Transcriptomic Analysis of Complex Tissue Microenvironments

By

MATTHEW B. CURTIS
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Engineering

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Steven C. George, Chair

Sepideh Gholami

Anthony Passerini

Committee in Charge

2022

Table of Contents

Abstract

Single cell RNA-Sequencing is a powerful technique to analyze the transcriptomic landscape of cells. Here we present two distinct *in vitro* systems (3D *in vitro* vascular networks in fibrin hydrogels; *ex vivo* organoid model of colorectal liver metastasis) and apply single cell RNA-Sequencing in order to better understand the extent of cellular heterogeneity in these 2 systems. For the 3D *in vitro* vascular network work, we are motivated by the fact that endothelial cells line all major blood vessels and serve as integral regulators of many functions including vessel diameter, cellular trafficking, and transport of soluble mediators. Despite similar functions, the phenotype of endothelial cells is highly organ-specific, yet our understanding of the mechanisms leading to organ-level differentiation is incomplete. We generated 3D vascular networks by combining a common naïve endothelial cell with six different stromal cells derived from the lung, skin, heart, bone marrow, pancreas, and pancreatic cancer. Single cell RNA-Seq analysis of the vascular networks reveals five distinct endothelial cell populations, for which the relative proportion depends on the stromal cell population. Morphologic features of the organotypic vessel networks inversely correlate with a cluster of endothelial cells associated with protein synthesis. The organotypic stromal cells were each characterized by a unique subpopulation of cells dedicated to extracellular matrix organization and assembly. Finally, compared to cells in 2D monolayer, the endothelial cell transcriptome from the 3D *in vitro* skin and lung vascular networks shows a closer match to the *in vivo* endothelial cells from the respective organs. We conclude that stromal cells contribute to endothelial cell and vascular network organ tropism, and create an endothelial cell phenotype that more closely

resembles that present *in vivo*. For the colorectal liver metastasis work, we are motivated by the fact that colorectal cancer is a leading cause of cancer-related death in the United States. A significant proportion of colorectal cancer cases present liver metastases at some point during the course of disease, with limited treatment options for these liver metastases. In order to better understand the liver metastases and their response to treatment, it is necessary to develop a robust model of these liver metastases. We generated a patient-matched *ex vivo* organoid model of colorectal cancer liver metastases and compared this organoid model to parental tumor samples using single cell RNA-Seq. Parental tumor samples have a rich diversity of cell types, whereas the organoid sample contains only epithelial cells. Additionally, we identify 3 sub-populations of epithelial cells with distinct transcriptomic profiles that are present in different amounts between the parental and organoid samples. We conclude that there is transcriptomic drift from the *in vivo* parental sample to the *ex vivo* organoid sample, which manifest in phenotypic drift between these samples. In both *in vitro* systems outlined above, single cell RNA-Seq is a powerful technique to understand both emergent cellular heterogeneity within the *in vitro* samples and transcriptomic differences between *in vitro* and *in vivo* tissues.

List of Figures and Tables

Abbreviations in Dissertation

BSA = bovine serum albumin

cPancreas = cancer pancreas

CRC = colorectal cancer

CRLM = colorectal cancer liver metastasis

DEGs = differentially expressed genes

DPBS = Dulbecco's phosphate buffered saline

EC = endothelial cell

ECFC-EC = endothelial colony forming cell-endothelial cell

ECM = extracellular matrix

GO = gene ontology

nPancreas = normal pancreas

OSC = organotypic stromal cell

PBS = phosphate buffered saline

scRNA-Seq = single cell RNA-sequencing

TiME = tissue microenvironment

TuME = tumor microenvironment

tSNE = t-distributed stochastic neighbor embedding

UMAP = uniform manifold approximation and projection

Dedication and Acknowledgements

This dissertation is dedicated to my family, without whom I would not be who and where I am today. Dad, Mom, Josh, and Brittany, thank you for all your love and support. This dissertation is also dedicated to my wonderful fiancée Lisa. There is too much to say about how you have helped me these past years and what you mean to me. I am lucky to have you in my life.

This dissertation is also dedicated to the members of the George Lab who I have had the pleasure of knowing during my time in the lab. I would like to specifically thank my PI, Dr. Steven George, for putting his faith in me and challenging me through my time as an undergraduate student, Masters student, and finally as a PhD student in his lab. Thank you.

Additionally, I would like to acknowledge the UC Davis DNA Technologies Core, the UC Davis Bioinformatics Core, Dr. Jamal Lewis and his lab, and Dr. Alyssa Panitch and her lab for technical assistance on this work. I would also like to thank Dr. Christopher Hughes for his collaboration on the organotypic 3D *in vitro* vascular network work. I would also like to thank Dr. Sepideh Gholami for her collaboration on the *ex vivo* colorectal cancer liver metastasis work. I would also like to thank Dr. Anthony Passerini for his help and guidance in completing the PhD. I would also like to thank Dr. Bhupinder Shergill and Dr. Venktesh Shirure in the George Lab for their help on the colorectal cancer liver metastasis work. Lastly, I would like to acknowledge my friends who have helped me get through graduate school here at UC Davis.

Chapter 1 – Studying Cellular Heterogeneity in *In Vitro* Systems

Single cell RNA-sequencing (scRNA-Seq) is a powerful technique which allows for the observation of gene transcript expression at the single cell level. This technology represents a marked improvement over bulk measurements of gene expression (i.e. bulk RNA-Seq, qPCR, etc.) in that a large amount of transcriptomic information can be obtained for individual cells. This allows an investigator to identify both large and subtle transcriptomic differences within (or between) samples, thereby allowing for the identification of rare sub-populations of cells and a broader picture of the heterogeneity within cell samples.

The central dogma of molecular biology, well known to most biologists and biomedical engineers, is DNA is transcribed to RNA which is then translated into protein. This process, while simply described here, does not capture the full picture of what is occurring within a cell. In fact, many molecules and processes actively work to enhance, repress, or otherwise modulate this larger DNA→RNA→protein process. For instance, proteins (i.e. transcription factors) can bind to regions of DNA to enhance the ability of RNA polymerase to transcribe specific regions of the DNA, causing downstream increases in mRNA transcripts for specific genes [1,2]. Additionally, small interfering RNA molecules (siRNA) can interfere with mRNA (RNA interference, or RNAi) to alter the availability of mRNA for translation into protein [3,4]. Finally, some specific proteins (i.e. kinases, phosphatases, etc.) can post-translationally modify other proteins to affect their function [5]. While not an exhaustive list of the different types of regulation that exist in transcription and translation, it is clear that a complex web of

interactions governs fundamental biological processes. Therefore, in order to understand how specific cells are functioning at a granular level, it is necessary to obtain a large amount of information for these cells. Depending on the duration of the sequencing (or the "sequencing depth"), large amounts of single cell-level information can be generated and utilized to paint a larger picture of a sample. While there are several methods to understand some degree of cellular heterogeneity on the protein level (i.e. flow cytometry, immunofluorescent microscopy, etc.), it has been considerably more difficult to measure the heterogeneity of cell populations at the gene/RNA level. These gene level measurements are critical to our understanding of a collection of cells, especially in the tissue contexts with known mutations in gene networks (i.e. cancer). Since the early years of scRNA-Seq [6], a wealth of information has been learned for a variety of tissues and pathologies, leading to new insights into how specific genes may be functioning with a specific cell, how larger gene networks may be functioning within a cell, and how different cells may be functioning and interacting with their environment.

In vitro systems are useful tools to allow for the study of complex biological phenomena. From relatively simple 2D *in vitro* cell monolayer cultures, to more complex 3D *in vitro* systems incorporating multiple cell types, ECMs, tissue organizational patterns, forces, etc., to *in vivo* animal models, there exists a large spectrum of model complexity, model cost, model reproducibility, and model biological relevance. The determination of how complex a model must be to study some biological phenomena is a critical question to answer while designing experiments to test specific hypotheses. While "simpler" 2D *in vitro* models have the benefit of more often being higher throughput, affording greater spatiotemporal control to the

investigator, and generally being less expensive (in both time and money), they may often fail to capture enough biological relevance to appropriately model a specific cellular or tissue phenomenon. Therefore, there exists a strong need to study and understand how *in vivo* tissues and their respective *in vitro* models differ in order to improve the models to better understand and predict *in vivo* phenomena.

The crux of this dissertation is a transcriptomic analysis of emergent cellular heterogeneity in two different *in vitro* model systems, and an attempt to understand the extent to which an *in vitro* model system can recapitulate the *in vivo* transcriptomes of each tissue. The first system that will be looked at is a 3D *in vitro* fibrin hydrogel model of vascular network formation using endothelial cells (ECs) and stromal cells. Specifically, we will describe how organ-specific stromal cells alter the EC transcriptome, and how these transcriptomic changes may be manifested in phenotypic changes in the tissue, along with a look at how these 3D *in vitro* vascular networks compare with organ-specific *in vivo* endothelial cells. The second system that will be looked at is an *ex vivo* organoid model of colorectal cancer liver metastasis, albeit for a dataset still in progress. Specifically, we will characterize the types of cells that are retained through the organoid culture process and describe several ways in which those cells differ from those found in the parental metastatic lesions (also known as "transcriptomic drift"). Overall, we find that the *in vitro* models leveraged in this dissertation differ from their *in vivo* counterparts to different extents in terms of their transcriptomes, emphasizing the need to understand an *in vitro* model in the context of its respective *in vivo* tissue, and to develop more biologically relevant *in vitro* models.

Chapter 2 – 3D *In Vitro* Models of Human Organotypic Capillary Networks

Introduction

Endothelial cells (ECs) line the vasculature and directly impact the structure and function of all vessels, including capillaries. Interestingly, the structure and function of capillary beds is markedly organ-specific, and generally contributes to the overall function of the organ [8]. For example, the capillary endothelium in the central nervous system is characterized by a reduced ability of molecules to transit from the blood to the tissue, and vice versa (i.e., the "blood-brain barrier") [9]. In contrast, the capillary endothelium in the liver exhibits large fenestrations which allow molecules to easily pass between the tissue and the blood [10]. In addition, increasing evidence shows that the vasculature also plays a role in patterning tissues [11–13], suggesting bi-directional crosstalk between the endothelium and the surrounding parenchyma. Much work has been done to characterize organ-specific ECs, but the mechanism for how these organ-specific differences emerge is unclear [14–17].

The local tissue microenvironment (TiME), including organotypic stromal cells (OSC), can potentially contribute to organ-specific capillary network structure and function. Stromal cells are a heterogeneous population of cells that form an integral part of the TiME. Stromal cells perform diverse functions in normal physiology and pathology, including extracellular matrix (ECM) production and remodeling, and secretion of growth factors and cytokines, both of which can potentially impact EC phenotype [18,19]. Some stromal cells, such as pericytes, directly associate with blood

vessels, and can regulate capillary permeability and sprouting angiogenesis [20–24]. Additionally, stromal-derived matricellular factors regulate EC sprouting angiogenesis and vascular lumen formation *in vitro* [25,26]. Taken together, it is clear that stromal cells are critical regulators of capillary network structure and function.

In order to better understand organ-specific features of the capillary endothelium, multicellular models of organotypic vasculature have been developed. 2D *in vitro* monolayers of organotypic EC are generally simple and inexpensive to create, but typically lack the multicellular complexity (i.e., are grown and studied in the absence of stromal cells) and geometry of the TiME. *In vivo* animal models are perhaps more physiologically relevant, but are often cost- and time-intensive, do not necessarily reflect the human endothelium, and have limited spatiotemporal resolution. We and others have reported on the development of 3D human vascular networks (both perfused and unperfused) as improved models to understand dynamic events such as angiogenesis [27,28] and cellular (immune and cancer) trafficking [29–31]. While some effort has been made to validate organ-specific characteristics of the vasculature in these 3D *in vitro* models [29,32–34], there is a significant gap in our knowledge of both the similarities and differences that emerge between *in vivo* and *in vitro* EC and capillary networks.

We hypothesized that OSC impact capillary EC phenotype, and that 3D capillary networks generate EC that are more representative (relative to 2D monolayer) of *in vivo* EC. To address our hypothesis, we examined both the EC and OSC transcriptome, using single cell sequencing, in 3D *in vitro* capillary networks created with OSC representing skin, lung, pancreas, pancreatic cancer, heart, and bone marrow. Our results demonstrate that: 1) five distinct EC phenotypes can be defined based on their

transcriptome; 2) the relative proportion of each EC phenotype in the 3D networks depends on the OSC; and 3) the proportion of the EC phenotype characterized by active protein synthesis is negatively correlated with key morphological features of actively growing capillary networks including total length and number of junctions. Furthermore, each OSC demonstrated a unique population of cells as well as two populations that were shared amongst the six different OSC. Finally, compared to 2D monolayers of ECs, the 3D capillary networks created with skin and lung stromal cells more closely resemble the *in vivo* transcriptome from publicly available datasets [35,36].

Methods

2D monolayer cell culture

Endothelial cell colony-forming endothelial cells (ECFC-ECs) were isolated from umbilical cord blood, as previously described [37,38], and were chosen as they can be considered "organ-agnostic" (derived from cord blood) and less differentiated (progenitor) than tissue-resident endothelial cells. The ECFC-ECs were grown on gelatin-coated tissue culture plastic in EGM-2 (Lonza #CC-3162). The same donor of ECFC-ECs was used for all experiments. All organotypic stromal cells (OSCs) were purchased commercially and grown according to vendor recommendations on tissue culture plastic. Human bone marrow stromal cells (Marrow OSC; Lonza #2M-302) were cultured in Myelocult H5100 media (STEMCELL Technologies #05150) supplemented with 1 μM hydrocortisone in a-MEM (STEMCELL Technologies #74142 and

ThermoFisher #12571063, respectively), 2 mM L-glutamine (ThermoFisher #25030081), and 50 unit/mL Penicillin-Streptomycin (ThermoFisher #15070063). Human cardiac fibroblasts (Heart OSC; Cell Applications #306AK-05a) were grown in HCF Growth Medium (Cell Applications #316K-500). Normal human dermal fibroblasts (Skin OSC; Lonza #CC-2511) were cultured in FGM-2 (Lonza #CC-3132). Normal human lung fibroblasts (Lung OSC; Lonza #CC-2511) were cultured in FGM-2 (Lonza #CC-3132). Pancreatic cancer associated fibroblasts (cPancreas OSC; Vitro Biopharma #CAF08) were cultured in Pancreatic Stellate CAF Maintenance Media (Vitro Biopharma #PC00B5). Human pancreatic stellate cells (nPancreas OSC; ScienCell Research Laboratories #3830) were cultured in Stellate Cell Complete Medium (ScienCell Research Laboratories #5301).

All cells were sub-cultured per vendor recommendations in a 37 ^{o}C, 5% CO_2 incubator. Once cells reached confluence, they were briefly washed with sterile DPBS without divalent cations (DPBS(-)) (ThermoFisher #14190094) and then treated with 0.05% Trypsin-EDTA (ThermoFisher #25300062). Once lifted, the cell suspension was neutralized using cell-specific media. Cells were then centrifuged at 300xg for 5 min and were either (1) propagated for additional sub-culture, (2) utilized in downstream experiments, or (3) frozen down in a 10% DMSO solution for long-term liquid nitrogen storage. All cells were between passage 3 and passage 10 prior to use in experiments.

Flow cytometry

Once cell monolayers reached confluence, cells were lifted using Trypsin as described above. Cells were blocked in a 0.1% BSA buffer and labeled using conjugated antibodies (**Supp. Table 1**). Cells were analyzed using a ThermoFisher Attune (model #A24858), with an average of 20,000 events recorded per sample. Gating for all relevant markers of interest was determined using fluorescence-minus-one (FMO) controls for each individual marker of interest. Raw FCS files were processed using FlowJo™. Cells were first gated for live, single cells prior to evaluating the expression of markers of interest. Flow cytometry analysis was repeated 3 times to verify consistency in marker expression.

Vendor	Cat #	Fluorophore	Target	Vol. per 100 µL Test (µL)
BD Biosciences	560411	FITC	CD144/ VE-Cadherin	20
BioLegend	303116	APC	CD31/ PECAM1	5
BioLegend	328122	BV 421	CD90/ THY1	5
BioLegend	323606	PE	PDGFRβ/ CD140b	5
ThermoFisher	L10119	Fixable Near IR	Live/Dead	1

Table 1
Antibodies used in the flow cytometry characterization of ECFC-EC and OSC monolayers prior to 3D fibrin hydrogel culture from Figure 1.

3D in vitro vascular networks

Fibrinogen (Sigma #F8630) was solubilized in DPBS(-) at a concentration of 10 mg/mL, and syringe-filtered prior to use. Thrombin (Sigma #T4648) was solubilized in DPBS(-) at a concentration of 50 units/mL, syringe filtered, and stored at -20 ^{o}C until thawed for use in experiments. Once ECFC-EC and OSC monolayers reached confluence, the cells were exposed to Trypsin and collected as previously described. To create the 3D vascular networks, we used a well-established previously published protocol[18,39]. In brief, the OSC and ECFC-EC were counted and mixed in sterile fibrinogen at a 2:1 ratio, respectively, such that there were 2 million OSCs per mL and 1 million ECFC-ECs per mL. 150 µL of the cell-fibrinogen solution was mixed with 6 µL of thrombin and the resultant volume introduced into individual wells of a 48 well plate. As a negative control, some fibrin hydrogels contained only 1 million ECFC-ECs per mL (no OSCs present in these hydrogels). The hydrogels were allowed to polymerize for 30 minutes in a 37 ^{o}C incubator prior to adding media. Fibrin hydrogels were cultured in EGM-2, with media changes every other day.

Immunofluorescence

After 7 days of culture, fibrin hydrogels were fixed in a 10% formalin buffer. Hydrogels were then permeabilized in PBS+0.5% Tween-20 before undergoing blocking in 2% BSA in PBS+0.1% Tween-20. Primary CD31 (PECAM1) antibody was diluted in blocking buffer and applied overnight to the hydrogels at 4 ^{o}C (**Table 2**). Secondary antibody was diluted in blocking buffer and applied for 45 min. at room temperature

(**Table 2**). Hydrogels were then washed in PBS+0.1% Tween-20 and imaged. All fibrin hydrogels were imaged using an Olympus IX83 widefield fluorescence microscope.

Vendor	Cat #	Species	Target	Fluorophore	Dilution
ThermoFisher	13031982	Mouse	CD31/ PECAM1	N/A	1:100
ThermoFisher	A21427	Rabbit	Anti-Mouse	Alexa Fluor 555	1:400

Table 2
Antibodies used in the immunofluorescent characterization of vascular networks from 3D in vitro fibrin hydrogels from Figure 2.

Quantitative evaluation of vascular networks

Resultant images of the CD31-labeled vascular networks were edited in ImageJ to further reduce background signal. Resultant JPG files were then quantified using Angiotool™, as previously described [39,40]. All images were quantified using the same settings, and the resultant data were further analyzed in the R computing environment (See *Statistical Analysis* below). Two parameters were highlighted, namely total vessel length and total number of junctions (or branching points).

Isolation of cells from fibrin hydrogels

After 7 days of culture, fibrin hydrogels were digested using nattokinase (Japan Bio Science Laboratory #NSK) diluted to 50 units/mL in EGM-2. After 1 hour of exposure at 37 ^{o}C, resultant degraded hydrogels were pooled for each condition (>5 hydrogels per condition), washed several times in DPBS(-), and exposed to an additional 10 min of 0.05% Trypsin-EDTA to ensure adequate release of cells from any remnant ECM fragments. Resultant solutions underwent additional washes, 70 μm filtration, and then finally a 40 μm filtration prior to scRNA-Seq analysis.

Transcriptome alignment and initial processing

All single cell library preps and sequencing was performed by the UC Davis DNA Technologies Core and UC Davis Bioinformatics Core Facilities. Cells were analyzed by 10X Genomics 3' sequencing v3. Raw fastq files were processed using CellRanger count (10X Genomics) for genome alignment via the Linux command line. Resultant filtered output files were brought into the R computing environment and analyzed further via the Seurat pipeline and by using a series of scRNA-Seq analysis packages (**Table 3**). All data files underwent quality control filtering to exclude cells with fewer than 200 unique genes, greater than 7,500 unique genes, and/or more than 10% of total gene expression derived from mitochondrial-specific genes (**Figure 5A-E**). In addition, genes that were not detected in at least 3 cells were excluded from downstream analysis. This quality control process yielded several thousand cells per 3D *in vitro* vascular network condition (**Figure 5A**). ECs and OSCs were identified by characteristic gene expression, and then separated and re-normalized for downstream analysis. The

following values were used for the “resolution” parameter to determine the number of clusters: 1 for the overall 3D *in vitro* vascular network-derived cell data (**Figure 6C**), 0.3 for the endothelial cells alone from the dataset (**Figure 9A**), and 1 for the stromal cells alone from the dataset (**Figure 14A**). Clustering was verified by observation and evaluation of the silhouette scores for each dataset (**Figure 6A-B** for overall 3D *in vitro* vascular network-derived cell data; **Figure 10E-F** for the endothelial cells alone; **Figure 16E-F** for the stromal cells alone). Additionally, cells were analyzed for their ribosomal gene-related gene expression by analyzing the total amount of *RPS* and *RPL* gene expression per cell, and percentage of *RPS/RPL* gene expression was reported (**Figure 12B-C** for endothelial cells; **Figure 16G-H** for stromal cells).

Gene ontology analysis

Differentially expressed genes (DEGs) were identified for all EC clusters and OSC clusters (after re-normalizing and re-clustering the data independently). For a specific cluster, the list of DEGs along with the list of all genes expressed was fed into the topGO package in R. In order to identify the relevant over-enriched biological process gene ontology (GO) terms associated with the DEGs for each cluster, Fisher’s exact test using the “elim” algorithm was performed. Resultant *p*-values were then log10-transformed and GO terms were rank ordered by the log10-transformed *p*-value.

Name	Environment	Version
Ubuntu	Linux	20.04.4
CellRanger	Linux	5.0.0
GRCh38	Linux	3.0.0
base	R	4.0.3
Matrix	R	1.4-1
org.HS.eg.db	R	3.12.0
topGO	R	2.42.0
SparseM	R	1.81
GO.db	R	3.12.1
circlize	R	0.4.14
AnnotationDbi	R	1.52.0
graph	R	1.68.0
stringr	R	1.4.0
SeuratWrappers	R	0.3.0
monocle3	R	1.0.0
SingleCellExperiment	R	1.12.0
SummarizedExperiment	R	1.20.0
GenomicRanges	R	1.42.0
GenomeInfoDb	R	1.26.7
IRanges	R	2.24.1
S4Vectors	R	0.28.1
MatrixGenerics	R	1.2.1
matrixStats	R	0.61.0
NMF	R	0.24.0
cluster	R	2.1.3
rngtools	R	1.5.2
pkgmaker	R	0.32.2
registry	R	0.5-1
CellChat	R	1.1.3
Biobase	R	2.50.0
BiocGenerics	R	0.36.1
igraph	R	1.3.0
cowplot	R	1.1.1
ggplot2	R	3.3.5
patchwork	R	1.1.1
dplyr	R	1.0.8
SeuratObject	R	4.0.4
Seurat	R	4.1.0
tidyverse	R	1.3.1

Table 3
System version information for Linux and relevant packages from R computing environment.

In vivo comparisons of in vitro vascular network data

For *in vivo* dataset comparisons [41], publicly available *in vivo* whole lung dissociate data and *in vivo* skin tissue data were utilized for comparison [35,36]. *In vivo* skin ECs were identified out of the skin *in vivo* dataset by EC characteristic gene expression. Briefly, *in vivo* skin EC data underwent an anchor-based integration with the corresponding skin-specific EC 3D *in vitro* vascular network data and ECFC-EC 2D *in vitro* monolayer data via FindIntegrationAnchors and IntegrateData in Seurat. The resultant integrated object was scaled and clustered. DEGs were identified for each EC type (*in vivo*, 3D *in vitro*, 2D *in vitro* monolayer), and then a score was assigned to each EC type based on the list of *in vivo* DEGs using AddModuleScore in Seurat. The resultant score was then normalized by the mean 3D *in vivo* score and the mean 2D ECFC-EC *in vitro* monolayer score, so as to better identify any relative shift of the 3D *in vitro* vascular network-derived ECs. The same process was repeated for comparisons of the *in vivo* lung dataset to lung-specific EC 3D *in vitro* vascular network data and ECFC-EC 2D *in vitro* monolayer data.

Statistical Analysis

Statistical analysis of vascular network data was carried out in the R computing environment. One-way ANOVA was used to compare the total vessel lengths and total number of junctions between the different 3D *in vitro* vascular networks. Reported *p*-values were deemed significant if $p < 0.05$. Evaluation of the linear regression EC cluster percentage and mean total vessel length was also performed in R. Pearson's

correlation coefficients were determined as part of the analysis and reported *p*-values were deemed significant if $p < 0.05$.

Results

Endothelial cells assemble into vascular network structures within 7 days when co-cultured with organotypic stromal cells in fibrin hydrogels

Prior to co-culture in a fibrin hydrogel, various OSC monolayers and EC monolayers were examined by flow cytometry to ensure characteristic marker expression (**Figure 1A-B**). ECFC-ECs were observed to be CD31+CD90-CD144+ while all OSCs were observed to be CD31- (with monolayer-specific differences in CD90 expression), consistent with previous observations [42,43].

ECFC-ECs assemble into vascular networks after 7 days of co-culture with each of the six different OSCs in 3D fibrin hydrogels as verified by IF microscopy for CD31/PECAM1 (**Figure 2A-B**). The vascular networks demonstrated morphological differences in terms of total vessel length and total number of junctions (**Figure 2C-D**). For example, the total vessel length per unit area for the skin network was approximately twice that of bone marrow or pancreatic cancer. ECs do not form stable vascular networks in the absence of OSCs in fibrin hydrogels (data not shown).

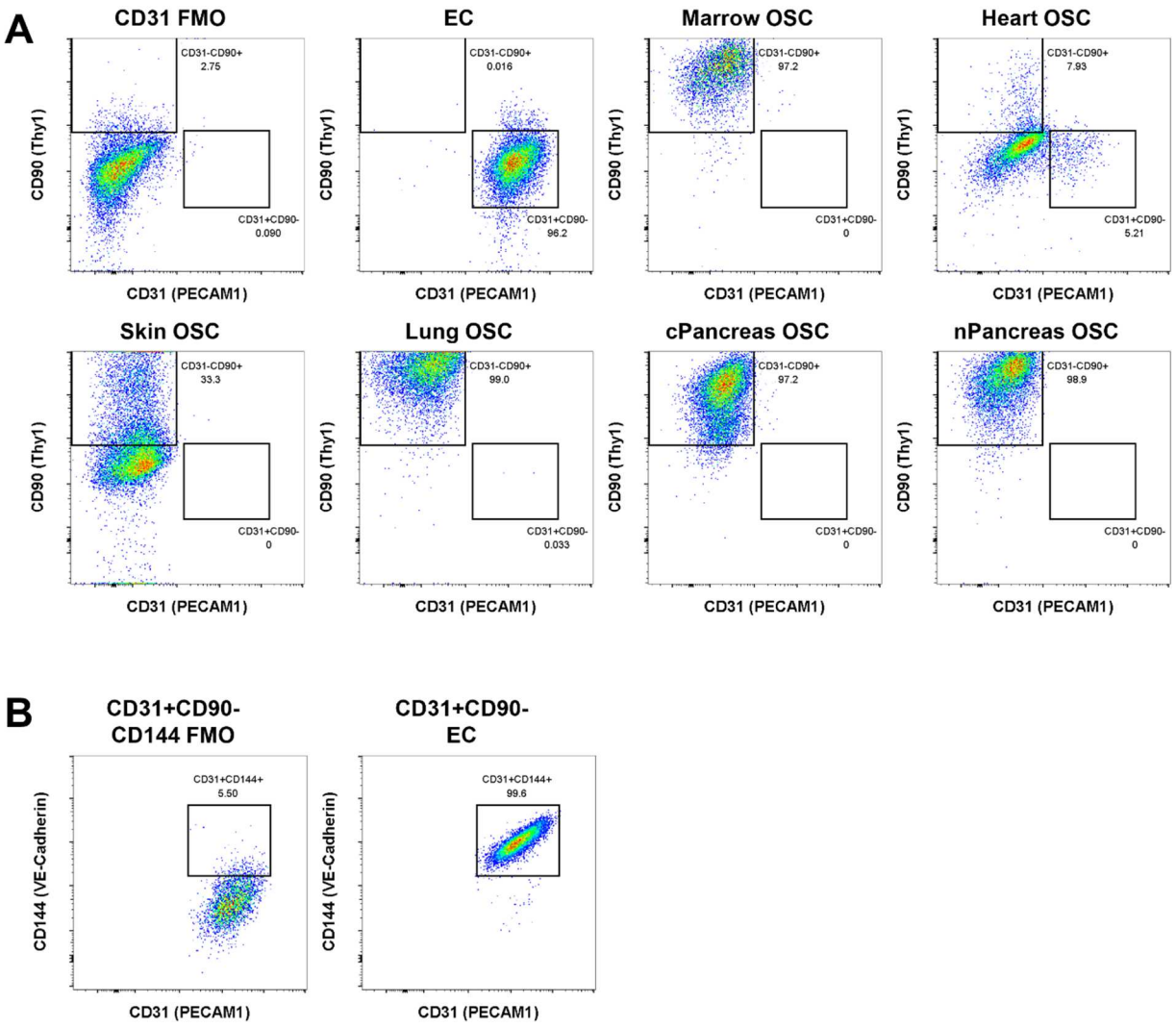

Figure 1

ECFC-EC monolayers are CD31+CD90-CD144+ by flow cytometry, while OSC monolayers have heterogeneous expression of CD90, but crucially do not express CD31. **A)** CD31 vs. CD90 expression for ECFC-EC and OSC monolayers. Average of 20,000 events recorded per cell type. Gating determined by FMO controls on ECFC-EC monolayers. **B)** CD31 vs. CD144 expression for CD31+CD90- ECFC-EC monolayers. Average of 20,000 events recorded. Gating determined by FMO controls on EC monolayers.

