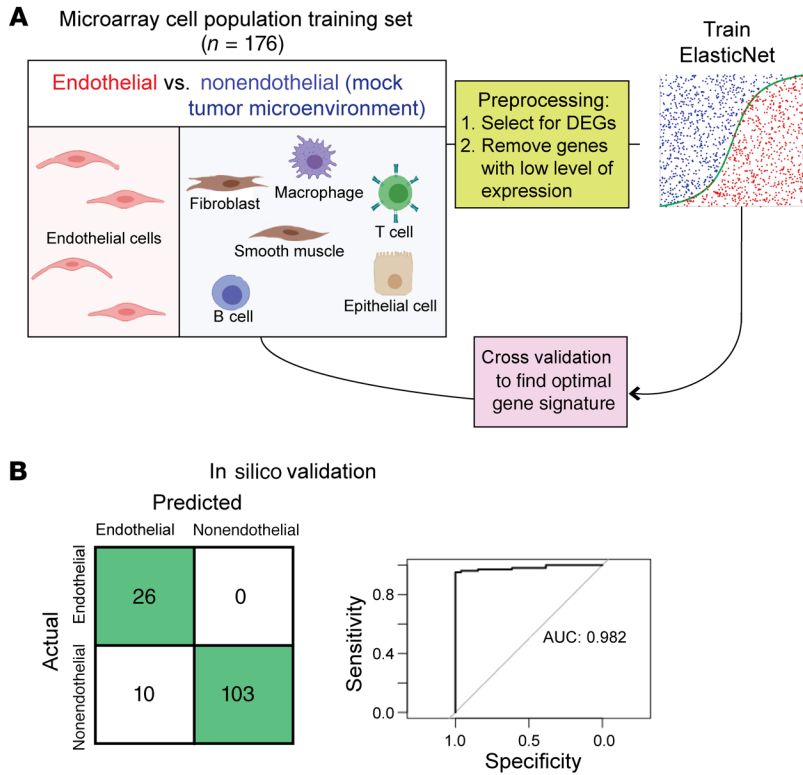
Figure 1. A 7-gene transcriptional classifier, the EI, recognizes endothelial cells. (**A**) Schematic of the Elastic-Net classifier used to design an endothelial classifier – the EI – consisting of *ANGPT2*, *CDH5*, *ESAM*, *ESM1*, *ERG*, *ICAM2*, and *TIE1*. (**B**) Results from in silico validation on an independent test set of 139 microarrays. The EI only committed false-positive errors in testing, with an overall prediction accuracy of 92% in silico and an area under the ROC curve of 0.982.

arate the endothelial from nonendothelial samples. Seven genes constituted the features of an "endothelial" versus "nonendothelial" classifier: vascular-endothelial cadherin (*CDH5*), angiopoietin-2 (*ANGPT2*), ETS-related gene (*ERG*), endothelial cell–specific adhesion molecule (*ESAM*), endothelial cell–specific molecule 1 (*ESM1*), intercellular adhesion molecule 2 (*ICAM2*), and tyrosine kinase with immunoglobulin-like domains 1 (*TIE1*) (Table 1).

To assess the performance of the 7-gene endothelial classifier, which we termed the endothelial index (EI), we used an independent test set of 139 microarrays ($n = 26$ endothelial; $n = 113$ nonendothelial; Supplemental Data File 1). When applied to this data set, the EI had an AUC of 0.982 (accuracy of 92%), committing only false-positive errors (Figure 1B and Supplemental Figure 1, B–E). Furthermore, interrogation of the RefExA human tissue database (www.lsbm.org/site_e/database/index.html) — a collection of expression data from normal human tissues, cultured cells, and cancer cell lines — revealed the expression of all 7 classifier genes to be either enriched in or exclusive to endothelial cell lines (Supplemental Figure 2). Thus, the EI can readily discriminate endothelial cells from other cell types.

In vivo validation of the EI. To independently evaluate the performance of the EI, we used 2 resources in which matched RNA-Seq data and histological sections were available from the same tumors, enabling a comparison of microvessel density (MVD) predicted by the EI with that directly measured by immunostaining. First, we obtained samples of human pancreatic ductal adenocarcinoma (PDAC), in which laser capture microdissection (LCM) was used to generate separate gene expression profiles for epithelial and stromal compartments from 60 tumors (11) (Figure 2A and Supplemental Figure 3A). EI scores (range from 0 to 1) were independently calculated for each compartment. Samples with an

EI of greater than 0.9 were classified as "hypervascular," whereas tumors with an EI of less than 0.1 were classified as "hypovascular" (the remainder being classified as "intermediate"). Within samples derived from the stromal compartment, 9% of tumors were classified as hypervascular, 51% as hypovascular, and 40% as intermediate (Figure 2B), consistent with previous reports that most pancreatic tumors are poorly vascularized (12–15). As expected, almost all samples derived from the epithelial compartment had EI scores near zero (Supplemental Figure 3B). We then directly measured MVD by staining for *CD31*, an endothelial cell marker, in the 44 tumors for which slides were available (Figure 2, C and D, and Supplemental Figure 4). The EI score tracked closely with *CD31* staining in all tumors tested ($R = 0.48$, $P = 0.0009$), with PDACs classified as hypervascular harboring, on average, a MVD twice that of those classified as hypovascular ($P = 0.01$) (Figure 2, E and F). These results suggest that estimates of tumor vascularity from stromal RNA-Seq data (using the EI score) are highly predictive of actual vessel density.

To further assess the performance of the EI, we used a panel of congenic C57BL/6 tumor cell clones derived from the *Kras*[LSL-G12D/+] *Trp53*[LSL-R172H/+] *Pdx1-Cre Rosa26*[LSL-YFP] (KPCY) mouse model of PDAC (Figure 2G) (16). Using the mouse orthologs of the 7 genes in the human endothelial classifier, we calculated EI scores from tumor RNA-Seq data for each of the 19 clones: 3 tumors (16%) were classified as hypervascular, 12 tumors (63%) were classified as hypovascular, and 4 tumors (21%) were classified as intermediate (Figure 2H). We then determined MVD by staining for the endothelial cell marker endomucin (*EMCN*) (Figure 2, I and J). As with the human samples, the EI score tracked closely with MVD in all 19 murine tumors, with hypervascular tumors having a vessel density twice that of hypovas-

Table 1. EI genes

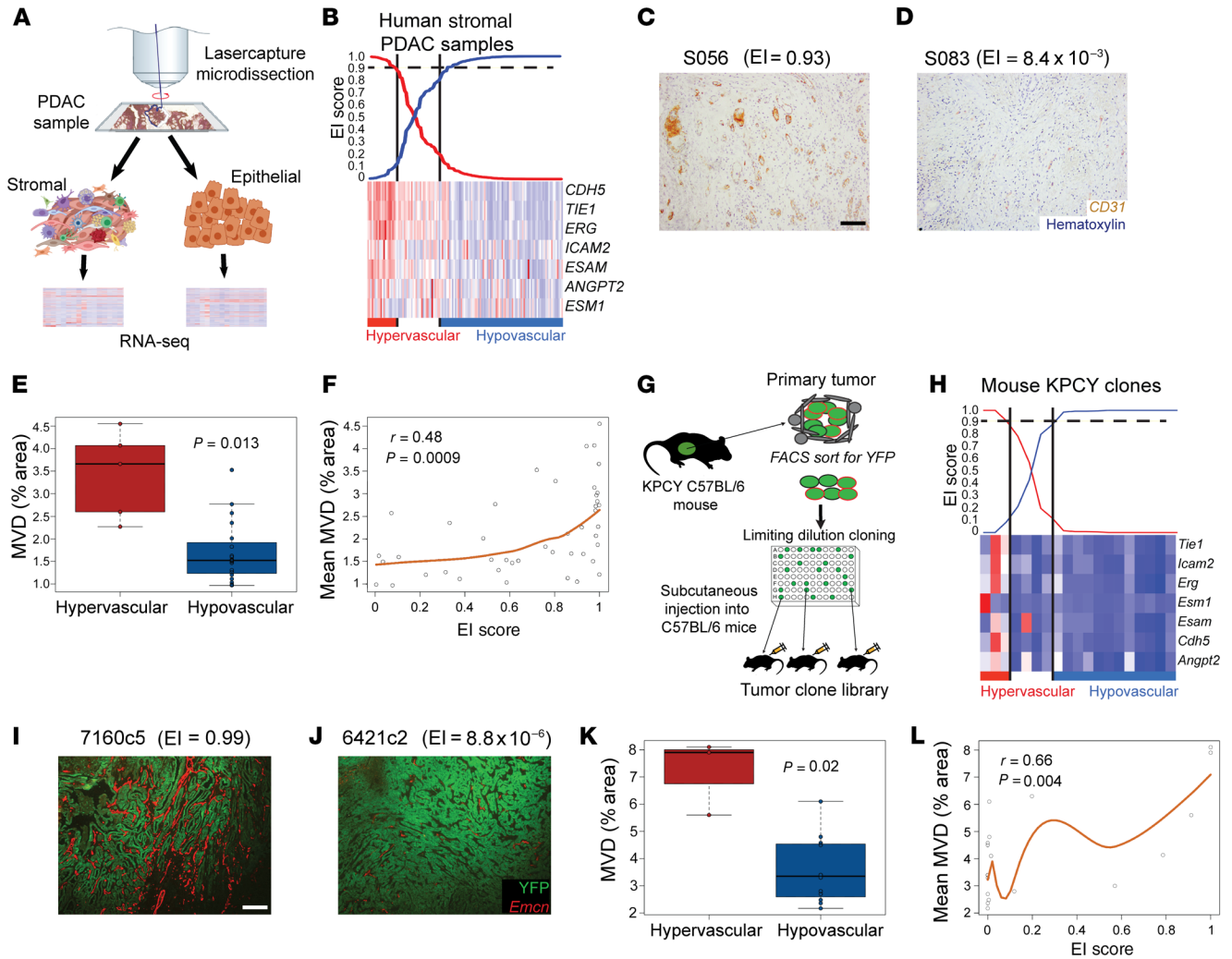| Gene	Full name
ANGPT2	Angiopoietin 2
CDH5	Cadherin 5
ERG	ETS-related gene
ESAM	Endothelial cell–selective adhesion molecule
ESM1	Endothelial cell–specific molecule 1
ICAM2	Intracellular adhesion molecule 2
TIE1	Tyrosine kinase with immunoglobulin-like domains 1

Figure 2. The EI is a transcriptional predictor of tumor vascularity. (**A**) Schematic showing the separation of stroma and epithelium from 124 PDAC samples by LCM followed by RNA-Seq. (**B**) EI scores plot, ranging from 0 to 1, derived by applying the classifier to RNA-Seq data generated from 124 stromal PDAC samples. For visualization of hyper- and hypovascular classes, both the EI score (red line) and the 1-EI score (blue line) representing the epithelial and nonepithelial probabilities, respectively, are plotted. (**C** and **D**) Representative images showing immunohistochemical staining for CD31 in tumors predicted by the classifier to be hypervascular (**C**) or hypovascular (**D**). (**E** and **F**) MVD quantification plotted as box plots for 24 tumors classified as either hypervascular or hypovascular (**E**) or by smoothened Loess regression for all 44 tumors in which CD31 staining was performed (**F**). (**G**) Schematic of the congenic tumor clone library generated from PDACs arising in KPCY C57BL/6 mice (16). (**H**) EI scores plot, ranging from 0 to 1, derived by applying the classifier to RNA-Seq data generated from 19 clonal tumors. (**I** and **J**) Representative images showing immunohistochemical staining for endomucin (*EMCN*) in tumors predicted by the classifier to be hypervascular (**I**) or hypovascular (**J**). (**K** and **L**) MVD quantification shown as box plots for 15 tumors classified as either hypervascular or hypovascular (**K**) or by smoothened Loess regression for all 19 tumors in which endomucin staining was performed (includes tumors with an intermediate vascular status) (**L**). An unpaired Student's *t* test was used for comparisons of MVD density between hypervascular and hypovascular tumors (**E** and **K**). Pearson's correlation coefficient and the corresponding 2-tailed probability values were calculated to test the relationship between the EI score and MVD in all tumors (**F** and **L**). In the box plots, the center line marks the median, the box limits span the IQR, and the whiskers extend 1.5 times the IQR range. Scale bars: 100 μm.

cular tumors (Figure 2, K and L). Taken together, these results suggest that the EI score can provide an accurate estimate of MVD from gene expression measured by different technologies (microarray and RNA-Seq) in both human and murine tumors.

