Technische Universität München
TUM School of Natural Sciences

Transcriptional monitoring of endothelial development at the single-cell level

Simon T. Rosowski

Vollständiger Abdruck der von der TUM School of Natural Sciences der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Gil Westmeyer

Prüfer*innen der Dissertation:

1. Prof. Dr. Michael Sattler
2. Prof. Dr. Matthias Meier

Die Dissertation wurde am 30. 01. 2023 bei der Technischen Universität München eingereicht und durch die TUM School of Natural Sciences am 01. 06. 2023 angenommen.

Acknowledgements

First and foremost, I thank **Prof. Dr. Matthias Meier** for giving me the opportunity to work in his research group. Without his support, patience, and understanding I would not have been able to accomplish the work presented here. Further, I thank **Dr. Julius Wiener** and **Daniel Kokotek** introducing me into techniques of antibody labeling and proximity ligation assay. Especially, **Julius** supported me with his mentoring during the first year of my PhD and further gave his specialist advice and discussed ideas in shaping my projects. I want to thank **Prof. Dr. Heiko Lickert** and his group for the friendly uptake into their lab in the very begin of my PhD. Special thanks go to **Dr. Annika Bötcher**, **Dr. Michael Sterr**, and **Ines Kunze** for supporting me and coordinating the single-cell transcriptomic pipeline.

Prof. Dr. Michael Sattler and **Dr. Celia Martinez-Jimenez** have my gratitude for being my university supervisor and additional advisor, respectively. Your support, fruitful input, and critical questions during the thesis committee meetings helped me a lot to get a red line into my work and how to communicate it to people with different background. Also, I am very pleased that **Prof. Dr. Gil Westmeyer** agreed to be the chair in my thesis defense committee.

I want to thank **Dr. Volker Bergen** and **Prof. Dr. Fabian Theis** for their input on the analysis of the single-cell mRNA sequencing analysis and especially the advices of how to apply their mRNA velocity algorithm on my data.

Additionally, I want to thank the whole group for all the laughing and crying inside and outside the lab.

With special thanks to **Dr. Sandra Wiedenmann**, **Maren Marder**, **Caroline Brähler**, and **Alina Platen** for their generous and lasting support in plenty of the experiments. I thank **Dr. Misao Akishiba** for fruitful discussions about the planning of experiments and funneling analysis strategies. **Johannes Wirth** has my thankfulness for our splendid work on developing and establishing the SPLiT-seq protocol in our laboratory. **Dr. Scott Atwell** is thanked for his exceptional support and explanations regarding imaging and imagej scripting. Further thankfulness goes to **Judith Bushe** and **Annette Feuchtinger** who performed the imaging and staining of cryo-sections from day 12 and 18 Matrigel culture.

Dr. Francisco Ruiz Ojeda, **Siegfried Ussar**, **Samira Zamani**, and **Michelle Jentzsch** for the ability to cooperate with them on the interesting topic of cellular signaling in type 2 diabetes.

All members of the Helmholtz Pioneer Campus for general supportive structures, the scientific exchange and exciting social events. Especially, the two retreats gave me a lot of impulses for future perspectives, in research and in life.

Finally, I thank **Mika Heimerl** for proofreading of my thesis. I keep my acknowledgements here on the professional level this should not exclude my gratefulness for **all the beautiful people** I shared special moments but not everyone who reads this thesis needs to know about. As a human, I am a social animal that relies on positive interactions undergo prospering development. I am aware of the fortunateness to which I have been exposed to in my life and am deeply thankful for all the support I am getting from the society, all the friends I have and my family.

Zusammenfassung

Vaskulare Strukturen sind unabdingbar für die Versorgung größerer Gewebestrukturen und daher von grundlegender Bedeutung für die künstliche Gewebezüchtung. Diese ist momentan ein limitierender Faktor in der Organoid Forschung. Weiterhin wird die Effektivität der Gewebeannahme nach einer Transplantation durch vorherige Vaskularisierung erheblich gesteigert. Für diese Arbeit wurden Endothelzellen aus humanen induzierten pluripotenten Stammzellen in einer 3D Suspensionskultur differenziert. Um die Neovaskularisierung in unterschiedlichen 3D Kleinstumgebungen zu untersuchen, wurde eine Einzelzelltranskriptionsuntersuchung konzipiert. Die zeitaufgelöste Probennahme erlaubte ein Verfolgen der gleichzeitigen und untrennbaren Entwicklung von Endothelzellen und Muralzellen, sowie Einsichten in deren Stabilität und phänotypische Formbarkeit. Die Überführung in eine 3D Matrixumgebung induzierte Neovaskularisierung und erlaubte die Charakterisierung von austreibenden, verschmelzenden und gefäßformenden Endothelzellphänotypen. Während der Verfolgung von Reifeprozessen des entstandenen vaskularen Netzwerks konnten zwei Unterpopulationen von Perizyten identifiziert werden, die sich aus den Muralzellen herausgebildet hatten. Ein Filtern nach Zell-Zell Interaktionen auf Basis der Einzelzelltranskripionsdaten zwischen Perizyten und gefäßformenden Endothelzellen ermöglichte ein Bestätigen von bereits beschriebenen Angiogenesesignalen im Organismus wie auch ein Aufdecken neuer Zytokine während der Gefäßbildung. Durch die Daten und deren Analyse können zukünftige Entwicklungen für Vaskularisierungsbesterbungen in der künstlichen Gewebezüchtung verbessert und verstanden werden. Weiterhin wurde die Integrierbarkeit von antikörperbasierten Techniken in die Prozesskette der Einzelzellsequenzierung untersucht. Diese können auf Proteinebene Aufschluss über Einzelzellmerkmale geben, die über die Erhebung von Transkriptomdaten hinaus geht. Eine Anwendung auf die Endothelzellentwicklung könnte so zusätzliche Daten hinsichtlich vorhandener Proteine und des Zustands der zellulären Signaltransduktion liefern.

Abstract

The formation of vascular structures is fundamental for *in vitro* tissue engineering. Vascularization can enable the nutrient supply within larger structures and increase transplantation efficiency, which are currently limiting factors in organoid research. We differentiated human induced pluripotent stem cells toward endothelial cells in 3D suspension culture. To investigate *in vitro* neovascularization and various 3D microenvironmental approaches, we designed a comprehensive single-cell transcriptomic study. Time-resolved single-cell transcriptomics of the endothelial and co-evolving mural cells gave insights into cell type development, stability, and plasticity. Transfer to a 3D hydrogel microenvironment induced neovascularization and facilitated tracing of sprouting, coalescing, and tubulogenic endothelial cells states. During maturation, we monitored two pericyte subtypes evolving of mural cells. Profiling cell-cell interactions between pericytes and endothelial cells confirmed *in vivo* angiogenic signaling and emphasized new cytokine signals during tubulogenesis. Our data, analyses, and results provide an *in vitro* roadmap to guide vascularization in future tissue engineering. Additionally, the integration of antibody-based detection methods into the single-cell mRNA sequencing workflow was investigated. These can provide single-cell characteristics on the protein level that go beyond the single-cell transcriptomic data. Their application on the ECs development would present further details about present proteins and the cell signaling status.

Table of Contents

Abbreviations

acronym	written out
AA	ascorbic acid
AF	alexa flour
APR	auto-cycling proximity recording
BC	barcode
BSA	bovine serum albumin
CD	cluster of differentiation
cDNA	complementary deoxyribonucleic acid
cEC	coalescing endothelial cells
CITE-seq	cellular indexing of transcriptomes and epitopes by sequencing
DAPI	4′,6-diamidino-2-phenylindole
DBCO	dibenzocyclooctyne
DDG	dynamic driver gene
DEG	differentially expressed gene
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double stranded deoxyribonucleic acid
DTT	dithiothreitol
EC	endothelial cell
ECM	extracellular matrix
ELISA	enzyme-linked immunosorbent assay
EndoMT	endothelial to mesenchymal transition
FACS	fluorescence-activated cell sorter
FRET	Förster resonance energy transfer
GEM	gel beads in emulsion
GO	gene ontology
hiPSC	human induced pluripotent stem cell
IntAct-seq	interaction sequencing
iPSC	induced pluripotent stem cell
KEGG	Kyoto encyclopedia of genes and genomes
MEndoT	mesenchymal to endothelial transition
mRNA	messenger ribonucleic acid
NGS	next generation sequencing
NHS	N-hydroxysuccinimide
OXPHOS	oxidative phosphorylation
P1	pericyte cluster 1

P2	pericyte cluster 2
PBS	phosphate buffered saline
PCA	principle component analysis
PCR	polymerase chain reaction
PDMS	polydimethylsiloxane
PEG	polyethylene glycol
PER	Primer exchange reaction
PFA	paraformaldehyde
PLA	proximity ligation assay
RCA	rolling circle amplification
RLU	relative luminescence units
RTK	receptor tyrosine kinase
SB	sodium borate
SCF	subcuanous fraction
scRNA-seq	single-cell mRNA sequencing
snRNA-seq	single-nuclei mRNA sequencing
SEC	size exclusion chromatography
mEC	migrating endothelial cells
smart	switching mechanism at 5’ end of the RNA transcript
SPLiT-seq	split-pool ligation-based transcriptome sequencing
ssDNA	single stranded deoxyribonucleic acid
tEC	tubulogenic endothelial cells
TF	transcription factor
TFEA	transcription factor enrichment analysis
tSNE	t-distributed stochastic neighbor embedding
TSO	template switching oligonucleotide
UMAP	Uniform Manifold Approximation and Projection
UMI	unique molecular identifier

1. Introduction

1.1. The cardiovascular system

Each and every cell in the body needs nutrients and a system for the disposal of cellular waste. At a certain size of a multicellular organism, diffusion fails maintaining this need. For this reason, in vertebrates the cardiovascular system with blood as supply liquid has developed. As its centerpiece, the heart maintains the circulation of the blood while blood vessels are the structural components to facilitate blood transport through the body. In humans, tubes start from up to 1 cm in diameter around the heart and level off in capillaries with the diameter of a hair to reach and supply every cell in the body. The total length of all vessels lined up mounts more than 100,000 km in an adult human[1]. The inner coating of each vessel consists of a single layer of endothelial cells (ECs) while its stability is assured by pericytes that encase these tubular structures against mechanical stress. In the larger vessels, vascular smooth muscle cells (VSMCs) grant further stability. Both cell types are summarized in the term mural cell. In the literature pericytes are mainly stated as microvascular periendothelial mesenchymal cells. Further, it has to be highlighted that no unequivocal marker can be used to differentiate between pericytes and VSMCs or other mesenchymal cells[2].

1.2. Cardiovascular development

In embryonal development, the heart is the first functional organ of the body, however, it is not pumping before a primary vascular system has formed. The heart arises, like other organs, through the migration of precursor cells that are specializing by differentiation, initiated through signaling and interactions with the surrounding tissue[3]. The precursors for the heart are termed multipotent cardiovascular progenitors and derive from the mesoderm. Here, FGF signaling is important to induce the expression of transcription factors (TFs) of the GATA, and HAND-family[4,5]. The multipotent cardiovascular progenitors embody the cell type that differentiates to all the different cell types in the heart, namely cardiomyocytes, smooth muscle cells, ECs, and endocardium cells[6–8]. The intermediate stage between the multipotent cardiac precursor cells and ECs that are structurally arranged in vessels is represented by hemangioblasts or angioblasts which are in loose association. The generation of vessels starts with vasculogenesis, the *de novo* generation of blood vessels. This accompanies with the development of blood islets from the angioblasts, their expansion and coalescing of blood vesicles into vessel structures, and formation of the primary cardiovascular system[9]. During development of new tissue and organs these existing vessels have to be extended to ensure the supply of the emerging structures which is guaranteed by the process of angiogenesis.

Angiogenesis is initiated by signaling molecules of the vascular endothelial growth factor (VEGF) family that are recognized *via* receptors on the ECs. Hereby, a single leading cell, termed tip cell, starts to sprout out of the present vessel and signals neighboring cells to support its migration by expanding the tube in

direction of sprouting[10]. The following and tube extending cell phenotype is termed stalk cell. The initial step for tip cell development, phenotypically noticeable by formation of long cytoplastic extensions, the filopodia, is VEGF-signaling (**Figure 1.1**). The development of the tip cell phenotype is a dynamic process that is induced in all ECs that are exposed to a VEGFA gradient. However, as soon as one cell is committed to the tip phenotype, it suppresses the tip development of neighboring cells *via* the Notch-pathway and thereby establishes their stalk tone (**Figure 1.1**)[11]. Whereas a gradient of VEGFA leads to angiogenesis, a high homogenous concentration of VEGFA induces proliferation[12]. The secretion of soluble (s)FLT1 by non-tip ECs leads to a corridor thereby provides additional guidance and prevents turn around back to the parental vessel[13].

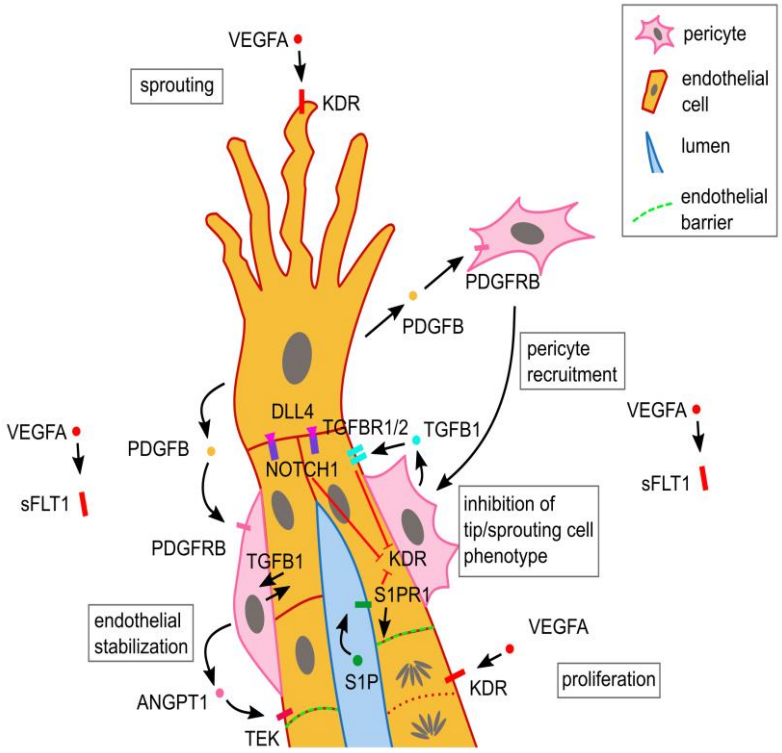

Figure 1.1 | Process of angiogenesis with subtypes of ECs and recruitment of pericytes. Different soluble and surface bound ligand-receptor interaction between ECs pericytes and surrounding cells are mandatory for angiogenesis. The figure is modified from Betsholtz and Ramsauer & D'Amore[14,15].

Besides inter EC signaling, communication between different cell types gives another branch of mechanistical understanding of vascularization. Besides ECs, mural cells are involved in that procedure. A subgroup of these, which facilitate the stabilization of capillaries, are termed pericytes. *In vitro* models have shown that pericyte recruitment is necessary for EC maturation and stabilization of vessels[16]. Hereby, one main signaling factor is PDGFB which is highly expressed by the ECs in the initial recruitment steps. PDGFB is a ligand for PDGFRB that also acts as a main marker for pericyte phenotype. When pericytes have attached to newly formed vessel structures, they express ANGPT1 which is a ligand for TEK that is required in vascular remodeling and maturation[17]. A reverse mechanism is caused by ANGPT2, a partial

agonist or even antagonist of TEK[18], it induces vessel destabilization by leading to the dissociation of pericytes from these vessel structures[19]. In addition, pericytes have a major impact on ECs in vascular development, maintenance, as well as in inflammation or disease[2]. Another important signaling pathway between ECs and pericytes is the TGF pathway. TGFB1 is the main ligand in this manner and inhibits EC proliferation[20] and migration[21]. It acts on TGFBR2 and TGFBR1 to induce a multimerization which also causes a signal transduction to decrease KDR expression[22,23]. Additionally, TGFB1 plays a bidirectional role by being expressed by ECs and pericytes and acting reciprocally[2]. Further factors in vessel stabilization are laminar shear stress as well as S1P which both induce S1PR1 to stabilize adherens junctions and barrier function[24]. Moreover, S1PR1 plays an important role in the development and homeostasis of blood vessels and prevents excessive sprouting[25].

1.3. Diseases with involvement of ECs

Since ECs form the walls of the cardiovascular system that supplies every cell in the body with nutrition and facilitates the removal of cellular waste, they have omnipresent functions and undergo various stress situations. Multiple protective mechanisms and signaling pathways allow a vessel maintenance to handle changing conditions. Therefore, states of injury and acute inflammation in terms of microbial infection can be overcome effective by a healthy microenvironment. For instance, in the state of an injury of large vessels a subtype of ECs in those, termed vascular endothelial stem cells, is able to perform an immediate transcriptomic adaption to a changed environment and initiate the recovery of affected areas[26,27].

However, persistent stress as for example chronical osmotic pressure by hyperglycemia which is a side effect of diabetes mellitus leads to lasting impairment of this microenvironment. This is especially fatal to ECs in small capillaries as in the retina and leads to diabetic retinopathy that causes blindness on the scale of the organism[28,29]. The current treatments of diabetes are multiple daily insulin injections that batch wise reduce hyperglycemic conditions or the use of an insulin pump that allows a more continuous subcutaneous insulin infusion[30]. However, compared to a cellular response of pancreatic cells in a healthy body, both treatments are less sensitive adjustments and do not allow a comprehensive cure and endothelial protection of chronical stress[31]. Therefore, direct transplantations of pancreatic islet and their encapsulation in devices are investigated to meet these needs[32,33]. Since glucose measurement and is insulin released directly into blood flow, pancreatic islets are highly vascularized and β-cells in contact to capillaries. On this account and for the general nutrient supply a transplantation of vascularized tissue is also increasing the efficiency of graft survival and integration in the donor cardiovascular system[34].

Further, in the progression of tumor growth, angiogenesis is indispensable for nutrient support[35]. Unlike angiogenesis in development, cancer angiogenesis occurs in an unstructured or chaotic manner. Nevertheless, VEGF signaling is a major pathway in tumor angiogenesis[36,37]. Thus, treatment with anti-VEGF antibodies is a proven method of countering tumor progression[38]. Besides growth rate acceleration, tumor vascularization also introduces the formation of metastasis and spreading of cancer cells all over the vascular system[39].

1.4. Stem cells for EC research

The limited availability of ECs for the treatment and modeling of disease with EC involvement necessitates the different sources. The use of stem cells and their differentiation to the cell type or purpose. Thus, stem cell therapy offers prospects in curing diseases that relate to the loss of tissue specific cells that cannot regrow from the surrounding cells. By applying stem cells or stem cell-derived cells to the location of cell loss, a restoring of the healthy state is aimed. Beside their usage in the already mentioned pancreatic islet replacement, heart scar or cartilage repair as well as bone marrow transplantation as a treatment of leukemia are areas of application[40]. Stem cells exist in different types. In the adult body, organ specific stem cells are able to differentiate towards a small number of related cell types. An example here would be the regular renewing of skin or growth of hair but this mechanism is not available for all cells and tissues as the loss of β-cells in type I diabetes cannot be replaced by the body.

The umbilical cord blood of a newborn contains stem cells that have an increased potential of regeneration. These are capable of a full hematopoietic reconstitution and can be used for a bone marrow transplantation or even treat cerebral palsy[41,42]. This highlights that with moving backwards in development, cell fates are less defined and have a higher potency to differentiate towards multiple cell types. The blastocyst state develops five to twelve days after fertilization and is a partly hollow sphere that has formed out of trophoblast cells, which contains the inner cell mass (ICM). While trophoblasts form the embryonic portion of the placenta, the ICM develops the actual embryo. Cells of the ICM are pluripotent (Latin, “capable of many things”) therefore in principle used to differentiate into any kind of desired cell type. The only higher level in potency is totipotent (Latin, “capable of all”) that refers develop not only the embryo but also extraembryonic tissue and defines cells from the zygote to the morula (16-cell state)[3]. The highly potent cell types share the characteristics of existing only in early stages of embryonal development wherefore the usage is coupled with their extraction in these early stages and extermination of the developing organism. The need for huge amounts of stem cells in the research of personalized medicine and disease models *in vitro* clashes with the ethical aspects of ending the life of potential humans growing out of embryonic stem cells.

Vast research on understanding the plasticity of body cells has brought up the question whether cells can be re-programmed towards a state of pluripotency. In 2006 the group of Yamanaka published their work in which they had induced pluripotency in mouse fibroblast[43]. While one year later, they as well as the group of Thomson experimentally approved the reprogramming for human fibroblasts[44,45]. The induction was performed by testing a big set of transcription factors and two sets of TFs were identified that enable the re-programming to human induced pluripotent stem cells (hiPSCs): OCT4 (POU5F1), SOX2, KLF4, and MYC by Yamanaka and POU5F1, SOX2, NANOG, and LIN28A by Thomson. This development defused the ethical debate about the use of embryonic stem cells and brought new momentum to stem cell research. Further, the developmental biology flourished with the opportunity of using iPSCs in emerging stem cell therapy. Taking into account the embryonal microenvironment of individual cell types, differentiation protocols were established by adapting specific extracellular matrix (ECM) components as

well as signaling molecules, present in the particular step of development. This is of high interest in personalized stem cell therapy and in the generation of accurate *in vitro* models for disease of drug testing. In all these approaches, a stable generation of preferably mature cell types, organoids or tissues is a field of scope.

1.5. Endothelial development *in vitro*

For the reproduction of the endothelial lineage, key stages have to be passed through which are oriented at the embryonal development. Approaches to gain this focuses on the design of a beneficial microenvironment which can be obtained by addition of morphogens to the media, a viral transduction, or by a co-culture with other cells. In the co-culture stromal cells have been established to induce human embryonic stem cells towards endothelial linage[46]. Upon lentiviral transduction, an overexpression of modified mRNAs of the transcription factor pair TAL1 and GATA2 or rather GATA2 and ETV2 is also able to cause the differentiation from hiPSCs toward ECs[47]. The usage of morphogens to generate ECs exists in various ways and was reviewed by Xu *et al*[48]. A widely applied protocol uses the two stage induction of first mesodermal and in a second step endothelial linage[49]. The first uses of BMP4 for induction of the mesoderm[50] which is paired with the inhibition of the Wnt signaling pathway with CHIR99021. In the second phase, the endothelial linage is induced by adding VEGFA and forskolin to the media. Patsch *et al.* designed their protocol for the traditional 2D monolayer cell culture differentiation while Olmer *et al.* showed a successful translation to 3D by using low attachment wells and shaking of assembled aggregates in suspension[51]. This 3D suspension culture gives new possibilities for a large-scale generation of patient specific ECs from hiPSCs.

The general approach for measuring the differentiation success focuses on several markers and the usage of flow cytometry. Methods with deeper resolution would give deeper knowledge to the state of the generated ECs. This is of special interest in clinical usage where a flawless quality of ECs is indispensable. One method for obtaining deep insights into cellular states is single-cell mRNA sequencing (scRNAseq) that will be discussed in the following chapters. For the differentiation from embryonic stem cell to EC in a 2D cell culture format, there are already published scRNA-seq datasets that confirm generation of ECs *in vitro*[52]. In contrast to tissue specific ECs, these stem cell derived ECs do not show a direct correlation to any tissue[53]. In embryonal development, with the differentiation from angioblasts to EC the specification into arterial or venous tone is made[54]. However, the establishment of these phenotype in 2D cell culture is questionable since the development to mature ECs is initiated by the establishment of vascular networks which is dependent on a supportive microenvironment, including a matrix structure and cellular signaling. The *in vivo* organization of ECs is guided by surrounding cells and the signals these cells send. This concept has to be transferred to *in vitro* models in order to facilitate the formation of vessel structures.

In bodies with an established cardiovascular system, new vessels emerge from already existing ones in a process termed angiogenesis. Without that, a process termed neovascularization occurs. In embryonal development, blood islets, developed in the mesoderm, coalesce to form a primitive vascular plexus, a primary structure of microvessels[9]. This process can be reproduced by embedding stem cell-derived ECs into a suitable matrix environment[55]. A further interesting example is that after an injection of stem cell-derived ECs into zebrafish embryos, they generated vascular structures which even integrated into the already developed vascular system of the organism[56], whereas the experiment with HUVECs displayed a formation vessel structure that remained separated of the embryo vasculature. To study the underlying mechanisms and further capabilities of stem cell-derived ECs, *in vitro* cell culture systems in a 3D environment are methods to achieve these goals.

These environments need to allow a diffusion of fresh medium, oxygen and cellular waste. Additionally, both stability and flexibility for motility and expansion is required for healthy cell growth. An appropriate scaffold allows ECs to organize and synthesize their ECM. This structural orientation allows the establishment of new structures and contacts between the surrounding cells. Different hydrogels match this purpose and offer a suitable scaffold for the growth of ECs alone or in co-culture. Here, Matrigel, a collagen type IV and laminin rich polyglycan of Engelbreth Holm Swarm sarcoma, collagen type I or a mixture of both are commonly used in cell differentiation and cultivation[55,57]. While Matrigel contains additional growth factors promoting cell proliferation and survival, pure collagen I hydrogels constitute the matrix for cellular adhesion only. Both represent an environment in which ECs evince proliferation and sprouting, but they are non-synthetic produced components and underlie batch effects. In addition, a chemically defined microenvironment and stability is not provided as needed for clinical research. This are reasons why hydrogels are under development that consist of synthetic materials only and match these criteria[58]. Thereby, difficulties arise to identify matrix characteristics that support neovascularization[59]. A variety of vascular models on a chip base were developed to reproduce specific microenvironments[60–62]. Also, the formation of lumens was achieved in synthetic matrices[63–65]. One step further is the generation of perfusable endothelial lumens inside synthetic hydrogels that allows nutrient and oxygen supply throughout vasculature and a vascularization of large-scale organoids or tissue structures without appearance of necrotic cores[66].

1.6. Analytics of EC development

All new designed methods have to be analyzed for their reliability and fit for purpose. The standard technique for a quick differentiation control is flow cytometry, however, it offers information about several proteins of interest only. Since the purpose of this work goes beyond a differentiation protocol and includes prolonged 3D cultivation under different conditions, a more detailed and untargeted analysis method is appropriate. One analysis method that expanded in the recent decade that fulfills these demands is single-cell mRNA sequencing. The foundation for this method was laid more than half a century earlier with understanding the mechanistic of how DNA is replicated. Identifying the responsible enzymes and

applying their function to amplify DNA *in vitro* provided further possibilities in molecular biological research[67]. The first usage decoding of DNA sequences with an analytical method took place with the invention of the sequencing approach by Sanger in 1973 that revealed new horizons in molecular biology[68]. It requires the purification of a high number of identical DNA molecules to provide clear results. This excludes applications that generate large numbers of widely varying and complex DNA sequences that would be necessary to resolve whole genomes of organisms[69,70].

1.6.1. Next generation sequencing

During the early 2000s several approaches of how to sequence single DNA molecules were developed and commercialized. The market leader nowadays is Illumina with its fluorophore-based strategy. Hereby, the core consists of high-resolution cameras and reversible protective groups on single deoxyribonucleic acid nucleotides coupled with four different fluorophores. The technique depends on providing specific adaptor sequences, namely P5 and P7, to the ends of the DNA molecules, in a process termed library construction. A constructed library can then be loaded to the flow cell where the actual sequencing is conducted. Prior to the loading process, the DNA was denaturized and can hybridize with its free adaptor to the reverse complementary sequences that are immobilized on the flow cell surface (**Figure 1.2a**). An isothermal amplification generates the second strand which is then covalently liked to the flow cell surface (**Figure 1.2b**). In an additional denaturation step, the original template is washed away. Brownian motion leads to a hybridization of the strand's free end with neighboring P5 or P7 oligonucleotides, respectively (**Figure 1.2c**). In a so termed bridge amplification, the complementary strand is synthesized and forms a double stranded bridge (**Figure 1.2d**). Multiple cycles of bridge amplification and denaturation generate clusters of identical DNA strands (**Figure 1.2e**-**g**). In the last preparation step, one type of strands, depending on the read direction, is cleaved and washed away, and three-prime ends are blocked to prevent unwanted priming. In the actual sequencing, the flow cell is filled with the four dNTPs coupled to a specific fluorophore and containing a protection group on their three-prime end. A polymerase can only integrate the single complementary nucleotide. The flow cell is washed, and an image is taken representing the cluster fluorescence that identifies the incorporated base (**Figure 1.2h**). After the cleavage of the three prime protective group, another cycle can be performed. Illumina offers sequencing kits of up to 300 cycles for different sequence lengths as well as library complexity and sequencing depth. The difference between the paired- and single ended sequencing strategy is based on whether the read cycles are only performed from one or both sides and depends on the library construction.

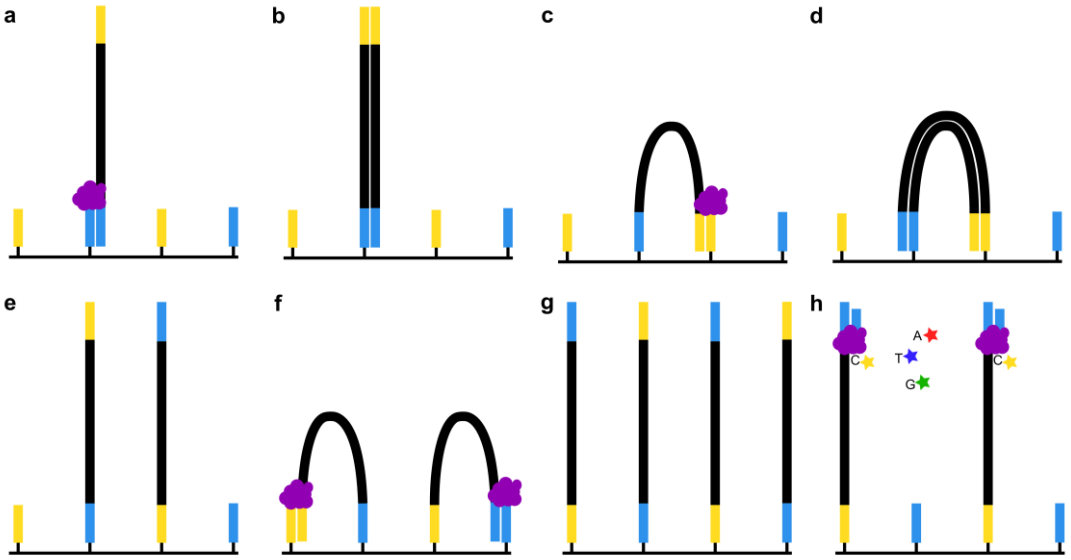

Figure 1.2 | Scheme of the Illumina next generation sequencing approach. Consisting of **a**, flow cell binding, **b**, an initial amplification, **c**, **d**, bridge amplification, **e**, **f**, **g**, cluster generation, and **h**, sequencing by synthesis. The figure is based on descriptions of Illumina, Inc.[71].

Besides the one of Illumina, there are multiple other sequencing platforms for NGS. The system of Helicos worked with the similar principle as Illumina but without the bridge amplification[72]. It represented the first commercial system, however the company became insolvent in 2012. With IonTorrent, Thermo Fisher offers a well based sequencing technology that uses beads that catch a single processed sequence which is amplified in a bridge amplification until the bead is covered. The beads with the different sequences of the library float into wells on a chip where the actual sequencing occurs. Unlike the fluorescent probe on reversible termination nucleotides which were used in the previous techniques, here the current is measured that is set free as a proton by the polymerase during formation of the phosphate backbone of the amplified sequence. PacBio commercialized a technique for sequencing of reads up to 25 kb. Key features here are immobilized DNA polymerases in microwells of a chip and the amplification of circular templates [73]. The MinION from Oxford Nanopore is a portable device for real-time sequencing[74]. It measures current differences by the amplification of a single DNA molecule that is pushed through a protein nanopore during sequencing. The multiplexing capacity totals to 512 pores on a device. These are all sequencing by synthesis approaches which include the amplification of the desired sequences to generate the read. The approach that is driven by polony DNA sequencing or commercialized as Sequencing by Oligonucleotide Ligation and Detection (SOLiD) by AppliedBioscience is a bit different[75]. While the preparation of sequence covered beads resembles the IonTorrent approach, the usage of ligation steps in five off sets for the sequence varies from the other stated NGS techniques.

These techniques are designed for the analysis of multiple different DNA sequences. However, in a differentiation process, we start from one cell type which is why no difference of the genome is expectable. Detectable changes that define the cell phenotype are defined by the accessible genome sections and the mRNA molecules that are transcribed from these. The mRNA molecules have to be reverse transcribed

and amplified to obtain a double stranded complementary DNA (cDNA) pool that can be processed to the sequencable library format.

1.6.2. RNA sequencing

This brings us to RNA or more specific mRNA sequencing that takes advantage of the mRNA-characteristic poly (A) tail to create a snapshot of what a cell population is transcribing from their genomes (**Figure 1.3b**). To this effect, poly (T) sequences coupled to magnetic beads, for example, can be used to capture the tails in an untargeted manner. A purified cell population is lysed and the whole amount of mRNA transcripts, the transcriptome of each cell, is collected as a bulk. Thereby, conclusions about which genes are active and in what amount a transcription takes place can be drawn. A pervious cell sorting step enables the mRNA isolation of different subpopulations based on defined surface markers. Since mRNA forms the link between the genomic information of a cell and the phenotype (**Figure 1.3**), mRNA sequencing provides insight into the actual state of a cell population. Consequently, this technique provides way deeper analysis with a higher throughput than an investigation of cell proteins *via* staining.

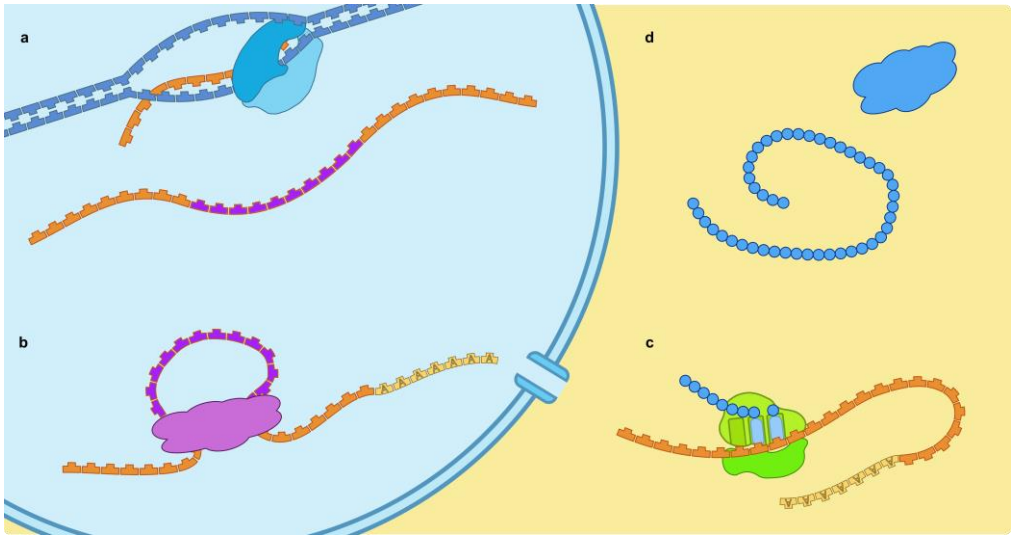

Figure 1.3 | Simplified process of protein biosynthesis that highlights the meaningfulness of mRNA. a, A specific RNA-polymerase recognizes gene encoding regions in the genomic DNA (blue) and transcribes a mRNA (orang and purple) from it. **b**, This contains non-coding sequences (purple) that are cleaved of in the process of splicing. Moreover, a poly (A) tail (yellow) is attached before the mRNA is transferred from the nuclei into cytoplasm. **c**, Ribosomes translate the mRNA sequence into an amino acid sequence (light blue). **d**, These polypeptides fold into proteins.

Even though, mRNA sequencing provides broad key insights about the regarded cell population, another level of magnification is needed to distinguish between subpopulations inside a sample. Such subpopulations are expectable during a differentiation and emerge through shifts in the developing cell population. These differences would be blurred by the collective transcriptome analysis of a cell population. Therefore, a transcriptome analysis on the level of individual cells is essential to dissolve these transcriptional changes in subpopulations.

1.6.3. Single-cell mRNA sequencing

Single-cell mRNA sequencing (scRNA-seq) is an updated version of mRNA sequencing that enables the transcriptomic readout of single cells and therefore allows the analysis of cell heterogeneity inside of populations, which introduces another layer of information. Several techniques were developed to answer questions of how to singularize cells prior to transcriptome capturing and processing the reverse transcribed mRNA towards a library for NGS. The common aspects here are that mRNA is targeted at its poly°(A) tail using a poly°(T) sequence, to each individual transcript a barcode is attached that relates it to the originating cell, and a unique molecular identifier (UMI) is incorporated allowing to subtract amplification bias during sample processing steps. These three elements are core characteristics of scRNA-seq.

One early method that was initially published in 2012 is smart-seq and represents a scRNA-seq method that has a main focus on full-length transcriptome sequencing[76]. The name originates from switching mechanism at 5' end of the RNA transcript (smart) and covers the treatment of mRNA only therefore the protocol starts after the singularization of the cells. The capture of full-length transcripts is facilitated by the usage of transposase and attachment of the sequencing adaptors for NGS to all generated fragments. This generates two types of sequences, on one hand sequences that refer to the individual cell and transcriptome (UMI reads) and the transposase digested fragments inside of the fragments (internal reads). After sequencing, the UMI reads enable cell and gene count while internal reads can be used to infer the allelic origin and identifying specific isoforms of mRNA transcripts. The second version, smart-seq2 improved protocol sensitivity and accuracy as well as the full-length coverage across transcripts[77]. In 2020, the third version was presented that has an about 10% increased cell-to-cell correlation compared to smart-seq2[78].

A different method to separate individual cells can be achieved using a microfluidic chip platform. This technique of scRNA-seq that was published by Macosko in 2015 and termed Drop-seq, used microfluidics to generate nanoliter droplets for the encapsulation of individual cells[79]. Transcriptome capturing occurs in the droplets by using microparticles that carry a barcode, UMI and poly (T) tail on their surface. During cDNA amplification also template switching is applied, similar to the smart-seq. The Drop-seq method enables the recovery of around 50 000 cells divided into eight separate libraries. 10x Genomics has commercialized the Drop-seq method to a broadly used pipeline for scRNA-seq by replacing the microparticles with "Gelbeads-in-Emulsion" (GEM, **Figure 1.4**)[80]. Their Chromium Next GEM Single Cell 3' kit uses the same cell transcriptome barcoding and droplet generation strategy.

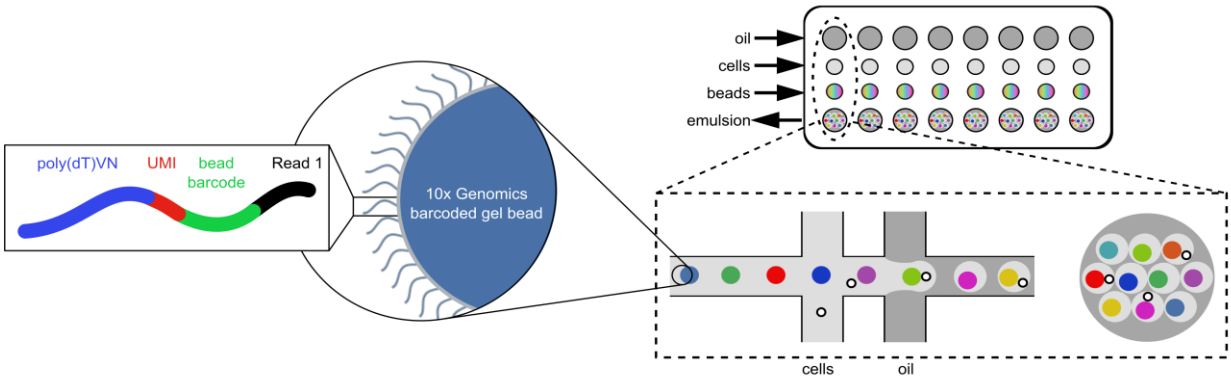

Figure 1.4 | Schematic of 10x Genomics gel bead and Chromium Next GEM Single Cell 3' chip loading and microfluidic process of droplet generation. Gel bead solution and cell suspension are encapsulated inside oil to generate an emulsion with separate compartments for mRNA capture, unique barcoding, and reverse transcription for the single cells individually. Barcoding is enabled by the bead specific barcode and a unique molecular identifier (UMI). Eight samples may run on a chip simultaneously. (modified from 10x Genomics, Manual Part Number CG00052, Rev D and Zheng *et al.* [80])

Since every captured mRNA sequence is labeled uniquely during reverse transcription, droplets are dissolved afterwards, and all sequences are processed together. The sequence length distribution of the amplified cDNA peaks at around 2000 bp whereas to annotate specific mRNAs to a reference genome, only a partial sequence is sufficient. In addition, every cycle is costly and with a length above 300 bp sequencing quality is reducing severely which is also represented by the available Illumina sequencing kits. Therefore, the cDNA is enzymatically fragmented. The here used enzyme is predominantly digesting longer sequences and therefore shortening the sequence pool over time to the desired length distribution. After the adaptor ligation, an indexing barcode to distinguish between different libraries as well as the adaptors P5 and P7 for flow cell attachment are attached by PCR amplification to complete the library construction (compare **Figure 1.2** and see **Figure 2.1** for a scheme of the individual steps of library construction).