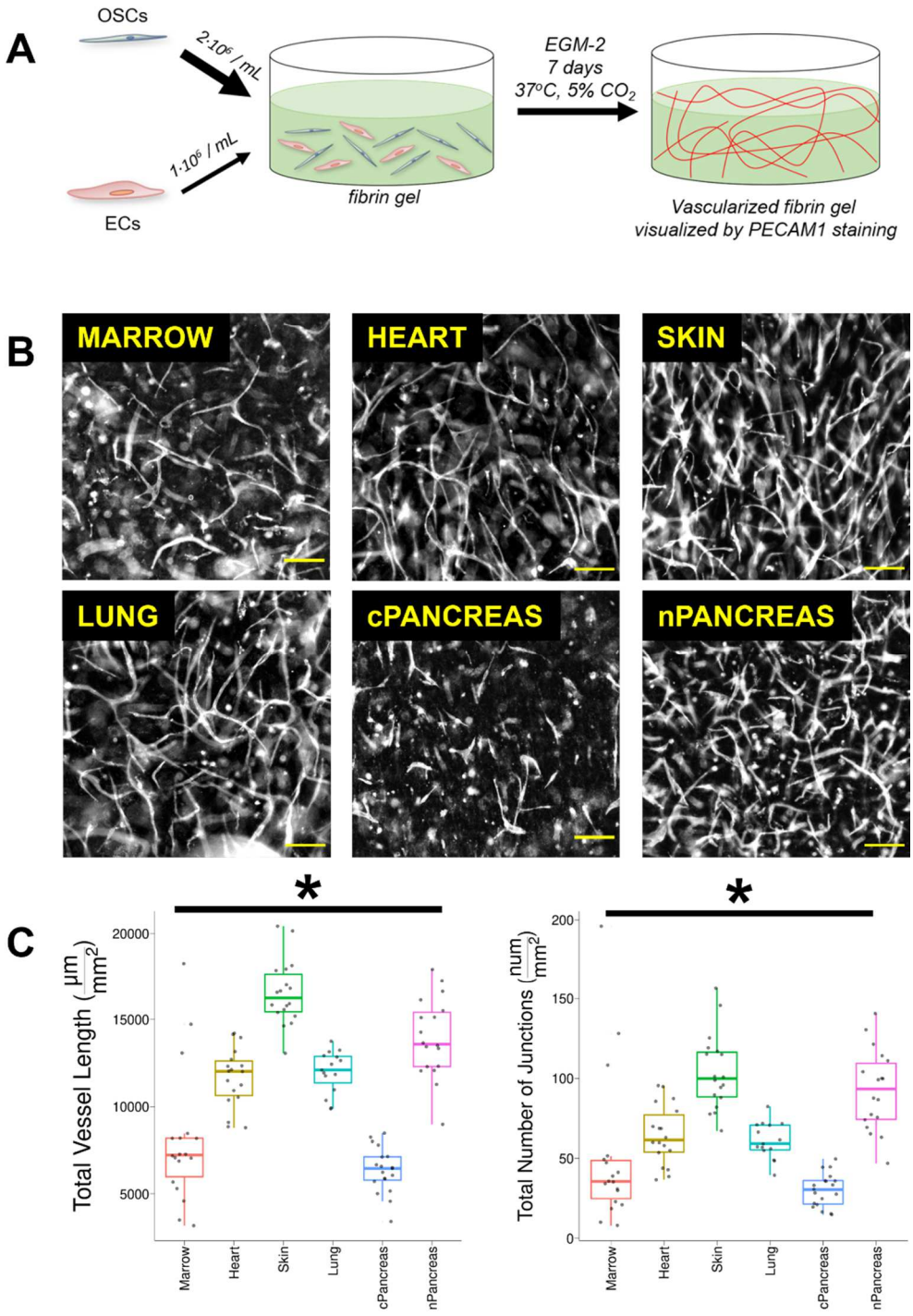

Figure 2

ECFC-ECs form into CD31+ vascular networks after 7 days of co-culture with a variety of OSCs in 3D *in vitro* fibrin hydrogels. **A)** Schematic detailing formation of 3D *in vitro* vascular networks in fibrin hydrogels. **B)** Representative images of CD31+ (PECAM1+) vascular networks formed using same parental ECFC-EC monolayer and OSCs as indicated in each image. Scale bar represents 200 μm. **C)** Quantification of CD31+ (PECAM1+) vascular networks by Angiotool. Total vessel length and total number of junctions normalized by total image area. *$p < 0.05$ by One-Way ANOVA.

ECs and OSCs can be identified by characteristic gene expression

3D *in vitro* fibrin hydrogels underwent enzymatic digestion by nattokinase followed by Trypsin (outlined in the Methods). Nattokinase is a potent fibrinolytic enzyme, which can degrade fibrin hydrogels quickly (**Figure 3**) and has previously been shown to be gentle enough on cells so as to not compromise cell viability and surface protein expression [29,44].

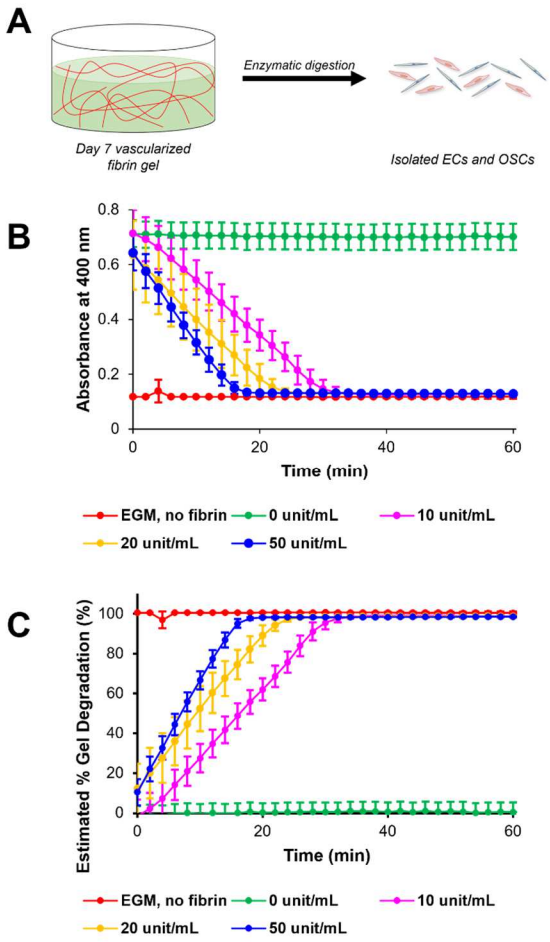

Figure 3
3D *in vitro* fibrin hydrogels can be degraded by leveraging fibrinolytic activity of nattokinase. **A)** Schematic detailing enzymatic digestion of 3D *in vitro* fibrin hydrogels. **B)** Absorbance values at 400 nm for 150 µL 3D *in vitro* blank fibrin hydrogels (no cells; fibrin-only) exposed to a variety of concentrations of nattokinase diluted in EGM-2 over the course of 2 hours. Sample size of 3 for each measurement. **C)** Estimated percent fibrin hydrogel degradation for 150 µL 3D *in vitro* blank fibrin hydrogels based on absorbance values at 400 nm.

After enzymatic digestion, the resultant samples were filtered to obtain cells for scRNA-Seq (**Figure 4A**). Resultant UMAP plots demonstrate significant transcriptomic heterogeneity both within and between 3D *in vitro* vascular network conditions (**Figure 4B**). The combined 3D *in vitro* fibrin hydrogel-derived scRNA-Seq dataset underwent quality control (detailed in the Methods; **Figure 5**) and unsupervised k-means clustering as part of the Seurat pipeline. Parameter sweeps were performed to ensure an appropriate number of clusters was achieved for downstream analysis (**Figure 6A-B**). Twenty-five distinct clusters were identified (**Figure 6C**). As previously mentioned, prior to co-culture in the fibrin hydrogels, ECFC-ECs were CD31+CD90-CD144+ and OSCs were CD31-CD90+ by flow cytometry. Therefore, we proceeded to examine known EC-specific and stromal-specific genes to distinguish the ECs and OSCs in the dataset [25,26,45,46]. Clusters 2, 16, 18, and 24 were notable as they express EC-specific genes including *CDH5*, *CLDN5*, *EGFL7*, *MCAM*, and *PECAM1* (**Figure 7A**), albeit with some heterogeneity in marker expression between the clusters. Clusters 2, 16, 18, and 24 also demonstrated heterogeneous expression of other known EC-specific genes (**Figure 8A**) [45]. These four clusters were also negative for a series of stromal-specific genes including *COL1A1*, *COL1A2*, *PDGFRA*, *PDGFRB*, and *TAGLN* (**Figure 7B**, **Figure 8B**). The remaining cell clusters (all clusters but 2, 16, 18, and 24) did not express EC-specific genes, while at the same time expressed (with some heterogeneity) stromal-specific genes (**Figure 7A-B, Figure 8A-B**). We therefore lumped clusters 2, 16, 18, and 24 into a set representing ECs, while all other clusters were considered OSCs (**Figure 7C**). This analysis resulted in a range of ECs (75-688)

and OSCs (1065-5220) from each of the six vascular networks (**Figure 7D** and **7E**, respectively).

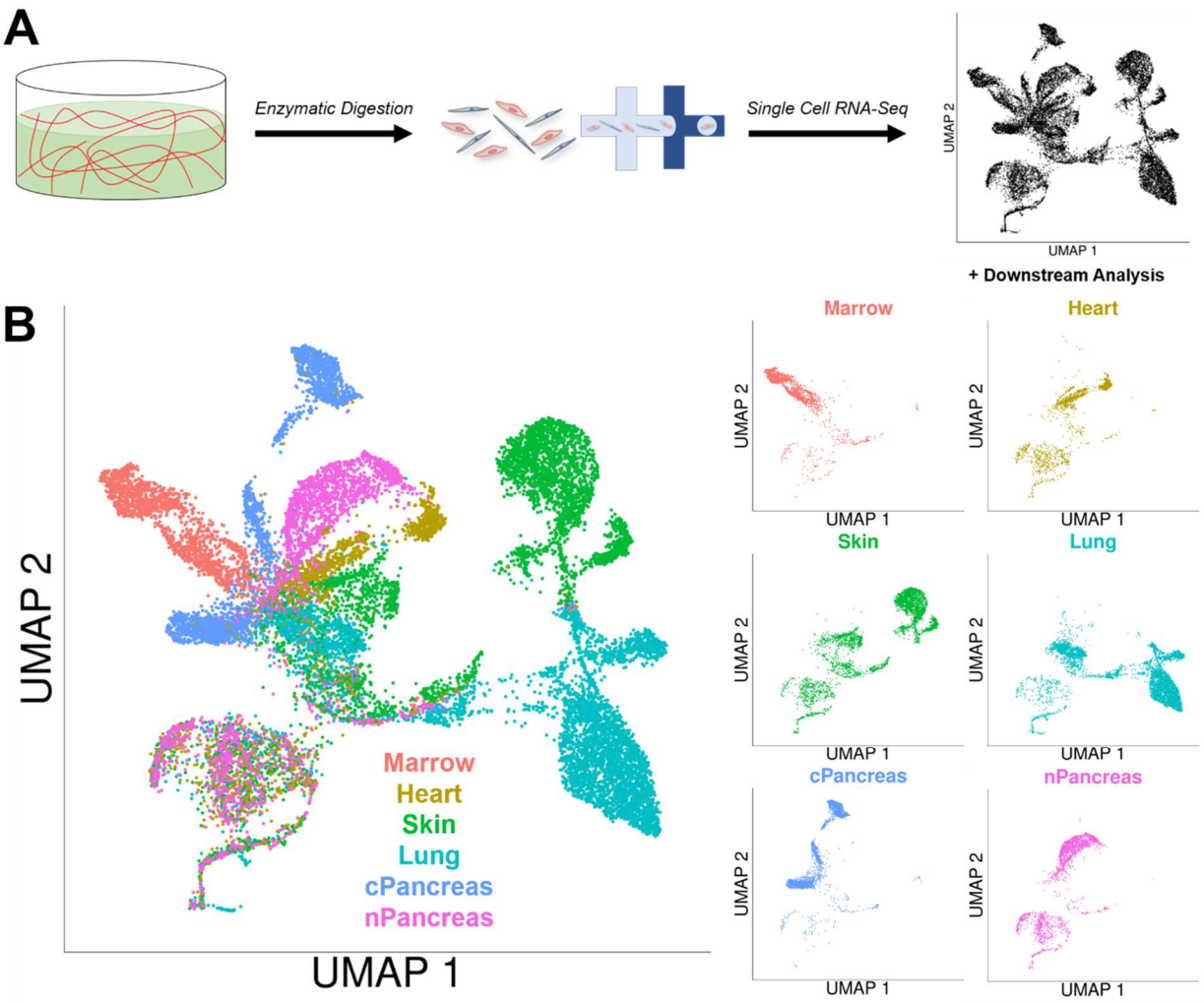

Figure 4
Cells isolated from 3D *in vitro* vascular networks in fibrin hydrogels can be subjected to scRNA-Seq analysis and are arranged into distinct clusters in UMAP space. **A)** Schematic detailing processing of 3D *in vitro* fibrin hydrogels for scRNA-Seq. **B)** Resultant UMAP plots of cells isolated from 3D *in vitro* vascular networks in fibrin hydrogels comprised of the 6 unique vascular network types. Additional UMAP plots are broken down by vascular network type.

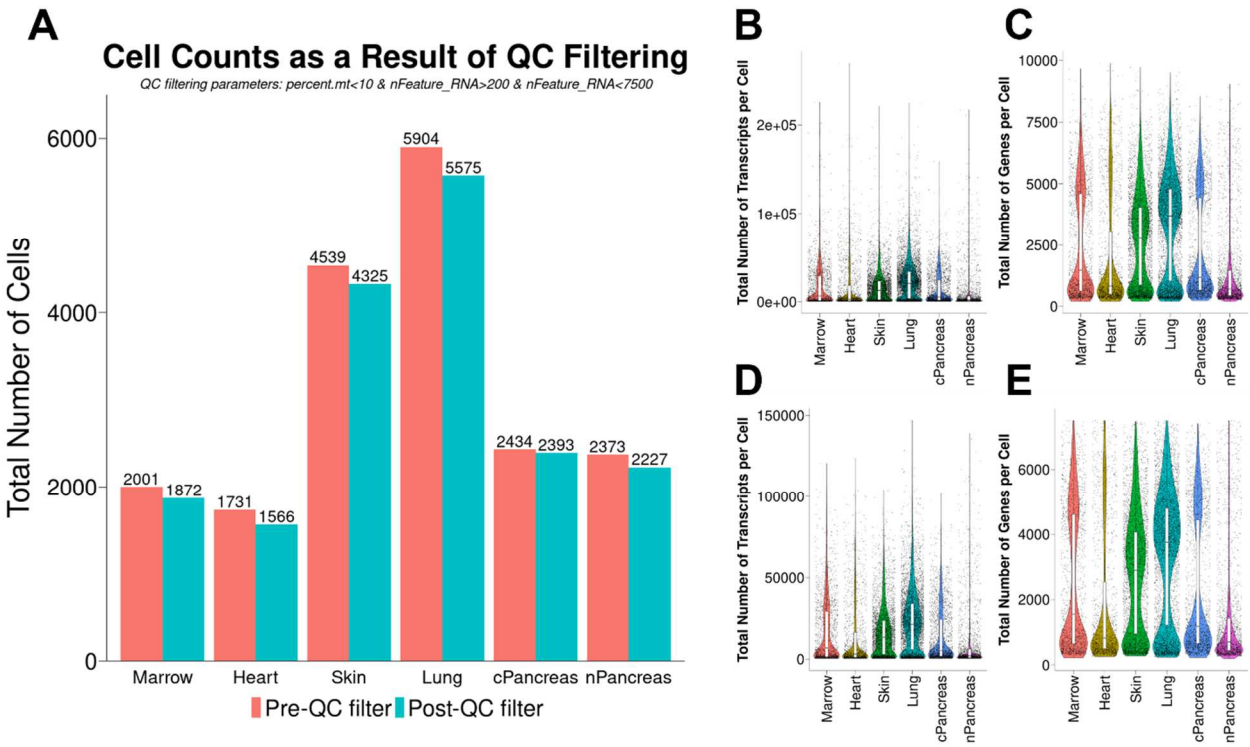

Figure 5

Quality control of complete 3D *in vitro* vascular network scRNA-Seq dataset. **A)** Bar chart indicating number of cells available for analysis from 3D *in vitro* fibrin hydrogels for each vascular network type before and after QC filtering as part of the scRNA-Seq analysis. **B)** Violin plot of total number of transcripts for each sample before QC filtering. **C)** Violin plot of total number of unique genes (UMIs) for each sample before QC filtering. **D)** Violin plot of total number of transcripts for each sample after QC filtering. **E)** Violin plot of total number of unique genes (UMIs) for each sample after QC filtering.

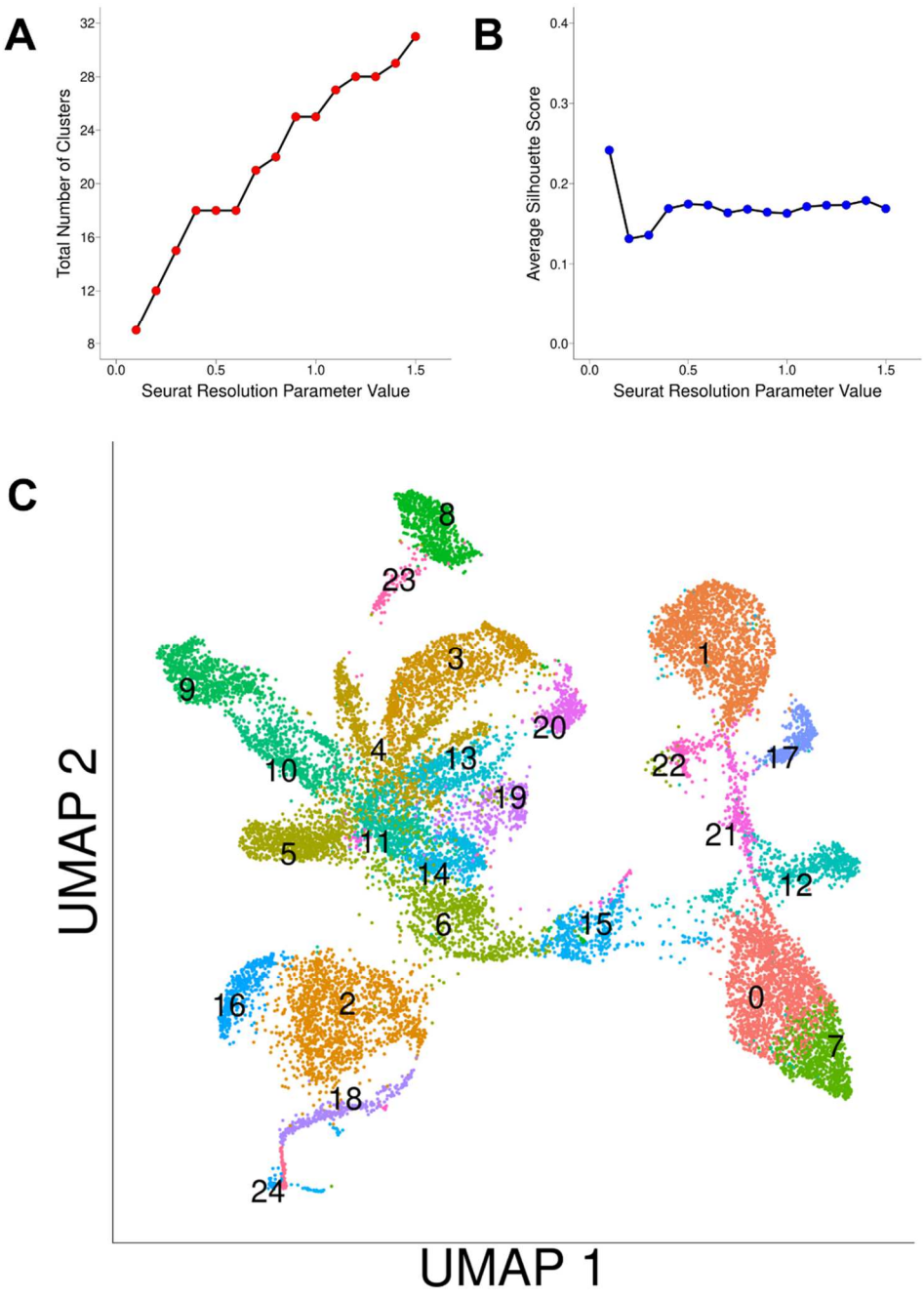

Figure 6

Clustering analysis of 3D *in vitro* vascular network scRNA-Seq dataset. **A)** Scatterplot of resultant number of cell clusters to Seurat resolution parameter (values between 0.1-1.5). **B)** Scatterplot of average silhouette score for select values of the Seurat resolution parameter (values between 0.1-1.5). **C)** Resultant clustering of the 3D *in vitro* vascular network scRNA-Seq dataset using Seurat resolution parameter = 1.0. A total of 25 distinct clusters are identified.

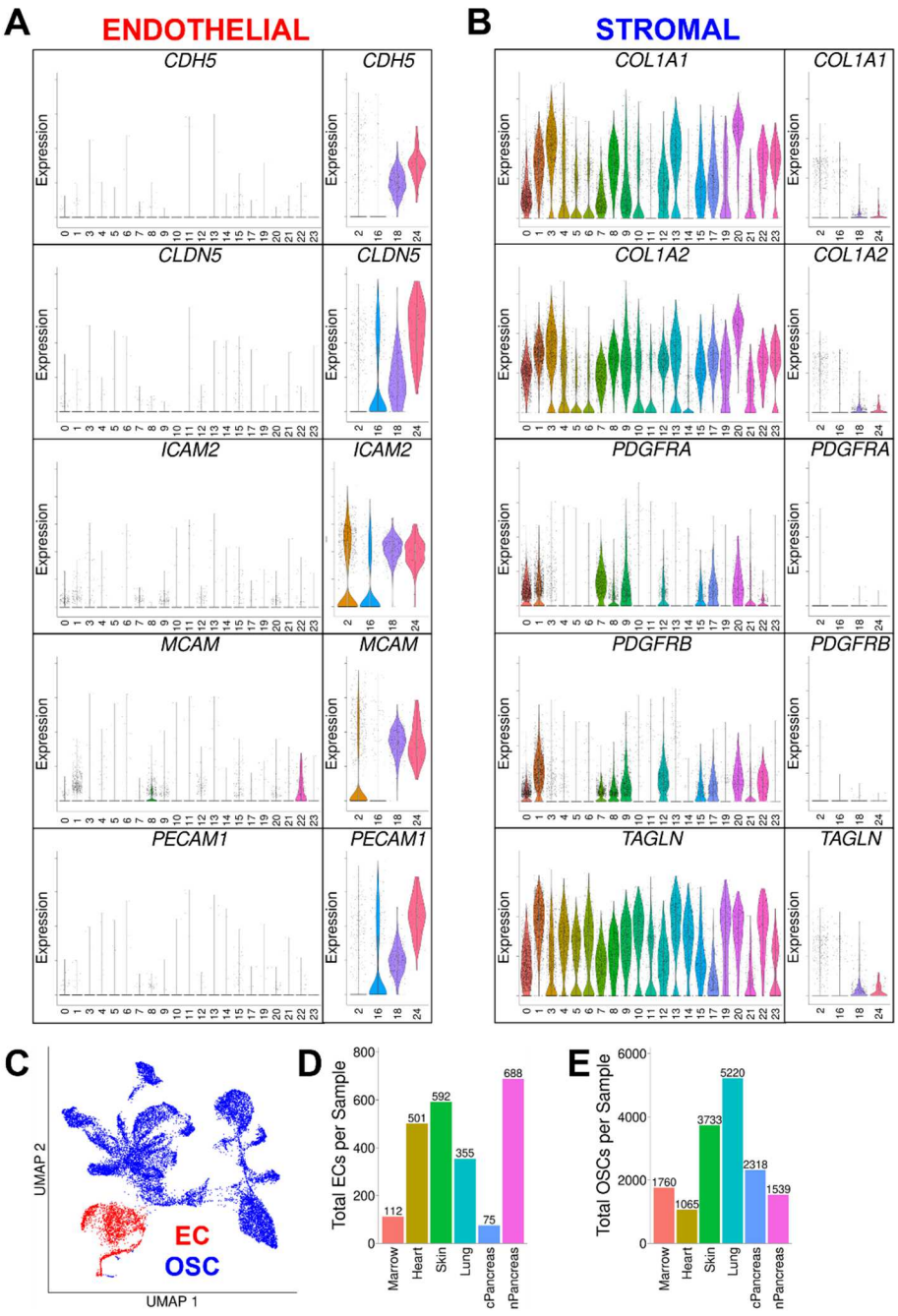

Figure 7

ECs and OSCs can be identified from 3D *in vitro* vascular network dataset by expression of characteristic EC and OSC genes. **A)** Violin plots of characteristic EC genes *CDH5, CLDN5, ICAM2, MCAM*, and *PECAM1*. **B)** Violin plots of characteristic OSC genes *COL1A1, COL1A2, PDGFRA, PDGFRB*, and *TAGLN*. **C)** Overall UMAP plot for complete 3D *in vitro* vascular network dataset broken down by EC/OSC classification (EC=red, OSC=blue). **D)** Bar chart of resultant number of ECs per vascular network type from the complete 3D *in vitro* vascular network dataset. **E)** Bar chart of resultant number of OSCs per vascular network type from the complete 3D *in vitro* fibrin hydrogel dataset.

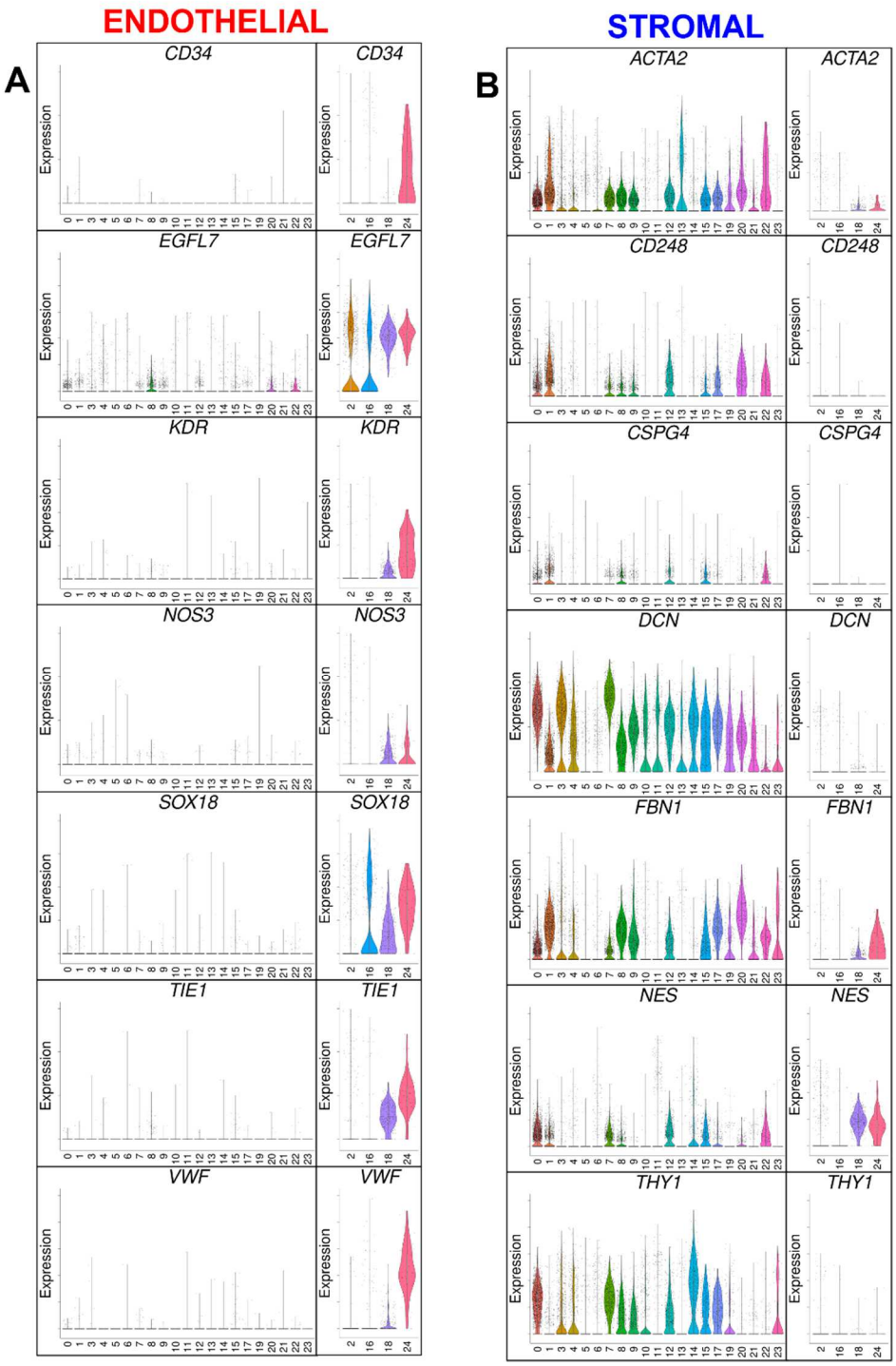

Figure 8

Additional characteristic EC and OSC gene expression of clusters from 3D *in vitro* vascular network dataset. **A)** Violin plots of additional characteristic EC genes *CD34, EGFL7, KDR, NOS3, SOX18, TIE1,* and *VWF* for each cluster identified in complete 3D *in vitro* vascular network dataset. **B)** Violin plots of additional characteristic OSC genes *ACTA2, CD248, CSPG4, DCN, FBN1, NES*, and *THY1* for each cluster identified in complete 3D *in vitro* vascular network dataset.

Distinct EC sub-populations are present in the organotypic vascular networks

The total population of ECs was next re-normalized and re-clustered in order to elucidate specific transcriptomic differences between potential sub-populations of ECs. ECs cluster into 5 distinct groups (labeled “EC-1” through “EC-5”), each with distinct transcriptomic profiles (**Figure 9A-B, Figure 10**). Interestingly, each EC cluster is present in each of the six 3D *in vitro* organotypic vascular networks, although the relative fraction of each EC cluster depends on the OSC (**Figure 9C-D**).

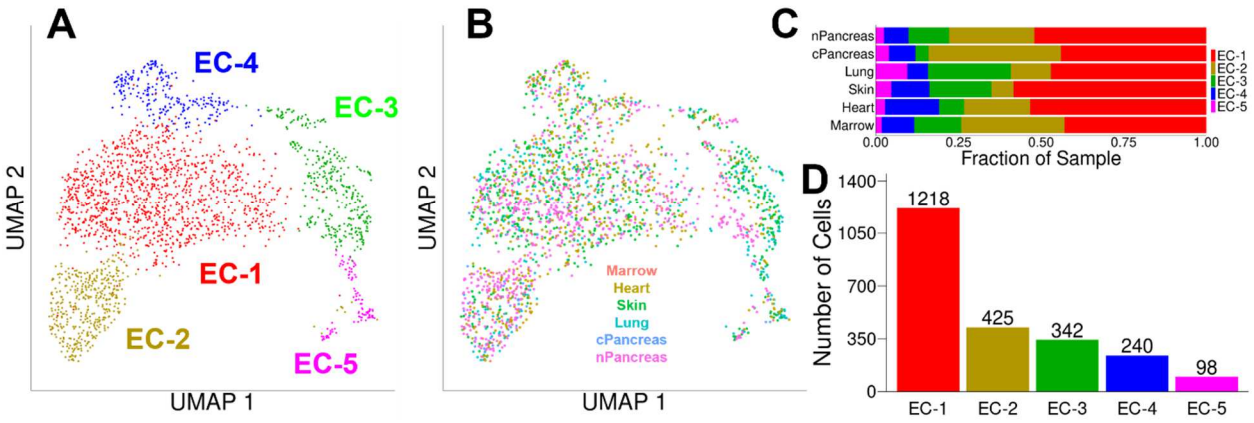

Figure 9

ECs isolated from 3D *in vitro* vascular networks separate into distinct clusters. **A)** UMAP plot of renormalized 3D *in vitro* EC dataset separates into 5 distinct clusters (EC-1, EC-2, EC-3, EC-4, and EC-5). **B)** UMAP plot of renormalized 3D *in vitro* EC dataset grouped by vascular network type. **C)** Grouped bar chart shows relative percentages of each EC cluster per vascular network type. **D)** Bar chart of total number of ECs per EC cluster.