Vessel quantitation across human solid tumors. We next used the EI to determine the extent to which MVD differs across human tumors. EI scores were calculated for 10,767 solid tumors from 31 distinct cancer types using RNA-Seq data available through TCGA (Supplemental Data File 2), which revealed considerable variability between and within cancer types (Figure 3A). Among

all tumor types, kidney renal clear cell carcinoma (KIRC) had the highest median EI score, consistent with the known hypervascular features of this tumor type (17). Likewise, thyroid cancers and paragangliomas had the fourth and fifth highest median EI scores, consistent with prior reports that these tumors are typically hypervascular (18–22). At the other end of the vascular spectrum, PDAC had the lowest median EI score, consistent with the known hypovascular features of most pancreatic cancers (12, 14, 15). Nevertheless, approximately 12% of pancreatic tumors were predicted to be hypervascular (EI score >0.9), a finding consistent with previously
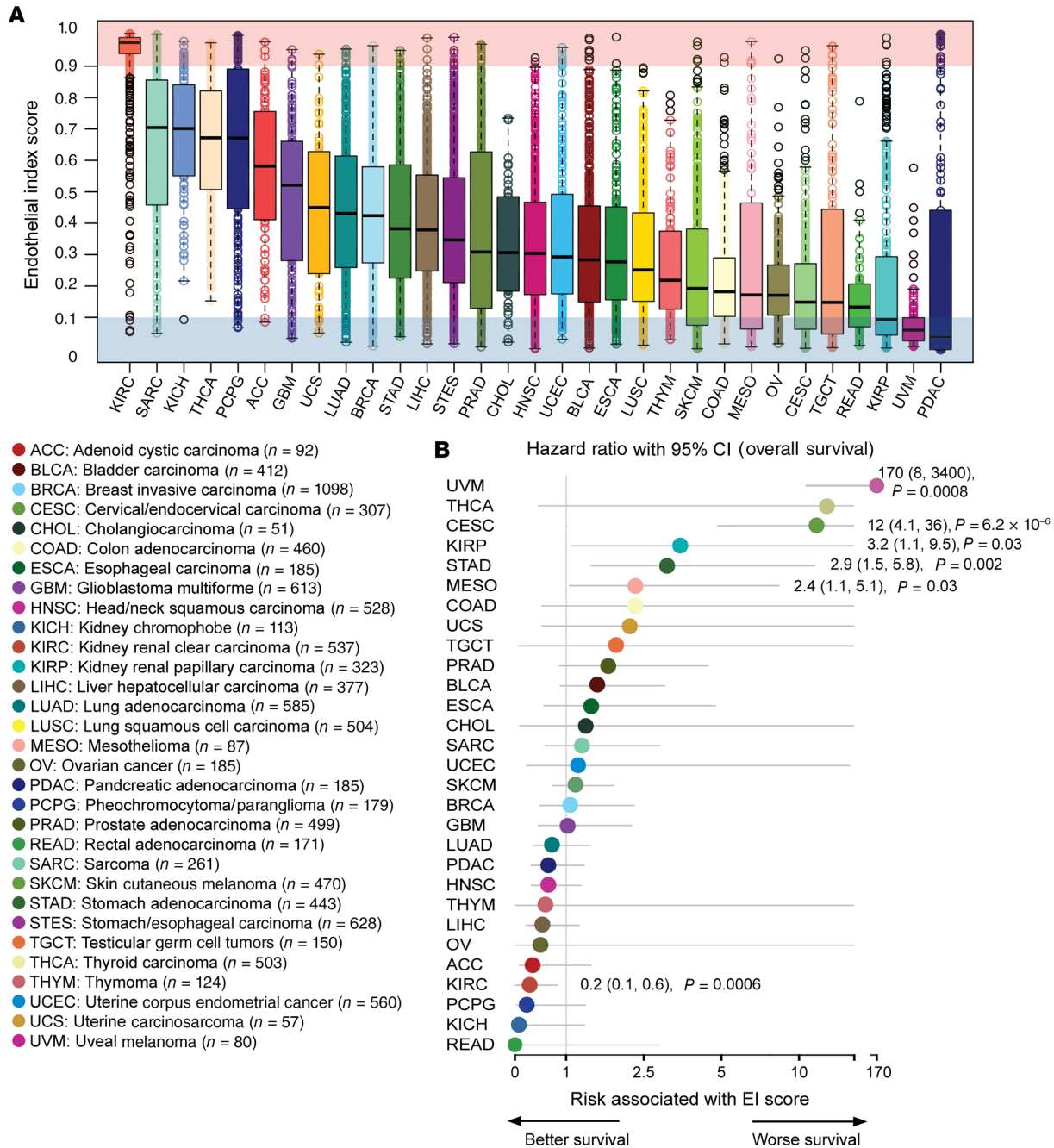
Figure 3. Vascular predictions in human tumors. (**A**) Box plots (center line marks the median, box limits span the IQR, whiskers extend to 1.5 times the IQR, and black circles indicate the outliers) showing estimates of tumor vascularity, based on the EI score, for 10,767 solid tumors from 31 different cancer types. Shaded regions highlight hypervascular (red, EI >0.9) and hypovascular (blue, EI <0.1) tumors. (**B**) Forest plots showing the association of the EI score with survival across 31 cancer types. A univariate Cox model was used to calculate the hazard associated with the EI score for each cancer type. HRs are plotted on a log scale and are listed with a 95% CI for cancer types with a statistically significant association. A HR of greater than 1 indicates increased risk (decreased survival), and a HR of less than 1 indicates decreased risk (increased survival).

observed heterogeneity in MVD for this disease (23) and our own findings from microdissected specimens (Figure 2B). Notably, vascular predictions for kidney cancer showed considerable variation across histological subtype, with most kidney chromophobe (KICH) tumors, like KIRC, having a high EI score (ranked third across all tumor types) and most kidney renal papillary cell carcinomas (KIRP) having a low EI score (Figure 3A). This is consistent

with reports showing that the KICH subtype, like the KIRC subtype, has a higher vessel density than the KIRP subtype (24, 25). Significantly, other tumor types, including many whose vascular features have not been previously explored, also exhibited wide variation in predicted tumor vascularity.

Pericytes play an important role in vascular integrity (26, 27). To determine the relationship between vascular density in tumors

and pericyte abundance, we used the pericyte-specific marker *RGS5* to quantify pericytes across TCGA database. First, we used single-cell RNA-Seq (scRNA-Seq) data from the PanglaoDB database (28) to confirm that *RGS5* was pericyte specific (Supplemental Figure 5A). Then, we compared *RGS5* expression with EI scores across TCGA. This revealed that, while *RGS5* expression correlated positively with EI scores across TCGA, its degree of correlation varied significantly between individual cancer types (Supplemental Figure 5, B and C). These data suggest that the degree of vessel coverage by pericytes may differ by cancer type.

Endothelial cells are known to have heterogeneous phenotypes. To determine whether tumors with a high EI score are enriched for specific endothelial cell subtypes, we used a recently published scRNA-Seq data set describing 13 distinct endothelial cell phenotypes in human lung cancer and normal human lung endothelium (29). Across TCGA, we found a significant overlap between the gene expression of tumors with high EI scores and the gene signatures of specific endothelial cell subsets (Supplemental Figure 5D). Gene expression in lung adenocarcinomas (LUADs) with high EI scores also overlapped significantly with signatures of specific endothelial cell subsets (Supplemental Figure 5E). All tumors with high EI scores, including LUADs, shared significant gene expression with the classically angiogenic "tip cell" phenotype (Supplemental Figure 5, D and E). These data indicate that the EI is sensitive to a diverse collection of endothelial cell phenotypes, including those thought to be capable of angiogenesis.

Association of EI score with clinical prognosis across human tumors. Blood vessels are required for tumor growth and hematogenous spread of tumor cells, 2 factors associated with reduced survival and worse clinical prognosis (30). To identify possible relationships between tumor vascularity and survival across TCGA data set, we performed univariate Cox regression analysis to calculate HRs associated with the EI score across all 31 solid cancer types (Figure 3B). A high EI score was found to correlate with worse overall survival (OS) in uveal melanoma (UVM), cervical/endocervical carcinoma (CESC), KIRP, stomach carcinoma (STAD), and mesothelioma (MESO), consistent with retrospective histological studies associating MVD with poor survival in CESC, STAD, MESO, and UVM (31–34). By contrast, a high EI score was found to correlate with better OS in KIRC (Figure 2B). These results indicate that total tumor vascularity, as assessed by the EI score alone, correlates with clinical outcomes in only a subset of tumor types. In addition, the EI score was weakly correlated with stromal content as measured by the ESTIMATE (estimation of stromal and immune cells in malignant tumor tissues using expression data) algorithm (ref. 35 and Supplemental Figure 5F), suggesting that the stromal content may potentially be a covariant in these survival analyses.

EI associates with distinct signaling environments in different tumor types. The above findings suggested that the EI provides an accurate estimate of tumor MVD but carries limited prognostic information for most tumor types. One possible explanation for this may be the exceptionally wide 95% CIs we observed when calculating the HRs relating the EI to OS (Figure 3B), consistent with studies showing that MVD alone is a poor predictor of response to antiangiogenesis therapy (36). These findings, along with the relative underperformance of antiangiogenic therapies, indicate

that additional tumor-specific contextual information might be required to understand vascular heterogeneity and its contribution to tumor growth and clinical course. We therefore hypothesized that measuring qualitative aspects of the tumor vascular microenvironment, especially features related to regulatory signaling pathways, might enable a better prediction of both clinical course and tumor-specific responses to antiangiogenic therapies.

To identify signaling pathways associated with variation in tumor vascularity, we identified genes overexpressed in hypervascular tumors (EI score >0.9) compared with hypovascular tumors (EI score <0.1) within each cancer type and used these genes as the input for a gene set enrichment analysis (GSEA). This analysis identified *VEGFA/VEGFR2* signaling as the pathway most associated with a high EI, although the strength of this association varied considerably by cancer type (Figure 4A). For example, *EGFR* and *PI3K/Akt/mTOR* signaling pathways were more strongly associated with hypervascular tumors in cholangiocarcinoma (CHOL) and thymoma (THYM) compared with *VEGFA/VEGFR2* signaling. Importantly, high EI scores were associated with multiple signaling pathways in most tumor types, supporting the notion that redundant signaling drives the formation of a tumor vasculature. Although some signaling pathways, such as that of *VEGFA/VEGFR2*, consistently correlated with high EI score across all cancer types, most signaling pathways correlated with elevated EI scores only in specific tumor types (Figure 4A).

Human tumors can be grouped according to 6 vascular microenvironment signatures. The marked heterogeneity revealed by this pathway analysis suggested that the vascular state of each tumor resulted from the expression of a defined repertoire of proangiogenic signals. Given this observation, we hypothesized that tumors could be categorized on the basis of signaling networks that were most strongly correlated with their vascular state. To this end, we developed an additional method for grouping tumors on the basis of underlying pathways associated with each tumor's vascular state (i.e., hyper- or hypovascular).

We began by generating a pan-cancer signaling network of EI-associated genes from which we could distinguish distinct clusters. We applied Fisher's method of combining *P* values with the 31 sets of genes upregulated in highly vascularized tumors (90th percentile EI score) compared with poorly vascularized tumors (10th percentile EI score) of each cancer type. This resulted in a ranked list of EI-associated genes across all cancers (Supplemental Data File 3). GSEA analyses showed a significant association of EI-associated ranked genes with Gene Ontology (GO) terms related to vascular growth (e.g., "vasculogenesis" and "angiogenesis" GO biological processes and a "vesicle lumen" GO cellular component) (Supplemental Figure 6, A and B). To translate our ranked list of EI-associated genes into signaling networks, we used Ingenuity's curated protein-protein interaction database (Ingenuity Pathway Analysis [IPA]) to build a connectivity network around genes whose expression is associated with hypervascular tumors (Supplemental Data File 4). We then curated the full interaction network to generate a more workable inventory of 24 "hub" genes representing the network's most highly connected nodes (Table 2).