A further technique for scRNA-seq is CytoSeq which has an array of picoliter wells as core element[81]. The array consists of polydimethylsiloxane (PDMS) and is imprinted from a photolithography generated mold. Cells for sequencing are applied on the platform and get captured in the wells while mRNA capture and barcoding are facilitated by magnetic beads. The singularization of the cells is achieved by the size of the wells and the dilution of the cell suspension.
In the initial forms of Drop-seq and CytoSeq, the barcode sequence of the beads was generated by combinatorial barcoding, to generate a high enough variability in unique sequences to individually label single cells. In contrast to that this, the split-pool ligation-based transcriptome sequencing (SPLiT-seq) approach turns this around by directly ligating barcodes to cell transcriptomes[82]. SPLiT-seq uses the combinatorial barcoding by applying several rounds of dividing cells in pots and pooling them again. Balancing the number of cells and total barcodes over the rounds of splitting and pooling vanishes the probability for cells with duplicated barcode. In the SPLiT-seq protocol cells are fixated by PFA to keep

mRNA transcripts inside of each cell while a permeabilization enables access of enzymes for the barcode tag generation, a mandatory step for having the similar barcode combination for the transcriptome of the respective cell. Its strength lays in having no requirement for specialized devices as microwell plates, FACS, or microfluidic platforms and the number of cells that are barcoded. This high number of barcoded cells exceeds the throughput of single experiments of the other presented platforms, results in a cost reduction for the transcriptomes of single cells and gives a further alternative.

1.6.4. Analysis of scRNA-seq data

After the readout of the transcriptomic data of single cells by NGS, a large pool of sequences has to be traced back to its original cell. This is facilitated by several DNA barcodes (BCs) incorporated during library preparation. The first layer, indexing BCs distinguishes between multiple samples/libraries that were sequenced on the same flow cell. The second BC refers to the specific cell – bead BC on the 10x Genomics or the combinatorial BC tag in the SPLit-seq approach. A third one is the unique molecular identifier (UMI) that allows the subtraction of amplification bias during sample processing. The last step is the annotation of the single transcriptome sequences to a reference genome and the generation of a count matrix, containing the single cells on one and the expressed genes on the other axis. To reduce this high dimensional space and make it representable, principal component analysis (PCA) is used to group cells that have similar gene expression patterns together[83]. This linear approach only is capable of visualizing major differences in gene expression between cells and distances but still draws direct conclusions concerning the distance relations between the cells projected in 2D, however, slighter differences between cell populations are not visually separated. Therefore, non-linear algorithms as t-distributed stochastic neighbor embedding (t-SNE) or uniform manifold approximation and projection (UMAP) are used to resolve these minor transcriptomic differences between cell populations[84,85]. Both of them are applied on a chosen number of PCAs to achieve their dimensional reduction model. Here, the distances between the cells do not refer to the difference in the gene expression in a linear way.

Besides algorithms for visualization, several others are used for assigning cells to clusters based on their expression pattern[86,87]. Louvain and Leiden, named after the cities where they were developed, are the ones mainly applied on scRNA-seq datasets. To determine the actual cell type or function behind the determined clusters, a method termed differential gene expression analysis is applied to them. This identifies genes that are predominantly expressed in the individual clusters and ranks the genes by generated p-values. Hereby, common marker genes pop up that allow the identification of certain cell types while further DEGs can refer to additional cellular characteristics of subclusters.

To bring structure into and revealing output from the multitude of genes, they may be annotated to gene ontology, a semantic repository to provide systematical order and relevance[88]. The commonly used method of GO term enrichment analysis uses the cluster specific DEGs and ranks the GO terms by highest match to the respective term. GO terms can consider biological processes, molecular function, and cellular component. A slight variation of this structure of knowledge is the 12urr12 encyclopedia of genes and

genomes (KEGG) database that focuses on pathways with the linked genes, gene products, and molecules[89]. Moreover, the transcription factor enrichment analysis (TFEA) follows a similar process by determining transcription factors (TFs) that are reported to regulate the DEGs. The number of genes annotated to the individual TF is reflected in its score. All these techniques aim to improve the amount of interpretable information out of multidimensional datasets and link it to our understanding of cellular mechanisms.

A further approach in expanding possible predictions of scRNA-seq output is achieved by including mRNA dynamics. Since the abundance of a transcript gives a static snapshot at a certain time point, it allows only limited statements about time-resolved phenomena and cell fate decisions. The concept of RNA velocity revolutionizes this and enables to describe trajectories of cell differentiation[90]. Since scRNA-seq measures two main stations in the life cycle of mRNA, namely transcription and splicing (**Figure 1.3b** and **Figure 1.5a**), algorithms can infer dynamics between these two processes[91]. By comparing the ratio of unspliced versus spliced mRNA of an individual gene, the time point of induction or repression of the respective gene can be determined, and the cell assigned to a certain state in the gene expression cycle (**Figure 1.5b** and **c**). Next, it is possible to identify genes that drive the differentiation process or cell fate, termed dynamical driver genes (DDGs). These facilitate an annotation of cells to a hidden or latent time in their dynamic development.

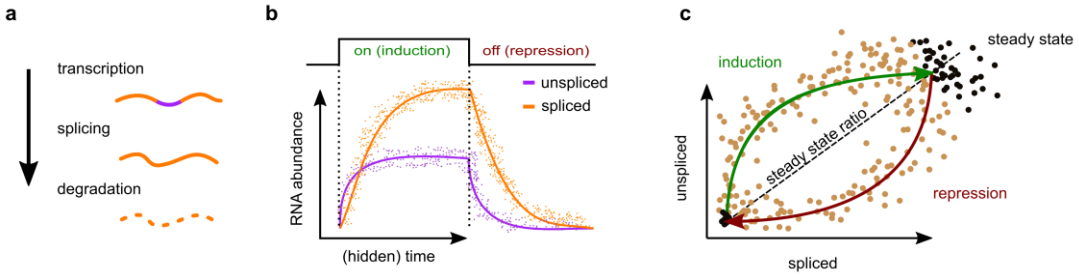

Figure 1.5 | Dynamics of the mRNA life cycle in cells. a, Schematics of the three main mRNA processing steps. **b**, Chronological abundance of un- and spliced mRNA in a cycle of gene expression induction and repression. **c**, Scatter plot of spliced versus unspliced mRNA to highlight the steady state of constant gene expression as well as the ratios of induction or repression phase. The figure is based on La Manno *et al.*[90].

1.6.5. Interaction Profiling on the basis of transcriptome data

Besides the use of scRNA-seq data to model cell trajectories, the prediction of cell-cell interactions is possible by linking the transcriptome to the involved ligand-receptor pairs. This offers an approach to allow statements about the cellular signaling states. Cells are always in contact with their environment and obtain signals *via* surface receptors and adapt their behavior or phenotype to these signals. This cellular crosstalk plays critical roles in biological processes including inflammation and cell development. Whereas scRNA-seq provides a comprehensive amount of data regarding cell types and expression of genes that are potentially involved in cellular signaling, multiple databases for protein-protein interactions and

regulatory pathways were built up. The goal is now to bring these different parts together to generate meaningful and predictive outcome. Several computational methods implemented this link between scRNA-seq data and protein-protein interaction databases.

scTensor uses tensor-based calculation to identify multiple to multiple cell-cell interactions by finding ligand -receptor pairs between different clusters inside a scRNA-seq dataset[92], these patterns for each cluster combination can be represented in hypergraphs afterwards. Ligand-receptor lists were built by collecting reported protein-protein interactions from multiple databases. The tool is completely conductible in R even though high memory is needed during tensor calculation with more complex datasets.

CellphoneDB is a further tool for discovering ligand-receptor interaction likelihood in scRNA-seq data and has a focus on more than two proteins that form complexes[93]. The algorithm uses the scRNA-seq count matrix and the set of cell type annotations to compare cell type expressions with their interaction database. The tool is designed for the Linux terminal and returns with a p-value and a rank for every interaction in their database for all annotated cell type combinations. A pipeline for selecting interacting cell types, filter for highest interaction potential and a representation of the data is available for R.

NicheNet takes a different approach by not using the direct count matrices of the scRNA-seq data but a set of DEGs specific for the cell type that is classified to receive signals[94]. These DEGs are regarded as potentially upregulated genes throughout ligand binding. The expression values are only used as a cutoff for minimal expression of the considered gene. While ligand expressing cells are defined as sender cell type, receptor and target gene expressing cells are considered as receivers. Multiple sender cell types can be considered in regulating a receiver cell type. To build the database on which NicheNet relies, multiple databases for ligand–receptor interactions, gene regulatory interactions, as well as signaling and protein-protein interactions were combined to generate on the one hand a ligand-receptor and on the other hand a ligand-target gene matrix. Both contain precalculated ligand-receptor interaction and ligand–target regulatory potential scores, respectively. Potential scores for the ligand-target gene matrix were validated with the data of 111 microarray studies. In these, the transcriptional response of ligand treatment was analyzed. By collating DEGs with the ligand-target gene matrix, valid target genes as well as ligands are predicted. Additionally, a ligand activity is calculated, scored by Pearson correlation to make a statement for the ligand meaningfulness.

1.6.6. Multi-omics

The goal of multi-omics is to extract more than just the transcriptome out of single cells but to collect additional signal besides the mRNA. This goes above the computational methods of establishing relations between mRNA expression data and databases. An example here is the droplet single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq)[95]. Thereby, not only the transcriptomes of single cells are collected but also the regions of genomic DNA identified that are less condensed in these cells. This gives insides into gene regulatory logic and epigenetic modality of the eukaryotic genome.

Additionally, the inclusion of spatial information improves analysis power of scRNA-seq. To this purpose, mRNA of tissue slides is annotated to defined special coordinates to generate transcriptomic landscapes. One possibility to achieve this is applying the tissue slide on a chip with spatially coded capturing spots as described by Ståhl[96]. This technique was adapted by 10x Genomics and perfectionated towards a commercially available spatial transcriptome sequencing platform. Deterministic barcoding in tissue for spatial omics sequencing (DbiT-Seq) is a further method for spatial transcriptomics and works with combinatorial barcoding, similar as SPLiT-seq[97]. It relies on formaldehyde-fixed tissue slides and a chip with multiple channels for applying the barcodes for the *in situ* ligation. Since cells are kept in place, splitting and pooling is skipped but barcoding cells on the slice *via* horizontal and vertical channels facilitates the spatial labeling in which the channel diameter reflects the resolution. However, the single-cell resolution in both techniques is replaced by the spot of mRNA capture or rather channel intersection of combinatorial labeling. The transcriptomes of this special coordinate are labeled the same and cannot be ascribed back to individual cells that were located there.
Moreover, the usage of antibodies to include protein detection into scRNA-seq experiments was implemented in a drop-seq method termed cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq)[98]. Thereby, an oligonucleotide labeled antibody is applied on cells. The oligonucleotide refers to the targeted protein and is captured along with the mRNA of the cell in the scRNA-seq pipeline. Presence and number of the CITE-seq signal provides information for the protein expression of the individual cell with a direct link to the transcriptomic data. Even ready to use antibodies from BioLegend are provided for the CITE-seq application. Also, DbiT-seq provides the incorporation of protein detection *via* usage of oligonucleotide tagged antibodies[96].

Antibodies are one of the mammalian immune system weapons against any type of extracellular microbes and microbial toxins. They appear as membrane bound receptors on B-lymphocytes as well as secreted antibodies in a soluble form, depending on their purpose. Antibodies are part of the globulin family, one of the two plasma protein families. The interchangeable term for antibodies, immunoglobulin (Ig) – derived from the Latin word *immunis* that denotes exempt or clean – refers to this affiliation[99]. In humans, the primary repertoire, the unique B-lymphocyte clones that are in circulation, amount to 10^8 while the genetically encoded variation is in the magnitude of 10^{11} antibody variants[100]. This enormous customizability to specific target molecules increased the interest in the usage of antibodies in multiple applications.
The aim for medical application led to the first industrial production of monoclonal antibodies facilitated by Milstein and Köhler in 1975[101]. In the following decades, monoclonal antibodies (mab) have risen to an effective therapy method against cancer, one of the most lethal diseases in the world[102]. Solely the cardiovascular diseases outpace cancer regarding the mortality in the western world as well as in developing countries[103]. More than fifty antibodies and antibody conjugates are meanwhile approved by the US Food and Drug Administration (FDA) and the European Medicines Agency (EMA)[104]. In cancer

therapy one strategy is to direct the immune system to the pathogen tissues and thereby supporting an active immune reaction by the body itself[105].

Moreover, due to their high specificity and binding strength monoclonal antibodies become indispensable for multiple biological applications, reaching from detection methods, as immunohistochemistry (IHC), immunofluorescence (IF) staining, or enzyme-linked immunosorbent assay (ELISA), over cell sorting approaches towards therapeutic applications. A major source for antibody generation is still the immunization of lab animals. In case human antibodies are desired transgenic – with their antibody encoding genes replaced by the human repertoire – animals are used. The chosen animal type is injected with a specific antigen and produces antibodies against it. To reach those antibody sequences the animal is sacrificed and B-lymphocyte storage tissues, *i.e.* spleen and lymphatic nods, are harvested. Genome extraction and amplification of the antibody containing regions and the generation of antibody libraries allow a screening, *e.g.* phage display or yeast surface display, for highly specific candidates[106,107]. After retrieval of a highly specific antibody, an adjustment for the specific assay is necessary. The first fluorescent antibody was engineered by Coons *et al.* in 1941 that enables the staining of its targets *via* microscopy[108]. Further conjugations include proteins, enzymes, small chemical molecules, metallic beads, and single stranded DNA (ssDNA)-oligonucleotides. The ssDNA-oligonucleotides give an easily amplifiable signal with DNA-polymerases.

To have a clean, biocompatible, and adaptable technique to label antibodies of choice, a used method has to be globally appliable on different antibodies. Here, the reactive side chains of amino acids, namely cysteine and lysine, are welcomely utilized. In antibodies free cysteines are rare wherefore the free amino group of the lysine side chains are used as a globally appliable method. Their functionalization is well characterized in the reaction with N-hydroxysuccinimide (NHS) esters in which a pseudopeptide bond or amide is formed under the release of NHS (**Figure 1.6**). The functionalization is mainly performed with a cross-linker in a high stoichiometric excess. Here variants of dibenzocyclooctyne (DBCO)-NHS-esters are commonly used (**Figure 1.6** (**2**)). The cross-linker contains DBCO on its second side which is a highly specific functional group for the addition of azides. The underlying reaction is the strain-promoted azide-alkyne cycloaddition (SPAAC), a biocompatible variant of the copper catalyzed click chemistry. The catalytic activity of Cu is replaced by the strained cyclooctyne[109,110]. The reaction with an azide-group facilitates steric discharge to the cyclooctene and represents the driving force for the highly specific covalent bond (**Figure 1.6**).

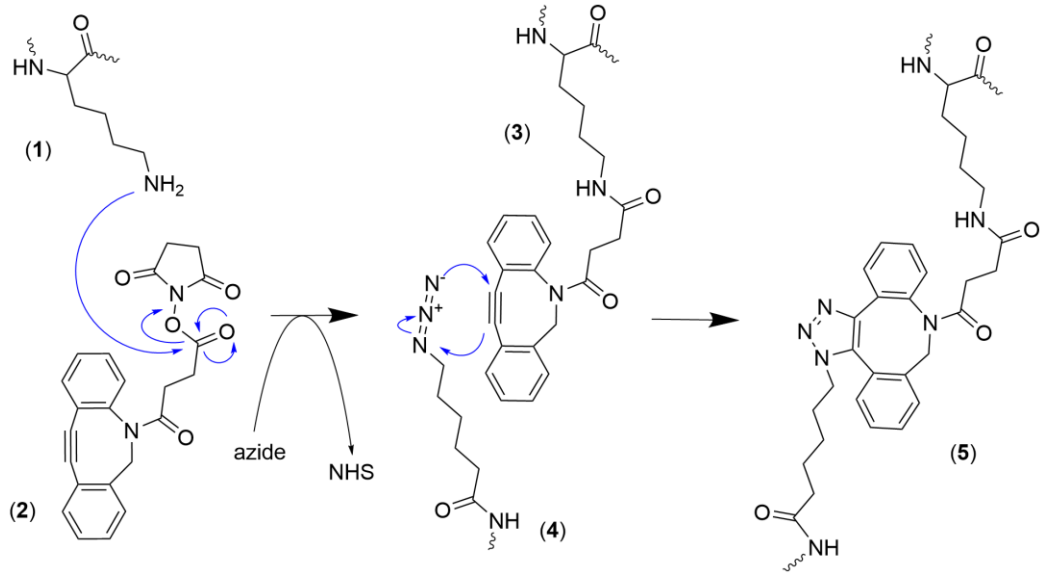

Figure 1.6 | Functionalization of antibodies and SPAAC to couple oligonucleotides. Reaction of a free lysin side chain from an antibody (**1**) with the DBCO-NHS-ester cross-linker (**2**) under the release of NHS (left). Functionalized antibody side chain (**3**) that reacts with an azide-functionalized oligonucleotide (**4**) in an SPAAC (center). Covalently cross-linked structure of antibody with an oligonucleotide (**5**, right).

The cleanup of the conjugates is necessary for an error-free and specific application. The stoichiometric excess of crosslinker is used to achieve high labeling ratio, however, leads to antibodies with multiple attached oligonucleotides as a side effect. This leads to unspecific binding events and disturbs applications of the oligonucleotide tagged antibodies[111]. Cleanup methods as the ion exchange chromatography overcome these issues by separating unconjugated antibodies, single and double conjugated ones, and free oligonucleotides. Multiple assays using these oligonucleotide-labeled antibodies in sequencing- or imaging-based detection were developed in the recent years[98,112–115]. Some of these assays which go beyond the detection of protein presence but provide statements about the neighborhood relations of the antibody-target proteins are presented in the following section.

1.7. Measuring proximity

Proximity increases the probability of interactions. This concept goes from gravitation between celestial objects over organisms with their sensory perception down to molecular and even subatomic forces. On the cellular level all molecules are in an ongoing entropy driven motion. Life can be seen as a mechanism of generating cavities to separate different environmental conditions and bringing molecules of interest in proximity to catalyze beneficial chemical reactions. The tools here are phospholipid bilayers, enzymes, and the DNA as persistent blueprint.

Therefore, it is important to be aware of the proximity of players behind a certain phenomenon to reveal its underlaying mechanisms. An example to achieve this is the coimmunoprecipitation that uses antibodies or labeled proteins to pull down their interaction partners[116]. This method is often coupled with a covalent cross-link reaction to sustain the interaction over the analysis *via* gel electrophoresis and western blotting.

A completely untargeted method is isothermal calorimetry that can recognize changes in the thermal energy level that occurs during formation of an interaction. Furthermore, the bio-layer interferometry and surface plasmon resonance are both optical techniques that use the attributes of changes in the reflected light on a surface under binding and dissociation conditions[117,118]. These assays are all measuring a direct interaction. In case of a cross-linker usage during the co-immunoprecipitation it has to be validated what exactly has been crosslinked in order to justify a direct interaction. A disadvantage of all these techniques is that the proteins of interest have to be isolated from their environment or even specially synthesized and purified for the performance of the assay. An *in situ* assay to measure proximity is the Förster resonance energy transfer (FRET)[119]. It is based on the emission-free transmission of energy from an excited fluorophore to a partner in close proximity. This transfer occurs on distances of up to 10 nm and decreases in intensity with the distance between the two partners, also stated as donor and acceptor[120,121]. A successful generated signal complies with the emission wavelength of the FRET acceptor fluorophore while the excitation wavelength of the FRET donor fluorophore is used as a trigger. The acceptor emission wavelength is only revealed in case of proximity to the donor whereas otherwise the donor emission wavelength is measured. With the perspective towards sequencing, the focus here goes towards assays that generate a sequencable amplification signal.

The proximity ligation assay (PLA) was developed by Söderberg *et al.* and generates a protein proximity signal[112]. The *in situ* method works with two ssDNA-conjugated antibodies that facilitate with a long- and a short-connector oligonucleotide, the formation of a circular ring – termed padlock probe, which is closed by a ligase (**Figure 1.7a**). The antibody bound oligonucleotides, termed adaptors, can span a distance up to 30 nm. The polymerase ϕ29 primes at one of the antibody adaptors and amplifies the padlock probe sequence several hundred times in a process termed rolling circle amplification (RCA, **Figure 1.7b**). The RCA product is a ssDNA aggregate having a diameter of around 0.4 µm and a repeating binding site for a fluorescent probe[122]. Applying this probe generates an enhanced signal that is detectable as sharp dots in fluorescence microscopy (**Figure 1.7c** and **Figure 3.30a**).

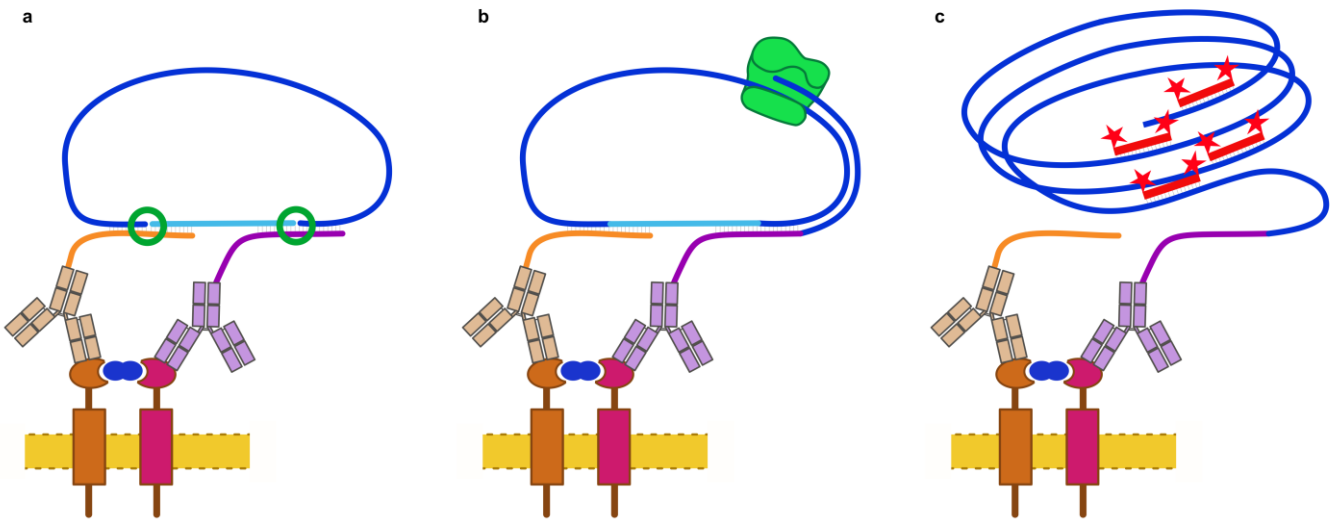

Figure 1.7 | Steps of the proximity ligation assay with adaptor labeled primary antibodies between two cell surface receptors. **a**, Hybridization of long and short connector (dark and light blue) to the adaptors (orange and purple) and ligation to the padlock probe. **b**, Rolling circle amplification of the padlock probe by a ϕ29 polymerase. **c**, staining with a fluorescent probe.

A further method to record proximity between proteins using hybridization and amplification similar as the PLA is the proximity extension assay (PEA). It uses two oligonucleotide-conjugated antibodies, however, their oligonucleotide sequences contain a hybridization side that acts as a primer sequence if both antibodies bond in proximity[123]. After fill-up and preamplification, different sequences are analyzed *via* quantitative (q)PCR. A modified version of the company Olink Proteomics couples the PEA with NGS which allows an increased throughput. Nevertheless, both analysis methods of the PEA are analogous and share a high correlation in direct comparison[124].

Auto-cycling proximity recording (APR) is a technique that uses hairpin probes to facilitate autonomous synthesis of DNA strands. It basically pairs the PEA with autonomous cyclic amplification of the proximity signal. The hairpin structures, also termed APR probes, allow the binding of a primer (**Figure 1.8a**, **i**). The amplification by using this primer molecules leads to a displacement of the original hairpin stem strand and the generation of the half-record (**Figure 1.8a**, **ii**). Brownian motion in turn supports the replacement of the amplified strand by the stem strand. In case of proximity to another hairpin molecule, a hybridization of palindromic sequence on the sequence ends occurs and in a further amplification step a full-record is produced (**Figure 1.8a**, **iiv**). The release of full-records restores the initial state of the APR probes and allows a new cycle (**Figure 1.8a**, **iv**). Both, full- and half-records, are released from the APR probes in a random manner and can be analyzed by gel electrophoresis or qPCR. The ability of cyclic half-record amplification allows a hybridization with not only one, but multiple APR probes located in proximity (**Figure 1.8b**). Thereby, the number of full-records directly correlates with the distance between the two APR probes.

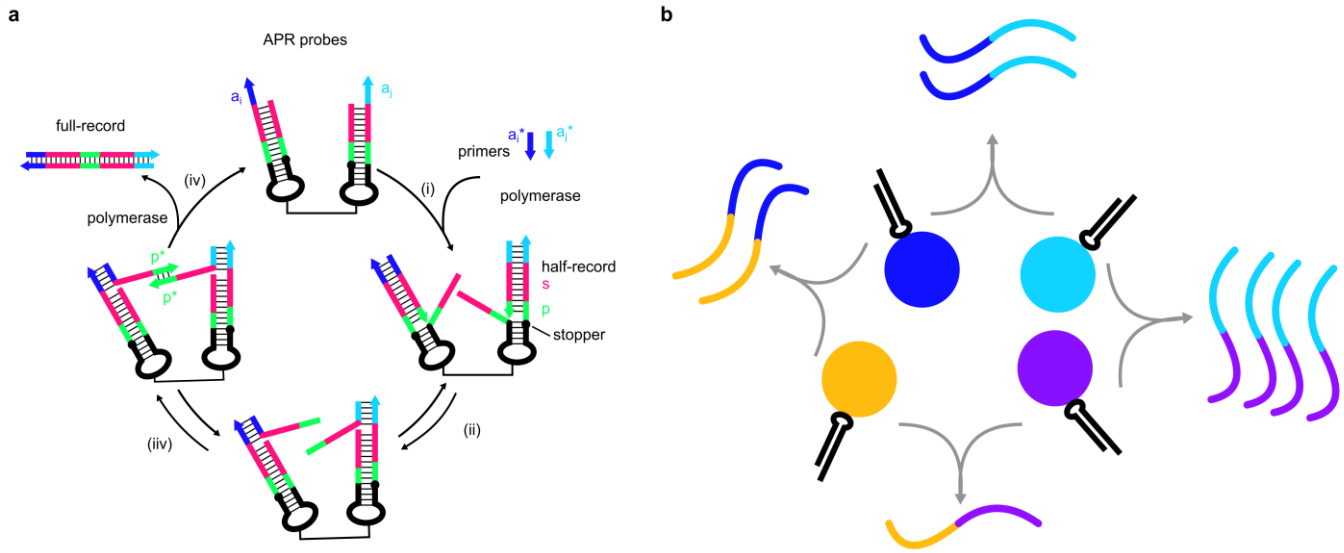

Figure 1.8 | Schematic of mechanistic APR procedure and recording of multiple reactions. a, Cycle of primer binding, step strand separation, displacement of amplified strand, hybridization of the palindromic sequences, and generation of the full-record with recovery of APR probes. **b**, Recording of multiple protein proximity relations. Number of generated sequences mirrors distance of between reaction partners. Reconstructed from Schaus *et al.*[125]

1.8. Objectives

The scope of this thesis is to reveal molecular mechanisms driving endothelial cell (EC) development and vessel formation within an *in vitro* cell culture system with the use of human induced pluripotent stem cells (hiPSCs). For this, I applied single-cell mRNA sequencing (scRNA-seq) techniques on ECs differentiated from hiPSCs in 3D suspension culture and on ECs during the process of neovascularization in an *in* vitro 3D hydrogel environment. Beyond the state-of-the-art scRNA-seq analytics, enhanced protein-based analytical tools and methods have been developed, to further improvement the readout of single-cell states in general. Two methods of adding a protein signal and protein proximity record to scRNA-seq were designed, performed, and are discussed in this thesis.

Chapter 3.1 covers the transcriptomic analysis of hiPSC-derived ECs using scRNA-seq. Starting with the monitoring of the differentiation in 3D suspension culture, cell type evolving, splicing velocities, and corresponding TFs are investigated. Differences of single-cell transcriptomes between the 3D suspension and a 2D monolayer culture, previously published by the McCracken and coworkers[52], were illuminated. Furthermore, transcriptomes of endothelial aggregates embedded into Matrigel as a 3D hydrogel environment were compared to the 3D suspension culture. Hereby, transcriptomic states in neovascular sprouting and formation of tubular structures were examined. Additionally, the influence of ascorbic acid, as a cofactor in formation of collagen bundles for basal lamina construction was taken into account. Several approaches highlighted the co-evolution of mural cells besides ECs and the plasticity in their cell phenotypes. Finally, scRNA-seq data after eight days of maturation in Matrigel was used to infer ligand-receptor signaling between mural cells that expressed pericyte genes and tube forming ECs. The pericyte character suggested a recruitment towards the formation of vascular structures therefore direct cell-cell communications became likely since these are required for cell type maturation on EC and pericyte side, respectively.

Chapter 3.2 describes the addition of a specific protein readout technology for single cells operating in parallel to standard mRNA-seq. Upon including protein expression information, I aimed at achieving a deeper understanding of the cell state, which is not necessarily correlated to the mRNA level. Moreover, protein detection by low mRNA expression and posttranslational or signaling-induced protein modifications are capturable in this manner. For that purpose, a combinatorial barcoding method was used that is termed SPLiT-seq[82]. The method is designed for scRNA-seq while it here was altered to allow an application to protein detection with oligonucleotide labeled antibodies. The whole approach led into the direction of sequencing a higher number of cells for lower costs. While the transcriptome read depth is reduced on the one hand, on the other further statements about surface and intracellular protein expression are provided.

The last result part, chapter 3.3 evaluates the PLA as a method for detection of interacting proteins. The application of the imaging-based method was presented on the interaction of the insulin receptor (Insr) and two members of the integrin family (Itgb1 and Itgb3) in mouse adipocytes. The underlying question was the influence of kindlin-2 on the interactions and their impact on insulin resistance as a major cause of type 2 diabetes. In a further step, the incorporation of a PLA variation into the 10x Genomics scRNA-seq pipeline was considered in chapter 3.3.2. This was facilitated by a digestion of the RCA amplicon in sequences that mimic the mRNA poly (A) tail and were captured additionally by the 10x Genomics bead sequences. This link of transcriptome and protein-protein interaction went beyond protein detection but gave insides into actually activated signaling cascades on a single-cell level.

2. Materials and Methods

The methods which are shared between this thesis and the publication[126] that forms the basis of this thesis were adopted.

2.1. Cell culture and differentiation

All used cell types were cultivated at 37 °C with a CO_2 atmosphere of 5%.

2.1.1. 2D hiPSC cell culture

HiPSCs were cultured on hESC Matrigel-precoated 6-plates according to manufacturer's recommendations (Corning) in mTeSR1 medium (Stemcell Technologies) with daily medium change and split twice a week in a 1:6 ratio using 0.05% Trypsin-EDTA (Sigma). The stem cell line was provided by Wang *et al.*[127] and is registered under the human pluripotent stem cell registry (https://hpscreg.eu) under the name HMGUi002-A, which includes all cell line details. The hiPSC culture were free of mycoplasma contamination as tested by the MycoSensor PCR assay kit (Agilent Technology). The general scientific use of the HMGUi002-A cell line was approved by the local ethics committee at the Technical University Munich (reference no. 400/21 S-KH).

2.1.2. 3D suspension culture

For the transfer into a 3D hiPSC cell culture, the medium was aspirated, cells were washed with 2 mL PBS -/- and 500 µL Accutase were added. During incubation at 37 °C for 3 to 7 min, cells detached while adding 2.5 mL mTeSR stopped the reaction. The well was washed with 1 mL mTeSR before centrifugation for 5 min at 200 x g and resuspended in 500 µL mTeSR with Rock inhibitor (Y-27632 dihydrochloride, Abcam, 10 µM) and 1% Penicillin/Streptomycin (P/S, Thermo). The cells were incubated in ultra-low attachment plates at 37°C and 100 rpm (Fisher).

2.1.3. 3D suspension culture differentiation to ECs

At day -1 of the differentiation, cells were transferred into low attachment 6 well plates in order to form embryonic bodies (1.5 x 10^6 cells/well). Differentiation was performed according to the protocols form Olmer *et al.* and Patsch *et al.*[49,51]. Briefly, from day 0 to 3 N2B27 medium supplemented with BMP4 (25 ng/mL) and CHIR (7.5 µM) was used without media exchange. Media was changed daily from day 3 to 7 using StemPro-34 with VEGFA (200 ng/mL) and forskolin (2 µM). Afterward, StemPro-34 with VEGFA (30 ng/mL) and FGF2 (30 ng/mL) was used with exchange after two days.

2.1.4. 3D microwell chip

For matrix-free cultivation, the microwell chip platform, described by Wiedenmann *et al.*, was used to allow the growth of 1196 aggregates in parallel[128]. The chip was coated with 10% Pluronic F-127 (Sigma)

overnight and seeded with approximately 1x10^6 cells. StemPro-34 with VEGFA (100 ng/mL), FGF2 (100 ng/mL), and FBS (15%) was used for cultivation. The medium was exchanged every second day.

2.1.5. Hydrogel cell culture

For cultivation in 24-well plates, 100 µL of Matrigel were added into each well of the plate that whole ground is covered and incubated at 37°C for 1 h. Aggregates were centrifuged at 800 rpm for 5 min at 4°C, resuspended in 80 µL of Matrigel and spread on top of the first layer. Gel polymerization was done at 37°C inside a container with a bit of ice to allow a slow temperature adjustment. After one hour 0.5 mL media were added. Media composition complies with microwell chip incubation. +AA samples contained 60 µg/mL of ascorbic acid. For imaging, the aggregates were embedded in ibidi slides with 10 to 20 per well. The amount of Matrigel and media was adjusted to the smaller volume.

2.1.6. Cell type stability experiments

On day six, aggregates were harvested and prepared for FACS, ECs (PECAM1$^+$, PDGFRB$^-$) and VSMCs (PDGFRB$^+$, PECAM1$^-$) were sorted. 2 x 10^5 of the sorted cell types and the unsorted mixture were seeded and cultured in 6-well plates coated with fibronectin bovine (LIFE Technologies). The medium composition was equivalent to the 3D differentiation.

2.2. Fluorescence-activated cell sorting (FACS)

Harvested cells were washed three times with PBS (200 x g for 5 min) and singularized using accutase. Five volumes of FACS buffer (10% FBS in PBS) were added and centrifuged at 300 x g for 5 min. Blocking was done for 20 min with StemPro34 + 10% FBS on ice. The cell suspension was filtered using a 70 µm nylon cell strainer (Corning). Live dead staining was performed with Trypan Blue while cells were counted. 20 µL of each antibody (FITC Mouse Anti-Human CD31 (BD Pharmingen™, 555445) and PE Mouse Anti-Human CD140b (BD Pharmingen™, 558821)) and 1 µL of violet fluorescent reactive dye (Invitrogen, REF: L34963A, LOT: 2179253) were added per 10^6 cells and incubated for 30 min at room temperature (RT). After washing once with FACS buffer, cells were resuspended in Tyto buffer (10^6 cells/mL, MACSQuant Tyto I Running Buffer – Milteniy (Cat. No: 130-107-207; Lot: 5200608355)). The sorting was done on a MACSQuant® Tyto™. For ensuring unspecific binding, FITC Mouse IgG1, κ Isotype Control (BD Pharmingen™, 554679) and PE Mouse IgG2a, κ Isotype Control (BD Pharmingen™, 559319) were used.

2.3. Flow cytometry

Singularized cells were washed once with PBS (centrifugation conditions: 300 x g, 5 min). Cells were fixated at RT for 15 min with 4% PFA in PBS and afterward washed twice with FACS buffer. 10^5 cells were transferred into a U-bottom-shaped 96-well plate (Greiner) and an antibody (2 µL per 15^5 cells) diluted in

100 µL FACS buffer was applied per well. After incubation for 30 min at RT, cells were washed twice with FACS buffer. For the measurement, the pellet was resuspended in 200 µL of FACS buffer and transferred through a cell strainer into FACS-tubes (5 mL round-bottom tubes, Coring). Flow cytometry analysis was performed on a MACSQuant® VYB.

2.4. Cryoembedding

For cryo embedding of 3D suspension culture aggregates, 500 µL of 4% PFA were added and incubated for 15 min on ice. The disc was detached from the walls with a needle. After two washing steps with PBS, an incubation with, first, 10%, second, 30% sucrose at RT for 2 h and, third, with a 1:1 mixture of 30% sucrose and OCT medium at 4°C overnight followed. All incubation steps were implemented on a wave shaker at 47 rmp. The medium was replaced by pure OCT and the sample was frozen on dry ice. The Slicing was done with a Leica CM1860 cryostat and a thickness of 20 µm. We used the following antibodies to prepare the immunofluorescence stainings: PECAM1 (Thermofisher, Ab WM59), PDGFRB (Thermofisher, AbAPB5), DLL4 (Cell Signaling, mAb 96406), DEPTOR (Cell Signaling, mAB 11816), CLDN5 (Thermofisher, MA5-32614), and ICAM2 (Cell Signaling, mAB 13355), and Collagen IV (Abcam, ab6586).

2.5. Fluorescence imaging

Slides were washed with PBS, permeabilized in PBS with 0.1% Triton for 30 min at RT, washed with 0.2% Tween 20 in PBS (PBST) and blocked with 2% BSA (Proliant) in PBST for 1 h. Antibodies were applied in the concentration as recommended by the manufacturer specifications in the blocking solution. After primary and secondary antibody staining, five washing steps with 5 min PBST were performed. Prior to confocal imaging (Zeiss Axio Observer LSM 880), Vectashield® Mounting Medium was added to the sample, it was covered and sealed by a coverslip.

2.5.1. Image analysis

IF and bright-field images were corrected for brightness and contrast with ImageJ. Z-projection of fluorescence images was performed using maximal intensity. ImageJ version 1.52p was used [129].

2.6. Workflow of scRNA-seq

2.6.1. Sample preparation for scRNA-seq

Matrix-free samples were washed with PBS, resuspended in Accutase and incubated for cell detachment at 37°C for 30 min. The reaction was stopped by adding five volumes of media. Afterward, cells were washed once with PBS. Prior to this process, Matrigel embedded samples were washed with PBS, and incubated with 0.5 mL Collagenase/Dispase Solution (1:100 to a final concentration of 1 mg/mL in StemPro34) for around 4 h, until organoids detached from the Matrigel. The enzymatic reaction was

stopped with 1 mL Neutralisation Buffer (1% BSA, 1% P/S in DMEM:F12). The single cells were cryo-preserved in DMEM with 10% heat-inactivated FBS (Thermo Fisher Scientific) and 10% DMSO based on a previously described scRNA-seq sample preparation protocol [130]. For sequencing, cryo-preserved cells were thawed in DMEM:F12. RNA libraries were generated using Chromium Single Cell 3' library and gel bead kit v3.1 (10x Genomics). The amplified cDNA library was sequenced on a NovaSeq 6000 S2 flow cell from Illumina. The sequenced cell numbers can be found in **Table 3.1**. (adopted from Rosowski *et al.*[126])

2.6.2. Concept of library preparation for scRNA-seq

To label mRNA molecules individual and access their sequence, scRNA-seq uses a sequencing approach that reads the respective information from one or the other side of the constructed library. The sequencing Read 1 is used to decode the cell barcode and the unique molecular identifier (UMI) and is part of the capturing sequences on the 10x Genomics gel beads (**Figure 1.4** and **Figure 2.1a**). The synthesis of the second strand during the reverse transcription needs a second primer for a subsequent PCR amplification. Therefore, the template switching oligo (TSO) is integrated to incorporate it as a primer binding site. The TSO uses the three cytosines, which are attached to the amplified sequence by the reverse transcriptase, to bind and is only needed in this first amplification step (**Figure 2.1b**). During the following enzymatic degradation, it is cleaved off (**Figure 2.1c**). The sequencing Read°2 is incorporated during end repair, A-tailing, and adaptor ligation (**Figure 2.1d**) The indexing PCR attaches the flow cell adaptors (P5 and P7) and the sample index (**Figure 2.1e**). Inside of an Illumina flow cell, sequencing from Read 1 gives the bead barcode and UMI that identify the associated cell and transcript molecule, from Read 2 returns the sequence of the captured mRNA, and the sample index relates to the library or rather lane of the 10x Genomics Chromium chip.

Figure 2.1 | Scheme of the different steps in der 10x Genomics library construction. **a**, The reverse transcription occurs inside the gel bead solution, directly after the microfluidic process (**Figure 1.4**) and includes the incorporation of the template switching oligonucleotide (TSO). **b**, cDNA amplification with partial Read 1 and partial TSO as primer. **c,** Enzymatic fragmentation. **d**, End repair and adaptor ligation to include Read 2. **e**, The index PCR attaches the P5 and P7 adaptors for binding to illumine flow cell while the 8 pb index allows identification of different libraries on the same flow cell. During sequencing, priming at Read 1 reveals cell/bead barcode and UMI in 28 cycles, i7 priming gives the library specification in 8 cycles and transcript readout occurs with 91 cycles priming at Read 2. The graphic is adapted and modified of the ChromiumNextGEMSingleCell3_v3.1_Rev_D manual (10x Genomics).

2.6.3. ScRNA-seq data pre-processing

Sequencing raw files were demultiplexed, aligned (reference genome hg38_ensrel97), filtered, barcodes and UMIs counted, and subjected to a quality filter with CellRanger (version 3.0.1, 10x Genomics). The pre-processing and downstream analysis were performed with the package 'Scanpy API' in python with default parameters, if not stated differently [131]. First, dead or stressed cells, identified by a percentage of mitochondrial genes higher than 10%, were filtered out. Next, cells with less than 200 and genes expressed in less than three cells were excluded. Afterward, the datasets of different days and experiments were concatenated, normalized to 10^4 gene counts per cell and log-transformed. Batch effects were corrected using ComBat. Further on, the highly variable genes were used for the downstream analysis. As discussed by Luecken and Theis, we corrected for the total gene counts, percentage of mitochondrial genes, and the cell cycle distribution of S, G2 and M phases to investigate differentiation-dependent changes on the transcriptome level [83].

2.6.4. Dimensionality reduction, clustering, and cell-type annotation

The single-cell nearest neighborhood graph was computed with the first 50 principal components and ten nearest neighbors. The cells were clustered with the Leiden algorithm with a resolution of 0.5. For visualization, the dimensionality of the data was reduced using Uniform Manifold Approximation and Projection (UMAP) [85]. For cell-type annotation, 300 DEGs for each of the clusters were calculated by ranking the clusters against all remaining cells with the t-test method. Clusters with proteasome-related genes scored at the top or a significantly reduced gene count were removed from the dataset as representing dying or damaged cells. The remaining clusters were annotated based on known marker genes.

