

# UC Irvine

## UC Irvine Previously Published Works

### Title

Single-cell RNA sequencing reveals gene expression signatures of breast cancer-associated endothelial cells

### Permalink

<https://escholarship.org/uc/item/9gg9g3r8>

### Journal

Oncotarget, 9(13)

### ISSN

1949-2553

### Authors

Sun, Zhengda  
Wang, Chih-Yang  
Lawson, Devon A  
[et al.](#)

### Publication Date

2018-02-16

### DOI

10.18632/oncotarget.23760

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

# Single-cell RNA sequencing reveals gene expression signatures of breast cancer-associated endothelial cells

Zhengda Sun<sup>1,\*</sup>, Chih-Yang Wang<sup>2,5,\*</sup>, Devon A. Lawson<sup>3</sup>, Serena Kwek<sup>4</sup>, Hugo Gonzalez Velozo<sup>2</sup>, Mark Owyong<sup>2</sup>, Ming-Derg Lai<sup>5</sup>, Lawrence Fong<sup>4</sup>, Mark Wilson<sup>1</sup>, Hua Su<sup>6</sup>, Zena Werb<sup>2</sup> and Daniel L. Cooke<sup>1</sup>

<sup>1</sup>Division of Neurointerventional Radiology, Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA 94143, USA

<sup>2</sup>Department of Anatomy, University of California, San Francisco, CA 94143, USA

<sup>3</sup>Department of Physiology and Biophysics, University of California, Irvine, CA 92697, USA

<sup>4</sup>Division of Hematology and Oncology, Department of Medicine, University of California, San Francisco, CA 94143, USA

<sup>5</sup>Department of Biochemistry and Molecular Biology, Institute of Basic Medical Sciences, College of Medicine, National Cheng Kung University, Tainan 70101, Taiwan

<sup>6</sup>Center for Cerebrovascular Research, Department of Anesthesia and Perioperative Care, University of California, San Francisco, CA 94143, USA

\* Co-first author

**Correspondence to:** Daniel L. Cooke, **email:** Daniel.cooke@ucsf.edu  
Zena Werb, **email:** zena.werb@ucsf.edu

**Keywords:** breast cancer; endothelial cell; single-cell RNA sequencing; extracellular matrix

**Received:** November 08, 2017

**Accepted:** December 22, 2017

**Published:** December 29, 2017

**Copyright:** Sun et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## ABSTRACT

**Tumor endothelial cells (TEC) play an indispensable role in tumor growth and metastasis although much of the detailed mechanism still remains elusive. In this study we characterized and compared the global gene expression profiles of TECs and control ECs isolated from human breast cancerous tissues and reduction mammoplasty tissues respectively by single cell RNA sequencing (scRNA-seq). Based on the qualified scRNA-seq libraries that we made, we found that 1302 genes were differentially expressed between these two EC phenotypes. Both principal component analysis (PCA) and heat map-based hierarchical clustering separated the cancerous versus control ECs as two distinctive clusters, and MetaCore disease biomarker analysis indicated that these differentially expressed genes are highly correlated with breast neoplasm diseases. Gene Set Enrichment Analysis software (GSEA) enriched these genes to extracellular matrix (ECM) signal pathways and highlighted 127 ECM-associated genes. External validation verified some of these ECM-associated genes are not only generally overexpressed in various cancer tissues but also specifically overexpressed in colorectal cancer ECs and lymphoma ECs. In conclusion, our data demonstrated that ECM-associated genes play pivotal roles in breast cancer EC biology and some of them could serve as potential TEC biomarkers for various cancers.**

## INTRODUCTION

Tumor endothelial cells (TEC) or tumor-derived EC, including both blood vessel origin and lymphatic origin, play an indispensable role in tumor growth and metastasis [1, 2]. The tumor environmental changes,

such as hypoxia and chronic growth factor stimulation, could induce a series of endothelial dysfunctions including irregular diameters, fragility, leakiness and abnormal blood flow [3–6]. A series of *in vitro* studies indicated that the direct and indirect communications between TEC and tumor cells create and maintain a

specific EC trait for the tumor niche and allows for the reciprocal growth factor exchange between endothelial and malignant cells that may also stimulate angiogenesis and neovascularization [7–12]. Although all these studies suggest the promising application of anti-angiogenic therapy for different kinds of tumors, the limited effects of this therapy in clinical trials suggest the functional and phenotypical heterogeneity of TEC from their normal counterpart [13–19]. For example, transforming growth factor beta (TGF $\beta$ ) induces a spectrum of EC phenotypes with different functions that could underlie the plasticity and heterogeneity of the tumor vasculature [20]. All of these underscore the necessity of characterizing the heterogeneous gene expression profile of TEC.

A handful of studies have addressed the gene expression patterns of TEC compared with different control ECs by using two popular high-throughput gene expression technologies, microarray and serial analysis of gene expression (SAGE) [21–28]. Although these valuable studies have shed much light on characterizing gene expression patterns and finding potential biomarkers of TEC, both DNA microarray and SAGE need voluminous mRNA extracted from millions of ECs that is often difficult to attain given limitations of available surgical tissue. Furthermore, growing evidence illustrates that seemingly homogeneous cell populations can show considerable heterogeneous expression patterns because of multiple factors inside and outside the cells [29, 30]. The DNA microarray results from expression analyses of bulk tissue or large cell populations cannot reflect cell-to-cell variability and thus discover the specific cellular subtypes. A more complete picture of individual TEC functional condition under specific cancerous environments requires an assay that can analyze the whole gene expression profile or global transcriptome at the single-cell level [31–36].

The recent development of single-cell RNA sequencing (scRNA-seq), which combines single cell isolation techniques with RNA-seq, facilitates detecting the global transcriptome of thousands of isolated cells on single cell level by exhaustively quantifying the studied transcripts [37]. During the past few years, several groups have developed single-cell cDNA library development techniques [38, 39], including the “Smart-seq2” technology by Sandberg group that can generate quantitative and reproducible data from both single cells and small amounts of purified RNA [40]. Compared with other protocols, Smart-seq2 has improved reverse transcription, template switching and pre-amplification to increase both yield and length of cDNA libraries generated from individual cells [41].

In this study, we address the feasibility of using the Smart-seq2 scRNA-seq technique to profile the transcriptome of single ECs isolated from human breast cancer tissues and build up a bioinformatics analysis flowchart for acquired libraries based on a series of state-of-the-art computational biology toolsets.

## RESULTS

### scRNA-seq libraries making from breast tissue isolated ECs

The flowchart of scRNA-seq library development and analysis is shown in Figure 1. To ensure that the isolated ECs were viable for library making, only propidium iodide (PI)-negative single cells were gated for further EC markers sorting (Figure 2A). The final percentage for CD31 and CD34 double-positive ECs was less than 1% (0.2% in Figure 2A). cDNA pre-amplification PCR could only give obvious cDNA peaks on the Bioanalyzer by  $\geq 22$  cycles that is higher than the recommended 18 cycles previously published [40] (Figure 2B). This may also indicate the difficulty of making single-cell cDNA libraries from patient samples that have been stored for an extended period. The Bioanalyzer figure of pooled cDNA libraries after adaptor PCR amplification showed the correct size (450 bps) and reasonable concentration ( $\sim 10$ nM) that suggested the acceptable quality of the cDNA libraries for further sequencing (Figure 2C).

### Library quality control and screening

280 viable ECs were isolated from patient tissues, including 88 cells from two breast cancer samples and 192 cells from four reduction mammoplasty samples, and their transcriptomes were sequenced. Quality control analysis of scRNA-seq library indicated that only 146 (52%) libraries had more than one million reads including 74 cancerous EC libraries (84%) and 72 normal EC libraries (38%). This different library making efficiency between cancerous and control ECs (84% versus 38%) suggests that it might be harder to isolate high quality viable ECs from normal breast tissue than cancerous breast tissue, partly because of the lower EC component in control tissue. The percentages of reads aligned to genome are  $59 \pm 9.3\%$  and  $52 \pm 12.2\%$  for cancerous and control ECs, respectively. 6,865 genes were detected by the standard of RPKM $>1$  from these 146 EC cDNA libraries and only these genes in the libraries with more than one million reads were selected for further computational biology analyses.

### Principal component analysis separates the cancerous and normal ECs into two clusters

To address the differences of transcriptome profile between the cancerous ECs and control ECs, unsupervised principal component analysis (PCA) was first used to study the potential clusters of these ECs based on the detected genes. The PCA screening plot visually indicated that the first three principal components (PC1-3) could well explain the whole variance of the database with PC1 itself explaining 29.8% of total variance (Figure 3A).

2D PCA plot also showed that PC1 and 2 separated the cancerous versus control ECs as two distinctive clusters with eight cancerous ECs overlapping into the control EC cluster (Figure 3B). This PCA result indicated that these two phenotypes of ECs could be separated by their transcriptome profile by unsupervised clustering. The top 100 genes that had the highest sums of score PC1-3 were also used to produce the heat map-based unsupervised

hierarchical clustering analysis. The developed heat map and dendrogram also separated these two EC phenotypes into two distinctive clusters with only 5 overlapped cells (Figure 4). Most of the ECs originating from different patients clustered together in both cancerous and control EC clusters and developed subgroups of ECs. This may indicate the subtle gene expression differences exist between patients that might belong to different breast

**Alignment & generation of counts**

CLC Genomics Workbench

**Quality Control**

Library size & percentage of aligned reads

**Normalization**

Sequencing depth by RPKM

**Identification of differential expression**

SINGuLAR package in R console

**Build biological networks**

GeneGo MetaCore

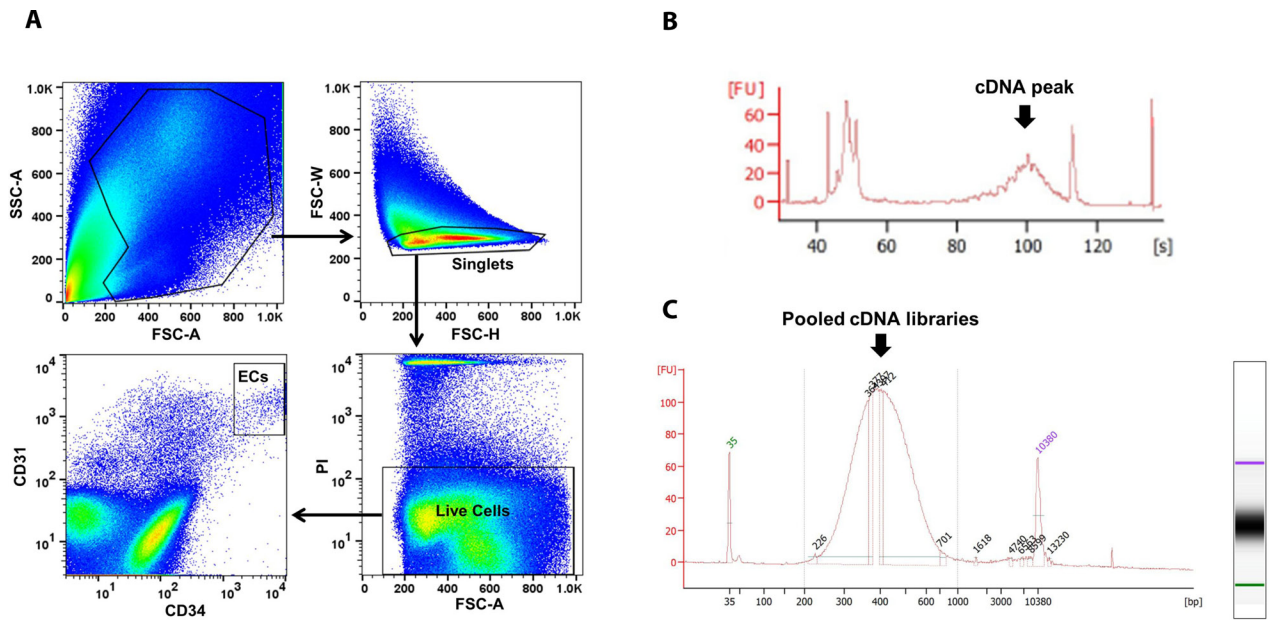
**Enriching gene sets**

Gene Set Enrichment Analysis software

**External validation**

TCGA breast cancer database  
&  
Oncomine cDNA microarray database  
&  
NCBI GEO datasets

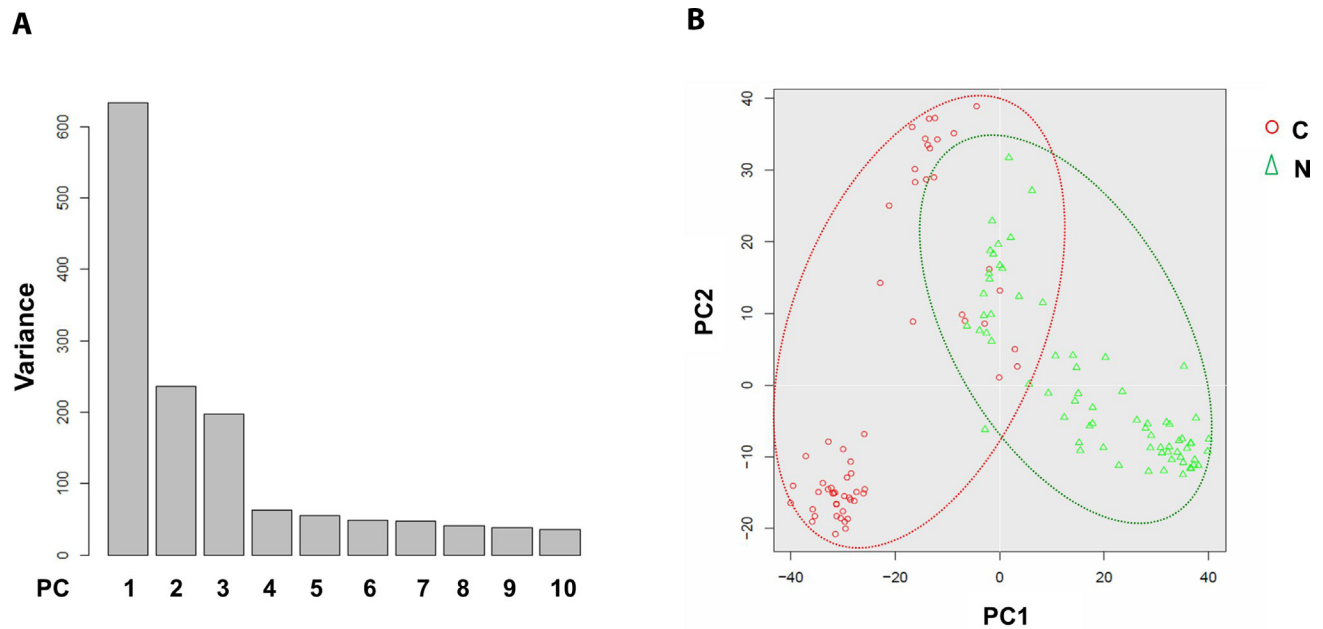
**Figure 1: Analytical strategies of scRNA-seq library database.** A flow chart of the steps in analysis of scRNA-seq data from single ECs is shown. The analysis steps are listed on the left. The computational biology toolsets used for the steps are listed in the boxes. RPKM, reads per kilobase of transcript per million mapped reads; TCGA, The Cancer Genome Atlas.



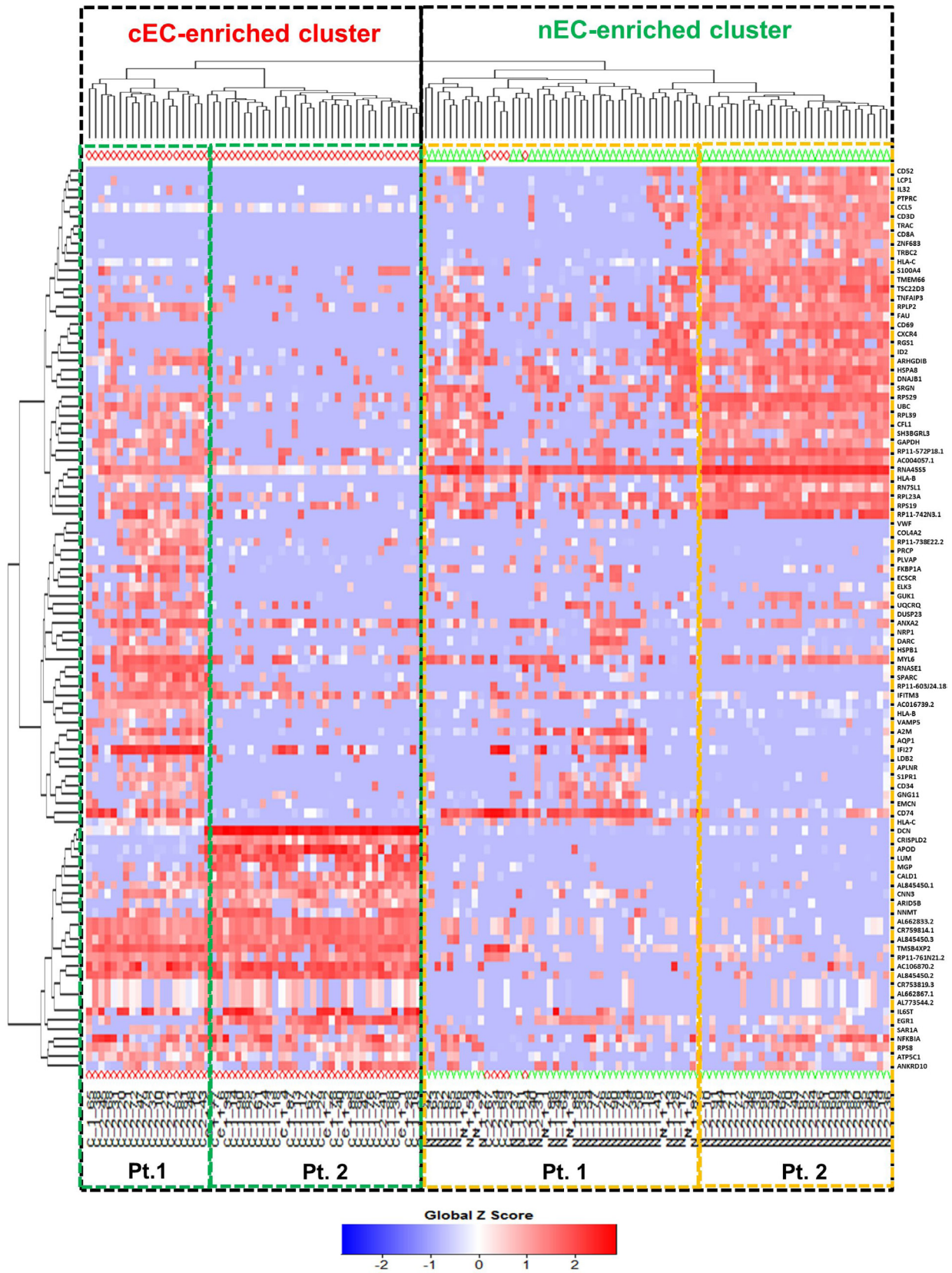
Region table for sample 10 : sample 10

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]
200	1,000	2,099.2	95	421	24.0	3,022.84	11,677.0

**Figure 2: EC sorting strategy on FACS and single-cell cDNA libraries quality control by Bioanalyzer.** ECs were sorted by CD31 and CD34 markers on viable single cells identified by PI (A). Representative Bioanalyzer electropherograms show an obvious cDNA peak at around 1 kb after cDNA pre-amplification PCR of one single EC (B), and the correct size and reasonable concentration of pooled single EC cDNA libraries (C). PI, propidium iodide.



**Figure 3: Principal component analysis of cancerous versus control EC libraries.** PCA screen plot indicates that PC1-3 could well explain the whole variance of the database (A). 2D PCA plot shows two distinct clusters along the PC1 axis that correspond to the cancerous ECs (red circle, C) and control ECs (green triangle, N) with 8 cancerous ECs overlapping into the control EC cluster (B).



**Figure 4: Heat map-based hierarchical clustering separates cancerous versus control ECs.** Cancerous ECs (cEC, red diamond) and control ECs (nEC, green triangle) are separated by heat map-based unsupervised hierarchical clustering produced from their gene expression profile of the top 100 genes with highest score PC1-3. The two clusters are highlighted by two black dot line boxes. Only 5 cancerous ECs clustered together with control ECs. Under each cluster, the cells are sub-grouped by their patient of origin.

pathological types of breast cancer or benign breast hyperplasia. ANOVA showed that 82 of these 100 genes had significantly different expression levels between cancerous and control ECs ( $p < 0.05$ )

### Disease biomarker analysis verifies breast cancer-correlated gene expression profile

SINGuLAR software identified 1,302 genes out of the total 6,865 genes were differentially expressed between cancerous and control ECs by the standard of  $p < 0.05$  and expression fold change  $>4$ . MetaCore disease biomarker analysis indicated that, among the first 20 disease annotations, three of the top five were breast neoplasm chemokines ( $p$  value =  $2.04 \times 10^{-4}$  and FDR =  $6.52 \times 10^{-3}$ ), breast neoplasm estrogen receptor 1 (ESR1,  $p$  value =  $2.04 \times 10^{-4}$  and FDR =  $6.52 \times 10^{-3}$ ) and breast neoplasm G-protein coupled receptor (GPCR) pathway regulation ( $p$  value =  $1.32 \times 10^{-3}$  and FDR =  $2.25 \times 10^{-2}$ ) (Figure 5). This suggested that these differentially expressed genes identified by this scRNA-seq database are highly correlated with the breast neoplasm diseases and could represent the characteristic gene expression profile of breast cancer ECs to some extent. In addition, MetaCore pathway analysis indicated that these differentially expressed genes were also highly correlated with dozens of signal pathways like cell adhesion, cytoskeleton remodeling and immune response (Figure 6).

### Extracellular matrix related pathways are enriched by GSEA

GSEA results showed that the differentially expressed gene set was enriched for some functional gene networks that are clearly associated with ECs, such as extracellular matrix (ECM) metabolism, vascular smooth muscle contraction and collagen formation, as well as some novel signal pathways such as complement and coagulation, drug metabolism, cancer pathways and axon guidance (Figure 7). What was most notable, three classic Gene Set Enrichment Analysis (GSEA) databases, Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome and Gene Ontology (GO), all indicated that the differentially expressed gene set was enriched for the ECM-related pathways (Figure 8). Specifically, 30 genes were enriched by KEGG database for ECM receptor interaction pathway, 20 genes were enriched by Reactome database for ECM organization pathway, and 116 genes were enriched by Gene Ontology database for ECM-related pathway. Altogether, these three databases enriched 127 genes that are related to ECM. This suggested the pivotal roles of ECM metabolism in the EC phenotype of breast cancer. Moreover, when Metacore thermometer figure was used to show the expression levels of GSEA enriched genes in one gene network of ECM remodeling, the thermometer reading levels indicated that many ECM associated genes in this network were differentially

#	Networks	0	0.5	1	1.5	2	2.5	3	$-\log(p\text{Value})$	$p\text{Value} \uparrow$	FDR
1	Breast neoplasm Chemokines								2.039e-4	6.517e-3	
2	Breast neoplasm ESR1								2.039e-4	6.517e-3	
3	Colorectal neoplasms Induction of apoptosis								2.300e-4	6.517e-3	
4	Ovarian Neoplasms (core network 2)								3.705e-4	7.874e-3	
5	Breast neoplasm GPCR pathway regulation								1.324e-3	2.251e-2	
6	Thyroid Neoplasms (core network)								1.605e-3	2.274e-2	
7	Colonic Neoplasms Cell cycle								2.449e-3	2.779e-2	
8	Ovarian Neoplasms (core network 1)								2.616e-3	2.779e-2	
9	Lung Neoplasms Signal transduction								4.841e-3	4.572e-2	
10	Amvotrophic Lateral Sclerosis (core network)								6.852e-3	5.824e-2	
11	Breast neoplasm p53								1.010e-2	7.807e-2	
12	Breast neoplasm Metalloproteases								1.284e-2	9.097e-2	
13	Carcinoma, Non-Small-Cell Lung								2.226e-2	1.455e-1	
14	Glaucoma (core network)								2.438e-2	1.480e-1	
15	Breast neoplasm IL6 and SP1								4.304e-2	2.280e-1	
16	Colorectal Neoplasms Regulation of progression through cell cycle								4.551e-2	2.280e-1	
17	Schizophrenia								4.561e-2	2.280e-1	
18	Prostatic Neoplasms Cell proliferation								5.177e-2	2.339e-1	
19	Breast neoplasm CREB1								5.229e-2	2.339e-1	
20	Breast neoplasm Inflammation								5.736e-2	2.438e-1	

**Figure 5: MetaCore disease biomarker analysis indicates the differentially expressed genes are correlated with breast cancer.** Genetic co-morbidity analysis indicated three of the top five disease annotations are breast neoplasm chemokines, breast neoplasm ESR1 and breast neoplasm GPCR pathway regulation. Orange bars represent the significance of difference by negative logical  $p$  value.  $p$  values and FDR values are listed on the right.

expressed and there are more genes had higher expression levels in cancerous ECs than in control ECs (Figure 9).

### Gene external validation by TCGA breast cancer database

To validate the differentially expressed gene set found by our database, we extracted the expression pattern of the GSEA-enriched ECM-associated genes in the TCGA breast cancer patient database through Oncomine. The heat map and hierarchical clustering indicated that 46 genes clustered the patients in the database into visual subgroups (Figure 10). The gene expression pattern also showed that 43 out of these 46 validated genes were upregulated and only three, collagen type VIII alpha 1 Chain (COL8A1), integrin subunit alpha 1 (ITGA1) and microfibril associated protein 5 (MFAP5), were downregulated in these breast cancer ECs compared with the control ECs (Figure 11).

### ECM associated genes are upregulated in a variety of cancers

To address if these 46 ECM-associated genes that were validated by TCGA breast cancer database are also important in other types of cancer, we extracted transcript expression data of these genes for 20 common cancer types from the Oncomine cDNA microarray database. 12 genes showed overall overexpression in tumors vs. normal tissues by the majority of the datasets (Figure 12), including Lysyl oxidase-like-2 (LOXL2,

57 vs. 4), Spondin-2 (SPON2, 26 vs. 3), fibronectin-1 (FN1, 86 vs. 8), bone morphogenetic protein-1 (BMP1, 20 vs. 1), Lumican (LUM, 36 vs. 14), Adipocyte enhancer-binding protein-1 (AEBP1, 41 vs. 10), Periostin (POSTN, 58 vs. 12), Collagen VI  $\alpha$ 3 (COL6A3, 49 vs. 9), Thrombospondin-2 (THBS2, 62 vs. 5), Versican (VCAN, 75 vs.15), Fibrillin-1 (FBN1, 27 vs. 13) and Asporin (ASPN, 30 vs. 9). These data suggest these 12 ECM-associated genes that may play oncogenic function in different cancer types and are not limited to breast cancer. Among these 20 cancers, colorectal cancer and lymphoma tissues showed remarkably high expression of these 12 genes (Figure 12).

We also investigated whether the mutation status of these 12 ECM related genes are favored during oncogenesis. We used meta-analysis to study the ECM gene mutation status in TCGA and other public breast cancer databases via cBioPortal platform [42]. Strikingly, there was an extremely low frequency of alterations disrupting these ECM function. In total 4063 breast tumor samples, we found only 239 point mutations were recorded at these ECM genes in these tumors (all frequencies are less than 1% except COL6A3 was 4%). These ECM mutation frequencies are thus considered very low, as high-frequency mutations are typically described as being over 20% and intermediate-frequency mutations between 2 and 20% [43]. Thus, the lack of large-scale genomic aberrations and the non-synonymous mutation frequency of 0.1–4% make it unlikely that alterations of these ECM genes impact the cancer cell phenotype (Supplementary Figure 1).

#	Maps	0	1	2	3	4	5	6	7	$-\log(p\text{Value})$	pValue $\uparrow$	FDR
1	Cell adhesion_Chemokines and adhesion										1.515e-9	1.153e-6
2	Cytoskeleton remodeling_Cytoskeleton remodeling										1.406e-8	5.351e-6
3	Cytoskeleton remodeling_TGF_WNT and cytoskeletal remodeling										3.056e-7	7.751e-5
4	Development_PIP3 signaling in cardiac myocytes										5.021e-7	9.553e-5
5	Immune response_Substance P-stimulated expression of proinflammatory cytokines via MAPKs										1.784e-6	2.715e-4
6	Development_Role of CNTF and LIF in regulation of oligodendrocyte development										3.712e-6	4.709e-4
7	Immune response_LTBR1 signaling										6.548e-6	6.430e-4
8	Androgen receptor activation and downstream signaling in Prostate cancer										6.760e-6	6.430e-4
9	Development_Gastrin in cell growth and proliferation										9.116e-6	6.937e-4
10	Glucocorticoid-induced elevation of intraocular pressure as glaucoma risk factor										9.116e-6	6.937e-4
11	Development_IGF-1 receptor signaling										1.109e-5	7.672e-4
12	Cell adhesion_ECM remodeling										1.325e-5	8.400e-4
13	Development_FGFR signaling pathway										1.575e-5	9.222e-4
14	PGE2 pathways in cancer										1.866e-5	1.014e-3
15	Regulation of lipid metabolism_Stimulation of Arachidonic acid production by ACM receptors										3.954e-5	1.880e-3
16	Immune response_IL-6 signaling pathway via JAK/STAT										3.954e-5	1.880e-3
17	Signal transduction_IP3 signaling										4.506e-5	1.896e-3
18	Cytoskeleton remodeling_Integrin outside-in signaling										4.506e-5	1.896e-3
19	Immune response_IL-15 signaling										4.766e-5	1.896e-3
20	Immune response_TSLP signalling										5.232e-5	1.896e-3

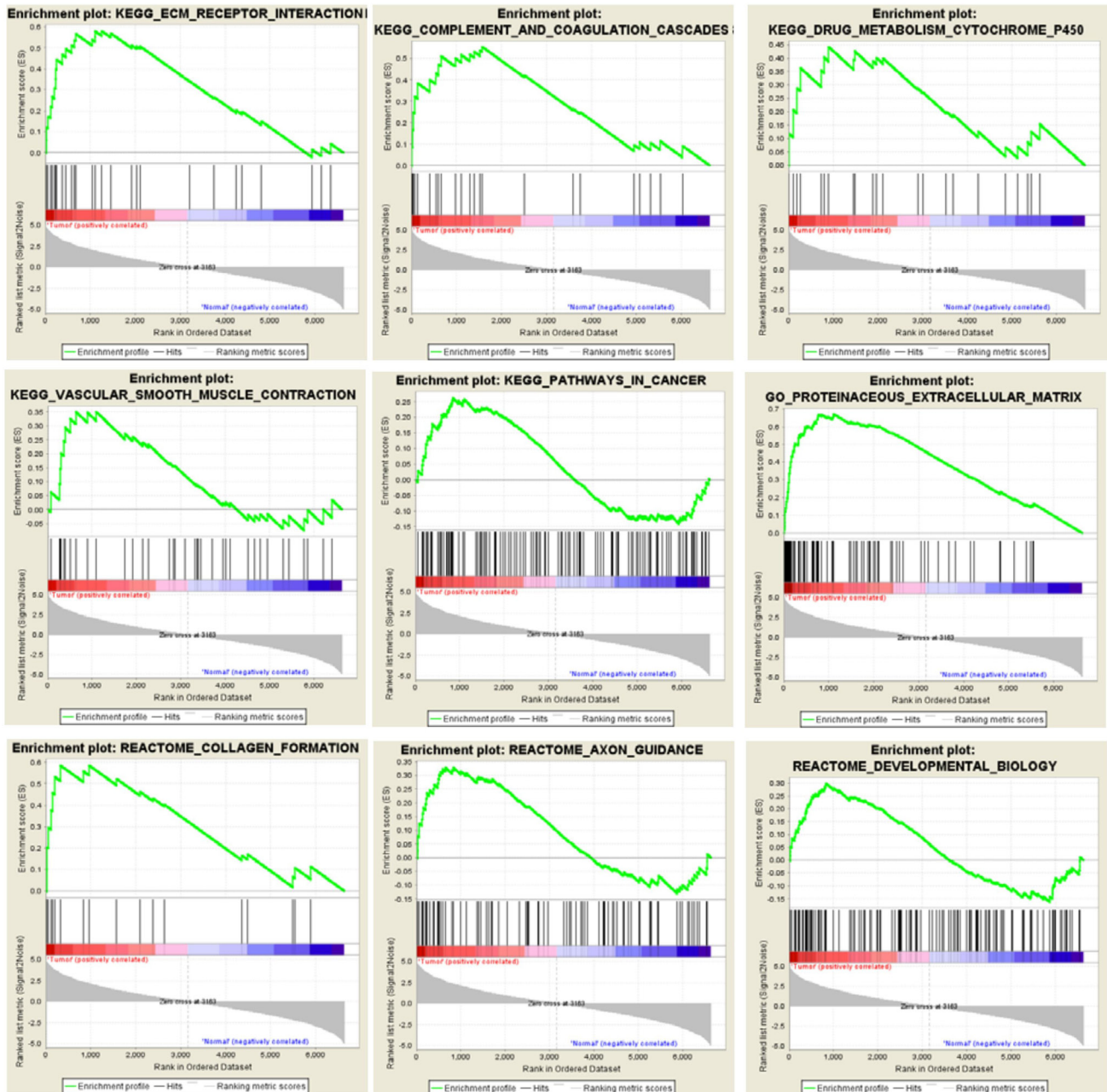
**Figure 6: MetaCore pathway analysis suggests dozens of signaling pathways.** Pathway analysis by MetaCore indicated that the differentially expressed genes are highly correlated with dozens of signal pathways; the 20 most significant pathways are listed. Orange bars represent the significance of difference by negative logical  $p$  value.  $p$  values and FDR values are listed on the right.



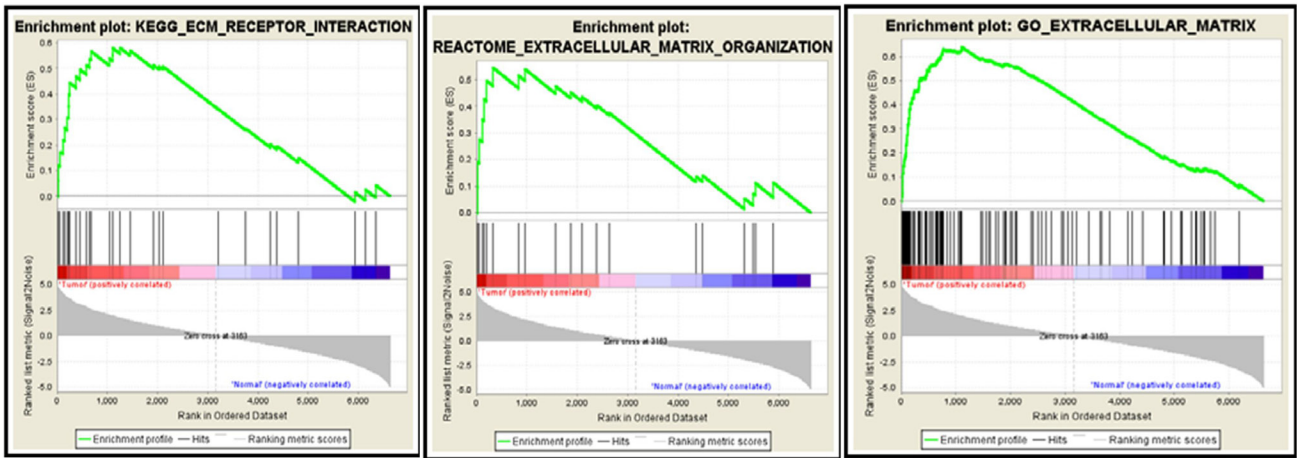
## ECM genes are highly expressed in TEC from other cancers

To address whether these validated genes are also differentially expressed in TECs from other cancer types, we go further to investigate their expression levels in colorectal cancer ECs and lymphoma ECs. Our data indicated that some of these genes specifically had high expression level in TECs from these two cancer types by analyzing available NCBI GEO database. Heat map of gene expression ranking indicated that these

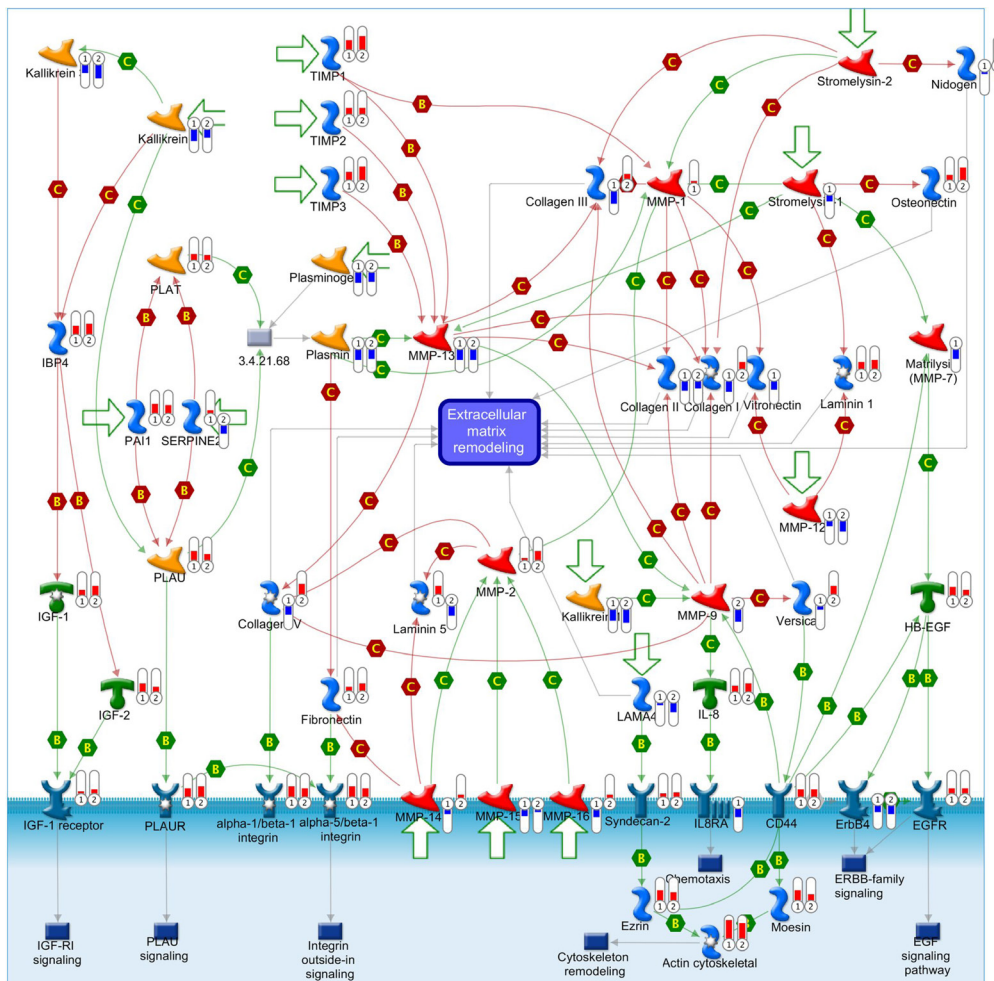
ECM associated genes were highly expressed in the colorectal cancer ECs (Figure 13A), especially with 4 genes ranked in the top 2000 genes out of over 44,000 detected microarray genes, i.e. ABI Family Member 3 Binding Protein (ABI3BP, No. 544), Coiled-Coil Domain Containing 80 (CCDC80, No.1378), SPON2 (No.1490) and THBS2 (No. 1883). Similarly in the lymphoma ECs, 7 genes ranked in the top 2000 genes, i.e., insulin like growth factor binding protein 7 (IGFBP7, No.83), laminin subunit alpha 4 (LAMA4, No.106), laminin subunit beta 1 (LAMB1, No.188), tissue inhibitor of metalloproteinase



**Figure 7: GSEA enriches differentially expressed genes into multiple functional gene networks.** GSEA results show that differentially expressed gene set in cancerous ECs are enriched for some EC-associated functional gene networks as well as some novel signaling pathways.



**Figure 8: GSEA databases enrich the differentially expressed genes to ECM-associated pathways.** 30 genes by KEGG database are enriched for ECM receptor interaction pathway (Left), 20 genes by Reactome database for ECM organization pathway (Middle) and 116 genes by Gene Ontology database for ECM-related pathway (Right). Red bar and blue bars represent upregulated and downregulated genes, respectively.



**Figure 9: ECM-associated genes are differentially expressed in cell adhesion ECM remodeling gene network.** This comprehensive picture shows gene expression changes and crosstalk of pathways in the gene network of cell adhesion ECM remodeling. The differentially expressed genes detected in our database are labeled with thermometers that indicate gene expression changes in TECs (1) and control ECs (2). Upward thermometers with red color reflect up-regulated expression and downward thermometers with blue color reflect down-regulated expression. There are more ECM-associated genes were highly expressed in TECs than control ECs. “B” and “C” mean binding and cleavage respectively. A legend explaining the symbols used by MetaCore is provided at [http://portal.genego.com/legends/legend\\_6.png](http://portal.genego.com/legends/legend_6.png).

3 (TIMP3, No.509), FN1 (No.1025), secreted protein acidic and cysteine rich (SPARC, No.1146) and POSTN (No.1890) (Figure 13B). These data suggested these ECM associated genes as possible universal TEC markers not only in breast cancer but also in colorectal cancer and lymphoma.

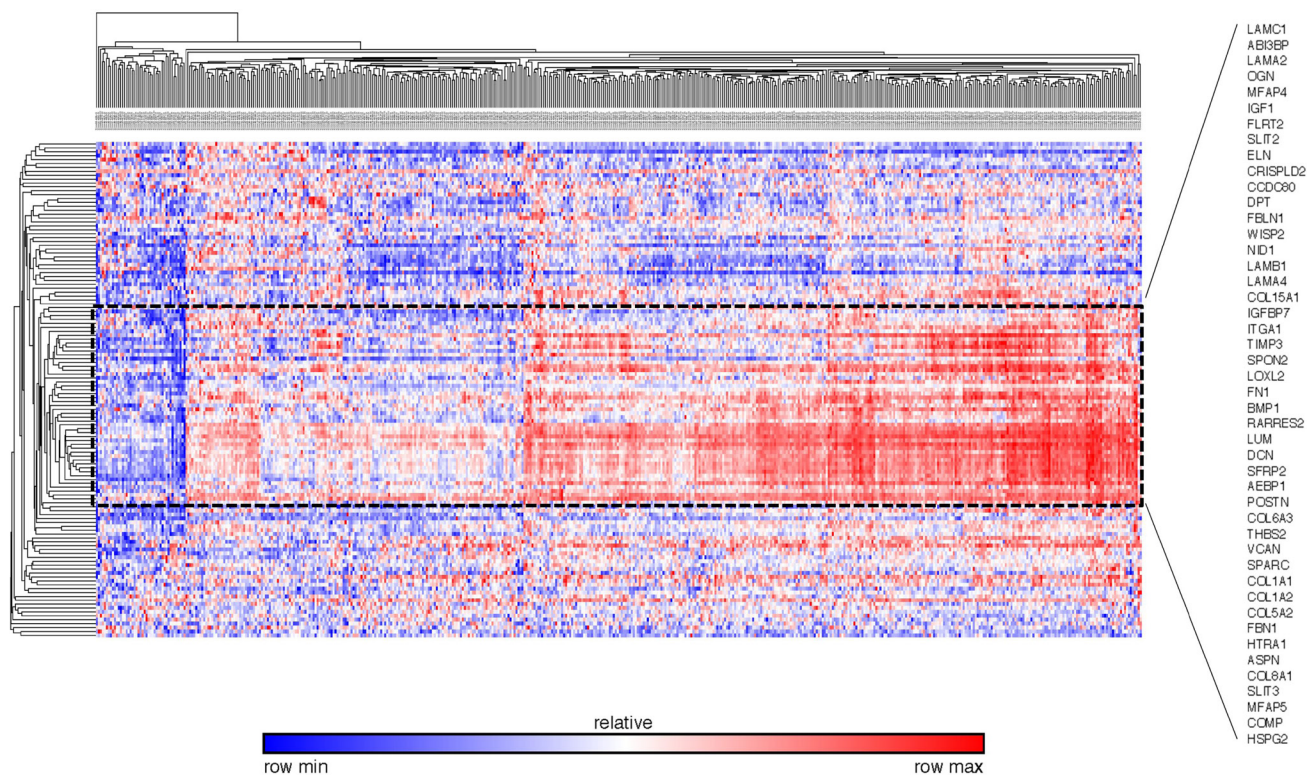
## DISCUSSION

In this study, we have shown the feasibility of developing scRNA-seq libraries from viable single ECs isolated from overnight stored clinical tissue. This is also the first report of the signature gene expression profile of breast cancer-derived ECs on a single-cell level. Our study showed that ECM-associated genes play a key role in tumor-derived ECs and identified some of these genes could serve as universal TEC biomarkers.

Although fresh tissue samples are always recommended for RNA sequencing to ensure the quality of RNA especially mRNA [40], practically, it is difficult to get such fresh tissues all the time because of the unpredictability of the surgical schedule and/or the accessibility of experiment resources. The detectable gene numbers in the transcriptome by scRNA-seq varies among studies because of different cell type, cell size, sequencing depth, etc. [44]. Furthermore, although there is still no

standard detectable gene number thus far, a common accepted standard is that 1 million reads per single cell library could detect >90% genes that has FPKM>1 [44]. In this study, using the sub-optimal breast cancerous and control reduction mammoplasty tissue, we found that such qualified single EC libraries can be made for transcriptome characterization even by relative “shallow” depth sequencing, though at a relatively lower percentage of read alignment to genome and a lower number of detected genes (about 7,000). We suggest several points are important for making cDNA libraries from sub-optimal tissues for scRNA-seq that echo the technique details emphasized in a previous publication [40, 44], such as identifying viable cells by viability markers as PI on FACS, increasing the cycle numbers to >20 for cDNA pre-amplification PCR and quality control of cDNA libraries and sequencing database. Our work indicates that sub-optimal clinical tissue, as kept overnight in storage buffer, may still be used to isolate single cells such as ECs that can be used for scRNA-seq. This may expand the usage scope for this state-of-the-art technique to study the transcriptome under different disease conditions.

In this scRNA-seq study, our results highlighted the important role of ECM in breast cancer EC biology. As a major component of the cancer local microenvironment, ECM has been reported to play important roles in



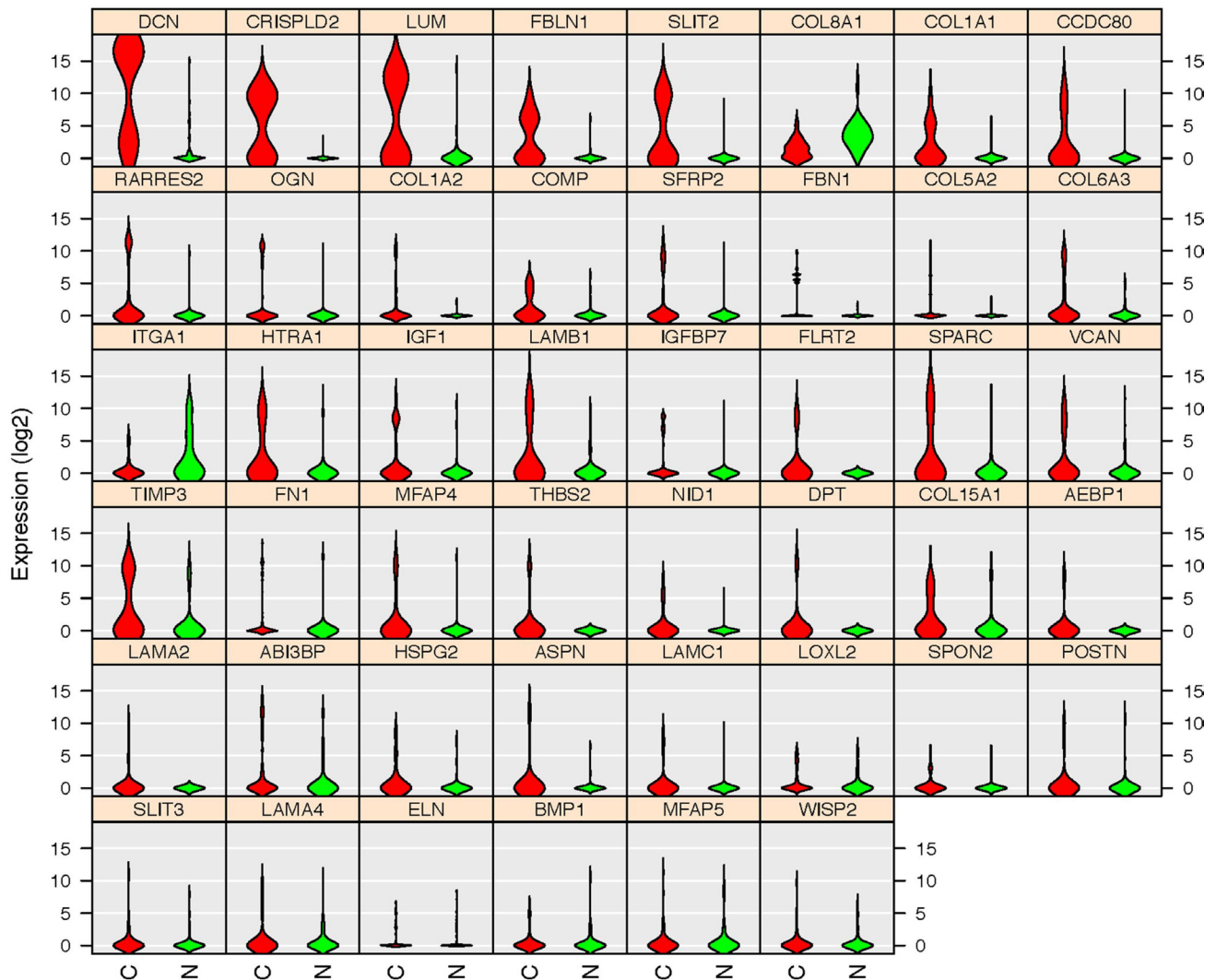
**Figure 10: External validation of ECM-associated genes by the TCGA breast cancer database** The heat map and hierarchial clustering presented the ECM-associated gene expression level of different breast patients in TCGA database. The patients and genes are listed along the X and Y axes respectively. The black dot line highlights the 46 genes that separated the patients in the database into visual subgroups.

cancer development [45–49]. The tightly controlled properties of ECM during embryonic development and tissue homeostasis are deregulated and disorganized in cancers. Besides directly acting on cancer progression by inducing cellular transformation and metastasis, ECM also facilitates development of a tumorigenic microenvironment by manipulating tumor-associated angiogenesis and inflammation. Therefore, it is expected that some ECM-associated genes may provide potential therapeutic targeting the tumor niche for treatment.

Overexpression of ECM-associated genes has also consistently been reported in previous studies on isolated TECs. In the colorectal cancer EC study of St. Croix *et al.* [21], seven out of the top 25 most differentially expressed biomarkers encode proteins involved in ECM formation or remodeling. Parker *et al.* found several ECM regulating genes overexpressed in isolated breast cancer ECs

including osteonectin, matrix metalloproteinase 9 (MMP9), and tissue inhibitor of metalloproteinase 1 (TIMP1) [23]. Madden *et al.* also found several ECM architecture regulating genes among the gene products identified as glioma endothelial markers, including heparan sulfate proteoglycan 2, several type IV collagen transcript variants, and matrix metalloproteinase 14 (MMP14) [50]. Similar to our study, Bhati *et al.* found 1176 genes were differentially expressed in human luminal-A breast tumors compared with normal vascular cells with the extracellular matrix gene ontology category was increased [24]. All these findings suggest that ECM changing is an essential component of alterations in tumor endothelium.

Although all of these 12 ECM-associated genes have been reported to be involved in the different stages of cancers like tumorigenesis, invasion and metastasis in different cancer types and were used as diagnostic



**Figure 11: Gene expression patterns of the 46 TCGA database validated genes.** Violin plots indicated that 43 genes are upregulated in cancerous ECs (Red) compared with control ECs (Green) in our database. Only three genes (COL8A1, ITGA1 and MFAP5) were downregulated. The gene name is indicated on top of each violin plot and the value on Y-axis represents the gene expression level in the binary logarithm (log<sub>2</sub>) value.

biomarker or prognostic indicator, only very limited studies have addressed their expression level in TEC specifically. LOXL2-neutralizing antibody, AB0023, was reported to inhibit angiogenesis in part by affecting VEGF signaling in ECs and simtuzumab, a humanized version of AB0023, is currently being evaluated in clinical trials for the treatment of advanced solid tumors [51]. BMP-1 overexpression in ECs was shown to be restricted to areas of tumor angiogenesis *in vivo* [52]. Recently, the EC-secreted fibronectin extra domain A (EDA) was demonstrated to promote the vasculogenesis, tumorigenesis and metastasis of colorectal carcinomas (CRCs) [53]. Thus far, no studies have reported the gene expression changes in SPON2, FN1, LUM, AEBP1, POSTN, COL6A3, THBS2, VCAN, FBN1, ASPN in TEC.

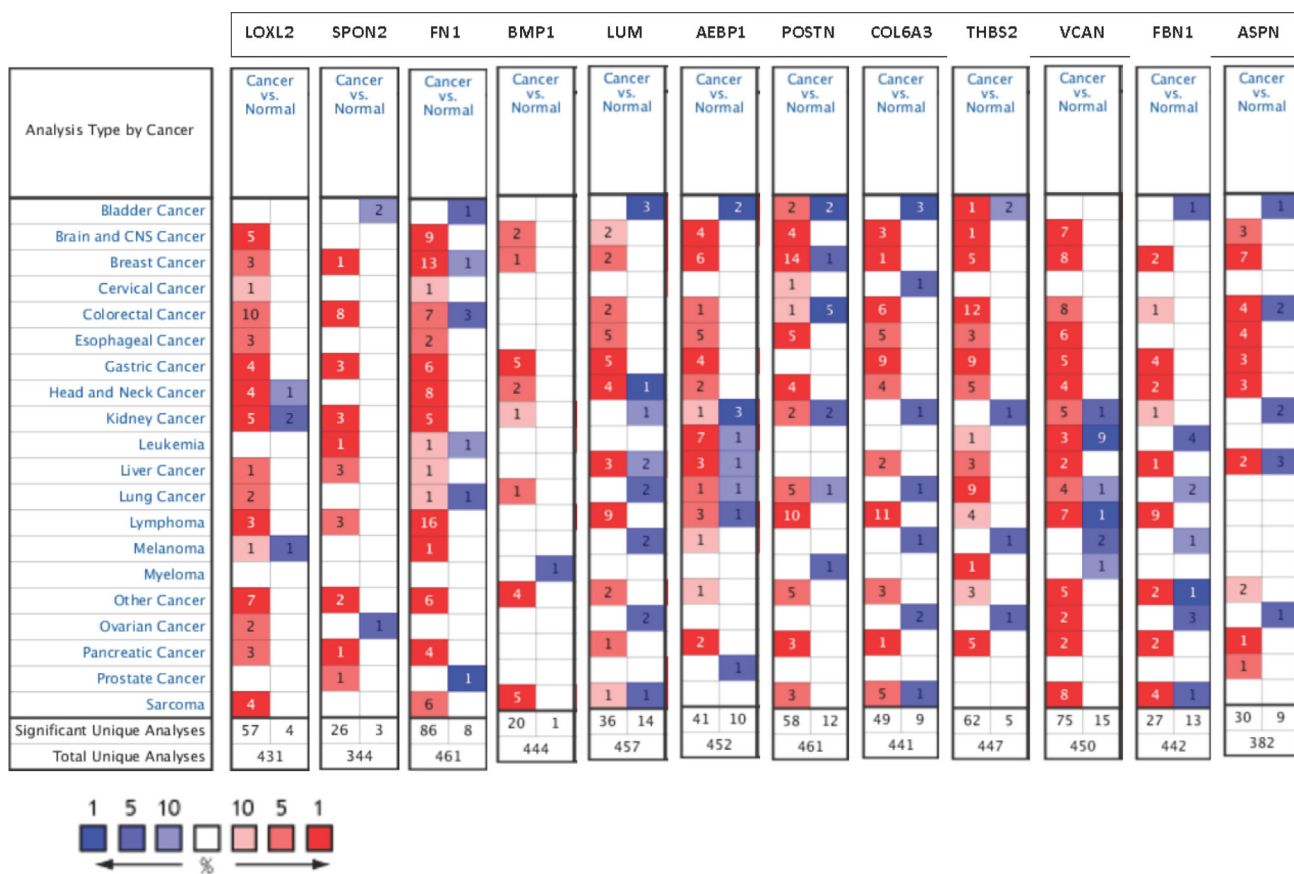
What is noteworthy is that, NCBI GEO gene expression database of TECs in colorectal cancer and lymphoma also verified the specific high expression of 7 out of these 12 ECM associated genes in TECs from colorectal cancer and lymphoma, i.e. SPON2, BMP-1, FN1, POSTN, THBS2, VCAN and AEBP1. These

data suggested these ECM associated genes as possible universal TEC markers not only in breast cancer but also in colon cancer, lymphoma and other cancers. Because of the limitation of sample size in this study, further analysis of scRNA-seq database from more TECs in different cancer tissues and according verification experiments are needed to justify these new biomarkers.

## METHODS

### Origin of tissue and single viable EC harvest

Both breast cancer and reduction mammoplasty tissues were overnight-shipped from Cooperative Human Tissue Network (CHTN), a program funded by the National Cancer Institute. Single viable ECs were isolated following the protocol of van Beijnum *et al.* with minor modification [54]. Briefly, about 0.5 g of tissue was minced using surgical blades. Mashed tissues were then enzymatically digested for 1 hour in 10 ml digestion buffer (0.1% collagenase II 9 ml, 2.5 U/ml

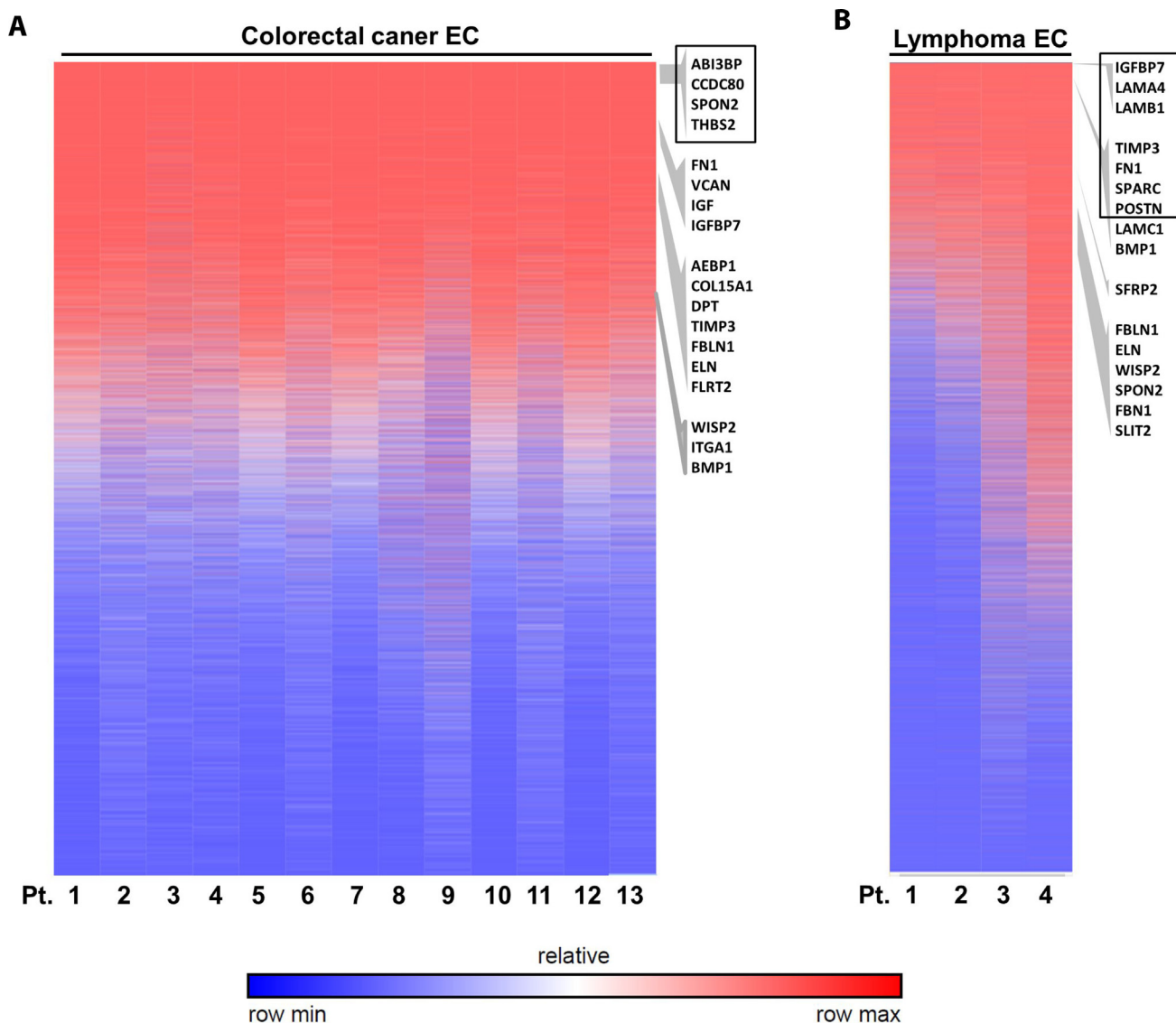


**Figure 12: Expression of 12 oncogenic ECM-associated genes in different cancer types.** The number of datasets that had significant mRNA overexpression (left column, red) and underexpression (right column, blue) of 12 specified genes are compared in 20 cancer types versus normal tissue. The color gradient correlates with decreasing gene rank percentile. The search criteria threshold was set at  $p$ -value  $< 0.0001$  with fold change  $> 2.0$  and gene rank percentile  $< 10\%$  for screening microarray datasets of cancer versus normal cases.

disperse 1ml and 0.1% DNase 75 ml) in a 37°C water bath under continuous agitation. Dissociated single cells were separated by sieving the samples through 100 mm cell strainer and 400 x g centrifugation. The re-suspended cells were then stained with two EC surface markers, CD31 Alexa647 (BD Biosciences 561654) and anti-CD34 PE-Cy7 (BioLegend 343516), and cell viability marker Propidium iodide (PI, Abcam, ab14083). After removing red blood cells by ACK lysis buffer (Life Technologies, A1049201), viable single ECs were sorted on a FACS Aria II (BD Biosciences, San Jose, CA) with 100 nm nozzle by the standard of CD31 and CD34 double positive single cells with negative PI staining. Each single EC was sorted directly into cell lysis buffer on a 96-well plate (Eppendorf, Hauppauge, NY).

### Single-cell RNA sequencing

Single-cell RNA isolation, reverse transcription and cDNA pre-amplification were performed following the Smart-Seq2 protocol of Picelli *et al* [40]. cDNA library preparation was performed following the Fluidigm C1 Protocol (100-7168 I1). Quantification of cDNA libraries was performed using Quant-iT PicoGreen dsDNA Assay Kits (Thermo Fisher Scientific, P11495) and high-sensitivity DNA chips (Agilent 5067-4626). The pooled libraries were cleaned up following the Double-sided Size Selection Protocol in KAPA Hyper Prep Kit (Kapa Biosystems, KR0961). To minimize the systemic errors, all libraries were loaded on the two lanes of one Rapid mode v2 SR1x100 flow cell and sequenced by Illumina



**Figure 13: Validation of ECM-associated genes in other cancer ECs.** Analysis of published microarray data of colorectal cancer ECs (A) and lymphoma ECs (B) indicated the high expression of some ECM associated genes. The heat map shows genes rankings ordered from highest to lowest for raw expression values across different patient samples. GSEA enriched ECM genes are listed on the right according to their ranking of expression level. Genes ranked in top 2000 are highlighted by black box.

HiSeq 2500 resulting in a calculated depth of ~1.0 million reads per cell (Illumina Rapid SR Cluster Kit v2 GD-402-4002 and Rapid SBS kit v2 FC-402-4022).

### Processing of scRNA-seq library dataset

The analytical strategies of scRNA-seq library database are shown in Figure 1. After de-multiplexing sequencing data to FASTQ files, the libraries that passed quality control assessment performed by FASTQC were further aligned to an indexed hg19 RefSeq genome using transcriptome analysis toolkits CLC Genomics Workbench (Qiagen, v6.0.4, CLCbio, Arhus, Denmark). Reads per kilobase of transcript per million mapped reads (RPKM) values of detected genes in each library were calculated for subsequent analyses. RPKM value is calculated by dividing the read counts of a certain gene in a library by the “per million” scaling factor and the length of the gene in kilobases sequentially to normalize both sequencing depth and gene length. The “per million” scaling factor is calculated by dividing the total reads number in the library by 1,000,000. For quality control filtering, only genes with a RPKM>1 in at least one sample were used for downstream computational biology analyses. Differentially expressed genes identification by one-way ANOVA, principal component analysis (PCA) and hierarchical clustering based on heat map were all performed in the Fluidigm SINGuLAR Analysis Toolset 2.0 R package after removing the outliers. A gene is considered to be differentially expressed when a probability  $P$  value <0.05 (with FDR correction) and expression fold change >4 between cancerous ECs versus control ECs. Genes are clustered on the basis of Pearson correlation and samples are clustered on the basis of a Euclidian distance matrix with complete linkage.

### Further computational biology analyses

The differentially expressed gene set was input into pathway analysis package GeneGo MetaCore (<https://portal.genego.com/>) to build biological networks and list the associated biological processes and diseases. A  $p$ -value of 0.05 was used as a cut off to determine significant enrichment of a pathway or annotated gene grouping present in the Metacore database. Gene Set Enrichment Analysis software (<https://software.broadinstitute.org/software/cprg/?q=node/14>) was also used to enrich gene sets or groups from these differentially expressed genes that share common biological function, chromosomal location, or regulation [55]. As a freely available software package together with an initial database of 1,325 biologically defined gene sets, GSEA is a computational method that determines whether a pre-defined set of genes

shows statistically significant, concordant differences between two biological states. GSEA-enriched genes were then externally validated in The Cancer Genome Atlas (TCGA) breast cancer database and the web-based Oncomine cDNA microarray database (<http://www.oncomine.org>) were also used to identify the clinical significance and expression level of these TCGA-validated genes in other cancer types. Published microarray data sets of TECs from 13 colon cancer patients (GSE89287) and 4 lymphoma patients (GSE8852) were downloaded from Gene Expression Omnibus (GEO). Microarray gene expression values were then calculated by global median normalization.

### Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding authors on reasonable request.

### Ethical approval statement

All the methods used in this study were carried out in accordance with the relevant guidelines and regulations of University of California San Francisco.

### ACKNOWLEDGMENTS AND FUNDING

Computational analyses and data mining were performed using the system provided by the Bioinformatics Core at the National Cheng Kung University, supported by the National Science Council, Taiwan. This study was supported in part by funds to Z. Werb from the National Cancer Institute (R01 CA057621 and U01 CA199315), D. Cooke from Society for Neurointerventional Surgery Foundation (125254A), H. Su from the National Institutes of Health (R01 NS027713, R01 HL122774 and R21 NS083788), Z. Sun from Department of Radiology and Biomedical Imaging, UCSF (Seed grant 14–29), M-D. Lai from Taiwan Ministry of Science and Technology (103-2325-B006-012), C.-Y.Wang from Taiwan Ministry of Science and Technology (104-2917-I-006-002) and L. Fong from Prostate Cancer Foundation. The authors would like to thank the members in Werb Lab, Fong Lab and Center for Cerebrovascular Research, UCSF for their technical support.

### Author contributions

Z.S, CYW, H.S., M.W., D.L.C., Z.W. conceived and designed the experiments, Z.S., D.A.L., S.K. and M.O. performed the experiments, CY.W. and MD.L. analyzed the data, H.G.V., MD.L., Z.W., L.F. and H.S. contributed reagents and analysis tools, Z.S., C.-Y.W., Z.W., D.L.C.

wrote the manuscript, and all authors reviewed the manuscript.

## CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

## REFERENCES

1. Dudley AC. Tumor endothelial cells. *Cold Spring Harb Perspect Med.* 2012; 2:a006536. <https://doi.org/10.1101/cshperspect.a006536>.
2. Cameron IL, Short N, Sun L, Hardman WE. Endothelial cell pseudopods and angiogenesis of breast cancer tumors. *Cancer Cell Int.* 2005; 5:17. <https://doi.org/10.1186/1475-2867-5-17>.
3. Kessenbrock K, Plaks V, Werb Z. Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. *Cell.* 2010; 141:52–67. <https://doi.org/10.1016/j.cell.2010.03.015>.
4. Ilmer M, Horst D. Pancreatic CSCs and microenvironment. *Genes Cancer.* 2015; 6:365–6. <https://doi.org/10.18632/genesandcancer.80>.
5. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nature Medicine.* 2013; 19:1423–37. <https://doi.org/10.1038/nm.3394>.
6. Van der Jeught K, Bialkowski L, Daszkiewicz L, Broos K, Goyvaerts C, Renmans D, Van Lint S, Heirman C, Thielemans K, Breckpot K. Targeting the tumor microenvironment to enhance antitumor immune responses. *Oncotarget.* 2015; 6:1359–81. <https://doi.org/10.18632/oncotarget.3204>.
7. Lopes-Bastos BM, Jiang WG, Cai J. Tumour-Endothelial Cell Communications: Important and Indispensable Mediators of Tumour Angiogenesis. *Anticancer Res.* 2016; 36:1119–26.
8. Ghiabi P, Jiang J, Pasquier J, Maleki M, Abu-Kaoud N, Halabi N, Guerrouahen BS, Raffi S, Raffi A. Breast cancer cells promote a notch-dependent mesenchymal phenotype in endothelial cells participating to a pro-tumoral niche. *J Transl Med.* 2015; 13:27. <https://doi.org/10.1186/s12967-015-0386-3>.
9. Lee E, Fertig EJ, Jin K, Sukumar S, Pandey NB, Popel AS. Breast cancer cells condition lymphatic endothelial cells within pre-metastatic niches to promote metastasis. *Nat Commun.* 2014; 5:4715. <https://doi.org/10.1038/ncomms5715>.
10. Lee E, Pandey NB, Popel AS. Lymphatic endothelial cells support tumor growth in breast cancer. *Sci Rep.* 2014; 4:5853. <https://doi.org/10.1038/srep05853>.
11. Buchanan CF, Szot CS, Wilson TD, Akman S, Metheny-Barlow LJ, Robertson JL, Freeman JW, Rylander MN. Cross-talk between endothelial and breast cancer cells regulates reciprocal expression of angiogenic factors *in vitro*. *J Cell Biochem.* 2012; 113:1142–51. <https://doi.org/10.1002/jcb.23447>.
12. Mierke CT. Cancer cells regulate biomechanical properties of human microvascular endothelial cells. *J Biol Chem.* 2011; 286:40025–37. <https://doi.org/10.1074/jbc.M111.256172>.
13. Albini A, Tosetti F, Li VW, Noonan DM, Li WW. Cancer prevention by targeting angiogenesis. *Nature Reviews Clinical Oncology.* 2012; 9:498–509. <https://doi.org/10.1038/nrclinonc.2012.120>.
14. Folkman J, Weidner N, Hanahan D. Switch to the Angiogenic State during Tumorigenesis. *Aids Research and Human Retroviruses.* 1991; 7:215–6.
15. Bergers G, Benjamin LE. Tumorigenesis and the angiogenic switch. *Nature Reviews Cancer.* 2003; 3:401–10. <https://doi.org/10.1038/nrc1093>.
16. Daniell K, Nucera C. Effect of the micronutrient iodine in thyroid carcinoma angiogenesis. *Aging (Albany NY).* 2016; 8:3180–4. <https://doi.org/10.18632/aging.101143>.
17. Wang J, Chen JP, Guo Y, Wang BC, Chu HL. Strategies targeting angiogenesis in advanced non-small cell lung cancer. *Oncotarget.* 2017; 8:53854–72. <https://doi.org/10.18632/oncotarget.17957>.
18. Piguet AC, Medova M, Keogh A, Gluck AA, Aebersold DM, Dufour JF, Zimmer Y. Impact of MET targeting on tumor-associated angiogenesis and growth of MET mutations-driven models of liver cancer. *Genes Cancer.* 2015; 6:317–27. <https://doi.org/10.18632/genesandcancer.74>.
19. Dodd KM, Tee AR. STAT3 and mTOR: co-operating to drive HIF and angiogenesis. *Oncoscience.* 2015; 2:913–4. <https://doi.org/10.18632/oncoscience.272>.
20. Xiao L, Kim DJ, Davis CL, McCann JV, Dunleavy JM, Vanderlinden AK, Xu N, Pattenden SG, Frye SV, Xu X, Onaitis M, Monaghan-Benson E, Burrige K, et al. Tumor Endothelial Cells with Distinct Patterns of TGFbeta-Driven Endothelial-to-Mesenchymal Transition. *Cancer Res.* 2015; 75:1244–54. <https://doi.org/10.1158/0008-5472.can-14-1616>.
21. St Croix B, Rago C, Velculescu V, Traverso G, Romans KE, Montgomery E, Lal A, Riggins GJ, Lengauer C, Vogelstein B, Kinzler KW. Genes expressed in human tumor endothelium. *Science.* 2000; 289:1197–202.
22. Ghilardi C, Chiorino G, Dossi R, Nagy Z, Giavazzi R, Bani M. Identification of novel vascular markers through gene expression profiling of tumor-derived endothelium. *BMC Genomics.* 2008; 9:201. <https://doi.org/10.1186/1471-2164-9-201>.
23. Parker BS, Argani P, Cook BP, Liangfeng H, Chartrand SD, Zhang M, Saha S, Bardelli A, Jiang Y, St Martin TB, Nacht M, Teicher BA, Klinger KW, et al. Alterations in vascular gene expression in invasive breast carcinoma. *Cancer Res.* 2004; 64:7857–66. <https://doi.org/10.1158/0008-5472.can-04-1976>.



24. Bhati R, Patterson C, Livasy CA, Fan C, Ketelsen D, Hu Z, Reynolds E, Tanner C, Moore DT, Gabrielli F, Perou CM, Klauber-DeMore N. Molecular characterization of human breast tumor vascular cells. *Am J Pathol.* 2008; 172:1381–90. <https://doi.org/10.2353/ajpath.2008.070988>.
25. Pepin F, Bertos N, Laferriere J, Sadekova S, Souleimanova M, Zhao H, Finak G, Meterissian S, Hallett MT, Park M. Gene-expression profiling of microdissected breast cancer microvasculature identifies distinct tumor vascular subtypes. *Breast Cancer Res.* 2012; 14:R120. <https://doi.org/10.1186/bcr3246>.
26. Xia P, Xu XY. Prognostic significance of CD44 in human colon cancer and gastric cancer: Evidence from bioinformatic analyses. *Oncotarget.* 2016; 7:45538–46. <https://doi.org/10.18632/oncotarget.9998>.
27. Hu M, Polyak K. Serial analysis of gene expression. *Nat Protoc.* 2006; 1:1743–60. <https://doi.org/10.1038/nprot.2006.269>.
28. Fernandez-Nogueira P, Bragado P, Almendro V, Ametller E, Rios J, Choudhury S, Mancino M, Gascon P. Differential expression of neurogenes among breast cancer subtypes identifies high risk patients. *Oncotarget.* 2016; 7:5313–26. <https://doi.org/10.18632/oncotarget.6543>.
29. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* 2008; 5:621–8. <https://doi.org/10.1038/nmeth.1226>.
30. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature.* 2008; 456:470–6. <https://doi.org/10.1038/nature07509>.
31. Lawson DA, Bhakta NR, Kessenbrock K, Prummel KD, Yu Y, Takai K, Zhou A, Eyob H, Balakrishnan S, Wang CY, Yaswen P, Goga A, Werb Z. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature.* 2015; 526:131–5. <https://doi.org/10.1038/nature15260>.
32. Zhu S, Qing T, Zheng Y, Jin L, Shi L. Advances in single-cell RNA sequencing and its applications in cancer research. *Oncotarget.* 2017; 8:53763–79. <https://doi.org/10.18632/oncotarget.17893>.
33. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, Louis DN, Rozenblatt-Rosen O, Suva ML, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science.* 2014; 344:1396–401. <https://doi.org/10.1126/science.1254257>.
34. Stucky A, Sedghizadeh PP, Mahabady S, Chen XL, Zhang C, Zhang G, Zhang X, Zhong JF. Single-cell genomic analysis of head and neck squamous cell carcinoma. *Oncotarget.* 2017; 8:73208–18. <https://doi.org/10.18632/oncotarget.18021>.
35. Abdelmohsen K, Panda AC, De S, Grammatikakis I, Kim J, Ding J, Noh JH, Kim KM, Mattison JA, de Cabo R, Gorospe M. Circular RNAs in monkey muscle: age-dependent changes. *Aging (Albany NY).* 2015; 7:903–10. <https://doi.org/10.18632/aging.100834>.
36. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, Cvejic A, Teichmann SA. Power analysis of single-cell RNA-sequencing experiments. *Nature Methods.* 2017; 14:381–387. <https://doi.org/10.1038/nmeth.4220>.
37. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, Ramalingam N, Sun G, Thu M, et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol.* 2014; 32:1053–8. <https://doi.org/10.1038/nbt.2967>.
38. Wilkinson DJ. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nat Rev Genet.* 2009; 10:122–33. <https://doi.org/10.1038/nrg2509>.
39. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature.* 2008; 453:1239–43. <https://doi.org/10.1038/nature07002>.
40. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc.* 2014; 9:171–81. <https://doi.org/10.1038/nprot.2014.006>.
41. Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods.* 2013; 10:1096–8. <https://doi.org/10.1038/nmeth.2639>.
42. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012; 2:401–4. <https://doi.org/10.1158/2159-8290.cd-12-0095>.
43. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature.* 2014; 505:495–501. <https://doi.org/10.1038/nature12912>.
44. Bacher R, Kendziorski C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* 2016; 17:63. <https://doi.org/10.1186/s13059-016-0927-y>.
45. Kessenbrock K, Wang CY, Werb Z. Matrix metalloproteinases in stem cell regulation and cancer. *Matrix Biology.* 2015; 44-46: 184-90. <https://doi.org/10.1016/j.matbio.2015.01.022>.
46. Insua-Rodriguez J, Oskarsson T. The extracellular matrix in breast cancer. *Advanced Drug Delivery Reviews.* 2016; 97:41–55. <https://doi.org/10.1016/j.addr.2015.12.017>.

47. Miao C, Liang C, Zhu J, Xu A, Zhao K, Hua Y, Zhang J, Chen W, Suo C, Zhang C, Liu Y, Su S, Wang Z. Prognostic role of matrix metalloproteinases in bladder carcinoma: a systematic review and meta-analysis. *Oncotarget*. 2017; 8:32309–21. <https://doi.org/10.18632/oncotarget.15907>.
48. Mellone M, Hanley CJ, Thirdborough S, Mellows T, Garcia E, Woo J, Tod J, Frampton S, Jenei V, Moutasim KA, Kabir TD, Brennan PA, Venturi G, et al. Induction of fibroblast senescence generates a non-fibrogenic myofibroblast phenotype that differentially impacts on cancer prognosis. *Aging (Albany NY)*. 2017; 9:128–46. <https://doi.org/10.18632/aging.101127>.
49. Maziveyi M, Alahari SK. Cell matrix adhesions in cancer: The proteins that form the glue. *Oncotarget*. 2017; 8:48471–87. <https://doi.org/10.18632/oncotarget.17265>.
50. Madden SL, Cook BP, Nacht M, Weber WD, Callahan MR, Jiang Y, Dufault MR, Zhang X, Zhang W, Walter-Yohrling J, Rouleau C, Akmaev VR, Wang CJ, et al. Vascular gene expression in nonneoplastic and malignant brain. *Am J Pathol*. 2004; 165:601–8.
51. Zaffryar-Eilot S, Marshall D, Voloshin T, Bar-Zion A, Spangler R, Kessler O, Ghermazien H, Brekhman V, Suss-Toby E, Adam D, Shaked Y, Smith V, Neufeld G. Lysyl oxidase-like-2 promotes tumour angiogenesis and is a potential therapeutic target in angiogenic tumours. *Carcinogenesis*. 2013; 34:2370–9. <https://doi.org/10.1093/carcin/bgt241>.
52. Wang RN, Green J, Wang Z, Deng Y, Qiao M, Peabody M, Zhang Q, Ye J, Yan Z, Denduluri S, Idowu O, Li M, Shen C, et al. Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis*. 2014; 1:87–105. <https://doi.org/10.1016/j.gendis.2014.07.005>.
53. Ou J, Peng Y, Deng J, Miao H, Zhou J, Zha L, Zhou R, Yu L, Shi H, Liang H. Endothelial cell-derived fibronectin extra domain A promotes colorectal cancer metastasis via inducing epithelial-mesenchymal transition. *Carcinogenesis*. 2014; 35:1661–70. <https://doi.org/10.1093/carcin/bgu090>.
54. van Beijnum JR, Rousch M, Castermans K, van der Linden E, Griffioen AW. Isolation of endothelial cells from fresh tissues. *Nat Protoc*. 2008; 3:1085–91.
55. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005; 102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.