It is important to note that, while some of these hub genes (e.g., *ANGPT2*, *FLT1*, *KDR*, *PTPRB*, *TGFBR2*) exhibited enriched expression in endothelial cells, the majority were not known to
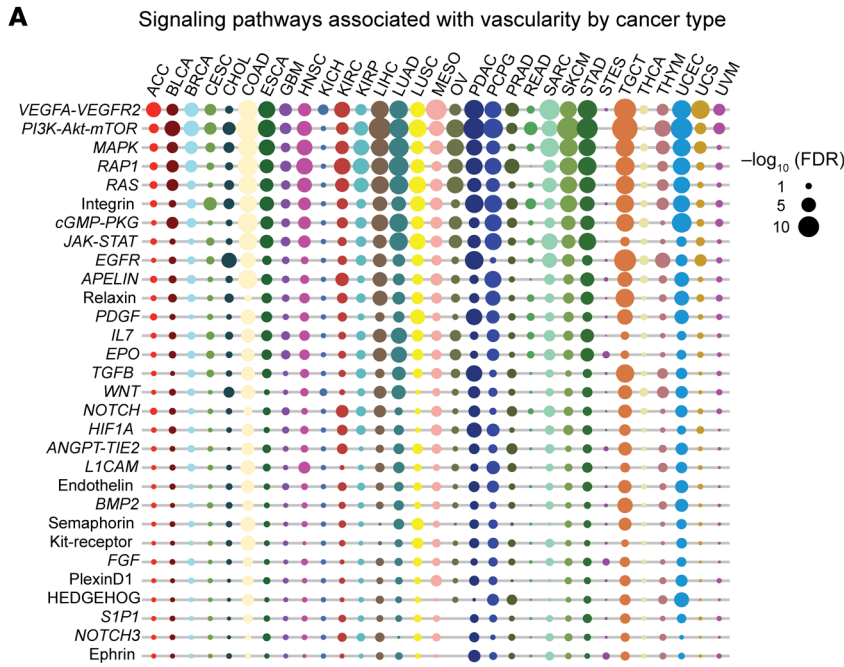
Figure 4. Signaling pathways and VMSs in cancer. (**A**) The signaling pathways most enriched in highly vascular tumors are shown. Bubble plot displays the pathways most significantly enriched in hypervascular tumors (90th percentile EI score) derived from GSEA (comparison of hyper- versus hypovascular tumors across 31 cancer types). All pathways listed as "signaling" were included in the analysis. Cancers are color-coded, and the size of each bubble represents the strength of association between the indicated signaling pathway and vascularity in the indicated cancer type [−log$_{10}$(FDR)]. (**B**) Unsupervised clustering of 10,767 solid tumors based on the expression of 24 vascularity network hub genes identified 6 VMSs (VMS1–VMS6). The color displayed for each of the 24 genes in the heatmap represents the gene expression value after normalization by both row (all tumors) and column (col) (all hub genes). Marked over- or underexpression of either *VEGFA*, *CDH1*, *NOTCH3*, or *MMP9* was characteristic of individual VMSs. Normalized levels of these genes are displayed as box plots (0 represents the normalized baseline level of this gene across cancers, the center line marks the median, box limits span the IQR, and whiskers extend to 1.5 times the IQR). min, minimum; max, maximum. The expression levels of these genes were compared across all 6 subtypes, and significant differences were detected by Kruskal-Wallis rank-sum test. Significance of the over- or underexpression of these genes in specific subtypes was tested by pairwise multiple comparisons with a Benjamini-Hochberg–adjusted post hoc Conover's test *P* value (***P < 0.001).
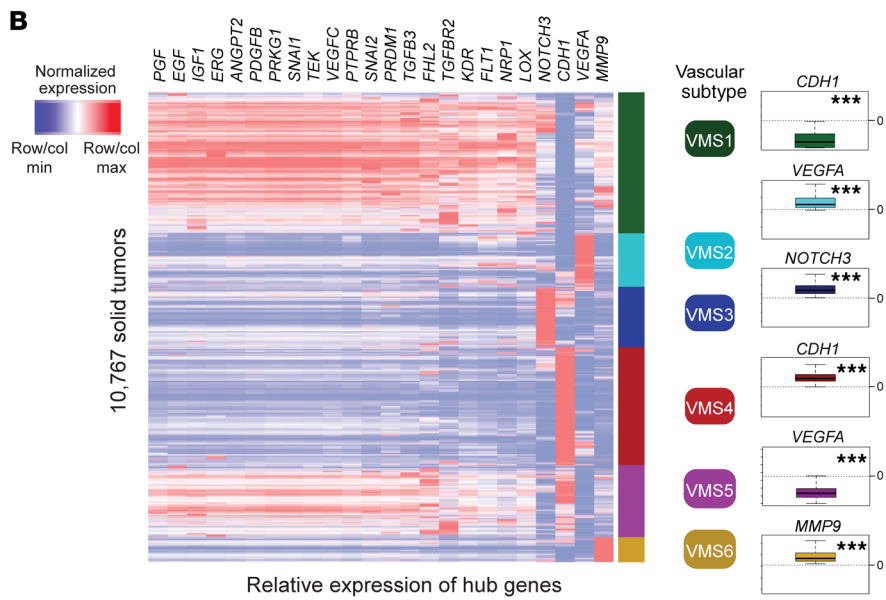
exhibit endothelial specificity. To determine whether these genes might be expressed in a subset of endothelial cells, we again evaluated the expression of these genes in individual endothelial cells using the Goveia et al. scRNA-Seq data set (29). These analyses showed that many vascular hub genes lacked endothelial cell expression (Supplemental Figure 7), suggesting that cells other than endothelial cells are responsible for their association with the EI and MVD. This interpretation is consistent with the known importance of other cells in the tumor microenvironment — pericytes, inflammatory cells, stromal cells, and tumor cells — in shaping tumor vasculature (37–39).

To identify tumors sharing patterns of vascular-associated gene expression, we clustered all 10,767 solid tumors on the basis of their relative expression of these 24 EI-associated vascular hub genes. Gene values were normalized across patients (rows) and

hub genes (columns) (see Methods). This analysis identified 6 distinct clusters, which we designated as vascular microenvironment signatures (VMSs) (Figure 4B, Supplemental Figure 6C, and Supplemental Data Files 5 and 6). Interestingly, each of the 6 VMSs were characterized by the marked over- or underexpression of 1 of 4 EI-associated hub genes — *CDH1*, *VEGFA*, *NOTCH3*, and *MMP9* — relative to the other VMSs (Figure 3C). VMS1 was the most highly represented vascular microenvironment, identified in approximately 30% of all solid tumors. VMS1 was characterized by the relative underexpression of CDH1 (E-cadherin) and the overexpression of other hub genes, including endothelial cell surface receptors, growth factors, and the genes *SNAI1* and *SNAI2*, which are associated with endothelial epithelial-mesenchymal transition (EMT) (40–42). VMS2 tumors accounted for 12% of solid tumors and were characterized by marked overexpression of *VEGFA*

Table 2. Hub genes of the tumor vascularity network

Gene	Relationships
VEGFA	94
EGF	85
IGF1	64
TGFBR2	39
KDR	31
SNAI1	31
ERG	29
PDGFB	25
PGF	23
TGFB3	23
ANGPT2	22
FLT1	22
MMP9	20
CDH1	19
VEGFC	17
SNAI2	17
LOX	16
NOTCH3	16
NRP1	16
PRKG1	16
FHL2	15
PRDM1	15
PTPRB	15
TEK	15

List of 24 vascularity network hub genes with the most relationships in a tumor vascularity network generated by IPA (see text and Methods for details).

(Supplemental Figure 8). VMS3 tumors accounted for 13% of solid tumors and were characterized by overexpression of *NOTCH3*. VMS4 tumors accounted for 25% of solid tumors and were characterized by the overexpression of *CDH1*. VMS5 tumors accounted for 15% of solid tumors and were characterized by underexpression of *VEGFA* and overexpression of other hub genes, including endothelial cell surface receptors and various growth factors. Finally, VMS6 tumors accounted for 5% of solid tumors and were characterized by overexpression of *MMP9*.

VMS distribution across human tumor types. We next sought to understand how these 6 VMS subtypes are distributed according to tumor type. To this end, we regrouped the tumors subtyped in Figure 4B according to tissue of origin. As shown in Figure 5, VMSs were unevenly distributed across different types of cancer. VMS1 tumors comprised the majority of glioblastomas, adrenocortical carcinomas, MESOs, liver hepatocellular carcinomas (LIHCs), sarcomas (SARCs), pheochromocytomas/paragliomas (PCPGs), adrenocortical carcinomas, skin melanomas, papillary renal cell carcinomas, testes/germ cell tumors, and uterine carcinosarcomas (UCSs). VMS2 tumors comprised a large fraction of KIRCs, glioblastomas, LUADs, and pheochromocytoma/paragliomas. Of note, breast cancers belonging to the basal-like molecular subtype were significantly overrepresented among VMS2 breast cancers, whereas basal tumors made up a small proportion of breast cancers possessing other VMSs (Figure 5, inset). VMS3

tumors comprised a large fraction of head and neck squamous cell carcinomas, esophageal carcinomas (ESCAs), and cervical and ovarian cancers (OVs). VMS4 tumors comprised a majority of CHOLs, colon adenocarcinomas (COADs), rectal adenocarcinomas (READs), prostate adenocarcinomas (PRADs), and thyroid cancers. VMS5 tumors made up a significant portion of UVMs, rectal, prostate, pancreatic, stomach, and COADs. Finally, VMS6 tumors comprised a significant portion of testes/germ cell tumors and head and neck squamous cell carcinomas.

Next, we assessed the relationship between VMS and vessel density by measuring EI scores associated with the 6 VMSs in each of the 31 cancer types (Figure 6, A and B). VMS1 was associated with low vessel density in KIRPs (FDR < 0.01 for all pairwise comparisons, or "all FDR") and pheochromocytoma/paraglioma (all FDR < 0.0001). VMS2 was associated with high vessel density in stomach cancers (all FDR < 0.0001), esophageal cancers (all FDR < 0.001), pheochromocytoma/paragliomas (all FDR < 0.0001), KIRCs (all FDR < 0.0001), papillary carcinomas (all FDR < 0.05), and thyroid cancers (all FDR < 0.001). Of note, VMS2 tumors also had the highest expression of the pericyte marker *RGS5* (Supplemental Figure 9). VMS3 was associated with high vessel density in testicular/germ cell tumors (all FDR < 0.01). VMS4 did not significantly associate with vessel density in any individual cancer types. VMS5 was associated with low vessel density in thyroid cancers (all FDR < 0.001) and LUADs (all FDR < 0.05). VMS6 was associated with high vessel density in COADs (all FDR < 0.05). These findings are consistent with a complex and diverse relationship between vessel density and the vascular microenvironment across tumor types.

Prognosis associates with VMS independent of the EI score. We hypothesized that a tumor's VMS might provide additional prognostic information beyond that afforded by MVD. To test this, we used multivariate Cox proportional hazards regression models to determine the association of the 6 VMSs with OS and the progression-free interval (PFI). Following correction for multiple-hypothesis testing, we found VMS to be a significant prognostic factor for OS in 3 cancer types (Figure 7, A and B) and for a PFI in 4 cancer types (Figure 7, C and D). VMS2 SARCs were linked to a significantly shortened OS (Figure 7, A and B) and PFI (Figure 7, C and D), consistent with prior studies identifying VEGF protein expression as a predictor of poor survival in soft-tissue SARCs (43–46). VMS2 was also associated with a shorter PFI in uterine cancer and the chromophobe subtype of renal cell carcinomas (Figure 7, C and D). In addition, VMS3 was associated with shortened OS in skin melanomas, VMS1 was associated with significantly prolonged survival and PFI in UVM (Figure 7, A and B), and VMS6 was associated with shortened PFI in SARCs (Figure 7, C and D). VMS4 and VMS5 were not significantly associated with prognosis in any cancer type. Importantly, statistical testing revealed no significant interaction between the categorical variable "VMS" and the continuous variable "EI score" in predicting survival. Thus, the tumor VMS represents a qualitative measure of tumor vascular biology that is clinically relevant and complementary to EI.