2.6.5. RNA velocity through dynamical modeling

We analyzed the RNA velocity to investigate developmental trajectories by recovering directed dynamic gene information through splicing kinetics. Information like clustering and UMAP coordinates were retrieved from the Scanpy analysis. The pre-processing and downstream analysis were performed with scVelo using default parameters [91]. Splice variants and cells were filtered, normalized, and logarithmized with the function scv.pp.filter_and_normalize (parameters: min_cells=3, min_counts=200, min_shared_counts=20, n_top_genes=2000). The moments based on the connectivities were calculated with 40 PCAs and 10 neighbors in the next step. After recovering the dynamics, the latent time was calculated and the velocity was calculated as a dynamical model.

2.6.6. Integration of datasets from different sequencing approaches

For integration and correction of datasets from different sequencing runs, we applied bbknn to the datasets (neighbors_within_batch=10, n_pcs=40, trim=0, copy=True). We then reclustered the cells with the Leiden algorithm at a resolution of 0.8[86,132].

2.6.7. Transcription factor enrichment analysis

The identified cluster-specific DDGs or DEGs were entered in the ChEA3 web tool[133] and the mean rank was plotted using R.

2.6.8. Pathway and Gene Ontology enrichment

DEGs were filtered by their unique expression over all clusters (standard deviation above 0.5) and an expression value above 0.5. For the GO term enrichment, the R package enrichR was used with the “GO Biological Process 2018” database and plotted in R[134,135].

2.6.9. Implementation of CellphoneDB

The count matrix and cluster annotation were exported from scanpy, imported into R, and processed as recommended by the authors[93]. Cell-cell interactions were selected by the highest mean score and lowest p-value.

2.6.10. Implementation of NicheNet

As Target gene input, top 300 DEGs have been used. The calculation was done in R, converting the anndata element into a Seurat object. The process was performed as recommended by the authors[94].

2.7. Agarose gel electrophoresis

A 1.2% agarose solution in sodium borate (SB) buffer[136] was heated until agarose was dissolved and the gel poured into shape. Running was performed in SB buffer at 240 V for 20 min.

2.8. Antibody labeling

The antibody oligonucleotide coupling was performed as stated in optimized labeling condition for single oligonucleotide conjugated antibodies by Wiener *et al.* [111]. In short, antibodies and 3’-amino-oligonucleotides functionalized using 10-fold excess of DBCO-PEG4-NHS Ester (Sigma) and 3-Azidopropionic Acid Sulfo-NHS Ester (ClickChem Tools) for , respectively. Unbound cross-linkers were removed *via* dialysis in 7k MWCO Slide-aLyzer cups (LIFE Technologies) in 1L PBS, at 4°, stirring, overnight, and PBS was once changed after the first hour. Protein and oligonucleotide concentrations were determined by the Qubit™ 4 fluormeter. The click chemistry reaction was conducted by incubating the mix of DBCO-activated antibody with an 3-fold excess of azide-activated oligonucleotide for 6 h at RT. Single labeled antibodies were purified using the ion-exchange chromatography system, proFIRE® (Dynamic Biosensors) and fractions of single labeled antibodies were concentrated using Amicon Centrifugal 4ml 10K NMWL (Merck Millipore).

2.9. Antibody SPLiT-seq

On 1 x 10^6 cells of Min6 wild type and Iir knock-out 25 ng/µL of oligonucleotide labeled anti-Iir antibody were applied in antibody diluent (Epredia™ Lab Vision™) containing 10 u/µL RNAse inhibitor, murine (NEB). The SPLiT-seq Protocol, Version 3.0[82] was followed until cDNA amplification was conducted. Sequences to used oligonucleotides can be find in **Table 2.1**. The cDNA was transformed in libraries using the NEBNext Singleplex Oligonucleotides for Illumina® (NEB) while "Tagmentation and Illumina Amplicon Generation" was not performed since the aim was to sequence full length "cDNA" or rather the full antibody tag. WT and KO libraries received individual indices and were pooled equimolar to a final concentration of 4 nM. Concentration was determined *via* BioAnalyzer with the High Sensitivity Kit (Agilent) and Qubit®-4 with the dsDNA HS Assay Kit (Invitrogen) while the results were averaged. The whole library has a length of 233 bp while the excluding Read 1 and 2 amounts to 123 bp. For sequencing a MiSeq Reagent Nano Kit v2 (300-cycles, Illumina) and device read specifications of Read 1: 150, Read 2: 150 cycles, and i7 Index 6 cycles was used.

Table 2.1 | Sequences used for antibody-SPLiT-seq.

Sequence name	Sequence (5´- 3´)
Ab adaptor oSR321201_1	[AmC6]CAGCGTCAGATGTGTATAAGAGACAGAACGGGCTCACCTATTAGCGGCTAAGGCT
Round01_ab BC rt primer oSR321402o	AGGCCAGAGCATTCGAACGTGATtcCCTTAGCCGCTAATAGGTGAGC
Round02_barcode linker oSR321206_R2_bridge_BC_0215	CGAATGCTCTGGCCTCTCAAGCACGTGGAT
Round02_block oSR321207_R2_block_BC_0216	ATCCACGTGCTTGAGAGGCCAGAGCATTCG
Round03_barcode linker oSR321209_R3_bridge_BC_0060	AGTCGTACGCCGATGCGAAACATCGGCCAC
Round02_block oSR321210_R3_block_BC_0066	GTGGCCGATGTTTCGCATCGGCGTACGACT
Round02_barcode oSR321205_R2_01	[Phos]CATCGGCGTACGACTAACGTGATATCCACGTGCTTGAG
Round03_barcode oSR321208_R3_01	[Btn]CAGACGTGTGCTCTTCCGATCTNNNNNNNNNNAACGTGATGTGGCCGATGTTTCG
Read 1 primer oSR321215_part_R1_pr_ab	CAGCGTCAGATGTGTATAAGAGACAG

Underlined non-binding nucleotides to prevent priming and elongation.

2.10. Proximity ligantion assay (PLA)

Duolink™ In Situ PLA®

All incubation and wash steps were performed using BioShake iQ at 500 rpm. For washing, 50 µL PBS were applied per well for 5 min. Cells were permeabilized in TBS-T buffer (0.05% Tween) for 10 min, blocked with 2% BSA in PBS for 30 min and washed twice with PBS, respectively. Primary antibodies were applied for 1 h at RT in antibody diluent. Antibodies were used in following: mouse anti-insulin receptor (1:5, ThermoFisher CT-1), rabbit anti-integrin beta 1 (1:10, ThermoFisher, SA40-08) and rabbit anti-integrin beta 3 (1:40, ThermoFisher, SJ19-09). As a negative control, no primary antibody was applied. The PLA assay was implemented as stated in the manufacturer's protocol, using Duolink™ In Situ PLA® ProbeAnti-Rabbit PLUS, Anti-Mouse MINUS and the In Situ Detection Reagents Orange (Sigma, DUO92101). The staining solution had a final concentration of BSA 25 µg/mL, DAPI (Thermo) 1µg/mL, Lipitox 488 (HCS LipidTOX™ Green Neutral Lipid Stain (1:200, InvitrogenTM, H34775) in 1x SSC buffer. Cell wells were washed three times and analyzed by confocal imaging (Zeiss Axio Observer LSM 880).

Customized PLA

Incubation and steps were performed equivalent to the Duolink PLA. Secondary antibodies (goat anti-rabbit IgG and goat anti-mouse IgG, ThermoFisher) were labeled with the adaptor sequences stated in **Table 2.2** and applied in a final concentration of 25 ng/µL. Ligation and RCA mixes were prepared as stated in **Table 2.3** and **Table 2.4** using the enzymes and associated buffers provided by NEB. For the final staining step, a final concentration of 6.25 nM for the fluorescent probe was used.

Table 2.2 | Sequences for IntActSeq PLA. Colors refer to hybridization sites between adaptors, connectors, and the probe.

Description	Sequence
Adapror 650	[N_3]AAAAAAAAAAGAATGGAACCTCGCTAGAACGT
Adaptor Uni	[N_3]CAAAAAAAAAAATAGTTCGGTCGAAGTTAGTCCT
Long connector	[Phos]GACCGAACTATCTAGTGCTGGATGATCGTCCCCCCTGCACCTCAAAACA CCCTAACGTTCTAGCG
Short connector	[Phos]AGGTTCCATTCAAAGGACTAACTTC
probe	[Atto-647]CTAGTGCTGGATGATCGTCC[Atto-647]

Table 2.3 | PLA ligation mix.

	Final Concentration
400U/µl T4 DNA ligase	1 U/µl (1:400)
10x T4 buffer	1x
10 µM LC	125 nM
10 µM SC	125 nM
20 mg/ml BSA	125 ng/ml

Table 2.4 | PLA RCA mix.

	Final Concentration
phi29 polymerase	0.25 U/µL (1:40)
phi 29 buffer	1x
10 mM dNTP	200 µM
20 mg/ml BSA	200 µg/ml

2.10.1. IntActSeq PLA

Steps of the PLA were performed equivalent to the standard PLA (chapter 2.10) if not stated different. But performed with methanol fixated (Manual Part Number CG000136, Rev B, 10x Genomics) cells in suspension. All incubation and washing buffer after fixation contained 10 u/µL RNAse inhibitor, murine (NEB). Antibodies were applied in in antibody diluent (Epredia™ Lab Vision™). The primary antibody in this experiment was an anti IGF-I Receptor β (9750s, Cell Signaling) applied in a final concentration of 1:200 while the secondary antibody was a goat anti-rabbit IgG (H+L, 31210, ThermoFisher) applied in a final concentration of 25 ng/µL, conjugated with Adaptor 1 and 2, respectively (**Table 2.5**). The splint was incubated overnight on ice.

Table 2.5 | Sequences for IntActSeq PLA. Colors refer to hybridization sites between adaptors, connectors, and the probe. The underlined bases mark the MluI recognition site.

Description	Sequence
Adaptor 1	[N_3]GACAACAACAACAACAACAAAAAACTCTGGGGTATGCTATGT
Adaptor 2	[N_3]GACAACAACAACAACAACAATACCCGATTGAGCATTTGGACT
Long connector	[Phos]CCCAGAGTTTTNNNNNNNNNNCTAGTGCTGGATGATCGTCCTACACGC GTTTTTTTTTTTTTTTTTTTTTAGTCCAAATGC
Short connector	[Phos]TCAATCGGGTAATGACATAGCATAC
Splint	CGTCCTACACGCGTTTT

The scRNAseq was conducted using the Chromium™ Single Cell 3' Reagent Kits v2 (10x Genomics) and according to the user guide if not stated different (Manual Part Number CG00052, Rev D). The loading cell number for the *in situ* PLA aimed at a target recovery of 10 000 cells. During loading of the Single Cell A Chip, 10 µL of the 40 µL Single Cell 3' Gel Bead solution was replaced with Mull-HF (NEB) and prior to the reverse transcription a 30 min incubation step at 37 °C was included to facilitate the digestion. Subsequent to 10 cycles of the cDNA amplification, half of the sample was removed while the remaining half underwent a total number of 30 cycles. The two conditions were processed separately. In the post cDNA amplification with SPRIselect, a 2.5X reagent ratio was applied to include fragments down to a size of 100 bp. The SPRIselect steps during library preparation were conducted with a 1.5X reagent ratio. The

four generated libraries were pooled and sequenced with a MiSeq Reagent Nano Kit v2 (300-cycles, Illumina) and device read specifications of Read 1: 26, Read 2: 274 cycles, and i7 index 8 cycles.

2.10.2. Software specifications

The scRNA-seq alignment was run in CellRanger version 3.0.1 and the analyses were run in python 3.7.4 with Scanpy API version 1.4.4 or 1.5.1, anndata version 0.6.22 or 0.7.6, umap version 0.4.6, numpy version 1.17.4, scipy version 1.5.2, pandas version 1.1.3 or 1.0.5, scikit-learn version 0.22, statsmodels version 0.10.1, python-igraph version 0.7.1, scvelo version 0.2.1, matplotlib version 3.2.1, seaborn version 0.9.0, loompy version 3.0.6, XlsxWriter version 1.2.6, bbknn version 1.3.6 and scrublet version 0.2.1.

The plots of TFEA and GO term analysis were generated in RStudio with R version 3.6.0 with the usage of the R packages enrichR_3.0, ggpubr_0.4.0, ggplot2_3.3.3, stringr_1.4.0, EBImage_4.32.0, and bioimagetools_1.1.5.

NicheNet analysis was performed using following package versions: xlsx_0.6.5, ggpubr_0.4.0.999, cowplot_1.1.1, RColorBrewer_1.1-2, circlize_0.4.13, forcats_0.5.1, stringr_1.4.0, dplyr_1.0.7, purrr_0.3.4, readr_2.1.1, tidyr_1.1.4, tibble_3.1.0, ggplot2_3.3.5, tidyverse_1.3.1, SeuratObject_4.0.4, Seurat_4.0.2, nichenetr_1.0.0, gridBase_0.4-7, and ComplexHeatmap_2.6.2

FastQ files of antibody and IntAct signals were annotated to reference sequences in R using ShortRead_1.52.0 package with following annotated packages. GenomicAlignments_1.30.0, SummarizedExperiment_1.24.0, Biobase_2.54.0, MatrixGenerics_1.6.0, matrixStats_0.61.0, Rsamtools_2.10.0, GenomicRanges_1.46.1, BiocParallel_1.28.3, Biostrings_2.62.0, GenomeInfoDb_1.30.0, Xvector_0.34.0, Iranges_2.28.0, S4Vectors_0.32.3, BiocGenerics_0.40.0.

2.11. Antibodies

Target	Used concentration	Manufacturer	Catalog #
PECAM1	1:50	Invitrogen	PA5-16301
α-Smooth Muscle Actin	1:100	Invitrogen	MA1-06110
PDGFRB	1:50	Sigma	WH0005159M8
Lumican	1:50	Novusbio	NBP2-89941
Collagen I	1:200	Novusbio	NB600-408
Collagen IV	1:100	Abcam	ab6586
anti-rabbit IgG, AF 555	1:500	Life Technologies	A31572
anti-mouse IgG, AF 488	1:500	Life technologies	A21202
IGF-I Receptor β	1:200	Cell Signaling	9750s
goat anti-rabbit IgG (H+L)	1:1000/1:500	ThermoFisher	31210
goat anti- mouse IgG (H+L)	1:500	ThermoFisher	31160
mouse anti-Insr	1:5	ThermoFisher	MA5-13778
Rabbit anti-Itgb1	1:10	ThermoFisher	MA5-31964
anti-Itgb3	1:40	ThermoFisher	MA5-32077
PECAM1	cryo-staining by facility*	Thermofisher	AbAPB5
DLL4	cryo-staining by facility*	Cell Signaling	mAb 96406
DEPTOR	cryo-staining by facility*	Cell Signaling	mAB 11816
CLDN5	cryo-staining by facility*	Thermofisher	MA5-32614
ICAM2	cryo-staining by facility*	Cell Signaling	mAB 13355
PDGFRB	cryo-staining by facility*	Thermofisher	AbWM59

*The cryo-sectioning and staining was performed by the Core Facility Pathology and Tissue Analytics – HMGU and the used antibody concentration was not provided.

2.12. Used cell lines

name	specification
hiPSCs	provided by Wang *et al.*[127]
nucleus reporter hiPSCs	AICS-0034-062 (Coriell Institute for Medical Research)
cytoplasmic reporter hiPSCs	AICS-0054-091 (Coriell Institute for Medical Research)
mouse preadipocytes	provided by Ruiz-Ojeda *et al.* [137]
mouse preadipocytes kindlin2 KO	provided by Ruiz-Ojeda *et al.* [137]
Min6 wt	provided by Ansarullah *et al.*[138]
Min6 Iir KO	provided by Ansarullah *et al.*[138]

3. Results and Discussion

3.1. Endothelial differentiation and transcriptomic analysis

In this study, we used single-cell transcriptomics to consecutively investigate the development and neovascularization of human induced pluripotent stem cells (hiPSCs) in an *in vitro* 3D microenvironment. In the first step, we used single-cell transcriptomics to explore the differentiation trajectory of co-evolving endothelial cells (ECs) and mural cells in a 3D suspension cell culture format. Comparison of single-cell transcriptomics of ECs, evolved of a monolayer and 3D suspension culture, revealed differences in extracellular matrix (ECM) gene expression and optimal differentiation parameters. In the second step, the single-cell transcriptomics approach was used to analyze the neovascularization in the heterogeneous 3D suspension culture upon transfer into a hydrogel culture. Ligand-target links between ECs and subcellular pericyte populations were predicted from the single cell transcriptomes during vessel maturation. Finally, the plasticity of generated cell types was investigated. The method of choice was single-cell mRNA sequencing (scRNA-seq) to obtain deep insights into cellular state. The scope on the single-cell level is of special importance due to the heterogeneity that may occur during differentiation. Moreover, mature ECs in the body represent extensive heterogeneity not only between different tissues but also within the same vessel[139]. With transcriptomic analysis insights in the EC population can be revealed in bulk[140]. However, this method is based on the selection of considered cell population and depends on differentiation efficiency or rather sorting strategy. Therefore, heterogeneity inside that bulk can only be raveled by single-cell approaches. The current increase of scRNA-seq data reveals further insights in EC heterogeneity in specific tissues and under different conditions[141,142]. This study builds on their development and provides descriptive aspects for embryonal development, angiogenesis, and the use of generated cells in therapeutic or tissue engineering applications.

This part of the thesis is substantially based on the publication: “Rosowski, S., Brähler, C., Marder, M., Akishiba, M., Platen, A., Ussar, S., Theis, F., Wiedenmann $, S., and Meier $, M. (2022). Single-cell characterization of neovascularization using hiPSC-derived endothelial cells in a 3D microenvironment. BioRxiv 2022.02.15.480506.” that forms the requirement for obtaining the PhD degree. Paragraphs from this version of the publication, as well as from the one that is currently under revision, have been included literally into this thesis.

3.1.1. Single-cell analysis of endothelial differentiation in a 3D suspension culture

To investigate the differentiation of hiPSCs into ECs in a 3D cell culture format at the single-cell level, we adopted the chemical two-step induction protocol[49,51,55]. Therefore, hiPSCs were differentiated towards the mesoderm germ layer and EC development was induced in the second step (**Figure 3.1a**) In suspension, the 3D cell cultures were stable over the differentiation and grew from a diameter of 150 µm to 300 µm. While hiPSC-derived aggregates exhibited a uniform spheroidal shape, aggregates from day four showed a more prolate shape[126].

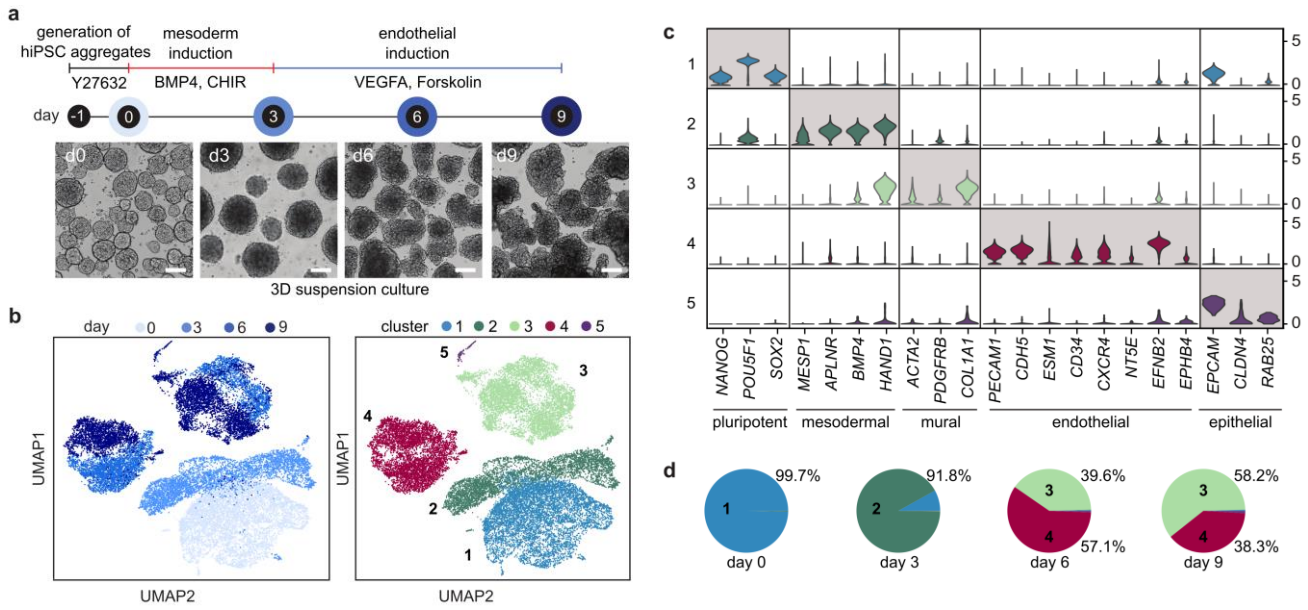

Figure 3.1 | Single-cell transcriptomics reveals the differentiation trajectory of hiPSCs into endothelial cells in 3D suspension culture. a, Schematic of the endothelial differentiation timeline with sampling points and chemical induction protocol. Bright-field images show representative 3D suspension cultures at the corresponding time point. Scale: 100 µm. **b**, UMAP plot of the single-cell transcriptomes. Left: Light to dark blue denotes the time points of sampling. Right: Six unique cell clusters were identified during the endothelial differentiation, 1: hiPSCs, 2: mesoderm, 3: mural cells, 4: angioblast-like cells, 5: epithelial cells. **c**, Violin plot shows the cluster expression levels of differentially expressed genes for the six cell clusters and the commonly used cell markers for cell type assignment. **d**, Cell type distribution analysis along the differentiation trajectory.

Flow cytometry was performed to analyze the robustness of the differentiation protocol. On day nine, 33.2% of the cells expressed the endothelial marker *CD31* (*PECAM1*), with a standard variation of 5.3% over three biological repeats (**Figure 3.4**)[126].

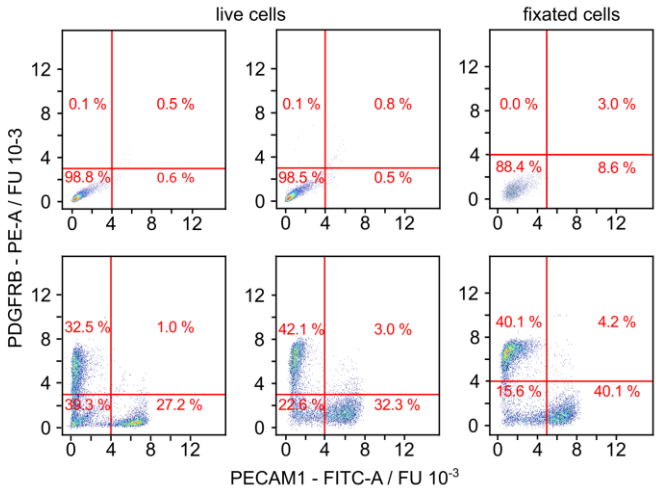

Figure 3.2 | Flow cytometric analysis of three independent EC differentiation experiments. Cells were stained with PECAM1-FITC and PDGFRB-PE antibodies on day 9 of differentiation. Upper left, unstained cell sample control; upper center and right, isotype control of the first EC differentiation experiment; lower row, three independent EC differentiation experiments.

To reconstruct EC development in the 3D suspension culture and define time-resolved cell composition, we performed single-cell mRNA sequencing (scRNA-seq) analysis on 22,192 cells (**Table 3.1**)[126].

Table 3.1 | Cell sample and sequencing statistics.

sample ID	condition	dead cells / %	mean # of genes per cell	mean # of sequencing read counts	number of analyzed cells
day 0	3D suspension culture	17.4	4440	22361	6411
day 3	3D suspension culture	36.5	4132	19660	4549
day 6	3D suspension culture	19.1	3644	12080	4807
day 9_1	3D suspension culture	25.0	4051	15200	6425
day 9_2	3D suspension culture	30.8	4388	19199	3011
day 12	3D suspension culture	16.1	5354	24817	4114
day 12	3D Matrigel	4.3	3835	12165	13530
day 18	reaggregated 3D suspension culture	17.0	4083	12869	4396
day 18	3D Matrigel	12.3	4807	19542	6462
day 18	3D Matrigel + ascorbic acid	7.6	4673	19369	8743

Upon dimensional reduction[85] and Leiden clustering[83], the cells were assigned into five clusters. With the progression of the differentiation process, the recorded single-cell transcriptomes changed, as indicated by the time-dependent emergence of distinct cell clusters (**Figure 3.1b**). All cell clusters could be assigned to cell types by matching known mesodermal and endothelial developmental markers to the differentially expressed genes (DEGs) in the respective cluster (**Figure 3.1c**). The cell populations were assigned to pluripotent stem cells (cluster 1), mesodermal cells (cluster 2), mural cells (cluster 3), angioblast-like cells (cluster 4), and epithelial cells (cluster 5). At the start of differentiation (day 0), the cell population consisted of homogenous undifferentiated hiPSCs, where over 96% of the cells expressed the pluripotency markers *OCT4*, *SOX2*, and *NANOG*. Cells assigned as mesoderm appeared on day three of differentiation and expressed markers for the lateral plate mesoderm[143], including *HAND1, MESP1*, and *APLNR.* Mural cells observed on day six of differentiation showed reduced *HAND1* expression level, while smooth muscle marker *ACTA2*, pericyte marker *PDGFRB*, and mesenchymal marker *COL1A1* were consistently expressed. Only a small fraction (0.9%) of epithelial cells was observed (**Figure 3.1d**). Immunohistochemical staining of 3D suspension cell cultures with the markers *PECAM1* and *PDFGRB* showed de-mixing of the two cell populations; however, there was no induction of vessel formation, supporting the angioblast cell state (**Figure 3.4**)[126].

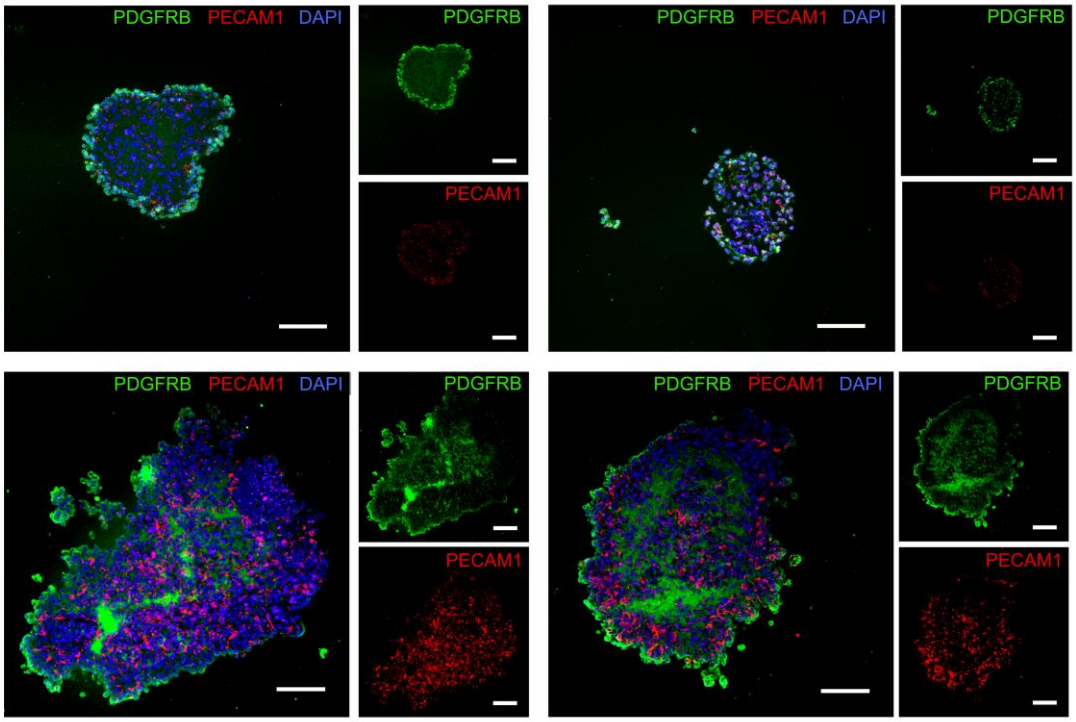

Figure 3.3 | Immunofluorescence images of sections of 3D suspension culture aggregates from day 6 (upper row) and day 9 (lower row) stained for DAPI (blue), PDGFRB (green), and PECAM1 (red). Scale bar denotes 100 µm.

Evaluation of cell cycle states showed that angioblast-like cells were entirely in the G1 phase, whereas approximately 50% of the mural cells were in G2 and S and thus proliferating (**Figure 3.4a**). This explains the increasing proportion of mural cells from day six to nine in the 3D cell culture. To test the robustness of the differentiation approach at the single-cell level, we sequenced the cells from day nine of two independent differentiation experiments. In both cases, a bimodal distribution of angioblast-like cells and mural cells was observed, with comparable distribution numbers (**Figure 3.4b** and **c**)[126].

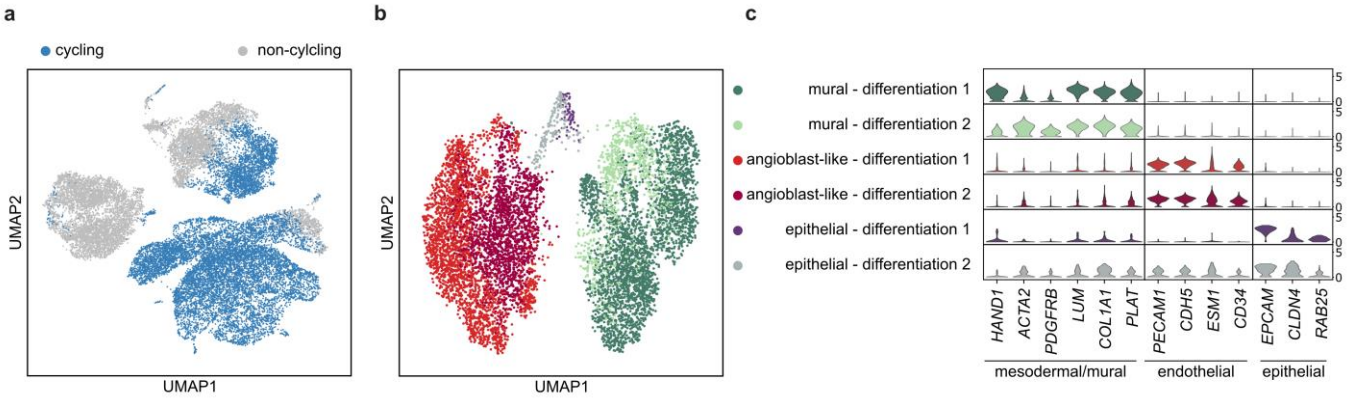

Figure 3.4 | Cell cycle analysis and reproducibility of the endothelial differentiation in a 3D suspension culture. a, UMAP plot of the scRNA-seq dataset from experiment 1 (experiment from the main Fig. 1), where cells in the G2, S, and M phases were colored in blue, and cells in the G1 phase in grey. **b**, UMAP plot of single-cell transcriptomes from two independent differentiation experiments. The color code denotes the cell types found in the two differentiations. 6618 cells were analyzed in the first and 5035 in the second sequencing experiment. **c,** Violin plot of common cell type marker genes for annotation of the clusters. The density distribution indicates the normalized cluster mean expression.

The results demonstrate a stable and effective differentiation of ECs from hiPSCs. Flow cytometry data as well as single-cell transcriptomic data ensured the stability of the protocol in 3D suspension culture. The second differentiation appears to have a higher maturity because the mesodermal marker *HAND1* is decreased in the mural cells of this experiment **Figure 3.4c**. Further in this cell stage, the transcript number of *PDGFRB* and the presence of the receptor on the cell surface represents at most a weak correlation, since flow cytometry and fluorescence images (**Figure 3.4a** and **e**) display significant presence of the receptor whereas mRNA expression levels are comparably low (**Figure 3.1c** and **Figure 3.4c**).

Interestingly angioblast-like cells did not show any proliferation which suggests that 3D suspension culture induces cell cycle arrest. Since no difference between day six and day nine is occurring it may not be influenced by maturation but the by cell type itself. A further possibility could be that the surrounding cell-cell contacts between the angioblast-like cells and the corresponding signaling in the 3D aggregate prevents proliferation. In the mural cell population, for both days a comparable number of cells is still proliferative. This proliferative subtype could be the mural cells in the surface layer of the aggregate while the mural cells which are not directly at the surface may also be arrested in their cell cycle (compare **Figure 3.4e**).

In embryonal development, MESP1 is repressing brachyury, SOX17, and goosecoid expression for specifying mesodermal cells towards cardiogenic precursor cells and induces migration to form the heart tube[144,145]. Further, the presence of BMP and absence of Wnt signaling are described to induce NKX2.5 and MESP1, that are both critical in the faith decision towards heart cells. During mesoderm induction in our differentiation protocol, similar conditions are present by having BMP4 and CHIR as additives in the media. While *MESP1* showed high expression in the mesoderm cluster (**Figure 3.1c**), *NKX2.5* does not show any. Comparing this with a cardiac differentiation protocol, the mesodermal cluster in this study might

correspond to the MESP1$^+$ cardiac mesoderm but the further stage of NKX2.5$^+$ cardiac progenitors was not reached[146].

Lineage specification on day six of differentiation into an arterial tone could be observed when taking the stronger expression level of *EFNB2* into account[147] (**Figure 3.1c**). For scaling EC production with a 3D suspension culture, the transcriptional cell type analysis argues that six days of differentiation is optimal due to the largest EC to mural cell ratio[126] and no major transcriptomic differences compared to day nine.

3.1.2. Time trajectory of endothelial development

To resolve the time-dependent relationships of the cell clusters, we performed a dynamic RNA velocity analysis[90] and identified the dynamic driver genes (DDG) in endothelial differentiation[91]. For this, we first calculated the latent time based on the balance of spliced and unspliced RNA transcripts within the single-cell transcriptomes (**Figure 3.5a**)**.** The corresponding RNA velocity is weakly streamlined between the transition states of the cell clusters. Nevertheless, a PAGA analysis on the RNA velocity demonstrated the connectivity (**Figure 3.5b**) between the angioblast-like and mural cells and the mesodermal progenitors. Subsequently, we plotted the DDGs along the velocity latent time to trace the central genes for the development of the respective cell types in 3D (**Figure 3.5c**)[126]. The scvelo algorithm identified the angioblast-like cluster as final endpoint of the latent time while mural cells were annotated chronologically between mesodermal and angioblast-like cluster. Following the gene expression of the cells aligned by the latent time, a pattern appears that links distinct DDGs to the annotated cell types. The top DDGs for the mural cell progenitors were involved in cell migration, attraction, or repulsion (*UNC5C, SLIT3*, and *TGFB2*)[148,149]. For ECs, known developmental genes of vasculogenesis were upregulated, including the *VEGF* receptors (*KDR, FLT1*) and the interacting receptors *TIE1* and *TEK*[150,151]. The unique cluster specific DDGs for angioblast-like and mural cell clusters were used in a transcription factor enrichment analysis (TFEA), to identify transcription factors (TFs) that control their development (**Figure 3.5d**). The highest-ranked TFs for EC development were *BCL6B* [152]*, ETS1* [153]*, ELK3* [154]*, ERG* [155], and members of the SOX family. All of them are reportedly associated with the process of early vasculogenesis with context-dependent function but integrate the VEGF and Notch signaling pathways [152]. Notably, the extracted TFs are putatively responsible for the development of the two identified cell types, but not for neovascularization due to the missing vessel organization within the 3D cell cultures at day six and nine of differentiation. TFEA of the DDGs for the mural cell progenitors revealed *TBX18*, *CENPA*, and *HAND2* as the top regulatory TFs [156]. Expression of *TBX18* was not detected in the scRNA-seq dataset; however, its transcriptional activity matched with its recently identified expression pattern in mice pericytes and vascular smooth muscle cells of the retina, brain, heart, skeletal muscle, and adipose fat depots [157]. It has to be considered that the velocity analysis with spare time points has to be understood as transcriptional correlation rather than real dynamics[126]. The spliced versus unspliced mRNA scatter plots of several TFs represent this correlation (**Figure 3.5e**). The plots show the top eight TFs that were included in the whole RNA velocity analysis. Several TFs, especially the once that were inferred for the mural cluster, are not

considered since the algorithm uses a cutoff to exclude genes with no significant impact of the modeling. Further, the dynamical model of induction and repression was not predictable only for all TFs. For instance, *SOX17* has cells with high counts of spliced but no unspliced transcripts, which might have prevented the modeling. In angioblast-like cells, all scatterplots of obtained TFs display an enriched number of spliced and unspliced mRNA compared to the other cell clusters that denote their upregulation (**Figure 3.5e**, left). However, the ratio refers mainly to a steady expression, since cells accumulate around the dashed steady state straight instead of around the induction or repression curve (compare **Figure 1.5c**). The TFs of the mural cell cluster represent with the exception of *PRRX1* and *TBX18*, which only shows a low number of cells in the plot, an induction state (**Figure 3.5e**, right). Nevertheless, the turn-on and -off of genes derived from the velocity analysis match with the expectation, for example, for the cell cycle regulator *HMGA2*. In mural cells, *HMGA2* is in the turned-on state, whereas in angioblast-like cells in the turned-off state but still detectable[126].

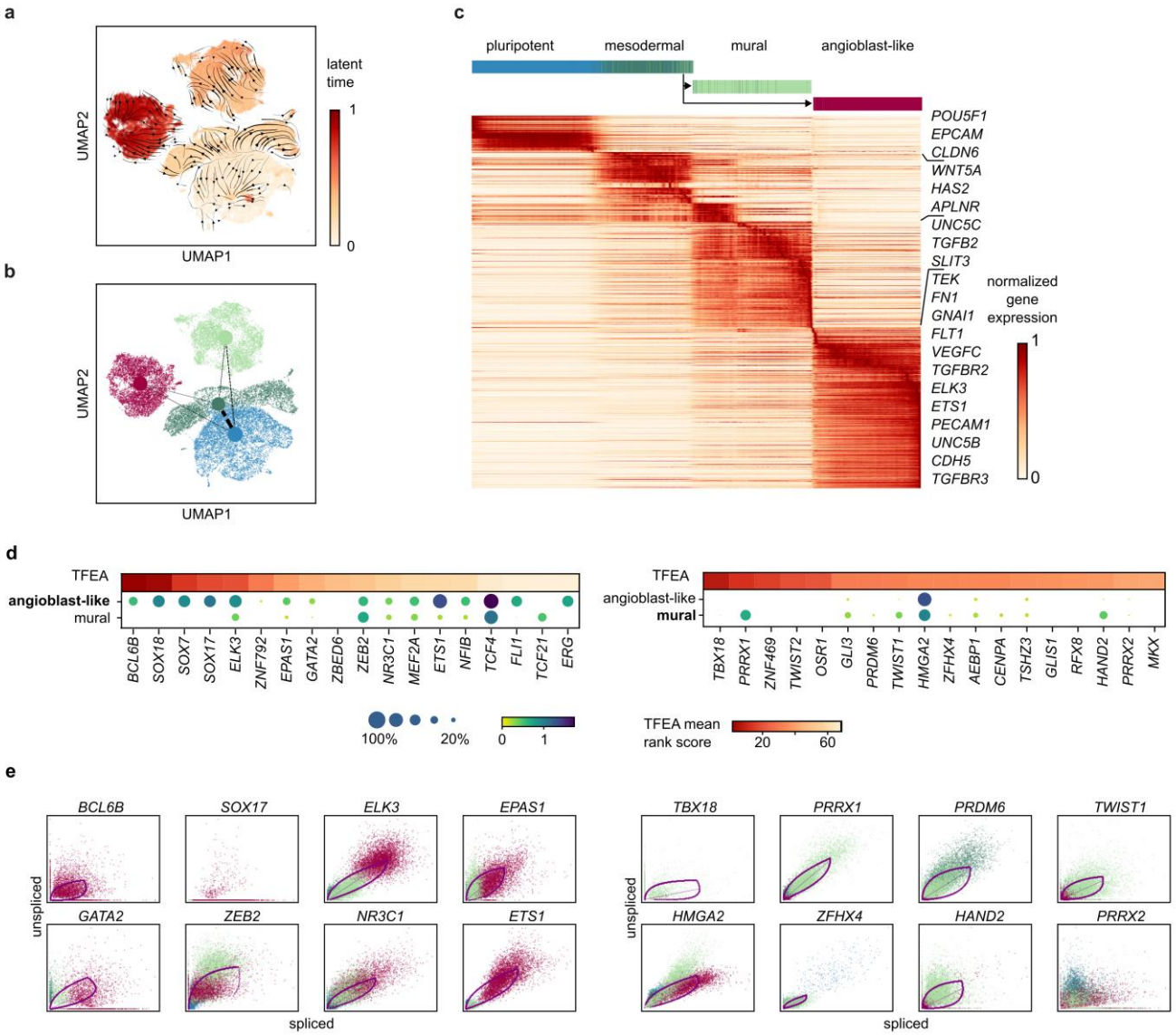

Figure 3.5 | Transcriptomic dynamics predict the differentiation path for endothelial and mural cells. a, UMAP cluster plot colored with the latent time calculated based on RNA velocity analysis. The velocity streamlines are given by the black arrows. **b**, UMAP plot colored by annotated cell type colors and with an overlay of PAGA connectivity. **c**, The top 300 DDGs sorted according to their likelihood scores and latent time. **d**, TFEA on the cluster-specific and unique DDGs for angioblast-like and mural cells. The dot plots show the gene expression level in the respective single-cell cluster, whereas the upper heat-colored bar shows the TFEA score. Color intensity and dot size denote the normalized cluster mean expression and the fraction of cells expressing the gene, respectively. **e**, and Scatter plots of spliced versus unspliced mRNA of enriched TFs that were considered in in the velocity analysis.