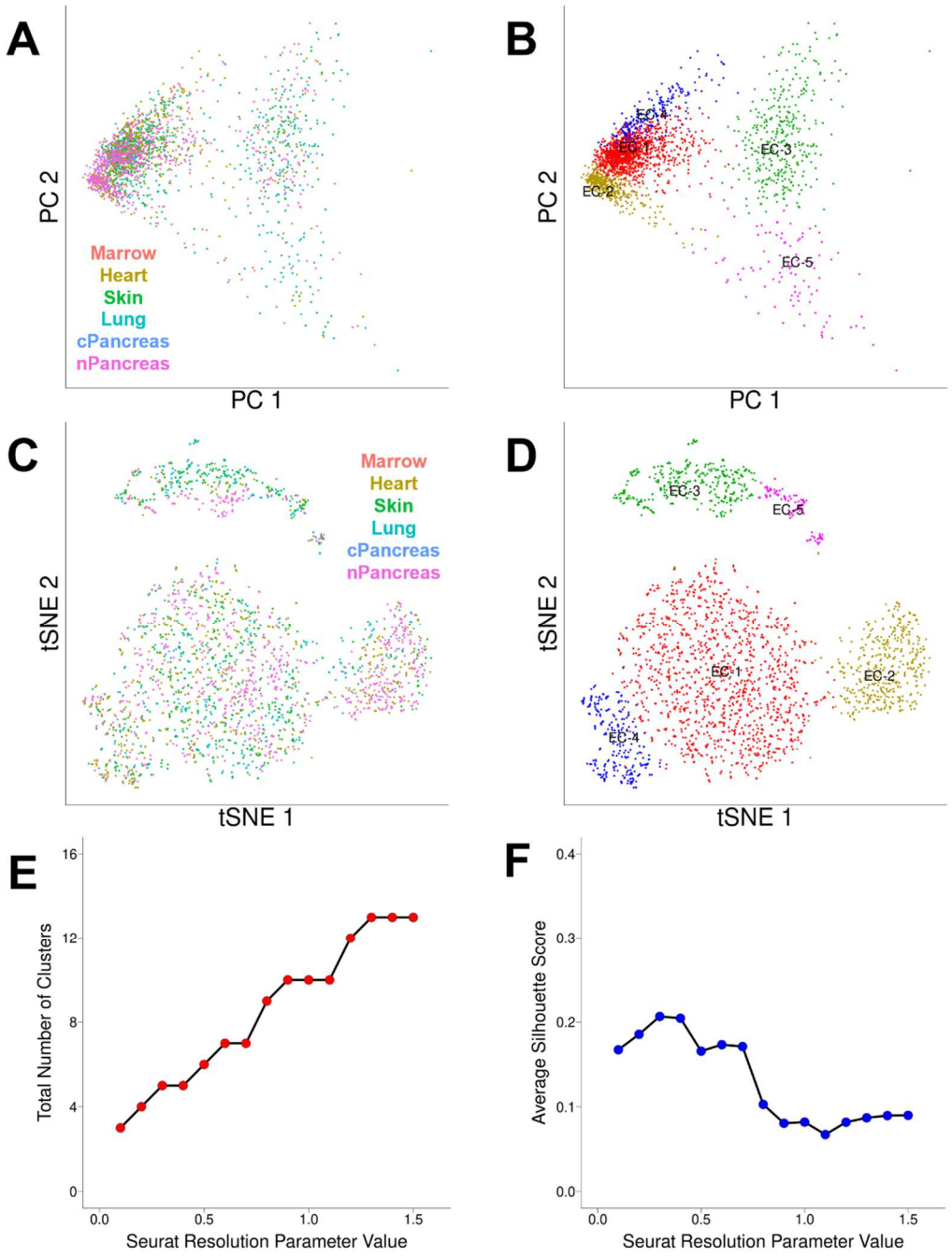

Figure 10

Dimensionality reduction of analysis of clustering results of ECs isolated from 3D *in vitro* vascular network dataset. **A)** Plot of first two principal components separated by vascular network type. **B)** Plot of first two principal components separated by EC cluster identity. **C)** Plot of first two tSNE components separated by vascular network type. **D)** Plot of first two tSNE components separated by EC cluster identity. **E)** Scatterplot of resultant number of cell clusters to Seurat resolution parameter (values between 0.1-1.5). **F)** Scatterplot of average silhouette score for select values of the Seurat resolution parameter (values between 0.1-1.5).

After obtaining a list of differentially expressed genes (DEGs) for each cluster of ECs, we performed gene ontology (GO) analysis to identify distinct biological processes characteristic of each EC cluster (**Figure 11A-B, Table 4**). EC-1 is characterized by DEGs that include redox-related genes (e.g., *MT2A*, *MT1E*, *TXN*), and is associated with enhanced metabolism based on GO terms including "mitochondrial respiratory chain complex I assembly" and "mitochondrial electron transport, NADH to ubiquinone." EC-2 is characterized by DEGs that include ribosomal-related genes (e.g., *RPS27*) associated with protein regulation and synthesis based on GO terms including "translational initiation" and "ribosomal large subunit assembly". EC-3 is characterized by DEGs that include cell-ECM regulatory genes (e.g., *THBS1, CTGF*), and is associated with migration and adhesion based on GO terms including "positive regulation of cell migration" and "cell adhesion." EC-4 is characterized by DEGs that include histone-related genes (e.g., *HIST1H1B*, *HIST1H13B*, *HIST1H2AG*) and cell cycle-related genes (e.g., *UBE2C*, *HMGB2*, *CKS2*), and is associated with cell proliferation based on GO terms including "cell division" and "microtubule cytoskeleton organization." Finally, EC-5 is characterized by DEGs that include several genes related to endothelial basement membrane (e.g., *COL4A2*, *HSPG2*, *COL4A1*, *COL18A1*), and is associated with angiogenesis based on GO terms including "angiogenesis" and "positive regulation of angiogenesis." When pseudotime values are mapped onto the UMAP plots (Monocle 3 [47]) different EC clusters follow a pattern from a less differentiated EC involved in migration, adhesion, and angiogenesis to a more fully differentiated and synthetic phenotype (EC3→EC5→EC4→EC1→EC2) (**Figure 12**).

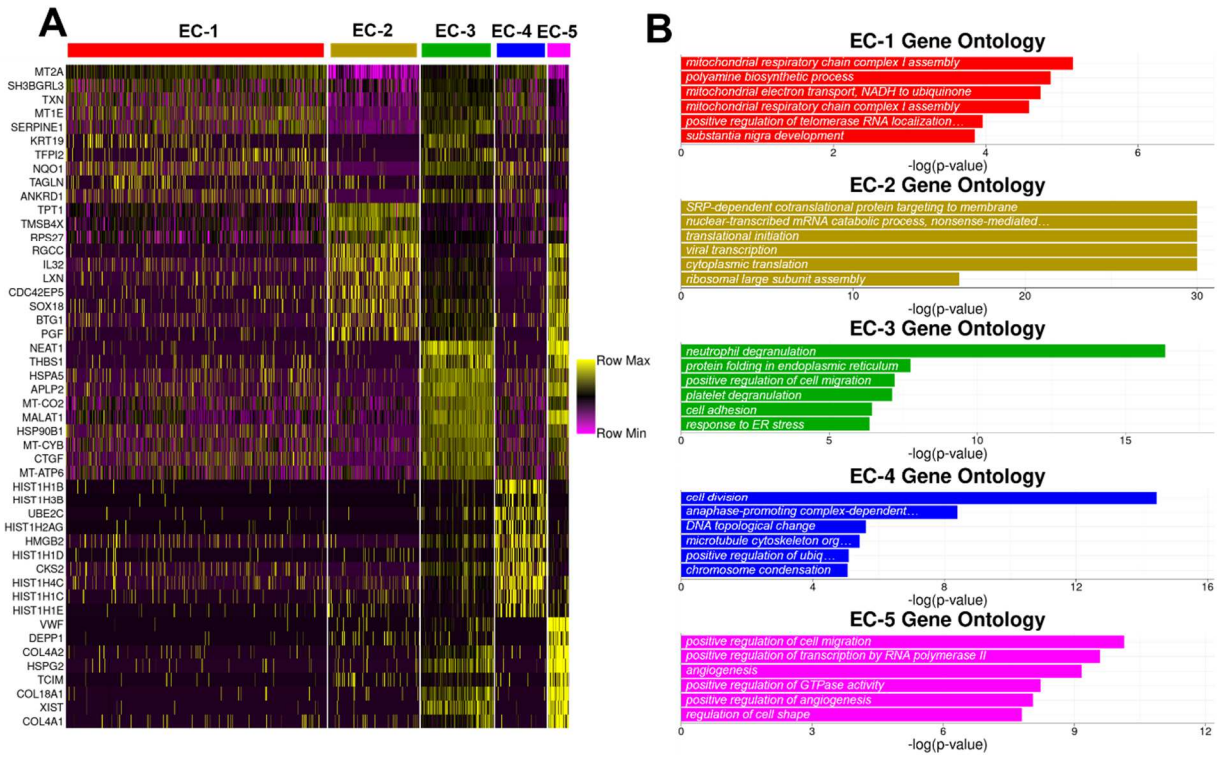

Figure 11

Renormalized EC clusters have unique transcriptomic profiles and unique GO terms associated with their respective DEGs. **A)** Heatmap of top 10 DEGs for each EC cluster. **B)** Top 6 most significant GO terms based on DEGs for each EC cluster.

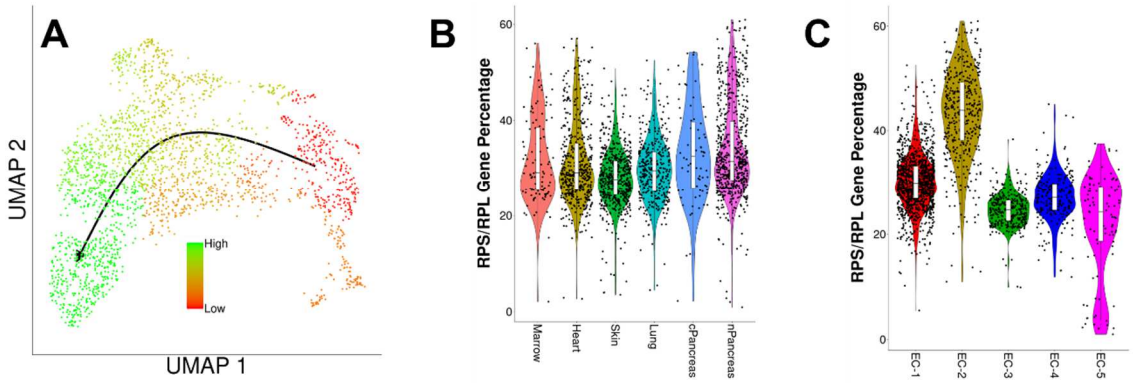

Figure 12

Renormalized EC clusters exhibit different pseudotime values for each cluster and have different expression of ribosomal-related genes. **A)** Pseudotime values of 3D in vitro EC dataset as determined by Monocle3, mapped onto the UMAP dimensionality reduction of the renormalized ECs (green = late/high pseudotime, red = early/low pseudotime). **B)** Violin plot of percentage of ribosomal genes (percentage of gene expression by *RPS* and *RPL* genes per cell) grouped by vascular network type. **C)** Violin plot of percentage of ribosomal genes (percentage of gene expression by *RPS* and *RPL* genes per cell) grouped by EC cluster identity.

In order to better understand the potential impact of EC clusters on the morphology of the individual vascular networks, we correlated the relative percentage of each EC cluster to the mean total vessel length for each 3D *in vitro* vascular network. We observed that the relative percentage of EC-2 in each vascular network is negatively correlated with the mean total vessel length of each vascular network (**Figure 13**; R = -0.857, *p = 0.029).

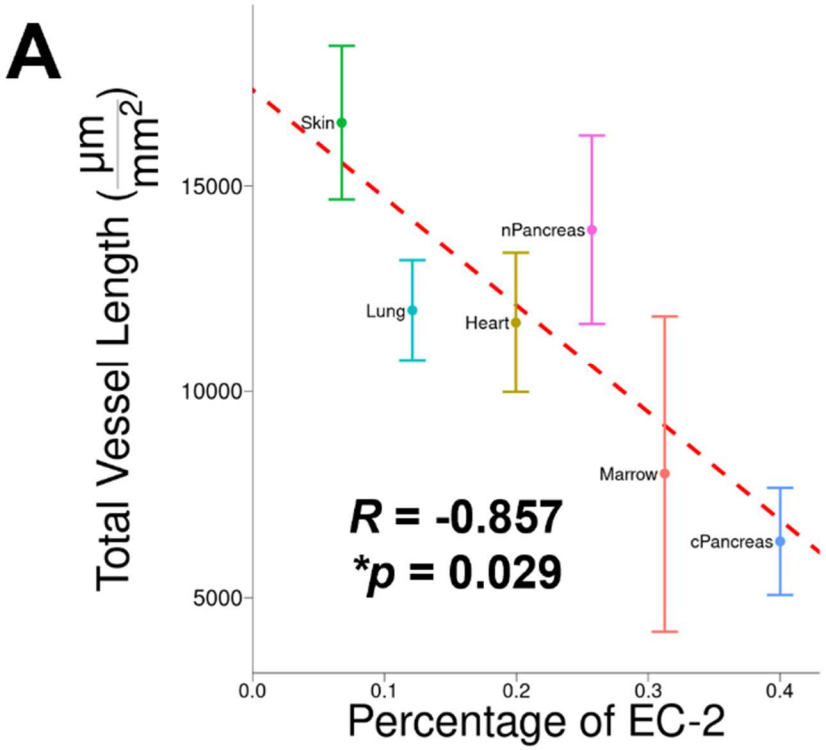

Figure 13
EC phenotype correlates with EC transcriptome. **A)** Correlation of the total vessel length measured for each 3D *in vitro* vascular network (Figure 2) compared with the relative percentage of EC-2 in each 3D *in vitro* vascular network (Figure 9). Error bars represent 1 standard deviation above and below the mean total vessel length. Pearson correlation coefficient R = -0.857 with associated p = 0.029.

#	EC-1	EC-2	EC-3	EC-4	EC-5
1	*MT2A*	*RGCC*	*NEAT1*	*HIST1H1B*	*MALAT1*
2	*MT1E*	*IL32*	*MT-CO2*	*HIST1H4C*	*NEAT1*
3	*TXN*	*LXN*	*HSPA5*	*UBE2C*	*VWF*
4	*SERPINE1*	*CDC42EP5*	*APLP2*	*HIST1H2AG*	*COL4A1*
5	*SH3BGRL3*	*SOX18*	*MT-CYB*	*HIST1H1C*	*HSPG2*
6	*RPL22L1*	*PGF*	*THBS1*	*HIST1H1D*	*COL4A2*
7	*BOLA3*	*TMSB4X*	*HSP90B1*	*CKS2*	*TCIM*
8	*HMGA1*	*TPT1*	*MT-ATP6*	*HMGB2*	*COL18A1*
9	*S100A6*	*RPS27*	*MALAT1*	*HIST1H3B*	*DEPP1*
10	*ANXA2*	*BTG1*	*CTGF*	*HIST1H1E*	*XIST*
11	*POLR2E*	*PNRC1*	*CYR61*	*HIST2H2AC*	*PECAM1*
12	*PRDX1*	*GNG11*	*AHNAK*	*CDK1*	*ARGLU1*
13	*GSTP1*	*RPL10*	*EDN1*	*CCNB1*	*CLDN5*
14	*PFN1*	*C12orf57*	*MT-CO3*	*TOP2A*	*WSB1*
15	*FKBP11*	*CALM1*	*HHIP*	*TPX2*	*TCF4*
16	*ADIRF*	*NOP53*	*MT-CO1*	*KPNA2*	*RHOB*
17	*LDHA*	*RPL28*	*ITGB1*	*NUSAP1*	*UACA*
18	*PSMD2*	*RPS28*	*PTX3*	*TUBB4B*	*PXDN*
19	*PPIA*	*RPL13*	*CRIM1*	*RRM2*	*MACF1*
20	*OAZ1*	*COMMD6*	*CLEC14A*	*TUBA1C*	*PNISR*
21	*MLLT11*	*RPL34*	*MT-ND4*	*CKS1B*	*SPTBN1*
22	*UROD*	*RPL26*	*ENG*	*AURKB*	*CLEC14A*
23	*ENO1*	*RPS4X*	*EFEMP1*	*UBE2S*	*LAMB1*
24	*FKBP3*	*RPS8*	*CKAP4*	*DLGAP5*	*THBS1*
25	*SRGN*	*RPS15A*	*MT-ND3*	*CDC20*	*NRP2*
26	*MYL12B*	*RPL39*	*SLC3A2*	*UBE2T*	*CDH5*
27	*S100A11*	*RPS26*	*SLC7A11*	*TUBA1B*	*UTRN*
28	*EFHD2*	*SELENOW*	*PRSS23*	*HIST1H4E*	*MYH9*
29	*VPS29*	*RPL15*	*CLDN11*	*SMC4*	*MMRN1*
30	*SDCBP*	*OST4*	*FLNB*	*H2AFZ*	*GABPB1-AS1*
31	*DKK1*	*RPS14*	*NR2F2*	*H2AFX*	*RSRP1*
32	*VEPH1*	*ZFAS1*	*MACF1*	*TYMS*	*HLA-B*
33	*TAF10*	*CCDC85B*	*AKAP12*	*PTTG1*	*KIAA1551*
34	*UCHL3*	*RPS3A*	*MMRN1*	*BIRC5*	*LUC7L3*
35	*NDUFA3*	*RPL37*	*APP*	*G0S2*	*RNF213*
36	*MYL6*	*RPL30*	*MCAM*	*PCLAF*	*PLK2*
37	*NDUFB6*	*GUK1*	*HMOX1*	*HMGN2*	*DLL4*
38	*IL1RL1*	*RPL32*	*ITGA5*	*STMN1*	*SOX4*
39	*NDUFAF2*	*RPL22*	*CD59*	*HMGB1*	*APP*
40	*CTNNAL1*	*RPS19*	*MT-ND2*	*RANBP1*	*RGCC*

Table 4

Top 40 DEGs for each EC cluster, ordered by fold change over all other clusters (as determined by Seurat). Top 10 DEGs for each cluster are shown in Figure 11.

Common and unique OSC sub-populations are present in 3D organotypic vasculature

The population of OSCs was next re-normalized and re-clustered in order to elucidate specific transcriptomic differences between potential sub-populations of OSCs present in the 3D *in vitro* models of organotypic vasculature. OSCs cluster into 8 distinct groups (labeled "OSC-1" through "OSC-8"), each with distinct transcriptomic profiles (**Figure 14A-B, Figure 16, Table 5**). Interestingly, each coculture is represented by a combination of these 8 clusters with each having a unique, dominant OSC cluster. All, however, share 2 clusters – OSC 1 and OSC-7 – albeit with the relative fraction of the common OSC clusters being different for each vascular network (**Figure 14C-D**). OSC-2 primarily consists of Lung OSCs; OSC-3 primarily consists of Skin OSCs; OSC-4 primarily consists of nPancreas OSCs; OSC-5 primarily consists of cPancreas OSCs; OSC-6 primarily consists of Marrow OSCs; and OSC-8 primarily consists of Heart OSCs. Thus, each of the OSC populations is transcriptionally distinct from the others, indicating they retain the "memory" of their origin tissue.

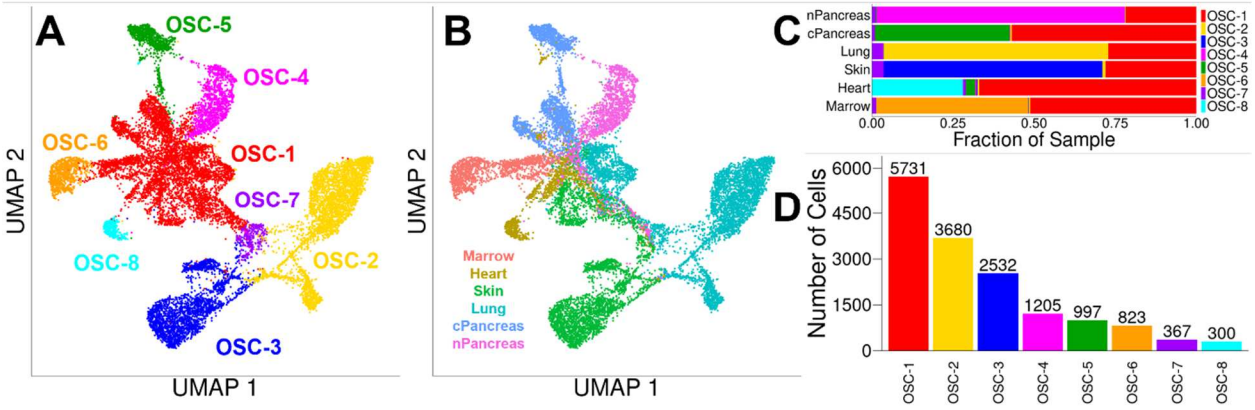

Figure 14

OSCs isolated from 3D *in vitro* vascular networks separate into distinct clusters. **A)** UMAP plot of renormalized 3D *in vitro* OSC dataset separates into 8 distinct clusters (OSC-1, OSC-2, OSC-3, OSC-4, OSC-5, OSC-6, OSC-7, and OSC-8). **B)** UMAP plot of renormalized 3D *in vitro* OSC dataset grouped by vascular network type. **C)** Grouped bar chart shows relative percentages of each OSC cluster per vascular network type. **D)** Bar chart of total number of OSCs per OSC cluster.

All of these unique OSC clusters differentially express matricellular genes (e.g., *DCN, LOX, PLAT*) and extracellular matrix proteins (e.g., *COL4A2*, *COL1A1*, *COL6A3*), and have GO terms related to extracellular matrix organization, wound healing, cytokine secretion, and angiogenesis (**Figure 15**), reflective of their fibroblast-like identity. OSC-1 is a large population of OSCs present in all cocultures and is relatively poorly defined transcriptionally. OSC-7 is a small population of OSCs present in all vascular networks and has GO terms related to cell division, suggesting that a relatively small number of OSCs in each coculture are actively proliferating.

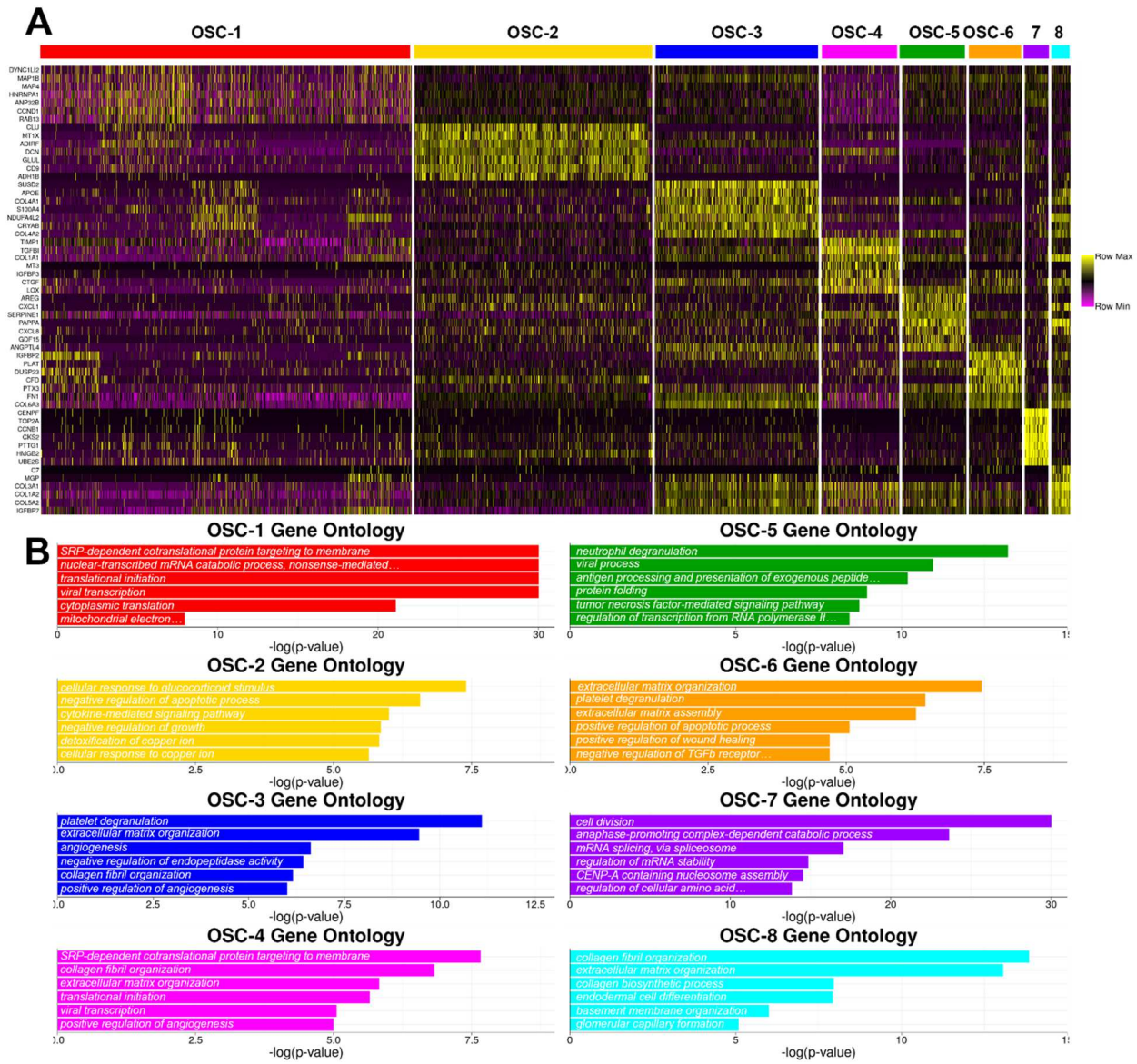

Figure 15

Renormalized OSC clusters have unique transcriptomic profiles and unique GO terms associated with their respective DEGs. **A)** Heatmap of top 7 DEGs for each OSC cluster. **B)** Top 6 most significant GO terms based on DEGs for each OSC cluster.

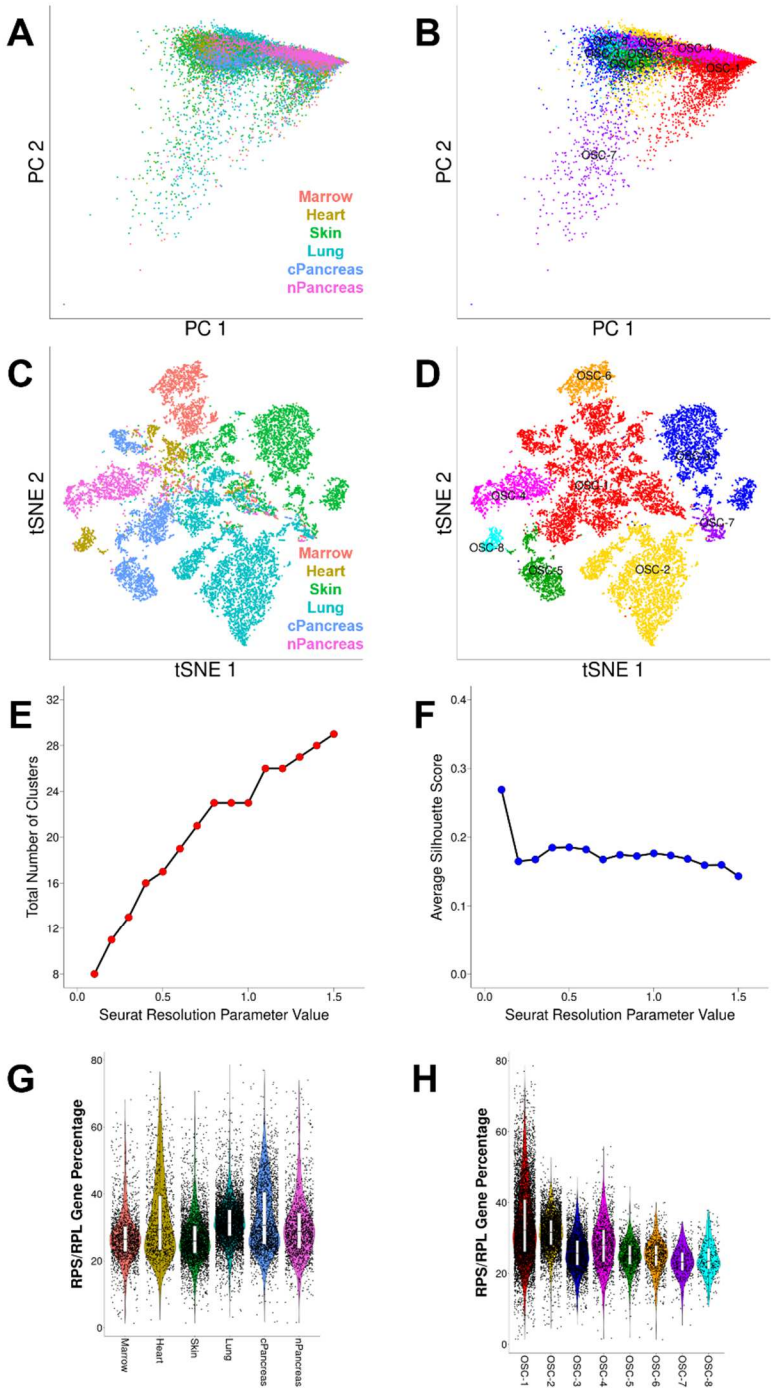

Figure 16

Dimensionality reduction of analysis of clustering results of OSCs isolated from 3D *in vitro* vascular network dataset. **A)** Plot of first two principal components separated by vascular network type. **B)** Plot of first two principal components separated by OSC cluster identity. **C)** Plot of first two tSNE components separated by vascular network type. **D)** Plot of first two tSNE components separated by OSC cluster identity. **E)** Scatterplot of resultant number of cell clusters to Seurat resolution parameter (values between 0.1-1.5). **F)** Scatterplot of average silhouette score for select values of the Seurat resolution parameter (values between 0.1-1.5).