Non-VMS2 classification predicts a favorable response to bevacizumab in OV. There is a need for predictive biomarkers to guide antiangiogenesis therapy for cancer. Some studies have linked increased tumor *VEGFA* expression with favorable patient responses to anti-*VEGFA* therapies, including bevacizumab (47–
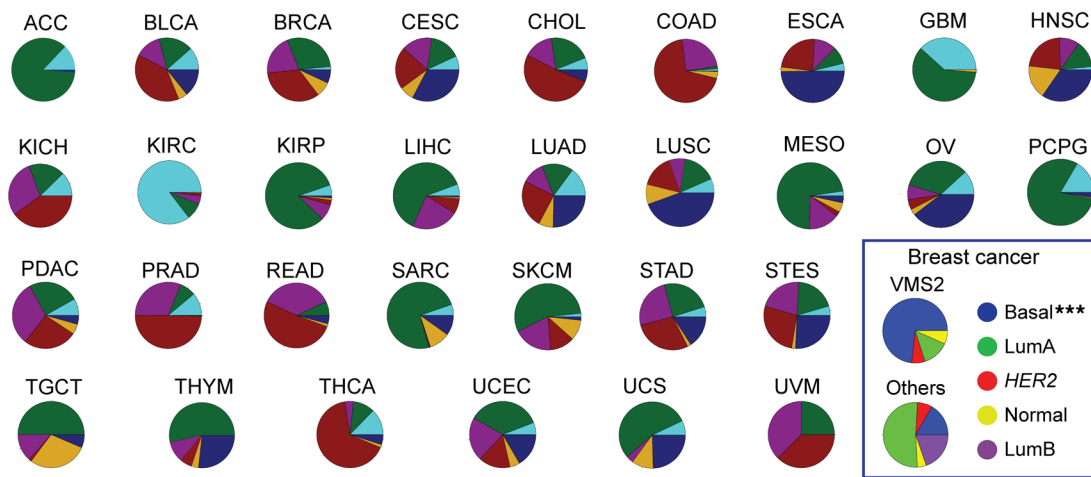
Figure 5. Distribution of VMSs across different cancer types. Pie charts showing the distribution of vascular subtypes for 31 individual cancer types. Inset: Pie charts showing that the basal subtype was significantly overrepresented in VMS2 breast cancers (74%). Fisher's exact test was used to compare the abundance of the basal subtype in VMS2 breast cancers with its abundance in all other breast cancers. ***$P < 0.001$. VMS1 (green), VMS2 (turquoise), VMS3 (dark blue), VMS4 (red), VMS5 (magenta), VMS6 (gold).

51). We thus hypothesized that the VMS2 tumor classification, characterized by overexpression of *VEGFA* relative to 23 other vascular hub genes, may predict a favorable response to the anti-*VEGFA* agent bevacizumab.

To test this hypothesis, we performed vascular subtyping of tumors from the ICON7 trial (International Collaboration on Ovarian Neoplasms), a phase III study of patients with OV treated with carboplatin and paclitaxel with or without the addition of bevacizumab (52). Given that unsupervised clustering algorithms (Figure 4B) cannot be used to classify individual tumors, we first trained a classifier to predict VMS2-associated tumors by applying expression values for the vascular hub genes to 7000 randomly selected TCGA tumors (Figure 8A). This VMS2 classifier performed with 97% accuracy on a test set consisting of the remaining 3000 TCGA tumors (Figure 8A). We then applied the VMS2 classifier to pretreatment expression data from 380 OVs from ICON7, classifying 31% of the tumors as VMS2 and 69% of the tumors as non-VMS2 (Figure 8B).

Next, we compared the OS and PFS of VMS2 and non-VMS2 tumors within the bevacizumab treatment arm using a univariate analysis by standard Kaplan-Meier testing and a multivariate analysis using Cox models adjusted for age and stage. Unexpectedly, we found that patients with OVs classified as non-VMS2 showed a trend toward greater benefit from bevacizumab compared with those classified as VMS2 in terms of OS ($P = 0.09$, adjusted for age and stage) and significantly greater benefit in PFS ($P = 0.03$, adjusted for age and stage) (Figure 8, C and D). We observed no significant difference in survival or disease progression between individuals with VMS2 or non-VMS2 tumors who were treated with carboplatin and paclitaxel without bevacizumab (Supplemental Figure 10, A and B), suggesting that the influence of vascular subtyping was specific to *VEGFA* blockade. Importantly, neither the EI score nor *VEGFA* expression alone predicted a response to bevacizumab (Figure 8, C and D).

Next, we extended this analysis by determining whether the added clinical benefit of bevacizumab over chemotherapy alone —

which led to FDA approval for its use in OV (53) — can be accounted for by patients with tumors classified as non-VMS2. The addition of bevacizumab conferred a progression-free survival (PFS) advantage over chemotherapy alone when all trial participants were included (Figure 9, A and D), confirming earlier findings (52). Importantly, patients with non-VMS2 ovarian tumors derived virtually all the benefit from the addition of bevacizumab (Figure 9, B and E), while patients with VMS2 tumors derived no benefit from the addition of the drug (Figure 9, C and F). We further observed that VMS6 tumors, which have elevated *MMP9* levels, specifically accounted for much of the collective non-VMS2 response to bevacizumab (Supplemental Figure 11). While the basis for the finding that VEGF-high VMS2 ovarian tumors responded more poorly to anti-VEGF therapy than did non-VMS2 ovarian tumors remains uncertain (see Discussion), these data suggest that VMS classification may provide a useful metric for predicting the response to bevacizumab in OV that outperforms vascular density or *VEGFA* levels.

Discussion

Quantitative and qualitative differences in tumor vasculature are believed to influence important aspects of tumor biology that contribute to clinical prognosis and therapeutic response. Here, we describe 2 metrics — an EI and 6 VMSs (VMS1–6) — that can be applied to transcriptomic data from both large cohorts and individual tumors, allowing stratification based on vascular characteristics. Analysis of the tumor vasculature using the EI and VMS reproduced several established paradigms regarding tumor vascular density, including the identification of renal clear cell and chromophobe carcinomas as hypervascular and pancreatic carcinomas as hypovascular. Importantly, EI and VMS analysis also revealed subtle features of vascular heterogeneity both within and between tumor types, allowing differences in tumor vascular biology to be probed using widely available transcriptional data. Furthermore, the reliance of these metrics on a relatively small number of genes highlights their potential application as clinical biomarkers. In the future, it should be possible to adapt both the EI score and VMS
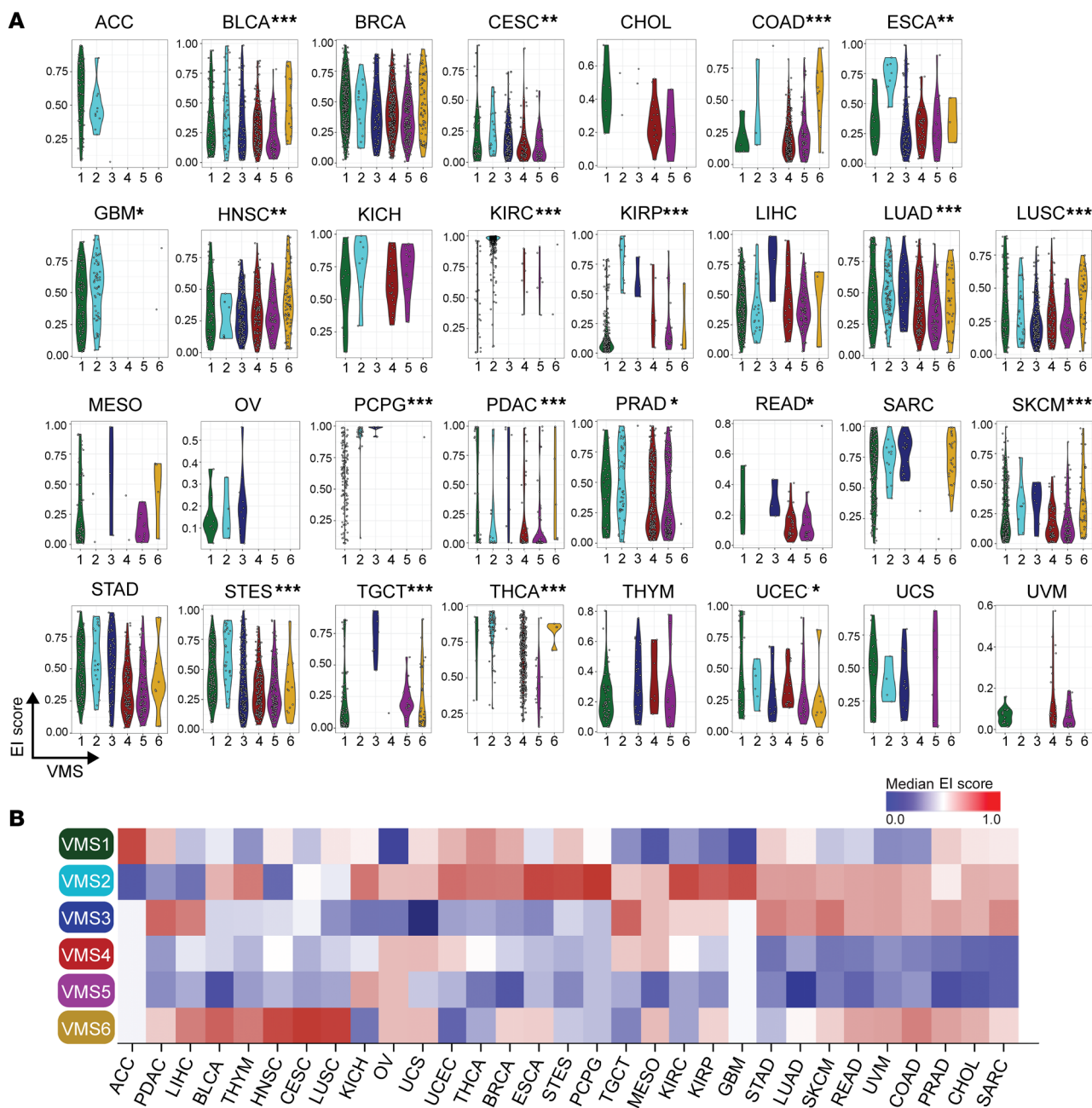
Figure 6. Relationship between the VMS and the EI differs across cancer types. (**A**) Violin plots showing EI scores (y axis) of tumors belonging to VMS1–VMS6 (x axis) in 31 individual cancer types. Significant differences in EI score of tumors with different VMSs (within each cancer type) were detected using a Kruskal-Wallis rank-sum test (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). Kruskal-Wallis testing was followed by Conover testing of pairwise comparisons to identify specific associations between the VMS and EI score within cancer types. The threshold FDR met by all pairwise comparisons is reported in the text for each VMS-EI score association. (**B**) Heatmap showing the median EI score (0–1, red, high; blue, low) of each vascular microenvironment within 31 solid tumor types. Tumor types are clustered by similarity with respect to their VMS proportions.

categorization as quantitative PCR (qPCR) tests on tumor biopsies. Such an advance could provide a real-time characterization of tumor vascularity and vessel environment, which in turn could aid in clinical decision making.

The EI score was able to predict OS for several tumor types. Surprisingly, the vessel density estimates from this measure were associated with opposing prognoses in 2 histologic subtypes of kidney cancer. Specifically, the EI score predicted better survival

for KIRC but worse survival for KIRP. These findings may explain why the prognostic value of MVD in kidney cancers — which were only rarely stratified according to histological subtype in prior studies — has remained controversial (54).