The transcriptional dynamics calculated by the RNA velocity analysis matched the time of differentiation and the scvelo algorithm for calculating the latent time generated a single endpoint in the EC population (**Figure 3.5b**). This single endpoint generation seems to be a characteristic of scvelo. In the publication of the velocity toolkit, the development of endocrine cells in mouse embryos was used as a model[91]. Here, the emergence of the different cell types; alpha, beta, delta, and epsilon cells were analyzed. While the velocity streamlines indicate their separation from the pre-endocrine cluster, the beta cell cluster was

assigned as endpoint of the latent time. Apparently, the latent time algorithm has difficulties in generating multiple final stages. The ECs have stronger differences in the expression pattern to mesoderm than mural cells and this suggests why ECs were assigned to the latent time endpoint. On the other hand, the generation of multiple latent time endpoints and its representation in a two directional time line would lead to a clash of cells with same latent time points. Here, a cluster-wise separation has to be performed, similar to the manually performed separation of the mural cluster in **Figure 3.5b**.

Based on the velocity model of gene induction and repression, ratios of unspliced versus spliced transcripts for a gene should follow the convex curves (**Figure 1.5c**). For the inferred endothelial TFs this aligning is not observed for the majority of the angioblast-like cells (**Figure 3.5e**). The dashed straight line between the convex gene induction and repression curves represents the steady states of gene expression and describes the endothelial expression of the shown TFs in a more accurate way. This suggests that angioblast-like cells in the cluster have a heterogeneity in the expression of these genes, but they are not in an induction or repression phase of transcription. Further, this aspect highlights the endothelial developmental arrest because no changes in gene expression are described for the individual cells.

In addition, the expression patterns of TFs highlight the weaknesses of scRNA-seq since the presence of the TFs cannot be confirmed for all that were determined by TFEA. Here, a more precise approach would be the usage of protein readout linked single-cell transcriptome sequencing. The proteins of the TFs with low mRNA copy number could be targeted by oligonucleotide coupled antibodies to infer their presence. Approaches to enable intracellular or intranuclear protein staining for multi-omics are described and discussed in (chapters **1.6.6** and **3.2.1**).

3.1.3. Comparison of endothelial differentiation in 3D versus 2D cell culture formats

In the next step, we sought to compare the *in vitro* development of ECs in 3D suspension culture to a previously performed stem cell differentiation approach with an adherent 2D cell culture on the single-cell level[52]. Chemical compounds for the endothelial induction protocol were the same; however, minor concentration differences of the individual compounds existed (**Table 3.2**)[126].

Table 3.2 | Culturing conditions of the differentiation in 2D and 3D.

	day of differentiation		**this study**	**McCracken *et al*.**	**day of differentiation**
lateral mesoderm induction	0-3	media BMP4 CHIR99021	N2B27 25 ng/ml 7.5 µM	N2B27 25 ng/mL 7 µM	1-4
endothelial induction	3 to 7	media VEGF-A Forskolin	StemPro-34 200 ng/ml 2 µM	StemPro-34 200 ng/ml 2 µM	4
late stage	from 8 on	media VEGF-A FGF-2 human AB serum	StemPro-34 30 ng/ml 30 ng/ml	EGM-2 50 ng/mL 1%	until 8

First, a comparison of the scRNA-seq datasets after mesodermal induction was performed (**Figure 3.6a**). This aimed for comparing the progenitor cell population of the ECs and mural cells. In literature, a mesoderm-derived precursor is described that is termed mesenchymoangioblast. It derives from a mesoderm population that is positive for *APLNR*, *KDR*, and *PDGFRA*. Interestingly, these markers only were present in the 3D differentiation dataset but not in the 2D mesodermal cells (**Figure 3.6b**). Further markers that support the mesenchymoangioblast cell type ($PDGFRB^{+}NGFR^{+}EMCN^{+}DLK1^{+}NT5E^{-}$)[143,158] were not expressed in none of the two datasets. The mesodermal marker *HAND1* was expressed in both culturing conditions. Moreover, genes that are described in cardiac mesoderm development, as *MESP1* and *ISL1*, are only expressed in the 3D cultivation[159].

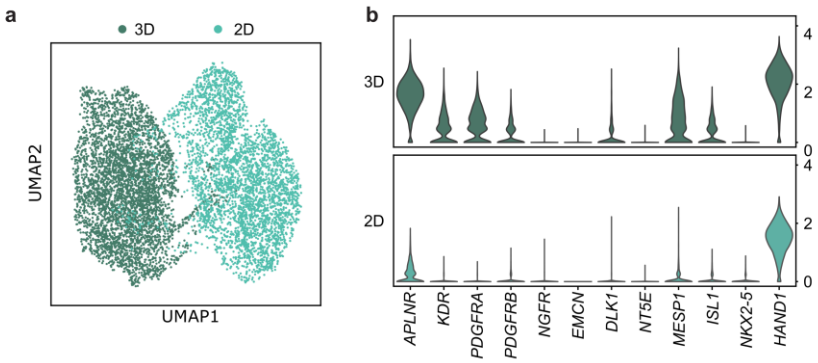

Figure 3.6 | Comparison of EC differentiation in 2D and 3D cell culture formats on day four and three, respectively. **a**, UMAP plot of single-cell transcriptomes acquired at days three (this study) and four (study of McCracken *et al.*) of the EC differentiation. **b**, Violin plots of mesodermal, mesenchymoangioblast, and heart development marker genes. The density distribution indicates the normalized cluster mean expression.

The more relevant time point for comparison is on day eight (reference study) or day nine (this study) when the differentiated ECs and coevolved mural cells were examined. Therefore, the two scRNA-seq datasets were combined (**Figure 3.7a**) and DEGs were calculated between the EC clusters of 3D and 2D culturing to quantify the differences in expression (**Figure 3.7b**). Of the 14,383 genes, 575 and 608 showed expression level differences with a p-value lower than 10^{-100}, in the 2D and 3D cell cultures, respectively. Dominantly, the expression patterns of the ECM genes were distinct in different culture formats. Within the 2D cell culture format, ECs exhibited a strong collagen phenotype with high expression levels of basal lamina proteins, such as *COL4A1/2, COL6A2,* or *COL18A1* (**Figure 3.7c**). In 3D cell culture, ECs expressed hyaluronic acid and the corresponding binding proteins. In contrast, angioblast-like cells within the 3D suspension culture upregulated cell-cell interaction and actin remodeling genes, such as *CLD5*, *DOCK4*, and *CTNNB1* (Wnt signaling)[160], and *RAP1B*[161]*, RAPGEF5*[162], and *RASIP1*[163] (Rap1 signaling), respectively. Using the DEGs as a query for a GO term analysis revealed an upregulations of genes associated with migration and motility for the ECs in the 2D cell culture while the angioblast-like cells represented an enrichment of Rap signaling was (**Figure 3.7d**). “Within the 3D cell culture angioblast-like cells stop proliferating from day six of differentiation, whereas stem cell-derived EC in the 2D cell culture format proliferated as indicated by the combined scRNA-seq data (**Figure 3.7d**).

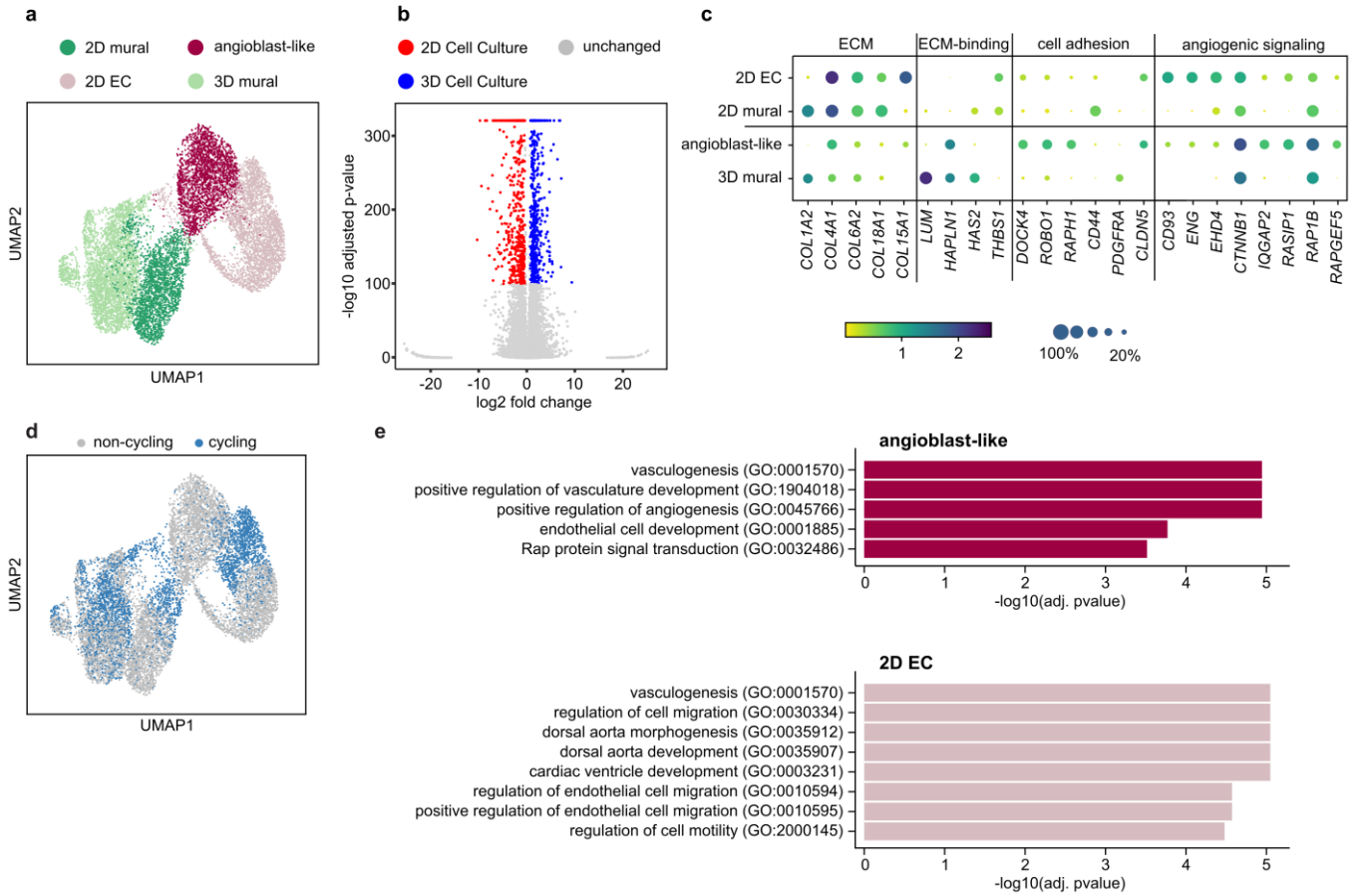

Figure 3.7 | Comparison of EC differentiation in 2D and 3D cell culture formats on day eight and nine. a, UMAP plot of single-cell transcriptomes of iPSCs derived endothelial cells differentiated on a 2D adhesion monolayer (5267 cells) and as 3D suspension culture (6492 cells). **b**, Volcano plot representing differentially expressed genes between ECs differentiated in 2D and 3D cell cultures. **c**, Dot plot of representative DEGs with assigned cellular functions and biological processes. Color intensity and dot size denote for the normalized cluster mean expression and fraction of cells expression the corresponding gene, respectively. **d**, UMAP plot shows the cells with gene expression relating to S, G2, or M-phase (blue). Cells expressing genes indicative of the G1 phase are denoted with a grey color. **e**, Gene ontology term analysis based on DEGs from the single-cell transcriptomes of ECs derived from stem cells cultured in a 2D monolayer and 3D suspension culture.

A dominantly expressed gene in the 3D mural cell cluster is *LUM* that is reported to suppress cell proliferation[164]. To evaluate if LUM might have an impact on the proliferation arrest of the angioblast-like cells, an immunofluorescence staining was conducted. However, the LUM signal mainly overlaid with the PDGFRB signal of mural cells while the center of the aggregates that represents the PECAM1 positive angioblast-like cells (compare **Figure 3.4**) was negative for LUM (**Figure 3.8a**). This result contradicts an impact of LUM on the angioblast-like cells. A similar approach was done with COL1 chains where the genes are highly expressed in mural cells (**Figure 3.7c**). Here, the immunofluorescence staining points out that COL1 is only built around mural cells while the COL1 signal vanished towards the inside of the aggregates (**Figure 3.8b**).

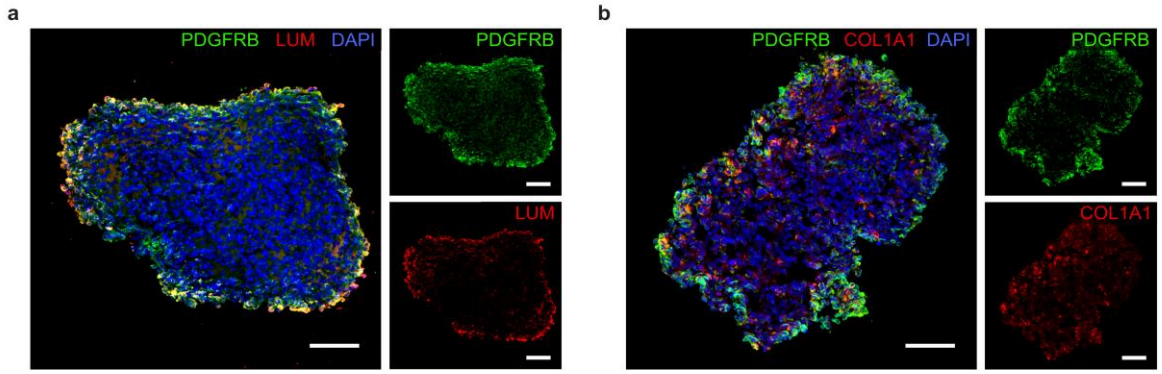

Figure 3.8 | Immunofluorescence images of sections of 3D suspension culture aggregates from day nine. Stained for DAPI, PDGFRB and either LUM (left) or COL1 (right). Scale bar denotes 100 µm.

The proliferation arrest of angioblast-like cells in 3D cell culture raised the question of whether the separation of the angioblast-like cells from the mural cells or the transfer to a 2D culture format reinstates the proliferation of the angioblast-like cells. FACS-sorted PECAM1^{+} cells from the 3D suspension culture proliferated and were passaged over eight generations upon plating on a culture dish (**Figure 3.9**).

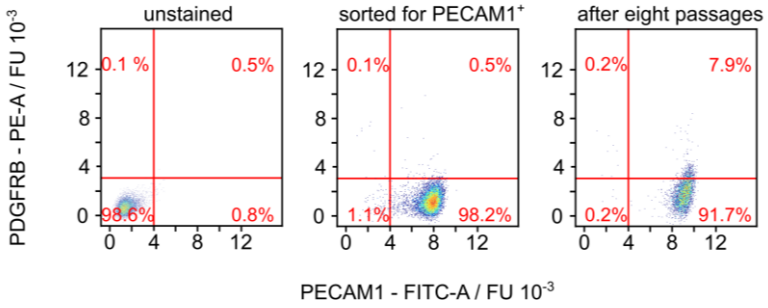

Figure 3.9 | Cell type stability and long-term culturing of stem cell-derived endothelial cells. a, FACS analysis of sorted PECAM1 positive cells from 3D suspension cultures at day nine of differentiation. Cells are plated on fibronectin coated well plates for long-term culturing in 2D cell culture format within EC maturation media. FACS analysis of the same PECAM1 positive cell culture after eight passages showed a comparable fraction of PECAM1 positive cells.

Comparison between sorted PECAM1^{+} cells within a 3D and 2D cell culture was not possible due to the observation that ECs alone did not form a stable aggregate in suspension culture. However, the transfer of the unsorted mixed 3D aggregates to a 2D cell culture format on day six of differentiation led to the proliferation of ECs, demonstrating that growth arrest is associated with the 3D suspension culture format[165].

The mesodermal cell population on differentiation day three of the 3D suspension culture represents high similarity to the literature reported mesenchymoangioblast phenotype that symbolizes the common precursor cell type for endothelial and mural linage[143]. Vodyanik *et al.* used colony-forming culture in 2D for their examination. The mesodermal populations in the study of McCracken *et al* showed already the

absence of the two mesodermal marker genes *PDGFRA* and *KDR* (**Figure 3.6b**). In this study, these markers were expressed at a certain level, however, an expression of marker genes supporting the mesenchymoangioblast state were not present. This indicates on the one hand that the 2D monolayer cultivation on day four is less mature than a 3D suspension culture. On the other hand, in the 3D suspension culture the mesenchymoangioblast state might be passed between day three and six and need the endothelial induction.

In embryonal heart development, a multipotent cardiovascular progenitor is described that gives rise to all cell types in the mature heart[6–8]. This progenitor shares certain similarities with mesenchymoangioblast that was described *in vitro*. Both have the capability to develop towards the endothelial and mural linage, respectively. The factors that were identified *in vivo* are KDR, ISL1 and NKX2-5. Anton *et al.* summarized the three publications stated above and drafted a triangular dependency in which the expression of two of the three proteins initiates a certain development[159]. While the expression pairs with NKX2-5 lead to cardiac linages, the co-expression of KDR and ISL1 indicates endothelial and mesenchymal development. On the gene level, the cells in the 3D aggregates express *KDR* and *ISL1* while both were absent in the 2D monolayer cell culture (**Figure 3.6b**). This leads to the suggestion, that cardiac mesoderm was passed through, and a predetermined state of the multipotent cardiovascular progenitors is reached only in the 3D suspension cell culture. A further question in these terms would be if the mesenchymoangioblasts can be placed on the same level in the development as the KDR^+ and $ISL1^+$ phenotype of the multipotent cardiovascular progenitors.

We found that the general differentiation trajectory of ECs and co-evolving mural cells in a 3D suspension culture resembled the development of that in a monolayer format. In contrast to the 2D cell culture approach, ECs become quiescent in the 3D suspension culture without entering the neovascularization process. Within the 3D suspension culture, the ECM in the surrounding of angioblast-like cells did not contain COL1. Furthermore, COL1 interacting proteins as the ITGA1 receptor were low expressed in angioblast-like cells. *ITGA1* signaling is essential for *in vivo* angiogenesis and could explain the observed proliferation arrest. *COL1* expressed by the mural cells could not compensate the missing COL1 since its deposition was spatially restricted to the mural cells in the 3D suspension culture. This also holds true for anti-angiogenic factors, such as Lumican, which colocalized only with mural cells (**Figure 3.8**).

3.1.4. Single-cell transcriptomics of *in vitro* neovascularization

The angioblast-like and mural cells obtained in the 3D shaking culture were evaluated for their ability to form structures in a 3D hydrogel environment. Upon the embedding into Matrigel, a dissociation of the aggregates on day 9 and reaggregation on a microwell chip platform was performed to ensure a homogenization in size (**Figure 3.10a**). Already in the first hour of cultivation in Matrigel initial sprouts or filopodia are visible (**Figure 3.10b**, d10). 24 h after embedding, sprouting cells have doubled the size of the former aggregate while the core represents high density owing to its dark color in the bright field image (d11). This has changed after further 24 h as an increased cell migration out of the aggregate core took place. The neighborhood of embedded former aggregates induces their neovascularization (compare **Figure 3.11a** and **Figure 3.10b**, d12). Prolonged incubation time leads to increase of microvessel length and thickness. Moreover, the spots of aggregate embedding become unrecognizable in the expanded network (**Figure 3.10b**, d14-18).

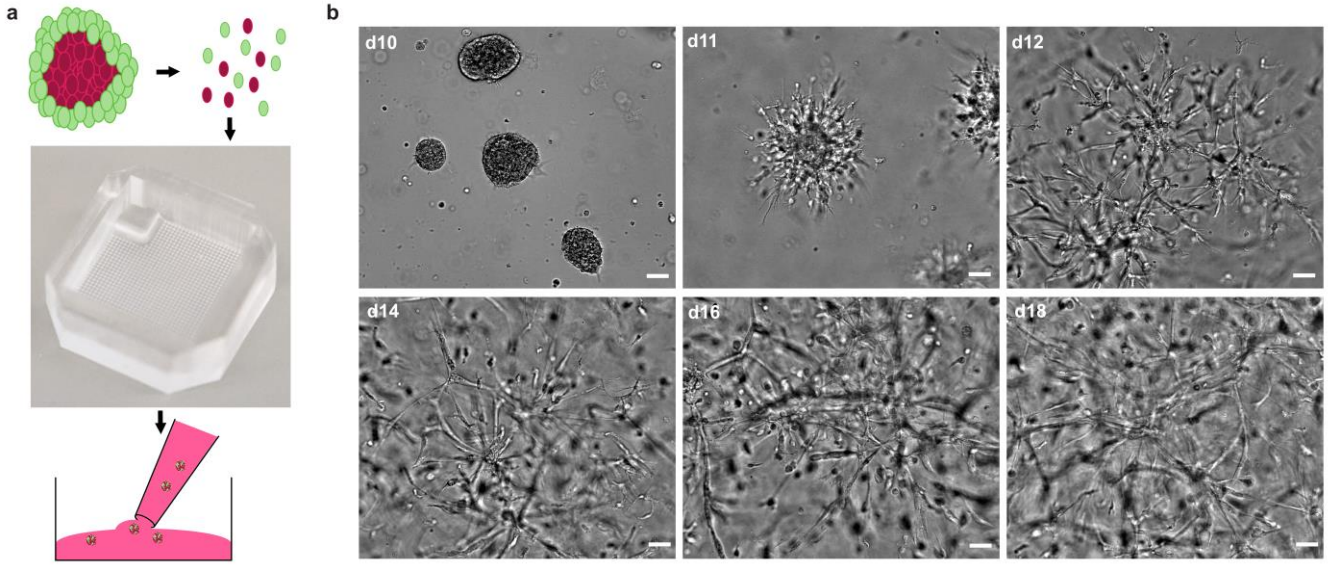

Figure 3.10 | Transfer of hiPSC-derived endothelial aggregates into Matrigel. a, Scheme of aggregate dissociation, seeding onto a microwell chip platform and the embedding into Matrigel. With mural cells in green and angioblast-like cells in red in correspondence to **Figure 3.1**. **b**, Bright-field images of the Matrigel culture along the timeline of day 10 to 18. Scale: 50 µm.

Single-cell transcriptomes of the sprouting Matrigel culture were determined 48 h after transfer and compared to single-cell transcriptomes of cells within 3D suspension cultures kept to the same day of differentiation. Three distinct transcriptomic EC subclusters were detected within the Matrigel culture (cluster 4-6), which separated from the angioblast-like cells in the 3D suspension culture (**Figure 3.11b**). In addition, mural cells formed two transcriptomic subclusters, where one cluster overlapped with the transcriptomic state of the mural cells in the 3D suspension culture. The expression levels of general cell type markers for the assignment of mural and EC clusters are shown in **Figure 3.13a**. The fraction of mural cells and EC was comparable to the that found in the 3D suspension culture (**Figure 3.11c**). A

corresponding PAGA analysis on the RNA velocity showed the connectivity between the transcriptional states of ECs and mural cells in the Matrigel and suspension culture cell types (**Figure 3.11d**).

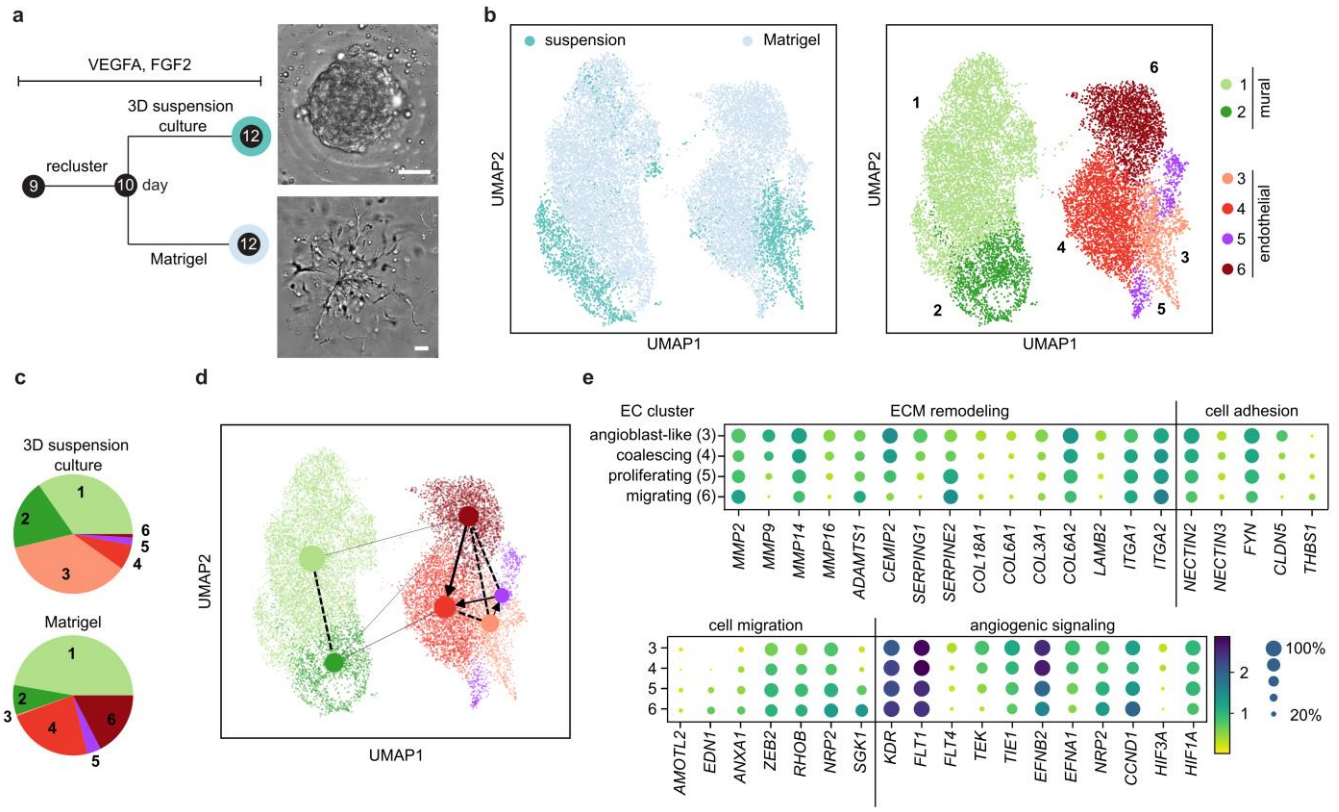

Figure 3.11 | **Single-cell analysis of hiPSC-derived ECs undergoing neovascularization in Matrigel. a**, Experimental overview of the culturing conditions for microvessel formation induction and representative bright-field images of cell culture morphologies. Scale: 50 µm. **b**, UMAP plots containing single-cell transcriptomes of cells from 3D suspension and Matrigel cultures. The color code denotes conditions (left) and Leiden cell clusters (right; 1 and 2: mural cells, 3: angioblast-like cells, 4: EC coalescing, 5: EC proliferating, 6: EC migrating). **c**, Cell type composition of the two culturing conditions is represented as a pie chart. **d**, Velocity analysis of the single-cell transcriptomic data from **a**. **e**, Expression levels of representative DEGs sorted by function for the three EC transcriptional states. The color intensity and dot size denote the normalized cluster mean expression level and the fraction of cell expression for the corresponding gene, respectively.

For the formation of microvessels, Matrigel-embedded angioblast-like cells must migrate into the hydrogel and coalesce. We plotted exemplarily DEGs related to ECM remodeling, cell migration and motility, cell interaction, and VEGF signaling (**Figure 3.11e**) to determine the genes that induced the processes and assigned transcriptional states to the main EC cluster. Genes associated with cell migration gradually increased from the angioblast-like state within the 3D suspension culture to the EC cluster 4 and 6 in Matrigel (e.g., *ANXA1*[166], or *ZEB2*[167]). On the contrary, gene expression for cell adhesion (e.g. *NECTIN2/3* or *CLDN5*) and ECM (e.g. *COL3*, *6*, and *18*) gradually decreased from the angioblast-like cells to ECs of clusters 4, and 6. This argued that ECs of clusters 4 and 6 represent the coalescing and migrating cell states, respectively. To investigate this further, we stained DEGs between the two clusters within cyro-sections of hydrogel culture from day 12 of differentiation. For ECs of cluster 4, DLL4 and CLD5 were

among the DEGs. Both proteins showed stronger expression in cells with interacting partners than in isolated ECs (**Figure 3.12**).

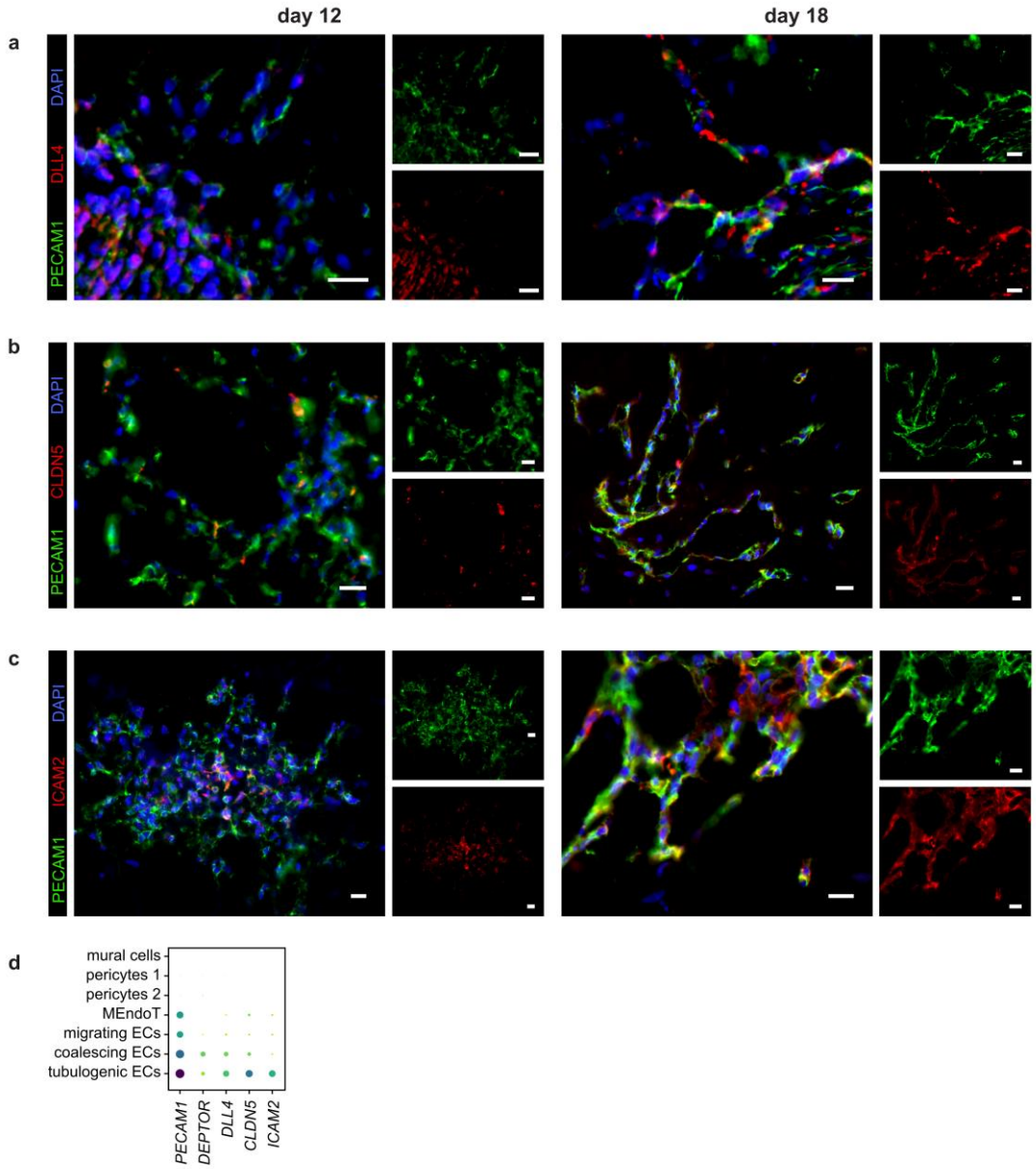

Figure 3.12 | Immunofluorescence images of cryo-sections prepared from 3D hydrogel cultures on day 12 and 18 of differentiation. Immunofluorescence stainings show the protein signal of the differentially expressed genes indicative for the migrating (**a**), coalescing, and tubulogenic ECs (**b**, **c**). Scale: 20 µm. **d**, Dot plot of mRNA expression of genes whose corresponding protein staining is shown above. The color intensity and dot size denote the normalized cluster mean expression and the fraction of cells expressing the gene, respectively.

This further indicated that ECs of cluster 4 are coalescence EC (cECs). Unfortunately, IF staining of DEGs of cluster 6 was not successful. For example, the strongest DEG of cluster 4 was the serum and glucocorticoid kinase 1 (SGK1), for which antibodies in the IF staining were not specific. However, it is known that ablation of SGK1 in mice leads to a strong reduction of EC migration. The additional EC cluster 5 can be assigned to proliferating ECs (**Figure 3.13b**).

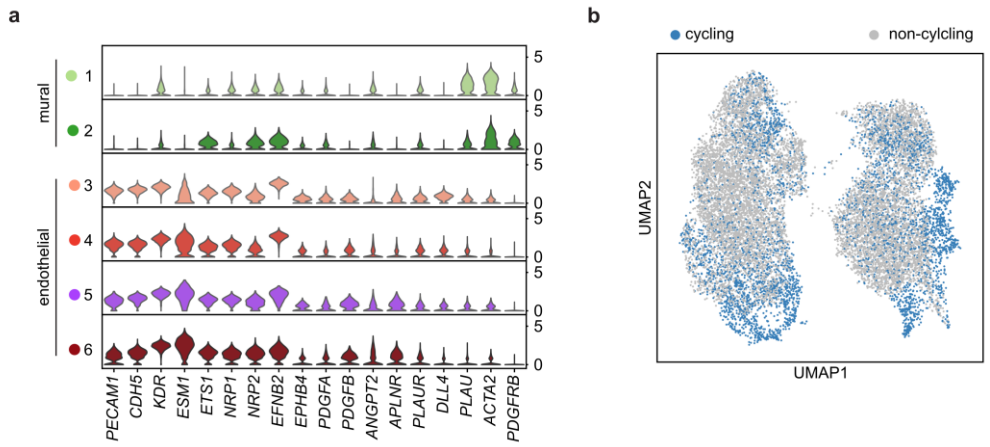

Figure 3.13 | Single-cell transcriptomes analysis of ECs and mural cells in Matrigel and 3D suspension culture. a, Violin plot of representative marker genes used to assign the cell clusters in the UMAP plot in main Fig 4. The density distribution indicates the normalized cluster mean expression. **b**, UMAP plot shows the cells with gene expression relating to S, G2, or M-phase (blue). Cells expressing genes indicative of the G1 phase are denoted with a grey color.

Within the DEGs of the ECs were key proteins of the VEGF, Notch, and mTOR signaling pathways. The most obvious was the upregulation of *KDR* and *NRP2* in mECs, which both act on endothelial motility, sprouting, and survival[12,168]. Within the Notch signaling pathway mECs downregulated *NOTCH1/4*, *DLL4*, and *JAG1* (**Figure 3.14**). This could be expected since Notch inhibition in cellular model systems has been shown to induce sprouting, branching, and filopodia induction [11,169,170]. Most interestingly, the mTOR pathway proteins, particularly those of the mTOR complex 2 (*RICTOR*), were strongly downregulated in mECs (**Figure 3.14**).

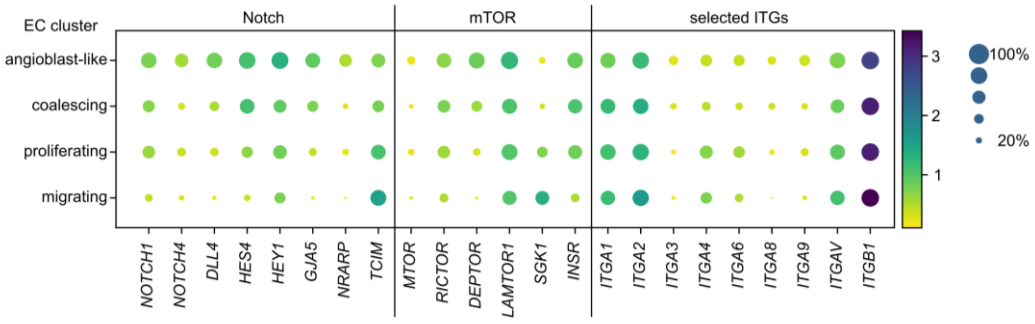

Figure 3.14 | Candidates with importance in angiogenesis – Notch and mTOR signaling pathway and Integrins. Dot plot of expression of genes within the mTOR and Notch pathway regulated during the cell state transition from angioblast to coalescing ECs and sprouting ECs. The color intensity and dot size denote the normalized cluster mean expression and the fraction of cells expressing the gene, respectively.

One downstream target of mTORC2 is the serum and glucocorticoid kinase 1 (*SGK1*), which is the top upregulated gene in mECs, indicating strong metabolic regulation in this motile cell state [171,172].

Furthermore, *DEPTOR*, an adaptor protein for mTOR complexes 1 and 2 was downregulated and concomitantly a DEG for cECs. IF images showed that in mECs DEPTOR exhibits a nuclear location and in cECs it could be detected in the nucleus and in the cytoplasm, demonstrating regulatory involvement of the mTOR pathway during EC migration (**Figure 3.15**). Notably, the expression levels of integrins changed only slightly between the EC transcriptional states.

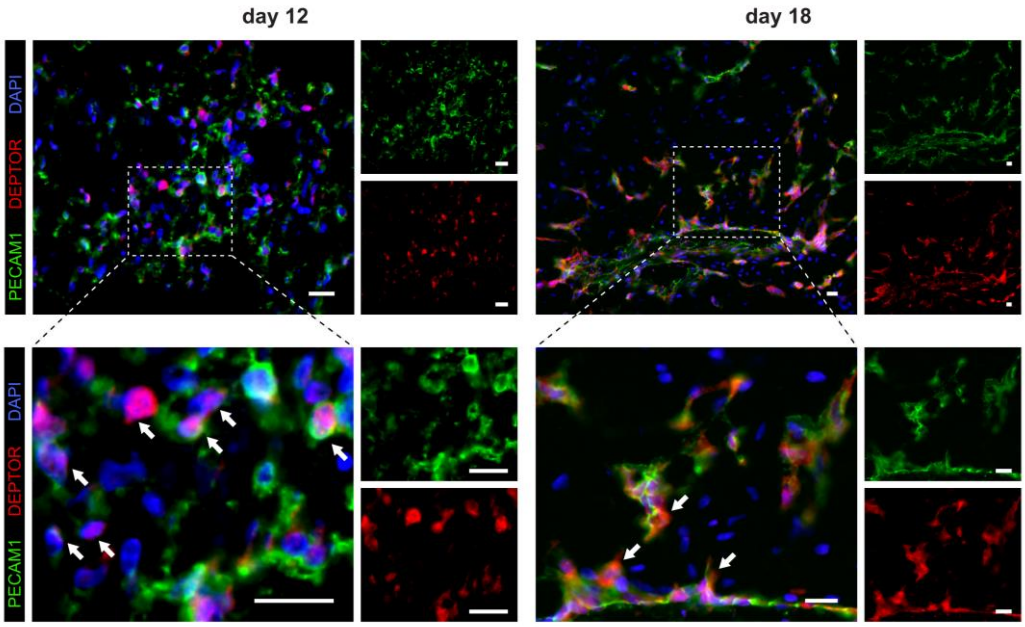

Figure 3.15 | Immunofluorescence images of cryo-sections prepared from 3D hydrogel cultures on day 12 and 18 of differentiation. Arrows in **a**, day 12 (lower left panel) highlight the location of DEPTOR in the nucleus, while on day 18 (lower right panel) DEPTOR is expressed in the nucleus and cytoplasm. Scale: 20 µm.

Corroboratively, a GO-term analysis of the DEGs of cluster 6 also indicated the enrichment of genes associated with cell migration (**Figure 3.16a**). Furthermore, a TFEA of the DEGs of mECs showed enrichment of *ELK3*, *KLF6*, and *SNAI2* (**Figure 3.16b**). The latter is a TF well described in the context of the epithelial-mesenchymal transition [173]. Based on these findings, we assigned the cluster 6 to migrating ECs (mECs). Additionally, *ERK* and *KLF6* control cell proliferation, which was indeed reinstated in ECs after Matrigel transfer[174].

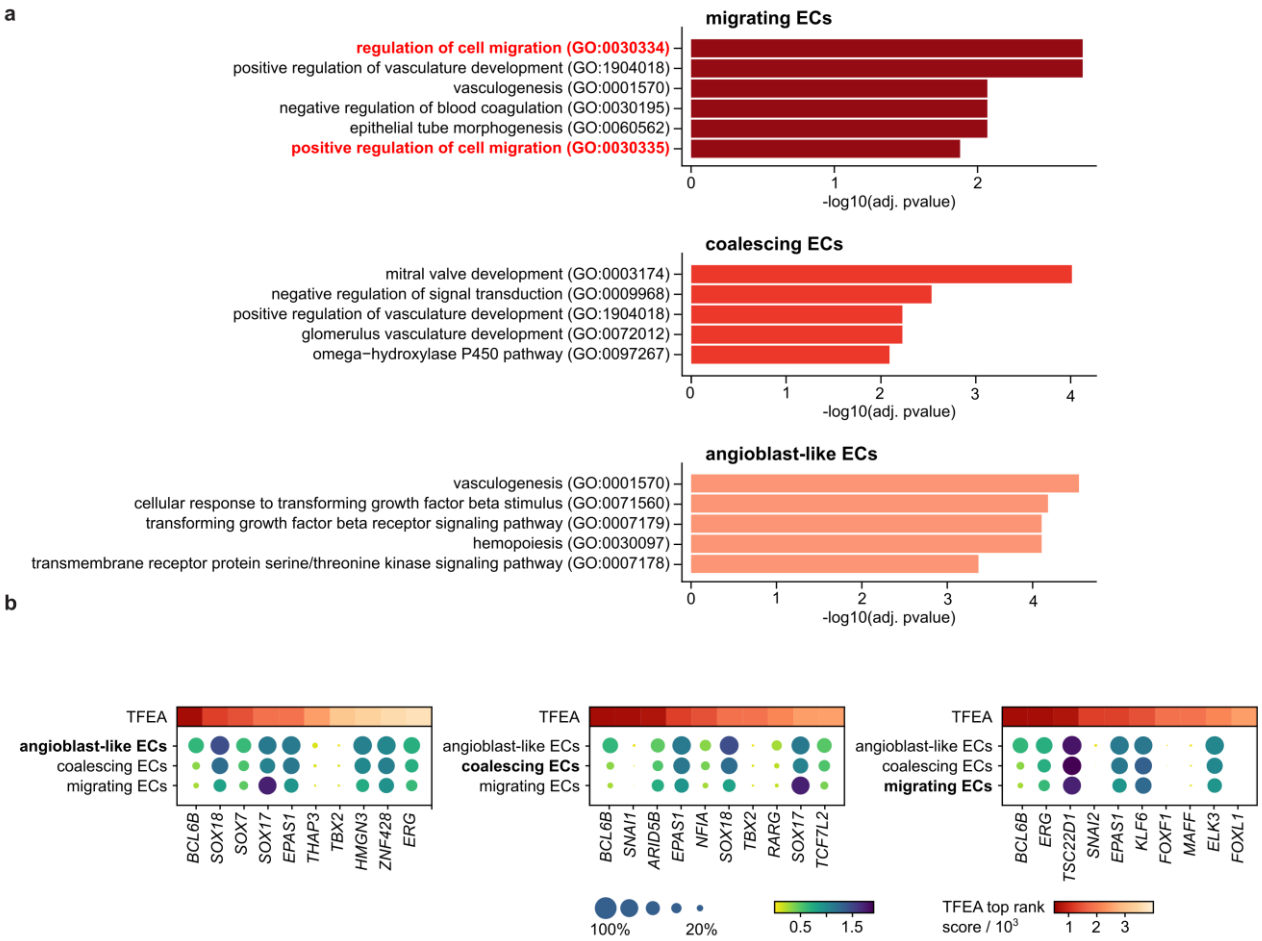

Figure 3.16 | Single-cell transcriptomes analysis of ECs and mural cells in Matrigel and 3D suspension culture. a, Gene ontology term analysis based on DEGs from the single-cell transcriptomes of angioblast to coalescing EC and sprouting ECs. **b**, Transcription factor enrichment analysis based DEGs the angioblast (left), coalescing (middle), and sprouting (right) EC cluster. DEGs between the EC clusters were filtered for expression and standard deviation before TFEA. TFEA scores are represented in a color code, whereas the mean expression levels of the corresponding TFs as dot plots.

Transfer of the 3D suspension culture into Matrigel led to sprouting of the ECs. Single-cell transcriptomics untangles migrating, coalescence, and tubulogenic EC cell states during neovascularization. In comparison to the *in vivo* reported data, after laser ablation in the choroid layer of mouse eyes, the cellular complexity is far lower. In addition to the expected VEGF and Notch signaling pathways, the mTOR pathway is regulated during neovascularization. mTOR kinase controls various processes, when complexed in the mTORC1, it controls predominantly the cell metabolism. The metabolism for example a switch between OXPHOS and glycolysis for mECs could not be observed as described for tip cells *in vivo*[175]. However, in mECs, mTORC2 adaptor proteins were downregulated and the downstream effector kinase *SGK1* was strongly upregulated, which controls cell survival during angiogenesis[176] and EC shape[177]; its ablation led to reduced neovascularization and impaired cell migration[171]. The function and regulation of *SGK1* through mTORC2 are unknown but have become the focus of further investigations.

TCIM is the only Notch pathway associated DEG that shows an increased expression in mECs (**Figure 3.14**). It suppresses the NOTCH2 signaling and therefore gives insights into regulatory

mechanisms of the Notch pathway downregulation in mECs[178]. From the structural point of view, that sprouting is an act of loosening cell-cell contacts to neighboring cells and Notch ligands as well as receptors are both membrane-bound, this signaling cannot be maintained during sprouting.

Comparing the TFEAs, we first have to take into account that the method of selecting the genes for query was different. The cluster specific DDGs between the three endothelial subtypes on day 12, angioblast, cECs and mECs, had a high conformity. Therefore, the unique DDGs between the clusters amounted between 15 and 41, which seemed to be an insufficient number of genes for the TFEA. Consequently, filtered DEGs were used instead. Nevertheless, inferred TFs showed similarities between angioblasts on day six/nine and the EC clusters on day 12. The highest overlap was between the two angioblast clusters, underpinning their phenotypical similarity and the arrest state of the ECs in suspension culture.

An interesting GO term that occurred for the angioblast cluster is hematopoiesis (**Figure 3.16a**). It suggests that in 3D suspension culture angioblast-like cells are not entirely committed to the endothelial linage but are able to commit a linage towards blood cells under opportune conditions. In embryonal development this would refer to a developmental stage before blood islets are established. This further strengthens the assumption of cell cycle arrest of the angioblast-like cells in 3D suspension culture and the reaggregation on day nine has not changed that phenomenon.

3.1.5. Cell type maturation in Matrigel

Upon prolonging the culturing time, the vessels grew and branched in the Matrigel microenvironment. Using fluorescent reporter cell lines, it was possible to monitor spatial organization of ECs and mural cells in the 3D hydrogel environment. For this, a nuclear and a cytoplasmic reporter cell line were differentiated towards ECs. On day nine the nuclear reporter cell line was FACS-sorted for PECAM1$^+$, the cytoplasmic for PDGFRB$^+$, and following combined in a 1:1 ratio. The reaggregation in 3D suspension culture and embedding into Matrigel were performed equivalent to the differentiation of the scRNA-seq datasets. While on day 15 only cells with nuclear reporter cells were lining up to sprout into Matrigel, on day 17 also cytoplasmic reporter cells aligned in microvascular structures (**Figure 3.17**). This suggests a recruitment of mural cells or pericytes to microvascular structures that have been established by sprouting ECs in the first place.

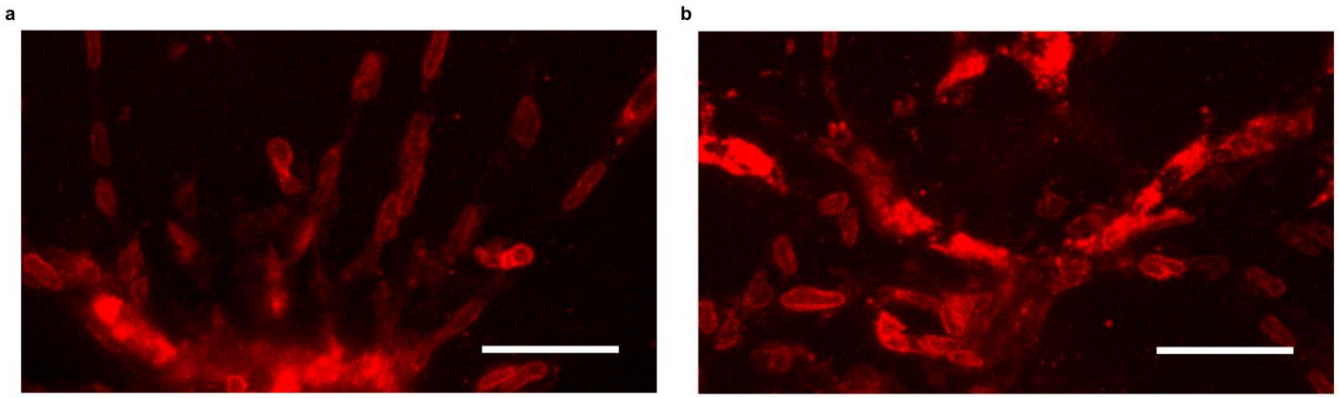

Figure 3.17 | Fluorescence imaging of reporter endothelial and mural cell lines. Two hiPSC cell lines, one with cytoplasmic and nucleus RFP marker were differentiated until day nine in the EC differentiation protocol while the nucleus reporter cell line was sorted for PECAM1^{+}, the cytoplasmic were sorted for PDGFRB^{+}. The sorted cells were combined in a 1:1 ratio for the reaggregation and embedding into Matrigel. Images were taken on day 15 (left) and (right) day 17. Scale: 50 µm.

To determine the genes activated during vessel maturation, we acquired single-cell transcriptomes of day 18 Matrigel cultures. In addition, we investigated the effect of ascorbic acid (AA) on EC maturation. AA increases the synthesis of the basal laminal protein collagen IV and reduces vessel permeability[179,180]. Bright-field imaging showed that vessel length and branching network were comparable in the presence and absence of AA (**Figure 3.18a**). To extract DEGs responsible for EC maturation, single-cell transcriptomes of all day 12 and 18 Matrigel cultures were clustered together (**Figure 3.18b**). ECs formed on day 18 of differentiation in the presence and absence of AA, where a low fraction of migrating and coalescing cells was still observed on day 18 (**Figure 3.18c**).

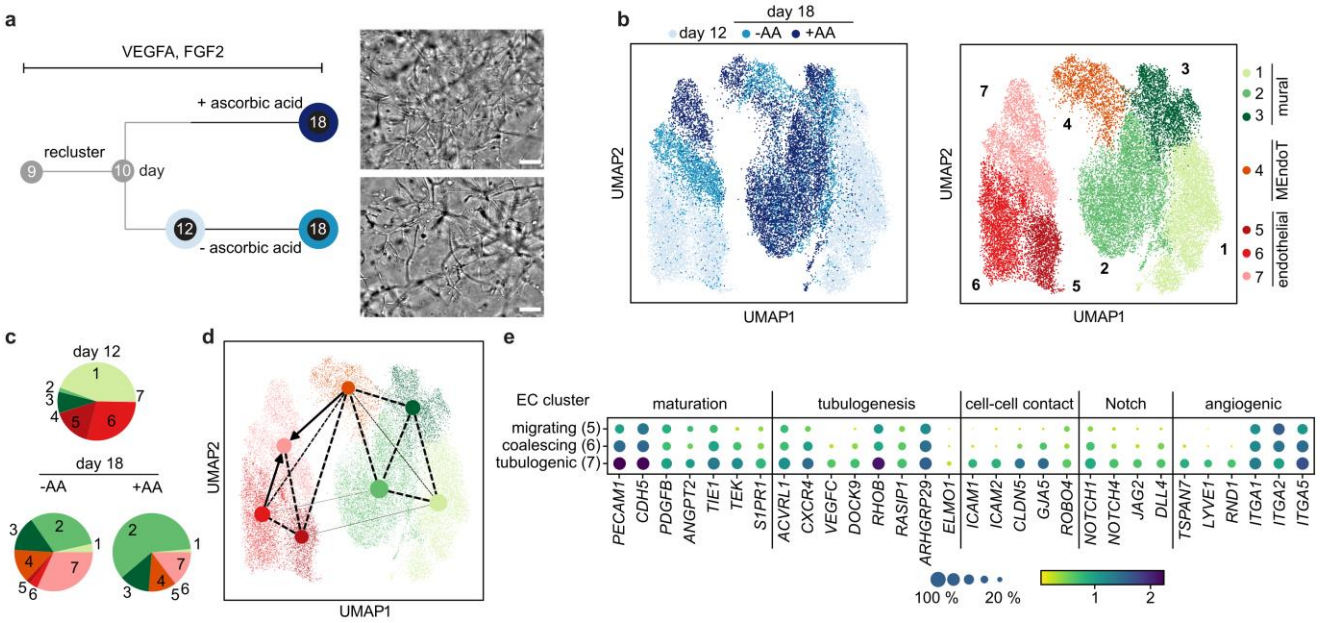

Figure 3.18 | Maturation of endothelial cells in Matrigel in presence and absence of ascorbic acid. a, Experimental timeline and applied conditions with corresponding bright-field images of stem cell-derived EC and mural cell cultures in Matrigel on day 18. Scale: 50 µm. **b,** Left: UMAP projection of scRNA-seq data from day 12 and two samples from day 18. Right: UMAP projection colored for the annotated cluster 1: mural cell, 2 and 3: pericytes (P1 and P2), 4: MEndoT, 5: migrating ECs, 6: coalescing ECs, 7: tubulogenic ECs. The dataset contains 13159, 6373, and 8684 cells for days 12 and 18 without and with ascorbic acid in the media, respectively. **c,** Cell type composition in the samples. **d,** Velocity analysis of the single-cell transcriptomes, where the latent time is colored on the UMAP plot. **e,** Expression levels of representative DEGs sorted by function for the three EC transcriptional states. The color intensity and dot size denote the normalized cluster mean expression level and the fraction of cell expression for the corresponding gene, respectively.

Upon addition of AA to the cell culture medium the fraction of mural cells increased compared to cell culture without AA (**Figure 3.18c**). A cell cycle analysis of scRNA-seq data set showed that addition of AA led to an increase of the number of proliferating cells in all cell clusters, however over proportionally in the mural cell cluster (**Figure 3.19a** and **b**). The highest impact is on the MEndoT population amounts to 7.7% more cells in active cell-cycle. The difference in the proliferation of the mural or pericyte clusters with 6.7% while for the ECs it only differs around 3.3%.

The transcriptomes of the mural cells from day 18 converged partially with the transcriptomes of day 12 and formed in total three clusters (cluster 1-3). Next to the EC and mural cell clusters the scRNA-seq data revealed a cell cluster, in which cells expressed mural and endothelial markers including *PDGFRB* and *PECAM1* (**Figure 3.19c**). Besides the higher expression of endothelial marker genes, a Pearson correlation analysis of the variable genes revealed higher proximity to mural cell clusters than the endothelial ones (**Figure 3.19d**). The PAGA connectivity based on the RNA velocity indicated that this population developed from the mural cells towards ECs (**Figure 3.18d**), wherefore we assigned these cells to mesenchymal to endothelial transitional cells.

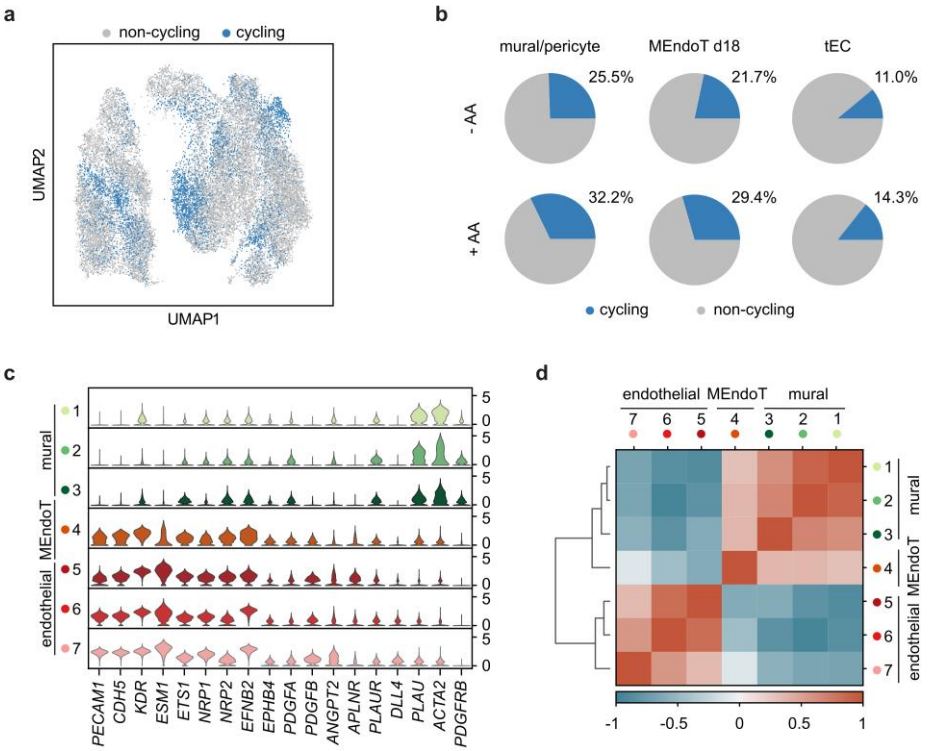

Figure 3.19 | Single-cell transcriptomes analysis of stem-cell derived endothelial cells form microvessels in Matrigel culture on day 18 of differentiation with and without ascorbic acid. a, UMAP plot of sc-transcriptomes acquired from cells in the Matrigel cultures on days 12 and 18. Blue colored dots represent cells expressing genes relating to S, G2, or M-phase. **b**, Pie charts show the percentage of cells undergoing the cell cycle within the respective clusters in presence and absence of ascorbic acid (AA). **c**, Violin plot of marker genes used to assign the cell cluster. The density distribution indicates the normalized cluster mean expression. **d**, Cell cluster Pearson correlation plot.

Evaluation of the DEGs between ECs revealed that on day 18, ECs increased the expression levels of cell-cell contact genes as *ICAM*s and *CLDN5* (**Figure 3.18e**). Furthermore, the expression of key tubulogenesis genes, such as *RASIP1*, *RHOB*, *ELMO1*[181], and *ARHGAP29*[182], were upregulated. In addition, the *DOCK9* gene expression level increased, which is a *RAC1* activator responsible for vascular lateral branching[183]. Based on the expression pattern, we assigned day 18 ECs to tubulogenic EC (tEC). To confirm this we prepared IF staining of ICAM2 and CLD5 within cryosection of day 12 and 18 hydrogel cultures (**Figure 3.12b** and **c**). Both DEG markers increased clearly in vessel structures on day 18. In line with the PAGA analysis, a Pearson correlation of the EC transcriptional states from days 12 and 18 showed that the tubulogenic state was closer to the coalescing than to the migrating state (**Figure 3.19d**). In addition to the changes in the levels of genes controlling the structural change, Notch signaling was upregulated again compared to the migrating and coalescent states in the forms of *NOTCH1, 4,* and *DLL4*. Upon increasing the resolution of the Leiden clustering, it is possible to separate ECs cultured in the presence and absence of AA; however, DEG change was minimal. Transcription factor analysis of the DEGs from tECs did not reveal differences in cECs.

Ascorbic acid is described to improve vessel maturation by acting as a cofactor for the collagen hydroxylation. Its reductive functionality is used to generate hydroxylation of proline and lysine to hydroxyproline and hydroxylysine [184]. After secretion, collagen peptides are assembled to triple-helical structures whose formation depends on the processed amino acid side chains[185]. To investigate this effect in the Matrigel setup, we added AA to the medium for one condition from day 12 on. Even if the gene expression of *COL4A1,2* is just moderately elevated in the +AA condition, the immune fluorescence staining represented strong differences in the presence of the protein (**Figure 3.20**).

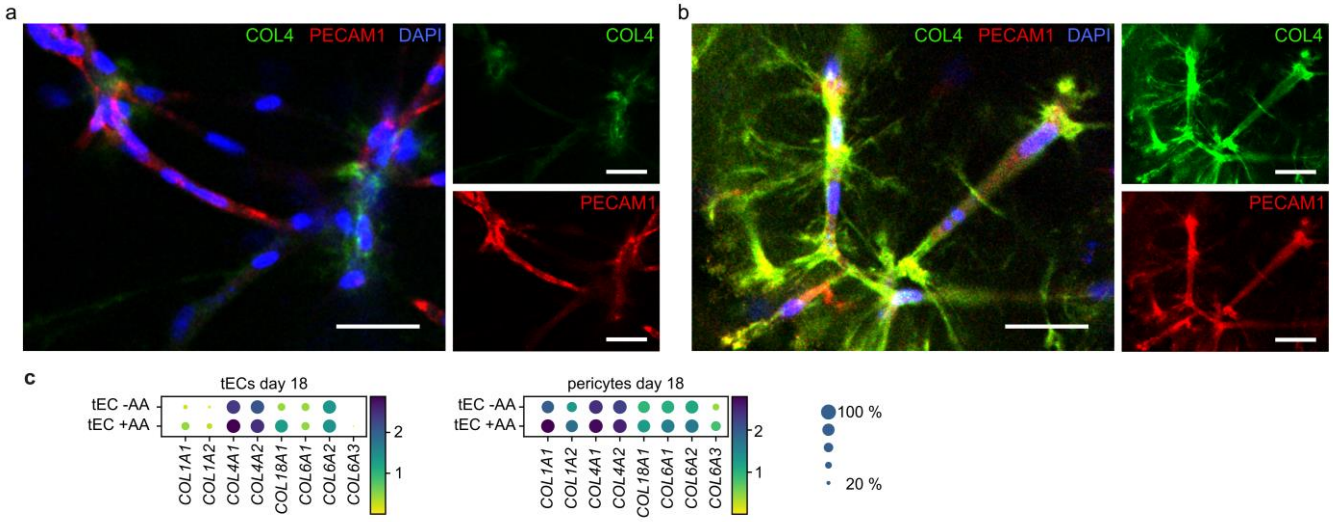

Figure 3.20 | Immunofluorescence images of microvessels at day 18 in the Matrigel culture with and without the addition of ascorbic acid. a, Immunofluorescence image of Matrigel culture on day 18 without and **b**, with AA addition to the media, where COL4, DAPI and PECAM1 were counterstained. Scale: 20 µm. **c**, Cluster-specific proliferation rate for day 18 specimen with and without ascorbic acid. **d**, Expression levels of the collagen IV chains between the condition with and without AA inside the tEC cluster. Color intensity and dot size denote the normalized cluster mean expression level and fraction of cell expression for the corresponding gene, respectively.

The six days of prolonged Matrigel cultivation accomplished a maturation in EC phenotype and structure of established microvessels. Further, by adding ascorbic acid the increased expression of collagen IV on the level of gene and protein was registered. Vessel structure formation within the hydrogel culture was accompanied by a change in ECM-integrin expression level. While *ITGA2* was upregulated in mEC, a gradual increase of *ITGA1* and *ITGA5* was detected towards the cEC and tEC cell stages (**Figure 3.18e**). Blocking collagen-binding ITGA1 and ITGA2 with antibodies has been shown to reduce angiogenesis[186]. The EC and mural cell stage-assigned integrin profiles provide new dynamic insights. […] ECs on day 18 of differentiation did not show a specification marker for arterial or any other vascular cell type. Notably, the scRNA-seq data from day 18 of the Matrigel cell cultures indicated a mesenchymal to endothelial transition. Such a mural cell plasticity has been observed before in heart tissue after injury[187], but not in in vitro conditions.

The increase of the basal lamina protein COL4 (**Figure 3.20a** and **b**) suggests a higher maturation state in the ECs of the +AA what is also highlighted in the velocity analysis (**Figure 3.18b** and **d**). Without AA as a cofactor the enzymes are not capable to generate triple-helical structures and form collagen bundles[185]. Consequently, singular peptide chains are situated loosely in the hydrogel around the cells and may be independently diffused into the media. This could explain the tremendous difference for protein presence in comparison to the moderate gene expression variation (**Figure 3.20a** and **b**). Besides *COL4*, the expression of *COL1* is upregulated by AA. This occurred in ECs as well as mural cells. In fibroblasts the addition of AA increased the collagen protein amount 2-3-fold[188]. Chojkier *et al.* used 200 µM AA while in this study 60 µg/mL which corresponds to 341 µM were used. Further, the addition of 100 µM AA to culture media amounted to significant increase of COL1 and COL4 expression on the mRNA and protein level[189]. Since maturation of ECs is a complex interplay of inter cellular and matrix signaling, the addition of AA as a single factor is unlikely to have a major impact even though it impacts the expression and assembly of the two basal membrane collagens effectively.

The increase of proliferative mural or pericyte populations in the +AA condition has to be further analyzed (**Figure 3.19b**). While on the one hand the increase of the mural phenotype might be a negative side effect in tissue engineering, on the other hand it might be linked to the process of maturation in which pericyte recruitment occurs. From this point of view, proliferative pericytes are beneficial for establishing maturated microvessels in engineering approaches. This assumption is further underlined by data from other studies that represent a link of the pericyte recruitment and their proliferation[16].

The data in our study suggests a mesenchymal to endothelial transition (MEndoT) while mainly the reverse direction is described in embryonal development, wound healing, state of inflammation and cancer progression[190–192]. In states of tissue damage and inflammation, ECs undergo EndoMT to support the reconstruction of surrounding tissue environment. One example here is the transition to cardiomyocytes after an acute cardiac ischemic injury[193]. However, the mechanism also contributes to the generation of cancer-associated fibroblasts[191]. Nevertheless, the MEndoT seems to play a role in neovascularization and was recently investigated in cardiac hypertrophy[194]. Here, it helped to preserve cardiac function *via* induction of fibroblasts towards an endothelial lineage. Since the Matrigel environment and media composition supports neovascular development, these findings underpin the suggestions of having MEndoT phenotype in the cell population.

3.1.6. Cell type plasticity

ECs and mural cells co-evolve in the differentiation from hiPSCs (**Figure 3.1**), undergo transitions between each other (**Figure 3.1**, **Figure 3.11**, and **Figure 3.18**), and adapt their gene expression pattern to environmental and functional changes (**Figure 3.11**). This refers to a high cell type plasticity which is examined in this part of the study. First, a 3D suspension culture differentiation was performed but instead of an endothelial induction on day three, ActivinA (INHBA) and PDGFB were used to induce the mural lineage (**Figure 3.21a**)[49]. This led to the generation of PDGFRB$^+$ mural cells in the ratio of 80% while a PECAM1$^+$ EC population emerged as a fraction of 7% on day six and 6% on day nine (**Figure 3.21b**). This highlights that the different morphogens used for induction *in vitro* lead to a majority of one of the cell types however cellular signaling may induce the co-evolvement of the other one.

Second, FACS was applied to separate PDGFRB$^+$ and PECAM1$^+$ populations on day six (**Figure 3.21c**). This approach targets the question of how stable the endothelial and mural cell type is maintained while growing separate of the other one which is of special interest in cell type engineering. Next, the sorted cell types as well as an unsorted control were cultivated in 2D since pure EC aggregates were not able to survive under the 3D suspension culture condition.

The PECAM1$^+$/PDGFRB$^-$ sorted ECs maintained stable expression of PECAM1$^+$ whereas no PDGFRB$^+$ population established over six days in the 2D monolayer culture format (**Figure 3.21d**, center). The unsorted control registered an increase of the PECAM1$^-$/PDGFRB$^+$ population on day nine whereas it was decreased again on day 12. The population of PECAM1$^+$/PDGFRB$^-$ cells represented a steady increased from 61% on day six to 72% on day 12 (**Figure 3.21d**, left). The cells that were sorted for PECAM1$^-$/PDGFRB$^+$, displayed a continuous decrease of the PDGFRB$^+$ phenotype towards a double negative marker expression (**Figure 3.21d**, right). The 61% of PDGFRB$^+$ cells amounted in 14% while the PECAM1$^+$ population fluctuated between a ratio of 0.7 to 2.5%. Double positive populations were a small subgroup in all conditions and at all timepoints. Morphologically, the ECs and mural cells can be distinguished in bright field microscopy (**Figure 3.21c**, lower). Whereas ECs align to each other in the 2D monolayer culture, mural cells express a more convoluted morphology. The black arrows mark noticeable morphological different cells in the unsorted control of the 2D cell culture that resemble the cells in the PDGFRB$^+$/PECAM1$^-$ sorted condition. Further, this cell morphology is not visible in the PECAM1$^+$/PDGFRB$^-$ sorted monolayer.

In a further step, the adjustment ability of longer cultivated ECs between different growth conditions was investigated. Therefore, the ECs mural cell populations that were neovascularizing Matrigel were dissolved and reaggregated in 3D suspension culture and incubated under these conditions for four more days (**Figure 3.21e**). The combined scRNA-seq analysis with the 3D suspension culture on day 12 represented a clustering together of both, endothelial and mural populations for the two conditions and highlights the adaptivity of mural and endothelial cells under changed conditions.

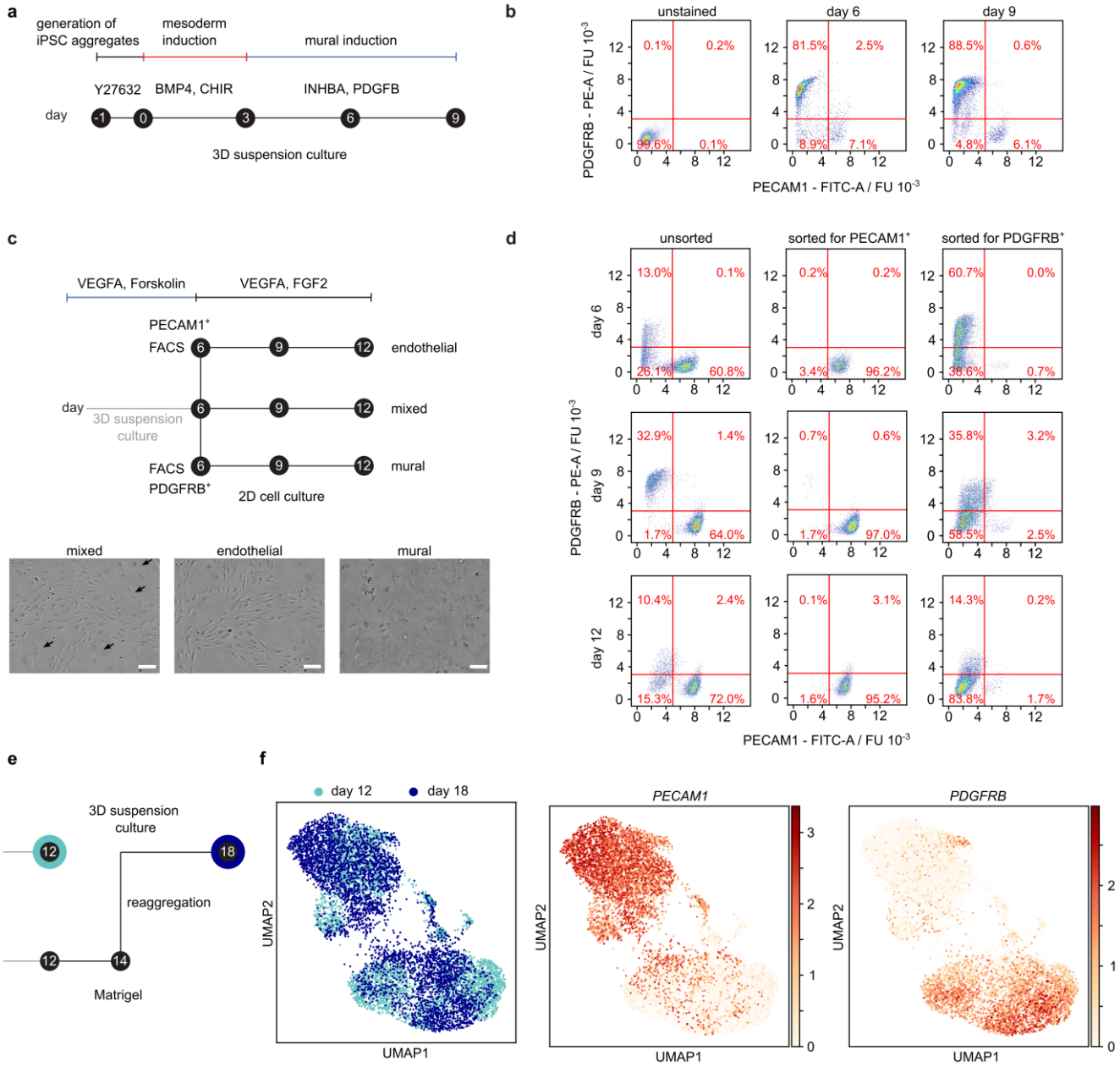

Figure 3.21 | Cell plasticity of ECs shown upon various culturing conditions. a, Schematic of mural cell differentiation in 3D suspension culture, **b**, Flow cytometry analysis of days six and nine of the differentiation. **c**, Upper: Experimental overview of cell sorting on day 6 of the EC differentiation for endothelial and mural marker PECAM1 and PDGFRB and the three different culturing conditions. Lower: Bright-field images of the three conditions, after three days in 2D cell culture. **d**, Flow cytometry analysis of sorted and unsorted cell populations from days six, nine, and 12 cultured under EC differentiation media. **e**, Schematic of the experimental timeline and applied culture conditions. First, microvessels were formed upon transfer of 3D suspension culture into Matrigel on day 10 of differentiation. Second, the microvessels were disaggregated by harvesting single-cell solution on day 14. Third, single cells were reaggregated and cultured as 3D suspension culture until day 18. **f**, UMAP plots of single-cell transcriptomes measured from 3D suspension cultures cultured to day 12 of differentiation (cyan) and reaggregated 3D suspension-cultured to day 18 (dark blue, left), colored for the gene expression of the EC marker *PECAM1* (middle) and mural cell marker *PDGFRB* (right).

In this part, the plasticity of ECs and mural cells could be highlighted in different approaches. Mesodermal cells express *TGFB1* and *PDGFB*, where the combination of these two factors is used to induce vascular smooth muscle cell development *in vitro* in the same mesodermal precursor cells [49]. Interestingly, stimulation of mesodermal cells with ActivinA (INHBA), a member of the TGFB family, and PDGFB led directly to PDGFRB$^+$ mural cells as reported before. Additionally to the PDGFRB$^+$ cells, we detected in the differentiation approach a fraction of 7% of PECAM1$^+$ cells on day six and 6% on day nine (**Figure 3.21b**). This demonstrates that EC and mural cells co-evolve in both differentiation directions *in vitro*.

Further, in prolonged 2D cell culture of unsorted and PDGFRB$^+$/PECAM1$^-$ sorted cells a decreasing of the PDGFRB$^+$ population was observable (**Figure 3.21d**). This finding contrasted with that in the 3D suspension culture, where the PDGFRB$^+$ cell fraction was even increased within the same culturing interval (compare **Figure 3.1**, **Figure 3.4**, **Figure 3.11**, and **Figure 3.21e**). We highlighted that cell-cell communication occurring inside of the aggregates of angioblasts and mural cells in the 3D suspension culture which is minimized in the 2D monolayer culture. Therefore, we suggest that this signaling is important for the maintenance of the PDGFRB$^+$ phenotype. In conclusion, cell-cell signals between angioblasts and mural cells lead to stabilization of the mural cell fraction.

The remaining plasticity of the ECs can be once more exemplified by disseminating of the Matrigel culture and reformation of the 3D suspension culture, where corresponding single-cell transcriptomes show the interconversion of the cell states (**Figure 3.21e** and **f**). On the one hand, this can lead to the suggestion that the differentiated cell types are not in a mature state, on the other hand to the high adaptivity of ECs and changes in their transcriptome in acute injury or inflammation[26]. In intact vessels, ECs are in a quiescent modus while certain environmental factors, as inflammation, angiogenic signaling or vessel injury, disturb this endothelial quiescence and initiate EC activity. Vessel injury induces a neighboring ECs to re-enter in the cell cycle as a response to stress signaling that underpins endothelial plasticity[27]. The degradation of the Matrigel microenvironment and dissociation of the formed microvessels is a situation that induces stress and probably has activated this ability for transcriptional changes while the transfer to 3D suspension culture led to the adaption to the similar phenotype as on day 12 under these conditions.

3.1.7. Cell-cell interaction profiling

The single-cell transcriptomes of the microvascular culture on day 18 revealed that mural cells of clusters 2 and 3 (**Figure 3.18b**) expressed pericyte markers *NG2*, *RGS5*, or *NT5E*. Among the DEGs were further genes previously found to be enriched in pericytes, such as *POSTN* and *PDLIM3*. IF stainings of cryosection from day 18 hydrogel cultures showed that indeed PDGFRB positive cells aligned on the PECAM1 cells (**Figure 3.22**). Therefore, we assigned mural clusters 2 and 3 as pericytes, P1 and P2, respectively.

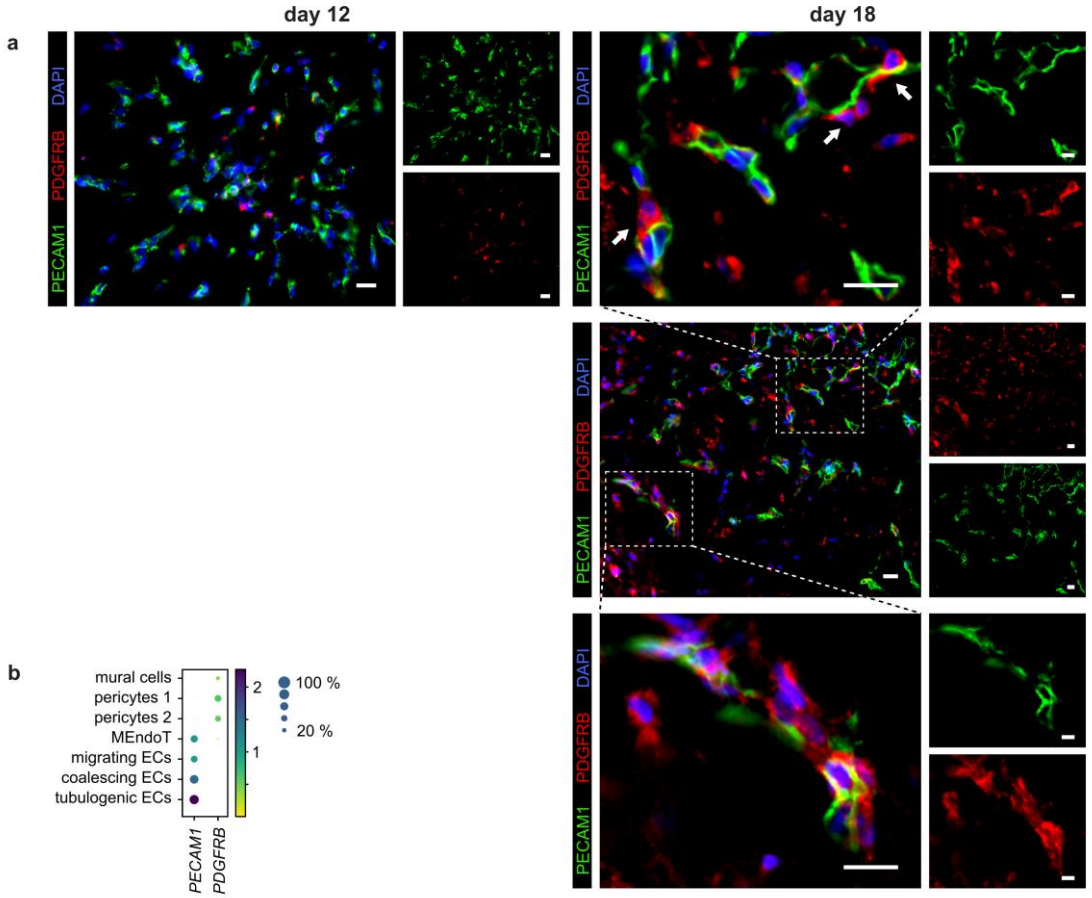

Figure 3.22 | IF images of cryo-sections prepared from 3D hydrogel cultures on day 12 and 18 of differentiation. a, Location of mural cells to ECs are shown upon co-staining of PDGFRB and PECAM1, respectively. On day 18 (right panel) pericyte recruitment is highlighted by arrows. Scale: 20 µm. b, Dot plot of mRNA expression of genes whose corresponding protein staining is shown above. The color intensity and dot size denote the normalized cluster mean expression and the fraction of cells expressing the gene, respectively.

While cells of the P1 cluster expressed the transcription factor *FOXF1*, cells of the P2 expressed *GATA4* (**Figure 3.24a**), where both were found in primary pericytes[195]. Localization of the subcellular pericytes within the Matrigel culture failed either due to the low amount of specific antibodies for the DEGs or cross expression of the marker. Mural cells of cluster 1, consisted mainly of cells from day 12 and clustered with mural cells of the 3D suspension culture. The higher expression level for the mesenchymal TF *HAND1/2* and *ACTA2*, argued that mural cells of cluster 1 were in a premature stage. The velocity analysis of the

mural cell transcriptomes, however, did not show any direction or evolving latent time underlining the plasticity of the cells.

In the last step, we asked whether it is possible to infer cell-cell communication signals leading to vessel maturation from the scRNA-seq datasets of high plastic cell types, stability, reduced proliferation, and migration [16,196]. To trace cell-cell signaling between pericytes and tECs, we performed a ligand-receptor analysis on the single-cell transcriptomic data. Here, one prominent algorithm is CellphoneDB[93] that uses the count matrix and the list of annotated cell types for generating its prediction (**Figure 3.23a**). The highest ranked interaction predictions belonged to cell-cell contacts between different collagen and integrin complexes (**Figure 3.23b**). Also, FGF and BMP signaling showed high rankings but could not recapitulate literature confirmed signals between ECs and pericytes, for example, *TGFB1* or *VWF* in our dataset.

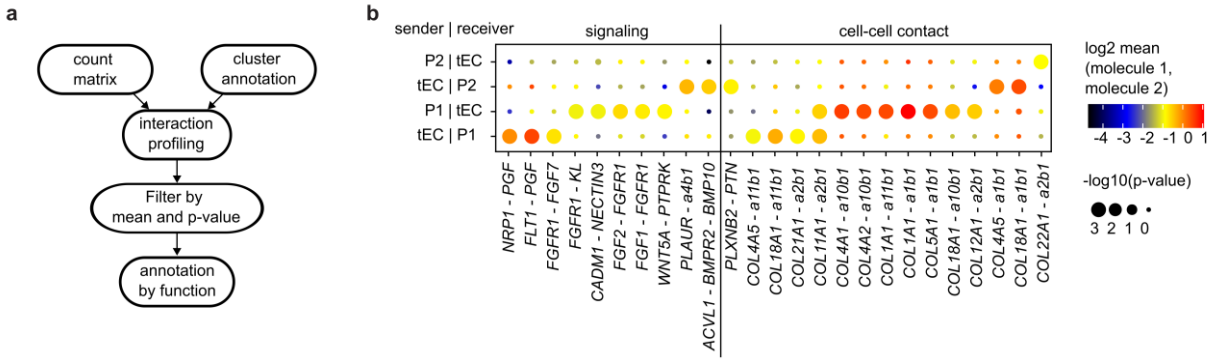

Figure 3.23 | Ligand-receptor interaction inferred by CellphoneDB. a, Schematic of CellphoneDB procedure. **b**, Dot plot with the highest calculated ligand-receptor links between tubulogenic ECs (tEC) and the two pericyte clusters (P1/2), respectively. The circle size indicates the p-values and the color of the mean expression levels.

Therefore, we performed a NicheNet analysis[94], which infers ligand-target links between interacting cells by combining DEG data with prior knowledge of signaling and gene regulatory networks (**Figure 3.24b**). NicheNet was applied bidirectionally by investigating ECs as ligand sender and receiver cells. According to the NicheNet results, that is, ligand activities and receptor and target gene interaction potential, were further filtered for gene expression levels to increase the probability of finding relevant interaction links (**Figure 3.24c**). For ECs as senders and pericytes as receivers, the algorithm predicted strong ligand activity for *TGFB1*, *BMP4*, and *VWF*. All three have been reported previously to be central for communication between both cell types[197]. *TGFB1* acts in paracrine and autocrine signaling, while tECs express the *TGFB1* receptor *ALK1* (*ACVRL1*) and pericytes express *ALK5* (*TGFBR1*), which is consistent with current reports using *in vitro* culture systems[198]. The inferred ligand activity for *PDGFB* was low and thus filtered out, although *PDGFB* was highly expressed. This is explained by the fact that we used DEGs of the entire differentiation trajectory to rank ligands, and since *PDGFB* is expressed strongly at all times, the corresponding target genes within the DEG cannot be expected.

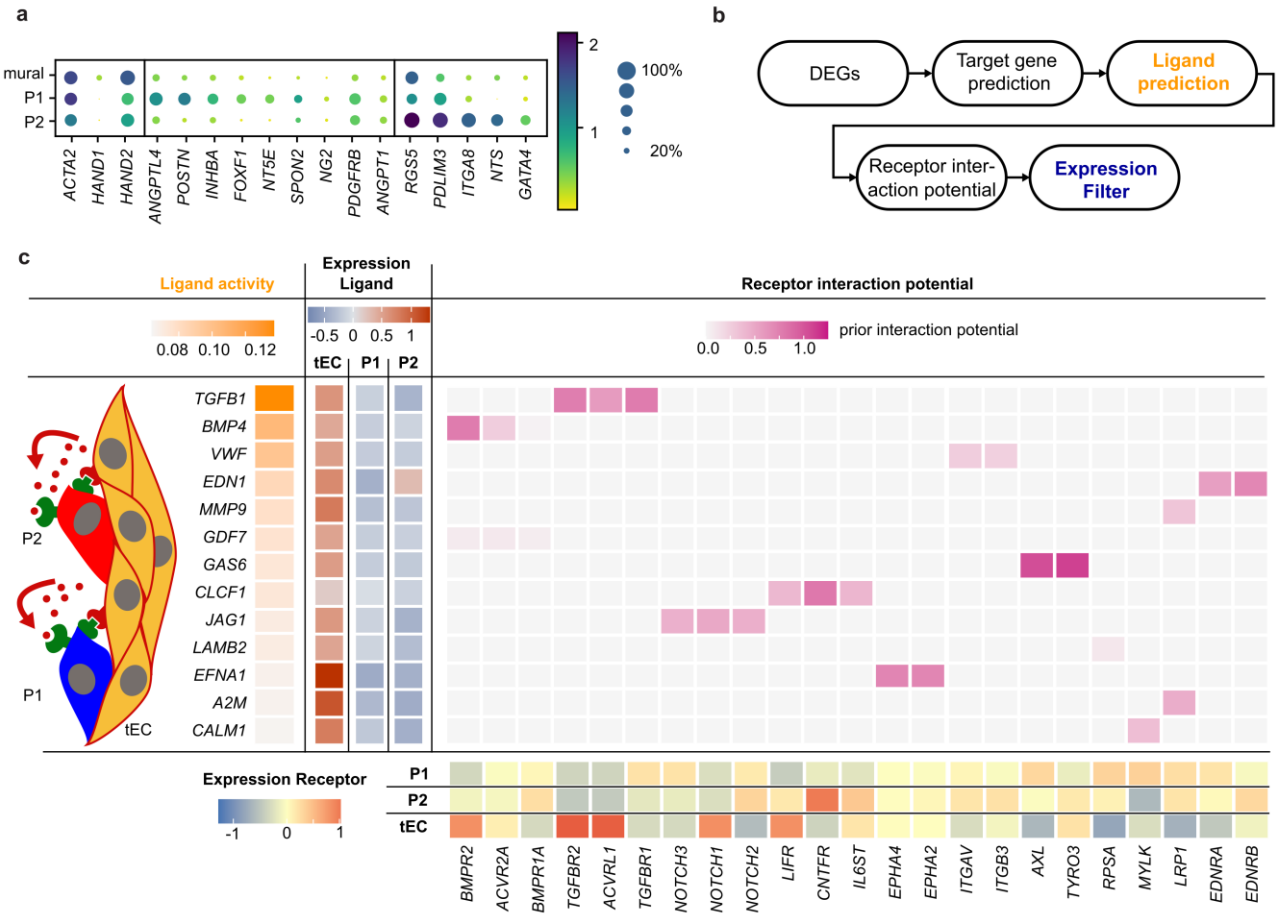

Figure 3.24 | Inferred signaling between endothelial cells and pericytes during vessel maturation from single-cell transcriptomics. a, Gene expression analysis of the mural cells cultured in Matrigel up to day 18 of differentiation. The P1 and P2 clusters show the expression of pericyte markers (bold). Color intensity and dot size denote the normalized cluster mean expression level and fraction of cell expression for the corresponding gene, respectively. **b**, NicheNet analysis workflow with the expression filter to infer ligand-receptor interactions. **c**, Ligand-receptor pairs inferred with NicheNet, where single-cell transcriptomes of tEC are used as senders and the combined P1 and P2 pericytes subpopulation as receivers for the analysis.

In turn, the top predicted ligand activities represented signals that occurred during vessel maturation rather than during vessel formation and the pericyte recruitment phase. Individual NicheNet analysis between ECs and P1 or P2 cells did not differ from the combined analysis. In contrast, the NicheNet analysis for P1 or P2 as signal senders and tECs as the receiver showed that P1 and P2 pericytes exhibited shared, as well as individual ligands (**Figure 3.25a**).

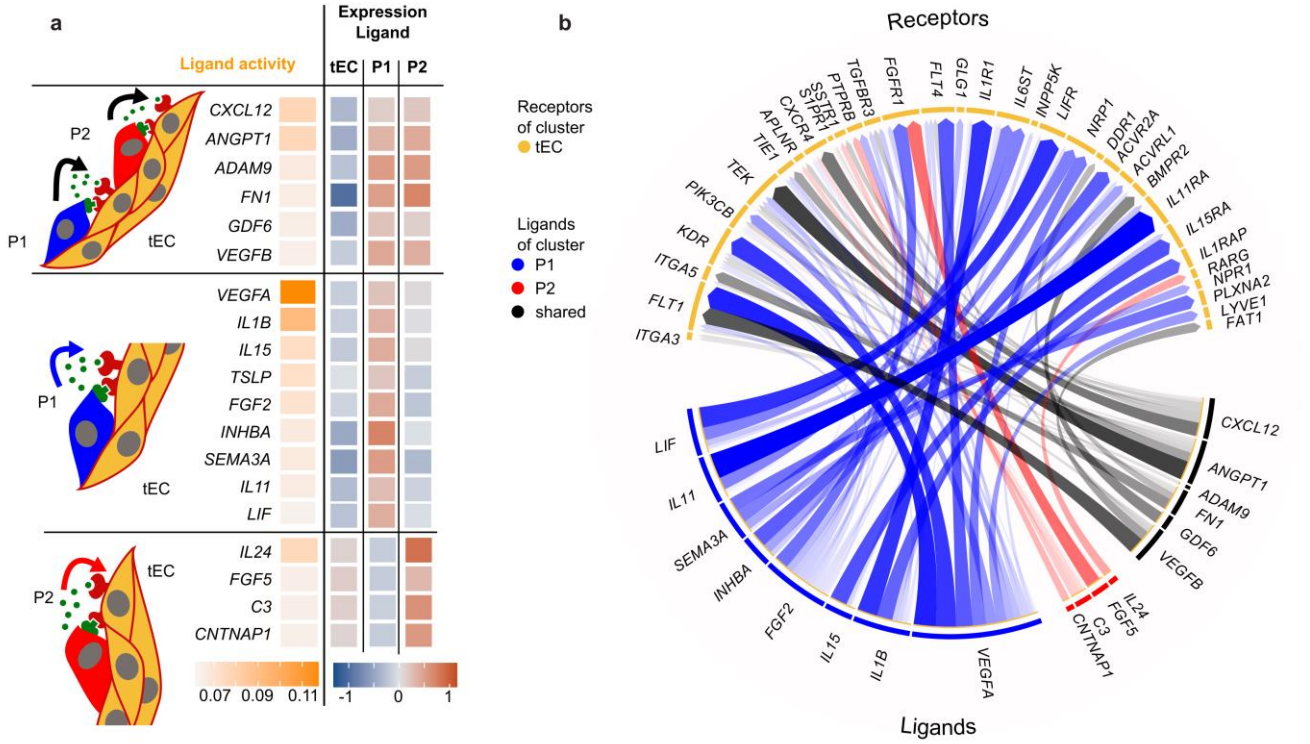

Figure 3.25 | Ligand-receptor interaction inferred by NicheNet with pericytes as sender and tECs as receiver. a, Ligands sorted by activity and grouped for shared and individual expression in the two pericyte clustes, respectively. Overlapping (top rows) and unique pericyte ligands (middle and bottom rows) are listed in table format. **b**, Ligand-receptor pairs are presented in the circular chord diagram.

The shared ligands included *CXCL12*, *ANGPT1*, fibronectin, and *GDF6*, which antagonize VEGF signaling to promote junctional stability and vascular integrity[199,200]. The strongest individual predicted ligand activity of P1 cells was *VEGFA*, followed by a set of cytokines (*IL1b* and *IL15*) and *FGF2*. For P2 cells, the ligands with the strongest individual activity were *IL24* and *FGF5*. The corresponding tEC receptors are shown in **Figure 3.24Figure 3.25b** and more detailed in matrix format of **Figure 3.26**.

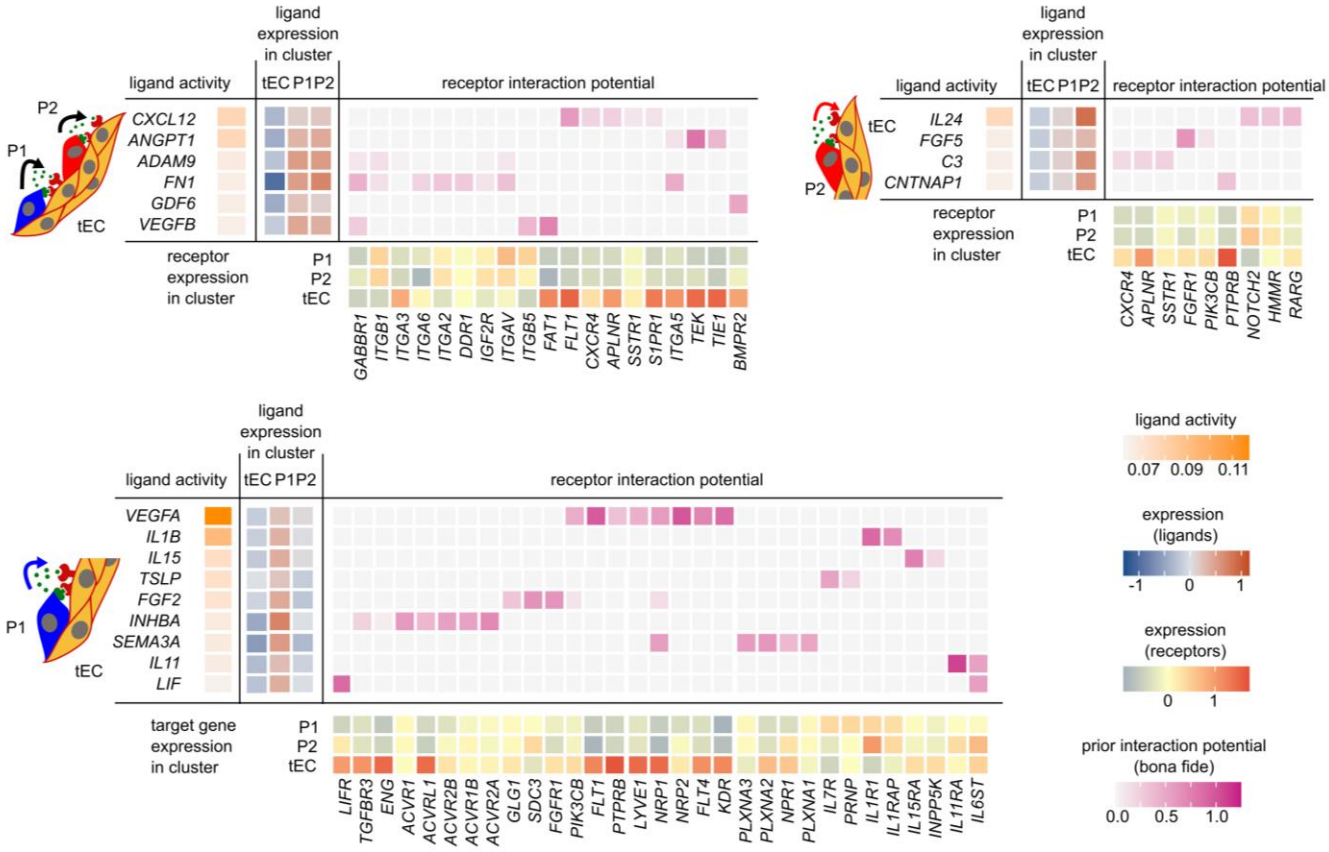

Figure 3.26 | Matrix plots of ligand-receptor interaction inferred by NicheNet with pericytes as sender and tECs as receiver.. Matrix plots of the predicted receptors for the tEC cluster as receiver cells and the ligands sent by either one or both of the two pericyte subpopulations. Detailed version of **Figure 3.25c**.

All inferred cytokines have proven pro- or anti-angiogenic function in cancer angiogenesis; however, their function during neovascularization and vessel maturation is unknown. The resolved heterogeneity of cytokine ligands between pericytes argues for functional differences between P1 and P2 during vessel maturation. The inferred ligands of the two pericyte clusters were compared in detail with literature to discover differences which may allow a functional annotation (**Figure 3.26**). The focus here laid on the evolvement of the ligands in endothelial development. Following ligands were assigned expression wise to pericyte cluster P1. Pericytes use **IL1B** for the regulation of microvascular permeability[201]. In the tumor microenvironment **IL1B** stimulates angiogenesis, by contributing to VEGF expression[202]. **IL15** promotes the expression of hyaluronan in EC that is an important element in endothelial cell–cell junctions defining the endothelial barrier[203,204]. **TSLP** needs interleukin 7 (IL-7)R-α chain in the functional TSLPR for signaling[205]. Further on, it is reported to simulate angiogenesis in cervical carcinoma[206]. **FGF2** binds on ECs to FGFR1 and increases PDGF-BB and PDGF-DD biosynthesis. Both are ligands for the PDGFRB that is important in pericyte recruitment. In tumor microenvironment it is reported that **FGF2** is secreted by tumor cells. Pericytes express FGFR2 as a receptor for **FGF2** and is involved in the upregulation of PDGFRB in response to signal transduction[207]. SDC3 is another receptor of **FGF2** that is assigned by NicheNet, is involved in retinal development in rats[208]. The homodimer of **INHBA** (ActivinA) induces TGF-signaling in ECs[209]. Activin receptors are essential in vascular development. While ACVR1 for example, is

necessary in development of vessels in the retina, a disturbance of ACVRL1 is leading to malformation of zebrafish cranial vessels[210,211]. **SEMA3A**/NRP1 signaling is important for maturation of lymphatic vascular networks and has a comparable influence on vascular permeability as VEGF[212,213]. The **IL11** and its receptor IL11RA are both higher expressed in the pericyte populations (**Figure 3.26c**). In literature, the function of **IL11** is described as a player in autocrine signaling of VSMCs in vascular inflammation[214]. **LIF** is a member of the IL6 family that is expressed by pericytes under hypoxia[215]. Its receptor, LIFR is reported to promote tumor angiogenesis[216].

Following ligands represented a specificity for cluster P2 based on their expression. **IL24** has been shown to play a suppressing role in the angiogenesis of cervical cancer[217]. One of its receptors inferred by NicheNet, RARG, increases the tightness of the blood brain barrier[218]. **FGF5** was shown to promote angiogenesis in human aortic ECs as well as it stimulates proliferation, migration and tube formation of human umbilical vein ECs[219,220]. **C3** as an interaction partner of APLNR could be involved in the polarization of the ECs and thereby regulates migration and aligning of the ECs in vessels[221]. PTPRB, assigned as an receptor of **CNTNAP1**, dimerizes with TEK as well as with CDH5 and is an important player in blood vessel morphogenesis and endothelial junctions[222,223]. This confirms that the inferred cytokines and remaining ligands all have a proven function in developmental or cancer angiogenesis. Even if their function resolved the heterogeneity between pericyte clusters P1 and P2 during vessel maturation, it has not allowed the functional characterization of these.

The most compelling results of the single-cell analysis during the vessel maturation phase in a reduced *in vitro* microenvironment were the resolved cell-cell communication signals between the two transcriptional pericyte subpopulations and tECs. Pericytes are recruited by PDGFB signaling to capillary walls to stabilize integrity and tube assembly[196]. The directionality of the NicheNet ligand-receptor analysis allowed for the resolution of *TGFB1* signals in the sender and receiver cells. Next to the known EC-pericyte signaling factors, the complex cytokine profiles of the pericytes were most obvious. The maturation factor for pericyte formation is deciphered as the CXCR4/CXCL12 signaling axis but in addition to a variety of chemokines with unreported functions during the neovascularization process[224]. Using the DEGs calculated from transcriptional time trajectories within the sender-receiver model, it is possible to extract cell-cell communication signals at defined processes, as applied in the vessel maturation phase. Furthermore, several ligands and receptors that are reported in angiogenesis are inferred in the same turn as from the pericyte side *VEGFA* with among others the receptors *FTL1* and *KDR*, *ANGPT1* with its receptors TIE1 and TEK, as well as *S1PR1* (compare **Figure 1.1**).

The cytokine ligands that were inferred by NicheNet all have a described role in cancer angiogenesis. Further, in stages of inflammation cytokines as TNF-α, TGFβ, and IL-1β initiate an EndoMT wherein ECs aim to replace impaired surrounding tissue[225]. However, in our study we only observed the opposite directionality of transition. One suggestion would be that TNF-α signaling as a strong ligand in inflammatory processes has a major impact on the EndoMT. *TGFB1* was inferred as a ligand in the signaling from ECs

towards pericytes whereas some of the corresponding receptors (*TGFBR2* and *ACVRL1*) were significantly higher expressed in the tEC cluster itself. On the one hand this might highlight an autocrine signaling however the receptors are only predicted bases on the presence of corresponding target genes which were present in the DEGs of the pericyte clusters. On the other hand, *TGFBR1* was higher expressed in the P1 cluster and may have similar target genes that caused the prediction of the other two receptors. On the side of the pericytes, the gene of the TGF-family member ActivinA (*INHBA*) was inferred as a ligand with *ACVRL1* stated as one of receptor. This ActivinA signaling might also be a reason for the upregulation of TGF-signaling associated receptors in the tECs. Further, literature describes the bidirectional TGF-signaling axis between ECs and pericytes in angiogenesis and vessel maturation[2]. The receptor ENG, also inferred by NicheNet and highly expressed in tECs, is hybridizing in ECs with complexes of TGFBR2-ACVRL1 or TGFBR2-TGFBR1 to induce or inhibit endothelial migration, respectively[226,227]. With all these possible signaling variations it has to be admitted that a specific interpretation of these predictions is not feasible.

In addition, it has to be stated that CellphoneDB was designed having a special attention on interaction of protein complexes which explains the high number of interactions of integrin complexes in its results. Since CellphoneDB has a different database for and is not using DEGs but the count matrix as input, differences in output compared to NicheNet are not surprising. Apart from that both are predictable algorithms that can only give hints for new investigations.

In this study, the release of cells out of the Matrigel environment was performed with incubation of a collagenase/dispase solution for about 4 h at 37°C. The examination of different preparation protocols for scRNA-seq of adult mouse liver revealed that methodical biases are introduced[228]. Special attention laid on the digestion to single cells in which the incubation on ice reduced the observed stress response in the scRNA-seq data compared to incubation at 37°C. This raises the question, in how far the NicheNet inferred chemokine signaling is related to the neovascularization and maturation process. The long exposure to digestive enzymes might have altered the cellular transcriptome to some extend by inducing stress response. Apart from that NicheNet is a computational tool, and therefore predicted signaling would need to be verified experimentally to draft solid statements for cellular crosstalk. The present single-cell transcriptomic data and analysis of the vascular structure formation process can be used as a benchmark set for future *in vitro* vascularization approaches, differentiation attempts to alter the specification of ECs, or investigation into disease-specific gene functions. A detection of the receptors inferred by NicheNet at the protein and single-cell level together with the transcriptomic data would improve the significance of the computational results. A possible assay to perform this is CITE-seq that is meanwhile commercially marketed by 10x Genomics. The next section evaluates a similar oligonucleotide-labeled antibody-based approach that is not depending on the commercial 10x Genomics platform. Computational tools for the integration of this protein readout into the scRNA-seq pipeline are already available and improve the meaningfulness of the generated data[229].

3.2. Single-cell antibody readout

The transcriptomic readout of scRNA-seq offers an immense dataset with multiple possibilities of computational analysis. The computational screening for protein-protein interaction as it was performed in the previous chapter is a perfect tool for identifying receptors that are possible targets. The presence of the actual protein would introduce a higher significance. Further, for some purposes no sufficient high sensitivity is provided by mRNA detection as for example the presence of certain TFs (see **Figure 3.5**). There, hypothetical TFs were calculated depending on the expression of other genes using a TFEA. Due to lower turnover rates, the actual presence of the protein can be verified by targeting the respective TF with a specific antibody. In other cases, the mRNA content is not necessarily correlated with the referring protein. Therefore, an additional protein detection would give a higher significance in this case. Antibodies with their high specificity and binding affinity have proven their capability in molecular biology *e.g.* in immune fluorescence, western blot, and therapeutic applications. To use antibodies in sequencing strategies, an amplifiable oligonucleotide tag is coupled to them which is caught side by side with the transcriptome. The possibility of including this antibody-based protein sequencing into a combinatorial barcoding technique for scRNA-seq is investigated in the following part.

3.2.1. Combinatorial Barcoding

The previously applied droplet based scRNA-seq method from 10x Genomics is focused on deep sequencing of up to 10 000 cells per experiment and as a commercial method it has a relative high cost per transcriptome of a single-cell. Combinatorial barcoding is a further method to uniquely label cells. The procedure of split and pool cells in different wells and attaching well-specific barcodes to their transcripts over several rounds generates high numbers of individually labeled cells. The initial concept of using this technique for sequencing transcriptomes of large cell numbers was implemented by Rosenberg *et al*. and is termed split-pool ligation-based transcriptome sequencing (SPLiT-seq)[82].

An important step to keep in mind is to consider the number of cells that can be uniquely labeled with the used barcoding system. According to Rosenberg *et al.*, not more than 5% of sequenced cells shall collide. The probability for a collision **p** is described by the following equation (1) where **x** refers to the total number of barcodes and **n** to the number of cells that are supposed to get sequenced.

$$n = p \times x \tag{1}$$

The total number of barcodes **x** results in the product of the rounds with the respective barcoding pots. Using this formula, in four barcoding rounds with a certain high number of pots, several hundred thousand up to over a million cells are uniquely labeled for Illumina NGS.

We adapted the SPLiT-seq strategy developed by Rosenberg *et al.*[82] and used an oligonucleotide, containing the Read 1 sequence, a barcode and a capture sequence, coupled to an antibody (**Table 2.1**

and **Figure 3.27**). The first barcoding round that acts as primer for the “reverse transcription” needs a different oligonucleotide than used for mRNA labeling however the later steps are equivalent to the original protocol of Rosenberg *et al.*.

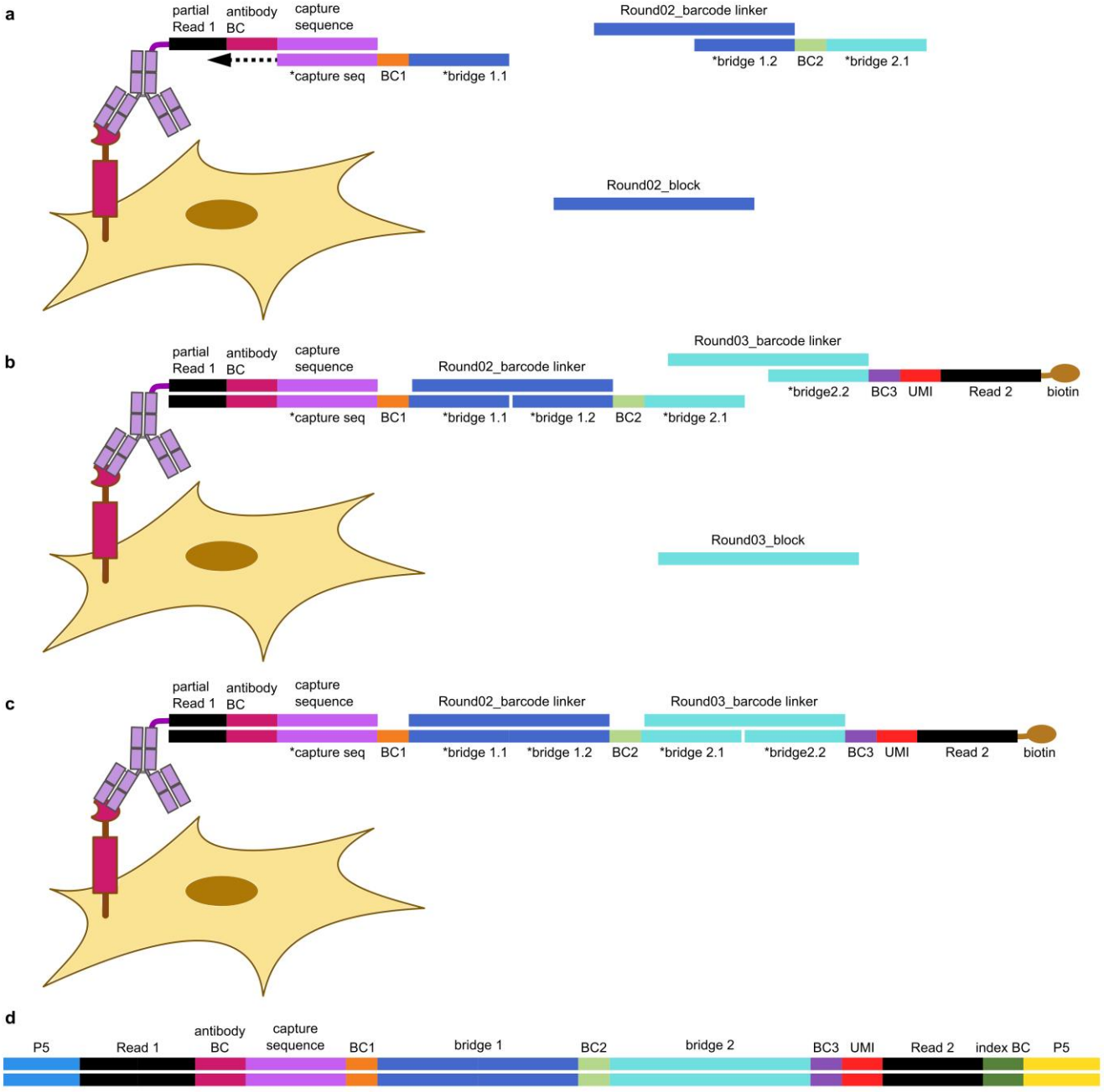

Figure 3.27 | Schematic of library generation in the antibody-SPLiT-seq technique. a, Hybridization of the first barcoding sequence (BC1) to the antibody coupled oligonucleotide (antibody BC) that acts as primer for reverse transcription and binding site for the second barcoding sequence (BC2). This one is covalent connected by ligation while unbound sequences are blocked by a reverse complement sequence. **b**, Similar to the first ligation step, the third barcoding sequence (BC3) and a UMI is incorporated. **c**, The complete tag can be captured by streptavidin beads **d**, and amplified in the index PCR to include the Illumina adaptor sequences as well as the library index (BC4).

To test the strategy, we used Min6 wild type (WT) and inceptor (Iir) knock-out (KO) as a cell model and an anti-inceptor antibody for the targeting[230]. Upon proving the concept of antibody-SPLiT-seq, we waived the combinatorial barcoding but solely used one type of barcoded oligonucleotides per round. The Iir KO condition should not give any signal, however, in the agarose gel electrophoresis a signal was observed for both after cDNA amplification (**Figure 3.28**).

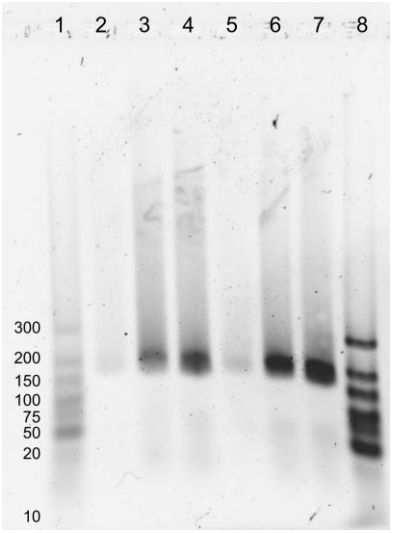

Figure 3.28 | Agarose gel electrophoresis after cDNA amplification. Following conditions were loaded: 2 µL of ultra low range (ULR, Invitrogen) DNA ladder (lane 1), 1 µL of WT cDNA (lane 2), 5 µL of WT cDNA (lane 3), 10 µL of WT cDNA (lane 4), 1 µL of Iir KO cDNA (lane 5), 5 µL of Iir KO cDNA (lane 6), 10 µL of Iir KO cDNA (lane 7), 4 µL of ULR DNA ladder (lane 8).

To investigate the identity of the two cDNA amplicons, the library construction was pursued. For sequencing, both libraries were pooled equimolar which resulted in similar amount of DNA per sample in the final pool. The base calls of the MiSeq run allowed an annotation of 82% of the obtained reads to the third-round barcode linker in the wild type condition while for the inceptor KO 73% of the reads contained this sequence (**Figure 3.29a**). In 16 to 21% of the cases only the third-round barcode linker was obtained however as soon as the second-round barcode linker was present, the third was also detected. With respect to the absence of Iir in the KO cell line, a certain degree of specificity has to be deduced. The percentage of duplicated sequences was for both experiments in the range of 60%. This is expectable for the WT condition since only the UMI in the third barcoding round is introducing heterogeneity.

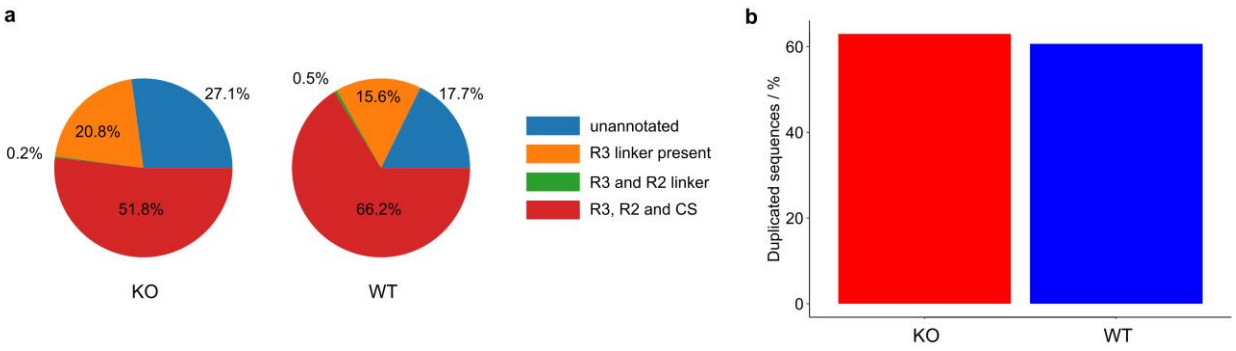

Figure 3.29 | Sequencing analysis of antibody-SPLiT-seq tag generation on Min6 WT and KO cells. a, Pie charts representing the ratio of reads from a MiSeq NGS run annotated to the three barcoding rounds. **b**, Bar plot that displays the percentage of duplicated reads from each of the condition.

The Iir antibody was applied in different approaches, as immunofluorescence staining or PLA, with high specificity and has proven its general usability[111]. As described by Wiener *et al*., labeled antibodies may drop in their specificity due to oligonucleotides of a length of 64 nt. In antibody-SPLiT-seq, the adaptor oligonucleotides conjugated to the antibody already have a length of 55 nt which might have reduced their specificity. On the one hand, the negative charge of the ssDNA may result in strong electrostatic interactions with positively charged elements in the cell, on the other hand a hybridization with cellular single stranded nucleotide sequences, even with unpaired bases or hairpins, is more probable with increase length.

An extended washing after antibody incubation with maybe a higher ion concentration in the washing buffer could help with this problem. The shortening of the adaptor oligonucleotide seems to be an alternative way to adjust unspecific binding but the key elements are immovable for the technique. The partial Read 1 sequence allows the amplification of the cDNA after reverse transcription and has to match with the corresponding primer binding site in the annealing temperature to allow an effective PCR. The antibody barcode is also mandatory in the case of using multiple protein targets. Also, the capturing sequence needs to have a certain length to give a sufficiently strong hybridization force and prevent dissociation. However, an increase in the GC ratio may allow a similar binding strength with a shorter sequence length. Here, the secondary structure formation of changed sequence composition has to be evaluated. Solely the last two bases could be left out by replacing the last deoxynucleotide with a dideoxynucleotide. Their aim is to prevent priming by the adaptor too keep the hybridization side for the round 2 barcode linker free. The purchase of oligonucleotides with modified bases increases their price tremendously and was avoided in this experiment. This is not an issue in terms of capturing mRNA transcripts since the capturing sequences contain a VN-motive as the final two nucleotides of the poly (T) tail which leads to a predominant binding of the mRNA with the last bases before and first of the poly (A) tail. The end of the longer mRNA poly (A) tail neither hybridize with the capturing oligonucleotide nor provide a priming side for the reverse transcriptase to elongate RNA sequences.

In a study for TF protein screening together with single-nuclei (sn)RNA-seq the technique intranuclear cellular indexing of transcriptomes and epitopes (inCITE-seq) was developed[231]. Thereby, antibodies were applied in a dextran sulfate solution to reduce unspecific binding which also could solve the specificity issue. The authors further mention, that their inCITE buffer performs way better then commercially available ones. Still the difference between snRNA-seq and scRNA-seq remains that the majority of mature mRNA that is located in the cytoplasm is not captured by inCITE-seq which also applies for protein detection.

In comparison to antibody-SPLiT-seq, the commercially available CITE-seq allows staining of surface proteins only. The methanol fixation on the 10x Genomics Chromium platform is possible[232,233] but paired with an increased occurrence of ambient mRNA[228]. This might be the reason why a further permeabilization is not performed. Hereby, the emission of cell mRNA might be additionally increased. Apart from that, PFA fixation shows multiple handling benefits in comparison to methanol fixation regarding the reduction of background signal during intracellular staining[234,235], keeping intracellular structures intact[236], and the fact that in some applications methanol fixation prevents a signal[237].

Fixed droplet RNA sequencing (FD-seq) also represents a scRNA-seq method that allows the staining of intracellular proteins by combining PFA fixation with the usage of droplet mRNA sequencing methods[238]. The permeabilization after PFA fixation allows lager molecules like antibodies the penetration of the cell while keeping mRNA in place better than methanol fixation which results in an increased number of read counts per cell and gene.

In comparison with literature, antibody-SPLiT-seq would have the benefit of being device independent. The strategy with PFA fixation also corresponds with other approaches described in literature that incorporate a combination of scRNA-seq with protein readout. The most promising adjustment would be the staining buffer adjustment and an extended washing after antibody staining to prevent the unspecific binding. These steps should allow the successful confraternization of transcriptome and proteome readout to improve data output regarding to low copy number mRNA, mRNA that is uncorrelated to the protein level or insights to posttranslational protein modification.

3.3. Protein proximity profiling

The capture of protein expression data combined with scRNA-seq data offers additional insights to present proteins and to the cell signaling state. This would have been beneficial for the screening low expressed transcription factors or receptors that were inferred by computational methods on the EC scRNA-seq data. One step further would be the experimental evidence of interacting proteins as an additional layer to scRNA-seq. To achieve this, I first present the implementation of the imaging-based proximity ligation assay (PLA) on fixated cells in 2D cell culture while subsequently the integration of the PLA in the scRNA-seq pipeline is experimentally verified.

The PLA was performed on mouse preadipocytes that were differentiated to adipocytes *in vitro* to study signaling changes during the emerging of insulin resistance in type 2 diabetes. Type 2 diabetes mellitus is a chronic metabolic disease that has risen to a global epidemic and affecting almost half a billion of people worldwide[239]. The main factors for this increase relate to increasing physical inactivity and the highest uptake of processed sugar in the human evolution. The resulting obesity contributes to over 50% to the development of type 2 diabetes[240]. While adipose tissue is an essential part of human metabolism, in obesity the adipose tissue faces a chronical inflammation that triggers fibrosis and adipose dysfunction[241]. A further consequence is the development of insulin resistance, a major attribute of type 2 diabetes[242]. The physical weight gain is reflected on a cellular level by adipocyte proliferation and expansion of existing adipocytes. Both processes lead to an ECM reorganization. Integrins are a group of membrane proteins that have a high significance in ECM organization by allowing ECM connection and intercellular signaling[243]. Enabling the active state of integrins, adaptor proteins of the talin and kindlin family are required[244]. The family member kindlin-2 (Kind2) plays a special role by being expressed ubiquitously and essential for activation of complexes containing integrin β1 and β3 (Itgb1,3) while the absence of Kind2 results in insulin resistance and glucose intolerance due to adipocyte apoptosis[245]. The mechanisms of a Kind2 knock-out in the mouse system was investigated by Ruiz-Ojeda *et al.*[137]. I applied the PLA targeting the interaction of insulin receptor with Itgb1 and Itgb3 integrin in the Kind2 wild type and knock-out, respectively.

3.3.1. Imaging-based proximity ligation assay

The standard PLA is designed as an imaging method working with a fluorescent probe binding to the amplified padlock probe sequence. The proximity of certain proteins indicates their interaction and the activation of related signaling pathways. The PLA is able to detect the co-localization of two proteins in a distance of up to 30 nm[246,247]. Due to these characteristics, it is suitable to infer mechanistics of potential protein-protein interactions.

This part evaluates the PLA assay as a method for measuring proximity and the interaction of surface receptors in an *in situ* model. It was applied on the differentiated mouse adipocytes to infer the interaction of Insr-Itgb1 and Insr-Itgb3 in dependency to Kind2. For this purpose, the assay was conducted on immortalized preadipocytes from subcutaneous adipose tissue of Kind2 flox mice (Kind2 fl/fl) and Kind2

flox; adiponectin-cre (Kind2$^{adipo\text{-}cre}$) mice were fed chow diet for 8 weeks. The preadipocytes were differentiated into adipocytes over a period of eight days and the differentiated adipocytes were identified by the size of lipid droplets. Notably, the differentiation effectivity of the Kind2$^{adipo\text{-}cre}$ cells was significantly reduced (data not shown). To allow a reliable analysis, areas without occurrence of big lipid droplets were excluded. In immunofluorescence staining, the distinct PLA dot signal was observed for all positive PLA combinations (**Figure 3.30a** and **b**).

Two different PLA experiments were conducted. In one the Duolink PLA kit was applied with its PLUS and MINUS secondary antibodies. To have a comparison, a customized PLA with self-labeled secondary antibodies was performed as described by Wiener *et al.*[111]. Besides enzyme mixes and secondary antibodies, the implementation was comparable. Both PLA approaches were leading to comparable output. While the interaction Insr-Itgb1 was reduced in the Kind2$^{adipo\text{-}cre}$, a higher number of dots was observed for the Insr-Itgb3 interaction for the Kind2$^{adipo\text{-}cre}$ genotype (**Figure 3.30c**). This result is represented in both approaches, even if the customized PLA showed a minimized dot count per cell. Here, the only exception is the Insr-Itgb3 interaction for the Kind2$^{adipo\text{-}cre}$ which shows a similar high average dot count per cell as the Duolink PLA.

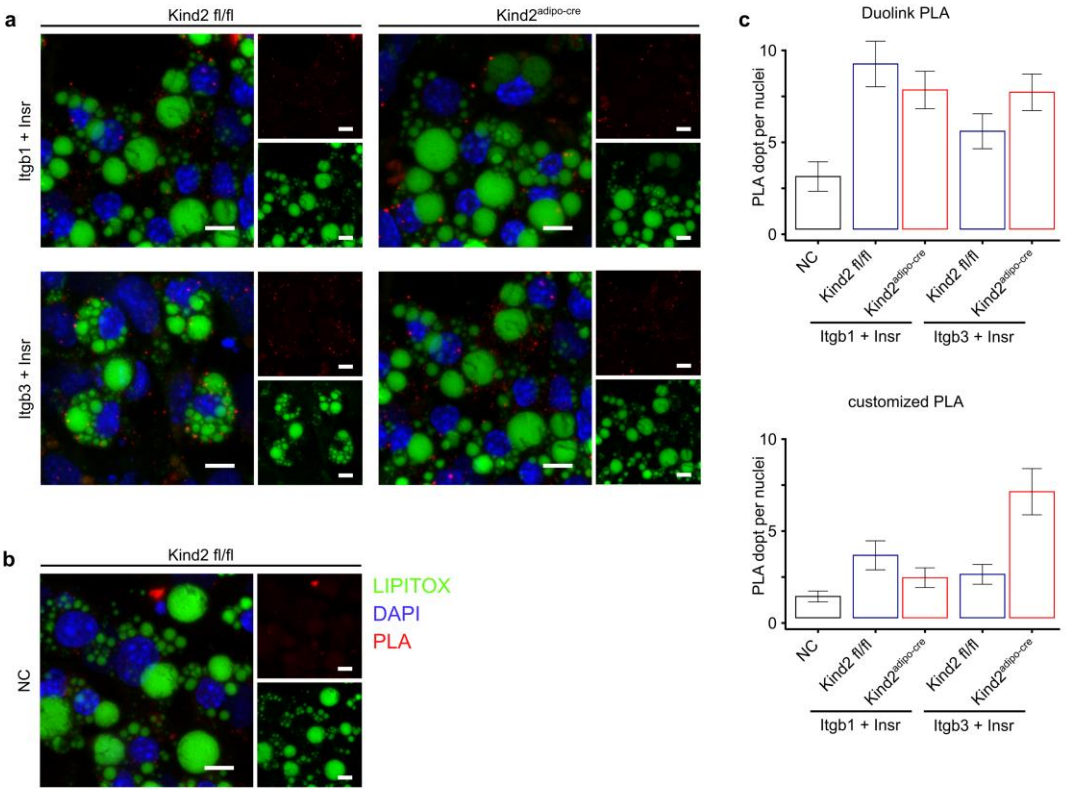

Figure 3.30 | PLA to determine the interaction of Insr with Itgb1 and Itgb3, respectively. Differentiated immortalized preadipocytes from subcutaneous adipose tissue of Kind2 flox mice (Kind2 fl/fl) and Kind2 flox; adiponectin-cre (Kind2$^{adipo\text{-}cre}$) mice were used as *in situ* model. **a**, representative immunofluorescence pictures of the Duolink PLA. PLA signal (red), DAPI nuclei staining (blue). Differentiation effectivity was determined by LIPITOX (green). Scale: 10 µm. **b**, Negative control of the Duolink PLA, without applying primary antibodies. **c**, Bar plots of PLA signal dot counted per nuclei for the respective condition. The error bars denote the standard deviation for the Duolink (upper) and customized (lower) PLA approach. A total number of 1063 cells were analyzed in the Duolink and 5156 cells in the customized PLA. (This data is also presented in the publication of Ruiz-Ojeda *et al.*[137])

The first point that has to be discussed in the beginning is the question if PLA is a suitable method to infer protein-protein interactions or if only statements about their proximity can be deduced. According to the inventers of the PLA, a distance in proximity of up to 30 nm can be measured by the assay[112]. The commercially available DuoLink PLA kit that works with primary and species specific oligonucleotide labeled secondary antibodies states a distance of 40 nm[248]. Another indirect technique for proximity as FRET gives a signal up to distances of up to 10 nm. This seems like a way more significant signal, however, the subcellular environment is defined by high dynamics in which it is highly unlikely that two molecules remain in a certain position without the presence of connecting molecular forces. Therefore, these differences in distance are not a crucial factor in that manner. The limitations of PLA significance are also discussed in the research community[249]. The most important point is that assays are accompanied by suitable controls which highlight occurrence of unspecific generation of a padlock probe paired with signal amplification. In this study, the negative controls without primary antibodies enabled this function and displayed a significantly decreased PLA count compared to the positive controls.

The Kind2$^{adipo\text{-}cre}$ mouse preadipocytes have the Cre recombinase coupled to the adiponectin gene, which is expressed when preadipocytes commit to the adipocyte differentiation path. This differentiation is phenotypically observable by the emerging and accumulation of increased size lipid droplets. Consequently, only differentiated cells, with a high lipid droplet accumulation could be examined for the PLAs. The lower differentiation efficiency suggests the involvement of Kind2 in the adipocyte differentiation. This matches the results of Pan *et al*. who ascertained that Kind2 regulates the differentiation of 3T3-L1 preadipocytes[250]. In their study, Kind2 knockdown inhibited the adhesion, migration, and adipogenic differentiation. In addition, adipocytes with Kind2 deficient suffered from insulin resistance and inflammation[137].

The comparison of the two approaches underpins the reproducibility under slightly different conditions since outcome was comparable even if the customized PLA displayed a general decreased dot count. It can be assumed that this is linked to lower enzymatic reactivity or decreased antibody specificity. A surprising observation was, however, that the Kind2$^{adipo\text{-}cre}$ condition of customized PLA was strongly increased and almost at the level of the Duolink PLA. A factor here could have been the lower differentiation efficiency of Kind2$^{adipo\text{-}cre}$ preadipocytes which was especially decreased in this condition. As a result, a minimized number of cells was analyzed in that condition which is also represented in an increased standard deviation denoted by the error bar. The signal of LIPITOX was used in a script to determine the differentiation efficiency using an imagej script. An overall low LIPITOX signal may have selected not fully differentiated adipocytes which have an increased signal readout. Here, it cannot be suggested that the Insr-Itgb1/3 dimerization is generally increased as a cause of the differentiation process or that the increased lipid droplet accumulation minimizes the PLA signal. Regarding the error bars for the other conditions, a 5-fold higher number of analyzed cells in the customized PLA could decrease them noticeably compared to the Duolink approach.

The PLA could highlight the relevance of interactions of Insr with Itgb1 and Itgb3 in adipocytes however a significant downregulation in the Kind2$^{adipo\text{-}cre}$ genotype was only observed for the Insr-Itgb1 and not for the Insr-Itgb1 interaction. The interaction of Insr-Itgb3 was even non-significantly increased for the Kind2 knock-out adipocytes. Nevertheless, the PLA could be presented as a suitable method to detect protein-protein interactions on a single-cell level. The imaging-based method was sufficient for this setup whereas a sequencable readout of protein-protein interactions that could be integrated into the scRNA-seq pipeline, would improve the significance.

3.3.2. PLA based protein-protein interaction profiling on single cells

Coupling scRNA-seq with sequencable protein-protein interaction readout of selected proteins in form of a PLA offers a direct link of the transcriptome to the signal transduction. Receptor tyrosine kinases (RTKs) create homo- and hetero-multimers in order to facilitate protein phosphorylation and signal transduction from the cell surface towards transcriptional changes. As mentioned in the previous chapter, RTKs as the insulin receptor also dimerize with various integrins. These dimerizations induce phosphorylation of the

involved receptors and following a signaling cascade inside of the cell that often leads to activation or inactivation of a certain gene expression. The ligand that binds to the receptor also affects the signal transduction and receptor dimerization characteristics. The majority of protein-protein interactions of RTKs and Integrins which are reported in literature, were investigated tediously, in multiple separate experiments from different research groups under several conditions and within various cell types. In here, a panel of 22 RTK and 14 integrins which are expressed in human pancreatic tissue, was investigated regarding reported protein-protein interaction. Therefore, the interaction data from the two database BioGrid and IntAct were collected and their interaction partners are presented in **Figure 3.31**.

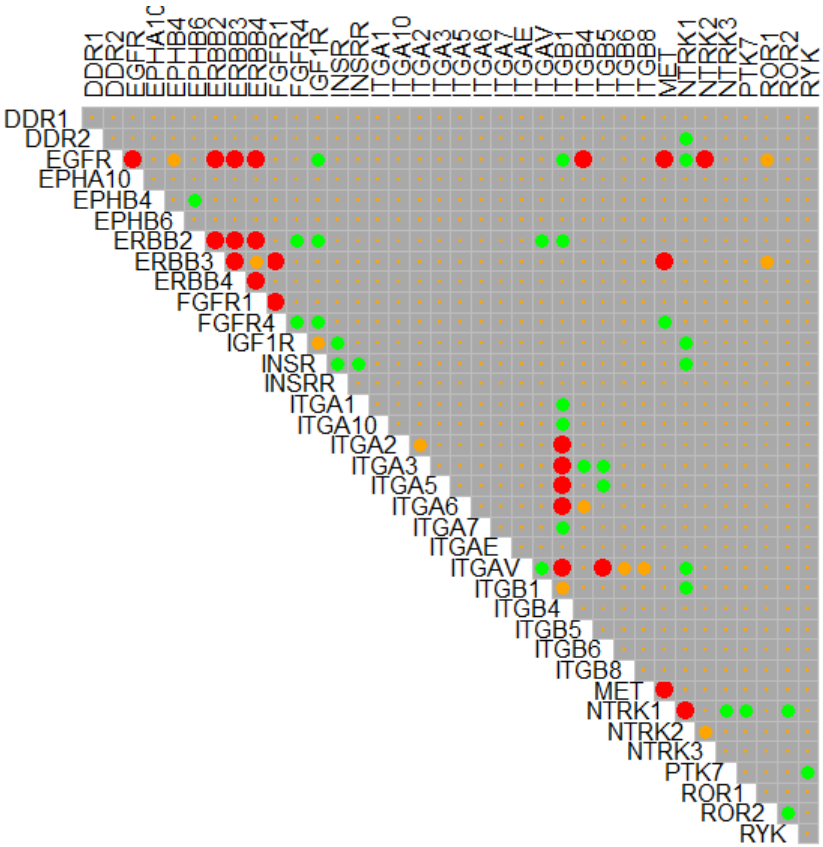

Figure 3.31 | Database analysis of literature reported interactions between RTKs and integrins. The BioGrid- and IntAct-databases were taken into account. Green dots indicate interactions reported by BioGrid-, orange ones by IntAct-database and red ones by both. 22 RTKs and 14 integrins that are expressed in the human pancreas were used as entry point. Data collection was performed in September 2019.

According to the database research from the 666 possible interactions 65 were revealed that are experimentally verified. An individual screening with the previous described imaging-based PLA would lead to a multitude of experiments and time-consuming high-resolution imaging. The combination of interaction screening with high throughput scRNA-seq allows confirmation of reported interactions and their importance in signal transduction inside different tissues and cell types. Further, screening of new possible interaction becomes feasible with high throughput screening techniques.

We have adjusted the PLA from a fluorescent microscopic readout to gain a sequencable signal, compatible with the Chromium scRNA-seq platform of 10x Genomics. Therefore, a poly (A) sequence and a MluI restriction site were integrated next to each other into the long connector (**Figure 3.32**). The hybridization of a splint and the digestion with the restriction enzyme produces oligonucleotides which we termed IntAct-records that mimic mRNAs with their poly (A) tails. Therefore, the IntAct-records are captured additionally to the mRNA transcriptome by the 10x Genomics gel beads. A further element is the IntAct-UMI, which allows the trace back to the individual detected interaction.

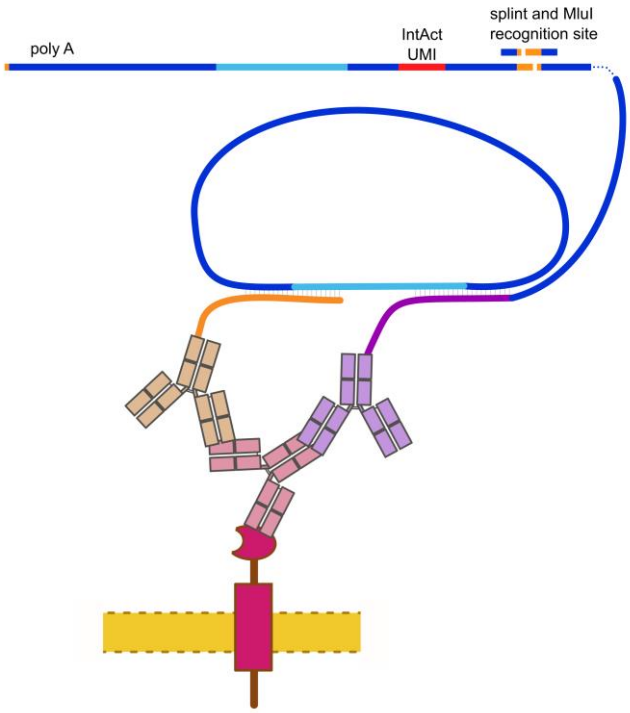

Figure 3.32 | IntActSeq PLA experimental setup and schematic of PLA amplicon digestion. Mouse Igf1r is targeted by a primary antibody from rabbit while the PLA adaptor tagged antibodies are polyclonal anti-rabbit IgG from goat. The incubation with a splint oligonucleotide generates a cleavage site for the restriction enzyme MluI. Each generated PLA amplicon fragment (IntAct-record) contains the unique molecular identifier (UMI) that refers to a specific PLA padlock probe formation and a poly (A) tail for a capturing together with cellular mRNA.

Two different approaches to generate the RCA product of the PLA were conducted. One *in vitro*, with an increased concentration of adaptor and connector sequences and one *in situ* on the pancreatic mouse cell line Min6. For the *in vitro*-PLA we used a 2-fold amount connector sequences and 5-fold amount of free adaptor oligonucleotides to facilitate random padlock probe formation. For the *in situ*-PLA, the ubiquitously expressed Igf1 receptor was used as the target. The adaptor-labeled antibodies were secondary antibodies against the primary antibody subtype. Next, not the proximity of cellular proteins was investigated but the one of polyclonal ones bound to the primary. Nevertheless, it composes a suitable concept to generate a reliable PLA signal as a proof of concept.

The harvested Min6 cells were fixated with methanol and the PLA performed on the cell suspension. The replacement of 25% of the GEM-bead solution with the restriction enzyme MluI allowed that MluI encounters its cutting sites of the PLA product on individual cells during GEM-Generation. Therefore, it is ensured that the digested single proximity sequences stay inside the droplet of their individual cell and get captured by the beads together with the cell transcriptome in the *in situ*-PLA.

After the reverse transcription, all sequence types that contain the poly (A) tail are labeled with the bead or cell barcode and contain primer binding sites for cDNA amplification which can be performed with all sequences together outside of the droplets. Upon ensuring a high DNA content after the cDNA amplification step, we amplified half of the sample for 30 cycles instead of the recommended 10 for the two conditions, respectively. The four conditions are in the following stated as *in situ*-PLA-10/30 or rather *in vitro*-PLA-10/30.

The main peak in the capillary gel electrophoresis of *in vitro*-PLA-10 is between 150 and 200 bp and matches the correctly processed fragment of 195 bp. The recurring pattern at around 300 to 500 may refer to multiple IntAct-records that were not digested by MluI (**Figure 3.33a**). In contrast to that, none of the other samples showed a peak that directly indicates the presence of the correctly formed IntAct-record. The amount of low size sequence fragments has significantly increased in the samples with 30 cycles and peaks were generally broadened up.

Upon proceeding in the library construction, the addition of two Read as well as P5 and P7 sequences (compare **Figure 2.1**) leads to a correctly processed IntAct-record with length of 297 bp. The capillary gel electrophoresis displays peaks at that length for the two *in vitro*-PLA conditions whereas the *in situ*-PLA samples show a broad accumulation of sequences between 200 and 500 bp (**Figure 3.33b**). Moreover, in both *in vitro*-PLA conditions an assemblage of peaks, separated from the IntAct-record peak, with sequence length in the range of 250 bp is visible that exhibits an even increased area under the curve.

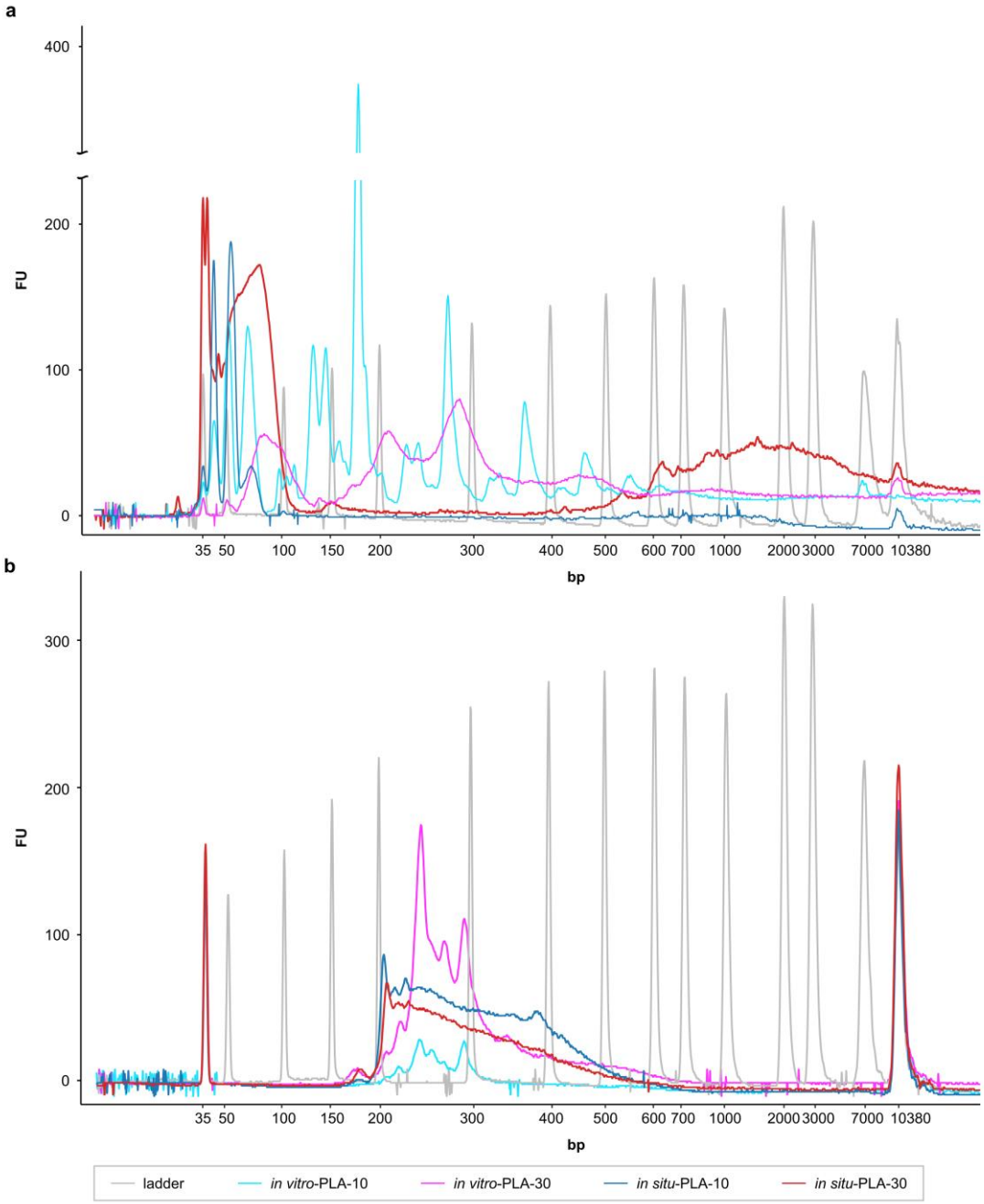

Figure 3.33 | Electropherogram of capillary gel electrophoresis of IntAct samples. a, Overlay of electropherograms from *in situ*-PLA-10/30 and *in vitro*-PLA-10/30 samples after cDNA amplification and **b**, after library construction performed with a Bioanalyzer High Sensitivity DNA Assay Kit (Agilent Technologies). The assays were performed as stated in the 10x Genomics protocol (Manual Part Number CG00052, Rev D).

The libraries were pooled equimolar and sequenced on a MiSeq Nano v2 flow cell kit. While Read 1 FastQ output files contain the 10x bead barcode and the 10x UMI, ones of Read 2 encode the mRNA transcripts and IntAct-records. The Read 2 sequences were screened for a 9 bp pattern next to the IntAct UMI to count numbers of annotated sequences as well as the ones that were unique. The *In vitro*-PLA conditions had a ratio of detected IntAct-records of around 40% while in the *in situ*-PLA conditions this ratio was with approximately 14% significantly lower (**Figure 3.34a**). The triplication of the cDNA amplification cycles had

a minor effect for lead to an increase of unmapped sequences in the *in vitro*-PLA approach of 6.5% whereas that is tending towards 0 for the *in situ*-PLA conditions. The *in situ*-PLA conditions contained transcriptomic data of the Min6 cells besides the IntAct signal. Therefore, the Cell Ranger software (10x Genomics) was used to annotate the reads to the mouse reference genome. The percentage of unique transcripts was at 1.1% for the *in situ*-PLA-10 while it had a value of 1.7% for the condition with 30 cycles (**Figure 3.34a**).

For the vast majority of the beads, only one IntAct-record was sequenced, and the number decreased exponentially (**Figure 3.34b**). Whereas one IntAct-record was found on up to 127 000 different beads, only about 7000 beads hat five IntAct-records. Summed up the IntAct-record counts from all beads from which more than five records were sequenced, it amounts to 11 000 beads. Considering the low number of IntAct-records that were captured per bead, mainly one sequence, a range of duplicated reads smaller than 1% for all conditions is not surprising (**Figure 3.34c**). Notably, the 30 cycles had a more than doubled duplication rate compared to 10 cycles of cDNA amplification. The occurrence of multiple size products which come from an insufficient MluI digestion are with 0.25% for the *in vitro*- and 0.0% for the *in situ*-PLA, surprisingly low since their peaks were clearly recognizable in the capillary gel electrophoresis after the cDNA amplification (**Figure 3.33d**). This is due to the enzymatic fragmentation step of the library construction sequences in which longer sequences have a higher probability of cleavage. In addition, the thermodynamic circumstances of polymerase drop down led to an increased amplification of shorter DNA fragments. This shifts the ratio of single to multiple IntAct-records in one DNA fragment further towards the shorter ones.

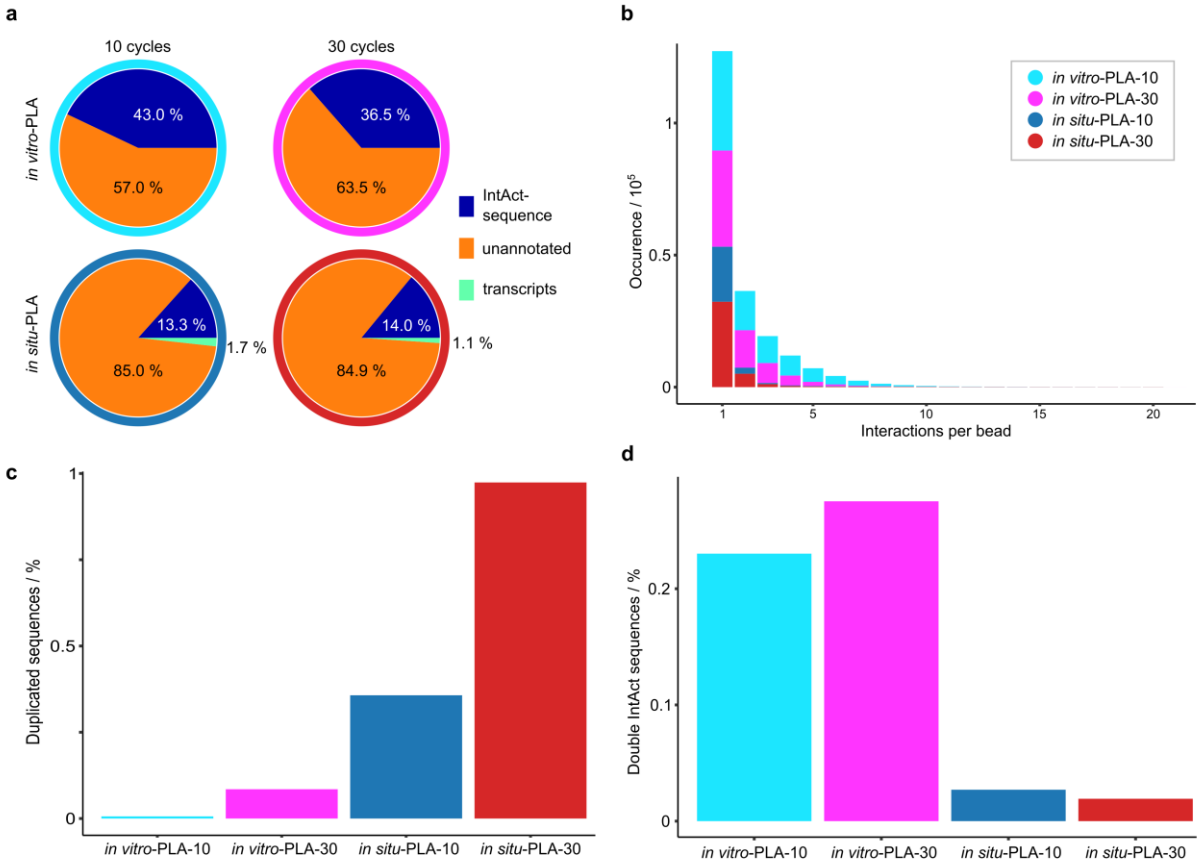

Figure 3.34 | MiSeq sequencing results for *in vitro* and *in situ* PLA with 10 and 30 cycles of cDNA amplification. a, Pie chart of sequences annotated to the IntAct-record and mRNA transcripts. **b**, Captured IntAct-records per bead. **c**, Percentage of IntAct-records that share the identical 10x bead barcode, 10x UMI, and IntAct UMI **d**, Percentage of IntAct-records that have the length of two rounds of RCA product which is caused by an insufficient MluI digestion.

For the *in situ*-PLA-10, we obtained a total count of annotated transcripts to the reference genome of 7214. This amounts to an average count per cell of 7 over a cell number of 1026. In comparison to the day nine dataset of the EC differentiation (chapter 3.1) where the average count for a single cell was around 14 000, the here captured transcripts cannot be described as a transcriptome. Furthermore, the initial filtering step of the standard scanpy processing pipeline removes cells with less than 200 expressed genes. According to that, no cell is passing the threshold and a proper analysis of the transcriptome is redundant. Therefore, we examined only the transcripts that were found the most in sequencing (**Figure 3.35**). Since Min6 is a pancreatic β-cell line, the high expression of the two mouse insulin subtype genes *Ins1* and *Ins2* can be assumed. Iapp and Malat are involved in the glucose monitoring and insulin secretion[251,252]. The peptide hormone Cck also play a role in pancreas by inducing the secretion of pancreatic enzymes into the intestine[253] while Pcsk2 contributes to processing proglucagon towards glucagon[254]. The mitochondrial genes *mt-Co3* and *mt-Atp6* refer to the subunits of the cytochrome c oxidase and ATP synthase enzyme complexes of the respiratory chain, basic cellular functions that proceed in every cell.

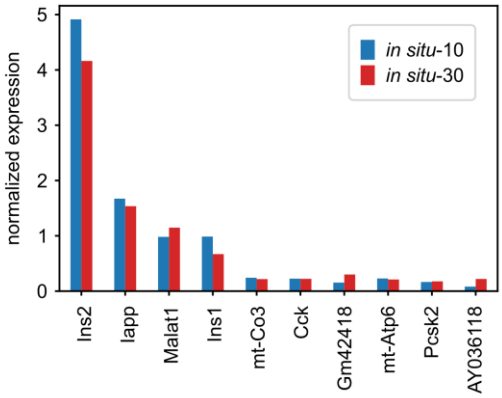

Figure 3.35 | Transcriptome analysis of the two IntAct-seq *in situ* conditions on Min6 cells. Top ten sequenced transcripts are listed by their normalized expression. The average expression is normalized over all cells in the respective condition.

Even if the fine tuning in cleanup and transcriptome capture displays points of improvement, the whole analysis shows a successful integration of IntAct-seq and common scRNA-seq on the 10x Genomics droplet platform. Consequently, screening for protein-protein interactions on the single-cell level is implementable by the PLA based concept.

The PLA-based screening for protein proximity incorporated to the scRNA-seq platform of 10x Genomics could be proven feasible. While the *in vitro* conditions showed general functionality of the pipeline, the samples of the *in situ*-PLA revealed additional transcriptomic data. Since a relatively homogenous cell line was used, a low sequencing depth applied, and the transcriptomic data amounted to 2% of the library sequences only, hence no further implications could be drawn. However, the approach reveals new possibilities for the screening of protein-protein interactions on a single-cell level.

The methanol fixation stem prior to the *in situ*-PLA has to be overseen and it has to be evaluated if a PFA fixation might be more suitable also in regard to targeting of intracellular proteins. The usage of methanol fixation in preserving cells with their genomes was published[232,233] for the 10x Genomics Chromium platform and is even included in its protocols (Manual Part Number CG000136, Rev B). However, there is an ongoing debate about how to store single cells without introducing biases[130,228]. This issue is discussed in more detail in chapter 3.2.1.

The high rate of unannotated sequences in the IntActSeq approach leads back to the SPRIselect cleanup in combination with the enzymatic fragmentation process during library construction. Investigating the area under the curve for the two *in vitro*-PLA samples in **Figure 3.33b**, it is evident that around a quarter of the sequences has the correctly processed length of 298 bp after library construction. A reduction of the short fragment number would be possible by lowering the amount of the SPRIselect reagent in the last cleanup step of the library construction. In the original protocol a double-sided cleanup is performed, aiming for a Gaussian distribution between 300 and 700 bp. We have increased the lower range SPRIselect reagent concentration from 0.8X to 1.5X to broaden the peak from 300 towards shorter fragments. Since the

SPRIselect cleanup is a range selective method a change of the reagent to 0.9 or 1.0X only instead of to 1.5X may have reduced the number of shorter sequences that represent unannotated reads after NGS. Even if we had minimized the peak of our IntAct-record at around 300 bp, the shorter fragments would have been discriminated in a more effective way and gave the library a changed size distribution towards higher fragments.

An advantage of newer 10x Genomics kits (v3.1) is that their gel beads contain capture sequences that enable a separation of additional readout signals and the transcriptome readout. This prevents an overloading of the gel bead poly T tags with IntAct-records while the transcriptomic readout is minimized. Further, by using SPRIselect reagent 0.6X in a cleanup, the longer mRNA can bind to the beads whereas shorter fragments as the lower length IntAct-records remain in the supernatant. The transfer of the supernatant to a new tube and formation of a repeated SPRIselect cleanup with 2.0X of SPRIselect reagent represents an elegant way to separate the two signals from each other and to allow an individual processing of the two libraries.

A step further is the examination of multiple interactions in one run. Firstly, this requires the more costly labeling of primary antibodies that is necessary to overcome the limitation of the finite host species and possible cross-reactivity. In the next step, two possibilities for a scale up come to mind. One is the use different sets of long and short connectors being specific to two antibody-coupled adaptors, respectively. Hereby, for each expected interaction the adaptor hybridization sides identify the specific interaction. If multiple interactions between one protein with others are the aim of an investigation, it is possible to design the different connectors. However, the downside is that if a long connector has hybridized to one adaptor that does not match the other one in proximity, no padlock probe is formed, and the signal is lost. To a certain degree this setup can be applied as long as the number of different interaction partners for single proteins is not overlapping the single interactions.

The other approach would be the design of a universal set of long and short connector while integrating additional barcode sequences in the antibody adaptors that would need an additional fill-up amplification or the addition of a reverse complementary sequence, which is integrated by ligation. However, this universal PLA could not produce specific signals (Kokotek and Wiener. unpublished data). Besides the additional amplification and ligation step that is needed, the increased length of the padlock probe could lead to the non-specificity.

By using a flow cell with increased read depth, the transcriptomic data could be further extended but that has not been a scope of this approach. The MiSeq Nano flow cell gives an output 1×10^6 reads while the NovaSeq 6000 System with the S2 flow cell, used for the EC differentiation (chapter 3.1), amounts up to 4×10^9 reads. Calculated back on the number of reads per cell, or bead in the *in vitro*-PLA, an average of 50 reads was obtained with the MiSeq Nano flow cell. The NovaSeq flow cell with 50 000 reads per cell provides a 1000-fold higher read depth.

Other approaches to infer proximity are auto-cycling proximity recording (APR)[125]. It uses antibodies tagged with hairpin probes that facilitate an amplification-based proximity signal in a cyclic manner (see **Figure 1.8**). This enables the generation of a repeated signal with the same hairpin probe or other probes in proximity and has the benefit, that not only one but multiple proximity partners are determinable. The authors used gel electrophoresis or qPCR as readout methods. The implementation into the 10x Genomics pipeline seems unachievable since the proximity records are not attached to the cell. Thus, the record generation would need to occur inside the GEM droplets side by side with the reverse transcription. Here, other polymerases would interfere with the reverse transcriptase and the other way round the reverse transcriptase may disturb the APR by adding guanosine to the 5' ends and introducing TSOs. Therefore, a successful application of APR to the 10x Genomics pipeline seems unachievable. In addition, the inclusion of adaptors for library construction is impractical due to missing capturing sides and the full-record length of around 75 bp might get lost in purification steps.

4. Conclusion

This thesis developed and applied scRNA-seq technologies and bioinformatic methods to investigate the differentiation of hiPSCs into ECs in a 3D suspension culture and the continuous conducted process of neovascularization by the derived ECs and co-evolving mural cells in a 3D Matrigel environment. Here, the commercially available scRNA-seq platform of 10x Genomics was used to compare the transcriptomes of individual cells at different time points and conditions. The progressive transcriptional monitoring over time allowed the identification of various cell types and cell stages that were assigned to different functions. The presence of the respective defining protein markers could be confirmed by immunofluorescent staining. Beside the transcriptomic readout of single cells, the possibilities of gaining an additional signal of protein expression and record protein-protein interactions were discussed and experimentally verified. These extended methods of analysis that can be included into the scRNA-seq workflow and increase the evidence of the results. The incorporation of the protein readout was conducted on the combinatorial barcoding method SPLiT-seq. The PLA was represented as an interaction detection tool and its feasibility for integration into scRNA-seq exemplified on the 10x Genomics Chromium platform.

The scRNA-seq with high read depth NGS identified numerous valuable information concerning development of ECs from hiPSCs in 3D suspension culture as well as biological mechanisms of neovascularization and maturation of microvessels in a 3D Matrigel environment. In the early stage of differentiation, a mesenchyme-angioblast population was identified that is described as a common progenitor of ECs and mural cells. It further has similarities with the cardiac mesoderm or multipotent cardiac progenitor cells that are described in embryonal development.

The usage of mRNA velocity by analyzing gene induction and repression through the ratio of unspliced versus spliced mRNA allowed the time-resolved trajectory of cells along the underlaying latent time. In the 3D suspension culture, the latent time matched with the time of sample taking and the evolving of the different cell types. The determination of DDG throughout mRNA dynamics identified cell type specific genes which are transcriptionally induced in the respective development. The use of those in TFEA inferred TFs that initiate EC and mural differentiation.

The developmental arrest of ECs in 3D suspension was detected by the absence of proliferation while aspects in missing induction or repression phases in mRNA velocity and the failure of survival of pure EC aggregates in 3D suspension culture underpinned this observation. Consequently, a matrix environment and supportive cell stimuli are required to implement a physiological cultivation of ECs *in vitro*. These are important elements to include in the design of all models that aim for endothelial proliferation and long-term survival.

Transcriptional differences between 3D suspension and 2D monolayer cultivation were carved out and these suggest that the main differences lay in the presence of an ECM for the 2D cell culture model. Both formats lead to endothelial phenotype presenting cells whereas considering a large-scale production of ECs the 3D suspension culture represents an easier possibility for a scale up since a 3D cultivation of ECs

in bioreactors with larger volumes is more efficient than the planar variant. Additionally, in the 3D suspension culture allows the development of the initial endothelial cell type without a provided matrix that is necessary in 2D and associated with not inconsiderable cost. McCracken and coworkers for example, were using a Matrigel coating that has the further problematic of batch effects and incompatibility with clinical guidelines. The transfer of the ECs into a supportive matrix environment that induces their proliferation and survival, is reversing the proliferation arrest in which the cells are in the 3D suspension culture. Therefore, this is no disadvantage of the differentiation approach.
Upon transfer to Matrigel as a 3D hydrogel environment, sprouting and coalescing EC phenotypes were characterized by scRNA-seq two days after culture condition change. This highlights that the presence of a matrix gives angioblast-like cells from 3D suspension culture new signaling input and triggers their proliferation and survival. The growth in a 3D hydrogel, allows the formation of three-dimensional microstructures generated by mECs. This initially highly motile subpopulation of mECs changes in their transcriptional profile upon prolonged cultivation in Matrigel. The result was a stable microvascular network that qualifies this 3D differentiation approach also for co-culture with other cells up until the design of vascularization models for clinical purposes.

Additionally, the co-evolving of endothelial and mural cell types was verified by using a differentiation protocol for mural cells and discovering the formation of ECs as a side population. Further investigation of cell plasticity were conducted by separating populations *via* FACS and cultivation in 2D under endothelial differentiation medium as used in 3D. It highlighted the stabilization of EC populations in pure $PECAM1^{+}/PDGFRB^{-}$ sorted condition and in the unsorted control. The $PDGFRB^{+}/PECAM1^{-}$ sorted as well as the unsorted cells displayed a loss of their mesenchymal surface marker under EC differentiation medium. However, no large-scale growth of $PECAM1^{+}$ populations was observed. From this it can be concluded that the sole medium additives that encourage the endothelial differentiation from stem cells are not sufficient for maintaining the endothelial phenotype over time.
Another example for plasticity was detected for ECs and mural cells which were cultivated in a 3D hydrogel, taken out of this environment, and reaggregated in 3D suspension culture. The scRNA-seq analysis together with cells cultivated in 3D suspension culture represented a clustering of the transcriptomes for ECs and mural cells, independent of the culturing time in the 3D hydrogel environment. On the one hand, this explains why stem cell derived ECs do not represent a specificity to a particular tissue while on the other hand it supports their plasticity and adaption of expression to changed environmental conditions and is a valuable fact for engineering applications. Furthermore, it can be assumed that stem cell derived ECs adapt to any tissue specific environment. Thus, their use in *in vitro* disease models and vascularization strategies is conceivable.
In a prolonged cultivation in Matrigel, a further EC population was identified that expressed maturation and tube formation markers, that was named tEC cluster. Indeed, imaging data confirmed the establishment of stable microvessel networks in the 3D hydrogel. Further, mural cells developed two subclusters

expressing distinct pericyte marker genes. A profiling for cell-cell communication was conducted on basis of the single-cell transcriptomic data to gain further information on a characterization of these subclusters and their interplay with the tEC population. However, neither the expression pattern nor a ligand-receptor interaction profiling between tECs and the two pericyte cluster could reveal their function. Nevertheless, the inferred cell-cell signaling towards both directions between the cell types are reported in context of angiogenesis as well as inflammation state. With the here developed 3D EC cell culture model it is now possible to grow hiPSC-derived ECs in co-culture with different organoids in order to achieve a vascularization of these as seen already in organdies of pancreatic islets.

The proof of principle for the antibody-split-seq experiment could not ensure a specific binding of the antibodies. Here, an improved washing, blocking or the use of adjusted staining solutions may help to overcome this issue. Even if the combinatorial barcoding was not implemented at that point since it would have resulted in a not feasible increase of costs whereas the transfer of the SPLiT-seq technique to an antibody-based protein readout was achieved also in this setup. In a further step, the combination of the antibody-derived protein signal with the cellular transcriptomes has to be evaluated and verified. This incorporation is of special interest if mRNA transcripts have a low copy number, have no correlation to the protein level or if antibody-detectable posttranslational modifications are the scope of the application.

The PLA was represented as an assay to confirm protein-protein interactions in an imaging-based output. The adaption of the method as an addition to transcriptome readout was successful since the proof of concept confirmed the principal feasibility of the sequencable PLA integrated in the 10x Genomics droplet scRNA-seq pipeline. Fine tuning concerning the purification steps during library construction is necessary for an effective application. Further, an increased sequencing depth is needed to perform an actual scRNA-seq analysis. The 10x Genomics Kits of the third generation, in which the gel beads include additional capturing sequences besides the poly°(A) tail, allow a separation of the IntAct-records and transcriptomic signal and prevent an overlap. For multiplexing, additional modifications will be necessary to allow parallel detection of multiple protein-protein interactions. Nevertheless, IntActSeq represents a method to confirm protein proximity on a sequencable single-cell level.

The invention of new techniques that improve scRNA-seq with additional readout is immense. The approach of combining all into one assay may lead to an overload and biased output of the individual techniques. More important will be the selection of analysis elements that can reveal most meaningful output for the individual research question. Finally, the comparison of data acquired with different methods will become fundamental and provides a stage for suitable bioinformatical tools that rely on mathematical but also on biological foundations.

5. Bibliography

1. Loe, M. J. & Edwards, W. D. A light-hearted look at a lion-hearted organ (or, a perspective from three standard deviations beyond the norm) Part 1 (of two parts). *Cardiovasc. Pathol.* **13**, 282–292 (2004).
2. Armulik, A., Genové, G. & Betsholtz, C. Pericytes: Developmental, Physiological, and Pathological Perspectives, Problems, and Promises. *Developmental Cell* vol. 21 193–215 (2011).
3. Gilbert, S. F. *Developmental biology*. (Sinauer Associates, 2010).
4. Davidson, B. & Levine, M. Evolutionary origins of the vertebrate heart: Specification of the cardiac lineage in Ciona intestinalis. *Proc. Natl. Acad. Sci.* **100**, 11469–11473 (2003).
5. Simões-Costa, M. S. *et al.* The evolutionary origin of cardiac chambers. *Dev. Biol.* **277**, 1–15 (2005).
6. Wu, S. M. *et al.* Developmental Origin of a Bipotential Myocardial and Smooth Muscle Cell Precursor in the Mammalian Heart. *Cell* **127**, 1137–1150 (2006).
7. Moretti, A. *et al.* Multipotent Embryonic Isl1+ Progenitor Cells Lead to Cardiac, Smooth Muscle, and Endothelial Cell Diversification. *Cell* **127**, 1151–1165 (2006).
8. Kattman, S. J., Huber, T. L. & Keller, G. M. M. Multipotent Flk-1+ Cardiovascular Progenitor Cells Give Rise to the Cardiomyocyte, Endothelial, and Vascular Smooth Muscle Lineages. *Dev. Cell* **11**, 723–732 (2006).
9. Noden, D. M. Embryonic origins and assembly of blood vessels. *Am. Rev. Respir. Dis.* **140**, 1097–1103 (1989).
10. Geudens, I. & Gerhardt, H. Coordinating cell behaviour during blood vessel formation. *Development* vol. 138 4569–4583 (2011).
11. Hellström, M. *et al.* Dll4 signalling through Notch1 regulates formation of tip cells during angiogenesis. *Nat. 2006 4457129* **445**, 776–780 (2007).
12. Gerhardt, H. *et al.* VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161**, 1163 (2003).
13. Chappell, J. C., Taylor, S. M., Ferrara, N. & Bautch, V. L. Local Guidance of Emerging Vessel Sprouts Requires Soluble Flt-1. *Dev. Cell* **17**, 377–386 (2009).
14. Ramsauer, M. & D'Amore, P. A. Getting Tie(2)d up in angiogenesis. *J. Clin. Invest.* **110**, 1615 (2002).
15. Betsholtz, C. Cell–cell signaling in blood vessel development and function. *EMBO Mol. Med.* **10**, (2018).
16. Stratman, A. N., Schwindt, A. E., Malotte, K. M. & Davis, G. E. Endothelial-derived PDGF-BB and HB-EGF coordinately regulate pericyte recruitment during vasculogenic tube assembly and stabilization. *Blood* **116**, 4720–4730 (2010).
17. Brudno, Y., Ennett-Shepard, A. B., Chen, R. R., Aizenberg, M. & Mooney, D. J. Enhancing Microvascular Formation And Vessel Maturation Through Temporal Control Over Multiple Pro-

Angiogenic And Pro-Maturation Factors. *Biomaterials* **34**, 9201–9209 (2013).

18. Yuan, H. T., Khankin, E. V., Karumanchi, S. A. & Parikh, S. M. Angiopoietin 2 is a partial agonist/antagonist of Tie2 signaling in the endothelium. *Mol. Cell. Biol.* **29**, 2011–2022 (2009).
19. Hammes, H. P. *et al.* Angiopoietin-2 causes pericyte dropout in the normal retina: evidence for involvement in diabetic retinopathy. *Diabetes* **53**, 1104–1110 (2004).
20. Orlidge, A. & D'Amore, P. A. Inhibition of capillary endothelial cell growth by pericytes and smooth muscle cells. *J. Cell Biol.* **105**, 1455 (1987).
21. Sato, Y. & Rifkin, D. B. Inhibition of endothelial cell movement by pericytes and smooth muscle cells: activation of a latent transforming growth factor-beta 1-like molecule by plasmin during co-culture. *J. Cell Biol.* **109**, 309 (1989).
22. Mandriota, S. J., Menoud, P. A. & Pepper, M. S. Transforming Growth Factor β1 Down-regulates Vascular Endothelial Growth Factor Receptor 2/flk-1 Expression in Vascular Endothelial Cells. *J. Biol. Chem.* **271**, 11500–11505 (1996).
23. Pepper, M. S. Transforming growth factor-beta: Vasculogenesis, angiogenesis, and vessel wall integrity. *Cytokine Growth Factor Rev.* **8**, 21–43 (1997).
24. Lee, M. J. *et al.* Sphingosine-1-phosphate as a ligand for the G protein-coupled receptor EDG-1. *Science* **279**, 1552–1555 (1998).
25. Jung, B. *et al.* Flow-regulated endothelial S1P receptor-1 signaling sustains vascular development. *Dev. Cell* **23**, 600–610 (2012).
26. Wakabayashi, T. *et al.* CD157 Marks Tissue-Resident Endothelial Stem Cells with Homeostatic and Regenerative Properties. *Cell Stem Cell* **22**, 384-397.e6 (2018).
27. McDonald, A. I. *et al.* Endothelial Regeneration of Large Vessels Is a Biphasic Process Driven by Local Cells with Distinct Proliferative Capacities. *Cell Stem Cell* **23**, 210-225.e6 (2018).
28. Gabbay, K. H. The Sorbitol Pathway and the Complications of Diabetes. *http://dx.doi.org/10.1056/NEJM197304192881609* **288**, 831–836 (1973).
29. Fowler, M. J. Microvascular and Macrovascular Complications of Diabetes. *Clin. Diabetes* **29**, 116–122 (2011).
30. Yardley, J. E. *et al.* Insulin Pump Therapy Is Associated with Less Post-Exercise Hyperglycemia than Multiple Daily Injections: An Observational Study of Physically Active Type 1 Diabetes Patients. *https://home.liebertpub.com/dia* **15**, 84–88 (2013).
31. Rickels, M. R. *et al.* Restoration of Glucose Counterregulation by Islet Transplantation in Long-standing Type 1 Diabetes. *Diabetes* **64**, 1713–1718 (2015).
32. Korsgren, O. Islet Encapsulation: Physiological Possibilities and Limitations. *Diabetes* **66**, 1748–1754 (2017).
33. Cayabyab, F., Nih, L. R. & Yoshihara, E. Advances in Pancreatic Islet Transplantation Sites for the Treatment of Diabetes. *Front. Endocrinol. (Lausanne).* **12**, (2021).
34. Aghazadeh, Y. *et al.* Microvessels support engraftment and functionality of human islets and hESC-

derived pancreatic progenitors in diabetes models. *Cell Stem Cell* **28**, 1936-1949.e8 (2021).

35. Sherwood, L. M., Parris, E. E. & Folkman, J. Tumor angiogenesis: therapeutic implications. *N. Engl. J. Med.* **285**, 1182–1186 (1971).
36. Folkman, J., Merler, E., Abernathy, C. & Williams, G. Isolation of a tumor factor responsible for angiogenesis. *J. Exp. Med.* **133**, 275–288 (1971).
37. Ferrara, N. & Henzel, W. J. Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells. *Biochem. Biophys. Res. Commun.* **161**, 851–858 (1989).
38. Hurwitz, H. *et al.* Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N. Engl. J. Med.* **350**, 2335–2342 (2004).
39. Fidler, I. J. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat. Rev. Cancer 2003 36* **3**, 453–458 (2003).
40. Mahla, R. S. Stem Cells Applications in Regenerative Medicine and Disease Therapeutics. *Int. J. Cell Biol.* **2016**, (2016).
41. Gluckman, E. *et al.* Hematopoietic reconstitution in a patient with Fanconi's anemia by means of umbilical-cord blood from an HLA-identical sibling. *N. Engl. J. Med.* **321**, 1174–1178 (1989).
42. Jensen, A. & Hamelmann, E. First autologous cell therapy of cerebral palsy caused by hypoxic-ischemic brain damage in a child after cardiac arrest-individual treatment with cord blood. *Case Rep. Transplant.* **2013**, 1–6 (2013).
43. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126**, 663–676 (2006).
44. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science (80-.).* **318**, 1917–1920 (2007).
45. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
46. Vodyanik, M. A. & Slukvin, I. I. Hematoendothelial differentiation of human embryonic stem cells. *Curr. Protoc. cell Biol.* **Chapter 23**, (2007).
47. Elcheva, I. *et al.* Direct induction of haematoendothelial programs in human pluripotent stem cells by transcriptional regulators. *Nat. Commun. 2014 51* **5**, 1–11 (2014).
48. Xu, M., He, J., Zhang, C., Xu, J. & Wang, Y. Strategies for derivation of endothelial lineages from human stem cells. *Stem Cell Res. Ther. 2019 101* **10**, 1–14 (2019).
49. Patsch, C. *et al.* Generation of vascular endothelial and smooth muscle cells from human pluripotent stem cells. *Nat. Cell Biol.* **17**, 994–1003 (2015).
50. Winnier, G., Blessing, M., Labosky, P. A. & Hogan, B. L. M. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.* **9**, 2105–2116 (1995).
51. Olmer, R. *et al.* Differentiation of Human Pluripotent Stem Cells into Functional Endothelial Cells in Scalable Suspension Culture. *Stem Cell Reports* **10**, 1657–1672 (2018).
52. McCracken, I. R. *et al.* Transcriptional dynamics of pluripotent stem cell-derived endothelial cell

differentiation revealed by single-cell RNA sequencing. *Eur. Heart J.* **41**, 1024–1036 (2020).

53. Paik, D. T. *et al.* Single-Cell RNA Sequencing Unveils Unique Transcriptomic Signatures of Organ-Specific Endothelial Cells. *Circulation* **142**, 1848–1862 (2020).
54. Adams, R. H. Molecular control of arterial–venous blood vessel identity. *J. Anat.* **202**, 105 (2003).
55. Wimmer, R. A. *et al.* Human blood vessel organoids as a model of diabetic vasculopathy. *Nature* **565**, 505–510 (2019).
56. Orlova, V. V. *et al.* Functionality of endothelial cells and pericytes from human pluripotent stem cells demonstrated in cultured vascular plexus and zebrafish xenografts. *Arterioscler. Thromb. Vasc. Biol.* **34**, 177–186 (2014).
57. Kleinman, H. K. *et al.* Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma. *Biochemistry* **21**, 6188–6193 (1982).
58. Seliktar, D. Designing cell-compatible hydrogels for biomedical applications. *Science (80-.).* **336**, 1124–1128 (2012).
59. Blache, U. & Ehrbar, M. Inspired by nature: Hydrogels as versatile tools for vascular engineering. *Adv. Wound Care* **7**, 232–246 (2018).
60. Menon, N. V., Tay, H. M., Wee, S. N., Li, K. H. H. & Hou, H. W. Micro-engineered perfusable 3D vasculatures for cardiovascular diseases. *Lab Chip* **17**, 2960–2968 (2017).
61. Mandrycky, C. J., Howard, C. C., Rayner, S. G., Shin, Y. J. & Zheng, Y. Organ-on-a-chip systems for vascular biology. *J. Mol. Cell. Cardiol.* **159**, 1–13 (2021).
62. Cochrane, A. *et al.* Advanced in vitro models of vascular biology: Human induced pluripotent stem cells and organ-on-chip technology. *Adv. Drug Deliv. Rev.* **140**, 68–77 (2019).
63. Kusuma, S. *et al.* Self-organized vascular networks from human pluripotent stem cells in a synthetic matrix. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12601–12606 (2013).
64. Natividad-Diaz, S. L. *et al.* A combined hiPSC-derived endothelial cell and in vitro microfluidic platform for assessing biomaterial-based angiogenesis. *Biomaterials* **194**, 73 (2019).
65. Hanjaya-Putra, D. *et al.* Controlled activation of morphogenesis to generate a functional human microvasculature in a synthetic matrix. *Blood* **118**, 804–815 (2011).
66. Liu, J. *et al.* Synthetic extracellular matrices with tailored adhesiveness and degradability support lumen formation during angiogenic sprouting. *Nat. Commun.* **12**, 1–12 (2021).
67. LEHMAN, I. R., BESSMAN, M. J., SIMMS, E. S. & KORNBERG, A. Enzymatic Synthesis of Deoxyribonucleic Acid: I. PREPARATION OF SUBSTRATES AND PARTIAL PURIFICATION OF AN ENZYME FROM ESCHERICHIA COLI. *J. Biol. Chem.* **233**, 163–170 (1958).
68. Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H. & Fischer, D. Use of DNA Polymerase I Primed by a Synthetic Oligonucleotide to Determine a Nucleotide Sequence in Phage f1 DNA. *Proc. Natl. Acad. Sci.* **70**, 1209–1213 (1973).
69. Wilson, R. K. How the worm was won. The C. elegans genome sequencing project. *Trends Genet.* **15**, 51–58 (1999).

70. Watson, J. D. & Cook-Deegan, R. M. Origins of the Human Genome Project. *FASEB J.* **5**, 8–11 (1991).
71. Sequencing: Illumina Technology. https://support.illumina.com/content/dam/illumina-support/courses/Sequencing_Illumina_Technology/story_html5.html?iframe. accesses: 2022-01-12.
72. Bowers, J. *et al.* Virtual Terminator nucleotides for next generation DNA sequencing. *Nat. Methods* **6**, 593 (2009).
73. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science (80-.).* **323**, 133–138 (2009).
74. Eisenstein, M. Oxford Nanopore announcement sets sequencing sector abuzz. *Nat. Biotechnol.* **30**, 295–296 (2012).
75. Porreca, G. J., Shendure, J. & Church, G. M. Polony DNA Sequencing. *Curr. Protoc. Mol. Biol.* **76**, 7.8.1-7.8.22 (2006).
76. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol. 2012 308* **30**, 777–782 (2012).
77. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc. 2013 91* **9**, 171–181 (2014).
78. Hagemann-Jensen, M. *et al.* Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nat. Biotechnol. 2020 386* **38**, 708–714 (2020).
79. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202–1214 (2015).
80. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
81. Fan, H. C., Fu, G. K. & Fodor, S. P. A. Combinatorial labeling of single cells for gene expression cytometry. *Science (80-.).* **347**, 1258367 (2015).
82. Rosenberg, A. B. *et al.* Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
83. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, (2019).
84. Van Der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
85. Mcinnes, L., Healy, J. & Melville, J. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. (2018).
86. Traag, V., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, (2018).
87. Nguyen, L. Van *et al.* Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).

88. Schulze-Kremer, S. Adding Semantics to Genome Databases: Towards an Ontology for Molecular Biology. (1997).
89. Goto, S. *et al.* Organizing and computing metabolic pathway data in terms of binary relations. *Pac. Symp. Biocomput.* 175–186 (1997).
90. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
91. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol. 2020 3812* **38**, 1408–1414 (2020).
92. Tsuyuzaki, K., Ishii, M. & Nikaido, I. Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data. *bioRxiv* 566182 (2019) doi:10.1101/566182.
93. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
94. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
95. Lareau, C. A. *et al.* Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility. *Nat. Biotechnol.* **37**, 916–924 (2019).
96. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science (80-.).* **353**, 78–82 (2016).
97. Liu, Y. *et al.* High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**, 1665-1681.e18 (2020).
98. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* **14**, 865–868 (2017).
99. Murphy, K., Travers, P. & Walport, M. *Janeway's immunobiology*. (Garland Science, 2008). doi:10.1111/j.1467-2494.1995.tb00120.x.
100. Beerli, R. R. & Rader, C. Mining human antibody repertoires. *MAbs* **2**, 365–378 (2010).
101. KOHLER, G. & MILSTEIN, C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495–497 (1975).
102. Frenzel, A., Schirrmann, T. & Hust, M. Phage display-derived human antibodies in clinical development and therapy. *MAbs* **8**, 1177–1194 (2016).
103. Jemal, A. *et al.* Global cancer statistics. *CA. Cancer J. Clin.* **61**, 69–90 (2011).
104. *PROTEIN THERAPEUTICS.* http://www.antibodysociety.org/mouse-vs-pan/ (2017).
105. Weiner, G. J. Building better monoclonal antibody-based therapeutics. *Nat. Rev. Cancer* **15**, 361–70 (2015).
106. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15**, 553–557 (1997).
107. McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. Phage antibodies: filamentous phage displaying antibody variable domains. *Nat. 1990 3486301* **348**, 552–554 (1990).

108. Coons, A. H., Creech, H. J. & Jones, R. N. Immunological Properties of an Antibody Containing a Fluorescent Group. *Exp. Biol. Med.* **47**, 200–202 (1941).
109. Vsevolod V. Rostovtsev, Luke G. Green, Valery V. Fokin, and K. B. S. A Stepwise Huisgen Cycloaddition Process: Copper(i)-Catalyzed Regioselective 'Ligation' of Azides and Terminal Alkynes. *Angew. Chemie Int. Ed.* (2002).
110. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A. & Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644–648 (2006).
111. Wiener, J., Kokotek, D., Rosowski, S., Lickert, H. & Meier, M. Preparation of single- and double-oligonucleotide antibody conjugates and their application for protein analytics. *Sci. Rep.* **10**, (2020).
112. Söderberg, O. *et al.* Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* **3**, 995–1000 (2006).
113. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).
114. Tao, X., Wang, X., Liu, B. & Liu, J. Conjugation of antibodies and aptamers on nanozymes for developing biosensors. *Biosens. Bioelectron.* **168**, (2020).
115. Jones, J. A. *et al.* Oligonucleotide conjugated antibody strategies for cyclic immunostaining. *Sci. Rep.* **11**, (2021).
116. Identification of associated proteins by coimmunoprecipitation. *Nat. Methods 2005 26* **2**, 475–476 (2005).
117. Abdiche, Y., Malashock, D., Pinkerton, A. & Pons, J. Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Anal. Biochem.* **377**, 209–217 (2008).
118. Vashist, S. K., Dixit, C. K., MacCraith, B. D. & O'Kennedy, R. Effect of antibody immobilization strategies on the analytical performance of a surface plasmon resonance-based immunoassay. *Analyst* **136**, 4431–4436 (2011).
119. Förster, T. *Delocalized excitation and excitation transfer.* (Florida State University, 1965).
120. Stryer, L. Fluorescence Energy Transfer as a Spectroscopic Ruler. *http://dx.doi.org/10.1146/annurev.bi.47.070178.004131* **47**, 819–846 (2003).
121. Algar, W. R., Hildebrandt, N., Vogel, S. S. & Medintz, I. L. FRET as a biomolecular research tool - understanding its potential while avoiding pitfalls. *Nat. Methods* **16**, 815–829 (2019).
122. Wu, X. *et al.* In situ characterization of the mTORC1 during adipogenesis of human adult stem cells on chip. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4143-50 (2016).
123. Assarsson, E. *et al.* Homogenous 96-Plex PEA Immunoassay Exhibiting High Sensitivity, Specificity, and Excellent Scalability. *PLoS One* **9**, e95192 (2014).
124. Zhong, W. *et al.* Next generation plasma proteome profiling to monitor health and disease. *Nat. Commun. 2021 121* **12**, 1–12 (2021).
125. Schaus, T. E., Woo, S., Xuan, F., Chen, X. & Yin, P. A DNA nanoscope via auto-cycling proximity

recording. *Nat. Commun.* (2017) doi:10.1038/s41467-017-00542-3.

126. Rosowski, S. *et al.* Single-cell characterization of neovascularization using hiPSC-derived endothelial cells in a 3D microenvironment. *bioRxiv* 2022.02.15.480506 (2022) doi:10.1101/2022.02.15.480506.
127. Wang, X. *et al.* Genome-wide analysis of PDX1 target genes in human pancreatic progenitors. *Mol. Metab.* **9**, 57–68 (2018).
128. Wiedenmann, S. *et al.* Single-cell-resolved differentiation of human induced pluripotent stem cells into pancreatic duct-like organoids on a microwell chip. *Nat. Biomed. Eng.* **5**, 897–913 (2021).
129. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods 2012 97* **9**, 676–682 (2012).
130. Guillaumet-Adkins, A. *et al.* Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol.* **18**, 45 (2017).
131. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
132. Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020).
133. Keenan, A. B. *et al.* ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Res.* **47**, W212–W224 (2019).
134. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
135. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
136. Brody, J. R. & Kern, S. E. Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. *Biotechniques* **36**, 214–216 (2004).
137. Ruiz-Ojeda, F. J. *et al.* Active integrins regulate white adipose tissue insulin sensitivity and brown fat thermogenesis. *Mol. Metab.* **45**, (2021).
138. Ansarullah *et al.* Inceptor counteracts insulin signalling in β-cells to control glycaemia. *Nat. 2021 5907845* **590**, 326–331 (2021).
139. Atkins, G. B., Jain, M. K. & Hamik, A. ENDOTHELIAL DIFFERENTIATION: MOLECULAR MECHANISMS OF SPECIFICATION AND HETEROGENEITY. *Arterioscler. Thromb. Vasc. Biol.* **31**, 1476 (2011).
140. Zhao, M. T. *et al.* Molecular and functional resemblance of differentiated cells derived from isogenic human iPSCs and SCNT-derived ESCs. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11111–E11120 (2017).
141. Chavkin, N. W. & Hirschi, K. K. Single Cell Analysis in Vascular Biology. *Front. Cardiovasc. Med.* **7**, 42 (2020).
142. Liu, Z. *et al.* Single-Cell RNA Sequencing Reveals Endothelial Cell Transcriptome Heterogeneity

under Homeostatic Laminar Flow. *Arterioscler. Thromb. Vasc. Biol.* **41**, 2575–2584 (2021).

143. Vodyanik, M. A. *et al.* A Mesoderm-Derived Precursor for Mesenchymal Stem and Endothelial Cells. *Cell Stem Cell* **7**, 718–729 (2010).
144. Bondue, A. & Blanpain, C. Mesp1: a key regulator of cardiovascular lineage commitment. *Circ. Res.* **107**, 1414–1427 (2010).
145. Lazic, S. & Scott, I. C. Mef2cb regulates late myocardial cell addition from a second heart field-like population of progenitors in zebrafish. *Dev. Biol.* **354**, 123–133 (2011).
146. Rajala, K., Pekkanen-Mattila, M. & Aalto-Setälä, K. Cardiac Differentiation of Pluripotent Stem Cells. *Stem Cells Int.* **2011**, (2011).
147. Zhang, J. *et al.* Functional characterization of human pluripotent stem cell-derived arterial endothelial cells. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E6072–E6078 (2017).
148. Paul, J. D. *et al.* SLIT3-ROBO4 Activation Promotes Vascular Network Formation in Human Engineered Tissue and Angiogenesis in Vivo. *J. Mol. Cell. Cardiol.* **64**, 124–131 (2013).
149. Reuten, R. *et al.* Structural decoding of netrin-4 reveals a regulatory function towards mature basement membranes. *Nat. Commun. 2016 71* **7**, 1–17 (2016).
150. Eichmann, A. & Simons, M. VEGF signaling inside vascular endothelial cells and beyond. *Curr. Opin. Cell Biol.* **24**, 188–193 (2012).
151. Puri, M. C., Partanen, J., Rossant, J. & Bernstein, A. Interaction of the TEK and TIE receptor tyrosine kinases during cardiovascular development. *Development* **126**, 4569–4580 (1999).
152. Ohnuki, H. *et al.* BAZF, a novel component of cullin3-based E3 ligase complex, mediates VEGFR and Notch cross-signaling in angiogenesis. *Blood* **119**, 2688–2698 (2012).
153. Pham, V. N. *et al.* Combinatorial function of ETS transcription factors in the developing vasculature. *Dev. Biol.* **303**, 772–783 (2007).
154. Kim, K. S. *et al.* ELK3 expressed in lymphatic endothelial cells promotes breast cancer progression and metastasis through exosomal miRNAs. *Sci. Reports 2019 91* **9**, 1–10 (2019).
155. Birdsey, G. M. *et al.* Transcription factor Erg regulates angiogenesis and endothelial apoptosis through VE-cadherin. *Blood* **111**, 3498–3506 (2008).
156. Osterwalder, M. *et al.* HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. *Dev. Cell* **31**, 345–357 (2014).
157. Wu, S.-P., Dong, X.-R., Regan, J. N., Su, C. & Majesky, M. W. Tbx18 Regulates Development of the Epicardium and Coronary Vessels. *Dev. Biol.* **383**, 307 (2013).
158. Slukvin, I. I. & Kumar, A. The Mesenchymoangioblast, Mesodermal Precursor for Mesenchymal and Endothelial Cells. *Cell. Mol. Life Sci.* **75**, 3507 (2018).
159. Anton, R., Kühl, M. & Pandur, P. A molecular signature for the “master” heart cell. *BioEssays* **29**, 422–426 (2007).
160. Martowicz, A. *et al.* Endothelial β-Catenin Signaling Supports Postnatal Brain and Retinal Angiogenesis by Promoting Sprouting, Tip Cell Formation, and VEGFR (Vascular Endothelial

Growth Factor Receptor) 2 Expression. *Arterioscler. Thromb. Vasc. Biol.* **39**, 2273–2288 (2019).

161. Altschuler, D. L. & Ribeiro-Neto, F. Mitogenic and oncogenic properties of the small G protein Rap1b. *Proc. Natl. Acad. Sci.* **95**, 7475–7479 (1998).
162. Griffin, J. N. *et al.* RAPGEF5 regulates nuclear translocation of β-catenin. *Dev. Cell* **44**, 248 (2018).
163. Gingras, A. R., Puzon-McLaughlin, W., Bobkov, A. A. & Ginsberg, M. H. Structural Basis of Dimeric Rasip1 RA Domain Recognition of the Ras Subfamily of GTP-Binding Proteins. *Structure* **24**, 2152–2162 (2016).
164. Vij, N., Roberts, L., Joyce, S. & Chakravarti, S. Lumican suppresses cell proliferation and aids Fas–Fas ligand mediated apoptosis: implications in the cornea. *Exp. Eye Res.* **78**, 957–971 (2004).
165. Fang, J. S. *et al.* Shear-induced Notch-Cx37-p27 axis arrests endothelial cell cycle to enable arterial specification. *Nat. Commun. 2017 81* **8**, 1–14 (2017).
166. Pin, A. L. *et al.* Annexin-1-mediated Endothelial Cell Migration and Angiogenesis Are Regulated by Vascular Endothelial Growth Factor (VEGF)-induced Inhibition of miR-196a Expression. *J. Biol. Chem.* **287**, 30541 (2012).
167. He, Y. *et al.* The Cdc42/Rac1 regulator CdGAP is a novel E-cadherin transcriptional co-repressor with Zeb2 in breast cancer. *Oncogene 2017 3624* **36**, 3490–3503 (2017).
168. Favier, B. *et al.* Neuropilin-2 interacts with VEGFR-2 and VEGFR-3 and promotes human endothelial cell survival and migration. *Blood* **108**, 1243–1250 (2006).
169. Lobov, I. B. *et al.* Delta-like ligand 4 (Dll4) is induced by VEGF as a negative regulator of angiogenic sprouting. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3219 (2007).
170. Suchting, S. *et al.* The Notch ligand Delta-like 4 negatively regulates endothelial tip cell formation and vessel branching. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3225 (2007).
171. Zarrinpashneh, E. *et al.* Ablation of SGK1 Impairs Endothelial Cell Migration and Tube Formation Leading to Decreased Neo-Angiogenesis Following Myocardial Infarction. *PLoS One* **8**, e80268 (2013).
172. Hong, F. *et al.* mTOR-Raptor Binds and Activates SGK1 to Regulate p27 Phosphorylation. *Mol. Cell* **30**, 701–711 (2008).
173. Cano, A. *et al.* The transcription factor Snail controls epithelial–mesenchymal transitions by repressing E-cadherin expression. *Nat. Cell Biol. 2000 22* **2**, 76–83 (2000).
174. Das, A. *et al.* Disruption of an SP2/KLF6 repression complex by SHP is required for farnesoid X receptor-induced endothelial cell migration. *J. Biol. Chem.* **281**, 39105–39113 (2006).
175. Rohlenova, K. *et al.* Single-Cell RNA Sequencing Maps Endothelial Metabolic Plasticity in Pathological Angiogenesis. *Cell Metab.* **31**, 862-877.e14 (2020).
176. Catela, C., Kratsios, P., Hede, M., Lang, F. & Rosenthal, N. Serum and glucocorticoid-inducible kinase 1 (SGK1) is necessary for vascular remodeling during angiogenesis. *Dev. Dyn.* **239**, 2149–2160 (2010).
177. Tsuji-Tamura, K. & Ogawa, M. Inhibition of the PI3K-Akt and mTORC1 signaling pathways promotes

the elongation of vascular endothelial cells. *J. Cell Sci.* **129**, 1165–1178 (2016).

178. Zhu, P. *et al.* C8orf4 negatively regulates self-renewal of liver cancer stem cells via suppression of NOTCH2 signalling. *Nat. Commun.* **6**, 1–13 (2015).
179. Utoguchi, N. *et al.* Ascorbic acid stimulates barrier function of cultured endothelial cell monolayer. *J. Cell. Physiol.* **163**, 393–399 (1995).
180. Yoshikawa, K., Takahashi, S., Imamura, Y., Sado, Y. & Hayashi, T. Secretion of non-helical collagenous polypeptides of alpha1(IV) and alpha2(IV) chains upon depletion of ascorbate by cultured human cells. *J. Biochem.* **129**, 929–936 (2001).
181. Epting, D. *et al.* The Rac1 regulator ELMO1 controls vascular morphogenesis in zebrafish. *Circ. Res.* **107**, 45–55 (2010).
182. Xu, K. & Cleaver, O. Tubulogenesis during blood vessel formation. *Semin. Cell Dev. Biol.* **22**, 993 (2011).
183. Abraham, S. *et al.* A Rac/Cdc42 exchange factor complex promotes formation of lateral filopodia and blood vessel lumen morphogenesis. *Nat. Commun.* **6**, 1–14 (2015).
184. Cardinale, G. J. & Udenfriend, S. Prolyl hydroxylase. *Adv. Enzymol. Relat. Areas Mol. Biol.* **41**, 245–300 (1974).
185. Bansal, M., Ramakrishnan, C. & Ramachandran, G. N. A Triple-Helical Model for (Gly-Pro-Hyp) n with cis Peptide Units. **14**, 2457–2466 (1975).
186. Senger, D. R. *et al.* The alpha(1)beta(1) and alpha(2)beta(1) integrins provide critical support for vascular endothelial growth factor signaling, endothelial cell migration, and tumor angiogenesis. *Am. J. Pathol.* **160**, 195–204 (2002).
187. Ubil, E. *et al.* Mesenchymal–endothelial transition contributes to cardiac neovascularization. *Nat. 2014 5147524* **514**, 585–590 (2014).
188. Chojkier, M., Houglum, K., Solis-Herruzo, J. & Brenner, D. A. Stimulation of Collagen Gene Expression by Ascorbic Acid in Cultured Human Fibroblasts: A role for lipid peroxidation? *J. Biol. Chem.* **264**, 16957–16962 (1989).
189. Kishimoto, Y. *et al.* Ascorbic acid enhances the expression of type 1 and type 4 collagen and SVCT2 in cultured human skin fibroblasts. *Biochem. Biophys. Res. Commun.* **430**, 579–584 (2013).
190. Lin, F., Wang, N. & Zhang, T.-C. The role of endothelial-mesenchymal transition in development and pathological process. *IUBMB Life* **64**, 717–723 (2012).
191. Zeisberg, E. M., Potenta, S., Xie, L., Zeisberg, M. & Kalluri, R. Discovery of endothelial to mesenchymal transition as a source for carcinoma-associated fibroblasts. *Cancer Res.* **67**, 10123–10128 (2007).
192. Patel, J., Baz, B., Wong, H. Y., Lee, J. S. & Khosrotehrani, K. Accelerated Endothelial to Mesenchymal Transition Increased Fibrosis via Deleting Notch Signaling in Wound Vasculature. *J. Invest. Dermatol.* **138**, 1166–1175 (2018).
193. Braunwald, E. *et al. Harrison's principles of internal medicine*. (McGraw Hill, 2001).

194. Dong, W. *et al.* Mesenchymal-endothelial transition-derived cells as a potential new regulatory target for cardiac hypertrophy. *Sci. Reports 2020 101* **10**, 1–14 (2020).
195. Muhl, L. *et al.* Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat. Commun. 2020 111* **11**, 1–18 (2020).
196. Lindahl, P., Johansson, B. R., Levéen, P. & Betsholtz, C. Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science (80-.).* **277**, 242–245 (1997).
197. Sweeney, M. & Foldes, G. It Takes Two: Endothelial-Perivascular Cell Cross-Talk in Vascular Development and Disease. *Front. Cardiovasc. Med.* **5**, 154 (2018).
198. Dave, J. M., Mirabella, T., Weatherbee, S. D. & Greif, D. M. Pericyte ALK5/TIMP3 Axis Contributes to Endothelial Morphogenesis in the Developing Brain. *Dev. Cell* **44**, 665-678.e6 (2018).
199. Krispin, S. *et al.* Growth Differentiation Factor 6 Promotes Vascular Stability by Restraining Vascular Endothelial Growth Factor Signaling. *Arterioscler. Thromb. Vasc. Biol.* **38**, 353–362 (2018).
200. SALCEDO, R. & OPPENHEIM, J. J. Role of Chemokines in Angiogenesis: CXCL12/SDF-1 and CXCR4 Interaction, a Key Regulator of Endothelial Cell Responses. *Microcirculation* **10**, 359–370 (2010).
201. Kerkar, S. *et al.* TNF-alpha and IL-1beta increase pericyte/endothelial cell co-culture permeability. *J. Surg. Res.* **132**, 40–45 (2006).
202. Voronov, E. *et al.* IL-1 is required for tumor invasiveness and angiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2645–2650 (2003).
203. Estess, P., Nandi, A., Mohamadzadeh, M. & Siegelman, M. H. Interleukin 15 induces endothelial hyaluronan expression in vitro and promotes activated T cell extravasation through a CD44-dependent pathway in vivo. *J. Exp. Med.* **190**, 9–19 (1999).
204. Singleton, P. A. Hyaluronan Regulation of Endothelial Barrier Function in Cancer. *Adv. Cancer Res.* **123**, 191 (2014).
205. Park, L. S. *et al.* Cloning of the Murine Thymic Stromal Lymphopoietin (Tslp) Receptor: Formation of a Functional Heteromeric Complex Requires Interleukin 7 Receptor. *J. Exp. Med.* **192**, 659 (2000).
206. Xie, F. *et al.* Cervical Carcinoma Cells Stimulate the Angiogenesis through TSLP Promoting Growth and Activation of Vascular Endothelial Cells. *Am. J. Reprod. Immunol.* **70**, 69–79 (2013).
207. Hosaka, K. *et al.* Dual roles of endothelial FGF-2-FGFR1-PDGF-BB and perivascular FGF-2-FGFR2-PDGFRβ signaling pathways in tumor vascular remodeling. *Cell Discov.* **4**, 3 (2018).
208. Krempel, P. G. *et al.* Bevacizumab Reduces Neurocan Content and Gene Expression in Newborn Rat Retina In Vitro. *Invest. Ophthalmol. Vis. Sci.* **55**, 5109–5115 (2014).
209. Kaneda, H. *et al.* Activin A inhibits vascular endothelial cell growth and suppresses tumour angiogenesis in gastric cancer. *Br. J. Cancer 2011 1058* **105**, 1210–1217 (2011).
210. Lee, H. W. *et al.* Alk2/ACVR1 and Alk3/BMPR1A Provide Essential Function for Bone Morphogenetic Protein Induced Retinal Angiogenesis. *Arterioscler. Thromb. Vasc. Biol.* **37**, 657

(2017).

211. Roman, B. L. *et al.* Disruption of acvrl1 increases endothelial cell number in zebrafish cranial vessels. *Development* **129**, 3009–3019 (2002).
212. Jurisic, G. *et al.* An unexpected role of semaphorin3A/neuropilin-1 signaling in lymphatic vessel maturation and valve formation. *Circ. Res.* **111**, 426 (2012).
213. Hou, S. T. *et al.* Semaphorin3A elevates vascular permeability and contributes to cerebral ischemia-induced brain damage. *Sci. Reports 2015 51* **5**, 1–14 (2015).
214. Lim, W. W. *et al.* Interleukin-11 is important for vascular smooth muscle phenotypic switching and aortic inflammation, fibrosis and remodeling in mouse models. *Sci. Reports 2020 101* **10**, 1–18 (2020).
215. Chen, C. W. *et al.* Human Pericytes for Ischemic Heart Repair. *Stem Cells* **31**, 305 (2013).
216. Wu, H. xuan *et al.* LIFR promotes tumor angiogenesis by up-regulating IL-8 levels in colorectal cancer. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1864**, 2769–2784 (2018).
217. Wang, Z., Lv, J. & Zhang, T. Combination of IL-24 and cisplatin inhibits angiogenesis and lymphangiogenesis of cervical cancer xenografts in a nude mouse model by inhibiting VEGF, VEGF-C and PDGF-B. *Oncol. Rep.* **33**, 2468–2476 (2015).
218. Stebbins, M. J. *et al.* Activation of RARα, RARγ, or RXRα increases barrier tightness in human induced pluripotent stem cell-derived brain endothelial cells. *Biotechnol. J.* **13**, (2018).
219. Seo, H. R. *et al.* Intrinsic FGF2 and FGF5 promotes angiogenesis of human aortic endothelial cells in 3D microfluidic angiogenesis system. *Sci. Reports 2016 61* **6**, 1–11 (2016).
220. Allerstorfer, S. *et al.* FGF5 as an oncogenic factor in human glioblastoma multiforme: autocrine and paracrine activities. *Oncogene 2008 2730* **27**, 4180–4190 (2008).
221. Zhao, H., Yao, P., Li, L. & Chen, L. Apelin receptor signaling: a novel mechanism of endothelial cell polarization. *Acta Biochim. Biophys. Sin. (Shanghai).* **48**, 1138–1139 (2016).
222. Fachinger, G., Deutsch, U. & Risau, W. Functional interaction of vascular endothelial-protein-tyrosine phosphatase with the angiopoietin receptor Tie-2. *Oncogene* **18**, 5948–5953 (1999).
223. Nawroth, R. *et al.* VE-PTP and VE-cadherin ectodomains interact to facilitate regulation of phosphorylation and cell contacts. *EMBO J.* **21**, 4885 (2002).
224. Salcedo, R. *et al.* Vascular endothelial growth factor and basic fibroblast growth factor induce expression of CXCR4 on human endothelial cells. In vivo neovascularization induced by stromal-derived factor-1α. *Am. J. Pathol.* **154**, 1125–1135 (1999).
225. Cho, J. G., Lee, A., Chang, W., Lee, M. S. & Kim, J. Endothelial to mesenchymal transition represents a key link in the interaction between inflammation and endothelial dysfunction. *Frontiers in Immunology* vol. 9 294 (2018).
226. Cheifetz, S. *et al.* Endoglin is a component of the transforming growth factor-beta receptor system in human endothelial cells. *J. Biol. Chem.* **267**, 19027–19030 (1992).
227. Van Meeteren, L. A. & Ten Dijke, P. Regulation of endothelial cell plasticity by TGF-β. *Cell and*

Tissue Research vol. 347 177–186 (2012).

228. Denisenko, E. *et al.* Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* **21**, 1–25 (2020).
229. Hu, H. *et al.* CITEMOXMBD: A flexible single-cell multimodal omics analysis framework to reveal the heterogeneity of immune cells. *RNA Biol.* **19**, 290 (2022).
230. Ansarullah *et al.* Author Correction: Inceptor counteracts insulin signalling in β-cells to control glycaemia. *Nat. 2021 5927852* **592**, E1–E1 (2021).
231. Chung, H. *et al.* Joint single-cell measurements of nuclear proteins and RNA in vivo. *Nat. Methods 2021 1810* **18**, 1204–1212 (2021).
232. Alles, J. *et al.* Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.* **15**, 1–14 (2017).
233. Chen, J. *et al.* PBMC fixation and processing for Chromium single-cell RNA sequencing. *J. Transl. Med.* **16**, 1–11 (2018).
234. Krutzik, P. O. & Nolan, G. P. Intracellular phospho-protein staining techniques for flow cytometry: monitoring single cell signaling events. *Cytometry. A* **55**, 61–70 (2003).
235. Pollice, A. A. *et al.* Sequential paraformaldehyde and methanol fixation for simultaneous flow cytometric analysis of DNA, cell surface proteins, and intracellular proteins. *Cytometry* **13**, 432–444 (1992).
236. Hoetelmans, R. W. M. *et al.* Effects of acetone, methanol, or paraformaldehyde on cellular structure, visualized by reflection contrast microscopy and transmission and scanning electron microscopy. *Appl. Immunohistochem. Mol. Morphol. AIMM* **9**, 346–351 (2001).
237. Jung, T., Schauer, U., Heusser, C., Neumann, C. & Rieger, C. Detection of intracellular cytokines by flow cytometry. *J. Immunol. Methods* **159**, 197–207 (1993).
238. Phan, H. Van *et al.* High-throughput RNA sequencing of paraformaldehyde-fixed single cells. *Nat. Commun. 2021 121* **12**, 1–11 (2021).
239. Dianna J Magliano (Co-chair), Edward J Boyko (Co-chair), B. *et al. IDF Diabetes Atlas, 10th edition*. (2021).
240. Chetoui, A. *et al.* Prevalence of overweight/obesity and its associated factors among a sample of moroccan type 2 diabetes patients. *Afr. Health Sci.* **21**, 23–31 (2021).
241. Sun, K., Tordjman, J., Clément, K. & Scherer, P. E. Fibrosis and Adipose Tissue Dysfunction. *Cell Metab.* **18**, 470–477 (2013).
242. Boucher, J., Kleinridders, A. & Ronald Kahn, C. Insulin Receptor Signaling in Normal and Insulin-Resistant States. *Cold Spring Harb. Perspect. Biol.* **6**, a009191 (2014).
243. Takada, Y., Ye, X. & Simon, S. The integrins. *Genome Biol.* **8**, 1–9 (2007).
244. Calderwood, D. A., Campbell, I. D. & Critchley, D. R. Talins and kindlins: partners in integrin-mediated adhesion. *Nat. Rev. Mol. Cell Biol. 2013 148* **14**, 503–517 (2013).
245. Gao, H. *et al.* Lipoatrophy and metabolic disturbance in mice with adipose-specific deletion of

kindlin-2. *JCI Insight* **4**, (2019).

246. Weibrecht, I. *et al.* Proximity ligation assays: a recent addition to the proteomics toolbox. *http://dx.doi.org/10.1586/epr.10.10* **7**, 401–409 (2014).
247. Borroto-Escuela, D. O. *et al.* In Situ Proximity Ligation Assay to Study and Understand the Distribution and Balance of GPCR Homo- and Heteroreceptor Complexes in the Brain. *Neuromethods* **110**, 109–124 (2016).
248. Alam, M. S. Proximity Ligation Assay (PLA). *Curr. Protoc. Immunol.* **123**, e58 (2018).
249. Alsemarz, A., Lasko, P. & Fagotto, F. Limited significance of the in situ proximity ligation assay. *bioRxiv* 411355 (2018) doi:10.1101/411355.
250. Pan, Y. *et al.* Kindlin-2 regulates the differentiation of 3T3-L1 preadipocytes: implications for wound healing. *Ann. Transl. Med.* **9**, 348–348 (2021).
251. Macfarlane, W. M. *et al.* Glucose Regulates Islet Amyloid Polypeptide Gene Transcription in a PDX1- and Calcium-dependent Manner. *J. Biol. Chem.* **275**, 15330–15335 (2000).
252. Ding, H. *et al.* LncRNA MALAT1 induces the dysfunction of β cells via reducing the histone acetylation of the PDX-1 promoter in type 1 diabetes. *Exp. Mol. Pathol.* **114**, (2020).
253. Harper, A. A. & Raper, H. S. Pancreozymin, a stimulant of the secretion of pancreatic enzymes in extracts of the small intestine. *J. Physiol.* **102**, 115–125 (1943).
254. Rouillé, Y., Bianchi, M., Irminger, J. C. & Halban, P. A. Role of the prohormone convertase PC2 in the processing of proglucagon to glucagon. *FEBS Lett.* **413**, 119–123 (1997).

6. Appendix

6.1. List of tables

6.2. List of figures