#	OSC-1	OSC-2	OSC-3	OSC-4	OSC-5	OSC-6	OSC-7	OSC-8
1	*CCND1*	*CLU*	*SUSD2*	*TIMP1*	*AREG*	*IGFBP2*	*CENPF*	*COL3A1*
2	*ANP32B*	*MT1X*	*APOE*	*IGFBP3*	*CXCL1*	*PTX3*	*CKS2*	*COL1A1*
3	*DYNC1LI2*	*ADIRF*	*COL4A1*	*TGFBI*	*CXCL8*	*CFD*	*PTTG1*	*COL1A2*
4	*MAP4*	*DCN*	*S100A4*	*MT3*	*SERPINE1*	*FN1*	*UBE2S*	*IGFBP7*
5	*HNRNPA1*	*GLUL*	*NDUFA4L2*	*LOX*	*ANGPTL4*	*PLAT*	*HMGB2*	*C7*
6	*MAP1B*	*CD9*	*CRYAB*	*COL1A1*	*PAPPA*	*COL6A3*	*TOP2A*	*MGP*
7	*S100A11*	*ADH1B*	*COL4A2*	*CTGF*	*GDF15*	*DUSP23*	*CCNB1*	*COL5A2*
8	*SH3BGRL3*	*MT1E*	*BGN*	*GLRX*	*HSPA5*	*RPS26*	*ASPM*	*FBN1*
9	*NPM1*	*SFRP1*	*CCL2*	*CTHRC1*	*RALA*	*COL8A1*	*UBE2C*	*PAPPA*
10	*PTPN11*	*CAMK2N1*	*TAGLN*	*LMCD1*	*GREM1*	*STC2*	*HIST1H4C*	*COL6A3*
11	*ZEB1*	*ID1*	*ID3*	*F3*	*SOD2*	*CCDC68*	*TPX2*	*CXCL2*
12	*UACA*	*ADAMTS1*	*COL6A3*	*FBXO32*	*INHBA*	*GAS6*	*H2AFZ*	*COL6A1*
13	*SET*	*DUSP1*	*COL6A1*	*COL3A1*	*HMGA2*	*LEPR*	*CKAP2*	*NDUFA4L2*
14	*MAP4K4*	*PHLDA1*	*NR4A1*	*CCDC80*	*SLC16A3*	*SRGN*	*STMN1*	*COL5A1*
15	*KIF1C*	*C1S*	*XIST*	*SPARC*	*PDLIM4*	*TXNIP*	*CDKN3*	*CXCL1*
16	*PALLD*	*MALAT1*	*LY6K*	*PLOD2*	*P4HA2*	*PLA2G16*	*TUBB4B*	*SPARC*
17	*TCEAL4*	*C1R*	*CYR61*	*SERPINE1*	*MME*	*CHST2*	*MKI67*	*COL6A2*
18	*CFL1*	*CCDC85B*	*COL6A2*	*IER3*	*CDKN1A*	*ITGBL1*	*SMC4*	*LOXL2*
19	*FKBP1A*	*PLIN2*	*C11orf96*	*COL1A2*	*CXCL2*	*PIP*	*NUSAP1*	*COL11A1*
20	*C12orf75*	*CEBPD*	*HLA-DRB1*	*COL5A1*	*HILPDA*	*DSEL*	*CENPE*	*CXCL3*
21	*MYL9*	*TBX3*	*TNC*	*NNMT*	*PLOD2*	*STMN2*	*CKS1B*	*CTGF*
22	*PFN1*	*MT1G*	*ACTA2*	*GREM1*	*SPON2*	*WNT5B*	*ARL6IP1*	*HTRA1*
23	*PPIA*	*NFKBIA*	*LOXL2*	*CTSC*	*VCAN*	*CCPG1*	*HMMR*	*PCOLCE*
24	*HSP90AA1*	*SELENOP*	*JUNB*	*THBS1*	*G0S2*	*CPM*	*BIRC5*	*COL4A1*
25	*MYL12A*	*IGFBP6*	*EMP1*	*POSTN*	*UCHL1*	*THBS1*	*TUBA1C*	*C11orf96*
26	*OAZ1*	*SPON2*	*COL8A1*	*HSD17B1*	*NDRG1*	*NPR3*	*KPNA2*	*CRLF1*
27	*MYL12B*	*MT1M*	*EPS8*	*CCL20*	*UGCG*	*TFPI*	*KIF20B*	*SRGN*
28	*TFPI2*	*TXNIP*	*PDGFRB*	*HLA-B*	*PLAUR*	*LEPROT*	*TUBA1B*	*EFEMP1*
29	*YBX1*	*SRPX*	*FOS*	*INHBA*	*ERO1A*	*NUPR1*	*CCNB2*	*MT-CO1*
30	*PTMS*	*GPX3*	*MT1M*	*TPM1*	*MEG3*	*RTN4*	*CDK1*	*GSTT2B*
31	*NCL*	*CTSK*	*SCUBE3*	*TFPI2*	*CCL20*	*DKK1*	*CDC20*	*ID3*
32	*TMSB4X*	*LGALS3*	*SERPINF1*	*HSD11B1*	*METRNL*	*GALNT15*	*ANLN*	*COL5A3*
33	*ATP5F1E*	*PDE5A*	*NID1*	*FKBP11*	*DCBLD2*	*TRIB3*	*DLGAP5*	*SERPINH1*
34	*CYCS*	*TBX2-AS1*	*NOTCH3*	*SAT1*	*TNFAIP6*	*DNAJC15*	*AURKA*	*POSTN*
35	*TPM4*	*NR2F1*	*TINAGL1*	*RGCC*	*CALR*	*EPSTI1*	*HMGB1*	*ACTA2*
36	*EIF5B*	*CKB*	*NR2F2*	*HIF1A*	*FAM162A*	*NETO2*	*MAD2L1*	*MT-CYB*
37	*RPL22L1*	*F2R*	*MALAT1*	*LDHA*	*VEGFA*	*HTRA1*	*BUB3*	*HGF*
38	*HNRNPA3*	*CXCL3*	*LAMC1*	*RECK*	*CXCL6*	*GADD45A*	*SGO2*	*COL7A1*
39	*TPM2*	*NEAT1*	*ID1*	*KDELR3*	*HIF1A*	*SLC3A2*	*JPT1*	*CXCL8*
40	*TPM3*	*PTN*	*HTRA1*	*KDELR2*	*KRT18*	*LMO7*	*CENPF*	*THBS2*

Table 5

Top 40 DEGs for each OSC cluster, ordered by fold change over all other clusters (as determined by Seurat).

3D in vitro ECs adopt organ-specific features of in vivo ECs

We next compared the transcriptome of our 2D ECFC-EC monolayer and 3D *in vitro* vascular networks to publicly available *in vivo* datasets consisting of cells isolated from skin [36] and whole lung [35]. For the skin *in vivo* data, we performed k-means clustering, which yielded 18 total clusters. We identified the ECs (clusters 3, 13, and 16) by differential expression of EC characteristic genes *CDH5*, *CLDN5*, and *EGFL7* (**Figure 17A-B**; **Figure 18**, **Figure 19**). Other cell types present in the skin *in vivo* dataset included *PDGFRA*+ stromal cells (clusters 0, 2, 4, 9, 10), *PTPRC*+ leukocytes (clusters 1, 7, 12, 14), *KRT*+ keratinocytes (clusters 5, 8, 11), *TAGLN*+ stromal cells (clusters 6, 17), and *MITF*+ melanocytes (cluster 15). We therefore included clusters 3, 13, and 16 for the downstream analysis of ECs. Based on the similarity score, 3D *in vitro* Skin vascular network ECs are closer to the *in vivo* skin ECs, and also demonstrate a much lower inter-cellular variation (less cell to cell heterogeneity in the transcriptome) compared to the 2D *in vitro* monolayer ECs ($p < 0.05$) (**Figure 17C-D**). We also performed a similar analysis with ECs identified from an *in vivo* whole lung dataset and obtained a similar result (**Figure 20-22**).

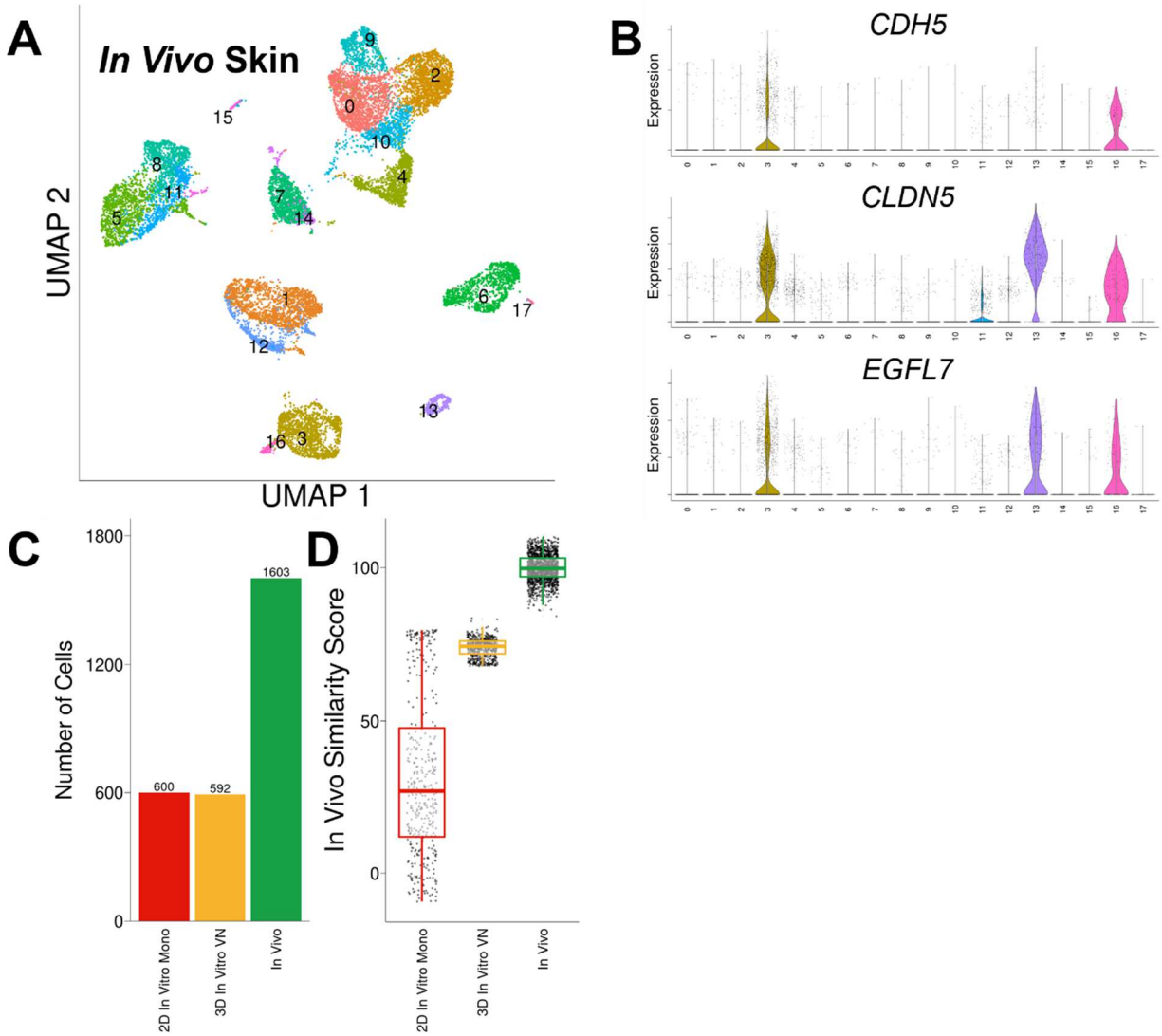

Figure 17

3D *in vitro* skin vascular network-derived ECs more mimic ECs from previously collected *in vivo* datasets compared to 2D *in vitro* ECFC-EC monolayer. **A)** UMAP plot of skin *in vivo* data from Solé-Boldo, *et al*. Clusters determined in Seurat. **B)** Violin plots showing expression of EC-characteristic genes *CDH5, CLDN5,* and *EGFL7* for clusters identified in the *in vivo* skin dataset. **C)** Bar chart of total number of ECs evaluated from each dataset (red = 2D *in vitro* ECFC-EC monolayer; yellow = 3D *in vitro* EC from Skin vascular network; green = *in vivo* skin ECs). **D)** Normalized similarity score metric based on top 20 DEGs from *in vivo* dataset.

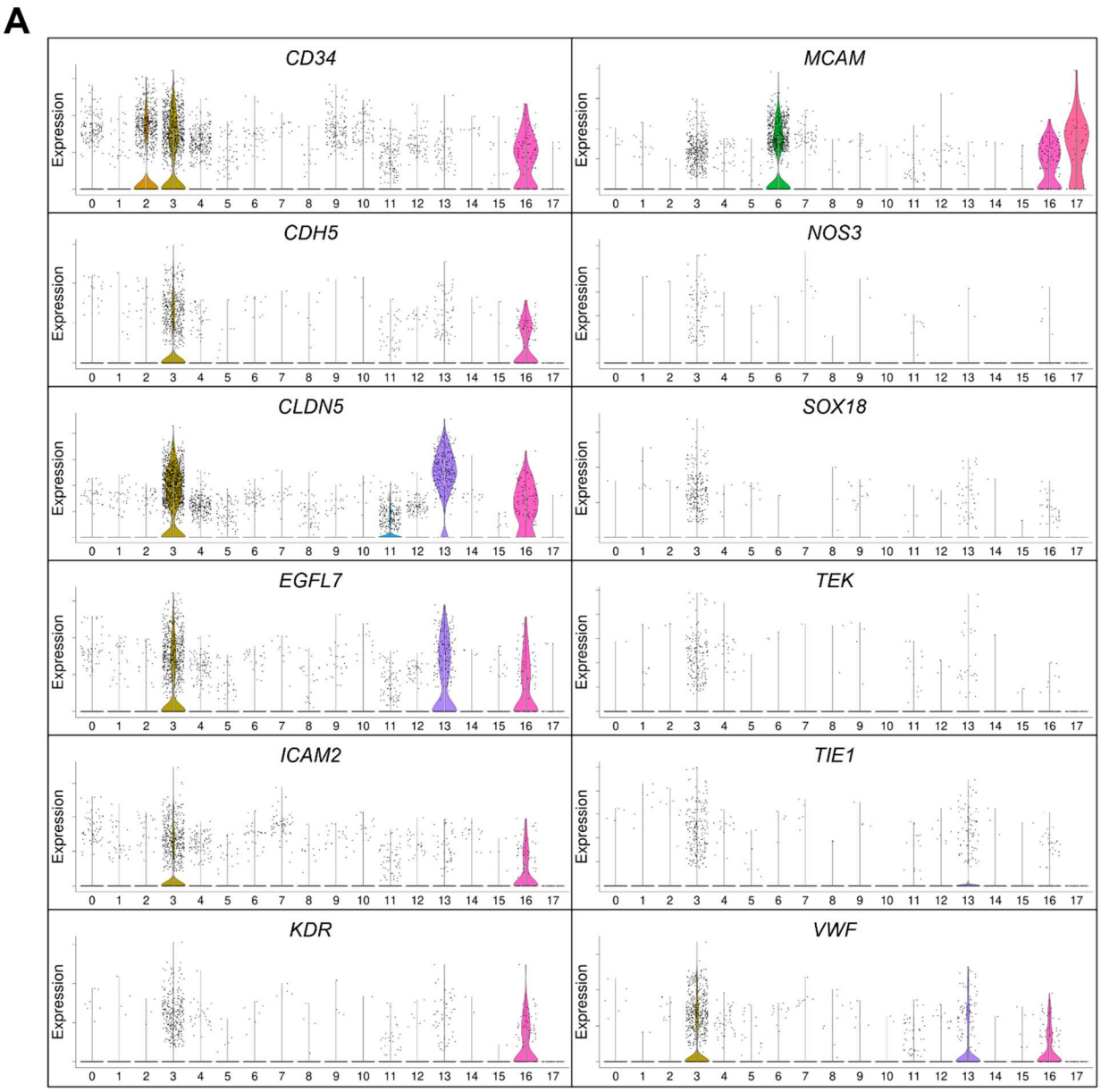

Figure 18

Additional violin plots of EC-characteristic genes from cluster identified in skin *in vivo* dataset (data from Solé-Boldo, *et al*.). **A)** Violin plots of EC-characteristic genes *CD34, CDH5, CLDN5, EGFL7, ICAM2, KDR, MCAM, NOS3, SOX18, TEK, TIE1,* and *VWF*.

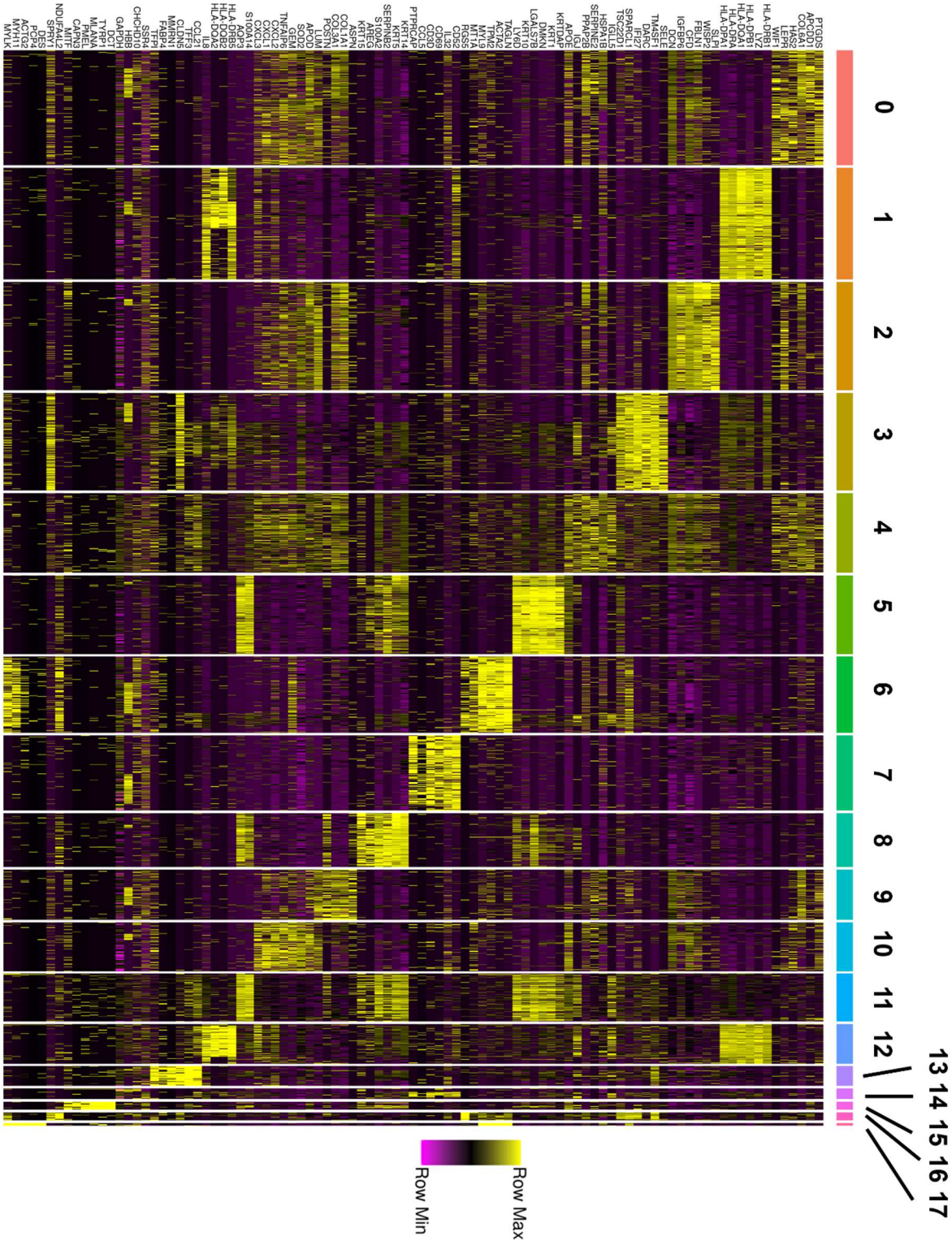

Figure 19
Heatmap of top 6 DEGs (per cluster) for *in vivo* whole skin dataset from Solé-Boldo, *et al*.

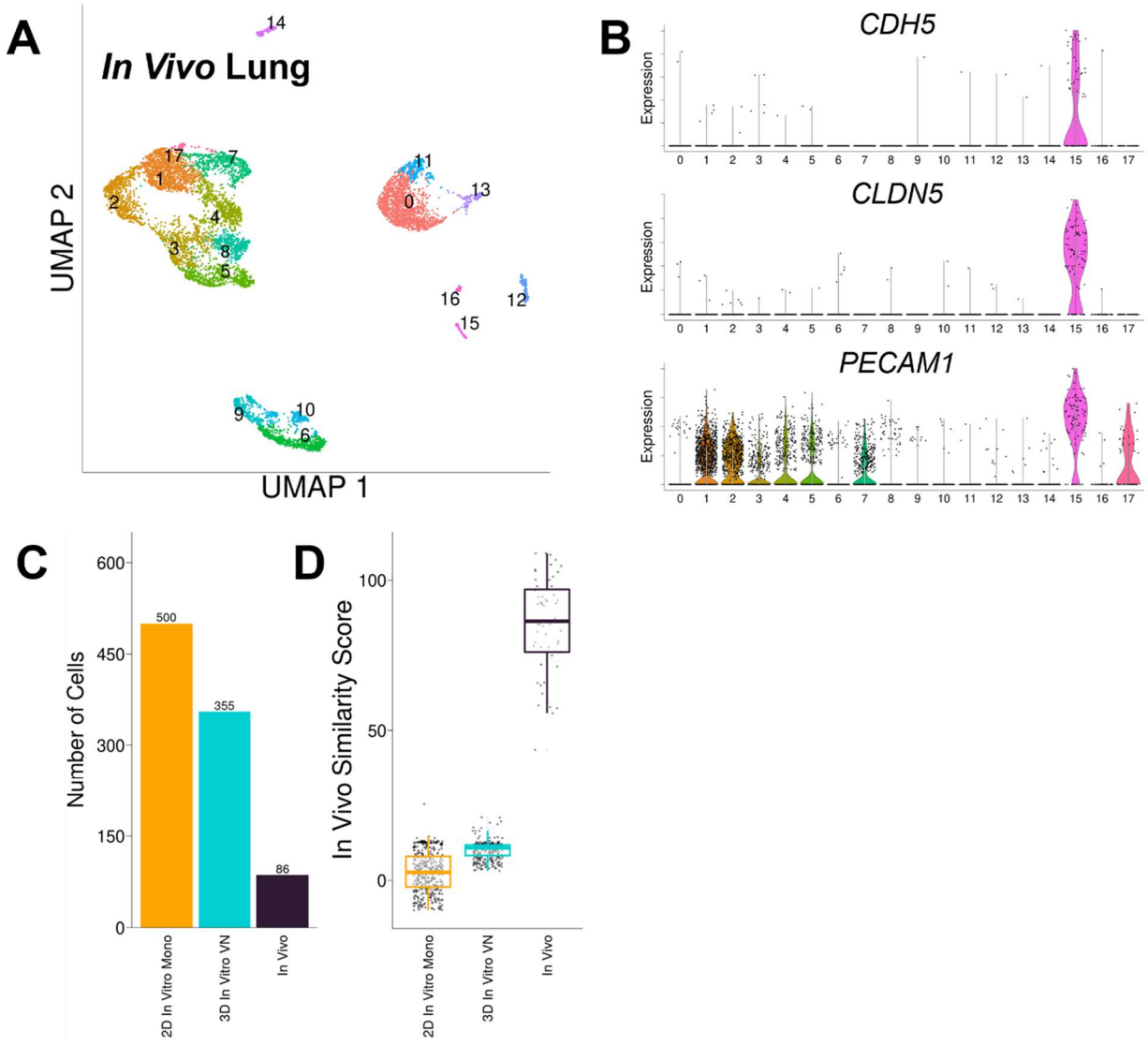

Figure 20

3D *in vitro* lung vascular network-derived ECs more mimic 2D *in vitro* monolayer ECFC-ECs compared to ECs isolated from previously collected *in vivo* lung ECs (Schupp, *et al.*). **A)** UMAP dimensionality reduction plot of whole lung dissociate *in vivo* data from Schupp, *et al*. Clusters determined using Seurat. **B)** Violin plots of EC-characteristic genes for clusters identified in the *in vivo* lung dataset. **C)** Bar chart of total number of ECs evaluated from each dataset (orange = 2D *in vitro* ECFC-EC monolayer; cyan = 3D *in vitro* EC from Skin vascular network; indigo = *in vivo* skin ECs). **D)** Normalized similarity score metric based on top 20 most differentially expressed genes from *in vivo* dataset.

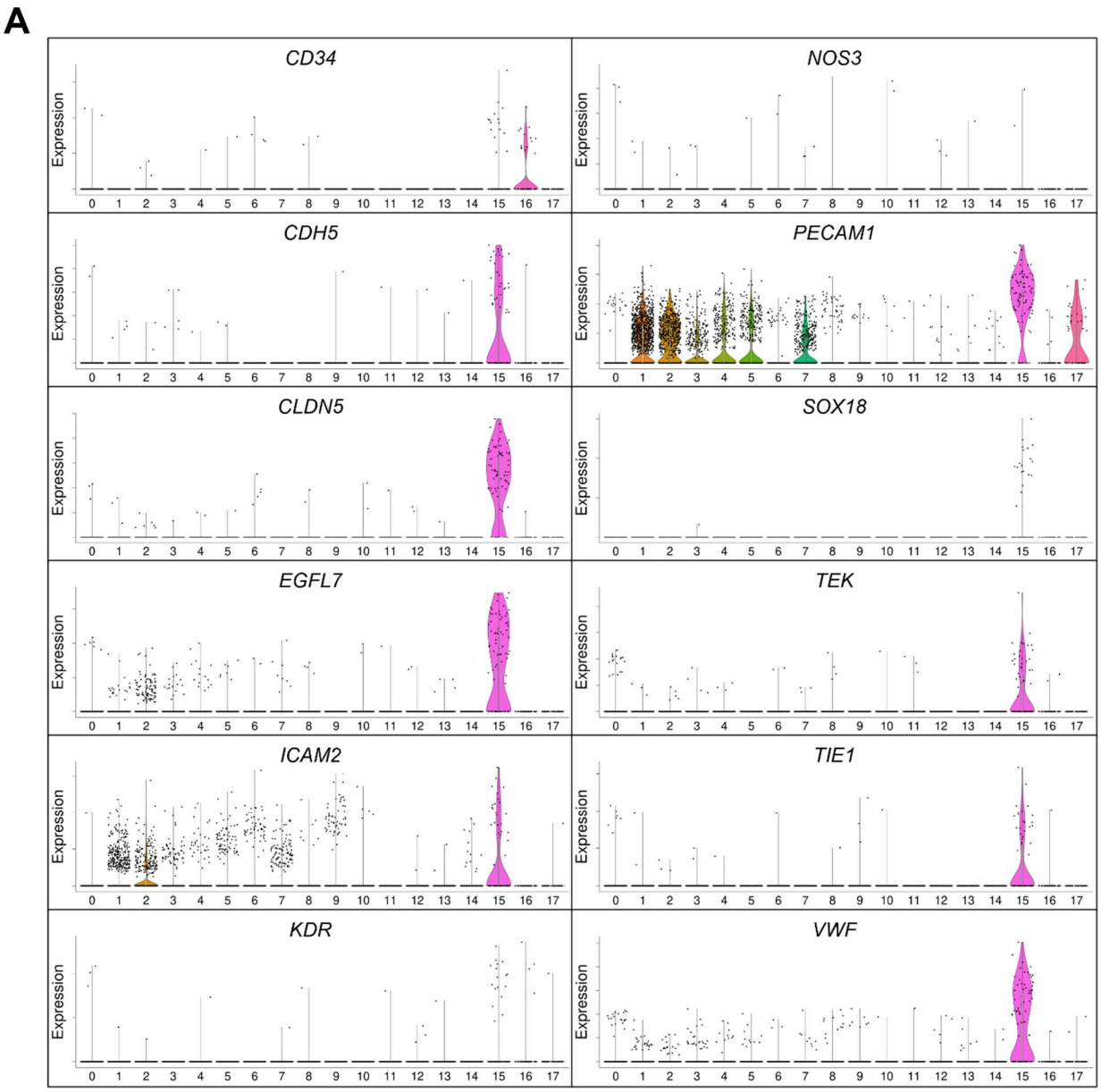

Figure 21

Additional violin plots of EC-characteristic genes from clusters identified in whole lung *in vivo* dataset (Schupp, *et al*.). **A)** Violin plots of EC-characteristic genes *CD34, CDH5, CLDN5, EGFL7, ICAM2, KDR, NOS3, PECAM1, SOX18, TEK, TIE1*, and *VWF*.

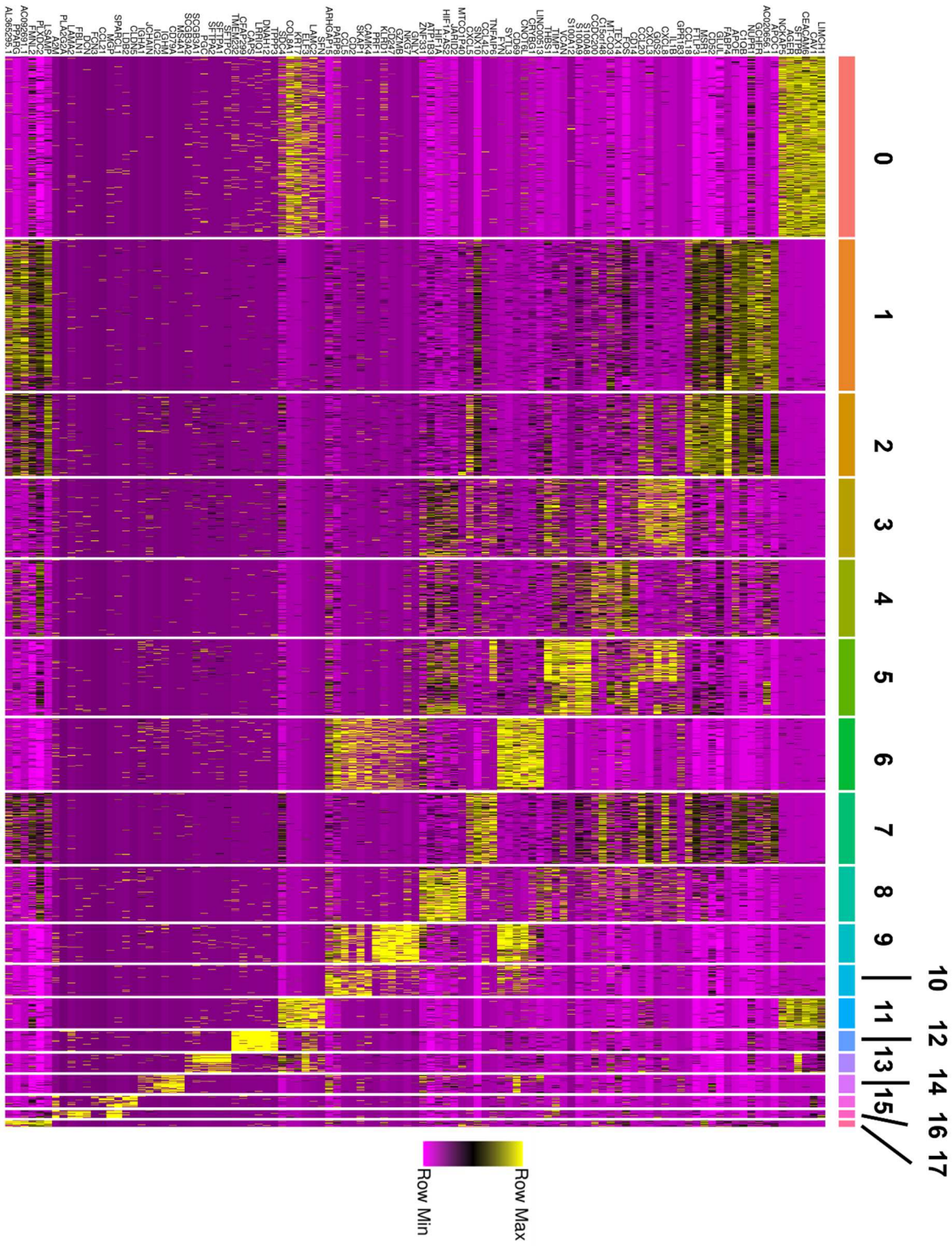

Figure 22
Heatmap of top 6 DEGs (per cluster) for *in vivo* whole lung dataset from Schupp, *et al*.