A primary motivation for developing the EI was to identify distinct signaling environments that associate with tumor vasculature in large numbers of tumors. We used the EI score to stratify human tumors with the highest and lowest vascular density,
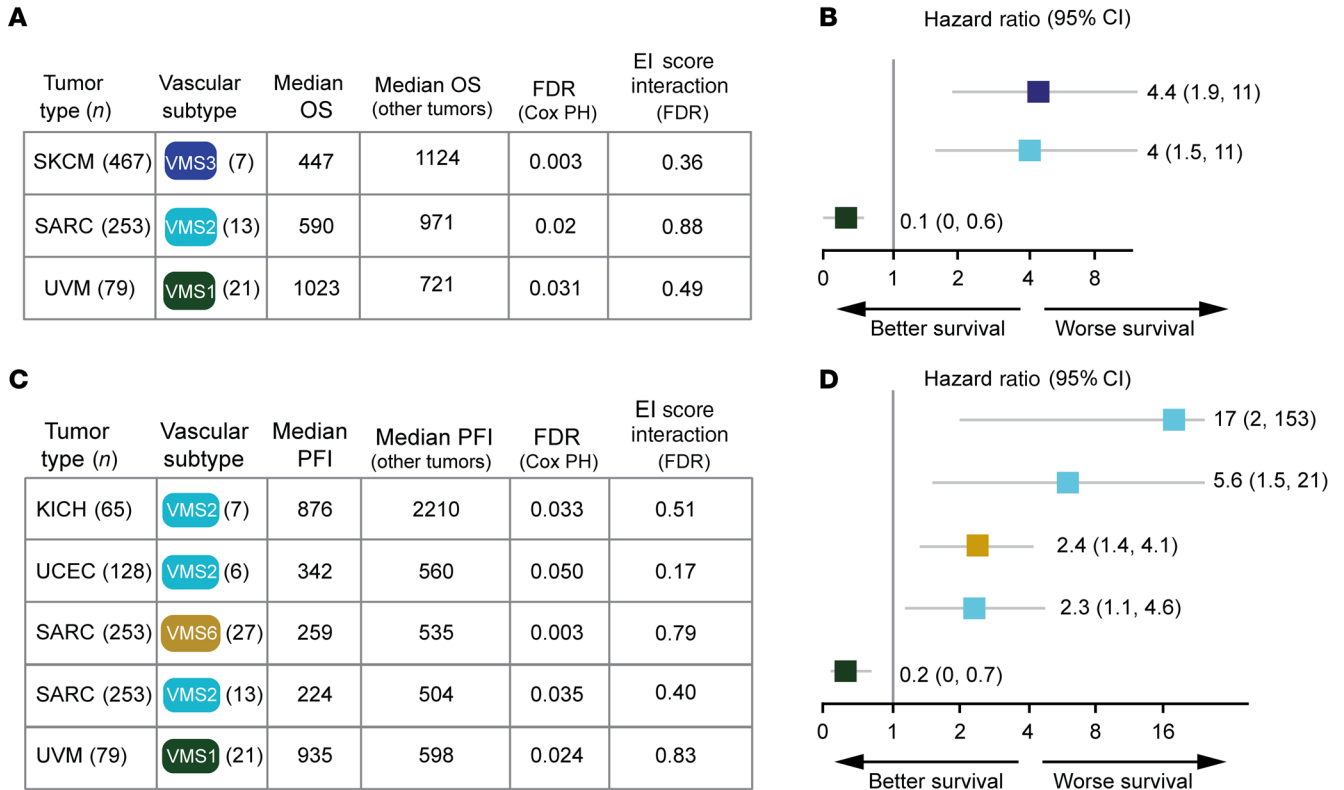
Figure 7. VMS predicts OS and the PFI in certain cancer types. (**A**–**D**) A multivariate Cox proportional hazards (PH) regression model, correcting for multiple comparisons, was used to calculate the hazard associated with each microenvironment and OS (**A** and **B**) or the PFI (**C** and **D**). HRs are plotted on a log scale and are listed with 95% CIs. The interaction of each vascular microenvironment with the EI score was calculated to test the dependence of these clinical associations on bulk vascularity (EI score interaction).

allowing the identification of 24 hub genes that best represented the signaling networks enriched in well-vascularized tumors. The 6 VMSs derived from clustered expression of these 24 hub genes represent different signaling environments associated with vascular state. Importantly, VMSs do not reflect signaling pathways operating solely within or upon endothelial cells. Rather, a tumor's VMS represents the contribution of the entirety of the signaling milieu to a tumor's vascular state — including potential contributions from endothelial cells, fibroblasts, hematopoietic cells, extracellular matrix, and pericytes.

While some retrospective analyses have reported a relationship between higher MVD and an improved responses to anti-VEGF therapy (55, 56), such correlations have not conferred predictive value in larger phase III clinical trials (36). Because VMSs provide a qualitative metric of the tumor/vessel signaling environment that is orthogonal to vascular density (EI score) and reflects the sum of numerous cellular and molecular contributions to the vascularity of a tumor, VMS categorization may be a better predictive biomarker. Indeed, we found that VMS2 classification — but not the EI score or simple *VEGFA* expression level — predicted the response to bevacizumab in a retrospective analysis of OVs. Non-VMS2 ovarian tumors benefitted most from the anti-*VEGFA* drug bevacizumab, whereas those with VMS2 status showed no difference in OS or PFS. Although the molecular basis for this association is unclear, our results highlight the complex and dynamic nature of the tumor's angiogenic response.

The finding that relatively VEGF-low non-VMS2, but not VEGF-high VMS2, ovarian tumors conferred a significant survival advantage for patients treated with the anti-VEGFA drug bevacizumab is counterintuitive and highlights the need to better understand the biology underlying the VMS categorization. One potential explanation for this result is that VMS2 tumors, which express higher levels of the pericyte marker *RGS5* than do non-VMS2 tumors (Supplemental Figure 9A), contain more mature vessels with higher pericyte coverage. Prior studies indicate that pericytes protect tumor vessels from loss of VEGF signaling and may be a primary cellular mechanism of tumor resistance to anti-VEGF therapy (57–62). Consistent with this hypothesis, we found that tumor *RGS5* expression significantly correlated with the survival of patients treated with bevacizumab, but not chemotherapy alone (Supplemental Figure 9). Other possible explanations for the poor response of the *VEGFA*-high VMS2 tumors to bevacizumab are that these tumors underwent a selection for bevacizumab-resistant clones or that their baseline level of *VEGFA* was too high for anti-VEGF drugs to effectively block. Future studies that dissect the serial genetic and transcriptional changes in tumors during antiangiogenesis therapy are likely to provide greater insight into the molecular mechanisms linking the vascular microenvironment and clinical outcome.

Finally, our study provides an entry point for the identification of patient subsets that might respond to angiogenesis inhibitors targeting pathways other than that of *VEGFA*. *NOTCH3* and
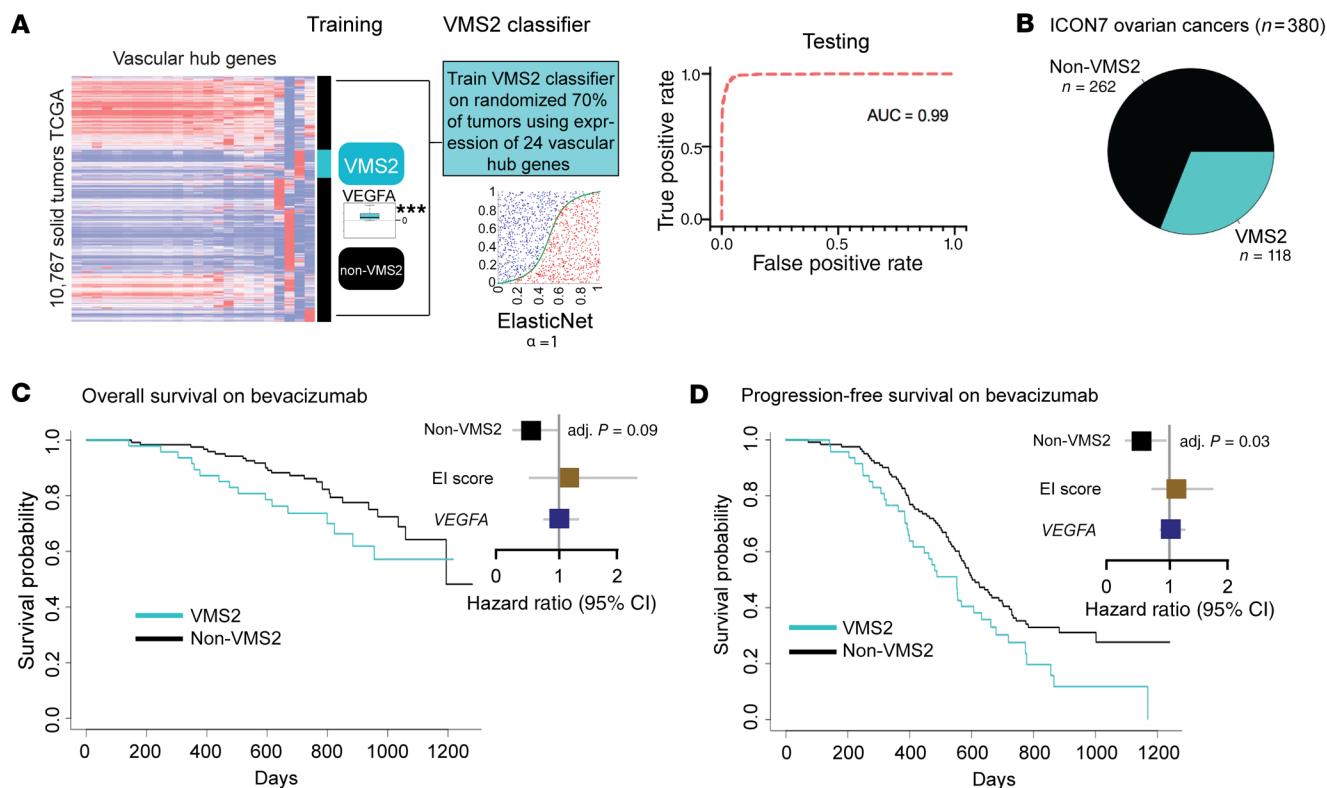
Figure 8. Non-VMS2 status predicts a favorable response to bevacizumab in OV. (**A**) Schematic of the samples used to train and test an ElasticNet classifier to identify tumors as VMS2 or non-VMS2 using expression values of the 24 vascular hub genes. The VMS2 classifier performed with 97% accuracy on an independent test data set (0.99 AUC). (**B**) The VMS2 classifier categorized 31% of OVs in the ICON7 trial data set as VMS2 and 69% as non-VMS2. Kaplan-Meier analysis comparing OS (**C**) and PFS (**D**) for patients with VMS2 or non-VMS2 tumors on bevacizumab. Forest plots show the association of non-VMS2 status, EI score, and VEGFA expression with survival. Adj., adjusted.

MMP9 play important roles in tumor angiogenesis (63–68), and drugs targeting them have shown promise in preclinical models (69–71). Therefore, the VMS3 and VMS6 tumor microenvironments, in which the expression of *NOTCH3* and *MMP9* dominates that of other vascular-associated hub genes, may represent novel populations of patients susceptible to new classes of antiangiogenesis therapies. Unfortunately, VMS1 and VMS5 tumors — which together comprise 45% of all tumors — are enriched for the expression of multiple vascular hub genes, suggesting that these tumors are driven by redundant angiogenic pathways that may be particularly difficult to target.

