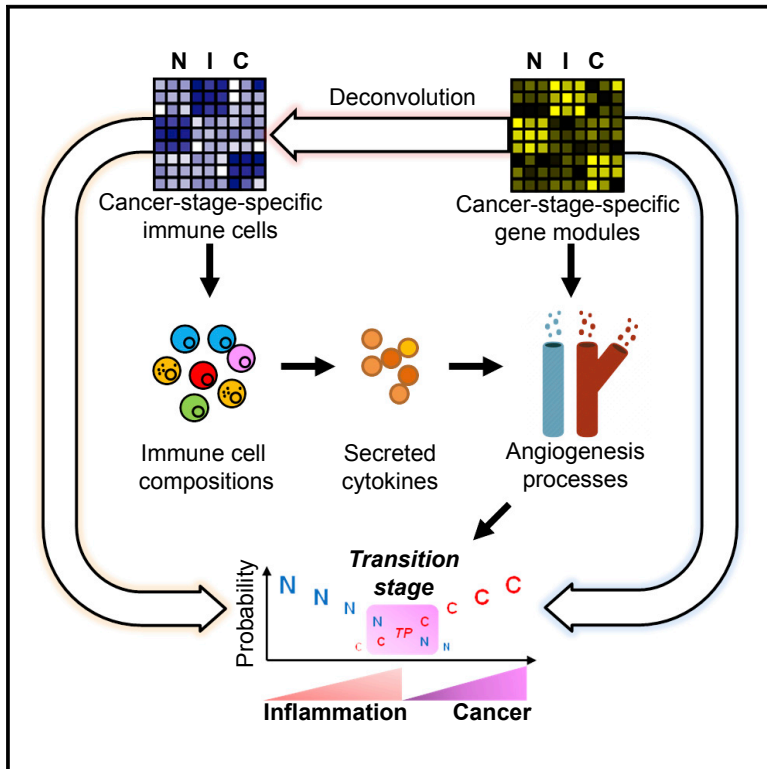


Cell Reports

Immune Cell Types and Secreted Factors Contributing to Inflammation-to-Cancer Transition and Immune Therapy Response

Graphical Abstract



Authors

Xingwei Chen, Chi Xu, Shengjun Hong, ..., Joseph McDermott, Yonglin Mu, Jing-Dong J. Han

Correspondence

jdhan@picb.ac.cn

In Brief

Chen et al. develop the SwitchDetector package for transcriptome module detection during inflammation-to-cancer (I2C) stage transitions. They show that angiogenesis is a common critical event for I2C in multiple cancers. The data also suggest that immune cells and secreted cytokines contribute to the I2C transition.

Highlights

- Angiogenesis is common during the inflammation-to-cancer (I2C) transition
- Data suggest link between immune cells, cytokines, and angiogenesis for I2C in the liver
- I2C cytokines and angiogenesis predict cancer survival and immune therapy response
- SwitchDetector package and I2C database are freely available



Immune Cell Types and Secreted Factors Contributing to Inflammation-to-Cancer Transition and Immune Therapy Response

Xingwei Chen,^{1,2,3} Chi Xu,^{1,2,3} Shengjun Hong,^{1,2,3} Xian Xia,^{1,2} Yaqiang Cao,^{1,2} Joseph McDermott,¹ Yonglin Mu,^{1,2} and Jing-Dong J. Han^{1,2,4,*}

¹Key Laboratory of Computational Biology, CAS Center for Excellence in Molecular Cell Science, Collaborative Innovation Center for Genetics and Developmental Biology, Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 320 Yue Yang Road, Shanghai 200031, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³These authors contributed equally

⁴Lead Contact

*Correspondence: jdhan@picb.ac.cn

<https://doi.org/10.1016/j.celrep.2019.01.080>

SUMMARY

Although chronic inflammation increases many cancers' risk, how inflammation facilitates cancer development is still not well studied. Recognizing whether and when inflamed tissues transition to cancerous tissues is of utmost importance. To unbiasedly infer molecular events, immune cell types, and secreted factors contributing to the inflammation-to-cancer (I2C) transition, we develop a computational package called "SwitchDetector" based on liver, gastric, and colon cancer I2C data. Using it, we identify angiogenesis associated with a common critical transition stage for multiple I2C events. Furthermore, we infer infiltrated immune cell type composition and their secreted or suppressed extracellular proteins to predict expression of important transition stage genes. This identifies extracellular proteins that may serve as early-detection biomarkers for pre-cancer and early-cancer stages. They alone or together with I2C hallmark angiogenesis genes are significantly related to cancer prognosis and can predict immune therapy response. The SwitchDetector and I2C database are publicly available at www.inflammation2cancer.org.

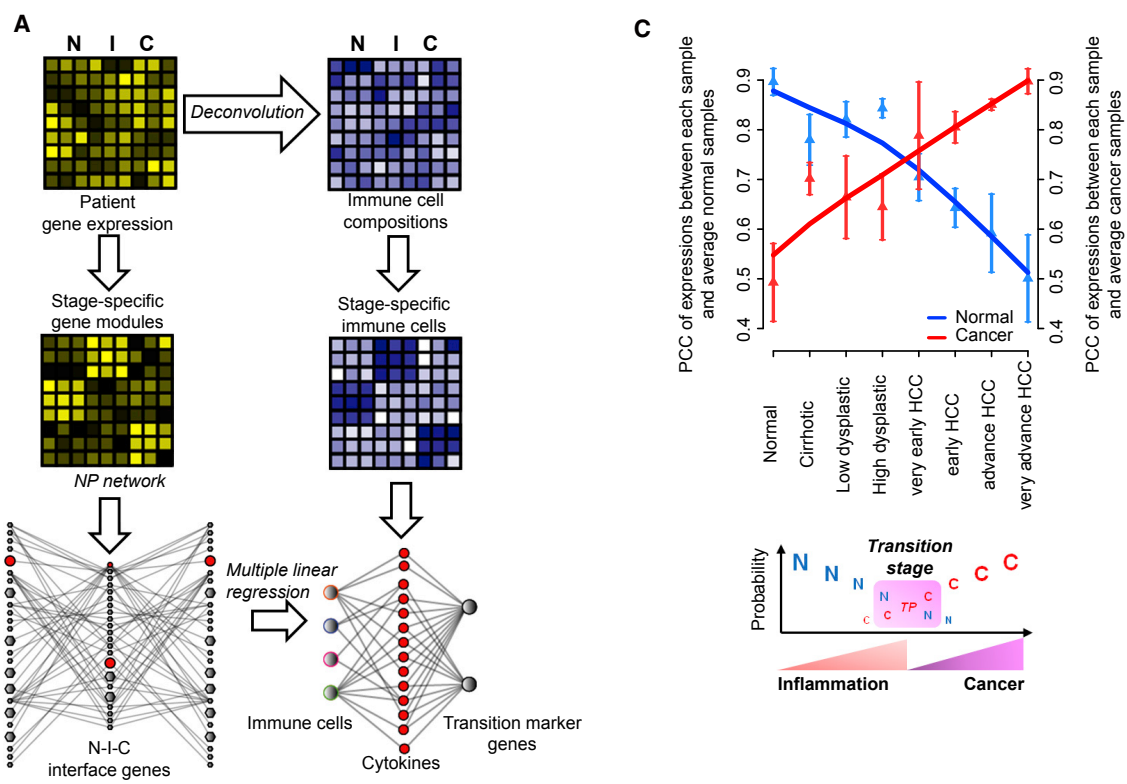
INTRODUCTION

Cancer development is a multifactorial and complex process, where both mutations in cancer cells and continuous stimulus in their microenvironment affect cancer development and prognosis. Inflammatory microenvironments are known to contribute to different stages of cancer development, including carcinogenesis, cancer development, and metastasis (Cousens et al., 2013; Grivennikov et al., 2010). Similarly, chronic and non-resolving inflammation, now recognized as a hallmark of cancer (Hanahan and Weinberg, 2011), can increase the risk of many cancers, such as hepatocellular carcinoma (HCC) and colon cancer, and can facilitate cancer progression (Marelli et al., 2017; Nickoloff et al., 2005). However, the molecular

mechanisms of how inflammation facilitates cancer development are still not well studied. Therefore, an urgent and challenging task is to develop the capability to recognize whether and when an inflamed tissue transitions to a cancerous tissue. Many studies have been carried out to tackle this challenge, and it has been found that cytokines and other soluble factors secreted by infiltrating immune cells can reprogram local cells in inflamed tissue microenvironments and are thus the focus of most biomarker studies (Grivennikov et al., 2010; Landskron et al., 2014; Lu et al., 2006). Secreted factors, such as cytokines and chemokines, are positively and negatively involved in cell transformation and malignancy (Landskron et al., 2014); many are strongly associated with cancer development, and some are used in cancer diagnosis and prognosis (Prasad and McCullough, 2013; Shaw et al., 2014; Yoo et al., 2009), such as IL-6, IL-4, and IL-8 (Culig, 2011), while others have even been used in cancer therapies (Dranoff, 2004; Lee and Margolin, 2011). One role of secreted factors is to regulate angiogenesis, such as with VEGF, IL-1, and IL-13 that promote angiogenesis, while others such as IL-10 and TIMP1 inhibit it (Coussens et al., 2013; Nishida et al., 2006). Angiogenesis ensures the supply of oxygen and nutrients for cancer cells and plays critical roles in cancer initiation and development (Nishida et al., 2006). However, systems-level analyses to unbiasedly uncover mechanisms and biomarkers for early-stage detection of the transition from inflammation to cancer are still lagging behind.

Here, using a network analysis approach, we integrated the expression profiles of normal tissues and tissues with inflammation, using samples from liver, esophageal, and colon cancers. Together with protein-protein interactomes and gene expression signatures of different immune cells, we identified interface genes between normal, inflammation, and cancer backgrounds and inferred regulatory immune cell types and their activated or repressed secreted factors and cytokine receptors (SFCRs) at the inflammation-to-cancer (I2C) transition (Figure 1A). We identified a transition stage between advanced inflammation and early cancerous transformation stages that not only governs the transformation process but also provides pre-cancer cellular and molecular markers and prognosis and immune therapy response predictors.





(legend on next page)

RESULTS

Dynamic Transcriptome Modules between Normal, Inflammation, and Cancer Stages

In order to study the molecular events of I2C transition, we collected public high-throughput omics datasets of inflammation and cancer samples from the Gene Expression Omnibus (GEO) database (Barrett et al., 2013), processed them from raw data, and individually examined data quality. We required each dataset to contain both inflammation and cancer samples (STAR Methods) and marked the sample stages according to the original sample descriptions. For this analysis, we collected 38 GEO datasets (Table S1A) as a byproduct of our literature search, and we made all datasets searchable and downloadable in our I2C database at www.inflammation2cancer.org. However, since not all datasets completely cover the I2C transitions stages, not all datasets are used for developing our computational methods to dissect the I2C transition.