Discussion

This study provides new insight into the interplay between stromal cells and endothelial cells at the transcriptomic level in 3D *in vitro* models of organotypic vasculature. By culturing the same population of organ-agnostic ECs with organ-specific stromal cells, we demonstrate that ECs modulate their transcriptome in response to the specific stromal cell population. The altered EC transcriptome is manifested, in part, by observable phenotypic differences of the vascular network. Furthermore, the ECs present in a 3D organotypic vascular network are not transcriptionally homogenous; rather the ECs cluster into five different phenotypes whose relative proportion depends on the OSC. Finally, the presence of the OSC encourages the 3D *in vitro* vascular network-derived EC transcriptome to resemble more closely that of the matching organ-specific *in vivo* ECs compared with the same ECs cultured as a 2D *in vitro* monolayer.

Following our initial unsupervised analysis of the transcriptome of all cells, we identified four clusters as endothelial cells (clusters 2, 16, 18, 24). However, of note was the significant heterogeneity in the expression of twelve known endothelial cell genes (**Figure 7A**, **Figure 8A**) amongst these four clusters. The EC clusters were all positive for some (but not all) characteristic endothelial cell genes; but were also identified as EC by the absence of gene expression (or minimal expression) of a panel of twelve stromal cell-specific genes. Of particular note was our observation that CD31 (*PECAM1*) expression in cluster 2 was essentially undetectable despite being present by flow cytometry in 2D culture prior to 3D vascular network formation. This could be

attributed to the sequencing depth of our analysis (i.e., *PECAM1* may be expressed in the cluster, just at a lower level than other EC clusters, but higher than stromal cells), or may reflect differences in gene (single cell sequencing) and protein (flow cytometry) expression. Alternatively, quiescent EC, with stable intercellular junctions may only require a low level of CD31 expression to maintain those junctions, whereas more actively dividing and migrating cells would require a higher level of turnover.

After re-normalizing and re-clustering only the ECs from the 3D *in vitro* vascular network dataset, we observed five distinct clusters (**Figure 9**). Four of these clusters (EC-2 through EC-5) demonstrated distinct transcriptomic profiles as shown by the relative overexpression of a small number (8-10) of genes (**Figure 11**). In contrast, EC-1, the largest EC cluster in each of the six cocultures (representing approximately 50% of the cells), is not well-defined; as such, cluster EC-1 may represent a population of EC performing generic EC functions that are common to all vasculature. Cluster EC-2 is negatively correlated with vascular network length and, based on GO and pseudotime analysis, is consistent with a more stable and differentiated population of ECs involved in active protein synthesis.

Since we only analyzed a single timepoint (7 days), it is possible that the organotypic vascular networks develop at different rates, and different proportions of the EC clusters would be observed at different timepoints. Cluster EC-5 is the smallest cluster, generally representing <10% of the ECs in any of the organotypic vascular networks. It is well-defined, and is characterized by genes associated with angiogenesis, in particular extracellular matrix proteins (e.g., *COL4A2*, *COL18A1*, and

COL4A1). The presence of clusters EC-4 and EC-5 suggests that 7 days may not be adequate to achieve a truly quiescent vascular network, or that the presence of growth factors in our media allows these phenotypes to persist.

Perhaps not surprisingly (given that each vascular network was formed with a specific OSC), each 3D organotypic vascular network was associated with a population of OSCs characterized by a unique set of DEGs (e.g., OSC-3 and skin OSCs). However, surprisingly this unique cluster in each coculture was associated with GO terms consistent with extracellular matrix production and organization (with notable exception of OSC-5, the only cancer associated fibroblast) -- processes generally considered to be shared across stromal cells. As such, these subpopulations of stromal cells may contribute to the morphological differences of the organotypic vascular networks (**Figure 2**), and differences in the EC transcriptome (**Figure 9**). Furthermore, this observation suggests that each stromal cell may invoke unique gene networks to create and support organ-specific extracellular matrix. The stromal cell clusters common to all of the organotypic vascular networks (OSC-1 and OSC-7) are associated with GO terms related to protein synthesis (OSC-1) and cell proliferation (OSC-7). OSC-1 is the largest stromal cell population in each network (~25-60% of stromal cell population), but as was the case with the largest EC cluster, is not well-defined by a unique set of genes. Thus, OSC-1 may represent a stromal cell population carrying out more generic stromal cell functions that are not unique to a specific tissue.

This study raises important questions regarding how closely 3D *in vitro* vascular networks resemble *in vivo* conditions. While a wide range of OSCs support vascular

network growth in this model system, the mere presence of an OSC does not fully convert, at the transcriptional level, an otherwise naïve EC into an organ-specific EC. While we were only able to compare the transcriptome of our organotypic 3D vascular networks to two *in vivo* EC datasets (skin and lung), our results suggest that the EC in the 3D organotypic vascular network is more similar to the *in vivo* EC compared to 2D ECFC-EC monolayers cultured alone, but that the degree of similarity may depend on the organ and OSC. The emergence of publicly accessible human organ-specific transcriptomic datasets is a welcome development over the past several years [35,36,48–53]; however, these datasets remain incomplete and do not yet paint a full picture of the EC transcriptome. There are numerous murine datasets [54,55] which aim to fill in these gaps, but known differences in murine and human EC biology limit their application.

Several limitations of the 3D organotypic vascular network model system likely contribute to transcriptomic differences with *in vivo* ECs. First, the model is supported entirely through diffusion of nutrients through the tissue (no convective flow). It is well known that intraluminal physiologic shear impacts EC phenotype [56,57], and interstitial shear can also impact both endothelial and stromal cell phenotype [27,58]. Second, additional common (e.g., immune) and organ-specific (e.g., keratinocyte, cardiomyocyte) cell types are abundant in the interstitium, and are not included in our simple 3D system. These cells are capable of impacting the mechanical microenvironment and/or secreting soluble mediators that could impact vascular network formation and stability [59–61].

In summary, we created organotypic 3D vascular networks by combining a common naive EC with six different OSCs. All six OSCs supported vascular network formation, suggesting that this simple model system may be a useful and more physiologically relevant model system to investigate processes such as organ-specific angiogenesis. After 7 days, the transcriptome of the ECs in the vascular networks could be characterized by five different populations, and the relative proportion of each was dependent on the OSC. Furthermore, morphologic features of the vascular networks, such as total vessel length, correlated strongly with an EC cluster associated with protein synthesis. Each of the OSCs were associated with a unique cluster of cells whose transcriptome was associated with extracellular matrix production and organization, suggesting that these processes, common to all organs, may have organ-specific gene pathways. Finally, while the transcriptome of the ECs in the 3D organotypic vascular networks more closely resembled the transcriptome of *in vivo* ECs compared to EC in 2D monolayer, there remains a significant gap, which is likely related to factors (cells, mediators, mechanical environment) in the interstitium not present in our model system known to impact EC function.

Chapter 3 – *Ex Vivo* Model of Colorectal Cancer Liver Metastasis

Introduction

Colorectal cancer (CRC) is a leading cause of cancer-related death in the United States, with almost 150,000 new cases and 53,000 deaths estimated for 2021 alone [62]. The liver is a primary site for CRC metastasis, with about 15% of CRC patients presenting colorectal liver metastases (CRLM) at the time of CRC diagnosis, and about 50% of patients developing CRLM during the course of disease [63–65]. CRLM compromise liver function (main cause of CRC-related death), with a 5-year survival of <14% [66,67]. A small percentage of patients are eligible for liver resection, and 75% of these patients experience a recurrence in disease [68,69]. Chemotherapeutic regimens (i.e. FOLFOX, FOLFIRI) have been shown to have limited benefit for patients with inoperable, advanced CRLM or as a neoadjuvant therapy in those patients eligible for resection [70–72]. Therefore, there exists a strong need to further study the pathogenesis of CRLM in order to develop more effective treatments for CRLM.

Patient-derived organoids (PDOs) represent a significant advancement as a model system to study tumors. PDOs allow for the retention of native extracellular matrix and a more complex 3D orientation of diverse cell types, as opposed to simple 2D monolayer cell culture systems, which do not adequately recapitulate the tumor microenvironment (TuME). At the same time, PDOs do not require the same investment of time or cost associated with patient-derived xenografts or other *in vivo* animal models, making them particularly useful when studying tumor growth,

development, and therapeutic response. PDOs have been established for a variety of cancers (i.e. breast cancer [73,74], pancreatic cancer [75,76], brain cancer [77], etc.), including CRC [78–80]. CRC PDOs have previously been shown to recapitulate the genetic and proteomic landscape of their parental tumors, and can be cryopreserved for later analysis [78,79,81,82]. While there has been significant progress in the establishment and use of CRC PDO models, to date, there have been few PDO models of CRLM [83].

Recent studies of CRLM show a significant amount of cellular heterogeneity and a distinctive TME from primary CRC TME [66,83]. Single cell RNA-Seq analysis has shown that primary CRC and CRLM have been shown to differentially express genes found in the IGFBP-IGF signaling pathway (such as *IGFBP1, SPARCL1, CDH2, ITIH2, F5,* and *APOA2*) and the complement-coagulation cascade (such as *C4BPA, F5, FGA, SERPINC1, F2,* and *SERPINA5*) [66]. Additionally, studies have shown that mutations are enriched in HGF and MET signaling pathways for CRLM (as opposed to patient-matched primary CRC), which could have implications for proliferation, angiogenesis, and cell migration in CRLM [84]. Primary CRC and CRLM have also been shown to exhibit significant differences in the cellular makeup of their respective TME before and after chemotherapy, such as the presence of different tumor-associated macrophage, cancer-associated fibroblast, and B-cell populations [83]. These data suggest that the TME of primary CRC and CRLM are distinct, and as a consequence may lead to different responses to therapy.

Taken together, there exists a strong need to establish and characterize a model system of CRLM to study the pathogenesis and treatment-responsiveness of CRLM.

Here, we present in progress data from patient-matched parental CRLM and CRLM PDOs to evaluate the extent to which CRLM PDOs are able to recapitulate the TME and cellular heterogeneity of the parental CRLM. Using single cell RNA-Seq, we were able to characterize the transcriptomic landscape of patient-matched parental CRLM tumors and an *ex vivo* organoid model of CRLM. As a preliminary finding, we observe a unique population of epithelial cells that is only present in the *ex vivo* organoid model, suggesting that there is some transcriptomic drift that could occur as a result of the *ex vivo* organoid culture process.

Methods

Patient information

For this work, 2 separate patients were used (herein referred to as Patient 7 and Patient 8). Patient information is summarized in Table 6, along with the presence of mutations in either *KRAS* or *BRAF*, along with whether the cancer was microsatellite stable or not. Importantly, each of these two patients receive neoadjuvant treatment with FOLFOX, with Patient 7 receiving an additional neoadjuvant round of treatment with FOLFIRI.

Patient #	Sex	Age	KRAS	BRAF	MSI
7	M	53	WT	WT	MSS
8	M	60	WT	MUT	MSS

Table 6
Patient information for this study. WT = wild-type, MUT = mutated, MSS = microsatellite stable.

Isolation of organoids from tumor resections

Upon receipt of the patient tissue samples, samples underwent both enzymatic digestion and mechanical disruption. Briefly, the parental tumors were split into 2 equal sized pieces. 1 piece would be enzymatically digested and mechanically disrupted, before being frozen down in LN2 as the bonafide "Parental" sample. The second piece would be enzymatically digested and mechanically disrupted, before beginning low-adhesion culture to form organoid sample (as outlined below; labeled as "Organoid"). Both pieces would generally follow the same enzymatic digestion and mechanical disruption protocol. Enzymatic digestion and mechanical disruption were each achieved using a Tumor Dissociation Kit and gentleMACS system in accordance with manufacturer's protocols (Miltenyi #130-095-929).

Ex vivo culture of CRLM organoids

Ex vivo organoids were maintained in low-adhesion 6-well plates throughout the duration of culture. The media formulation used to maintain and grow these organoids is based on several previously published protocols [80,85,86]. Organoids were maintained in DMEM/F12 + GlutaMAX (ThermoFisher #10565018) supplemented with 2% StemPro (ThermoFisher #A1000601), 10% BSA (ThermoFisher #A1000801), 10 ng/mL ROCK inhibitor (STEMCELL Technologies #Y-27632) , 100 ng/mL R-Spondin-1 (PeproTech #120-38), 10 ng/mL Noggin (PeproTech #120-10C), 10 ng/mL WNT3A (R&D Systems #5036-WN-010), 10 ng/mL EGF (PeproTech #AF-100-15), 5 ng/mL IGF-1 (ThermoFisher #PHG0078), 10 ng/mL FGF10 (PeproTech #100-26), 10 ng/mL hFGF

Fibroblastic Growth Factor (Lonza #CC-4068), and 10 ng/mL Endothelin 3 (Lonza #CC-4510). Organoids were fed every 2-3 days until confluent. Organoids were then removed and centrifuged, before undergoing (1) freezing down for LN2 storage, (2) further propagation in suspension culture, or (3) processing for scRNA-Seq analysis. P7 organoids underwent one freeze-thaw cycle prior to analysis by scRNA-Seq.

Preparation of samples for scRNA-Seq

In preparation for scRNA-Seq, previously frozen vials of organoids and parental tumors were thawed. After thawing, each sample underwent enzymatic digestion with TrypLE (ThermoFisher #A1217702) supplemented with DNAse I (ThermoFisher #18047019). After enzymatic digestion, cells underwent a dead cell filtration in order to remove any dead cells or debris in the sample using a commercially available magnetic bead sorting-based Dead Cell Removal Kit (Miltenyi #130-090-101). Finally, resultant cell solutions were passed through a 40 um FlowMi filter and the cell suspensions were counted for total viable cells. This resultant solution was then passed on for cDNA library preparation.

Transcriptome alignment and initial processing

All single cell library preps and sequencing was performed by the UC Davis DNA Technologies Core and UC Davis Bioinformatics Core Facilities. A total of 3 samples were provided: Patient 7 organoid ("P7 Organoid"), Patient 7 parental ("P7 Parental"), and Patient 8 parental ("P8 Parental"). Cells were analyzed by 10X Genomics 3'

sequencing v3. Raw fastq files were processed using CellRanger count (10X Genomics) for genome alignment via the Linux command line. Resultant filtered output files were brought into the R computing environment and analyzed further via the Seurat pipeline and by using a series of scRNA-Seq analysis packages (**Table 3**). All data files underwent quality control filtering to exclude cells with fewer than 200 unique genes, greater than 7,500 unique genes, and/or more than 10% of total gene expression derived from mitochondrial-specific genes (**Figure 24A-E**). In addition, genes that were not detected in at least 3 cells were excluded from downstream analysis. This quality control process yielded several thousand cells for the parental tumor samples and several hundred cells for the organoid samples (**Figure 24A**). The following values were used for the "resolution" parameter to determine the number of clusters for the overall CRLM dataset: 0.5 for the overall parental tumor and organoid cell data (**Figure 23C**). Clustering was verified by observation and evaluation of the silhouette scores for the dataset (**Figure 24F-G**).

Gene ontology analysis

Differentially expressed genes (DEGs) were identified for all epithelial cell clusters. For a specific cluster, the top 40 DEGs along with the list of all genes expressed was fed into the topGO package in R. In order to identify the relevant over-enriched biological process gene ontology (GO) terms associated with the DEGs for each cluster, Fisher's exact test using the "elim" algorithm was performed. Resultant *p*-values were then log10-transformed and GO terms were rank ordered by the log10-transformed *p*-value.

Results

CRLM exhibits rich transcriptomic landscape and reveals subtypes of leukocytes and epithelial cells

After enzymatic digestion, the resultant samples were filtered to obtain live, single cells for scRNA-Seq (**Figure 23A**). Resultant UMAP plots demonstrate significant transcriptomic heterogeneity between the organoid and parental tumor samples, and within each individual sample (**Figure 23B**). The combined CRLM scRNA-Seq dataset underwent quality control (detailed in the Methods; **Figure 24**) and unsupervised k-means clustering as part of the Seurat pipeline. Parameter sweeps were performed to ensure an appropriate number of clusters was achieved for downstream analysis (**Figure 24F-G**). 18 distinct clusters were identified, with each cluster having a distinct transcriptomic profile (**Figure 23B-C, Figure 25**). The parental tumor samples (P7 Parental and P8 Parental) were made up of cells from all 18 clusters, whereas the organoid tumor sample (P7 Organoid) only consisted of cells from clusters 2, 7, and 13 (**Figure 23D**).

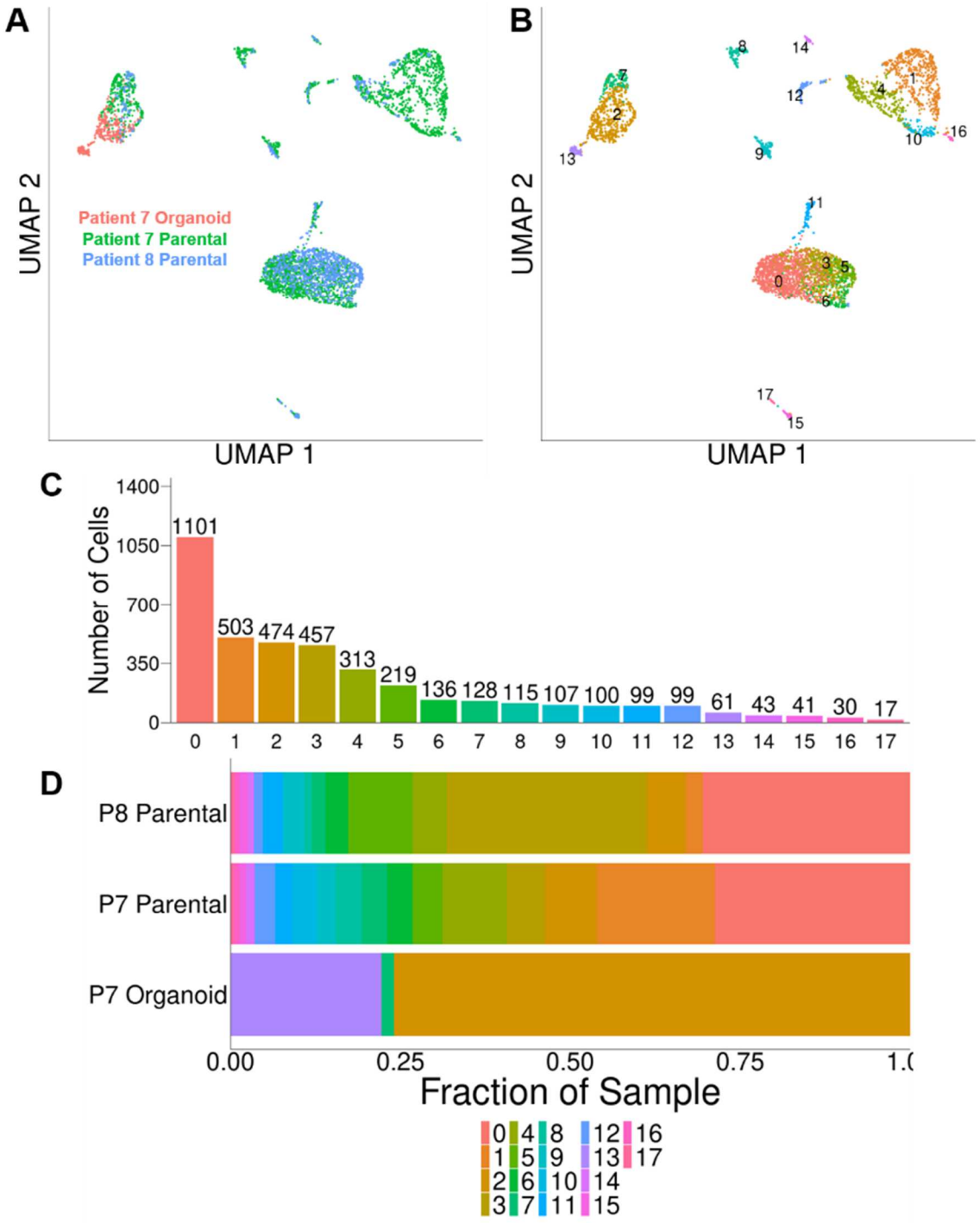

Figure 23

CRLM patient samples exhibit significant heterogeneity both within and between samples. **A)** UMAP dimensionality reduction plot grouped by sample type. **B)** UMAP dimensionality reduction plot grouped by cluster. **C)** Bar chart of total number of cells per cluster. **D)** Bar chart indicating the makeup of each CRLM sample (P7 Organoid, P7 Parental, P8 Parental) by the relative frequency of the overall CRLM clustering.

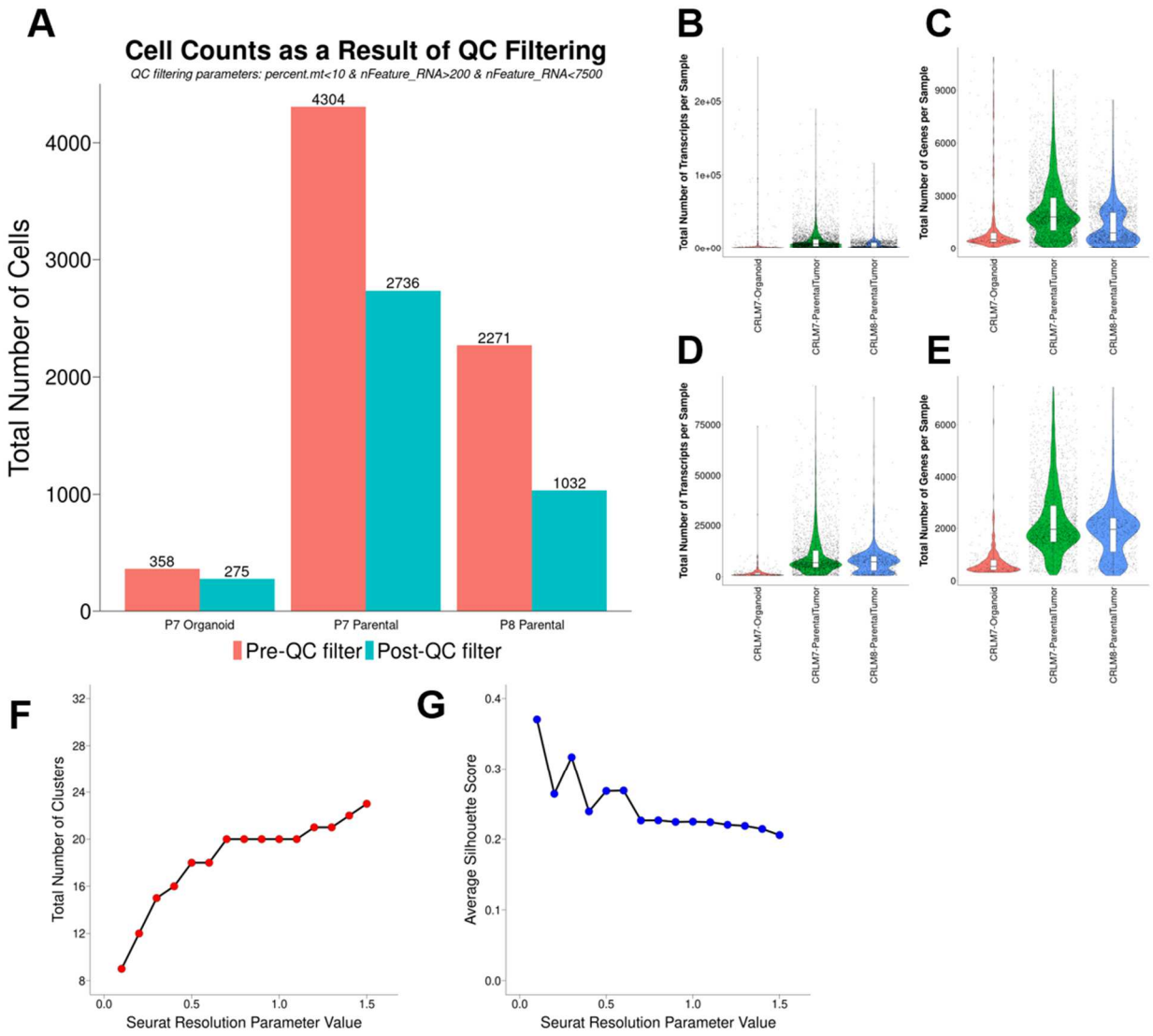

Figure 24

Quality control of complete CRLM scRNA-Seq dataset. **A)** Bar chart indicating number of cells available for analysis for each sample type before and after QC filtering as part of the scRNA-Seq analysis. **B)** Violin plot of total number of transcripts for each sample before QC filtering. **C)** Violin plot of total number of unique genes (UMIs) for each sample before QC filtering. **D)** Violin plot of total number of transcripts for each sample after QC filtering. **E)** Violin plot of total number of unique genes (UMIs) for each sample after QC filtering. **F)** Scatterplot of resultant number of cell clusters to Seurat resolution parameter (values between 0.1-1.5). **G)** Scatterplot of average silhouette score for select values of the Seurat resolution parameter (values between 0.1-1.5).

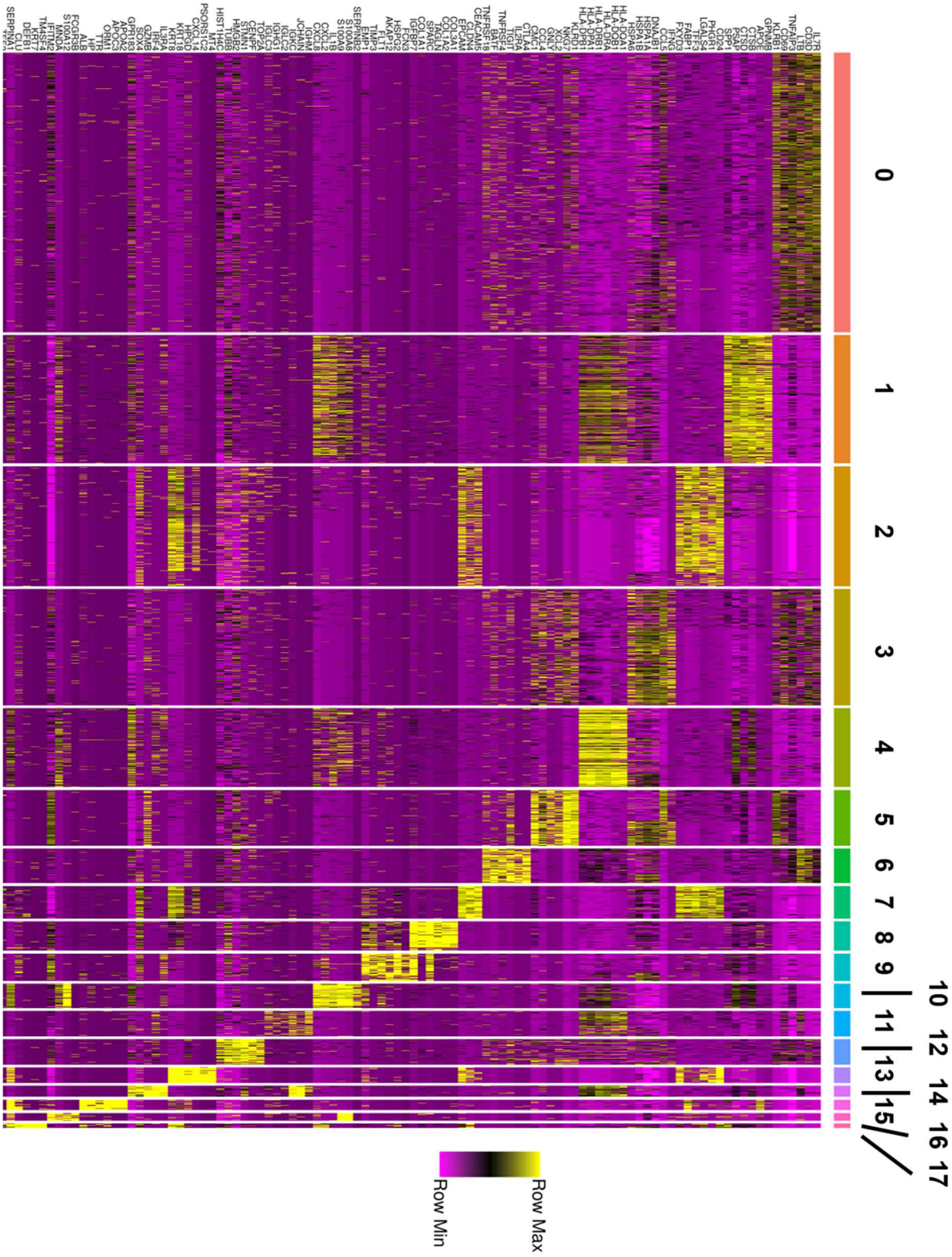

Figure 25
Heatmap of top 6 DEGs (per cluster) for CRLM dataset.

From P7 Parental and P8 Parental, it is possible to probe the immune landscape of CRLM (**Figure 26**). A sizeable proportion of cells in the CRLM scRNA-Seq dataset are *PTPRC*+ leukocytes (clusters 0, 1, 3 ,4, 5, 6, 10, 11, 12, and 14). More specifically, the dataset contains *CD3G*+ T lymphocytes (clusters 0, 3, 6, and 12), *CD3G*+*CD8A*+ cytotoxic T lymphocytes (clusters 0, 3, and 12), *CD14*+ monocytes (clusters 1, 4, and 10), *CD79A*+ B lymphocytes (cluster 11), and *FCGR3B*+ neutrophils (cluster 16). Additionally, it is possible to probe non-leukocyte populations from the P7 Parental and P8 Parental samples in this CRLM dataset (**Figure 27**). *CDH5*+*PECAM1*+ endothelial cells (cluster 9), *COL1A1*+*TAGLN*+ stromal cells (cluster 8), and *ALB*+ hepatocytes (clusters 15 and 17) are all present in the dataset. Interestingly, *EPCAM*+*CD24*+ epithelial cells separate into smaller clusters (clusters 2, 7, and 13), suggesting that these different clusters represent subtypes of epithelial cells and might perform different functions in the CRLM TuME.