Methods

Generating an endothelial cell classifier. Samples from the following publicly available, normal human tissue microarray data sets were downloaded from the NCBI's Gene Expression Omnibus (GEO) database (72) to form training and testing sets for the generation of a human endothelial cell classifier: GSE3239, GSE44926, GSE59326, GSE76340, GSE39396, GSE59326, and GSE21212 (Supplemental Data File 1). Microarray data were quantile normalized and log transformed (Supplemental Figure 1). Slight differences in distribution between the training and testing sets resulted from quantile-normalizing each separately; this was necessary to ensure reproducible performance of a classifier when applied to arbitrary external validation sets. Microarray samples originating from endothelial tissue were given the label

"E," and all other nonendothelial samples were labeled "N." Because the goal of the EI was to predict microvessel density in tumors, "N" samples were chosen to represent cell types that are commonly present in tumor microenvironments (Figure 1A). Training involved use of the 200 most differentially expressed genes between the "E" and "N" classes. ElasticNet (73) is a 2-response logistic regression method with an additive penalty term meant to restrict the number of features included in a model by adding each non-zero feature coefficient to the additive error. The objective function for this model is as follows:

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{P+1}} - \left[\frac{1}{N}\sum_{i=1}^{N} y_i \cdot (\beta_0 + x_i^T\beta) - \log(1 + e^{(\beta_0 + x_i^T\beta)}) \right] + \lambda[(1-\alpha)||\beta||_2^2/2 + \alpha||\beta||^1]\cdot$$

(Equation 1)

The character of this penalty term is parameterized by α (0 to 1), which balances between lasso ($\alpha = 1$) and ridge ($\alpha = 0$) methods, where the former will drive coefficients for groups of highly correlated genes to zero, while the latter allows linear combinations of correlated features in the model. Because the preprocessing step selects for 200 features, many of which are highly correlated, α was set to 1 to minimize redundancy in the feature set. The strength of this penalty term is parameterized by λ. Larger values of λ correspond to a smaller feature set, and the optimal value of λ is found by repeated k-fold cross-validation (Supplemental Figure 1E). These penalty terms enabled the
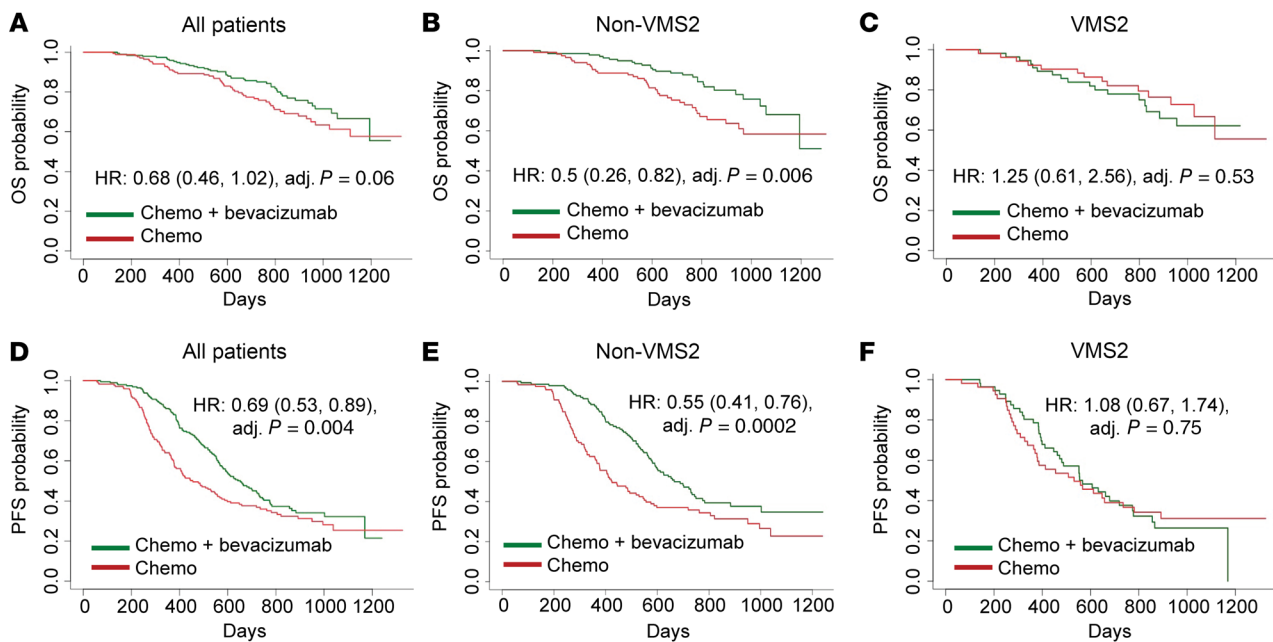
Figure 9. The benefit of bevacizumab in OV is mostly seen in patients with non-VMS2 tumors. Kaplan-Meier analysis comparing OS (**A**–**C**) and PFS (**D**–**F**) for all patients (**A** and **D**), patients with VMS2 tumors (**C** and **F**), and patients with non-VMS2 tumors (**B** and **E**), with and without the addition of bevacizumab. The HRs shown relate to the addition of bevacizumab. All *P* values are the result of multivariate Cox regression analysis, with patient age and tumor stage as covariates. Chemo, chemotherapy.

selection of a concise, 7-gene model. As with a standard 2-response logistic regression, the classifier uses the binomial log likelihood ratio to estimate the probability that an applied sample belongs to class "E." This score, which ranges from 0 to 1, was termed the EI and serves as a continuous metric of vascularity.

The EI was then applied blindly to an independent test set of 139 microarrays from 26 endothelial "E" and 113 nonendothelial "N" tissues. "N" tissues were selected to wholly comprise a mock tumor microenvironment, containing cell types such as fibroblasts, various immune cells, and epithelial cells from various sources. Samples assigned an EI score of greater than 0.9 were labeled "E," and those with an EI score of less than 0.1 were labeled "N." The EI score classified the test set with an overall accuracy of 92%. Interestingly, the EI score committed only false-positive errors in classifying the test set. For each of these errors, the model output a probability between 0.4 and 0.6 for class "E." All other samples were classified correctly to within 1% confidence, suggesting that any reasonably discriminating threshold we chose as our "activation function" to assign the binary "E" and "N" labels had no impact on the results (Figure 1A).

As a final step of in silico validation, we applied maximum relevance minimum redundancy (mRMR) feature selection, a different modeling method that uses a greedy heuristic to choose a maximal set of linearly independent features that explain variance, to the training set. This method selected 6 of the 7 EI genes found previously by ElasticNet, suggesting that the EI is a reliable measure of the endothelial cell classification independent of the classification method. The code for tumor EI prediction is available at https://github.com/faryabib/VMS.

Human PDAC cohort and EI scoring. A cohort of 60 human pancreatic ductal adenocarcinomas with paired gene expression data and archived tissue sections were chosen to validate the accuracy of the EI score in predicting intratumoral MVD. Information was provided

for a total of 122 patients with PDAC who underwent surgery at the Columbia Pancreas Center, as described previously (11). Freshly frozen PDAC samples were processed and sectioned for histologic review as described (11). The stromal and epithelial portions of each tumor were isolated by LCM for their separate expression and histologic analyses (Supplemental Figure 3A). RNA was collected from tissue slides for RNA-Seq, and adjacent slides were fixed and embedded in paraffin blocks for histologic analysis. The EI was applied to the gene expression profiles of both the stromal and epithelial portions of each tumor to generate EI scores for each. Because blood vessels are located in the tumor stroma, EI scores for the stromal samples were used to reflect the microvessel density of each tumor sample (Figure 2, A and B), and EI scores for the epithelial samples were used as a negative control (Supplemental Figure 3B). One epithelial sample was classified as hypervascular (EI >0.9) but was believed to represent a biological outlier with either stromal contamination or a stromal gene expression pattern based on principal component analysis and histological analysis (11).

Mouse PDAC cohort and EI scoring. A panel of 19 congenic pancreatic tumor cell clones, isolated from late-stage primary tumors from KPCY (*Kras^{LSL-G12D/+} Trp53^{L/+} Pdx1-Cre Rosa26^{YFP/YFP}* and *Kras^{LSL-G12D/+} Trp53^{LSL-R172H/+} Pdx1-Cre Rosa26^{YFP/YFP}*) mice was generated by limiting dilution as previously described (16). These tumor cell clones were orthotopically implanted into C57BL/6 mice to produce a panel of 19 mouse pancreatic tumors with heterogeneous microenvironments. Bulk tumor RNA was isolated from these 19 tumors, and RNA-Seq was performed and processed as described previously (16) to generate bulk tumor gene expression profiles. The EI was then applied to these gene expression profiles to generate EI scores for each of the 19 mouse tumors.

Quantification of MVD in mice and humans. A *CD31* mouse anti-human antibody (Dako, clone JC70A, 1:40) was used to measure MVD

in human tumor slides. Slides were deparaffinized and rehydrated, and antigen retrieval was performed using 6.5 mmol/L sodium citrate, pH 6.0. Slides were blocked with 5% donkey serum, 0.3% Triton-X 100 for 1 hour at room temperature (RT) and incubated with a primary antibody overnight at 4°C. The slides were washed, incubated in 3% hydrogen peroxide for 10 minutes at RT, washed again, and incubated with a biotinylated donkey anti-mouse secondary antibody for 1 hour at RT. Slides were washed, incubated in ABC reagent for 30 minutes at RT, incubated in DAB (peroxidase substrate) for 5 minutes, counterstained with hematoxylin for 3 minutes, and imaged using an Olympus IX71 inverted multicolor fluorescence microscope and a DP71 camera.

The endomucin rat anti-mouse antibody (Santa Cruz Biotechnology, V.7C7, 1:200) was used to measure MVD in the mouse tumor slides. After sectioning, tissues were fixed in Zn-formalin and paraffin embedded for histological analysis and immunofluorescence staining. Sections were deparaffinized, rehydrated, and prepared by antigen retrieval and then blocked in 5% donkey serum for 1 hour at RT, incubated with primary antibodies overnight at 4°C, washed, incubated with secondary antibodies for 1 hour at RT, and then washed again and mounted. Primary antibodies included goat anti-GFP (Abcam, ab5450) (tumor cells) and rat anti-EMCN (blood vessels). Slides were visualized and imaged as described above.

The Chalkley method (74) was used to quantify MVD in both mouse and human tumors. Neovascular regions from representative areas of each tumor were selected for imaging. ImageJ software (NIH) was used to quantify the area of fluorescence staining in mouse tumors and to deconvolute multicolored images of human tumors stained for *CD31* (DAB) and hematoxylin (purple) to isolate *CD31* staining. The percentage of area with *CD31* staining was quantified by the fixed threshold method.

TCGA data sets. To investigate the levels of vascularity across all solid tumor types, RNA expression data sets and the corresponding clinical information for 31 types of solid cancer were accessed from TCGA (http://cancergenome.nih.gov/). The following cancer types (project code and sample size) were selected for analysis: adenoid cystic carcinoma (ACC) ($n = 92$), bladder carcinoma (BLCA) ($n = 412$), breast invasive carcinoma (BRCA) ($n = 1098$), CESC ($n = 307$), CHOL ($n = 51$), COAD ($n = 460$), ESCA ($n = 185$), glioblastoma multiforme (GBM) ($n = 613$), head/neck squamous carcinoma (HNSC) ($n = 528$), KICH ($n = 113$), KIRC ($n = 537$), KIRP ($n = 323$), LIHC ($n = 377$), LUAD ($n = 585$), lung squamous cell carcinoma (LUSC) ($n = 504$), MESO ($n = 87$), OV ($n = 185$), PDAC ($n = 185$), PCPG ($n = 179$), PRAD ($n = 499$), READ ($n = 171$), SARC ($n = 261$), skin cutaneous melanoma (SKCM) ($n = 470$), STAD ($n = 443$), stomach/esophageal carcinoma (STES) ($n = 628$), testicular and germ cell tumors (TGCTs) ($n = 150$), thyroid carcinoma (THCA) ($n = 503$), THYM ($n = 124$), uterine corpus endometrial carcinoma (UCEC) ($n = 560$), UCS ($n = 57$), and UVM ($n = 80$). Tumor RNA-Seq data sets and clinical information were obtained for each disease type from the Broad Institute's Genome Data Analysis Center (GDAC) Firehose (http://gdac.broadinstitute.org). The EI was applied to the mRNA expression data set from each cancer type to generate EI scores.

Survival analysis. The prognostic significance of the EI score was retrospectively tested for 2 clinical endpoints: OS (time between diagnosis and death from any cause) and PFI (time between diagnosis and the first incidence of disease progression). For each of 31 individual cancer types, univariate Cox proportional hazard regression models (75) with the EI score as the variable were calculated along with log-

rank P values and HRs using the "surv'" and "coxph" functions of the "survival" package in R.