As tissue-matched samples for normal, inflammation, and cancer stages are the most available and comprehensively profiled for three common human cancers—liver (GSE6764) (Wurm-bach et al., 2007), colon (GSE4183) (Galamb et al., 2008), and esophageal cancer (GSE19529) (Saadi et al., 2010)—we focused on these to examine the I2C transition. Among them, we first used the liver cancer dataset that contains three and four pathologically progressive substages for inflammation and cancer states, respectively, to dissect the dynamic transcriptome changes during I2C transition using our proposed computational methods, then repeated the analysis on the colon and esophageal cancer datasets to verify the findings. The liver data contain 75 samples, with expression datasets covering 8 progressive stages of hepatitis C virus (HCV)-induced HCC, which was the most comprehensive profile of an I2C process. Samples were divided into three groups—10 for physiologically normal stages and 30 and 35 for three stages of progressive inflammation and four cancer stages, respectively (Table S1B). The well-established progression order of the samples allowed us to compare transitions between different stages and to identify the stage transition for the I2C process at a transcriptome level. We identified 1,417 differentially expressed genes (DEGs) between any two stages (STAR Methods) using RankProd (Hong et al., 2006). To detect dynamic transcriptional modules, we used an auto-optimizing BIC-SKmeans algorithm (Zhang et al., 2013) and assigned DEGs to six clusters (numbered from 0 to 5) according to their first appearance stages (Figure 1B; Figure S1A). Most clusters specifically showed high expression in only one of the three stages. Gene ontology (GO) function

enrichment analysis showed the normal stage, specifically highly expressed cluster 0 (normal module), was enriched for steroid and retinol metabolism. The inflammation stage, specifically highly expressed cluster 3 (inflammation module), was enriched for immune response, antigen processing and presentation, and cytokine-cytokine receptor interaction. The early cancer stage, highly expressed cluster 4 (cancer module), was enriched for regulation of cell adhesion, in particular, ECM-receptor interaction, in which abnormalities can deregulate stromal cells, cancer-associated angiogenesis, and inflammation, causing a tumorigenic microenvironment (Lu et al., 2012). Finally, the late cancer stage, highly expressed cluster 5 (late cancer module), was enriched for p53 signaling (Vazquez et al., 2008) and cell cycle, in which deregulation underlies abnormal cell proliferation and promotes genetic instability (Williams and Stoeber, 2012) (Figure 1B; Tables S2 and S3). Interestingly, secreted-factor-related processes were significantly enriched in inflammation-stage-specific modules. In addition, the I2C transition modules (clusters 3, 4, and 5) were similarly observed in the colon (GSE4183) and esophageal (GSE19529) cancer data with similar functional enrichment, as grouped together by principle-component analysis (PCA) of function enrichment (Figure S2C). All clusters 4 and 5 in the liver, colon, and esophageal cancers were enriched for cell adhesion and cell cycle, respectively; cluster 3 of liver and colon cancers were enriched for immune response, which implies common functional changes contributing to the I2C transition (Figures S2A and S2B; Tables S2 and S3; STAR Methods).

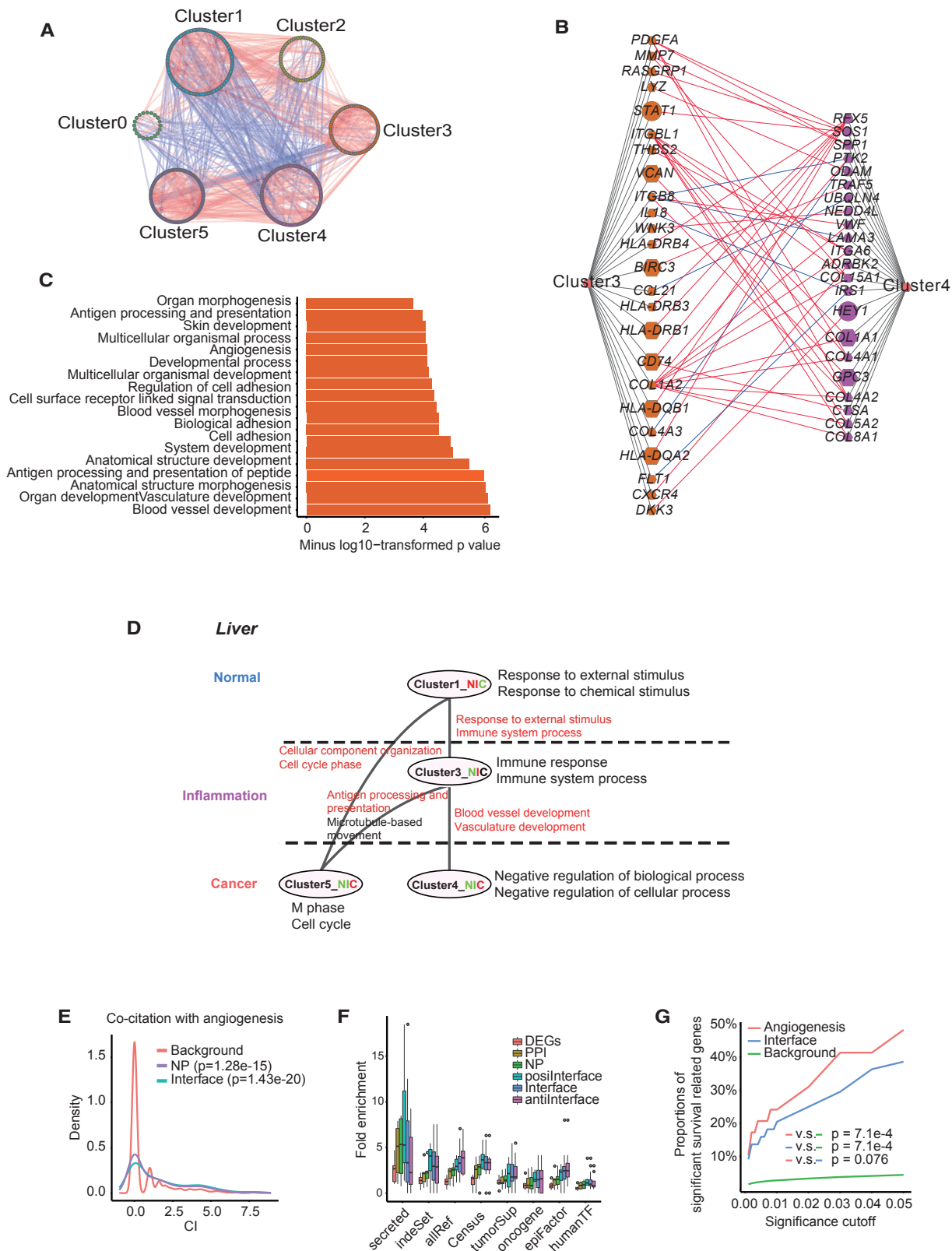
Intriguingly, when regarding the normal stage as the start of normal-inflammation-cancer (NIC) and the late cancer stage as the end, transcriptome similarities to the two endpoints cross at the junction between the advanced inflammation and the very early cancer stages, forming an expression pattern transition stage between two alternative transcription states (similarity to normal drastically dropped and to late cancer drastically rose), whereas within the inflammation or cancer stages, the substages are relatively similar but did show progressive dynamic changes (Figures 1B and 1C). Thus, the transition between inflammation and cancer modules (clusters 3 and 4) might be a transition stage for I2C, and early markers predicting cancer occurrence might be identified before or at this point.

Angiogenesis Gene Expression as a Hallmark and Key Regulator of I2C Transition

To find mechanisms between state transitions, regulators linking before and after a stage switch are the key to study. Our previous

Figure 1. Workflow of I2C Transition Stage Detection and Gene Expression Module Analysis of Human Liver Cancer Progression Stages

(A) Workflow of inflammation-to-cancer (I2C) transition stage detection. N, normal stage; I, inflammation stage; C, cancer stage.
(B) Gene expression modules specifically expressed at normal, inflammation, and cancer stages in hepatocellular carcinoma (HCC), identified from analysis of expression changes of DEGs pairwise comparisons between the three stages. I, inflammation (includes Ci, cirrhosis [13 samples]; Ld, low-grade dysplastic [10 samples]; and Hd, high-grade dysplastic [7 samples] liver samples); C, cancer (includes Ve, very early [8 samples]; Ea, early [10 samples]; Ad, advanced [7 samples]; and Va, very advanced [10 samples] HCC). Top 3 significantly enriched functional terms in each gene cluster are listed on the right. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes. See Tables S2 and S3 for more information.
(C) The Pearson correlation coefficient (PCC) of each sample's transcriptome with the average gene expression profiles of all genes in normal tissue or late-stage cancer displayed a crossover at the I2C transition, suggesting a turning point (TP) at the I2C transition. Each blue/red bar indicates the mean and standard deviation of the PCCs between samples' expression profiles and normal samples/late-stage cancer samples. Polynomial fitting curves of the mean values are also shown.