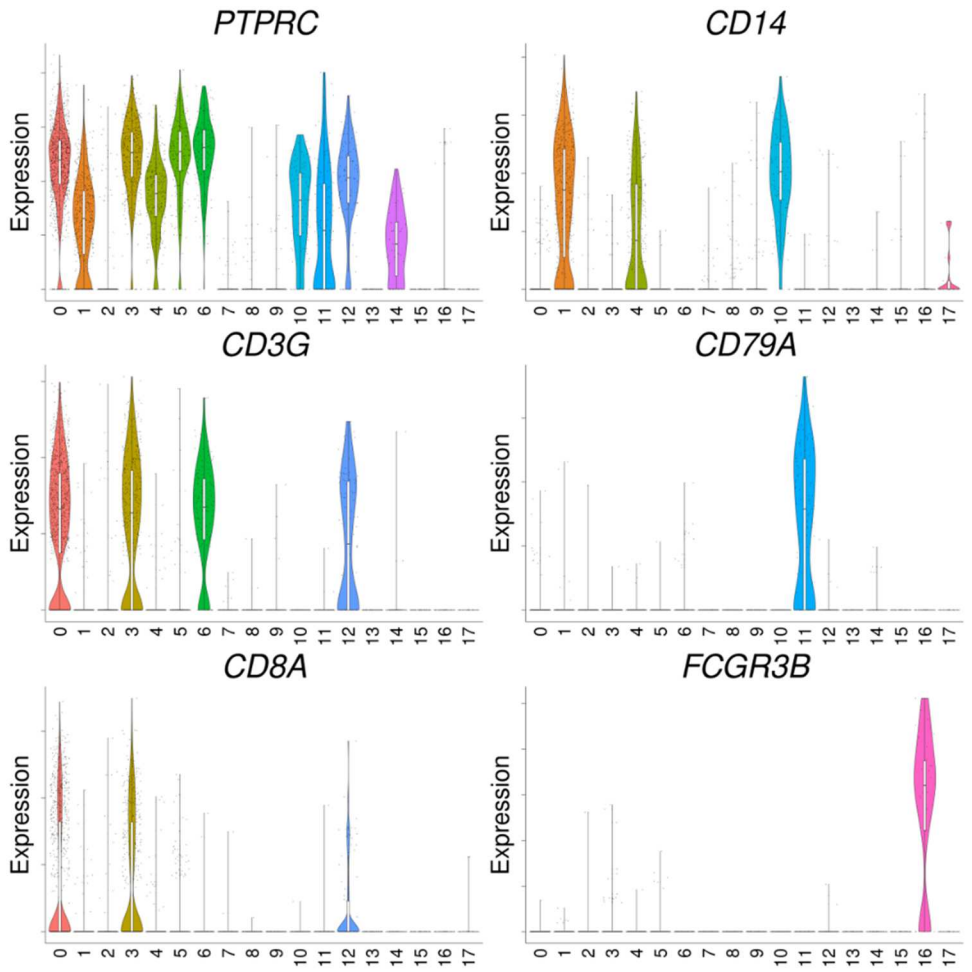

Figure 26
The immune landscape of CRLM samples is complex and consists of diverse cell types from both the innate and adaptive immune systems. Violin plots are shown for the pan-leukocyte marker *PTPRC* and leukocyte subtype-specific markers *CD3G* (T lymphocytes), *CD8A* (cytotoxic T lymphocytes), *CD14* (monocytes), *CD79A* (B lymphocytes), and *FCGR3B* (neutrophils).

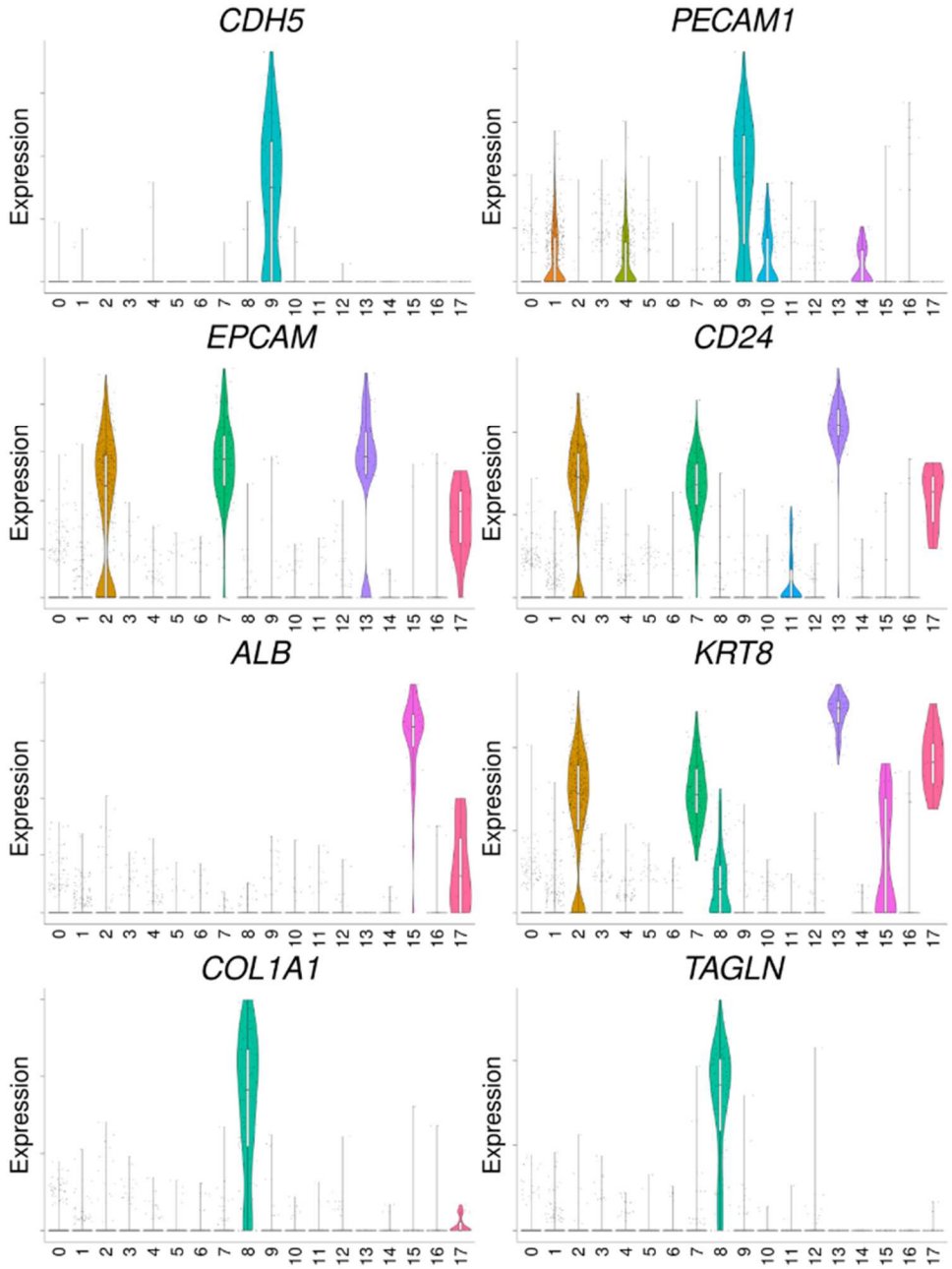

Figure 27

The non-immune landscape of CRLM samples shows cell type-specific clustering based on characteristic gene expression. Violin plots are shown for endothelial cell characteristic genes *CDH5* and *PECAM1* (cluster 9), epithelial cell characteristic genes *EPCAM*, *CD24*, and *KRT8* (clusters 2, 7, 13, and 17), hepatocyte characteristic gene *ALB* (cluster 15), and stromal cell characteristic genes *COL1A1* and *TAGLN* (cluster 8).

Ex vivo organoid model exhibits transcriptional drift from parental tumor

As stated previously, P7 Organoid consists of cells only from clusters 2, 7, and 13 (**Figure 23, Figure 27**). These clusters were previously shown to consist of *EPCAM+CD24+KRT8+* epithelial cells. To investigate the differences in the epithelial cells between the parental samples and the organoid sample, clusters 2, 7, and 13 were isolated from the dataset for closer inspection. Interestingly, the parental epithelial samples only consist of clusters 2 and 7 (~70% and ~30% for clusters 2 and 7, respectively). The organoid sample, however, has several marked differences to the parental sample. P7 Organoid consists of 76% cluster 2, 2% cluster 7, and 22% cluster 13 (**Figure 28**).

In order to better understand each of the epithelial cell clusters, GO pathway analysis was performed on the top 40 DEGs for each cluster (**Table 6**). Cluster 2 is characterized by DEGs that include cell-cell and cell-ECM interactions (e.g. *LGALS4*), and is associated with migration and adhesion based on GO terms including “negative regulation of anoikis”, “positive regulation of cell migration”, and “calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules” (**Table 7**). Cluster 7 is characterized by DEGs that include mucosal layer maintenance (e.g. *TFF3*, *MUC13*), and is associated with cell maintenance and several signaling pathways based on GO terms including “maintenance of gastrointestinal epithelium” and “C-type lectin receptor signaling” (**Table 8**). Cluster 13 is characterized by DEGs that include DNA damage detection/repair and intermediate filament anchors (e.g. *SFN*, *DSP*), and is associated with cytoskeletal organization and cell death pathways based on GO terms including

"cornification", "desmosome organization", and "hepatocyte apoptotic process" (**Table 9**).

#	Cluster 2	Cluster 7	Cluster 13
1	*CD24*	*TRIM31*	*MT4*
2	*PHGR1*	*CYP3A5*	*PSORS1C2*
3	*KRT8*	*MUC13*	*KRT31*
4	*KRT19*	*KRT20*	*S100A3*
5	*KRT18*	*CEACAM5*	*CXCL14*
6	*EPCAM*	*GPRC5A*	*C6orf15*
7	*S100P*	*CLDN7*	*KRT80*
8	*LY6G6F-LY6G6D*	*ELF3*	*HPGD*
9	*GPX2*	*CEACAM6*	*SFN*
10	*LGALS4*	*C19orf33*	*LY6G6F-LY6G6D*
11	*TFF3*	*SMIM22*	*DPEP1*
12	*TSPAN8*	*RND3*	*KRT19*
13	*CEACAM5*	*CLDN3*	*APCDD1*
14	*FABP1*	*EDN1*	*KRT17*
15	*FXYD3*	*LMO7*	*TPD52L1*
16	*SMIM22*	*MUC12*	*CCND1*
17	*FERMT1*	*LSR*	*TSPAN1*
18	*IFI27*	*GDF15*	*KRT18*
19	*PERP*	*AGR2*	*KRT8*
20	*CLDN3*	*CLDN4*	*CD24*
21	*S100A6*	*TFF1*	*DSP*
22	*CEACAM6*	*IFI27*	*SCD*
23	*ADIRF*	*EPCAM*	*SLC26A3*
24	*ASCL2*	*TSPAN8*	*CKB*
25	*SPINK1*	*SPINK1*	*S100P*
26	*C19orf33*	*PHGR1*	*SLPI*
27	*AGR2*	*KRT19*	*TMEM47*
28	*MT1G*	*FXYD3*	*EPCAM*
29	*CKB*	*CKB*	*AL445524.1*
30	*ELF3*	*KRT8*	*TSPAN8*
31	*CLDN4*	*CD24*	*FXYD3*
32	*CCND1*	*TFF3*	*PERP*
33	*TFF1*	*KRT18*	*CST3*
34	*TXN*	*SPINT2*	*CAB39L*
35	*ZFAS1*	*RRBP1*	*PCCA*
36	*DSTN*	*SOX4*	*ADIRF*
37	*PPDPF*	*FABP1*	*S100A14*
38	*SPINT2*	*PMEPA1*	*CAMK2N1*
39	*PRDX5*	*LGALS3*	*JUP*
40	*EIF2S2*	*S100A6*	*HEBP2*

Table 7
Top 40 DEGs for each epithelial cell cluster from the CRLM dataset, ordered by fold change over all other clusters in the combined CRLM dataset (as determined by Seurat). Top 6 DEGs for each cluster are shown in Figure 25.

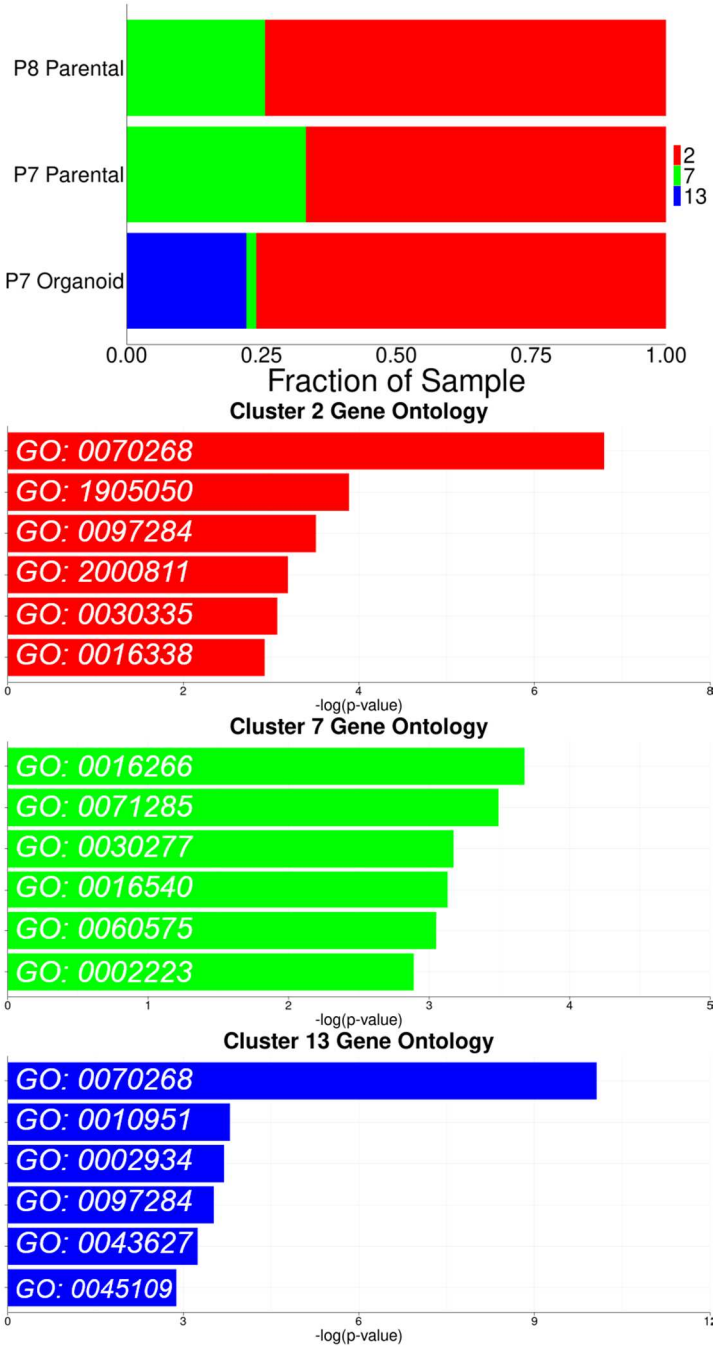

Figure 28
Closer look at the epithelial cell composition of P7 Parental, P8 Parental, and P7 Organoid (simplified view of Figure 23D). GO plots correspond with GO terms from Tables 7-9.

-log(p-value)	GO #	Term
6.80	0070268	Cornification
3.89	1905050	Positive regulation of metallopeptidase activity
3.51	0097284	Hepatocyte apoptotic process
3.19	2000811	Negative regulation of anoikis
3.07	0030335	Positive regulation of cell migration
2.93	0016338	Calcium-independent cell-cell adhesion via plasma membrane cell-adhesion molecules

Table 8
Top 6 GO terms for cluster 2 of the CRLM dataset based on top 40 DEGs.

-log(p-value)	GO #	Term
3.68	0016266	O-glycan processing
3.49	0071285	Cellular response to lithium ion
3.17	0030277	Maintenance of gastrointestinal epithelium
3.13	0016540	Protein autoprocessing
3.05	0060575	Intestinal epithelial cell differentiation
2.89	0002223	Stimulatory C-type lectin receptor signaling pathway

Table 9
Top 6 GO terms for cluster 7 of the CRLM dataset based on top 40 DEGs.

-log(p-value)	GO #	Term
10.06	0070268	Cornification
3.80	0010951	Negative regulation of endopeptidase activity
3.70	0002934	Desmosome organization
3.52	0097284	Hepatocyte apoptotic process
3.24	0043627	Response to estrogen
2.88	0045109	Intermediate filament organization

Table 10
Top 6 GO terms for cluster 13 of the CRLM dataset based on top 40 DEGs.

Epithelial cells in CRLM dataset exhibit heterogeneity of drug transporter and type I/II intermediate filament gene expression, but this heterogeneity is attenuated in cancer stem cell gene expression

In order to better understand practical limitations to using the *ex vivo* organoid as an *in vitro* model system to study CRLM, we looked more closely at three distinct classes of genes known to impact CRC and CRLM. Specifically, we analyzed the expression of relevant drug transporter genes [87–89], type I and type II intermediate filament genes [90–92], and cancer stem cell-related genes [93,94]. Analysis of the epithelial cells in the CRLM dataset (clusters 2, 7, and 13) revealed that several known drug and solute transporters (*ABCA1, ABCA2, ABCB1, ABCC3, ABCG1,* and *POU2F1*) appear to be consistently expressed in cluster 7, but not in clusters 2 and 13 (**Figure 29**). Additionally, further analysis of the epithelial cells revealed that several keratin-related genes (*KRT8, KRT17, KRT18, KRT19,* and *KRT31*), representing a subset of type I and type II intermediate filaments, are upregulated in cluster 13 over cluster 2 and 7 (**Figure 30**). However, there are other keratins (*KRT20*) which appear to be absent from cluster 13, while having significant expression in clusters 2 and 7. Lastly, we investigated the expression of known cancer stem cell-related genes in the epithelial cells of the CRLM dataset. Analysis of the epithelial cells revealed that several cancer stem cell-related genes (*BMI1, CD200, CLU, LGR5, PROM1*) are not expressed in any cluster, while other genes (*ANXA1, CD44, EPHB2*) have heterogeneous expression between the epithelial cell clusters (**Figure 31**).

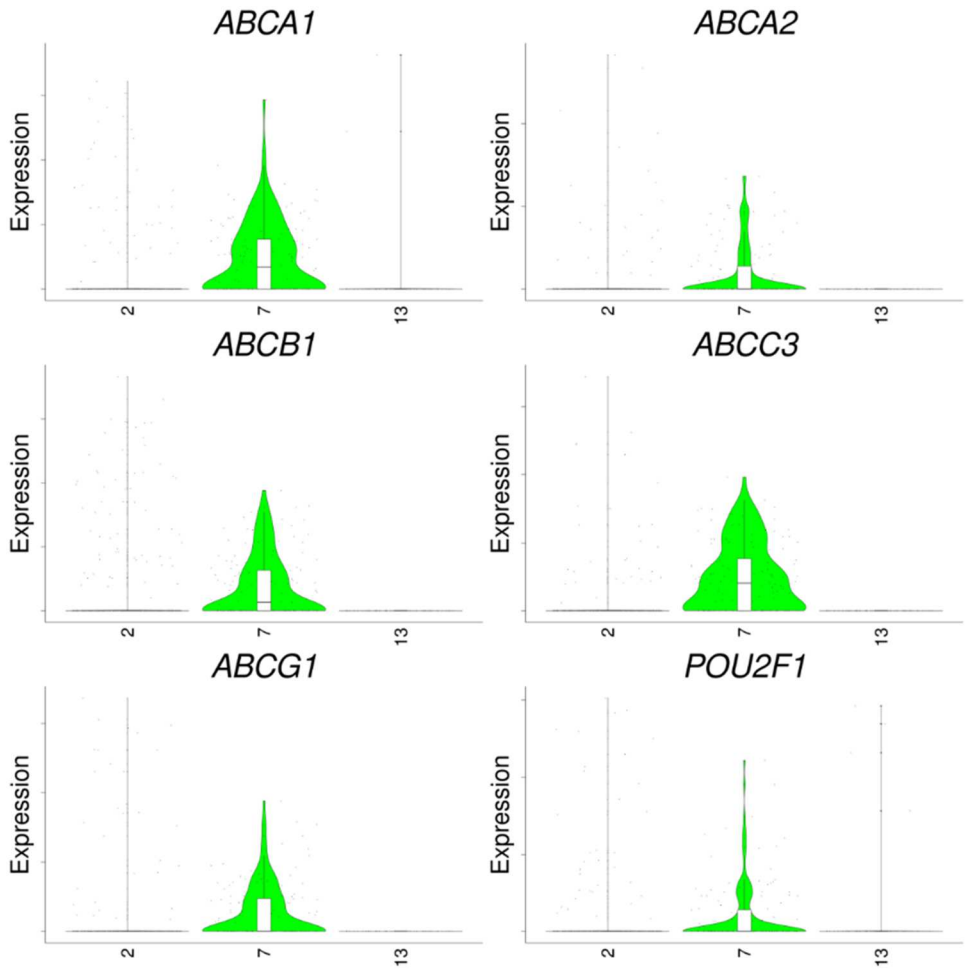

Figure 29
Expression of genes associated with known drug and solute transporters in CRLM: *ABCA1, ABCA2, ABCB1, ABCC3, ABCG1,* and *POU2F1*.

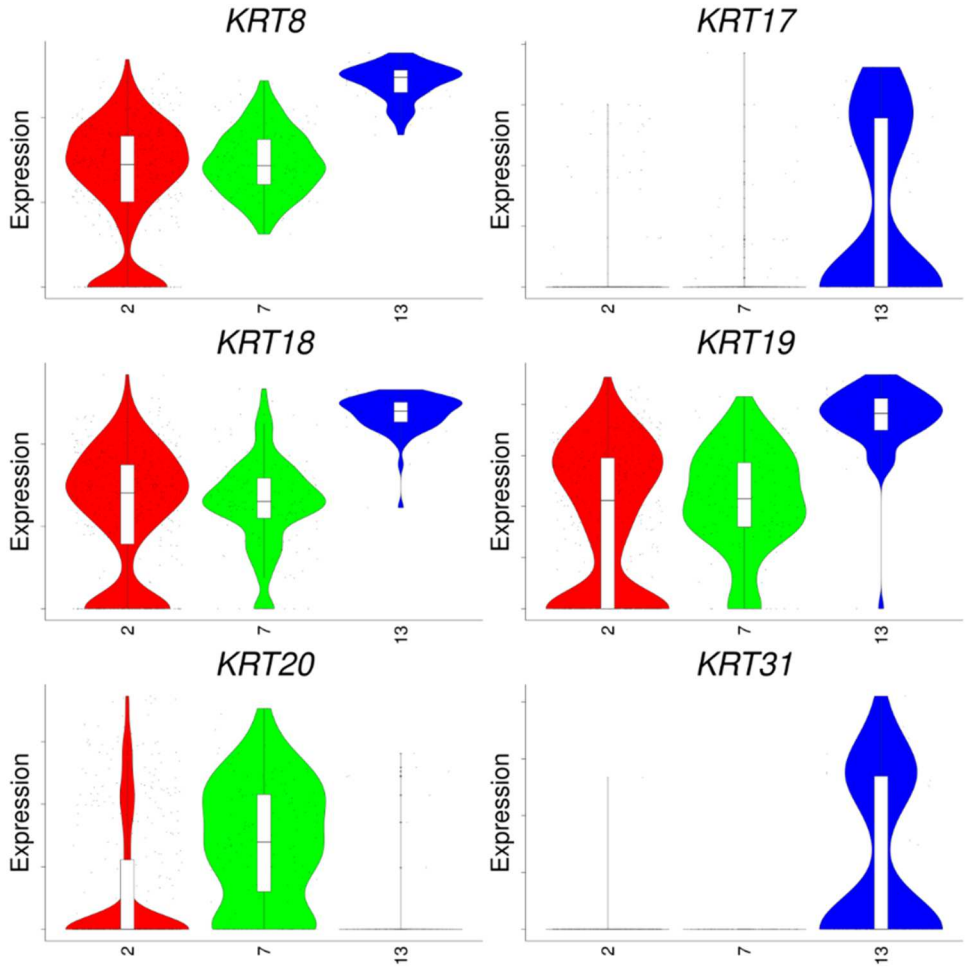

Figure 30
Expression of keratins (both type I and type II intermediate filaments) is heterogeneous between the epithelial clusters. Genes shown are *KRT8, KRT17, KRT18, KRT19, KRT20,* and *KRT31*.

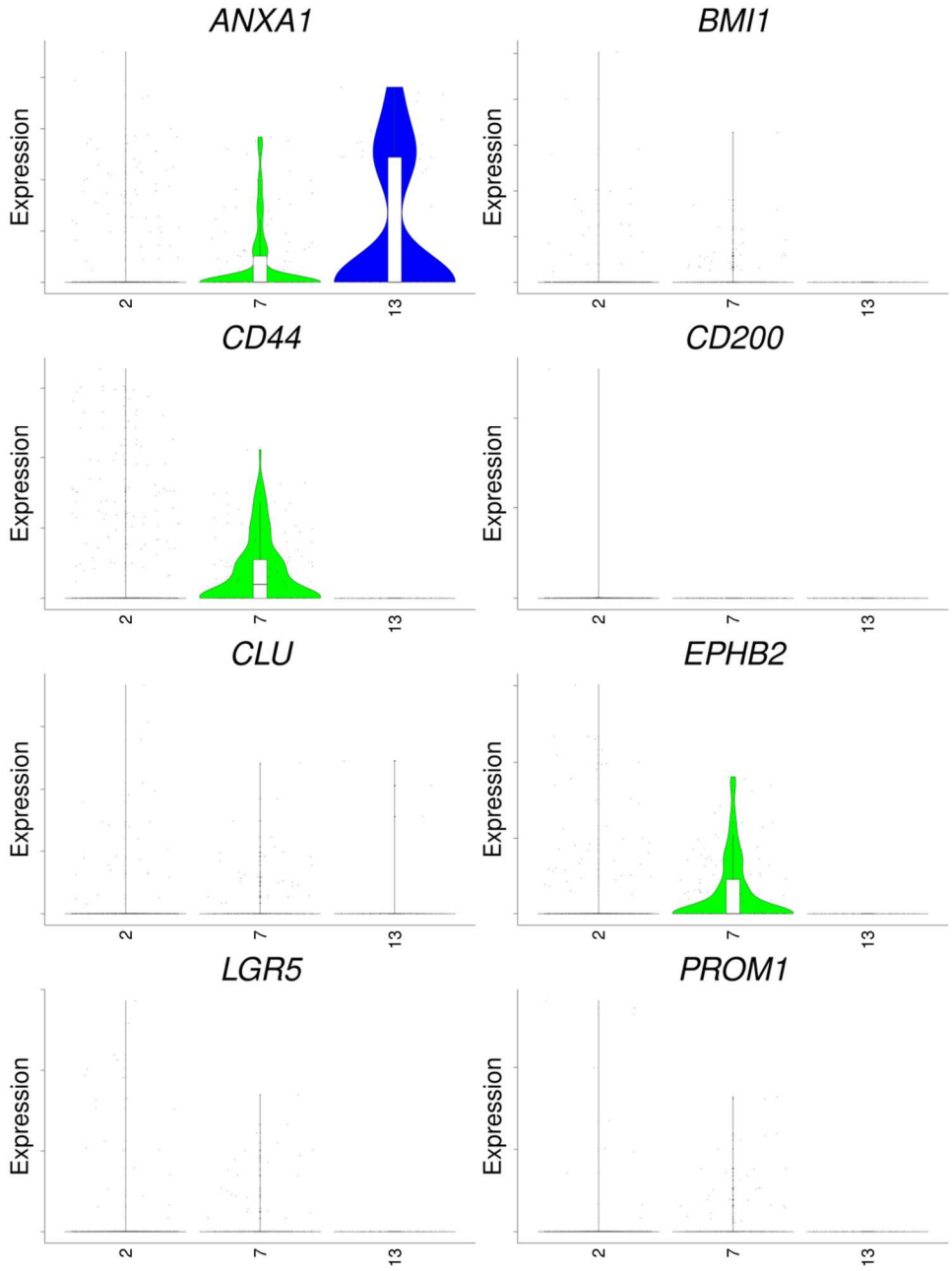

Figure 31
Expression of cancer stem cell-related genes is generally not found in epithelial cells from CRLM dataset. Genes shown are *ANXA1, BMI1, CD44, CD200, CLU, EPHB2, LGR5,* and *PROM1*.

Discussion

This study provides new insight into the transcriptomic heterogeneity in parental CRLM tumors and the transcriptomic drift that may emerge as a consequence of the *ex vivo* organoid culture of these tumors. The parental CRLM tumor is richly diverse, consisting of a wide range of cell types including a wide range of leukocytes, endothelial cells, stromal cells, epithelial cells, and hepatocytes. While we only analyzed a single patient's *ex vivo* organoid sample, the resultant *ex vivo* organoid model of CRLM exhibited less cellular diversity and are enriched for epithelial cells. These organoid epithelial cells also demonstrate transcriptomic drift from the parental tumor epithelial cells. This transcriptomic drift is manifested, in part, by alterations in the relative frequency of clusters of epithelial cells, which may have consequences on the expression of drug transporter-related genes and intermediate filament-related genes. This transcriptomic drift may also affect the expression of cancer stem cell-related genes, although this drift appears to be more muted.

The parental tumor samples (P7 Parental and P8 Parental) exhibit significant cellular heterogeneity and primarily consist of leukocytes (**Figure 23C**). From the analysis presented here, there are multiple subpopulations of T lymphocytes (clusters 0, 3, 6, and 12) and monocytes (clusters 1, 4, and 10). These cell populations are each implicated in the progression and therapeutic response of CRLM, and have recently become the subject of intense investigation by multiple groups [84,95–97]. In this work, we limit ourselves to investigating the differences between the parental tumor and the *ex vivo* organoid model, and as such the immune landscape of the parental CRLM, while

interesting, falls outside the scope of our study. This data does suggest, however, that there is ample opportunity for future work to study the immune landscape of CRLM.