Multivariate Cox proportional hazards models were generated to determine the prognostic significance of the 6 vascular microenvironments in predicting survival for each individual cancer type. Membership of each vascular microenvironment was treated as a "dummy" variable, and Cox PH regression modeling was performed for this variable with the other 5 vascular microenvironments serving as covariates. The Benjamin-Hochberg method was applied to the resulting P values to correct for multiple comparisons and generate FDRs. To test whether the prognostic power of vascular subtypes found to predict survival with statistical significance (FDR <0.05) was dependent upon the EI score, we generated an interaction term (EI score*vascular microenvironment) using the "coxph" function. All features fulfilled the proportional hazards assumption (χ^2 tests: $P > 0.05$). Testing used Schoenfeld residuals and was done with the "cox.zph" function.

Kaplan-Meier curves were used to visualize the survival differences between groups within the ICON7 data set. Multivariate cox proportional hazards models were used to evaluate the association between specific factors and survival in these analyses, and log-rank tests adjusted for 2 covariates (age and tumor stage) were reported as adjusted P values. All features fulfilled the proportional hazards assumption (χ^2 tests: $P > 0.05$), and testing used Schoenfeld residuals (76) and was done with the "cox.zph" function.

Signaling pathway and GSEA. To identify signaling pathways associated with increased vascularity, tumors within each of 31 cancer types were stratified according to "high" (>90th percentile EI) and "low" (<10th percentile EI) predicted vascularity. Differential gene expression analysis comparing these groups was calculated with the R/Bioconductor package "limma" with voom transformation to identify genes upregulated in the more vascular tumors of each cancer type (77). Genes overexpressed in the "high" vascularity tumors (adjusted $P > 0.01$) were mapped to the KEGG (Kyoto Encyclopedia of Genes and Genomes), BioCarta, and MsigDB pathways using GSEA (78). Only pathways labeled "signaling" were selected for analysis. The adjusted P values for all signaling pathways overexpressed in vascularity "high" tumors were combined across all 31 cancer types using Fisher's method to identify the pathways that most consistently associated with vascularity in cancer. The association of these signaling pathways with vascularity is represented in a bubble chart (Figure 4A).

Defining vascular microenvironments. The adjusted P values assigned to genes overexpressed in vascularity "high" tumors of all 31 cancer types were combined using Fisher's method (79) to generate a list of genes associated with increased vascularity across cancer. To confirm that this list reflected differences that were biologically relevant to vascularity, GSEA was performed on this gene set to evaluate enrichment of GO terms for "biological process" and "cellular component."

These genes differentially upregulated in vascularity "high" tumors across cancer were sent, along with their combined P value, combined adjusted P value, and mean (across all cancer types) $\log_2$ fold-change, to IPA (QIAGEN, www.qiagen.com/ingenuity) to generate a protein-protein interaction (PPI) network. IPA used the Ingenuity Knowledge Base, a collection of experimentally observed relationships between genes and gene products, to generate a PPI connectivity plot (Supplemental Data File 4) detailing the relationships between genes overexpressed in vascularity "high" tumors. The 24 genes that

served as the most connected nodes in the network were then chosen to be used as hub genes for further analysis (Table 2). The expression levels of these 24 hub genes were normalized to each other within each tumor, and then normalized across all 10,767 tumors, which were then clustered into 6 "vascular microenvironments" according to their expression of the 24 hub genes. The average silhouette width method (80) identified 6 as the optimal number of clusters, and k-means clustering was used to assign tumors their respective vascular microenvironments using the Cluster package in R.

Generating a VMS2 tumor classifier. A total of 10,767 solid tumors of 31 different types were clustered into 6 VMSs using the expression of 24 vascular hub genes, as previously described. All VMS2 tumors retained their label, and tumors of other microenvironments were labeled "non-VMS2." The 10,767 samples were then randomly split into a training set (70%) and a testing set (30%). Training and testing sets were generated such that they would both have the same ratio of VMS2 and non-VMS2 labels. Features for each sample were limited to expression values for 21 of the 24 vascular hub genes (3 were excluded because of probe set inconsistencies in the ICON7 expression data set). A logistic regression classifier with an ElasticNet penalty (λ=1) was fitted to the training data in order to classify VMS2 versus non-VMS2 tumors. This classifier was then applied to the independent test set consisting of the remaining, unlabeled, 30% of TCGA samples. Default thresholds for logistic regression were kept such that samples assigned a VMS2 classifier score of greater than 0.5 were labeled "VMS2," and scores of less than 0.5 were labeled "non-VMS2." The overall accuracy of the VMS2 classifier on the test set was 97% with a receiver operating characteristic (ROC) AUC of 0.99 (Figure 8A).

Generating a multiclass VMS tumor classifier. A total of 10,767 solid tumors of 31 different types were clustered into 6 vascular subtypes (VMSs) using the expression of 21 vascular hub genes, as described above. Features for each sample were limited to expression values for 21 of the 24 vascular hub genes (3 were excluded because of probe set inconsistencies in the ICON7 expression data set). All the tumors were labeled as belonging to each of these VMSs, with each tumor only being assigned to a single VMS. The 10,767 samples were then randomly split into a training set (70%) and a testing set (30%). Training and testing sets were generated such that they would both have the same number of samples belonging to each of the VMSs. Features for each sample were limited to expression values for the 21 vascular hub genes. A one-vs.-rest multivariate logistic regression classifier with an ElasticNet penalty ($\lambda = 1$) was fitted to the training data in order to classify tumors as belonging to each of the VMSs (VMS1 through VMS6). This classifier was then applied to the independent test set composed of the remaining, unlabeled, 30% of TCGA samples. Default thresholds for logistic regression were kept such that samples assigned a given VMS score of greater than 0.5 were labeled as belonging to that VMS. The class-weighted accuracy of the multiclass classifier on the test set was 95.7%, with ROC AUCs of 0.98, 0.97, 0.98, 0.99, 0.81, and 0.98 for VMS1, VMS2, VMS3, VMS4, VMS5, and VMS6, respectively.

Single-cell RNA-Seq data acquisition and analysis. Single-cell RNA-Seq data for endothelial cells isolated from human lung tumors was prepared as described previously (29) and were downloaded from ENDOTHELIOMICS (https://endotheliomics.shinyapps.io/) (Supplemental Figure 7A). *t*-Distributed stochastic neighbor embedding (*t*-SNE) plots demonstrating the expression of all 24 vascular signaling hub genes (Supplemental Figure 7, A and B) were generated using the Lung Tumor ECTax database (https://www.vibcancer.be/software-tools/lungTumor_ECTax).

Estimation of pericyte abundance. *RGS5* expression, which has been used as a quantitative marker of pericyte coverage (81), was used to estimate pericyte abundance in TCGA tumor samples. Single-cell RNA-Seq data for human testes were accessed from the PanglaoDB database (data set SRA667709; https://www.ncbi.nlm.nih.gov/sra/SRA667709) to confirm pericyte-specific expression of *RGS5*. Spearman's rank correlation coefficient and the corresponding 2-tailed probability values were calculated to test the relationship between the EI score and *RGS5* expression in all tumors, and Pearson's correlation coefficient was used to test the relationship between the EI score and *RGS5* expression in individual tumor types. Different methods were used to calculate correlation coefficients in individual cancer type data sets (Pearson's) and in all tumors aggregated (Spearman's), because the relationship between *RGS5* expression and the EI score behaved linearly in individual data sets but appeared nonlinear in the pan-cancer aggregate.

Estimation of stromal content. Stromal estimates were generated using the ESTIMATE (estimation of stromal and immune cells in malignant tumor tissues using expression data) algorithm (https://bioinformatics.mdanderson.org/public-software/estimate/), which predicts tumor cellularity and the infiltration of normal stromal cells using GSEA to estimate the enrichment of a "stromal signature" (35). The correlation between the EI score and stromal fraction or tumor purity in all human tumors was calculated using Pearson's correlation coefficient (*R* value) (Supplemental Figure 5F), which was then converted to a *t* statistic, and the corresponding *P* value was listed.

Statistics. Statistical comparisons of microvessel density between hypervascular and hypovascular groups were performed using an unpaired Student's *t* test (Figure 2, E and K). The correlation between the EI score and MVD in all human and mouse tumors was calculated using the Pearson's correlation coefficient (Figure 2, F and L). Pearson's *R* values were then converted to a *t* statistic, and the corresponding *P* value was listed.

The over- or underexpression of either *VEGFA*, *CDH1*, *NOTCH3*, or *MMP9* appeared as a characteristic of each vascular microenvironment (Figure 4B). To determine whether these genes were truly over- or underexpressed in specific vascular subtypes, the raw expression levels of these genes were compared across all 6 subtypes. Significant differences in the expression of these genes across vascular microenvironment subtype was first calculated using the Kruskal-Wallis rank-sum test. Significance of the over- or underexpression of these genes in specific vascular subtypes compared with the others was calculated using Conover's test of multiple pairwise comparisons. The FDR was determined using the Benjamin-Hochberg procedure for multiple-comparison correction.

Fisher's exact test was used to compare the proportion of VMS2 breast cancers that were of the "basal" molecular subtype (74%) compared with the abundance of the "basal" molecular subtype in breast cancers of vascular subtypes other than VMS2 (16%) (Figure 5). To determine whether specific vascular subtypes had abnormally high or low vessel density (as estimated by the EI score) in individual cancer types, the Kruskal-Wallis rank-sum test was performed. The Kruskal-Wallis test was chosen as opposed to 1-way ANOVA because Kruskal-Wallis is a nonparametric test, and the EI score does not follow a

normal distribution in every cancer type (Figure 3A). First, the presence of significant differences in EI scores across vascular subtypes in each of 31 cancer types was detected by Kruskal-Wallis rank-sum test and reported (Figure 6A). Associations between specific vascular subtypes and EI score were then tested using Conover's test of multiple pairwise comparisons. Vascular subtypes that accounted for fewer than 5 tumors in a given cancer type were removed from the analysis. FDRs were then calculated for each pairwise comparison using the Benjamin-Hochberg procedure for corrections and reported here.

Fisher's exact test was used to determine the significance of overlap between genes upregulated in hypervascular tumors (EI >0.9) across cancers with the gene signatures of 13 previously described endothelial cell phenotypes (29). The resulting P values were corrected for multiple hypothesis testing and reported as FDRs.

Study approval. For validation studies, samples were obtained from patients with PDAC who underwent surgery at the Columbia Pancreas Center using a protocol approved by the Columbia University Medical Center ethics committee (IRB no. AAAB2667), following informed consent.

Author contributions
BMK, RGA, KPO, RBF, and BZS designed the study. BMK, MDW, and HCM conducted experiments. BMK, MDW, and HCM acquired the data. BMK, RGA, GWS, JL, HCM, KPO, RBF, BZS, and AL analyzed the data. BMK, RBF, and BZS wrote the manuscript.

Address correspondence to: Ben Z. Stanger or Robert B. Faryabi, 421 Curie Blvd., 512 BRB II/III, Philadelphia, PA 19104, USA. Phone: 215.746.5559; Email: bstanger@upenn.edu (BZF). Phone: 215.573.8220; Email: faryabi@pennmedicine.upenn.edu (RBF).

HCM's present address is: Klinik und Poliklinik fur Innere Medizin II, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany.

1. Albini A, et al. Cancer prevention by targeting angiogenesis. *Nat Rev Clin Oncol*. 2012;9(9):498–509.

2. Gimbrone MA Jr., et al. Tumor dormancy in vivo by prevention of neovascularization. *J Exp Med*. 1972;136(2):261–276.

3. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–674.

4. Keating GM. Bevacizumab: a review of its use in advanced cancer. *Drugs*. 2014;74(16):1891–1925.

5. Nagy JA, Dvorak HF. Heterogeneity of the tumor vasculature: the need for new tumor blood vessel type-specific targets. *Clin Exp Metastasis*. 2012;29(7):657–662.

6. Zhao Q, et al. Single-cell transcriptome analyses reveal endothelial cell heterogeneity in tumors and changes following antiangiogenic treatment. *Cancer Res*. 2018;78(9):2370–2382.

7. Hasan J, et al. Intra-tumoural microvessel density in human solid tumours. *Br J Cancer*. 2002;86(10):1566–1577.

8. Offersen BV, et al. Comparison of methods of microvascular staining and quantification in prostate carcinoma: relevance to prognosis. *APMIS*. 2002;110(2):177–185.

9. Sandlund J, et al. Evaluation of CD31 (PECAM-1) expression using tissue microarray in patients with renal cell carcinoma. *Tumour Biol*. 2007;28(3):158–164.

10. Tanaka F, et al. Evaluation of angiogenesis in non-small cell lung cancer: comparison between anti-CD34 antibody and anti-CD105 antibody. *Clin Cancer Res*. 2001;7(11):3410–3415.

11. Maurer C, et al. Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. *Gut*. 2019;68(6):1034–1043.

12. Erkan M, et al. Cancer-stellate cell interactions perpetuate the hypoxia-fibrosis cycle in pancreatic ductal adenocarcinoma. *Neoplasia*. 2009;11(5):497–508.

13. Koong AC, et al. Pancreatic tumors show high levels of hypoxia. *Int J Radiat Oncol Biol Phys*. 2000;48(4):919–922.

14. Olive KP, et al. Inhibition of hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science*. 2009;324(5933):1457–1461.

15. Vasseur S, et al. Hypoxia induced tumor metabolic switch contributes to pancreatic cancer aggressiveness. *Cancers (Basel)*. 2010;2(4):2138–2152.

16. Li J, et al. Tumor cell-intrinsic factors underlie heterogeneity of immune cell infiltration and response to immunotherapy. *Immunity*. 2018;49(1):178–193.e7.

17. Grignon DJ, Che M. Clear cell renal cell carcinoma. *Clin Lab Med*. 2005;25(2):305–16.

18. Akslen LA, Livolsi VA. Increased angiogenesis in papillary thyroid carcinoma but lack of prognostic importance. *Hum Pathol*. 2000;31(4):439–442.

19. Bialas M, et al. Study of microvessel density and the expression of vascular endothelial growth factors in adrenal gland pheochromocytomas. *Int J Endocrinol*. 2014;2014:104129.

20. Goldenberg JD, et al. Well-differentiated thyroid carcinomas: p53 mutation status and microvessel density. *Head Neck*. 1998;20(2):152–158.

21. Segal K, et al. Angiogenesis in follicular tumors of the thyroid. *J Surg Oncol*. 1996;63(2):95–98.

22. Van den Berg R, et al. Management of paragangliomas. Clinical and angiographic aspects. *Interv Neuroradiol*. 2002;8(2):127–134.

23. Ntellas P, et al. Microvessel density and impact of angiogenesis on survival of resected pancreatic cancer patients: a systematic review and meta-analysis. *Pancreas*. 2019;48(2):233–241.

24. Aziz SA, et al. Vascularity of primary and metastatic renal cell carcinoma specimens. *J Transl Med*. 2013;11:15.

25. MacLennan GT, Bostwick DG. Microvessel density in renal cell carcinoma: lack of prognostic significance. *Urology*. 1995;46(1):27–30.

26. Gerhardt H, Betsholtz C. Endothelial-pericyte interactions in angiogenesis. *Cell Tissue Res*. 2003;314(1):15–23.

27. Bergers G, Song S. The role of pericytes in blood-vessel formation and maintenance. *Neuro*

Oncol. 2005;7(4):452–464.

28. Franzén O, et al. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)*. 2019;2019:baz046.

29. Goveia J, et al. An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates. *Cancer Cell*. 2020;37(1):21–36.e13.

30. Folkman J. Role of angiogenesis in tumor growth and metastasis. *Semin Oncol*. 2002;29(6 Suppl 16):15–18.

31. Edwards JG, et al. Angiogenesis is an independent prognostic factor in malignant mesothelioma. *Br J Cancer*. 2001;85(6):863–868.

32. Foss AJ, et al. Microvessel count predicts survival in uveal melanoma. *Cancer Res*. 1996;56(13):2900–2903.

33. Hong WG, et al. Clinicopathological significance and prognostic role of microvessel density in gastric cancer: A meta-analysis. *Pathol Res Pract*. 2017;213(12):1459–1463.

34. Hu X, et al. Prognostic value of microvessel density in cervical cancer. *Cancer Cell Int*. 2018;18:152.

35. Yoshihara K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.

36. Tewari KS, et al. Final overall survival of a randomized trial of bevacizumab for primary treatment of ovarian cancer. *J Clin Oncol*. 2019;37(26):2317–2328.

37. Abramsson A, et al. Endothelial and nonendothelial sources of PDGF-B regulate pericyte recruitment and influence vascular pattern formation in tumors. *J Clin Invest*. 2003;112(8):1142–1151.

38. Stockmann C, et al. The impact of the immune system on tumor: angiogenesis and vascular remodeling. *Front Oncol*. 2014;4:69.

39. Wang FT, et al. Cancer-associated fibroblast regulation of tumor neo-angiogenesis as a therapeutic target in cancer. *Oncol Lett*. 2019;17(3):3055–3065.

40. Barrallo-Gimeno A, Nieto MA. The snail genes as

inducers of cell movement and survival: implications in development and cancer. *Development*. 2005;132(14):3151–3161.

41. Cano A, et al. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat Cell Biol*. 2000;2(2):76–83.

42. Lee SW, et al. Snail as a potential target molecule in cardiac fibrosis: paracrine action of endothelial cells on fibroblasts through snail and CTGF axis. *Mol Ther*. 2013;21(9):1767–1777.

43. Hayes AJ, et al. Serum vascular endothelial growth factor as a tumour marker in soft tissue sarcoma. *Br J Surg*. 2004;91(2):242–247.

44. Kilvaer TK, et al. The VEGF- and PDGF-family of angiogenic markers have prognostic impact in soft tissue sarcomas arising in the extremities and trunk. *BMC Clin Pathol*. 2014;14(1):5.

45. Kilvaer TK, et al. Profiling of VEGFs and VEGFRs as prognostic factors in soft tissue sarcoma: VEGFR-3 is an independent predictor of poor prognosis. *PLoS One*. 2010;5(12):e15368.

46. Yudoh K, et al. Concentration of vascular endothelial growth factor in the tumour tissue as a prognostic factor of soft tissue sarcomas. *Br J Cancer*. 2001;84(12):1610–1615.

47. Berk V, et al. Predictive significance of VEGF and HIF-1α expression in patients with metastatic colorectal cancer receiving chemotherapy combinations with bevacizumab. *Asian Pac J Cancer Prev*. 2015;16(14):6149–6154.

48. Dowlati A, et al. Cell adhesion molecules, vascular endothelial growth factor, and basic fibroblast growth factor in patients with non-small cell lung cancer treated with chemotherapy with or without bevacizumab—an Eastern Cooperative Oncology Group Study. *Clin Cancer Res*. 2008;14(5):1407–1412.

49. Lebellec L, et al. Prognostic and predictive factors for angiosarcoma patients receiving paclitaxel once weekly plus or minus bevacizumab: an ancillary study derived from a randomized clinical trial. *BMC Cancer*. 2018;18(1):963.

50. Miles DW, et al. Biomarker results from the AVADO phase 3 trial of first-line bevacizumab plus docetaxel for HER2-negative metastatic breast cancer. *Br J Cancer*. 2013;108(5):1052–1060.

51. Van Cutsem E, et al. Bevacizumab in combination with chemotherapy as first-line therapy in advanced gastric cancer: a biomarker evaluation from the AVAGAST randomized phase III trial. *J Clin Oncol*. 2012;30(17):2119–2127.

52. Perren TJ, et al. A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med*. 2011;365(26):2484–2496.

53. Burger RA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med*. 2011;365(26):2473–2483.

54. Cheng SH, et al. Prognostic role of microvessel density in patients with renal cell carcinoma: a meta-analysis. *Int J Clin Exp Pathol*. 2014;7(9):5855–5863.

55. Bais C, et al. Tumor microvessel density as a potential predictive marker for bevacizumab benefit: GOG-0218 biomarker analyses. *J Natl Cancer Inst*. 2017;109(11):djx066.

56. McDermott DF, et al. Clinical activity and molecular correlates of response to atezolizumab alone or in combination with bevacizumab versus sunitinib in renal cell carcinoma. *Nat Med*. 2018;24(6):749–757.

57. Thomas M, et al. A novel angiopoietin-2 selective fully human antibody with potent anti-tumoral and anti-angiogenic efficacy and superior side effect profile compared to pan-angiopoietin–1/–2 inhibitors. *PLoS One*. 2013;8(2):e54923.

58. Jain RK. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. *Science*. 2005;307(5706):58–62.

59. Franco M, et al. Pericytes promote endothelial cell survival through induction of autocrine VEGF-A signaling and Bcl-w expression. *Blood*. 2011;118(10):2906–2917.

60. Bergers G, et al. Benefits of targeting both pericytes and endothelial cells in the tumor vasculature with kinase inhibitors. *J Clin Invest*. 2003;111(9):1287–1295.

61. Bergers G, Hanahan D. Modes of resistance to anti-angiogenic therapy. *Nat Rev Cancer*. 2008;8(8):592–603.

62. Benjamin LE, et al. A plasticity window for blood vessel remodelling is defined by pericyte coverage of the preformed endothelial network and is regulated by PDGF-B and VEGF. *Development*. 1998;125(9):1591–1598.

63. Bergers G, et al. Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nat Cell Biol*. 2000;2(10):737–744.

64. Egeblad M, Werb Z. New functions for the matrix metalloproteinases in cancer progression. *Nat Rev Cancer*. 2002;2(3):161–174.

65. Kayamori K, et al. NOTCH3 is induced in cancer-associated fibroblasts and promotes angiogenesis in oral squamous cell carcinoma. *PLoS One*.

2016;11(4):e0154112.

66. Limoge M, et al. Tumor-fibroblast interactions stimulate tumor vascularization by enhancing cytokine-driven production of MMP9 by tumor cells. *Oncotarget*. 2017;8(22):35592–35608.

67. Lin S, et al. Non-canonical NOTCH3 signalling limits tumour angiogenesis. *Nat Commun*. 2017;8:16074.

68. Liu H, et al. Notch3 is critical for proper angiogenesis and mural cell investment. *Circ Res*. 2010;107(7):860–870.

69. Fields GB. Mechanisms of action of novel drugs targeting angiogenesis-promoting matrix metalloproteinases. *Front Immunol*. 2019;10:1278.

70. Giraudo E, et al. An amino-bisphosphonate targets MMP-9-expressing macrophages and angiogenesis to impair cervical carcinogenesis. *J Clin Invest*. 2004;114(5):623–633.

71. Kang H, et al. Notch3-specific inhibition using siRNA knockdown or GSI sensitizes paclitaxel-resistant ovarian cancer cells. *Mol Carcinog*. 2016;55(7):1196–1209.

72. Barrett T, et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res*. 2013;41(Database issue):D991–995.

73. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc* B. 2005;67(2):301–320.

74. Karslioglu Y, et al. Chalkley method in the angiogenesis research and its automation via computer simulation. *Pathol Res Pract*. 2014;210(3):161–168.

75. Cox DR. Regression models and life-tables. *J R Stat Soc B*. 1972;34(2):187–220.

76. Schoenfeld D. Partial residuals for the proportional hazards regression-model. *Biometrika*. 1982;69(1):239–241.

77. Law CW, et al. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29.

78. Subramanian A, 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):15545–15550.

79. Fisher RA. Answer to question 14 on combining independent tests of significance. *Am Statistician 2*. 1948;2(5):30–31.

80. Rousseeuw PJ. Silhouettes — a graphical aid to the interpretation and validation of cluster-analysis. *J Comput Appl Math*. 1987;20(1):53–65.

81. Mitchell TS, et al. RGS5 expression is a quantitative measure of pericyte coverage of blood vessels. *Angiogenesis*. 2008;11(2):141–151.