(legend on next page)

studies find that an expression-based positive-negative-correlated protein-protein interaction network (NP network) directly linking the neighboring stages can significantly enrich the switch regulators. These networks often reside in toggle switches like feedback loops, balancing the critical temporal or state switches (Huang et al., 2015; Peng et al., 2016; Wang et al., 2017; Xia et al., 2006; Xue et al., 2007). We therefore tested whether NP network analysis could capture regulators of I2C transition. We extracted a protein-protein interaction (PPI) network between inflammation-specific modules and cancer-specific modules (between I and C modules) for liver, colon, and esophageal cancers to create I2C module NP networks (Huang et al., 2015; Xia et al., 2006). We similarly constructed NP networks for the normal-to-inflammation (N2I) stages. The liver NP network contained 514 genes linked by 1,086 positive and 256 negative edges (Pearson correlation coefficient [PCC] > 0.22 or < -0.22 corresponding to $p < 0.05$) (Figure 2A; Figure S1B). Inside the NP network, we define genes directly connected by edges between different clusters as interface genes. Such interfaces have been shown to be enriched for genes with roles to drive stage transitions in NP networks (Huang et al., 2015; Xia et al., 2006). We considered interactions between cluster 3 (highly expressed in early inflammation that gradually tapered) and cluster 4 (initially expressed in the early cancerous stage), to straddle right across the transition stage observed at the transcriptome level (Figure 1C), and therefore named it the “I2C interface.” The interface contained 46 genes, which were mainly enriched for blood vessel development, cell adhesion, and angiogenesis functions (Figures 2B and C; Tables S4 and S5). Angiogenesis has been shown to be essential for tumor initiation and development and I2C transition (Kobayashi and Lin, 2009; Tassi and Wellstein, 2006). 29 of the 46 genes are known angiogenesis genes according to GO annotation and literature. Similarly, I2C interfaces in the colon and esophageal cancer datasets (which were also between their cluster 3 and cluster 4) were also enriched for blood vessel development or angiogenesis functions (Figure 2D; Table S4;

STAR Methods), which are the common features of I2C across the three tumor types. Unlike the liver and colon datasets where gene expressions are from whole tissues, in the GSE19529 esophageal data, fibroblasts were isolated from fresh esophageal tissues of squamous, Barrett’s, and cancer groups. Surprisingly, the I2C marker genes could be also observed at the I2C transition in these purified fibroblasts, suggesting that as part of the microenvironment of tumors and sensing the microenvironment (Van Linthout et al., 2014), they memorize the cytokines in their microenvironment and even produce similar cytokines to contribute to the pro-angiogenesis microenvironment at I2C. For example, among the interface 3-4 genes of the esophageal fibroblast data, we identified a cytokine-related gene, *TIMP1*, and three angiogenesis-related genes, *MMP2*, *MMP14*, and *HIF1A*. *TIMP1* is a glycoprotein and highly inducible by many cytokines (Ries, 2014). *TIMP1* has cytokine-like activities and influences cell growth, apoptosis, differentiation, angiogenesis, and oncogenesis (Ries, 2014). *MMP2* and *MMP14* play an indirect role in angiogenesis by promoting VEGF mobilization and generating antiangiogenic factors (Mook et al., 2004). *HIF1A* is involved in initiating angiogenesis and promoting tumor growth. Additionally, *TIMP1* is known to inhibit MMP proteins (Ries, 2014), and silencing the *HIF1 α* gene can effectively inhibit hypoxia-induced upregulation of *TIMP1* (Yang et al., 2006). To further confirm the NP network selected modules’ association with angiogenesis, we used CoCiter (Qiao et al., 2013) to calculate co-citation impact (CI) between the genes and the terms “angiogenesis” and “blood vessel.” Compared with all 1,417 DEGs in liver, NP network genes showed significantly larger CI values (t test; $p = 1.28\text{e-}15$), and interface genes showed a further increase of CI values (t test; $p = 1.43\text{e-}20$). In particular, genes in the cluster 3-4 interface (Figures 1B and 2E; Table S1C) had significantly higher CI values with angiogenesis (CoCiter; $p < 0.001$). As in cluster 3 and 4, interface angiogenesis genes had peak expression before or after the I2C transcriptome transition stage (Figure S1C).

Figure 2. NP Network Module Interfaces Are Enriched in Cancer-Related Genes and Angiogenesis

(A) A NP network generated between the six gene clusters from Figure 1C. Nodes represent genes, red edges represent positively correlated interactions between genes ($\text{PCC} > 0.22$, $p < 0.05$), and blue edges represent negatively correlated interactions ($\text{PCC} < -0.22$, $p < 0.05$). See Figure S1B for details of PCC cutoff selection.

(B) Interface between clusters 3 and 4 (see Tables S2 and S3 for details) in liver cancer. Nodes with large size represent cancer-related genes (small size nodes are not cancer-related genes), and circular nodes represent human transcription factors (TFs) (hexagon nodes are not human TFs) (STAR Methods). Red edges represent positive correlations between the genes, and blue edges represent negative correlations.

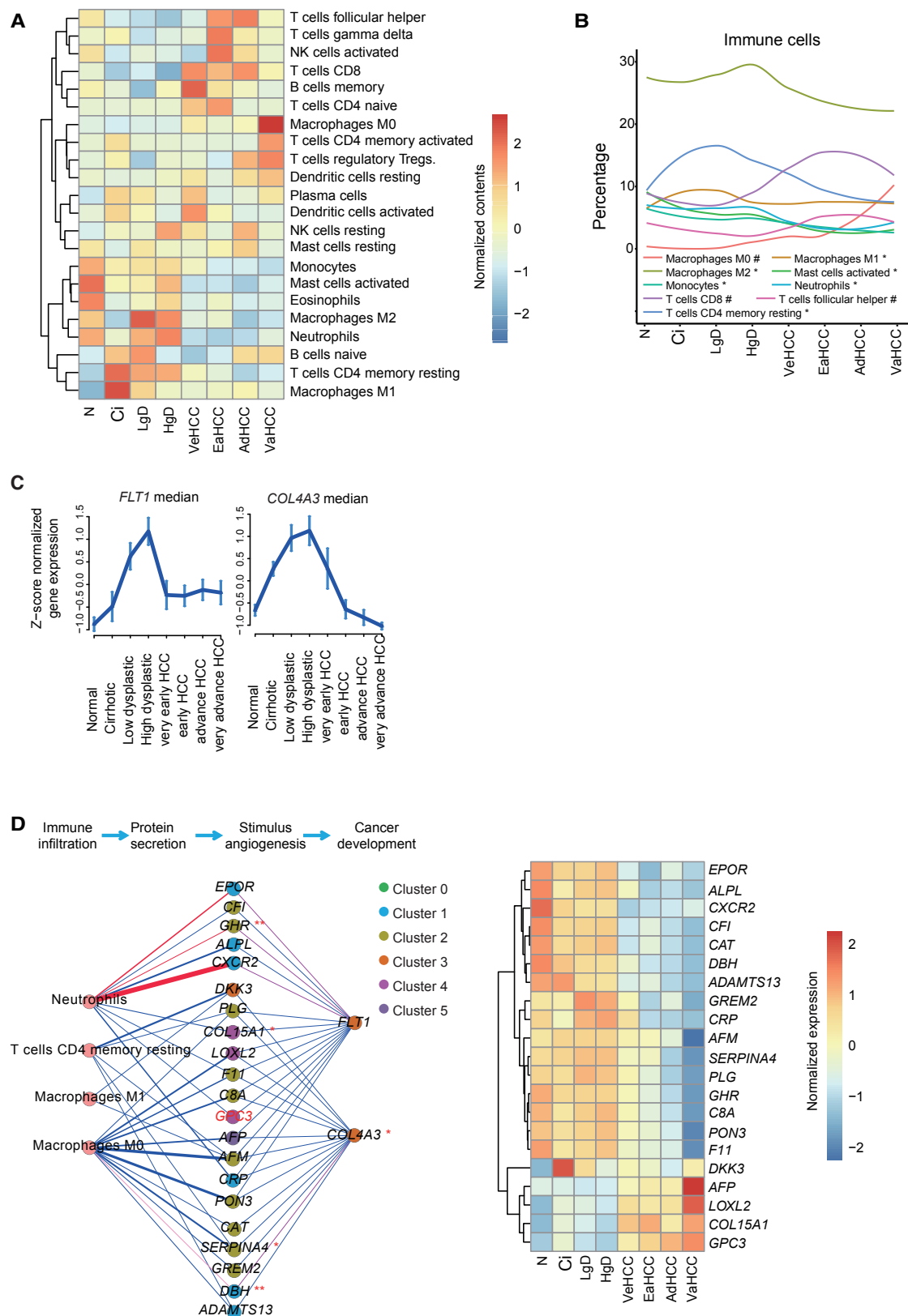
(C) GO terms enriched in interface genes between clusters 3 and 4. See Tables S4 and S5 for exhaustive information.

(D) Clusters and significantly interacted interface (interfaces between the 4 significantly interacted cluster pairs 1-3, 1-5, 3-4, and 3-5 [p value < 0.05] among all 15 interfaces; see STAR Methods) and their enriched GO terms during the I2C transition of HCC. Solid lines indicate interfaces between clusters that significantly interact (STAR Methods), and top 2 GO terms were listed for each interface and cluster. Red GO terms next to an interface indicate the functions enriched in that interface for at least one other cancer (colon or/and esophageal). Dashed lines separate the NIC process into three stages—normal, inflammation, and cancer—which include corresponding stage-specific gene clusters of each stage. N, normal stage; I, inflammation stage; C, cancer stage. Round nodes indicate gene clusters, where red/green letter N (I or C) indicates high/low expression of this cluster in the corresponding stages.

(E) Density distribution of co-citation impact (CI) values of DEGs, NP network genes, and interface genes in the NP network with “angiogenesis” and “blood vessel” terms.

(F) Fold enrichment of DEGs, PPI genes, NP network genes, interface genes, and interface genes with positive or negative interactions in the eight gene sets, which include secreted factors, human transcriptome factors, and reported cancer-related genes (STAR Methods).

(G) Proportions of genes that significantly predicted prognosis in at least 1 of 21 TCGA cancers that have more than 200 samples at different significance levels for I2C interface genes (blue), angiogenesis genes at the I2C interface (red), and all DEGs as background (green). Genes are sorted by each gene’s lowest prediction p values (Benjamini-Hochberg corrected) among the 21 TCGA cancers, at different p value cutoffs from 0.001 to 0.05, proportions of significant genes in each gene set were plotted, and paired Wilcoxon signed rank test was used to test the significance of proportion difference between different gene sets with the p values labeled on the graph.



(legend on next page)

As our analysis recaptured a well-known I2C transition event, angiogenesis, in three types of cancers, it suggests the NP network interface—in particular, the interface between inflammation to the very early cancerous stage—can capture important I2C mediators and biomarkers. Consistent with this, we found that the fold enrichment for cancer-related genes significantly increased in the NP network (t test; $p = 0.023$) and further increased in the interfaces (t test; $p = 0.002$), particularly in the I2C interface (fold enrichment = 4.59, Fisher exact test; $p = 2.3 \times 10^{-4}$) (Figure 2F). If interface genes are indeed important for cancer initiation and progression, we would also expect interface genes, including the angiogenesis genes, to predict cancer survival. As expected, when we examined the Cancer Genome Atlas (TCGA) database, the expression levels of the interface genes, especially the angiogenesis genes in the interface, significantly better predicted survival than when predicted with all DEGs (Figure 2G; STAR Methods; paired Wilcoxon signed rank test; $p = 7.1 \times 10^{-4}$ for predicting survival with either interface angiogenesis genes or all interface genes compared to with all DEGs; $p = 0.076$ for interface angiogenesis genes compared to with all interface genes). Interestingly, literature co-citation analysis showed that of all of the interface genes, 93% of them (43/46) showed at least one co-citation to “cancer” and “survival” (permutation test p value = 0.003 and 0.001), which further indicated a biological significance of the relationship between the interface genes and cancer survival.

Immune Cell Composition Change Associated with I2C Transition

As angiogenesis commonly appeared in I2C transition processes in different cancers, we investigated the major types of immune cells contributing to the rise of angiogenesis gene expression at the I2C transition. We inferred the relative immune cell type content in each of the samples during I2C transition using the CIBERSORT (Newman et al., 2015) deconvolution algorithm based on the expression signatures of 22 immune cell types (Figure 3A; Table S1D). 9 of the 22 cell types showed significant content differences between inflammation and cancer stages (t test; type I error set to 0.05), with resting CD4⁺ memory T cells, neutrophils, activated mast cells, monocytes, and M1 and M2 macrophages being more abundant in inflammation stages and CD8⁺ T cells, follicular helper T cells, and M0 macrophages being more abundant in cancer stages (Figure 3B; Figure S3A). For these nine I2C-associated immune cell types, we also observed content switches in other I2C datasets, including

four of liver cancer and four other cancer types (Table S1E; Figures S3B–S3D).

Immune Cell Secreted or Suppressed Factors Associated with I2C

We reasoned that if the I2C transition is a critical transition stage, genes showing peak expression before the transition might be early-detection biomarkers for the I2C transition and predictive of early events such as angiogenesis, whereas those peaking right at I2C might be cancer commitment markers and prognostic. Among all I2C interface angiogenesis genes, *FLT1* and *COL4A3* had peak expression right before the I2C transition (Figure 3C), and their expression level could be explained by immune-cell-type content (Figure S4A). Extracellular proteins secreted or suppressed by immune cells have also been known to contribute to angiogenesis (Watnick, 2012). When we replaced the angiogenesis gene set with all 46 I2C interface genes for analysis, we uncovered *FLT1* and *COL4A3* again and one additional gene, *DKK3*, which add up to 6.5% (3 out of 46) interface genes, similar to the ratio uncovered using I2C interface angiogenesis genes (6.9%, 2 out of 29). When extending the gene set to all DEGs, in total 10 genes were discovered, including *FLT1* and *COL4A3*, which is 0.7% of all DEGs (10 out of 1,417), relatively lower than the ratio using angiogenesis genes or the 46 I2C interface genes. These results are consistent with the assumption that the interface genes and in particular interface angiogenesis genes are likely I2C commitment markers.

We inferred a multiple linear regression (MLR) model between the 9 I2C-associated immune cell types' content and the expression levels of SFCRs in the liver I2C dataset (GSE6764) (STAR Methods; Figure S4B). The predicted SFCRs' expression values showed a significantly higher correlation to their real expression values (t test; $p = 4.25 \times 10^{-3}$) and explained more of the variance than the background (all 1,417 DEGs in Figure 1B; t test; $p = 0.025$; Figure S4D). When predicting each SFCR's high- or low-expression pattern, our model performed better than predicting randomly selected background genes, as indicated by the significantly higher area under curve (AUC) (t test; $p = 0.044$; Figure S4D). Altogether, 23 SFCRs were well predicted by the immune cells' contents with correlations (PCC) between predictions and real values > 0.68 ($p < 0.05$; STAR Methods), suggesting these 23 SFCRs are indicators of the immune cell composition in the I2C tissue environment (Figure S4B). Consistently, 22 of the 23 SFCRs have at least one literature co-citation with term “immune” (permutation test p value = 0.006) as identified

Figure 3. Immune-Cell-Type Deconvolution Revealed Immune Cell Content Switched between Inflammation and Cancer

(A) A heatmap showing the expression-based deconvolution-predicted relative contents of 22 immune cells across 8 stages of normal-inflammation-cancer transition.

(B) Average content of immune cells that change significantly between normal, inflammation, and cancer stages. Each immune cell's content was normalized to a percentage of composition in a total of 22 immune cells.

(C) Median expression pattern of *FLT1* and *COL4A3*.

(D) The interaction network of immune cells and SFCRs with content or expression that significantly contributed to predicting *COL4A3* and *FLT1* expressions. The interactions were determined by multiple linear regression model (left panel) and a heatmap showing expression changes of secreted factors and cytokine receptor genes (SFCRs) in DEGs in the 8 stages of normal-inflammation-cancer transition process (right panel; STAR Methods). Red node names represent genes annotated in the cancer gene databases.

* indicates the high- or low-expression patterns significantly correlated with HCC patients' prognosis. * indicates log-rank test $p < 0.1$ and ** for $p < 0.01$. N, normal samples; Ci, cirrhosis samples; LgD, low-grade dysplastic samples; HgD, high-grade dysplastic liver samples; VeHCC, very early HCC; EaHCC, early HCC; AdHCC, advanced HCC; VaHCC, very advanced HCC.

Table 1. Parameters of Multiple Linear Regression Models to Predict the Expression Level of COL4A3 and FLT1, Respectively

	COL4A3				FLT1			
	Estimated	Standard Error	t Value	Pr(> t)	Estimated	Standard Error	t Value	Pr(> t)
Intercept	2.760	1.498	1.842	0.070	5.454	1.265	4.313	5.6e-5
T cells CD4 memory resting	7.721	2.193	3.521	0.0008	2.400	1.851	1.297	0.199
Monocytes	−6.761	5.520	−1.225	0.225	−5.447	4.660	−1.169	0.247
Macrophages M1	7.328	4.661	1.572	0.121	2.945	3.935	0.748	0.457
Macrophages M2	−0.089	2.548	−0.035	0.972	−0.030	2.151	−0.014	0.989
Mast cells activated	−0.363	2.908	−0.125	0.901	0.932	2.455	0.379	0.706
Neutrophils	9.764	5.166	1.890	0.063	13.873	4.360	3.182	0.002
Macrophages M0	−2.699	2.443	−1.105	0.273	1.486	2.062	0.721	0.474
T cells CD8	2.777	2.850	0.974	0.333	1.144	2.406	0.475	0.636
T cells follicular helper	1.210	5.704	0.212	0.833	−0.797	4.815	−0.166	0.869

by CoCiter (Qiao et al., 2013), indicating a significant reported association of these genes with immune functions. For example, for the top co-cited “star” gene, C-X-C motif chemokine receptor 2 encoding gene, *CXCR2*, its deficiency or disruption of *CXCR2* suppresses pancreatic inflammation (Steele et al., 2015), attenuates murine model’s arthritis (Jacobs et al., 2010), and was found to participate in enhancing anti-PD1 efficacy (Highfill et al., 2014). For co-citation with different immune cell types, 18 of the 23 SFCRs were co-cited with at least one immune cell type (permutation test p value = 0.004). *CXCR2* also showed significant co-citations to the different cell types (permutation test p value = 0.001), with reported relationships to neutrophils, monocytes, and activated mast cells (literature co-citation count = 231, 43, and 1, respectively), covering three of the four immune cell types that we predicted to contribute significantly to the *CXCR2* gene. Surprisingly, among the immune cells, neutrophils and monocytes covered a majority of the co-citations (count = 658 and 344, respectively), while mast cells activated, T cell CD8, and macrophage M1 only showed 2, 2, and 1 co-citations, respectively, and no co-citations were detected for the other four immune cells. It suggests that there is a huge gap of knowledge between SFCRs and most immune cell types (except for neutrophils and monocytes). Our analyses and results, on the contrary, do not have such a study bias and thus can be more objective when filling such a gap.

Then, using the expression of SFCRs in a MLR to predict the expression of the two angiogenesis genes, *FLT1* and *COL4A3*, we found 17 and 14 out of the 23 immune-cell-associated SFCRs significantly predict the expression of *FLT1* and *COL4A3*, respectively (p < 0.05). Finally, to construct a network containing immune cells, SFCRs, and the two angiogenesis genes, we required that the immune cell types should significantly predict both the expression of *FLT1* and *COL4A3* and the SFCRs, thus the final network retained 4 immune cell types, 21 SFCRs, and the 2 key I2C interface angiogenesis genes (Figure 3D; Figure S4C; STAR Methods). Replacing the SFCRs by randomly selected gene sets or gene sets with unrelated GO terms does not give rise to such large set of input genes retained in the final network, making the false discovery rate (FDR) of getting the size of the SFCR network < 0.001 (Figure S4E). These SFCRs showed a switch in expression patterns from high to low or low to high at the I2C transition (Figure 3D; STAR Methods).

COL4A3 encodes type IV collagen, a major component of basement membranes that has cysteine-type endopeptidase activity and is involved in apoptotic processes, blood circulation, cell adhesion, and proliferation (Maeshima et al., 2000). Three types of immune cells, resting CD4+ memory T cells, M1 macrophages, and M0 macrophages, significantly contributed to predicting the expression level of *COL4A3* (Table 1). *FLT1*, fms related tyrosine kinase 1, encodes a VEGF receptor to induce angiogenesis, and neutrophils significantly contributed to predicting expression of *FLT1* (Table 1). Altogether, 17 out of the 21 SFCRs had at least 1 co-citation with the term “angiogenesis” (Qiao et al., 2013) (Table S1F). Among them *GHR*, *CXCR2*, *EPOR*, and *DBH* are known to promote angiogenesis, and *CXCR2*, *EPOR*, and *GHR* are cytokine receptors for interleukin, erythropoietin, and growth hormone, respectively. *CXCR2* was reported to promote angiogenesis and cancer growth (Keane et al., 2004; Matsuo et al., 2009; Yang et al., 2010). *DBH* encodes an oxidoreductase expressed in secretory vesicles and granules and exists in both soluble and membrane-bound forms. The expression of the *DBH* gene is regulated by cytokines and is required for angiogenesis (Chalothorn et al., 2005; Nakano et al., 2007; Ren et al., 2005; Xi et al., 2012; Yang et al., 2010).

To test the predictions’ association with angiogenesis, we examined these genes’ somatic-mutation-induced expression signatures from TCGA cancer transcriptomes and found that downregulated signatures in samples having the mutations of 12 of the 21 SFCRs identified above (*GPC3*, *FLT1*, *CXCR2*, *AFP*, *PON3*, *CRP*, *COL15A1*, *CAT*, *F11*, *COL4A3*, *ADAMTS13*, and *C8A*) were indeed significantly enriched for angiogenesis GO terms (Figure S5A; STAR Methods). 8 out of these 12 genes have been reported to be cancer biomarkers previously, while 3 out of the other 9 genes have been reported as cancer biomarkers (Table S1G).

We hypothesized that cancer microenvironment markers that act in an early pre-cancer stage might be not only useful for cancer prevention but could be also associated with cancer severity and prognosis if the same microenvironment plays a role in cancer initiation, progression, and recurrence or relapse. Therefore, we investigate whether the I2C transition and angiogenesis-associated SFCR expressions are in general associated with cancer severity by evaluating their ability to predict cancer survival in

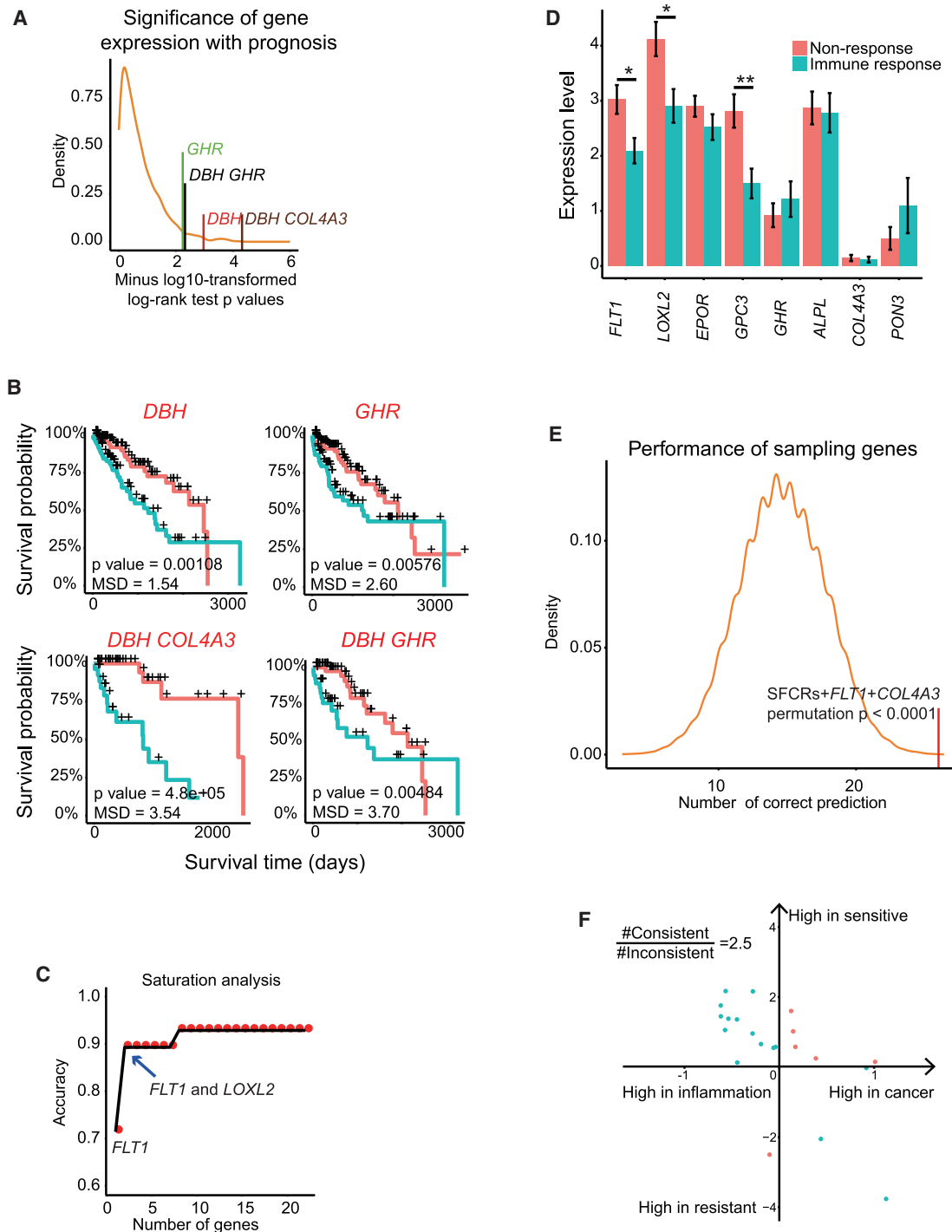


Figure 4. Contribution of Immune-Cell-Associated SFCRs to HCC Prognosis and Melanoma Response to Immune Therapy

(A) Distribution of survival analysis p values of TCGA liver cancer patients with high- or low-expression patterns of all DEGs. The survival analysis p values of SFCRs in Figure 3D are indicated by vertical lines.

(B) Survival curves of TCGA liver cancer patients divided by high or low expression (STAR Methods) of the SFCRs and angiogenesis genes in Figure 3D. Log-rank test p values and median survival time difference (MSD) fold changes are indicated in the bottom. Genes in red color represent high survival rates in groups with the respective genes' high-expression patterns. Benjamini-Hochberg (BH) correction was used to correct the multiple test p values.

(C) Prediction accuracy saturation analysis when using different number of SFCRs, together with *FLT1* and *COL4A3* genes to predict immune therapy response in melanoma (GSE78220).

(legend continued on next page)

the TCGA data. *DBH* and *GHR* significantly distinguished liver cancer patients' prognoses, with a mean of ~ 1.54 -fold difference in median survival time (log-rank test; $p = \sim 0.001$) for *DBH* and a mean of ~ 2.60 -fold difference in median survival time (log-rank test; $p = 0.011$ after Benjamini-Hochberg multiple testing correction) for *GHR* (Figures 4A and 4B). With a pairwise combination of these SFCRs and angiogenesis genes in Figure 3D, *DBH* and *COL4A3* further increased the difference between median survival times of bad and good prognoses for liver cancer patients to a mean of 3.54-fold (log-rank test; corrected $p = 4.8 \times 10^{-5}$; Figures 4A and 4B), which is much higher than the HCC pathological grade classification's 1.81-fold difference between stage I and stage II (or higher) in median survival time (Figure S5B; STAR Methods). The combination of *DBH* and *GHR* renders a 3.7-fold difference (log-rank test; corrected $p = 9.7 \times 10^{-3}$; Figures 4A and 4B), larger than pathological grade classification (average fold difference = 1.81). However, *FLT1* did not show a prognosis prediction ability in liver cancer maybe because the expression of *FLT1* exhibited relatively little variances in the TCGA data (the standard deviation of *FLT1* is 0.91, with coefficient of variance 0.09, while the standard deviation of *COL4A3* is 2.26, with coefficient of variance 0.75). In addition to liver cancer, *DBH* and *GHR* can also significantly distinguish bad and good prognoses for many other cancers (Figures S5C and S5D). Overall, our analysis revealed a panel of immune-cell-associated SFCR biomarkers for I2C-associated angiogenesis. In particular, as circulating or secreted molecules, *DBH* and *COL4A3* may serve in diagnosis, whereas the other genes might be therapeutic targets for intervening in I2C transitions.

Immune Therapy Response Prediction

Because the SFCRs together with *FLT1* and *COL4A3* identified are predictable from I2C differentially infiltrating immune cells, we wondered whether they have implications in immune therapy where immune cell infiltration is a key determinant for therapy response (Lim and June, 2017; Medler et al., 2015), for example, whether *DBH*, inferred to be secreted by M0 cells (Figure 3D), makes a cancer more or less susceptible to T cell immune therapy. We collected two RNA sequencing (RNA-seq) melanoma datasets with PD-1 blockade and CTLA-4 blockade to validate (Table S1H). To this end, we analyzed a published RNA-seq dataset (GSE78220) (Hugo et al., 2017), which in total has 28 melanoma samples with 15 responders and 13 non-responders to PD-1 blockade immune therapy. Using normalized expression (STAR Methods) of the SFCRs (and *FLT1* and *COL4A3*) in this dataset, we modeled immune therapy response prediction using logistic regression (Table S1I). Expression of *DKK3* predicted immune therapy response the most accurately among the 21 SFCRs identified above, with 71.4% accuracy. Additionally, the accuracy for *LOXL2* and *EPOR* was 67.9% and 64.3%, respectively. Taking all SFCRs together with *FLT1* (71.4% alone) and *COL4A3* (50%

alone), prediction accuracy reached 92.9% (26 out of 28 samples) (Figure 4C). Saturation analysis demonstrated that a minimum of an eight-factor combination (*FLT1*, *LOXL2*, *GPC3*, *GHR*, *PON3*, *EPOR*, *COL4A3*, and *ALPL*) could reach the same accuracy (92.9%; Figure 4C), and 3 of the 8 genes showed significantly different expression between to immune therapy response and the non-response group (Figure 4D). The two-factor combination, *FLT1* and *LOXL2*, already reached 89.3% accuracy (Figure 4C). Similarly, we also collected the CTLA-4 blockade immune therapy RNA-seq dataset (45), containing 15 responders and 22 non-responders. 12 out of these 23 genes were detectable by RNA-seq and thus used to predict the immune therapy response (STAR Methods). The accuracy of the 12 genes reached 83.8% (31 out of 37 samples), with the maximal accuracy attained by as few as 3 genes (Figure S5E), of which the *CFI* gene had significantly lower expression in responders than the non-responders (Figure S5F). The fact that our I2C marker SFCRs predicted immune therapy response with very high accuracy indicates that the I2C transition is intimately associated with a tumor's response to immune therapy and that the I2C marker genes, in particular, *DKK3*, *FLT1*, *LOXL2*, *GPC3*, *GHR*, *PON3*, *EPOR*, *COL4A3*, and *ALPL*, have further predictive value in the personalized application of immune therapeutics. Among them, *DKK3*, *LOXL2*, *GPC3*, *PON3*, and *COL4A3* are secreted factors or in circulation and thus may serve as easy to access diagnostic markers and intervention targets. Overall, compared to random gene sets, I2C-predicting secreted factors were highly predictive of immune therapy sensitivity (Figure 4E), suggesting that the I2C tumors are ideal candidates for immune therapy.

Interestingly, among the I2C-associated SFCRs, the inflammation phase highly expressed SFCRs mostly marked sensitivity to immune therapy, whereas the cancer phase highly expressed SFCRs marked resistance to immune therapy (fold difference = 2.5; Figure 4F). This suggests that for cancers that emerge from inflammation, the best treatment phase is before the I2C transition, further highlighting the importance of developing early-diagnosis markers.

DISCUSSION

Our analysis on I2C transition quantitatively explored the pathways and their potential molecular mechanisms and identified angiogenesis as a hallmark of I2C transition and I2C-related angiogenesis genes associated with cancer prognosis. We dissected immune cell composition and change during this transition stage and found 21 SFCRs associated with 4 immune cell types (Figure 3D) that significantly predicted the expression of two key I2C interface angiogenesis genes, *FLT1* and *COL4A3*, which peak right before, or at, the I2C transition (Figure 3D). The combinations of these genes better predicted cancer prognosis than single angiogenesis genes. Furthermore, in cancers

(D) Expression level of the eight genes in (C) in immune therapy response and non-immune therapy response groups; t test was used for expression level difference analysis, * indicates $p < 0.05$ and ** indicates $p < 0.01$; standard error of mean (SEM) shown by error bars.

(E) The performance of using SFCRs together with *FLT1* and *COL4A3* to predict PD-1 blockade immune therapy response compared to that of the same number of randomly selected genes in 10,000 trials.

(F) High expression of the 21 SFCRs in inflammation and cancer reflect sensitivity and resistance to immune therapy response, respectively. SFCRs consistent with such a pattern are shown in cyan; those that are inconsistent are shown in red.

that emerge from inflammation, as few as 2 of these 23 genes could predict immune therapy response with 89.3% accuracy, and 8 could predict with 92.9% accuracy. The TCGA and immunotherapy data, although they do not match the I2C stages, are good validations of the clinical potentials for these identified I2C markers. Their effectiveness in predicting cancer severity and treatment response suggests that these markers not only signify cancer development but also cancer severity and prognosis (mainly based on recurrence and relapse status), thus implicating a common immune microenvironment component of these processes. Finally, we packaged our analysis algorithms to a downloadable application, “SwitchDetector,” for investigators to examine other biological switches.

Angiogenesis is a hallmark of cancer by addressing the needs of nutrients and oxygen of cancer to prevent tumors from necrosis or apoptosis (Nishida et al., 2006). Vascular endothelial growth factor (VEGF) was found to be an indicator of angiogenesis and is also a tumor prognostic factor (Hoebe et al., 2004) in multiple cancers, such as colorectal cancer, breast cancer, lung cancer, and head and neck squamous cell carcinoma (Nishida et al., 2006), indicating the important role of angiogenesis in tumor prognosis. Drugs such as sunitinib that inhibit angiogenesis by targeting to VEGF receptors are used as anti-angiogenic therapy to cancers. Genes associated with sunitinib response have also been associated with cancer prognosis (Dornbusch et al., 2013). On the other hand, immune cell composition changes at different tumor stages, with specific impact of different cell types on patient survival (Bindea et al., 2013). Our computational analysis examined the relationships between these two important biological processes, angiogenesis and immune composition change, and systematically identified the key extracellular secreted factors linking these two processes, thus shedding light on potentially new clinical targets for cancer therapy. Although we found fibroblasts isolated from esophageal tissue (GSE19529) also reflected the I2C microenvironment changes, more data from whole tissues will be ideal to further test the commonality of I2C signatures across different cancers.

In recent years, targeting both immune cells and the secreted factors in the cancer microenvironment also has gathered momentum in anti-cancer therapies, which are mostly based on monoclonal antibodies (Marelli et al., 2017), as epitomized by the PD-1 and PDL-1 antibodies. Our immune cell and secreted factor centered analysis of cancer development, particularly at the I2C transition, provides a new angle for such therapeutic target identification and immune therapy response prediction, with the methods developed here applicable to many types of cancers. Furthermore, the “SwitchDetector” tool and the I2C database will facilitate discovery of critical events in other cancer and disease transitions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

● METHOD DETAILS

- Datasets
- I2C transition analysis with SwitchDetector
- Module detection of I2C transition
- NP network construction
- Significant NP-network interfaces
- Function similarity analysis
- Functional enrichment analysis
- Immune cell-SFCR-angiogenesis linear model
- Survival analysis
- ROC curves and AUCs
- Mutation signature enrichment
- Immune therapy response prediction

● QUANTIFICATION AND STATISTICAL ANALYSIS

● DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and five tables and can be found with this article online at <https://doi.org/10.1016/j.celrep.2019.01.080>.

ACKNOWLEDGMENTS

This work was supported by grants from National Natural Science Foundation of China (91329302, 91749205, 31210103916, and 91519330), China Ministry of Science and Technology (2015CB964803 and 2016YFE0108700), and Chinese Academy of Sciences (XDB19020301 and XDA01010303) to J.-D.J.H.

AUTHOR CONTRIBUTIONS

J.-D.J.H. conceived and designed the study. X.C., C.X., and S.H. performed the analyses. X.X. and Y.C. helped in collecting and preprocessing data. X.C., J.-D.J.H., C.X., and J.M. wrote the manuscript. All authors contributed to preparation of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 25, 2018

Revised: September 17, 2018

Accepted: January 22, 2019

Published: February 12, 2019

REFERENCES

- Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995.
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013). Spatio-temporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795.
- Chalothorn, D., Zhang, H., Clayton, J.A., Thomas, S.A., and Faber, J.E. (2005). Catecholamines augment collateral vessel growth and angiogenesis in hindlimb ischemia. *Am. J. Physiol. Heart Circ. Physiol.* 289, H947–H959.
- Coussens, L.M., Zitvogel, L., and Palucka, A.K. (2013). Neutralizing tumor-promoting chronic inflammation: a magic bullet? *Science* 339, 286–291.
- Culig, Z. (2011). Cytokine disbalance in common human cancers. *Biochim. Biophys. Acta* 1813, 308–314.

- Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* 155, 948–962.
- Dornbusch, J., Zacharis, A., Meinhardt, M., Erdmann, K., Wolff, I., Froehner, M., Wirth, M.P., Zastrow, S., and Fuessel, S. (2013). Analyses of potential predictive markers and survival data for a response to sunitinib in patients with metastatic renal cell carcinoma. *PLoS ONE* 8, e76386.
- Dranoff, G. (2004). Cytokines in cancer pathogenesis and cancer therapy. *Nat. Rev. Cancer* 4, 11–22.
- Forbes, S.A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., et al. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 43, D805–D811.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183.
- Galamb, O., Györfi, B., Sipos, F., Spisák, S., Németh, A.M., Miheller, P., Tulassay, Z., Dinya, E., and Molnár, B. (2008). Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature. *Dis. Markers* 25, 1–16.
- Grivennikov, S.I., Greten, F.R., and Karin, M. (2010). Immunity, inflammation, and cancer. *Cell* 140, 883–899.
- Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. *Cell* 144, 646–674.
- Highfill, S.L., Cui, Y., Giles, A.J., Smith, J.P., Zhang, H., Morse, E., Kaplan, R.N., and Mackall, C.L. (2014). Disruption of CXCR2-mediated MDSC tumor trafficking enhances anti-PD1 efficacy. *Sci. Transl. Med.* 6, 237ra67.
- Hoeben, A., Landuyt, B., Highley, M.S., Wildiers, H., Van Oosterom, A.T., and De Bruijn, E.A. (2004). Vascular endothelial growth factor and angiogenesis. *Pharmacol. Rev.* 56, 549–580.
- Hong, F., Breitling, R., McEntee, C.W., Wittner, B.S., Nemhauser, J.L., and Chory, J. (2006). RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics* 22, 2825–2827.
- Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57.
- Huang, Y., Yu, X., Sun, N., Qiao, N., Cao, Y., Boyd-Kirkup, J.D., Shen, Q., and Han, J.D. (2015). Single-cell-level spatial gene expression in the embryonic neural differentiation niche. *Genome Res.* 25, 570–581.
- Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2017). Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* 168, 542.
- Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4, 249–264.
- Jacobs, J.P., Ortiz-Lopez, A., Campbell, J.J., Gerard, C.J., Mathis, D., and Benoist, C. (2010). Deficiency of CXCR2, but not other chemokine receptors, attenuates autoantibody-mediated arthritis in a murine model. *Arthritis Rheum.* 62, 1921–1932.
- Johnson, W.E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127.
- Kandoth, C., McLellan, M.D., Vandin, F., Ye, K., Niu, B., Lu, C., Xie, M., Zhang, Q., McMichael, J.F., Wyczalkowski, M.A., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339.
- Keane, M.P., Belperio, J.A., Xue, Y.Y., Burdick, M.D., and Strieter, R.M. (2004). Depletion of CXCR2 inhibits tumor growth and angiogenesis in a murine model of lung cancer. *J. Immunol.* 172, 2853–2860.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., et al. (2009). Human Protein Reference Database–2009 update. *Nucleic Acids Res.* 37, D767–D772.
- Kobayashi, H., and Lin, P.C. (2009). Angiogenesis links chronic inflammation with cancer. *Methods Mol. Biol.* 511, 185–191.
- Landskron, G., De la Fuente, M., Thuwajit, P., Thuwajit, C., and Hermoso, M.A. (2014). Chronic inflammation and cytokines in the tumor microenvironment. *J. Immunol. Res.* 2014, 149185.
- Lee, S., and Margolin, K. (2011). Cytokines in cancer immunotherapy. *Cancers (Basel)* 3, 3856–3893.
- Lim, W.A., and June, C.H. (2017). The Principles of Engineering Immune Cells to Treat Cancer. *Cell* 168, 724–740.
- Lu, H., Ouyang, W., and Huang, C. (2006). Inflammation, a key event in cancer development. *Mol. Cancer Res.* 4, 221–233.
- Lu, P., Weaver, V.M., and Werb, Z. (2012). The extracellular matrix: a dynamic niche in cancer progression. *J. Cell Biol.* 196, 395–406.
- Maeshima, Y., Colorado, P.C., Torre, A., Holthaus, K.A., Grunkemeyer, J.A., Erickson, M.B., Hopfer, H., Xiao, Y., Stillman, I.E., and Kalluri, R. (2000). Distinct antitumor properties of a type IV collagen domain derived from basement membrane. *J. Biol. Chem.* 275, 21340–21348.
- Marelli, G., Sica, A., Vannucci, L., and Allavena, P. (2017). Inflammation as target in cancer therapy. *Curr. Opin. Pharmacol.* 35, 57–65.
- Matsuo, Y., Raimondo, M., Woodward, T.A., Wallace, M.B., Gill, K.R., Tong, Z., Burdick, M.D., Yang, Z., Strieter, R.M., Hoffman, R.M., and Guha, S. (2009). CXCR2-chemokine/CXCR2 biological axis promotes angiogenesis in vitro and in vivo in pancreatic cancer. *Int. J. Cancer* 125, 1027–1037.
- Medler, T.R., Cotechini, T., and Coussens, L.M. (2015). Immune response to cancer therapy: mounting an effective antitumor response and mechanisms of resistance. *Trends Cancer* 1, 66–75.
- Medvedeva, Y.A., Lennartsson, A., Ehsani, R., Kulakovskiy, I.V., Vorontsov, I.E., Panahandeh, P., Khimulya, G., Kasukawa, T., Consortium, F., and Drablos, F. (2015). EpiFactors: a comprehensive database of human epigenetic factors and complexes. *Database (Oxford)* 2015, bav067.
- Mook, O.R., Frederiks, W.M., and Van Noorden, C.J. (2004). The role of gelatinases in colorectal cancer progression and metastasis. *Biochim. Biophys. Acta* 1705, 69–89.
- Nakano, M., Satoh, K., Fukumoto, Y., Ito, Y., Kagaya, Y., Ishii, N., Sugamura, K., and Shimokawa, H. (2007). Important role of erythropoietin receptor to promote VEGF expression and angiogenesis in peripheral ischemia in mice. *Circ. Res.* 100, 662–669.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.
- Nickoloff, B.J., Ben-Neriah, Y., and Pikarsky, E. (2005). Inflammation and cancer: is the link as simple as we think? *J. Invest. Dermatol.* 124, x–xiv.
- Nishida, N., Yano, H., Nishida, T., Kamura, T., and Kojiro, M. (2006). Angiogenesis in cancer. *Vasc. Health Risk Manag.* 2, 213–219.
- Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., et al. (2016). Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. *Dev. Cell* 36, 681–697.
- Prasad, G., and McCullough, M. (2013). Chemokines and cytokines as salivary biomarkers for the early diagnosis of oral cancer. *Int. J. Dent.* 2013, 813756.
- Qiao, N., Huang, Y., Naveed, H., Green, C.D., and Han, J.D. (2013). CoCiter: an efficient tool to infer gene function by assessing the significance of literature co-citation. *PLoS ONE* 8, e74074.
- Ren, Y., Cao, B., Law, S., Xie, Y., Lee, P.Y., Cheung, L., Chen, Y., Huang, X., Chan, H.M., Zhao, P., et al. (2005). Hepatocyte growth factor promotes cancer cell migration and angiogenic factors expression: a prognostic marker of human esophageal squamous cell carcinomas. *Clin. Cancer Res.* 11, 6190–6197.
- Ries, C. (2014). Cytokine functions of TIMP-1. *Cell. Mol. Life Sci.* 71, 659–672.
- Saadi, A., Shannon, N.B., Lao-Siriex, P., O'Donovan, M., Walker, E., Clemons, N.J., Hardwick, J.S., Zhang, C., Das, M., Save, V., et al. (2010). Stromal genes discriminate preinvasive from invasive disease, predict outcome, and highlight

- inflammatory pathways in digestive cancers. *Proc. Natl. Acad. Sci. USA* **107**, 2177–2182.
- Santarius, T., Shipley, J., Brewer, D., Stratton, M.R., and Cooper, C.S. (2010). A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* **10**, 59–64.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.
- Shaw, V.E., Lane, B., Jenkinson, C., Cox, T., Greenhalf, W., Halloran, C.M., Tang, J., Sutton, R., Neoptolemos, J.P., and Costello, E. (2014). Serum cytokine biomarker panels for discriminating pancreatic cancer from benign pancreatic disease. *Mol. Cancer* **13**, 114.
- Steele, C.W., Karim, S.A., Foth, M., Rishi, L., Leach, J.D., Porter, R.J., Nixon, C., Jeffry Evans, T.R., Carter, C.R., Nibbs, R.J., et al. (2015). CXCR2 inhibition suppresses acute and chronic pancreatic inflammation. *J. Pathol.* **237**, 85–97.
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452.
- Tassi, E., and Wellstein, A. (2006). Tumor angiogenesis: initiation and targeting - therapeutic targeting of an FGF-binding protein, an angiogenic switch molecule, and indicator of early stages of gastrointestinal adenocarcinomas. *Cancer Res. Treat.* **38**, 189–197.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578.
- Van Linthout, S., Miteva, K., and Tschöpe, C. (2014). Crosstalk between fibroblasts and inflammatory cells. *Cardiovasc. Res.* **102**, 258–269.
- Vazquez, A., Bond, E.E., Levine, A.J., and Bond, G.L. (2008). The genetics of the p53 pathway, apoptosis and cancer therapy. *Nat. Rev. Drug Discov.* **7**, 979–987.
- Wang, X., Liu, D., He, D., Suo, S., Xia, X., He, X., Han, J.J., and Zheng, P. (2017). Transcriptome analyses of rhesus monkey preimplantation embryos reveal a reduced capacity for DNA double-strand break repair in primate oocytes and early embryos. *Genome Res.* **27**, 567–579.
- Watnick, R.S. (2012). The role of the tumor microenvironment in regulating angiogenesis. *Cold Spring Harb. Perspect. Med.* **2**, a006676.
- Williams, G.H., and Stoeber, K. (2012). The cell cycle and cancer. *J. Pathol.* **226**, 352–364.
- Wu, G., Feng, X., and Stein, L. (2010). A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* **11**, R53.
- Wurmbach, E., Chen, Y.B., Khitrov, G., Zhang, W., Roayaie, S., Schwartz, M., Fiel, I., Thung, S., Mazzaferro, V., Bruix, J., et al. (2007). Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* **45**, 938–947.
- Xi, H.Q., Wu, X.S., Wei, B., and Chen, L. (2012). Aberrant expression of EphA3 in gastric carcinoma: correlation with tumor angiogenesis and survival. *J. Gastroenterol.* **47**, 785–794.
- Xia, K., Xue, H., Dong, D., Zhu, S., Wang, J., Zhang, Q., Hou, L., Chen, H., Tao, R., Huang, Z., et al. (2006). Identification of the proliferation/differentiation switch in the cellular network of multicellular organisms. *PLoS Comput. Biol.* **2**, e145.
- Xue, H., Xian, B., Dong, D., Xia, K., Zhu, S., Zhang, Z., Hou, L., Zhang, Q., Zhang, Y., and Han, J.D. (2007). A modular network model of aging. *Mol. Syst. Biol.* **3**, 147.
- Yang, C., Zeng, S., and Lv, M. (2006). Effect of shRNA inhibiting Hif1alpha gene on TIMP1 expression in RPE cells. *J. Huazhong Univ. Sci. Technol., Med. Sci.* **26**, 133–136.
- Yang, G., Rosen, D.G., Liu, G., Yang, F., Guo, X., Xiao, X., Xue, F., Mercado-Urbe, I., Huang, J., Lin, S.H., et al. (2010). CXCR2 promotes ovarian cancer growth through dysregulated cell cycle, diminished apoptosis, and enhanced angiogenesis. *Clin. Cancer Res.* **16**, 3875–3886.
- Yoo, S.Y., Lee, S.Y., and Yoo, N.C. (2009). Cytokine expression and cancer detection. *Med. Sci. Monit.* **15**, RA49–RA56.
- Zhang, H.M., Chen, H., Liu, W., Liu, H., Gong, J., Wang, H., and Guo, A.Y. (2012). AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.* **40**, D144–D149.
- Zhang, W., Liu, Y., Sun, N., Wang, D., Boyd-Kirkup, J., Dou, X., and Han, J.D. (2013). Integrating genomic, epigenomic, and transcriptomic features reveals modular signatures underlying poor prognosis in ovarian cancer. *Cell Rep.* **4**, 542–553.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
SwitchDetector	This paper	www.inflammation2cancer.org
RankProd	Hong et al., 2006	http://bioconductor.uib.no/2.7/bioc/html/RankProd.html
DAVID	Huang et al., 2009	https://david.ncifcrf.gov/
Cytoscape	Shannon et al., 2003	https://cytoscape.org/
Cociter	Qiao et al., 2013	http://www.picb.ac.cn/hanlab/cociter
CIBERSORT	Newman et al., 2015	https://cibersort.stanford.edu/
BIC-SKMeans	Zhang et al., 2013	N/A
Tophat	Trapnell et al., 2012	https://ccb.jhu.edu/software/tophat/index.shtml
Cufflinks	Trapnell et al., 2012	http://cole-trapnell-lab.github.io/cufflinks/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jing-Dong Jackie Han (jdhan@picb.ac.cn).

METHOD DETAILS

Datasets

This part includes information of raw data used for analysis, including datasets selection, inclusion and usage. All I2C datasets were downloaded from GEO database. We queried the GEO database with terms such as “((inflammation) AND cancer) AND homo sapiens [Organism].” Term “inflammation” was also replaced with “inflammatory,” and “cancer” was replaced with “tumor” and “carcinoma.” Additionally “hepatitis” and “hepatocellular carcinoma” were also used for query liver cancer, and “gastritis” and “stomach carcinoma” for gastric cancer. All queried datasets were manually checked for data quality and included in the I2C database.

For all collected 38 I2C datasets, we selected datasets with samples across three stages (Normal, Inflammation and Cancer) for analysis method development, 5 liver cancer datasets, 1 colon cancer dataset, 1 esophageal cancer dataset, 1 lung cancer dataset and 1 cervical cancer dataset meet this requirement of covering all three stages. As GSE6764 has most of samples among the 5 liver datasets, was used for detailed analysis, while all other datasets were used for validation and confirmation. The colon dataset and esophageal dataset were used to validate the functional transition from inflammation to cancer. All these datasets were used in deconvolution analysis for the immune cell compositions, to validate the deconvolution results of GSE6764. We used another independent Melanoma RNA-seq dataset (GSE78220) containing 28 pretreatment melanoma biopsies including 15 immune therapy response samples and 13 non-response samples to validate the immune therapy response prediction.

I2C transition analysis with SwitchDetector

SwitchDetector contains 4 steps to complete the analysis, and can be downloaded from www.inflammation2cancer.org. It takes the expression profiles of samples from different states as input. First, these gene expression profiles were used in deconvolution analysis to generate the immune cells compositions in each sample with CIBERSORT ([Newman et al., 2015](#)). Second, differentially expressed genes (DEGs) across different states are detected, then these genes were clustered using BIC-SKmeans ([Zhang et al., 2013](#)). Third, Protein-protein interaction network were used as template (a default network is provided in the SwitchDetector software) to construct an NP-network, and generate module interfaces with functional enrichment annotations. Finally, immune cell compositions and gene expressions were combined to infer the transition network mediated by immune cells.

Module detection of I2C transition

Raw data of microarray expression profiles of 5 liver cancer datasets, 1 colon cancer dataset, 1 esophageal cancer dataset, 1 lung cancer dataset and 1 cervical cancer dataset with samples of normal, inflammation and cancer stages were downloaded from Gene Expression Omnibus (GEO) database (Table S1D). RMA package ([Irizarry et al., 2003](#)) was used to perform quantile normalization,

background correction and gene expression quantification for the expression profiles. Then pairwise comparison between samples of inflammation versus normal, cancer versus normal and cancer versus inflammation were performed to identify differentially expressed genes (DEGs) using RankProd package with false discovery rate (FDR) less than 0.05 and log2-transformed fold changes larger or equal to 2 (Hong et al., 2006).

NP network construction

To generate the negative-positive (NP) network, we first integrated a comprehensive human protein-protein interaction (PPI) network as reference, which was based on data from Human Protein Reference Database (HPRD) (Keshava Prasad et al., 2009), STRING (Szklarczyk et al., 2015) and Wu's parsed protein interaction network (Wu et al., 2010). PPIs from STRING were kept only when the confidence score was larger than 900. Then we extracted sub network consisting of all the I2C-transition DEGs from the integrated PPI network, and calculated the Pearson correlation coefficient (PCC) of expression for each edge-connected gene pairs using the I2C-transition DEGs' fold changes. Only edges with significance $p < 0.05$ (corresponding $|PCC| > 0.22$) were remained as the final NP network (Huang et al., 2015; Xia et al., 2006) (Figure 1B).

Significant NP-network interfaces

To elucidate the transition mechanism of the progression from inflammation to cancer stages, we focused the linkage of genes between the two stages' transcriptional modules. We defined gene pairs connected within the NP network and belonging to two different stages' modules as the interface genes of the two modules, and determined whether the interface gene linkages between them is significantly higher than randomly expected. The significances of interfaces between clusters were calculated using Fisher's exact test. P values were calculated following the formula:

$$p = \frac{\binom{b}{a} \binom{N-b}{c-a}}{\binom{N}{b}}$$

, where the number of edges between two gene modules is a , and the number of edges between the one of the two modules and all other module is b , and the number of edges between another module and all other modules is c , and there are in total N edges in the NP network, then the p value of interaction between the two modules was defined.

Function similarity analysis

With liver cancer, colon cancer and esophageal cancer clusters, we developed a principal component analysis (PCA) based method to quantify the similarity of function enrichment of the clusters across cancers. Each cluster has a list of enriched functions (GO terms) with corresponding fold enrichment and p value. All significant functions of the three cancers were concatenated together as a reference function vector. Then the function vector for each cluster is filled by the cluster's fold enrichment for a function when $p < 0.05$ and 0 when $p > 0.05$. We calculated pairwise Pearson Correlation Coefficient (PCC) of gene clusters between the three datasets (liver, colon and esophageal) using their enriched GO terms. GO terms of all clusters were first combined into a matrix with minus log2-transformed p value (DAVID hypergeometric test) and filled with zero when a term not significantly enriched in a cluster. Clusters having the highest PCC between datasets were considered to be the corresponding clusters (Table S1J) for comparison visualization using the PCA method (Figure S3A).

Functional enrichment analysis

The functional enrichment analysis of annotated terms from Gene Oncology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were performed with online tool DAVID (Huang et al., 2009) using human genome genes as background. Fisher's exact test was performed for the enrichment analysis of gene clusters against cancer hallmark gene sets, cancer associated transcription factors (TFs) and epigenetic factor gene sets. For the cancer associated genes, we obtained four genes sets for the functional enrichment analysis, as CENSUS, CENSUS_AMP, COSMIC and SMG. Gene set of CENSUS is human cancer genes and that of CENSUS_AMP is genes amplified or overexpressed in human cancers (Futreal et al., 2004; Santarius et al., 2010). COSMIC gene set contains genes mutated in at least 5 cancer samples (Forbes et al., 2015). SMG gene set contains genes significantly mutated among 12 TCGA cancers (Kandoth et al., 2013). Combination of the four gene sets generated the "allRef" set (Figure 2B). Combination of the COSMIC and SMG gene sets constituted the independent set (indeSet) excluding genes from CENSUS. Top 500 confident oncogenes and tumor suppressor genes reported in (Davoli et al., 2013) were obtained as the oncogene and tumor suppressor (tumorSup) sets respectively. Human TFs and epigenetic factors were downloaded from AnimalTFDB (Zhang et al., 2012) database and EpiFactors database (Medvedeva et al., 2015) respectively.

All accessible GO terms were used for functional enrichment analysis by DAVID. As interface 3-4 in liver cancer, colon cancer and esophageal cancer are also enriched for "angiogenesis" or "blood vessel development," to assess whether these overlapping functions can appear by chance, we randomly selected the same number of genes as interface 3-4 in liver cancer (46 genes), colon cancer (140 genes) and esophageal cancer (13 genes) for 10,000 times. For the liver 46 genes, we found only 1.6% of the 10,000 random

gene sets showed significant enrichment in “angiogenesis” or “blood vessel development” (hypergeometric test $p < 0.05$), and none out of the 10,000 times had a p value less than that of the 46 genes; for the colon 140 genes, 3.6% and none of the 10,000 random gene sets showed significant enrichment in “angiogenesis” or “blood vessel development,” or with a p value less than that of the 140 genes; for the esophageal 13 genes, 1.4% and 0.2% of the 10,000 random gene sets did so.

Immune cell-SFCR-angiogenesis linear model

We generated an SFCR gene set by collecting genes with GO terms containing “cytokine” and “extracellular space,” and angiogenesis gene set by collecting genes with GO terms containing “angiogenesis” and genes with co-citation count at least 2. We modeled the expression of these genes using the contents of 9 immune cells showing significant difference between inflammation-to-cancer transition following the formula:

$$\text{Gene Expression} = a_1C_1 + a_2C_2 + \dots + a_9C_9 + \varepsilon$$

Where C_1, C_2, \dots, C_9 are contents of the 9 immune cells. The linear model is used to judge if each of the 9 immune cells can significantly affect the expression level of an SFCR gene or angiogenesis gene. Using MLR, we modeled using the liver-data-generated cell compositions of the nine immune cells to predict the liver-data gene expression of each of the SFCRs, and which were also then used to predict the expression of each angiogenesis genes in the liver data, with the regression weight representing the cell-to-gene importance, a_1, a_2, \dots, a_9 . This step is to select SFCRs and angiogenesis genes that can be significantly contributed by the contents of immune cells. We assume these weights are contributed by SFCRs secreted from different immune cells, and these weights contribute to angiogenesis genes’ expression. Thus, these weights should be correlated to the expression of SFCRs and angiogenesis genes across different immune cells. So we integrated a set of public gene expression profiles in the 9 immune cells that significantly changed during I2C (Table S1D) through batch effect correction by combat (Johnson et al., 2007), and then used the weight vector of a gene as a query pattern to test whether the gene also have this pattern of expression across the 9 immune cell types. For example, assume we have an SFCR A_i and an angiogenesis gene B_j . First, the MLR model gives the immune cell types importance weights: $wA_i = (a_{i1}, a_{i2}, \dots, a_{i9})$, $wB_j = (b_{j1}, b_{j2}, \dots, b_{j9})$ and corresponding p values of the two genes: $pA_i = (p_{i1}, p_{i2}, \dots, p_{i9})$, $pB_j = (q_{j1}, q_{j2}, \dots, q_{j9})$ in the liver data. At the same time, we have generated the expression profile of these two genes in the nine immune cell types: $\text{expr}A_i = (e_{i1}, e_{i2}, \dots, e_{i9})$, $\text{expr}B_j = (e_{j1}, e_{j2}, \dots, e_{j9})$. We defined the SFCR A_i contributing to the angiogenesis gene B_j only if 1) wA_i is significantly correlated with $\text{expr}A_i$; 2) wB_j is significantly correlated with $\text{expr}B_j$; 3) for the nine p value pairs (p_{ik}, q_{jk}) , $k = 1, 2, \dots, 9$, there exists at least one pair (p_{ik}, q_{jk}) satisfies $p_{ik} < 0.05$ and $q_{jk} < 0.05$ to ensure at least one cell type is significantly associated with the expression of both genes. The same procedure is then applied the all SFCR-angiogenesis gene pairs to construct the immune cell-SFCRs-angiogenesis network.

Survival analysis

Expression profiles and clinical data of multiple cancers were downloaded from TCGA. We divided samples into three groups based on the expression values of given genes as high expressed (top 30%), middle expressed (middle 40%) and low expressed (bottom 30%) groups. Cox regression analysis was used to test the significance of survival difference between different groups. Log-rank test p value followed by Benjamini-Hochberg correction for multiple testing are calculated to determine significance. For pathological classification of liver cancers, samples were divided into two groups, stage I group and stage II and III group. To test if mutation can affect prognosis, samples were divided into two groups, mutated and not mutated, for survival analysis.

ROC curves and AUCs

Receiver operating characteristic (ROC) curves were performed following leave-one-out (LOO) procedures. Expression values of genes were divided into two groups, high expressed (top 50%) and low expressed (bottom 50%) respectively. Then a binary classification model for predicting the gene expression to belong high or low expressed group was constructed through multiple linear regression model.

Mutation signature enrichment

RNA-seq data of cancer patients and corresponding normal samples as well as mutation annotations of cancer patients were downloaded from TCGA. Cancer samples with somatic mutations of secreted cytokine coding genes were used to compare with normal samples for differentially expressed gene detection, which is defined as somatic-mutation-induced expression signatures of the SFCRs. To examine the relationships between the SFCRs and angiogenesis process, we used genes of GO term “GO:0001525 angiogenesis” to perform fold enrichment analysis with the up- or downregulated genes of the somatic-mutation-induced expression signatures of SFCRs.

Immune therapy response prediction

Gene expression data of 28 pretreatment melanoma biopsies including 15 immune therapy response samples and 13 non-response samples were downloaded from GEO database (GSE78220). After log-transformation and z-score normalization, logistic regression with L2 regularization under LOO schedule using Python sklearn package was used to perform the immune therapy response

prediction for each gene independently. Forward selection stepwise method was used to perform the saturation analysis of gene combinations to detect the optimal combinations with saturated prediction accuracy.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of the statistical tests used are present in the figures, figure legends, reiterated in the text. They are also summarized below:

- 1) RankProd package were used to identify DEGs with 100 permutations, and the false discovery control the proportion of false prediction genes not exceeding 0.05.
- 2) Pearson Correlation Coefficient (PCC) and corresponding p values were calculated with Python package 'scipy'.
- 3) GO / KEGG enrichment were performed with DAVID with Bonferroni false discovery rate at 0.05. No multiple test correction in cluster similarity analysis with GO enrichments across different cancers.
- 4) Co-citation difference, gene expression difference, cancer fold enrichment difference, and immune cell contents difference were tested with t test, Bonferroni correction were applied for multiple test and significant level was set at 0.05.
- 5) Interactions between clusters were tested with fisher-exact test, with significant level at 0.05.
- 6) Log-rank test was applied to survival analysis, with Bonferroni false discovery rate set at 0.05.
- 7) Paired Wilcoxon signed rank test was applied to the proportion difference test between number of genes with significant survival prediction abilities, with significant level at 0.05.
- 8) Permutation test were applied for co-citation analysis and immune therapy predictions analysis, with significant level at 0.05.
- 9) F-test was used to generate linear regression p values, with significant level at 0.05.

DATA AND SOFTWARE AVAILABILITY

Our software SwitchDetector and I2C database are publicly available at www.inflammation2cancer.org.