The epithelial cells in this dataset separated into 3 distinct clusters (**Figure 27**). Interestingly, the relative percentage of each epithelial cell cluster is not equivalent across the different samples (P7 Parental, P8 Parental, P7 Organoid). An organoid-specific population of epithelial cells (cluster 13) makes up > 20% of the total cells from P7 Organoid. These cells appear to have elevated expression of some type I and type II intermediate filament genes (**Figure 30**; keratins *KRT8, KRT17, KRT18, KRT19, KRT20,* and *KRT31*). Various cytokeratins (CK7, CK8, CK18, CK20) have previously been shown to be clinically relevant in identifying CRLM and subtyping primary CRC, and might serve as potential prognostic indicators of disease outcome [91,92,98]. The fact that cluster 13 is organoid-specific and has an elevated level of keratin gene expression may suggest that there are clinically relevant differences between the parental tumor and organoid model. Additionally, the P7 Organoid sample has a markedly reduced amount of cluster 7 epithelial cells compared to the P7 Parental and P8 Parental samples (2% for P7 Organoid, ~30% for P7 Parental and P8 Parental). These cells appear to be the only epithelial cells with some expression of drug transporter genes (**Figure 29**; *ABCA1, ABCA2, ABCB1, ABCC3, ABCG1, POU2F1*). That these cells appear to be underrepresented in the P7 Organoid sample may suggest that the ability of *ex vivo* organoids to respond to therapeutic challenge may be altered as a consequence of the culture process, although this requires additional patient samples to reach a definitive conclusion. Additionally, more work is needed to evaluate the drug responsiveness of the *ex vivo* organoid and to compare that response to *in vivo* data in

order to see if the observed drug transporter gene changes affect phenotypic responses of these two systems. Finally, cancer stem cell-related genes (**Figure 31**; *ANXA1, BMI1, CD44, CD200, CLU, EPHB2, LGR5, PROM1*) do not appear to be consistently expressed in any epithelial cell sub-population. While this may be due to the limited sample size, it is an interesting result as some *ex vivo* organoid models of primary CRC have previously been shown to upregulate expression of cancer stem cell-related genes [99,100]. The resultant changes in the relative percentage of epithelial cell sub-populations between the parental and organoid samples, taken with the differences in gene expression between these sub-populations, suggest that transcriptomic drift is occurring as a consequence of the *ex vivo* organoid culture. Importantly, this drift may not be visible to the same extent for specific families of genes (i.e. drug transporter genes and cancer stem cell-related genes). More patient parental and organoid samples are required in order to see whether these observations represent a broader trend in the transcriptomic drift, or a patient-specific phenomenon.

Several limitations of the organoid model system likely contribute to transcriptomic differences observed between the parental *in vivo* tumor and the organoid *ex vivo* model. First, the parental tumor samples have a wider cellular diversity, likely reflecting a more complete recapitulation of the TuME compared to the organoid sample. The organoid sample only contains epithelial cells, which is likely a consequence of the extended *in vitro* culture process prior to scRNA-Seq analysis. As outlined in the Methods, the parental tumor samples were enzymatically digested and mechanically disrupted, which yields many single cells that are frozen until scRNA-Seq analysis. However, the organoid sample then underwent additional weeks of

suspension culture in order to generate a substantial number of organoids for downstream analysis. The duration of the *ex vivo* culture, along with the media formulation used, likely enriches for epithelial cells. Further optimization of the culture conditions is likely required in order to retain other cell types. Additionally, the data presented here consists of 2 parental samples and 1 organoid sample. More work is needed to expand this dataset to verify that the observed transcriptomic differences are not merely unique to this single organoid sample, but may instead be a more widely observed phenomenon.

In summary, we characterized a patient-matched *ex vivo* organoid model of CRLM using scRNA-Seq. Parental CRLM samples exhibited rich cellular heterogeneity, while the organoid sample consisted of exclusively epithelial cells. 3 sub-populations of epithelial cells were identified in this dataset, and each of these sub-populations had distinct DEG and GO pathway profiles. Additionally, several classes of genes appeared to be upregulated in these epithelial cell sub-populations (drug transporters in cluster 7, keratins in cluster 13). Finally, while the transcriptomes between the 3 epithelial cell clusters shared some similarities, there were distinct differences in the relative percentage and DEGs for these epithelial cell sub-populations, highlighting transcriptomic drift that occurs over the course of *ex vivo* organoid culture.

Chapter 4 – Concluding Remarks

Single cell RNA-Seq is an incredibly powerful technique to understand and measure heterogeneity within cellular populations. scRNA-Seq allows for the identification of potentially rare sub-populations of cells within a tissue and (with the explosion in the number of publicly available *in vivo* scRNA-Seq datasets) is emerging as a more and more effective tool to understand differences between cellular samples.

Presented here in this dissertation is an examination of the transcriptomes of two different *in vitro* model systems and a comparison of the transcriptomes of these *in vitro* systems to relevant *in vivo* transcriptomes. In the case of the 3D *in vitro* vascular networks, we have demonstrated that commercially available organotypic stromal cells induce both phenotypic and transcriptomic changes in endothelial cells during the course of 7 days of co-culture in fibrin hydrogels (using the same parental endothelial cell for all samples). These phenotypic changes can be seen by the difference in the robustness of vascular network formation, with some stromal cells being quite adept at supporting and maintaining vascular networks, while others are considerably less capable. Likewise, organotypic stromal cells induce transcriptomic changes in the endothelial cells, with different amounts of 5 sub-populations of endothelial cells being present in each vascular network. Of particular note, we showed that a sub-population of transcriptionally distinct endothelial cells (expressing genes related to protein translation and ribosomal assembly) is strongly negatively correlated to quantitative metrics of the vascular network phenotype. While more work is needed to understand and validate the consequences of these transcriptomic changes on the phenotypes of the vascular networks, we believe that this work raises important questions about the

ability to model organ-specific functions of vasculature and the plasticity of the endothelium. In the case of the CRLM (caveat that the dataset is still in progress), we have demonstrated that an *ex vivo* organoid model appears to enrich for epithelial cells. Additionally, we observe that the organoids undergo some transcriptomic drift from the parental samples. This transcriptomic drift can be observed in the changes of the relative frequency of 3 sub-populations of epithelial cells in our samples. Genes related to drug and solute transporters, along with genes related to type I and type II intermediate filaments (keratins), appear to be differentially expressed between the sub-populations of epithelial cells. However, this drift does not appear to affect all clinically relevant panels of genes (at least not with this limited dataset). For instance, the epithelial cell sub-populations appear to have very minor differences in terms of their expression of cancer stem cell-related genes. More work is needed to acquire additional samples for analysis. This will allow for a better understanding of the magnitude of this transcriptomic drift as a consequence of the *ex vivo* organoid culture.

The overall goal of this dissertation was to utilize a burgeoning technology to better understand the types of cellular heterogeneity that may emerge in *in vitro* model systems, and to what extent that heterogeneity matches what is seen *in vivo* ("transcriptomic drift"). For 2 disparate model systems (organ-specific capillary networks and *ex vivo* colorectal cancer liver metastasis organoids), we observe noticeable differences between the *in vitro* and *in vivo* systems. Further work is required to better understand these differences, in order to more effectively utilize these systems.

Chapter 5 – Code for Analysis of 3D *In Vitro* Vascular Networks

Introduction

Below is a subset of the various code that was utilized for the 3D *in vitro* vascular network work described in Chapter 2. Version properties of the software are indicated in Table 3. Note that the early steps in the R computing environment code roughly follow the suggested pipeline for Seurat as previously described [41,101].

Linux Computing Environment

The code below allows for the use of CellRanger (10X Genomics) to align the raw fastq files obtained from an Illumina NovaSeq and then count the number of hits for each unique gene/molecular identifier using the human GRCh38-3.0.0 reference transcriptome [102].

```
cd ~/yard/scrna_analysis/apps/cellranger-5.0.0

export PATH=/home/user/scrna_analysis/apps/cellranger-5.0.0:$PATH

which cellranger

cd ~/yard/user/name

mkdir output

cd ~/yard/user/name/output
```

```
cellranger count --id=UNIQUENAME --fastqs=/home/user/name/RAWDATA --
      sample=SAMPLE --transcriptome=/home/user/scrna_analysis/refdata-
      cellranger-GRCh38-3.0.0
```

R Computing Environment

The code below describes the initial processing steps for the aligned output data along with the initial quality control filtering of the scRNA-Seq data.

```
library(Seurat)
library(ggplot2)
library(cowplot)
library(dplyr)
library(patchwork)
bmsc.vn.data <- Read10X(data.dir = “ #directory to Marrow 3D in vitro
      vascular network data in filtered_feature_bc_matrix folder ”)
nhlf.vn.data <- Read10X(data.dir = “ #directory to Lung 3D in vitro vascular
      network data in filtered_feature_bc_matrix folder ”)
nhdf.vn.data <- Read10X(data.dir = “ #directory to Skin 3D in vitro vascular
      network data in filtered_feature_bc_matrix folder ”)
```

```
caf.vn.data <- Read10X(data.dir = “ #directory to cPancreas 3D in vitro
      vascular network data in filtered_feature_bc_matrix folder ”)

hcf.vn.data <- Read10X(data.dir = “ #directory to Heart 3D in vitro vascular
      network data in filtered_feature_bc_matrix folder ”)

hpastec.vn.data <- Read10X(data.dir = “ #directory to nPancreas in vitro
      vascular network data in filtered_feature_bc_matrix folder ”)

bmsc.vn <- CreateSeuratObject(counts=bmsc.vn.data, project=“BMSC vascular
      network”, min.cells=3, min.features=200)

nhlf.vn <- CreateSeuratObject(counts=nhlf.vn.data, project=“NHLF vascular
      network”, min.cells=3, min.features=200)

nhdf.vn <- CreateSeuratObject(counts=nhdf.vn.data, project=“NHDF vascular
      network”, min.cells=3, min.features=200)

caf.vn <- CreateSeuratObject(counts=caf.vn.data, project=“CAF08 vascular
      network”, min.cells=3, min.features=200)

hcf.vn <- CreateSeuratObject(counts=hcf.vn.data, project=“HCF vascular
      network”, min.cells=3, min.features=200)

hpastec.vn <- CreateSeuratObject(counts=hpastec.vn.data, project=“HPaSteC
      vascular network”, min.cells=3, min.features=200)

vncomb1 <- merge(bmsc.vn, y=c(nhlf.vn, nhdf.vn, caf.vn, hcf.vn, hpastec.vn),
      add.cell.ids=c(“BMSC vascular network”, “NHLF vascular network”, “NHDF
      vascular network”, “CAF08 vascular network”, “HCF vascular network”,
      “HPaSteC vascular network”), project=“Combined Vascular Network Data”)
```

```
vncomb1$orig.ident <- Idents(vncomb1)

l.abbr <- vector(mode="character", length=length(vncomb1$orig.ident))

for (i in 1:length(vncomb1$orig.ident)) {

if (vncomb1$orig.ident[i]=="BMSC vascular network") {

l.abbr[i] <- "Marrow" }

else if (vncomb1$orig.ident[i]=="CAF08 vascular network") {

l.abbr[i] <- "cPancreas" }

else if (vncomb1$orig.ident[i]=="HCF vascular network") {

l.abbr[i] <- "Heart" }

else if (vncomb1$orig.ident[i]=="HPaSteC vascular network") {

l.abbr[i] <- "nPancreas" }

else if (vncomb1$orig.ident[i]=="NHDF vascular network") {

l.abbr[i] <- "Skin" }

else if (vncomb1$orig.ident[i]=="NHLF vascular network") {

l.abbr[i] <- "Lung" }

}

lev.fancy <- c("Marrow", "Heart", "Skin", "Lung", "cPancreas", "nPancreas")
```

```
lev <- c("BMSC vascular network", "HCF vascular network", "NHDF vascular
      network", "NHLF vascular network", "CAF08 vascular network", "HPaSteC
      vascular network")

vncomb1$fancy.orig.ident <- l.abbr

vncomb1$fancy.orig.ident <- factor(vncomb1$fancy.orig.ident,
      levels=lev.fancy)

vncomb1$orig.ident <- factor(vncomb1$orig.ident, levels=lev)

vncomb1[["percent.mt"]] <- PercentageFeatureSet(vncomb1, pattern="^MT-")

vncomb1[[“percent.ribo”]] <- PercentageFeatureSet(vncomb1,
      pattern=”^RP[SL][[:digit:]]|^RPLP[[:digit:]]^RPSA”)

vncomb2 <- subset(vncomb1, subset=percent.mt<10 & nFeature_RNA<7500 &
      nFeature_RNA>200) # filtering such that any cell with >10% of total
      genes as mitochondrial genes (non-informative genes), or >7500 unique
      genes (potentially multiple cells grouped together), or <200 unique
      genes (potentially poor/dying/dead cells or debris) are excluded from
      downstream analysis.
```

The next portion of code describes the normalization, dimensionality reduction, and clustering of the combined 3D *in vitro* vascular network object. Additionally, this code describes obtaining the silhouette scores for the sweep of the Seurat resolution parameter in order to evaluate the clustering fit.

```
vncomb2 <- NormalizeData(vncomb2)

vncomb2 <- FindVariableFeatures(vncomb2, selection.method="vst",
      nfeatures=2000) # standard Seurat settings

vncomb2 <- ScaleData(vncomb2, features=rownames(vncomb2)) # standard Seurat
      settings

vncomb2 <- RunPCA(vncomb2, features=VariableFeatures(vncomb2)) # standard
      Seurat settings

vncomb2 <- FindNeighbors(vncomb2, dims=1:15) # standard Seurat settings;
      looking at first 15 PCs

vncomb2 <- FindClusters(vncomb2, resolution=1) # standard Seurat settings;
      increasing resolution increases number of clusters

vncomb2 <- RunUMAP(vncomb2, dims=1:15) # standard Seurat settings; looking at
      first 15 PCs

vncomb2 <- RunTSNE(vncomb2, dims=1:15) # standard Seurat settings; looking at
      first 15 PCs
```

```
Idents(vncomb2) <- vncomb2$seurat_clusters

library(cluster)

dist.mat <- dist(Embeddings(vncomb2[['pca']])[,1:15])

res <- c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4,
      1.5)

vncomb2 <- FindClusters(vncomb2, resolution=res)

vncomb2.sum <- data.frame(matrix(NA_real_, nrow=length(res), ncol=4))

colnames(vncomb2.sum) <-
      c("resolution","total.num.clusters","avg.sil.score","sd.sil.score")

for (i in 1:15) {

vncomb2 <- FindClusters(vncomb2, resolution=res[i])

clust <- vncomb2@meta.data$seurat_clusters

sil <- silhouette(as.numeric(clust), dist=dist.mat)

name.sil <- paste0("silhouette_score_",res[i])

vncomb2[[name.sil]] <- sil[,3]

vncomb2.sum[i,1] <- res[i]

vncomb2.sum[i,2] <-
      as.numeric(last(rownames(table(vncomb2@meta.data$seurat_clusters)))) +
      1

vncomb2.sum[i,3] <- mean(sil[,3])
```

```
vncomb2.sum[i,4] <- sd(sil[,3])
print(i)
}
```

The next portion of code describes the re-normalization, dimensionality reduction, and re-clustering of the ECs identified from the combined 3D *in vitro* vascular network object. This code also describes obtaining the silhouette scores for the sweep of the Seurat resolution parameter in order to evaluate the clustering fit. Additionally, this code describes the identification of the top DEGs for each EC cluster. This code also describes determining the significant GO terms based on DEGs for each cluster. This portion of code is almost identical for both the re-normalized EC and OSC data.

```
vncomb2.ec.renorm <- subset(vncomb2, idents=c("2","16","18","24"))
      #determined by gene expression pattern of characteristic EC and OSC
      genes.
vncomb2.ec.renorm <- NormalizeData(vncomb2.ec.renorm)
vncomb2.ec.renorm <- FindVariableFeatures(vncomb2.ec.renorm,
      selection.method="vst", nfeatures=2000) # standard Seurat settings
vncomb2.ec.renorm <- ScaleData(vncomb2.ec.renorm,
      features=rownames(vncomb2.ec.renorm)) # standard Seurat settings
```

```
vncomb2.ec.renorm <- RunPCA(vncomb2.ec.renorm,
      features=VariableFeatures(vncomb2.ec.renorm)) # standard Seurat
      settings
vncomb2.ec.renorm <- FindNeighbors(vncomb2.ec.renorm, dims=1:15) # standard
      Seurat settings; looking at first 15 PCs
vncomb2.ec.renorm <- FindClusters(vncomb2.ec.renorm, resolution=0.3) #
      standard Seurat settings; increasing resolution increases number of
      clusters
vncomb2.ec.renorm <- RunUMAP(vncomb2.ec.renorm, dims=1:15) # standard Seurat
      settings; looking at first 15 PCs
vncomb2.ec.renorm <- RunTSNE(vncomb2.ec.renorm, dims=1:15) # standard Seurat
      settings; looking at first 15 PCs


Idents(vncomb2.ec.renorm) <- vncomb2.ec.renorm$seurat_clusters
dist.mat <- dist(Embeddings(vncomb2.ec.renorm[['pca']])[,1:15])
res <- c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4,
      1.5)
vncomb2.ec.renorm <- FindClusters(vncomb2.ec.renorm, resolution=res)
ec.sum <- data.frame(matrix(NA_real_, nrow=length(res), ncol=4))
colnames(ec.sum) <-
      c("resolution","total.num.clusters","avg.sil.score","sd.sil.score")
```

```
for (i in 1:15) {
vncomb2.ec.renorm <- FindClusters(vncomb2.ec.renorm, resolution=res[i])
clust <- vncomb2.ec.renorm@meta.data$seurat_clusters
sil <- silhouette(as.numeric(clust), dist=dist.mat)
name.sil <- paste0("silhouette_score_",res[i])
vncomb2.ec.renorm[[name.sil]] <- sil[,3]
ec.sum[i,1] <- res[i]
ec.sum[i,2] <-
      as.numeric(last(rownames(table(vncomb2.ec.renorm@meta.data$seurat_clust
      ers)))) + 1
ec.sum[i,3] <- mean(sil[,3])
ec.sum[i,4] <- sd(sil[,3])
print(i)
}
Idents(vncomb2.ec.renorm) <- vncomb2.ec.renorm$seurat_clusters
library(stringr)
l.abbr <- vector(mode="character",
      length=length(vncomb2.ec.renorm$orig.ident))
for (i in 1:length(vncomb2.ec.renorm$seurat_clusters)) {
if (vncomb2.ec.renorm$seurat_clusters[i]=="0") {
```

```
l.abbr[i] <- "EC-1" }
else if (vncomb2.ec.renorm$seurat_clusters[i]=="1") {
l.abbr[i] <- "EC-2" }
else if (vncomb2.ec.renorm$seurat_clusters[i]=="2") {
l.abbr[i] <- "EC-3" }
else if (vncomb2.ec.renorm$seurat_clusters[i]=="3") {
l.abbr[i] <- "EC-4" }
else if (vncomb2.ec.renorm$seurat_clusters[i]=="4") {
l.abbr[i] <- "EC-5" }
}
lev.ec <- c("EC-1", "EC-2", "EC-3", "EC-4", "EC-5")
vncomb2.ec.renorm$ec.class <- l.abbr
vncomb2.ec.renorm$ec.class <- factor(vncomb2.ec.renorm$ec.class,
      levels=lev.ec)


Idents(vncomb2.ec.renorm) <- vncomb2.ec.renorm$ec.class
ec.mark <- FindAllMarkers(vncomb2.ec.renorm, only.pos=TRUE, min.pct=0.25,
      logfc.threshold=0.25)
top10.ec <- ec.mark %>% group_by(cluster) %>% top_n(n=10, wt=avg_log2FC)
```

```
library(topGO)

library("org.Hs.eg.db")

Idents(vncomb2.ec.renorm) <- vncomb2.ec.renorm$ec.class

all.genes <- rownames(GetAssayData(vncomb2.ec.renorm,slot="counts")) #
      provide list of all genes in vncomb2.ec.renorm Seurat object

ec1.mark <- ec.mark[ec.mark$cluster=="EC-1" & ec.mark$p_val_adj<0.05,]$gene

geneList.ec1 <- ifelse(all.genes %in% ec1.mark, 1, 0) # generate logical
      vector to see which genes are differentially expressed for cluster EC-1

names(geneList.ec1) <- all.genes

ec1.godata <- new("topGOdata", ontology="BP", allGenes=geneList.ec1,
      geneSelectionFun=function(x)(x==1), annot=annFUN.org,
      mapping="org.Hs.eg.db", ID="symbol", nodeSize=5) # creating the
      topGOdata object

rf.ec1 <- runTest(ec1.godata, algorith="elim", statistic="fisher") #
      implementing the elimination algorithm to acquire a Fisher p-value (F-
      test based value)

ec1.res1 <- data.frame(GenTable(ec1.godata,Fisher=rf.ec1,
      topNodes=25,numChar=60))

ec1.res2 <- ec1.res1[,c('GO.ID','Term','Fisher')]

ec1.res3 <- ec1.res2[1:25,]
```

```
ec1.res3$neglogPval <- -log10(as.numeric(ec1.res3$Fisher)) # take the -log10
      of the Fisher p-value reported (will use this in future graphs)
ec1.res3[is.na(ec1.res3)] <- 30 # in the event there is no data, set the
      value to 30 (will exclude if is the case)
ec1.res3$abbrTerm <- rbind(str_trunc(ec1.res3$Term, 38, "right"))[1,]
ec2.mark <- ec.mark[ec.mark$cluster=="EC-2" & ec.mark$p_val_adj<0.05,]$gene
geneList.ec2 <- ifelse(all.genes %in% ec2.mark, 1, 0) # generate logical
      vector to see which genes are differentially expressed for cluster EC-2
names(geneList.ec2) <- all.genes
ec2.godata <- new("topGOdata", ontology="BP", allGenes=geneList.ec2,
      geneSelectionFun=function(x)(x==1), annot=annFUN.org,
      mapping="org.Hs.eg.db", ID="symbol", nodeSize=2) # creating the
      topGOdata object
rf.ec2 <- runTest(ec2.godata, algorith="elim", statistic="fisher") #
      implementing the elimination algorithm to acquire a Fisher p-value (F-
      test based value)
ec2.res1 <- data.frame(GenTable(ec2.godata,Fisher=rf.ec2,
      topNodes=25,numChar=60))
ec2.res2 <- ec2.res1[,c('GO.ID','Term','Fisher')]
ec2.res3 <- ec2.res2[1:25,]
```

```
ec2.res3$neglogPval <- -log10(as.numeric(ec2.res3$Fisher)) # take the -log10
      of the Fisher p-value reported (will use this in future graphs)
ec2.res3[is.na(ec2.res3)] <- 30 # in the event there is no data, set the
      value to 30 (will exclude if is the case)
ec2.res3$abbrTerm <- rbind(str_trunc(ec2.res3$Term, 38, "right"))[1,]
ec3.mark <- ec.mark[ec.mark$cluster=="EC-3" & ec.mark$p_val_adj<0.05,]$gene
geneList.ec3 <- ifelse(all.genes %in% ec3.mark, 1, 0) # generate logical
      vector to see which genes are differentially expressed for cluster EC-1
names(geneList.ec3) <- all.genes
ec3.godata <- new("topGOdata", ontology="BP", allGenes=geneList.ec3,
      geneSelectionFun=function(x)(x==1), annot=annFUN.org,
      mapping="org.Hs.eg.db", ID="symbol", nodeSize=5) # creating the
      topGOdata object
rf.ec3 <- runTest(ec3.godata, algorith="elim", statistic="fisher") #
      implementing the elimination algorithm to acquire a Fisher p-value (F-
      test based value)
ec3.res1 <- data.frame(GenTable(ec3.godata,Fisher=rf.ec3,
      topNodes=25,numChar=60))
ec3.res2 <- ec3.res1[,c('GO.ID','Term','Fisher')]
ec3.res3 <- ec3.res2[1:25,]
```

```
ec3.res3$neglogPval <- -log10(as.numeric(ec3.res3$Fisher)) # take the -log10
      of the Fisher p-value reported (will use this in future graphs)
ec3.res3[is.na(ec3.res3)] <- 30 # in the event there is no data, set the
      value to 30 (will exclude if is the case)
ec3.res3$abbrTerm <- rbind(str_trunc(ec3.res3$Term, 38, "right"))[1,]
ec4.mark <- ec.mark[ec.mark$cluster=="EC-4" & ec.mark$p_val_adj<0.05,]$gene
geneList.ec4 <- ifelse(all.genes %in% ec4.mark, 1, 0) # generate logical
      vector to see which genes are differentially expressed for cluster EC-2
names(geneList.ec4) <- all.genes
ec4.godata <- new("topGOdata", ontology="BP", allGenes=geneList.ec4,
      geneSelectionFun=function(x)(x==1), annot=annFUN.org,
      mapping="org.Hs.eg.db", ID="symbol", nodeSize=5) # creating the
      topGOdata object
rf.ec4 <- runTest(ec4.godata, algorith="elim", statistic="fisher") #
      implementing the elimination algorithm to acquire a Fisher p-value (F-
      test based value)
ec4.res1 <- data.frame(GenTable(ec4.godata,Fisher=rf.ec4,
      topNodes=25,numChar=60))
ec4.res2 <- ec4.res1[,c('GO.ID','Term','Fisher')]
ec4.res3 <- ec4.res2[1:25,]
```

```
ec4.res3$neglogPval <- -log10(as.numeric(ec4.res3$Fisher)) # take the -log10
      of the Fisher p-value reported (will use this in future graphs)

ec4.res3[is.na(ec4.res3)] <- 30 # in the event there is no data, set the
      value to 30 (will exclude if is the case)

ec4.res3$abbrTerm <- rbind(str_trunc(ec4.res3$Term, 38, "right"))[1,]

ec5.mark <- ec.mark[ec.mark$cluster=="EC-5" & ec.mark$p_val_adj<0.05,]$gene

geneList.ec5 <- ifelse(all.genes %in% ec5.mark, 1, 0) # generate logical
      vector to see which genes are differentially expressed for cluster EC-2

names(geneList.ec5) <- all.genes

ec5.godata <- new("topGOdata", ontology="BP", allGenes=geneList.ec5,
      geneSelectionFun=function(x)(x==1), annot=annFUN.org,
      mapping="org.Hs.eg.db", ID="symbol", nodeSize=5) # creating the
      topGOdata object

rf.ec5 <- runTest(ec5.godata, algorith="elim", statistic="fisher") #
      implementing the elimination algorithm to acquire a Fisher p-value (F-
      test based value)

ec5.res1 <- data.frame(GenTable(ec5.godata,Fisher=rf.ec5,
      topNodes=25,numChar=60))

ec5.res2 <- ec5.res1[,c('GO.ID','Term','Fisher')]

ec5.res3 <- ec5.res2[1:25,]

ec5.res3$Place <- rownames(ec5.res3)
```

```
ec5.res3$neglogPval <- -log10(as.numeric(ec5.res3$Fisher)) # take the -log10
      of the Fisher p-value reported (will use this in future graphs)
ec5.res3[is.na(ec5.res3)] <- 30 # in the event there is no data, set the
      value to 30 (will exclude if is the case)
ec5.res3$abbrTerm <- rbind(str_trunc(ec5.res3$Term, 38, "right"))[1,]
```

This next portion of code describes the pseudotime value determination and mapping using Monocle.

```
library(monocle3)
library(SeuratWrappers)
ec.cds <- as.cell_data_set(vncomb2.ec.renorm)
ec.cds <- cluster_cells(ec.cds, resolution=1e-3)
reducedDim(ec.cds, type="PCA") <-
      vncomb2.ec.renorm@reductions$pca@cell.embeddings
ec.cds@int_colData@listData$reducedDims$UMAP <-
      vncomb2.ec.renorm$umap@cell.embeddings
ec.cds@clusters$UMAP$clusters <- vncomb2.ec.renorm@meta.data$seurat_clusters
names(ec.cds@clusters$UMAP$clusters) <- rownames(vncomb2.ec.renorm@meta.data)
ec.cds <- learn_graph(ec.cds)
```

```
ec.cds <- order_cells(ec.cds)

vncomb2.ec.renorm$pseudotime <- pseudotime(ec.cds)
```

The next portion of code describes the integration of the skin *in vivo* single cell RNA-Seq data from Solé-Boldo, *et al* [36]. Additionally, this code describes the scoring of the 2D *in vitro* ECFC-EC monolayer, Skin 3D *in vitro* vascular network-derived ECs, and *in vivo* skin ECs. This code is almost identical to the code used for the comparison to whole lung data presented in Chapter 2.

```
library(Matrix)

library(readr)

library(Seurat)

setwd("/home/user/soleboldo-skin/pull_from_geo")

counts <- readMM("GSE130973_matrix_filtered.mtx.gz")

genes <- read_tsv("GSE130973_genes_filtered.tsv.gz", col_names=FALSE)

gene_ids <- genes$X2

cell_ids <- read_tsv("GSE130973_barcodes_filtered.tsv.gz",
      col_names=FALSE)$X1

rownames(counts) <- gene_ids

colnames(counts) <- cell_ids
```

```
skin.obj <- CreateSeuratObject(counts)
skin.obj[["percent.mt"]] <- PercentageFeatureSet(skin.obj, pattern="^MT-")

skin.obj.f <- subset(skin.obj, subset=percent.mt<10 & nFeature_RNA<7500 &
      nFeature_RNA>200)
skin.obj.f <- NormalizeData(skin.obj.f)
skin.obj.f <- FindVariableFeatures(skin.obj.f, selection.method="vst",
      nfeatures=2000)
skin.obj.f <- ScaleData(skin.obj.f, features=rownames(skin.obj.f))
skin.obj.f <- RunPCA(skin.obj.f, features=VariableFeatures(skin.obj.f))
skin.obj.f <- FindNeighbors(skin.obj.f, dims=1:15)
skin.obj.f <- FindClusters(skin.obj.f, resolution=0.5)
skin.obj.f <- RunUMAP(skin.obj.f, dims=1:15)
skin.obj.f <- RunTSNE(skin.obj.f, dims=1:15)
Idents(skin.obj.f) <- skin.obj.f$seurat_clusters
skin.mark <- FindAllMarkers(skin.obj.f, only.pos=F, min.pct=0.25,
      logfc.threshold=0.25)
top10.skin <- skin.mark %>% group_by(cluster) %>% top_n(n=10, wt=avg_log2FC)
ec.mono[["percent.mt"]] <- PercentageFeatureSet(ec.mono, pattern="^MT-")
```

```
ec.mono.qc <- subset(ec.mono, subset=percent.mt<10 & nFeature_RNA<7500 &
      nFeature_RNA>200)
ec.mono.qc$vv <- "ECFC Monolayer"
ec.mono.qc$fancy.vv <- "2D In Vitro Mono"
nhdf.obj <- subset(vncomb2.ec.renorm, idents="NHDF VN")
nhdf.obj$vv <- "In Vitro VN"
nhdf.obj$fancy.vv <- "3D In Vitro VN"
skin.obj.f$vv <- "In Vivo"
skin.obj.f$fancy.vv <- "In Vivo"
options(future.globals.maxSize = 1000 * 1024^2)
l.anchor1 <- FindIntegrationAnchors(object.list=list(ec.mono.qc.small,
      nhdf.obj, skin.obj.f.sub1), dims=1:20)
skin.comb.int1 <- IntegrateData(anchorset = l.anchor1, dims=1:20,
      k.weight=50)
Idents(skin.comb.int1) <- skin.comb.int1$vv
DefaultAssay(skin.comb.int1) <- "RNA"
skin1.markers <- FindAllMarkers(skin.comb.int1, only.pos=TRUE, min.pct=0.25,
      logfc.threshold=0.25)
top20.skin1 <- skin1.markers %>% group_by(cluster) %>% top_n(n=20,
      wt=avg_log2FC)
```

```
skin.vivo <- top20.skin1$gene[top20.skin1$cluster=="In Vivo" &
      top20.skin1$p_val_adj<0.05]
DefaultAssay(skin.comb.int1) <- "RNA"
skin.comb.int1 <- AddModuleScore(skin.comb.int1, features=list(skin.vivo),
      name="vivo_enriched")
x.mono <-
      mean(skin.comb.int1@meta.data[skin.comb.int1@meta.data$fancy.vv=="2D In
      Vitro Mono",]$vivo_enriched1)
x.vivo <-
      mean(skin.comb.int1@meta.data[skin.comb.int1@meta.data$fancy.vv=="In
      Vivo",]$vivo_enriched1)
skin.comb.int1$scale.vivo.module.score <- (skin.comb.int1$vivo_enriched1 -
      x.mono) / (x.vivo - x.mono) *100
```

Chapter 6 – Code for Analysis of *Ex Vivo* CRLM Organoids

Introduction

Below is a subset of the various code that was utilized for the *ex vivo* organoid model of CRLM work described in Chapter 3. Version properties of the software are indicated in Table 3. Note that the early steps in the R computing environment code roughly follow the suggested pipeline for Seurat as previously described [41,101].

Linux Computing Environment

The code below allows for the use of CellRanger (10X Genomics) to align the raw fastq files obtained from an Illumina NovaSeq and then count the number of hits for each unique gene/molecular identifier using the human GRCh38-3.0.0 reference transcriptome [102].

```
cd ~/yard/scrna_analysis/apps/cellranger-5.0.0

export PATH=/home/user/scrna_analysis/apps/cellranger-5.0.0:$PATH

which cellranger

cd ~/yard/user/name

mkdir output

cd ~/yard/user/name/output
```

```
cellranger count --id=UNIQUENAME --fastqs=/home/user/name/RAWDATA --
      sample=SAMPLE --transcriptome=/home/user/scrna_analysis/refdata-
      cellranger-GRCh38-3.0.0
```

R Computing Environment

The code below describes the initial processing steps for the aligned output data along with the initial quality control filtering of the scRNA-Seq data.

```
library(Seurat)
library(ggplot2)
library(cowplot)
library(dplyr)
library(patchwork)
library(tidyverse)
crlm7.sc.data <- Read10X(file.path(cellranger7.sc.dir))
crlm7.org.data <- Read10X(file.path(cellranger7.org.dir))
crlm8.sc.data <- Read10X(file.path(cellranger8.sc.dir))
```

```
crlm7.sc <- CreateSeuratObject(counts=crlm7.sc.data, project="CRLM7-
      ParentalTumor")

crlm7.org <- CreateSeuratObject(counts=crlm7.org.data, project="CRLM7-
      Organoid")

crlm8.sc <- CreateSeuratObject(counts=crlm8.sc.data, project="CRLM8-
      ParentalTumor")

crlm78 <- merge(crlm7.sc, y=c(crlm7.org, crlm8.sc), add.cell.ids=c("CRLM7-
      ParentalTumor", "CRLM7-Organoid", "CRLM8-ParentalTumor"))


crlm78[["percent.mt"]] <- PercentageFeatureSet(crlm78, pattern="^MT-")

crlm78.qc <- subset(crlm78, subset = nFeature_RNA>200 & nFeature_RNA<7500 &
      percent.mt<10) # filtering such that any cell with >10% of total genes
      as mitochondrial genes (non-informative genes), or >7500 unique genes
      (potentially multiple cells grouped together), or <200 unique genes
      (potentially poor/dying/dead cells or debris) are excluded from
      downstream analysis.
```

The next portion of code describes the normalization, dimensionality reduction, and clustering of the combined CRLM object. This code also describes obtaining the silhouette scores for the sweep of the Seurat resolution parameter in order to evaluate the clustering fit. Additionally, this code describes the identification of DEGs for each cluster of the overall CRLM dataset.

```
crlm78.qc <- NormalizeData(crlm78.qc)

crlm78.qc <- FindVariableFeatures(crlm78.qc, selection.method="vst",
      nfeatures=2000)

crlm78.qc <- ScaleData(crlm78.qc, features=rownames(crlm78.qc))

crlm78.qc <- RunPCA(crlm78.qc, features=VariableFeatures(crlm78.qc))

crlm78.qc <- FindNeighbors(crlm78.qc, dims=1:15)

crlm78.qc <- FindClusters(crlm78.qc, resolution=0.5)

crlm78.qc <- RunUMAP(crlm78.qc, dims=1:10)

crlm78.qc <- RunTSNE(crlm78.qc, dims=1:10)


l.abbr <- vector(mode="character", length=length(crlm78.qc$orig.ident))

for (i in 1:length(crlm78.qc$orig.ident)) {

if (crlm78.qc$orig.ident[i]=="CRLM7-Organoid") {

l.abbr[i] <- "P7 Organoid" }
```

```
else if (crlm78.qc$orig.ident[i]=="CRLM7-ParentalTumor") {
l.abbr[i] <- "P7 Parental" }
else if (crlm78.qc$orig.ident[i]=="CRLM8-ParentalTumor") {
l.abbr[i] <- "P8 Parental" }
}

lev.fancy <- c("P7 Organoid", "P7 Parental", "P8 Parental")
crlm78.qc$fancy.orig.ident <- l.abbr
crlm78.qc$fancy.orig.ident <- factor(crlm78.qc$fancy.orig.ident,
      levels=lev.fancy)

library(cluster)
Idents(crlm78.qc) <- crlm78.qc$seurat_clusters
crlm.sil <- crlm78.qc
dist.mat <- dist(Embeddings(crlm.sil[['pca']])[,1:15])
res <- c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4,
      1.5)
crlm.sil <- FindClusters(crlm.sil, resolution=res)
sil.sum <- data.frame(matrix(NA_real_, nrow=length(res), ncol=4))
```

```
colnames(sil.sum) <-
      c("resolution","total.num.clusters","avg.sil.score","sd.sil.score")


for (i in 1:15) {
crlm.sil <- FindClusters(crlm.sil, resolution=res[i])
clust <- crlm.sil@meta.data$seurat_clusters
sil <- silhouette(as.numeric(clust), dist=dist.mat)
name.sil <- paste0("silhouette_score_",res[i])
crlm.sil[[name.sil]] <- sil[,3]
sil.sum[i,1] <- res[i]
sil.sum[i,2] <-
      as.numeric(last(rownames(table(crlm.sil@meta.data$seurat_clusters)))) +
      1
sil.sum[i,3] <- mean(sil[,3])
sil.sum[i,4] <- sd(sil[,3])
print(i)
}


Idents(crlm78.qc) <- crlm78.qc$RNA_snn_res.0.5
```

```
crlm.mark <- FindAllMarkers(crlm78.qc, only.pos=TRUE, min.pct=0.25,
      logfc.threshold=0.25)
top6 <- crlm.mark %>% group_by(cluster) %>% top_n(n=6, wt=avg_log2FC)
top8 <- crlm.mark %>% group_by(cluster) %>% top_n(n=8, wt=avg_log2FC)
top20 <- crlm.mark %>% group_by(cluster) %>% top_n(n=20, wt=avg_log2FC)
top40 <- crlm.mark %>% group_by(cluster) %>% top_n(n=40, wt=avg_log2FC)
```

The next portion of code describes the identification of epithelial cells from the larger CRLM dataset and the identification of GO terms from each cluster based on the top 40 DEGs for each epithelial cell cluster.

```
Idents(crlm78.qc) <- crlm78.qc$RNA_snn_res.0.5
crlm.sub <- subset(crlm78.qc, idents=c("2","7","13"))
clus2.mark40 <- crlm.mark[crlm.mark$cluster=="2" &
      crlm.mark$p_val_adj<0.05,]$gene[1:40]
clus7.mark40 <- crlm.mark[crlm.mark$cluster=="7" &
      crlm.mark$p_val_adj<0.05,]$gene[1:40]
clus13.mark40 <- crlm.mark[crlm.mark$cluster=="13" &
      crlm.mark$p_val_adj<0.05,]$gene[1:40]
```

```
library(topGO)

library(strngr)

library("org.Hs.eg.db")


# Looking first at cluster 2

all.genes <- rownames(GetAssayData(crlm78.qc,slot="counts"))

geneList.clus240 <- ifelse(all.genes %in% clus2.mark40, 1, 0)

names(geneList.clus240) <- all.genes

clus240.godata <- new("topGOdata", ontology="BP", allGenes=geneList.clus240,
      geneSelectionFun=function(x)(x==1), annot=annFUN.org,
      mapping="org.Hs.eg.db", ID="symbol", nodeSize=5) # creating the
      topGOdata object

rf.clus240 <- runTest(clus240.godata, algorith="elim", statistic="fisher")

clus240.res1 <- data.frame(GenTable(clus240.godata,Fisher=rf.clus240,
      topNodes=25,numChar=60))

clus240.res2 <- clus240.res1[,c('GO.ID','Term','Fisher')]

clus240.res3 <- clus240.res2[1:25,]

clus240.res3$Place <- rownames(clus240.res3)

clus240.res3$neglogPval <- -log10(as.numeric(clus240.res3$Fisher))

clus240.res3[is.na(clus240.res3)] <- 30
```

```
clus240.res3$abbrTerm <- rbind(str_trunc(clus240.res3$Term, 38, "right"))[1,]

# Next look at cluster 7

all.genes <- rownames(GetAssayData(crlm78.qc,slot="counts"))

geneList.clus740 <- ifelse(all.genes %in% clus7.mark40, 1, 0)

names(geneList.clus740) <- all.genes

clus740.godata <- new("topGOdata", ontology="BP", allGenes=geneList.clus740,
      geneSelectionFun=function(x)(x==1), annot=annFUN.org,
      mapping="org.Hs.eg.db", ID="symbol", nodeSize=5) # creating the
      topGOdata object

rf.clus740 <- runTest(clus740.godata, algorith="elim", statistic="fisher")

clus740.res1 <- data.frame(GenTable(clus740.godata,Fisher=rf.clus740,
      topNodes=25,numChar=60))

clus740.res2 <- clus740.res1[,c('GO.ID','Term','Fisher')]

clus740.res3 <- clus740.res2[1:25,]

clus740.res3$Place <- rownames(clus740.res3)

clus740.res3$neglogPval <- -log10(as.numeric(clus740.res3$Fisher))

clus740.res3[is.na(clus740.res3)] <- 30

clus740.res3$abbrTerm <- rbind(str_trunc(clus740.res3$Term, 38, "right"))[1,]


# Next look at cluster 13
```

```
all.genes <- rownames(GetAssayData(crlm78.qc,slot="counts"))

geneList.clus1340 <- ifelse(all.genes %in% clus13.mark40, 1, 0)

names(geneList.clus1340) <- all.genes

clus1340.godata <- new("topGOdata", ontology="BP",
      allGenes=geneList.clus1340, geneSelectionFun=function(x)(x==1),
      annot=annFUN.org, mapping="org.Hs.eg.db", ID="symbol", nodeSize=5) #
      creating the topGOdata object

rf.clus1340 <- runTest(clus1340.godata, algorith="elim", statistic="fisher")

clus1340.res1 <- data.frame(GenTable(clus1340.godata,Fisher=rf.clus1340,
      topNodes=25,numChar=60))

clus1340.res2 <- clus1340.res1[,c('GO.ID','Term','Fisher')]

clus1340.res3 <- clus1340.res2[1:25,]

clus1340.res3$Place <- rownames(clus1340.res3)

clus1340.res3$neglogPval <- -log10(as.numeric(clus1340.res3$Fisher))

clus1340.res3[is.na(clus1340.res3)] <- 30

clus1340.res3$abbrTerm <- rbind(str_trunc(clus1340.res3$Term, 38,
      "right"))[1,]
```

Personal References

NOTE: The list is accurate as of Wednesday, June 8, 2022.

Curtis, MB, Kelly, N, Hughes, CCW, et al. 2022. Organotypic stromal cells impact endothelial cell transcriptome in 3D capillary networks. ***In Review***.

Rollins, ZA, Curtis, MB, Faller, R, et al. 2022. Automated homology modelling of TCRs to a target pMHC based on patient-specific sequences. ***In Review***.

Ang, LT, Nguyen, A, … Curtis, MB, et al. 2022. Generating human artery and vein cells from pluripotent stem cells highlights the arterial tropism of Nipah and Hendra viruses. *Cell*. ***In Press***.

Joshi, R, Hadley, D, … Curtis, MB. 2022. Concept mapping serves as metacognition tool in a problem-solving-based BME course during in-person and online instruction. Biomedical Engineering Education. ***In Press***. https://doi.org/10.1007/s43683-022-00066-3.

Glaser, DE, Curtis, MB, Sariano, PA, et al. 2022. Organ-on-a-chip model of vascularized human bone marrow niches. *Biomaterials*. **280**: 121245. https://doi.org/10.1016/j.biomaterials.2021.121245.

Campbell, K, Curtis, MB, Massey, JM, et al. 2021. Isolating and characterizing lymphatic endothelial progenitor cells for potential therapeutic lymphangiogenic applications. *Acta Biomaterialia*. **135**: 191-202. https://doi.org/10.1016/j.actbio.2021.08.005.

Shirure, VS, Bi, Y, Curtis, MB, et al. 2018. Tumor-on-a-chip platform to investigate progression and drug sensitivity in cell lines and patient-derived organoids. *Lab on a Chip*. **18**(23): 3687-3702. https://doi.org/10.1039/c8lc00596f.

Index of References in the Dissertation

1. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**, 613–626 (2012).
2. Mitsis, T. *et al.* Transcription factors and evolution: An integral part of gene expression (Review). *World Academy of Sciences Journal* **2**, 3–8 (2020).
3. Agrawal, N. *et al.* RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews* **67**, 657–685 (2003).
4. Svoboda, P. Key Mechanistic Principles and Considerations Concerning RNA Interference. *Frontiers in Plant Science* **11**, (2020).
5. Deribe, Y. L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nat Struct Mol Biol* **17**, 666–672 (2010).
6. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377–382 (2009).
7. Hirsch, C. & Schildknecht, S. In Vitro Research Reproducibility: Keeping Up High Standards. *Frontiers in Pharmacology* **10**, (2019).
8. Aird, W. C. Phenotypic heterogeneity of the endothelium: II. Representative vascular beds. *Circulation Research* **100**, 174–190 (2007).
9. Obermeier, B., Daneman, R. & Ransohoff, R. M. Development, maintenance and disruption of the blood-brain barrier. *Nature Medicine 2013 19:12* **19**, 1584–1596 (2013).
10. Auvinen, K. *et al.* Fenestral diaphragms and PLVAP associations in liver sinusoidal endothelial cells are developmentally regulated. *Scientific Reports 2019 9:1* **9**, 1–16 (2019).
11. Lammert, E., Cleaver, O. & Melton, D. Induction of Pancreatic Differentiation by Signals from Blood Vessels. *Science* **294**, 564–567 (2001).
12. Rafii, S., Butler, J. M. & Ding, B.-S. Angiocrine functions of organ-specific endothelial cells. *Nature* **529**, 316–325 (2016).

13. Daniel, E. & Cleaver, O. Chapter Six - Vascularizing organogenesis: Lessons from developmental biology and implications for regenerative medicine. in *Current Topics in Developmental Biology* (ed. Wellik, D. M.) vol. 132 177–220 (Academic Press, 2019).
14. Marcu, R. *et al.* Human Organ-Specific Endothelial Cell Heterogeneity. *iScience* **4**, 20–35 (2018).
15. Cleuren, A. C. A. *et al.* The in vivo endothelial cell translatome is highly heterogeneous across vascular beds. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 23618–23624 (2019).
16. Nolan, D. J. *et al.* Molecular Signatures of Tissue-Specific Microvascular Endothelial Cell Heterogeneity in Organ Maintenance and Regeneration. *Developmental Cell* **26**, 204–219 (2013).
17. Ricard, N., Bailly, S., Guignabert, C. & Simons, M. The quiescent endothelium: signalling pathways regulating organ-specific endothelial normalcy. *Nature Reviews Cardiology 2021 18:8* **18**, 565–580 (2021).
18. Griffith, C. K. & George, S. C. The Effect of Hypoxia on In Vitro Prevascularization of a Thick Soft Tissue. *Tissue Engineering Part A* **15**, 2423–2434 (2009).
19. Kreutziger, K. L. *et al.* Developing Vasculature and Stroma in Engineered Human Myocardium. *Tissue Engineering Part A* **17**, 1219–1228 (2011).
20. Eilken, H. M. *et al.* Pericytes regulate VEGF-induced endothelial sprouting through VEGFR1. *Nature Communications 2017 8:1* **8**, 1–14 (2017).
21. Ghajar, C. M. *et al.* Mesenchymal cells stimulate capillary morphogenesis via distinct proteolytic mechanisms. *Experimental Cell Research* **316**, 813–825 (2010).
22. Seynhaeve, A. L. B. *et al.* Spatiotemporal endothelial cell – pericyte association in tumors as shown by high resolution 4D intravital imaging. *Scientific Reports 2018 8:1* **8**, 1–13 (2018).
23. Lindahl, P., Johansson, B. R., Levéen, P. & Betsholtz, C. Pericyte Loss and Microaneurysm Formation in PDGF-B-Deficient Mice. *Science* **277**, 242–245 (1997).
24. Hellström, M. *et al.* Lack of Pericytes Leads to Endothelial Hyperplasia and Abnormal Vascular Morphogenesis. *Journal of Cell Biology* **153**, 543–554 (2001).

25. Newman, A. C., Nakatsu, M. N., Chou, W., Gershon, P. D. & Hughes, C. C. W. The requirement for fibroblasts in angiogenesis: Fibroblast-derived matrix proteins are essential for endothelial cell lumen formation. *Molecular Biology of the Cell* **22**, 3791–3800 (2011).
26. Newman, A. C. *et al.* Analysis of stromal cell secretomes reveals a critical role for stromal cell-derived hepatocyte growth factor and fibronectin in angiogenesis. *Arteriosclerosis, Thrombosis, and Vascular Biology* **33**, 513–522 (2013).
27. Shirure, V. S., Lezia, A., Tao, A., Alonzo, L. F. & George, S. C. Low levels of physiological interstitial flow eliminate morphogen gradients and guide angiogenesis. *Angiogenesis* **20**, 493–504 (2017).
28. Shirure, V. S. *et al.* Tumor-on-a-chip platform to investigate progression and drug sensitivity in cell lines and patient-derived organoids. *Lab on a Chip* **18**, 3687–3702 (2018).
29. Glaser, D. E. *et al.* Organ-on-a-chip model of vascularized human bone marrow niches. *Biomaterials* **280**, 121245–121245 (2022).
30. Chen, M. B. *et al.* Inflamed neutrophils sequestered at entrapped tumor cells via chemotactic confinement promote tumor cell extravasation. *Proceedings of the National Academy of Sciences* **115**, 7022–7027 (2018).
31. Kim, S., Lee, H., Chung, M. & Jeon, N. L. Engineering of functional, perfusable 3D microvascular networks on a chip. *Lab Chip* **13**, 1489–1500 (2013).
32. Sui, C., Zilberberg, J. & Lee, W. Microfluidic device engineered to study the trafficking of multiple myeloma cancer cells through the sinusoidal niche of bone marrow. *Sci Rep* **12**, 1439 (2022).
33. Ahn, S. I. *et al.* Microengineered human blood–brain barrier platform for understanding nanoparticle transport mechanisms. *Nat Commun* **11**, 175 (2020).
34. Petrosyan, A. *et al.* A glomerulus-on-a-chip to recapitulate the human glomerular filtration barrier. *Nat Commun* **10**, 3656 (2019).
35. Schupp, J. C. *et al.* Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung. *Circulation* **144**, 286–302 (2021).
36. Solé-Boldo, L. *et al.* Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Communications Biology 2020 3:1* **3**, 1–12 (2020).

37. Chen, X. *et al.* Rapid Anastomosis of Endothelial Progenitor Cell–Derived Vessels with Host Vasculature Is Promoted by a High Density of Cotransplanted Fibroblasts. *Tissue Eng Part A* **16**, 585–594 (2010).
38. Melero-Martin, J. M. *et al.* Engineering Robust and Functional Vascular Networks In Vivo With Human Adult and Cord Blood–Derived Progenitor Cells. *Circulation Research* **103**, 194–202 (2008).
39. Ehsan, S. M. & George, S. C. Vessel network formation in response to intermittent hypoxia is frequency dependent. *Journal of Bioscience and Bioengineering* **120**, 347–350 (2015).
40. Zudaire, E., Gambardella, L., Kurcz, C. & Vermeren, S. A computational tool for quantitative analysis of vascular networks. *PLoS ONE* **6**, (2011).
41. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
42. Vestweber, D. VE-Cadherin. *Arteriosclerosis, Thrombosis, and Vascular Biology* **28**, 223–232 (2008).
43. Woodfin, A., Voisin, M.-B. & Nourshargh, S. PECAM-1: A Multi-Functional Molecule in Inflammation and Vascular Biology. *Arteriosclerosis, Thrombosis, and Vascular Biology* **27**, 2514–2523 (2007).
44. Carrion, B., Janson, I. A., Kong, Y. P. & Putnam, A. J. A Safe and Efficient Method to Retrieve Mesenchymal Stem Cells from Three-Dimensional Fibrin Gels. *Tissue Engineering Part C: Methods* **20**, 252–263 (2014).
45. Goncharov, N. V., Nadeev, A. D., Jenkins, R. O. & Avdonin, P. V. Markers and Biomarkers of Endothelium: When Something Is Rotten in the State. *Oxidative Medicine and Cellular Longevity* **2017**, e9759735 (2017).
46. Yamazaki, T. & Mukouyama, Y. Tissue Specific Origin, Development, and Pathological Perspectives of Pericytes. *Frontiers in Cardiovascular Medicine* **5**, (2018).
47. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
48. Lin, W. *et al.* Single-cell transcriptome analysis of tumor and stromal compartments of pancreatic ductal adenocarcinoma primary tumors and metastatic lesions. *Genome Medicine* **12**, 80 (2020).
49. He, S. *et al.* Single-cell transcriptome profiling of an adult human cell atlas of 15 major organs. *Genome Biology* **21**, 294 (2020).

50. THE TABULA SAPIENS CONSORTIUM. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* **376**, eabl4896 (2022).
51. Winkler, E. A. *et al.* A single-cell atlas of the normal and malformed human brain vasculature. *Science* **375**, eabi7377 (2022).
52. Garcia, F. J. *et al.* Single-cell dissection of the human brain vasculature. *Nature* **603**, 893–899 (2022).
53. Yang, A. C. *et al.* A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022).
54. Paik, D. T. *et al.* Single-Cell RNA Sequencing Unveils Unique Transcriptomic Signatures of Organ-Specific Endothelial Cells. *Circulation* **142**, 1848–1862 (2020).
55. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
56. Helle, E., Ampuja, M., Antola, L. & Kivelä, R. Flow-Induced Transcriptomic Remodeling of Endothelial Cells Derived From Human Induced Pluripotent Stem Cells. *Frontiers in Physiology* **11**, 591450 (2020).
57. Roux, E., Bougaran, P., Dufourcq, P. & Couffinhal, T. Fluid Shear Stress Sensing by the Endothelial Layer. *Frontiers in Physiology* **11**, 861 (2020).
58. Grellier, M., Bareille, R., Bourget, C. & Amédée, J. Responsiveness of human bone marrow stromal cells to shear stress. *Journal of Tissue Engineering and Regenerative Medicine* **3**, 302–309 (2009).
59. Bruno, A. *et al.* Orchestration of Angiogenesis by Immune Cells. *Frontiers in Oncology* **4**, 131 (2014).
60. Kim, H., Wang, M. & Paik, D. T. Endothelial-Myocardial Angiocrine Signaling in Heart Development. *Frontiers in Cell and Developmental Biology* **9**, 697130 (2021).
61. Zhou, X., Chen, Y., Cui, L., Shi, Y. & Guo, C. Advances in the pathogenesis of psoriasis: from keratinocyte perspective. *Cell Death Dis* **13**, 1–13 (2022).
62. Surveillance Research Program at National Cancer Institute. SEER*Explorer: An interactive website for SEER cancer statistics.

63. Jegatheeswaran, S., Mason, J. M., Hancock, H. C. & Siriwardena, A. K. The liver-first approach to the management of colorectal cancer with synchronous hepatic metastases: a systematic review. *JAMA Surgery* **148**, 385–391 (2013).
64. Vatandoust, S., Price, T. J. & Karapetis, C. S. Colorectal cancer: metastases to a single organ. *World Journal of Gastroenterology* **21**, 11767–11776 (2015).
65. Siegel, R. L. *et al.* Colorectal cancer statistics, 2020. *CA: A Cancer Journal for Clinicians* **70**, 145–164 (2020).
66. Zhang, T. *et al.* Identification of Candidate Biomarkers and Prognostic Analysis in Colorectal Cancer Liver Metastases. *Frontiers in Oncology* **11**, 3051–3051 (2021).
67. Kopetz, S. *et al.* Improved Survival in Metastatic Colorectal Cancer Is Associated With Adoption of Hepatic Resection and Improved Chemotherapy. *Journal of Clinical Oncology* **27**, 3677–3677 (2009).
68. Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M. & Wallace, M. B. Colorectal cancer. *The Lancet* **394**, 1467–1480 (2019).
69. Chow, F. C. L. & Chok, K. S. H. Colorectal liver metastases: An update on multidisciplinary approach. *World Journal of Hepatology* **11**, 150–172 (2019).
70. Nordlinger, B. *et al.* Perioperative FOLFOX4 chemotherapy and surgery versus surgery alone for resectable liver metastases from colorectal cancer (EORTC 40983): long-term results of a randomised, controlled, phase 3 trial. *The Lancet Oncology* **14**, 1208–1215 (2013).
71. Tintelnot, J. & Stein, A. Immunotherapy in colorectal cancer: Available clinical evidence, challenges and novel approaches. *World Journal of Gastroenterology* **25**, 3920–3928 (2019).
72. Bi, Y. *et al.* Tumor-on-a-chip platform to interrogate the role of macrophages in tumor progression. *Integrative Biology* **12**, 221–232 (2020).
73. Sachs, N. *et al.* A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity. *Cell* **172**, 373-386.e10 (2018).
74. Rosenbluth, J. M. *et al.* Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. *Nature Communications 2020 11:1* **11**, 1–14 (2020).

75. Krieger, T. G. *et al.* Single-cell analysis of patient-derived PDAC organoids reveals cell state heterogeneity and a conserved developmental hierarchy. *Nature Communications 2021 12:1* **12**, 1–13 (2021).
76. Driehuis, E. *et al.* Pancreatic cancer organoids recapitulate disease and allow personalized drug screening. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 26580–26590 (2019).
77. Jacob, F. *et al.* A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. *Cell* **180**, 188-204.e22 (2020).
78. Okamoto, T. *et al.* Comparative Analysis of Patient-Matched PDOs Revealed a Reduction in OLFM4-Associated Clusters in Metastatic Lesions in Colorectal Cancer. *Stem Cell Reports* **16**, 954–967 (2021).
79. Yan, H. H. N. *et al.* Organoid cultures of early-onset colorectal cancers reveal distinct and rare genetic profiles. *Gut* **69**, (2020).
80. Fujii, M. *et al.* A Colorectal Tumor Organoid Library Demonstrates Progressive Loss of Niche Factor Requirements during Tumorigenesis. *Cell Stem Cell* **18**, 827–838 (2016).
81. Weeber, F. *et al.* Preserved genetic diversity in organoids cultured from biopsies of human colorectal cancer metastases. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13308–13311 (2015).
82. Cristobal, A. *et al.* Personalized Proteome Profiles of Healthy and Tumor Human Colon Organoids Reveal Both Individual Diversity and Basic Features of Colorectal Cancer. *Cell Reports* **18**, 263–274 (2017).
83. Buzzelli, J. N., Ouaret, D., Brown, G., Allen, P. D. & Muschel, R. J. Colorectal cancer liver metastases organoids retain characteristics of original tumor and acquire chemotherapy resistance. *Stem Cell Research* **27**, 109–120 (2018).
84. Che, L.-H. *et al.* A single-cell atlas of liver metastases of colorectal cancer reveals reprogramming of the tumor microenvironment in response to preoperative chemotherapy. *Cell Discovery* **7**, 80–80 (2021).

85. Sato, T. *et al.* Long-term Expansion of Epithelial Organoids From Human Colon, Adenoma, Adenocarcinoma, and Barrett's Epithelium. *Gastroenterology* **141**, 1762–1772 (2011).
86. Fujii, M. & Sato, T. Somatic cell-derived organoids as prototypes of human epithelial tissues and diseases. *Nat. Mater.* **20**, 156–169 (2021).
87. Nigam, S. K. What do drug transporters really do? *Nat Rev Drug Discov* **14**, 29–44 (2015).
88. Vasilogianni, A.-M. *et al.* Proteomics of colorectal cancer liver metastasis: A quantitative focus on drug elimination and pharmacodynamics effects. *British Journal of Clinical Pharmacology* **88**, 1811–1823 (2022).
89. Aguirre-Portolés, C., Feliu, J., Reglero, G. & Ramírez de Molina, A. ABCA1 overexpression worsens colorectal cancer prognosis by facilitating tumour growth and caveolin-1-dependent invasiveness, and these effects can be ameliorated using the BET inhibitor apabetalone. *Molecular Oncology* **12**, 1735–1752 (2018).
90. Strnad, P., Stumptner, C., Zatloukal, K. & Denk, H. Intermediate filament cytoskeleton of the liver in health and disease. *Histochem Cell Biol* **129**, 735 (2008).
91. Tot, T. & Samii, S. The clinical relevance of cytokeratin phenotyping in needle biopsy of liver metastasis. *APMIS* **111**, 1075–1082 (2003).
92. Polari, L. *et al.* Keratin intermediate filaments in the colon: guardians of epithelial homeostasis. *The International Journal of Biochemistry & Cell Biology* **129**, 105878 (2020).
93. Zhou, Y. *et al.* Cancer stem cells in progression of colorectal cancer. *Oncotarget* **9**, 33403–33415 (2018).
94. Engel, R. M. *et al.* Patient-Derived Colorectal Cancer Organoids Upregulate Revival Stem Cell Marker Genes following Chemotherapeutic Treatment. *Journal of Clinical Medicine* **9**, 128 (2020).
95. Wu, Y. *et al.* Spatiotemporal Immune Landscape of Colorectal Cancer Liver Metastasis at Single-Cell Level. *Cancer Discovery* **12**, 134–153 (2022).
96. Peng, J. *et al.* Immune Cell Infiltration in the Microenvironment of Liver Oligometastasis from Colorectal Cancer: Intratumoural CD8/CD3 Ratio Is a Valuable Prognostic Index for Patients Undergoing Liver Metastasectomy. *Cancers* **11**, 1922 (2019).

97. Masuda, K. *et al.* Multiplexed single-cell analysis reveals prognostic and nonprognostic T cell types in human colorectal cancer. *JCI Insight* **7**, (2022).

98. Kim, J. H., Rhee, Y.-Y., Bae, J. M., Cho, N.-Y. & Kang, G. H. Loss of CDX2/CK20 Expression Is Associated With Poorly Differentiated Carcinoma, the CpG Island Methylator Phenotype, and Adverse Prognosis in Microsatellite-unstable Colorectal Cancer. *The American Journal of Surgical Pathology* **37**, 1532–1541 (2013).

99. Fumagalli, A. *et al.* Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. *Cell Stem Cell* **26**, 569-578.e7 (2020).

100. Wang, R. *et al.* Systematic evaluation of colorectal cancer organoid system by single-cell RNA-Seq analysis. *Genome Biology* **23**, 106 (2022).

101. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e29 (2021).

102. Overview of Single Cell Software -Software -Single Cell Gene Expression -Official 10x Genomics Support. https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome.