UiT The Arctic University of Norway

Faculty of Health Sciences
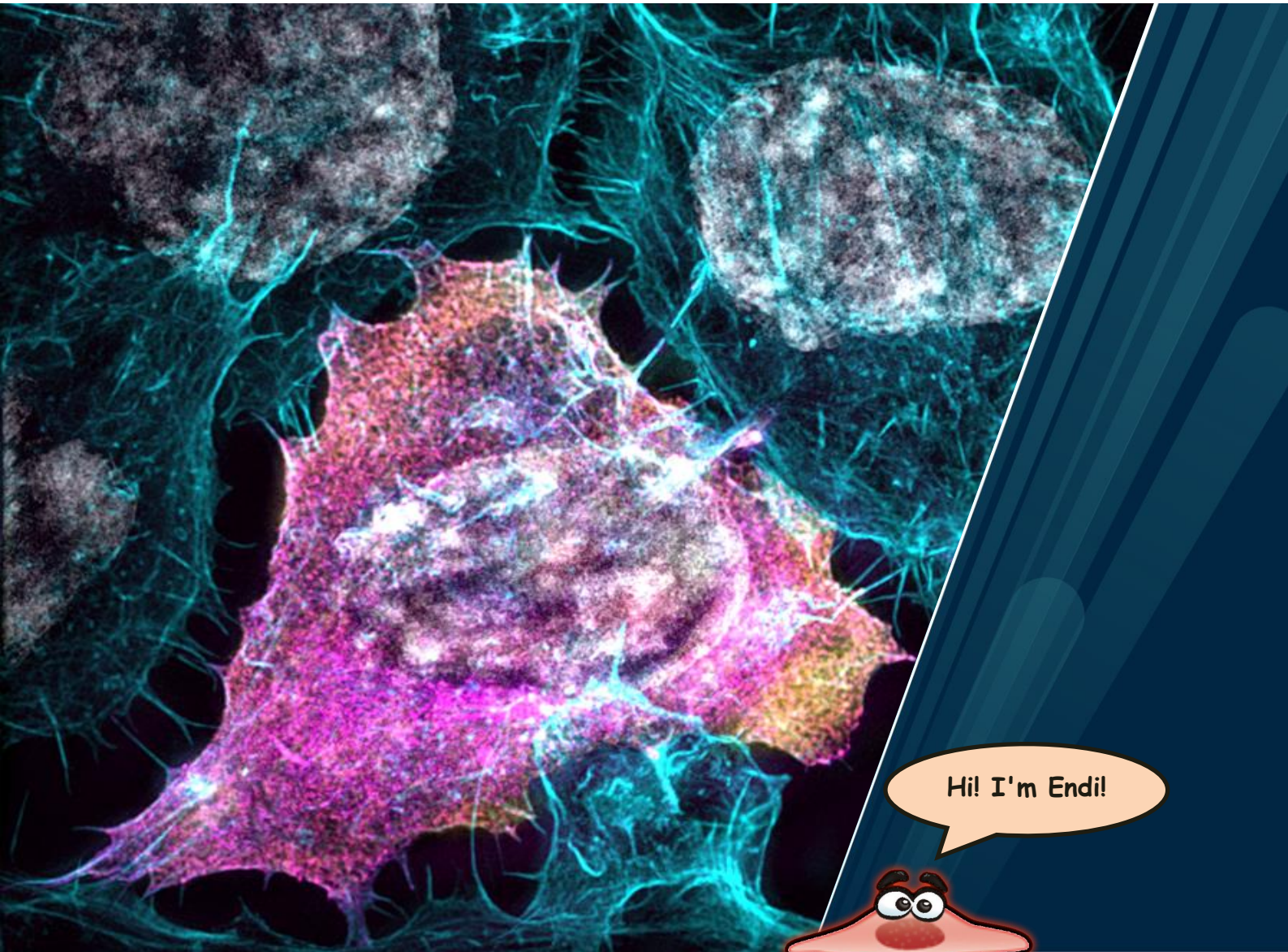
Department of Clinical Medicine

The Endothelial Cell Response to Inflammation, the Functional Role of the Endothelial-enriched Protein KANK3 and the Adipose Tissue Transcriptome

Unravelling the dynamics of endothelial cell biology and transcriptomics: Insights into cellular responses, temporal profiling, culture conditions, and tissue variation

Eike Christopher Struck

A dissertation for the degree of Philosophiae Doctor   August 2023

Hi! I'm Endi!

The Endothelial Cell Response to Inflammation, the Functional Role of the Endothelial-enriched Protein KANK3 and the Adipose Tissue Transcriptome

Unravelling the dynamics of endothelial cell biology and transcriptomics: Insights into cellular responses, temporal profiling, culture conditions, and tissue variation

Eike Christopher Struck

A dissertation for the degree of Philosophiae Doctor

Translational Vascular Research Group

Department of Clinical Medicine

Faculty of Health Sciences

UiT – The Arctic University of Norway

August 2023

"Do... or do not. There is no try."

Master Yoda, 'Star Wars Episode V - The Empire Strikes Back'

Acknowledgments

"Every champion was once a contender who refused to give up."

Rocky Balboa (Creed 2)

I am deeply grateful to have had the opportunity to embark on this research journey with an amazing team and great friends who were by my side. These projects were realised under the supervision of **Lynn Butler**, **Philip Dusart** and **Clément Naudin**. Fantastic researchers. Leaders. Humans. I thought of deep scientific quotes to include, but we have to live with Rocky, Pokémon and The Lego Movie.

"Everything is awesome! Everything is cool when you're part of a team!

Everything is awesome when you're living out a dream!"

Everything is awesome by Tegan and Sara, The Lego Movie

"If your actions create a legacy that inspires others to dream more, learn more, do more and become more, then, you are an excellent leader." - **To Lynn Butler**. I am incredibly grateful to have been your mentee. You are a fantastic scientist and a great leader. I cannot thank you enough for teaching me so much. Thank you for your encouragement, your patience, your understanding, your empathy when the stress was overwhelming. Which it rarely was, thanks to you. You inspired me to be a better scientist and a better person. Thank you for being the best boss I ever had. Additionally, to that, you are one of the best people that I have met in my life, and I could not be more happy with you as my supervisor. I would follow you anywhere. I would also hate the warm weather anywhere with you, discuss feminism (even if I am wrong some of the times and need to learn!), and chat about schnitzel or black forest gateau with you.

"In science, there are no shortcuts to truth." **- To Philip Dusart.** I would not be here and write this acknowledgment section if it weren't for you. You taught me so much when I was your master's student. Your padawan. And when I came here to do a PhD you immediately trusted me to be fit for the job. I learned so much from you and by trying to improve on your work. You have challenged me to think critically and constantly improve. I also love your bad hairs days. You are one of the most funny people I ever met and a loyal friend. Coffee is better than tea and I hope we will continue to play Wordle for a long time.

"Science is not only a disciple of reason but also one of romance and passion." **- To Clément Naudin.** When we first met, we clashed a little. But know now, that I feel nothing but respect and friendship for you. Our many brain stormings were probably the best scientific input I ever got. The spontaneous meetings and the pre-meeting-meetings were a joy. And you made my days a lot more happy. You and **Claudia** are some of the most wonderful people I've ever met. Thank you for your guidance, your trust, but primarily for your friendship. There is no party without you! Or egg cells. I will never forget you!

"Thank you for being a friend. Your heart is true, you're a pal and a confidant." **- To Sofia Öling!** From making party playlists, changing schedules so we can clean the lab together, sitting under the table, crying about 'mussel cells' or just doing cord extractions for 8 hours without drinking a sip of water! I could not have imagined this journey without you. Someone to vent. Someone to laugh with. Someone who became a friend for life. A bestie in a vestie! Or a bestie without a vestie. You do you! Someone who hits me with an axe while I hit her with a net. Whether we are crossing animals or poking the mons, we do it together!!! Besties forever!!! ROOOOBBBBIIIIIIIEE!!! LUCIFANS 4 LIFE!!!

"Everybody, rock your body!" - **To Marthe Norreen-Thorsen**, we started this journey together and were equally scared of what lies ahead. Our adventure may have had its ups and downs, but I am proud of what we have accomplished together. Thank you for always being honest with me and for your awesome taste in music, even though you are wrong on Billy Joel. May we have many more adventures together.

"Always, I know, you'll be at my show! Watching, waiting, commiserating!" - Great thanks also to my colleagues in TVR/CAP. **Maria**. You are by far the best human being on the planet. The nicest and most helpful person ever. You are amazing! Thank you for everything! **Jacob**. Thank you for a lot of guidance, for advice and constructive criticism and for making me a better researcher. **Jeong**! You will always be my wingman! My Ken! Thank you for your advice and the fantastic discussions! **Vera**, you are such a great human being, a great mother, super funny, and intelligent! I hope we can have many more coffees together and discuss life! Thanks to former members of CAP, that have helped me: **Laura, Julia, Leo.** You are wonderful!

> *"Every challenge along the way with courage, I will face.*
> *I will battle every day to claim my rightful place.*
> *Come with me, the time is right. There's no better team.*
> *Arm in arm, we'll win the fight, it's always been our dream!"*
>
> Pokémon theme song No. 1.

"Dashing and daring, courageous and caring, faithful and friendly, with stories to share." - I know that this journey would not have been the same without some great friends. To one of the most outstanding women, I have ever met. An intelligent, smart, and critical thinker, but more than anything else: A great friend: **Karolina Szafrańska**. Thank you for the laughs and the many discussions. For your scientific input. You will go far, and you completely deserve it. **Christopher Holte**, the darkness around the stars. Where the demons hide. Thank you for our many intelligent discussions and sometimes very dark jokes, and for contradicting me when I needed to be. You are smart and scary! To **Tom Boosch**: *"You are a man of few words."*

"The further on the edge, the hotter the intensity. Highway to the Danger Zone" - To **Casper Ullsten-Wahlund**, thank you for advice, jokes, and keeping me happy. Thank you for all the stories of being a dad and advice when I needed it, even though you think *"KeNnY rOgErS iS oLd PeOpLe mUsIc"*. You could not be more **wrong** about it. You are one of the most politically incorrect people I know, and I will officially disagree with everything you say! But I hope we will be friends forever.

"You are KENough!" - **Jakub** and **Tetyana**! You are wonderful people and great scientists! Thank you for tagging along! Tetyana, thank you for explaining every sauna detail and Jakub thank you for being one of the most beautiful beings in history of humankind. **Tine** and **Olivia**. You are way better people than I will ever be, and you truly make this planet a better place. Stay as you are and always stand up for your beliefs! Thank you for fantastic discussions. **Parwez**! It was fantastic to share office with you and that you joined our circle of crazy friends! **Basema**! You are amazing and a wonderful person! Just the way you are! You are smart and will make it far! **To the Skyfire Chronicles Group**! Up to the next adventure!

"A hero is a person intent on making this a better place for all people." Special thanks to **Christen Peder Dahl** for standing up for what is right and showing support, when we needed it the most. Without your office being a safe harbour for us, everything would have been so much harder.

"Never gonna give you up! Never gonna let you down!" - To **Peter McCourt** and his McTeam, McKids, and McEverything. You are **McAwesome**! Thank you for your help, guidance, and jokes.

"Power. Courage. Wisdom." - Big thanks to my twin brother **Ingmar** for possessing the ability to read and his understanding in statistics, which helped him to find the missing commas in my code. To my brother **Malte** for simply being my big brother. And to my nephews **Phileas** and **Elias** for lightening up my life. I love you.

"Hoffmann am Apparat!" - To **Annedore**, who would have told everyone off, who would criticise even a single word of this thesis! I will always be your "little one". Thank you for raising me and making me become the man I am today. You are my role model and I wish I turned out like you hoped.

"Simply the best!" - To **Anja & Ulf Kruse**. Thank you for helping me on the way and taking care of my Philip. Anja, thank you for telling me to *"come down to earth and explain research like a human being and not like an alien."* Ulf, thank you for always making me laugh *"Wir sehen uns auf der Fähre!."*

"Did I mention that I'm in love with you?" - To **Larissa**. Mein Lieblingsmensch. You have been my rock, my partner, and my best friend, and I cannot thank you enough for everything you have done for me. Without you, I would be lost in chaos. Thank you for the compassion, love, and critical advice when I needed it the most. You were not built for backing down and never will. Without you I would not have made it. I love you with every fibre of my being *"Mit Wellenrauschen, Sternenlicht, zartem Lachen im Gesicht".*

"Twinkle, twinkle, little star." - To **Philip Alexander Kruse**. My baby. *My love. My life.* The greatest thing I have accomplished in my life is simply to be your **"PAPA"**. I could not ask for a better title. You have taught me so much and I am grateful for every moment we have spent together. Your presence in my life has made me a better person. Thank you for being the light in my life and reminding me of what truly matters: To watch you explore the world, see the wonders of life, and learn new things every day*. May you never lose the wonder in your soul… May you always have a blanket for the cold… May the living light inside you be the compass as you go… May you always know you have my hand to hold*…

"Behind every child who believes in them, are parents who believed in them first." - To my parents, **Sigrid Falk-Struck** and **Klaus Struck,** who never missed an opportunity to tell me how proud they are of me. You have been my pillars of support throughout my life, and I cannot thank you enough for the love, guidance, and encouragement you have provided me with. Your unconditional support has been a constant source of comfort and inspiration for me, and I cannot imagine where I would be without you both. Thank you!

To **Philip Alexander van Vloten**

"We can be heroes in our own lives, every one of us, if we only have the courage to try.",
Optimus Prime in Transformers: The Last Knight

Our many adventures together ended too soon. It has been almost ten years now, but it feels like yesterday when you called me for the last time in the middle of the night. You would have told me that you knew I could do it. You would be proud of me. But you would not have said it, instead you would have said something much cooler, such as: *"You should include a Stargate quote that no one understands but us. Maybe something in Goa'uld."* And I would have done it. For you. And maybe, just maybe, it would have been something like this: ***"Tal met. Priem ta shree, tal ma." Our love does not end in death.*** Or something like ***"Tak mal tiak", You are remembered!*** Thank you for being a teacher, a friend and for all the adventures together. *May the falcon always fly high, Thoron!*

Now let's do some science!

Summary

The vascular endothelium is the innermost cell layer of all blood and lymphatic vessels, which reacts to diverse stimuli with a multitude of biological functions. This thesis explores various aspects of endothelial cell biology, including inflammatory responses, cell type enriched transcriptional profiles, and functional investigation of an uncharacterised gene predominantly expressed in this cell type .

Paper I focuses on the **response of endothelial cells to the pro-inflammatory cytokine tumour necrosis factor (TNF).** Endothelial cells were exposed to TNF and gene expression profiles were analysed at various time points up to 72 hours, to map global temporal changes. The analysis revealed two distinct response phases: an acute one occurring within 1-4 hours and a later one between 12-24 hours. We identified uncharacterised genes, including long non-coding RNAs and pseudogenes, which were regulated during the acute response. Genes regulated in the later phase were largely related to interferon signalling and anti-viral responses. We provide an online portal (www.endothelial-response.org) for user friendly access to all data.

Paper II utilises human **adipose tissue bulk RNA-sequencing data to generate a cell type specific transcriptomic atlas**. An integrative correlation method was used to identify over 2,300 cell-type-enriched transcripts, including 157 endothelial-enriched transcripts in visceral and 155 in subcutaneous adipose tissue. Comparative analysis identified cell types driving gene expression differences in between subcutaneous and visceral adipose tissue and sex-specific cell type profiles. Data is incorporated into the Human Protein Atlas (www.proteinatlas.org/). These results, in combination with other related studies from our group, were used as the basis for selection of an endothelial enriched gene for characterisation in Paper III.

Paper III focuses on the **functional investigation of the endothelial-enriched transcript *KANK3*.** KANK3 expression and distribution was modified by shear stress and *KANK3* knockdown induced enhanced cell motility. The expression levels of the coagulation factor tissue factor (*F3*) and of the anti-fibrinolytic factor PLAT were also modified by *KANK3* knockdown, with corresponding enhancement of thrombin generation in plasma, attributable to tissue factor up regulation.

This thesis contributes to our understanding of endothelial cell responses to inflammation, demonstrates how to extract useful information from existing datasets and provides the first functional characterisation of an endothelial enriched gene.

List of Papers

Paper I

Eike C Struck, Tatiana Belova, Ping-Han Hsieh, Jacob Odeberg, Marieke L Kuijjer, Philip J Dusart, Lynn M Butler: ***Global transcriptome analysis reveals distinct phases of the endothelial response to tumour necrosis factor.*** Pre-print: https://www.biorxiv.org/content/10.1101/2023.06.04.543378v1
Re-submitted following revision at ***Journal of Immunology***.

Paper II

Marthe Norreen-Thorsen, Eike Christopher Struck, Sofia Öling, Martin Zwahlen, Kalle Von Feilitzen, Jacob Odeberg, Cecilia Lindskog, Fredrik Pontén, Mathias Uhlén, Philip James Dusart, Lynn Marie Butler: ***A human adipose tissue cell-type transcriptome atlas.*** Cell Reports (Volume 40, Issue 2, 12 July 2022, 111046). https://doi.org/10.1016/j.celrep.2022.111046

Paper III

Eike Christopher Struck, Sofia Maria Öling, Philip James Dusart, Marthe Norreen-Thorsen, Julian Connor Eckel, Larissa Dorothea Kruse, Casper Ullsten-Wahlund, Jacob Odeberg, Clément Naudin, Lynn Marie Butler: **KANK3 is a shear stress regulated endothelial protein with a role in cell migration and tissue factor regulation.**
Manuscript.

Abbreviations

aa	Amino acids
AT	Adipose tissue
ATP	Adenosine triphosphate
BAT	Brown Adipose Tissue
BH	Benjamini-Hochberg
BSA	Bovine Serum Albumin
CAT	Calibrated Automated Thrombogram
CDH5	Cadherin 5 / VE-Cadherin
cDNA	Complementary DNA
CTRL	Control
CV	Coefficient of variation
DAPI	4′,6-diamidino-2-phenylindole
DEG	Differentially Expressed Gene
DNA	Deoxyribonucleic Acid
DSG2	Desmoglein-2
EC	Endothelial Cell
ECM	Extracellular Matrix
EDHF	Endothelium derived hyperpolarising factor
eNOS	Endothelial nitric oxide synthase (also NOS3)
ESAM	Endothelial selective adhesion molecule
FA	Focal adhesion
FACS	Fluorescence-activated cell sorting
FC	Fold change
FVII	Coagulation Factor 7
FVIIa	Coagulation Factor 7, activated
FIII, F3	Coagulation Factor 3, Tissue Factor (TF)
GTEx	Genotype-Tissue Expression project
GO	Gene ontology
HCAEC	Human Coronary Artery Endothelial Cells
HDMEC	Human Dermal Microvasculature Endothelial Cells
HPA	Human Protein Atlas
HPAEC	Human Pulmonary Artery Endothelial Cells
HREC	Human Retinal Endothelial Cells
HUVEC	Human Umbilical Vein Endothelial Cells
ICAM1	Intercellular Adhesion Molecule 1
IFN	Interferon
IHC	Immunohistochemistry
IL1B	Interleukin 1B
ISG	Interferon-stimulated Gene
JAM	Junctional adhesion molecule
JAM-A/ F11R	Junctional adhesion molecule A
KANK	KN motif and ankyrin repeat domain-containing
KANK3	KN motif and ankyrin repeat domain-containing 3
kb	Kilobases
KIF21A	Kinesin Family Member 21A
LC	Leukocyte
LIMMA	Linear Models for Microarray and RNAseq Data
LPS	Lipopolysaccharide

LSEC	Liver Sinusoidal Endothelial Cells
MEM	Minimal Essential Medium
mRNA	Messenger RNA
NO	Nitric oxide
NOS3	Nitric oxide synthase 3 (see eNOS)
ρ	Spearman's rho, Speramn's correlation coefficient
p	P value
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PLAT	Plasminogen Activator, Tissue Type
PLAU	Plasminogen Activator, Urokinase
PROCR	Protein C Receptor
PECAM1	Platelet endothelial adhesion molecule
qPCR	Quantitative polymerase chain reaction
Ref.T	Reference Transcript
RNA	Ribonucleic acid
RNAseq	RNA sequencing
RT-PCR	Reverse Transcription PCR
ROS	Reactive Oxygen Species
SAT	Subcutaneous Adipose Tissue
scRNAseq	Single cell RNA sequencing
SD	Standard deviation
SELE	E-Selectin
SELP	P-Selectin
SEM	Standard error over mean
SMC	Smooth Muscle Cell
SNPs	Single nucleotide polymorphism
TF	Tissue Factor *(not to be confused with transcription factor)*
TFPI	Tissue Factor Pathway Inhibitor
TM	Thrombomodulin
TNF	Tumour Necrosis Factor
TNF+	TNF-treated
tPA	Tissue-type plasminogen activator
TPM	Transcripts per Million
UMAP	Uniform Manifold Approximation and Projection
uPA	Urokinase plasminogen activator
UT	Untreated
UTP	Uridine-5-triphosphate
VAT	Visceral Adipose Tissue
VCAM1	Vascular Cell Adhesion Molecule 1
vWF	von Willebrand factor
WAT	White Adipose Tissue
WGCNA	Weighted Gene Coexpression Network Analysis
WPB	Weibel-Palade Bodies

Contents

List of Figures

1 Introduction

The articles summarised in this thesis explore different facets of endothelial cell biology and transcriptomics, shedding light on the complex processes involved in cellular responses to stimuli, tissue dependent transcriptomics, as well as the role of cell type enriched transcripts. The following introduction summarises endothelial participation in key vascular processes, important aspects of adipose tissue biology, the basics about transcriptomics, including normalisation methods and statistical analyses, and dives deeper into important parts from the respective studies.

1.1 The Vascular Endothelium
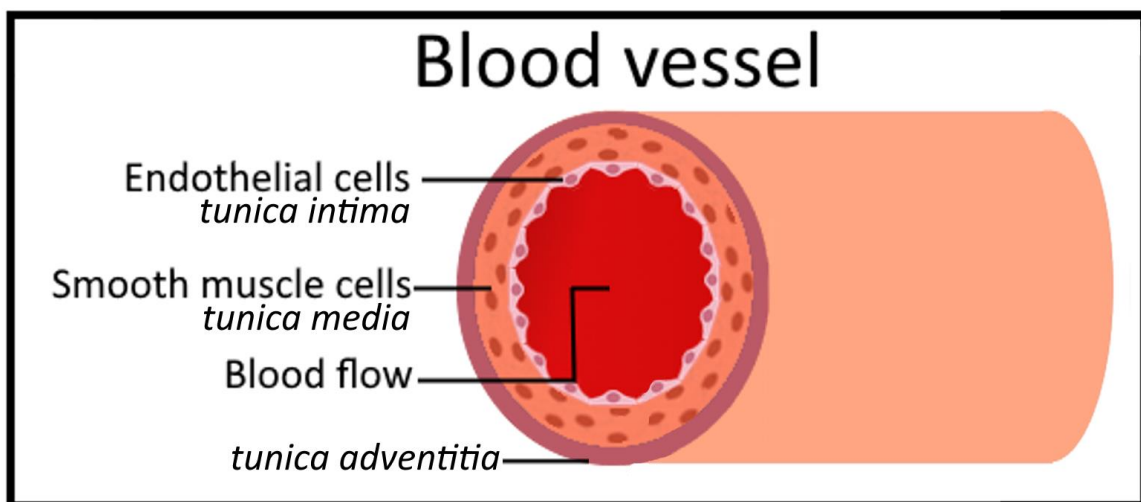
Figure 01: The general structure of blood vessels: The blood vessel is made up of three layers: the tunica intima, containing the endothelium and the basal lamina, the tunica media, which is made up of smooth muscle cells and elastic fibres, and lastly the tunica adventitia, that makes up the outermost layer, composed of connective tissue (1).

The vascular endothelium is the innermost cell layer of all blood and lymphatic vessels (Figure 01 & 02). Amongst other functions, it separates the blood from the surrounding tissue (2). With a surface area of 4000–7000 m^2 and an average weight of ~1kg the endothelium is one of the biggest organs in the average adult human body (3).

Earlier described as merely an inert separator between blood circulation and organs with a rather passive role (4), the endothelium is now well recognised as a dynamic organ that reacts to diverse stimuli and has a multitude of biological functions (2). Endothelial cells (EC) play pivotal roles in many vascular processes, such as control of haemostasis and thrombosis (5), mediation of inflammation (6), vascular tone (7), angiogenesis (8), and metabolic transport between blood vessels and tissues (9).
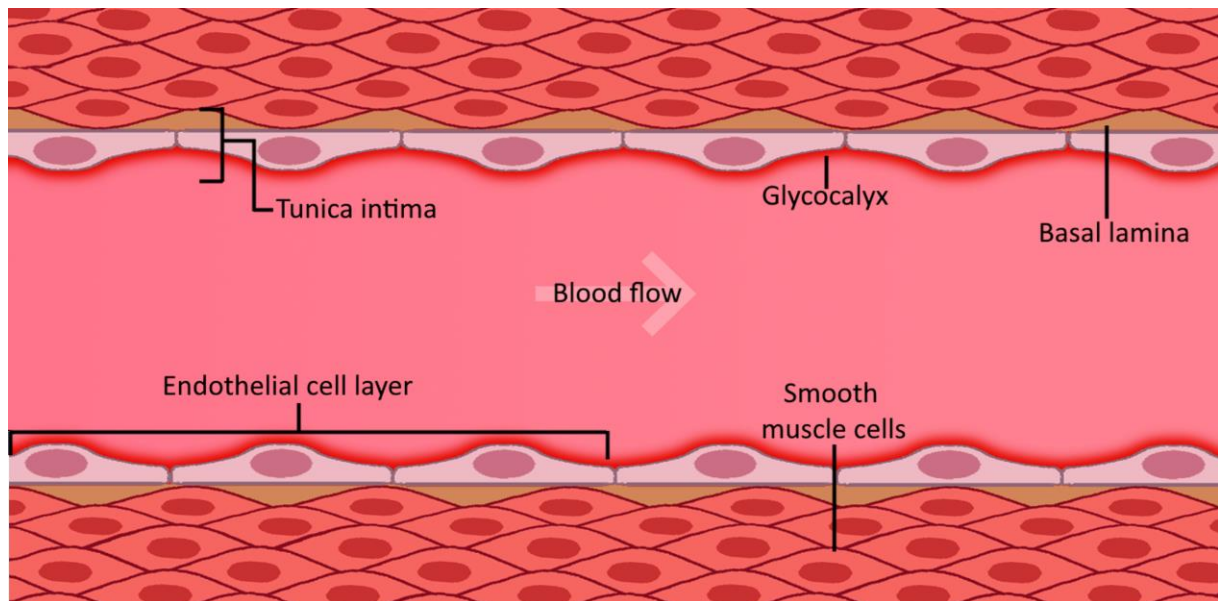
Figure 02: The vascular endothelium: *Cross section of a blood vessel highlighting the structure of the vascular endothelium and the surrounding smooth muscle cell layers. The vascular endothelium is the innermost layer of all blood vessels and separates blood from surrounding tissue (2).*

Healthy EC display anti-thrombotic and anti-inflammatory properties, while endothelial dysfunction leads to a reversal of these properties. The healthy endothelium is defined by the ability to maintain the vascular tone, and constitutively expresses vasodilators, such as nitric oxide (NO) and prostacyclin, alongside minimal levels of reactive oxygen species (ROS) and uric acid (10, 11). Endothelial dysfunction refers to an activated or damaged endothelium which exhibits an impairment in its regular function. It is characterised by a reduction in vasodilation, increase of ROS, and upregulation of adhesion molecules. EC activation leads to a prothrombotic, pro-inflammatory surface with increased permeability (12). Endothelial dysfunction can cause a range of vascular disorders, which include atherosclerosis, hypertension, coronary artery disease and thrombosis (13).

1.1.1 Endothelial Heterogeneity

Endothelial phenotype and function differs depending on location and tissue (14, 15) e.g., dynamic vascular fenestrations found in liver sinusoidal EC, facilitate bidirectional transport of substrates (16). The selectively-permeable EC in the blood brain barrier regulate in- and outflow of substances (17). Heart EC have specialised functions in cardiac physiology (18). EC exhibit differences along the vascular tree, for example EC dependent dilation depends on the vessel origin (11). Moreover, within capillaries of a single organ, distinct subpopulations of specialised EC can be identified (19).

Endothelial specific gene expression and functions have been previously described in its relation to different vascular beds and tissues. They are highly specific compared to housekeeping genes, constitutive genes that are essential for cellular functions, and other gene clusters (20, 21) and one could reasonably infer that tissue- and cell-type specific expression relates to tissue- and cell-type specific functions (22).
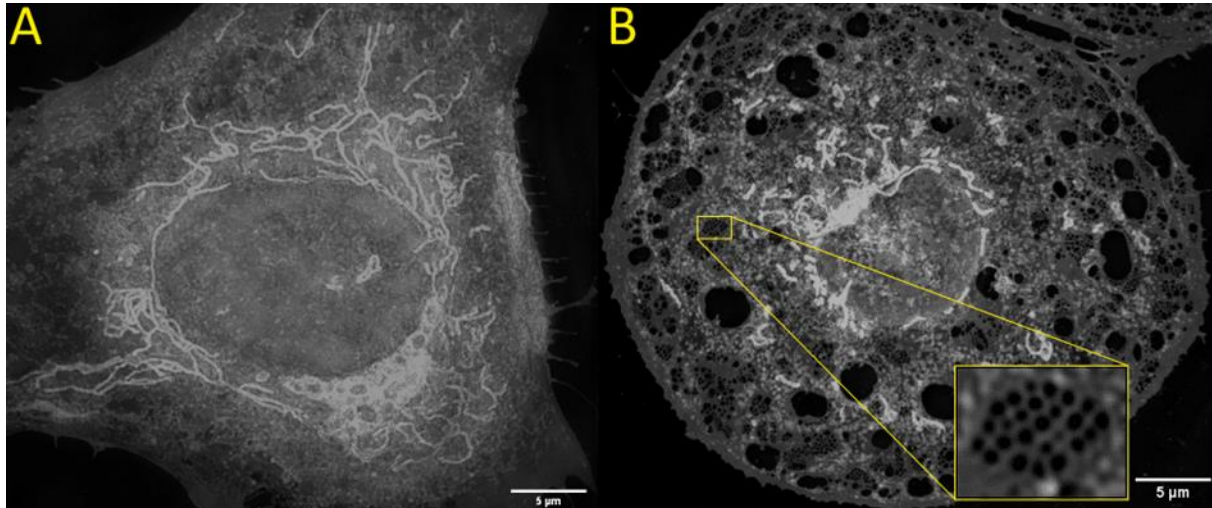
Figure 03: Phenotypic differences of endothelial cells from different sources: EC phenotype differs depending on vascular bed and cellular function. Human umbilical vein endothelial cells (HUVEC) and liver sinusoidal endothelial cells (LSEC) were stained for cell plasma membrane (Cell Mask™). (A) HUVEC display an even, unperforated surface, while (B) LSEC are fenestrated. Fenestrations (yellow box) are transmembrane nanopores that enable passive transportation between the blood components and underlying cells. They cluster in groups called "sieve plates" (23). These structures are absent in HUVEC [Structured Illumination Microscopy images received, with courtesy, from Larissa Kruse].

A comparison between human umbilical vein EC (HUVEC), juvenile EC extracted from the vein of the umbilical cord, with liver sinusoidal EC (LSEC), mature EC extracted from the small capillaries of the liver, is shown in Figure 03. Both cell types fulfil specialised functions in their respective organ and express different phenotypes. HUVEC have a smooth continuous surface (Figure 03 A), whereas LSEC have clusters of holes, called fenestrations (Figure 03 B), which allow for rapid exchange of macromolecules between blood and liver tissue (16, 24).

1.1.2 The Endothelium as Mediator of Vascular Tone

EC maintain a delicate balance between the relaxation of blood vessels (vasodilation) and the contraction of blood vessels (vasoconstriction) through messaging to smooth muscle cells (SMC) by releasing vasoactive factors. Thereby, they influence blood flow and blood pressure, a process known as "keeping the vascular tone" (25).
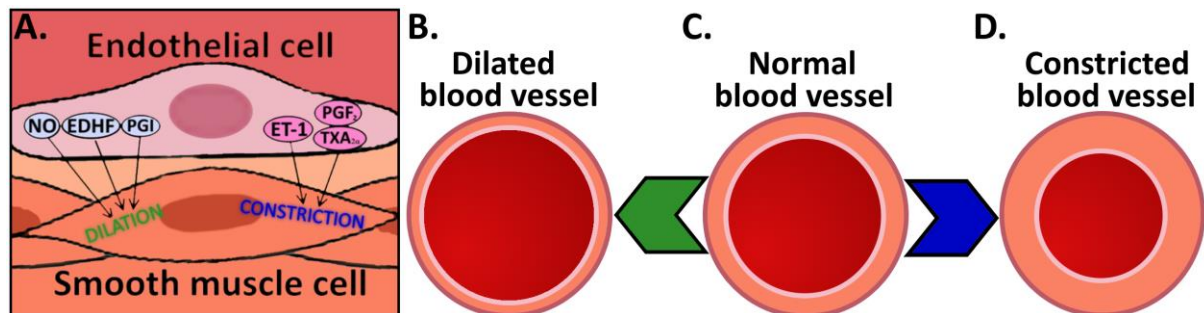
***Figure 04: The endothelium mediates blood pressure through communication with SMC: (A)** Endothelial and smooth muscle cell-cell communication controls blood pressure, by dilating and constricting the blood vessel through release of various relaxing and contracting factors. (C) The normal blood vessel displays a balanced lumen size and a normal diameter. (B) When dilated the vessel diameter is larger and the blood flow is increased. (D) A constricted blood vessel is tightened, and the diameter is reduced, leading to higher blood pressure and reduced blood flow* (25, 26)*.*

EC-dependent vasodilation is mediated by a multitude of factors, such as nitric oxide (NO), prostacyclin, and endothelium derived hyperpolarising factor (EDHF). EC vasoconstriction is facilitated by factors, such as thromboxane, prostaglandin 1, and endothelin-1 (7) (Figure 04). The release of NO in response to shear stress stimulates vasodilation, which is directly proportional to the amount of NO released. Additionally, NO inhibits platelet aggregation, monocyte adherence, and the proliferation *of* vascular smooth muscle cells (27).

In response to shear stress or cellular damage, EC release adenosine triphosphate (ATP), uridine-5'-triphosphate (UTP), and endothelin-1 which act as vasoconstrictors, and produce less NO and prostacyclin which leads to unopposed vasoconstriction (28).

1.1.3 The Endothelial Barrier Function

Although functionally diverse, the vascular endothelium primarily acts as a selective barrier between circulating blood and surrounding tissue by regulating tissue perfusion. The permeability of the endothelial monolayer is tightly regulated and varies depending on the organ, and is mainly maintained through endothelial junctions, extracellular matrix and the endothelial glycocalyx (29). EC allow passive diffusion of small substances, such as solutes, gasses and ions, but under resting conditions repels structures of higher molecular weight or size, such as proteins or cells (30).
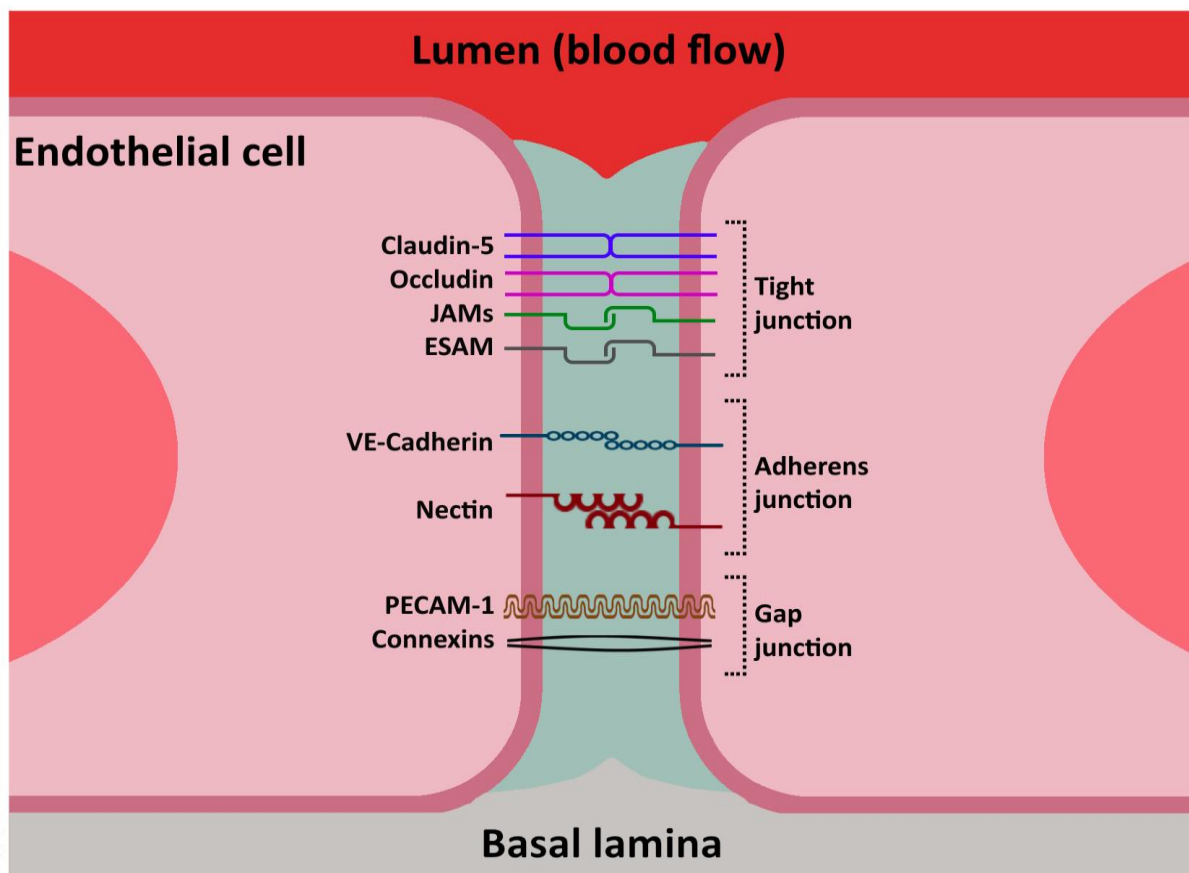
Figure 05: The endothelial junction: *The integrity of endothelial junctions are maintained by several proteins in three types of junctions: Tight junctions, adherens junctions and gap junctions. A variety of adhesion molecules, such as JAMs, ESAM, CLDN5, Occludin, Connexins, PECAM-1, NECTIN, and CDH5, are responsible for the regulation and maintenance of endothelial junctions (31, 32, 33).*

Endothelial junctions are the connecting structures between EC, providing vasculature integrity and restricting the extravasation of proteins, solutes, leukocytes, and metastatic cells across the EC barrier. Endothelial junctions are controlled and maintained through different adhesion molecules, including platelet endothelial cell adhesion molecule (PECAM1), junctional adhesion molecule A (F11R/(JAM-A)),

junctional adhesion molecule C (JAM3/(JAM-C)), endothelial cell-selective adhesion molecule (ESAM), claudin-5 (CLDN5), desmoglein-2 (DSG2), nectin-2 (NECTIN2) and cadherin-5 (CDH5)(34).

These molecules constitute three distinct types of junctional complexes: tight junctions, adherens junctions, and gap junctions (Figure 05). Tight junctions are positioned distal to the basal lamina. They create a continuous belt-like structure encircling the cells and regulating paracellular permeability, the ability to pass between adjacent cells, through upholding the endothelial barrier function (35). Adherens junctions are located beneath tight junctions. They facilitate cell-to-cell adhesion by interacting with the actin cytoskeleton (35, 36). Gap junctions serve as communicative structures between adjacent cells and enable the passage of small solutes with low molecular weight (36). The expression and arrangement of these junctions vary based on the specific type of blood vessels and the organ specific needs. Several EC types, such as liver sinusoidal EC, are fenestrated, allowing rapid exchange of macromolecules between blood and tissue (24) without need for junctional changes.

The extracellular matrix (ECM) is a network of proteins with multiple functions: it offers physical and mechanical support to cells, while also facilitating biochemical communication between them (37). Remodelling of the endothelial basement membrane through deposition of extracellular matrix proteins influences endothelial adhesion and barrier integrity by modulating basal cell contractility and force transduction at the CDH5 junctions. Hence, ECM remodelling is well recognised to play a major role in disease progression (38, 39, 40).
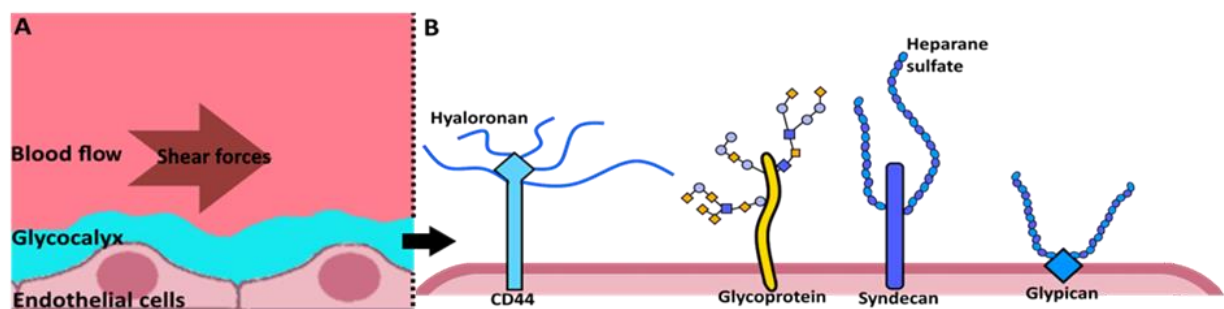
Figure 06: The endothelial glycocalyx: *(A) The endothelial glycocalyx is a layer on the lumen of EC that separates the cell from shear forces of the blood flow. (B) The glycocalyx itself is made up of several components, mainly glycoproteins and proteoglycans, hyaluronan and proteins (41, 42, 43).*

The endothelial glycocalyx is a negatively charged carbohydrate-rich layer (Figure 06) connected to the endothelium through proteoglycans and glycoproteins. With a

thickness of 200-400nm, it repels proteins, platelets, and blood cells. It prevents undesired interactions between blood and the surrounding tissue. Disturbance or alterations of the negative charge can result in increased vascular permeability. Furthermore, the inflammatory cytokine TNF has been shown to be a potent disturbing factor to the endothelial glycocalyx, leading to its shedding (44, 45, 46, 47).

1.1.4 The Vascular Endothelium in Acute Inflammation

Vascular endothelial cells are important regulators of inflammatory response. They providing an anti-inflammatory and anticoagulant surface in resting state (48) and undergo morphological and functional modifications when activated by various stimuli. Such changes include the presentation of pro-inflammatory mediators and the recruitment of leukocytes (LC) through the expression of adhesion molecules (49). LC are recruited in a series of sequential interactions between endothelial proteins and leukocyte ligands. The interactions include the capture of LC through selectins, followed by rolling, slow rolling, crawling, firm adhesion and transcellular or paracellular migration through the endothelial layer and the movement to the target tissue (Figure 07) (6, 50, 51).
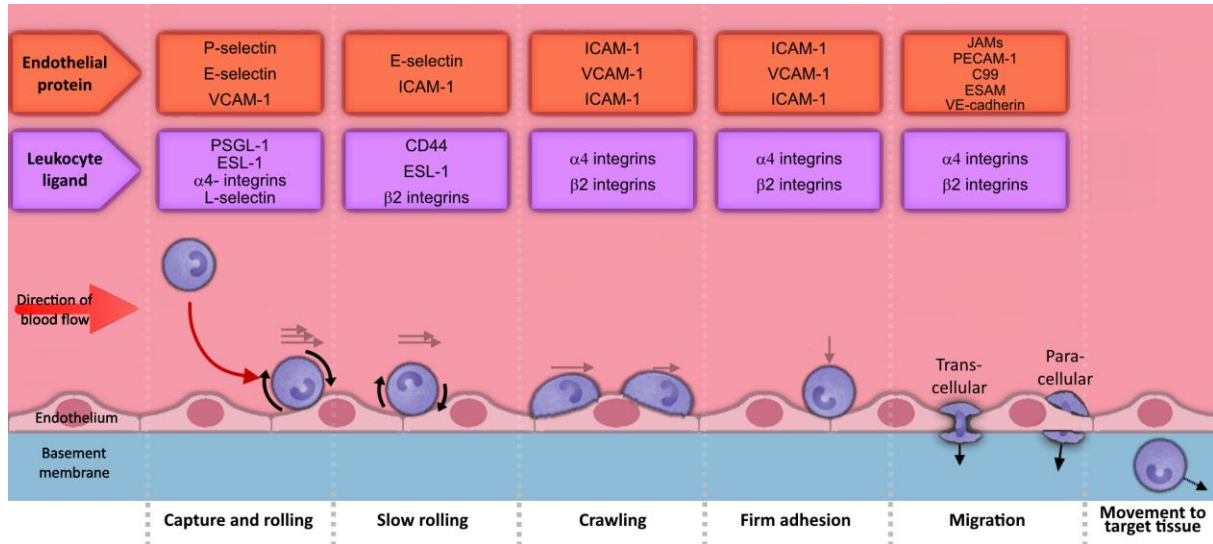
Figure 07: Leukocyte recruitment to the site of inflammation is facilitated by EC-LC interactions: *Leukocyte adhesion is a multistep cascade depending on several factors. Following activation, EC present adhesion proteins, such as P-selectin, E-selectin, ICAM1 and VCAM1 to recruit flowing leukocytes to the EC monolayer and mediate adhesion. Leukocyte recruitment is initiated by low-affinity interactions that allow rolling along the surface, followed by high affinity interactions that induce firm adhesion. Finally, the leukocyte can move through the endothelial layer by paracellular or transcellular migration into the target tissue (Figure adapted from Ley, Laudanna (50) (2007)).*

7

The interaction between EC and LC is initiated by different adhesion molecules, such as selectins (E-selectin (SELE), P-selectin (SELP)). Upon endothelial activation, P-selectin and Von Willebrand factor (VWF), are rapidly secreted from Weibel-Palade bodies (small granules located in vascular EC) onto the endothelial surface. The rapid release of P-selectin is then followed by *de-novo* transcription of E-selectin (52, 53). Selectins interact with their respective LC ligands to enable capture and rolling on the endothelial surface. Further interaction of LC integrins with EC adhesion molecules intercellular adhesion molecule 1 (ICAM1) and vascular cell adhesion protein 1 (VCAM1) triggers slow rolling, followed by crawling, firm adhesion and arrest. This processes is mainly mediated through chemokines which activate LC receptors and the endothelium to control homing of LC to the target tissue (54).

The adhesion then strengthens, spreads and LC transmigrate through the endothelial barrier via transcellular or paracellular diapedesis, facilitated by a set of different molecules, expressed at the adherent junctions of the endothelium, among others junctional adhesion molecules (JAMs), PECAM1 (55), CDH5 (56) and ESAM (57). While cell-cell junctions facilitate the passage of most leukocytes across the endothelial border in a paracellular way, studies have suggested that CDH5 mediated migration might occur in a transcellular fashion (56), although results are conflicting (58). Following transmigration, LC move on to the inflamed target tissue (59).

1.1.5 Cytokines in Vascular Biology

Study 1 of this thesis explores the endothelial response to TNF over time and presents findings regarding a secondary signalling pathway involving interferon-stimulated genes. Study 3 employs TNF and IL-1B as stimulation agents to induce tissue factor production.

Cytokines are small proteins that act as mediators of cell signalling by binding to receptors. Pro-inflammatory cytokines, such as tumour necrosis factor (TNF), interleukins, and interferons play important roles in vascular biology and inflammation.

TNF is a pro-inflammatory cytokine that is mainly produced by pathogen-activated macrophages and monocytes. It can activate EC in an analogous manner to bacterial lipopolysaccharide (LPS) and interleukin-1B (IL1B). Activation by TNF leads to increased vascular permeability and recruitment of leukocytes to the infected target tissue through expression of adhesion molecules (E-selectin, VCAM1, ICAM1) on the endothelial surface (60, 61, 62, 63).

Interferons (IFN) are mediators of innate and adaptive response in EC and key players in antiviral defence. In the presence of viruses, IFN variants are produced by multiple cell types, such as parenchymal cells, fibroblasts, macrophages, and dendritic cells (64). IFNs induce interferon stimulated genes (ISGs), proteins with antiviral properties that can inhibit cell proliferation, mainly through the JAK/STAT pathway, followed by the activation of nearby cells (65).

1.1.6 The Endothelium in Angiogenesis and Vasculogenesis

Endothelial cells are central in vascularisation, the formation of blood vessels. Two key processes are angiogenesis, the creation of new blood vessels from existing vasculature, and vasculogenesis, the de novo creation of new blood vessels (66, 67).
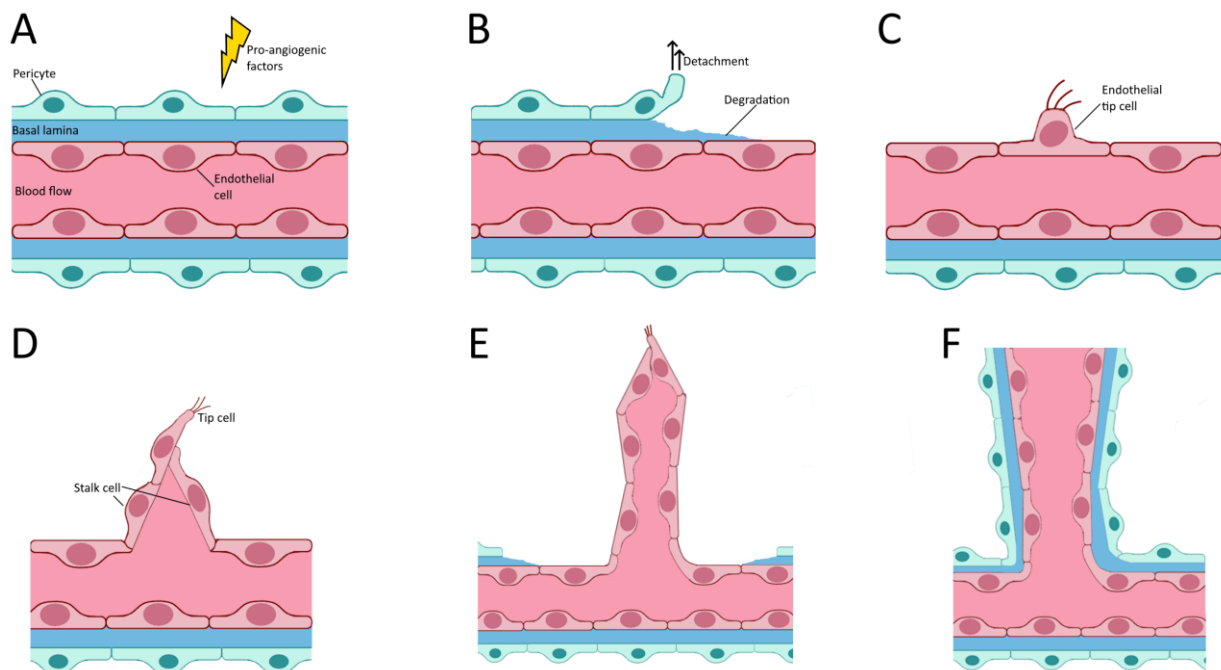
Figure 08: The endothelium in angiogenesis: (A) Angiogenesis is initiated by pro-angiogenetic factors such as VEGF. (B) Pericytes, basement cells lining the capillaries and venules, detach and the basal lamina is enzymatically degraded. (C) EC differentiate into tip cells and move out of the vessel wall. (D) They extend into sprouting vessels with "stalk cells" following the tip cell (E) and extend the vessel or sprout (not shown). (F) The vessel matures by attraction of new pericytes and formation of a new basement membrane. Figure design adapted from (66).

Angiogenesis (also known as neovascularisation) is the process of new blood vessel formation from pre-existing ones. This can occur when meet tissue oxygen and nutrient requirements are not met, to support wound healing, and form the vasculature during embryogenesis. Therefore, common triggers for angiogenesis are hypoxia and inflammation (66). Angiogenesis is kept in equilibrium by the interplay of pro-

angiogenic (VEGF, MMP) and angiostatic factors (Endostatin, TIMP), which control the "angiogenic switch".

The formation of new vessels in adults is a coordinated multi-step process (66, 68).

1. **Angiogenic activation:** The initiation is triggered by pro-angiogenic agents, including vascular endothelial growth factor (VEGF) and basic fibroblast growth factor (bFGF) (Figure 08A).
2. **Matrix degradation:** Pericytes, stabilising cells that surround capillaries, detach and release from the endothelium. EC then break down the basement membrane through the secretion of proteolytic enzymes, known as matrix metalloproteinases (Figure 08B).
3. **Endothelial differentiation:** EC differentiate out of the vessel wall under precise Notch/DLL4 control. These EC become tip cells, the leading cells at vascular sprouts (Figure 08C-D).
4. **Vessel formation:** The tip cells extend filopodia (Figure 08E) that penetrate the surrounding tissue and lead sprouting vessels towards angiogenic cues. Proliferating stalk cells follow, the sprouts elongate and create lumens and connect to form blood-perfused microvascular networks.
5. **Maturation:** New basement membranes and perivascular cell recruitment solidifies these networks (Figure 08F).

Angiogenesis can be analysed *in vivo*, where sprouting of new blood vessels can be observed using different markers and in different models, such as zebra fish or mouse. *In vitro* measurement of angiogenesis can be done using various assays, such as wound healing assays ("scratch assay"), tube formation assays, in microfluidic assays, spheroid sprouting assays, or trans well migration assays (66).

The term vasculogenesis, on the other hand, refers to the *de novo* generation of blood vessels from mesodermal precursors that differentiate into EC. This mechanism plays a pivotal role in embryonic development and tissue regeneration. It is essential for constructing a vascular network to supply tissues with oxygen and nutrients (67, 69).

1.1.7 The Vascular Endothelium in Haemostasis

The endothelium is a key player in the maintenance of vascular haemostasis. In healthy conditions, the endothelium is non-adhesive and anticoagulant, inhibiting the activation of platelets and the initiation of the coagulation cascade and maintaining the

fibrinolytic system. Molecular links between inflammation and thrombosis have been previously described, and the vascular endothelium is gaining increasing recognition as a key player in thrombo-inflammation (70, 71, 72, 73).
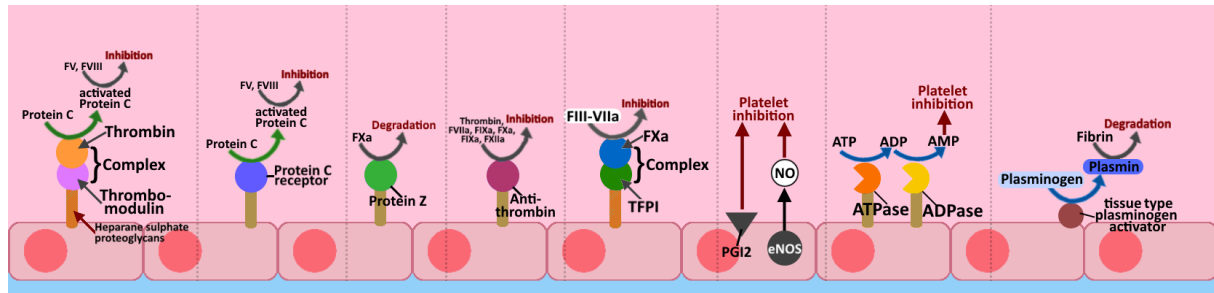
Figure 09: Overview of EC derived antithrombotic factors: *Endothelial cells express and secrete a multitude of anticoagulant factors that can inhibit platelets and coagulation factors (adapted from references as below).*

EC produce and secrete a multitude of anticoagulant and pro-fibrinolytic factors (Figure 09), such as:

- **Thrombomodulin (THBD, TM)** is a transmembrane glycoprotein that complexes thrombin, converting it to an anticoagulant enzyme, and furthermore activating the anticoagulant protein C (74).

- **Endothelial protein C receptor (EPCR, PROCR)** is a multi-ligand receptor, which acts as a key player in protein C mediated anticoagulation by induction of protein C activation (75, 76).

- **Protein Z / Vitamin K-dependent protein Z (PROZ)** is a glycoprotein that inhibits coagulation through degradation of Factor Xa (77, 78).

- **Antithrombin III (SERPINC1, ATIII)** is an endothelial expressed surface protein, which is mostly bound to endogenous heparin-like substances. It inhibits thrombin by forming a complex with it and coagulation factor Xa and inactivates them (70, 79).

- Additionally, EC express **heparan-sulphate proteoglycans** and **heparan-like molecules**, which act as cofactors for ATIII and increase its activity (80).

- **Tissue factor pathway inhibitor (TFPI)** is an anticoagulant that reduces the activity of the FIII-VIIa complex as well as forms of prothrombinase (81).

- **PGI2** and **NO** act as vasorelaxant molecules that inhibit platelet activation, with NO acting as additional enhancer for PGI2. PGI2 has been previously described as the most potent natural inhibitor of platelet aggregation (82, 83, 84). Endothelial nitric oxide synthases (eNOS, NOS3) produce nitric oxide (85).

11

- **ATP** activates adenylate cyclase, leading to elevated levels of cyclic adenosine monophosphate (AMP). The rise in **cyclic AMP** inhibits platelet activation Endothelial cells express ATPases and ADPases to reduce ATP to adenosine diphosphate (ADP) and ADP to AMP (86).

- **Tissue-type plasminogen activator (t-PA, PLAT)** is a serine protease and a key player in the breakdown of blood clots. It is found on endothelial cells and converts plasminogen to plasmin, the primary enzyme in fibrinolysis (87).
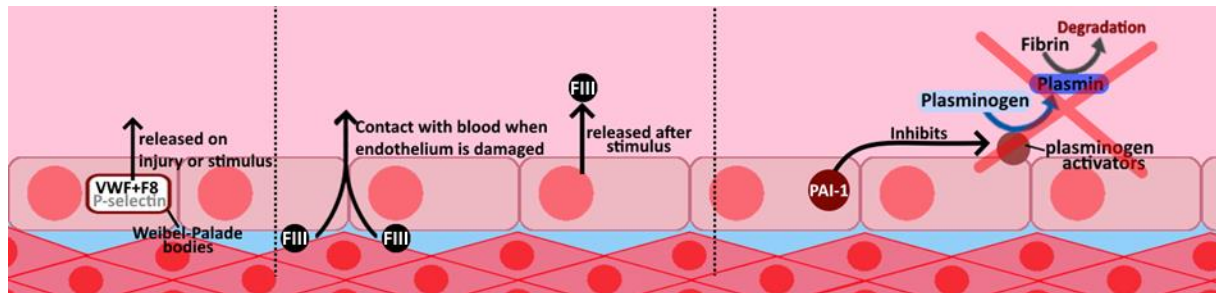
Figure 10: Overview of EC derived prothrombotic factors: *Upon activation endothelial cells express a multitude of procoagulant factors (adapted from references as below).*

EC release pro-coagulant factors such as tissue factor (F3), von Willebrand Factor (vWF), plasminogen activator inhibitor 1 (SERPINE1) and coagulation factor VIII (F8) (Figure 10).

- **von Willebrand Factor (vWF)** is a mediator of platelet adhesion and shear-induced aggregation, which is stored in Weibel-Palade bodies and rapidly secreted upon endothelial activation (88).

- **Factor 8 (FVIII, F8)** has been reported in multiple EC types and is packaged in Weibel-Palade bodies together with VWF. It is a key player in the intrinsic coagulation cascade and acts as cofactor in the activation of factor X by factor IXa (89, 90), leading to the common pathway of coagulation (89, 90).

- **Tissue Factor (FIII)** is the initiator of the extrinsic coagulation pathway by complexing and activating coagulation factor VII and activating FX and the common pathway of coagulation (91).

- **Plasminogen activator inhibitor-1** (PAI-1) is an endothelial expressed protein that inhibits fibrinolysis by acting as inhibitor of tPA and urokinase plasminogen activator (uPA, PLAU), the main activators of plasminogen. Additionally, to EC it is expressed in a multitude of cell types, such as: megakaryocytes, smooth muscle cells, fibroblasts, monocytes, macrophages, adipocytes, and others (71, 92).

1.1.8 The Coagulation Cascade
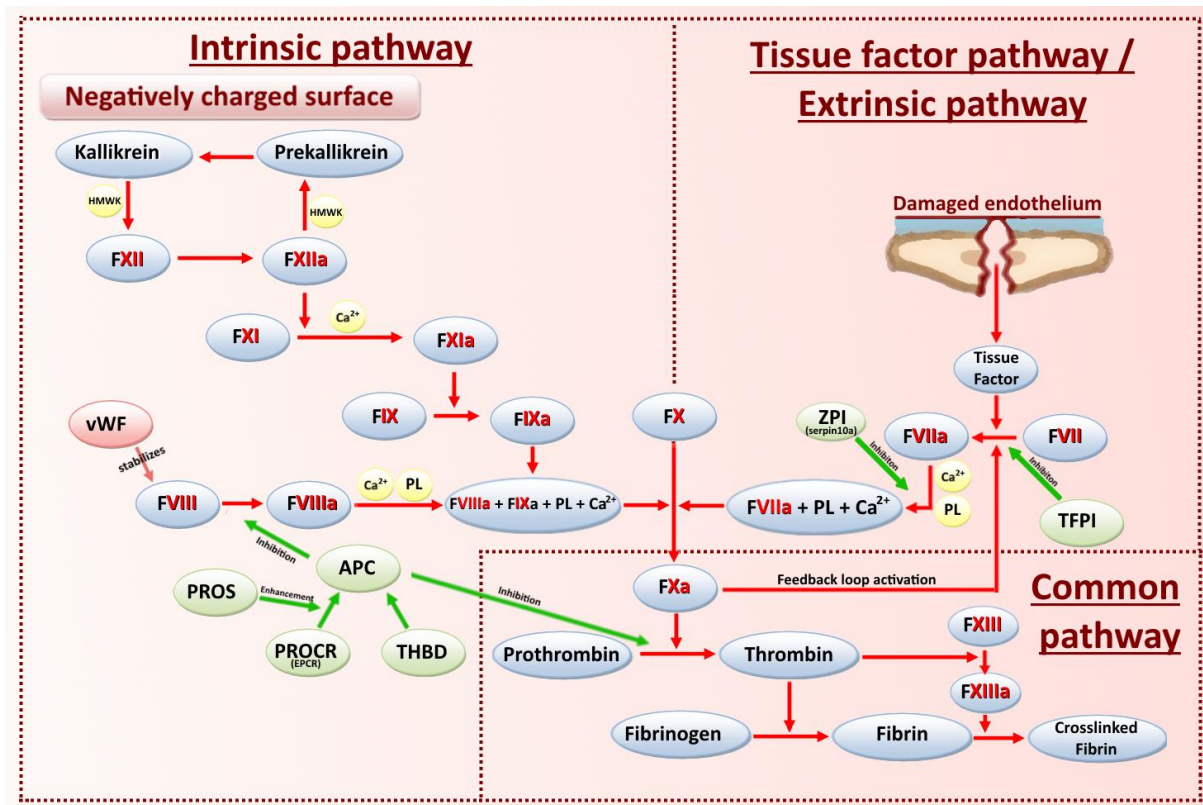
Figure 11: The coagulation cascade and role of the endothelium in coagulation: *The coagulation cascade is a series of protein activations and interactions that lead to the formation of a clot of cross-linked fibrin. The cascade is divided into three pathways: the intrinsic, the extrinsic and the converged common pathway. The intrinsic coagulation cascade is activated through exposure to negatively charged molecules (e.g., bacteria, lipids) and autoactivation of FXII on their surface. This activates prekallikrein, stabilised by high molecular weight kininogen (the contact activation system), followed by the activation of more factors, such as XII, XI, IX and a common activation of FX by FVIIa, phospholipids, calcium and FIXa. The extrinsic pathway (tissue factor pathway) is triggered by damaged endothelium and release of tissue factor, which activates FVII to FVIIa and induces FX activation in presence of phospholipids and calcium. Factor V acts in the intrinsic pathway and forms a prothrombinase complex with factor X. Both pathways converge in the common pathway, where prothrombin is converted to thrombin by FXa, simultaneously activating FXIII to FXIIIa. Thrombin converts fibrinogen to fibrin which is crosslinked by FXIIIa to stabilise the clot. Negative regulators of the extrinsic pathway are ZPI and TFPI, with TFPI inhibiting TF-FVIIa complex activation (adapted from Palta et al. (2014) and Adams et al. (2009) (93, 94)).*

Coagulation, or blood clotting, is the process of stopping bleeding by forming a stable blood clot. A damaged blood vessel induces vasoconstriction to limit the blood flow, activates platelets that adhere to the site of injury and form a temporary plug, followed by activation of the coagulation cascade (95). The coagulation cascade is a series of sequential enzymatic activations in which proenzymes and profactors are converted to their active forms (96, 97).

The coagulation cascade is divided into two pathways, the intrinsic and the extrinsic, that merge up to form a common pathway (Figure 11). Each of them consists of several factors that circulate the blood in inactive form. Activation of the intrinsic pathway occurs when endothelial collagen is exposed, while the extrinsic pathway is initiated after tissue factor release in response to external injury. Both pathways result in a burst of thrombin which in return activates fibrin, which is converted to a stable network, forming a blood clot. The clot acts as protective barrier over the injury site until healing completes. While intrinsic activation plays a significant role in clot formation *in vitro*, it seems to have no impact *in vivo*, as FXII deficient mice were shown to have regular haemostatic ability, albeit capability for thrombus formation was impaired (96, 98).

The extrinsic pathway of coagulation (tissue factor pathway)

The extrinsic pathway is initiated by impairment of the endothelium, e.g., from physical injury like cuts or penetration by sharp objects, as well as from the collapse of blood vessels during severe cases of sepsis (96). Tissue factor (FIII) is the primary initiator of the extrinsic pathway of coagulation and one of the key players in coagulation (91). Following vessel injury, FIII complexes and activates coagulation factor VII (FVII) to factor VIIa (FVIIa), which in return activates factor X (FX) to factor Xa (FXa) (96). While cells exposed to the blood flow normally do not express FIII on their surface to prevent undesirable clotting, this protein is found on a multitude of extravascular cell surfaces, including smooth muscle cells (99, 100). Endothelial cells have been shown to express FIII upon activation which can result in the development of blood clots (5).

The intrinsic pathway of coagulation (contact pathway)

The intrinsic pathway of coagulation is initiated by internal damage to the blood vessel and exposure of collagen by different factors, such as inflammation or trauma (96). Factor XII is activated to FXIIa as part of a mechanism that includes high-molecular-weight kininogen and plasma prekallikrein. FXIIa activates FXI to FXIa. FXI activates FIX to FIXa, forming the intrinsic tenase complex (consisting of FVIII, FIXa and Ca^{2+}). The tenase complex activates FX to FXa (96).

The common pathway of coagulation

Both pathways lead to factor X activation. FXa converts prothrombin (FII) to thrombin, which in return converts fibrinogen to fibrin. Fibrin is crosslinked by FXIIIa which is activated from FXIII by thrombin. Crosslinked fibrin then forms a stable clot (96).

14

1.1.9 Fibrinolysis – The Degradation of Blood Clots

Breakdown of blood clots is a crucial part in wound healing through a process called fibrinolysis. Fibrinolysis is controlled by plasmin, an enzyme that degrades blood clots at various places (101). This leads to the circulation of fibrin degradation products, such as D-dimer, an important marker for the presence of thrombosis (102).
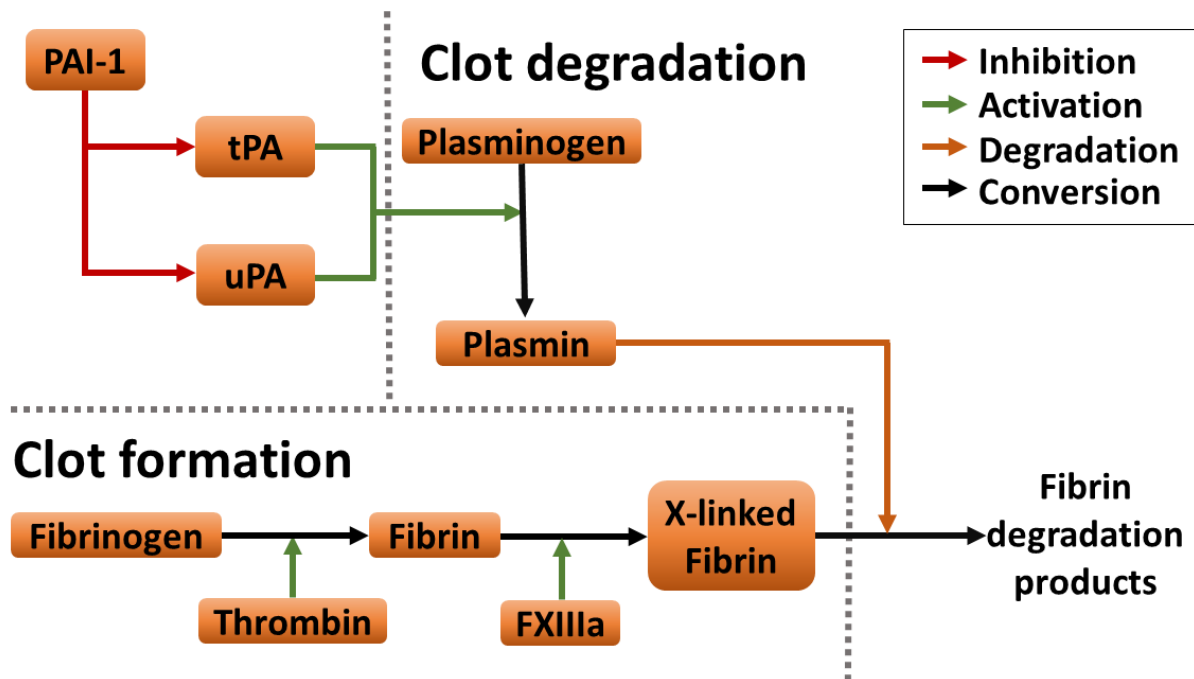
Figure 12: Principle of fibrinolysis: Cross linked fibrin is degraded by plasmin. Plasmin is activated through tissue-type plasminogen (tPA) activator and urokinase plasminogen activator (uPA). Inactivation of these through PAI-1 inhibit the downstream activation of plasminogen to plasmin (adapted from Meltzer et al. (2009) (103)).

Plasmin is mainly produced in the liver and circulates the blood in its inactive form plasminogen (104). The conversion of plasminogen to active plasmin, which enables the process of fibrinolysis, is facilitated by tissue plasminogen activator (t-PA) and urokinase plasminogen activator (uPA) (103). PAI1 protects, amongst other factors, against downstream plasmin activation (71) (Figure 12).

Fibrinolysis and the coagulation are intertwined processes that intricately maintain haemostasis. The initiation of fibrinolysis starts simultaneously with the coagulation system, working in tandem to control the development of blood clots. This ensures the fluidity of blood, preventing undesired clots, while simultaneously protecting against blood loss (103).

1.2 The Endothelial-enriched Transcriptome

The endothelial-enriched transcriptome describes all genes that display a high EC-specificity. It was only vaguely defined until Butler, Hallstrom (20) (2016) developed a then novel method to identify a set of pan-endothelial-enriched genes by profiling 124 body-wide tissue samples from 32 different organs using tissue transcript profiling of existing RNAseq data. It assumes that the number of EC specific transcripts relates to the amount of EC (Figure 13). While there were endothelial transcriptomes available prior to this approach, Butler's approach was distinguished by utilising whole-body transcriptome, instead of relying on microarray data from isolated cells.
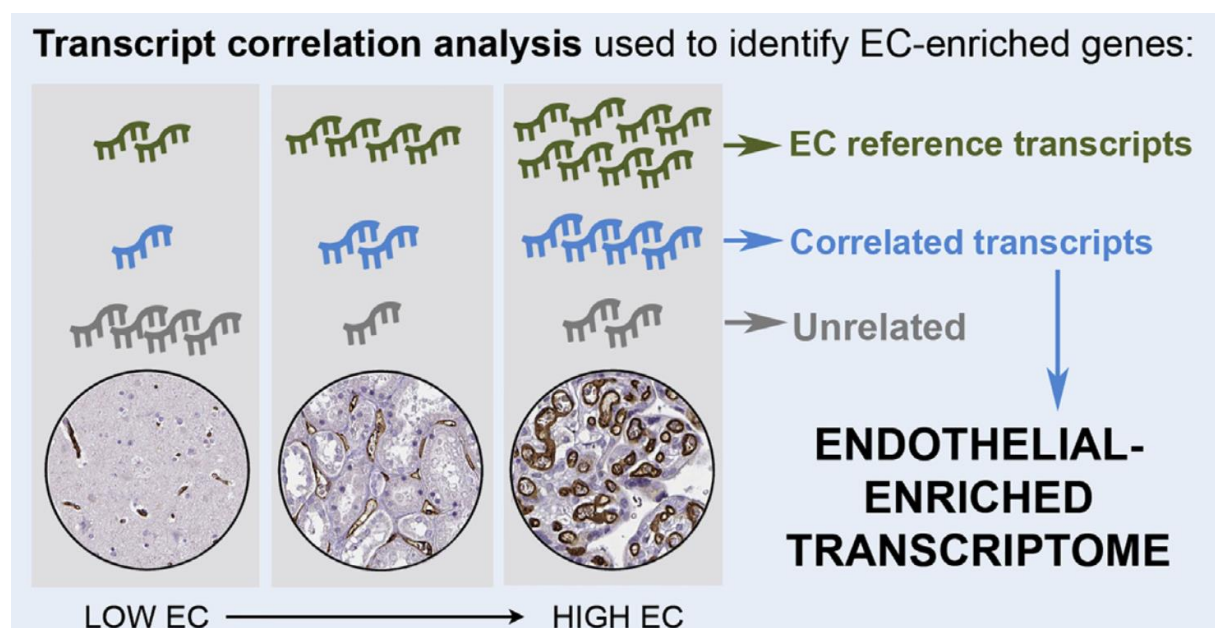
Figure 13: Definition of the endothelial-enriched transcriptome using a correlation-based approach: *Known endothelial specific reference transcripts were used to compare against the protein coding transcriptome. Genes with similar correlation behaviour over the whole sample set were defined as endothelial-enriched. Figure from Butler, Hallstrom (20) (2016).*

Three endothelial cell specific transcripts (vWF, CD34 and CLEC14A) were used as combined surrogate markers ("reference markers") to create an in vivo endothelial enrichment score by comparing correlation coefficients to 20.073 mapped protein coding genes. 234 transcripts were defined as endothelial-enriched. 88 of those gene transcripts were not previously reported as endothelial expressed, whereas 30 were completely uncharacterised proteins. One of these uncharacterised proteins is KANK3, the target of project 3 (20).

1.2.1 Adipose Tissue Cell Type Enriched Transcriptome

The adipose tissue cell type enriched transcriptome is presented in study 2.

However, the endothelial enriched transcriptome defined by Butler *et al. (2016)* has certain limitations. It exclusively focused on endothelial cells and did not categorise transcripts based on tissue or sub-tissue location. It lacked definition for other cell types. It utilised a now outdated version of the GTEx resource. Furthermore, it incorporated a smaller sample size in comparison to the more recent transcript atlases established by our group (105, 106, 107). One of the follow up projects, addressing these issues, was the adipose tissue enriched transcriptome described in study 2 of this thesis (105).

The adipose tissue (AT) enriched transcriptome describes the specificity of cell-type transcripts within adipose tissue (Figure 14). It was generated by Norreen-Thorsen et al. (2022) (105) using adipose tissue bulk RNAseq data from 527 samples of human visceral and 646 samples of subcutaneous adipose tissue from the Genotype Expression (GTEx) web resource (108).

AT comprises a variety of cell types, including adipocytes, adipocyte progenitor cells, endothelial cells, fibroblasts, smooth muscle cells, stromal cells, as well as immune cells such as macrophages and T cells (105, 109). Ten cell types were profiled using an integrative correlation approach, based on the one previously defined by Butler et al. (2016) (20). Using a correlative method of full tissue sequencing, challenges regarding single cell RNAseq analysis, were bypassed and more than 2,300 cell-type-enriched transcripts could be identified. Among those were 157 EC enriched transcripts in visceral and 155 EC enriched transcripts in subcutaneous adipose tissue.
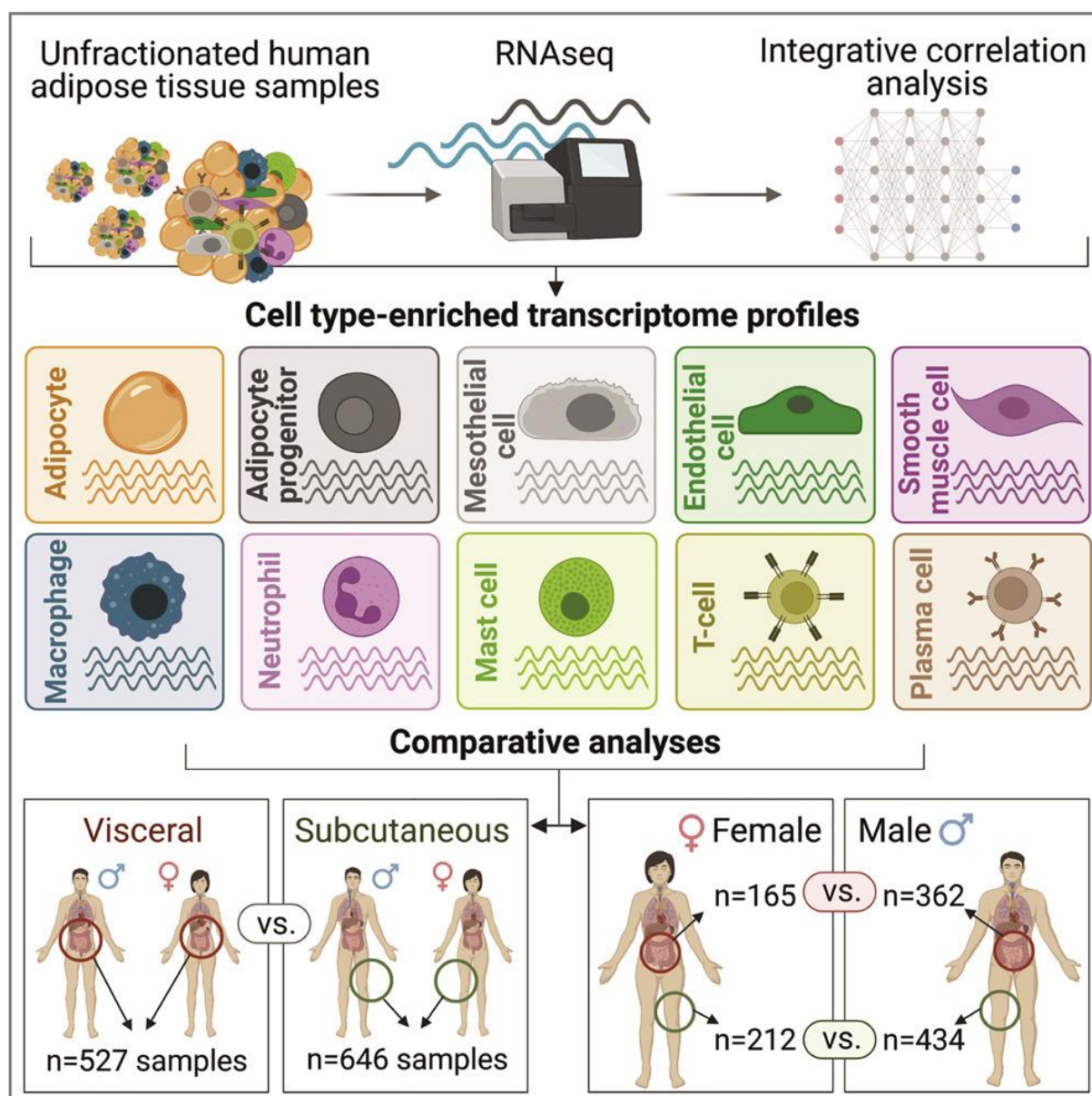
Figure 14: Analysis and definition of the adipose tissue transcriptome: *Unfractionated human adipose tissue RNAseq transcript levels (TPM) from visceral and subcutaneous adipose tissue were retrieved from the GTEx database. Cell-type enriched transcriptome profiles for 10 different cell types were calculated, using a correlative analysis base on three cell type specific markers per cell type, followed by tissue- and sex-specific comparison (from Norreen-Thorsen et al. (2022) (105)).*

While single cell RNAsequencing studies have been used for identification of adipose tissue cell specific transcriptomics (110, 111, 112), challenges with this method exist with regard to large cell types, cells with high buoyancy, or multinucleated cells (113), leading to the absence of adipocyte data from several single cell RNAseq datasets, including but not limited to *tabula muris* and *tabula sapiens* (114, 115, 116). These problems are partially addressed by single nuclei sequencing, which uses isolation of

single nuclei instead of cells, but it is important to note that transcript expression can vary between nuclei and cytoplasm (117).

1.2.2 Adipose Tissue

Adipose tissue (AT), also known as body fat, is a connective tissue type, which is mainly composed of adipocytes. It acts as energy depot, mainly in form of lipids, regulates metabolism and provides insulation (118). The earliest recorded mention of adipose tissue in published literature can be traced back to 1837, with subsequent mentions appearing until the 1940s. Long thought of as merely an energy depot and insulation, adipose tissue remained understudied for decades (119, 120).
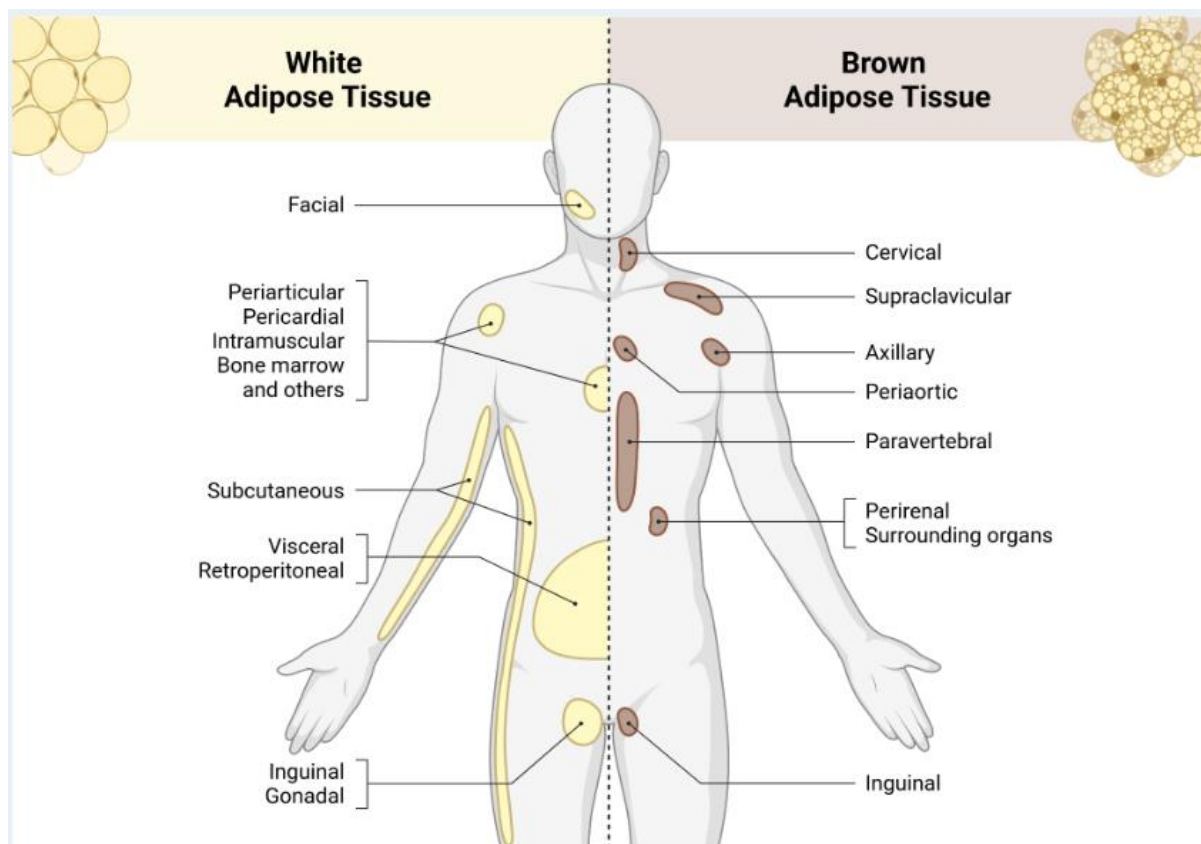
Figure 15: Distribution of white (WAT) and brown adipose tissue (BAT): *WAT is the main deposit of body fat, making up 95% of adipose tissue. WAT is widely distributed across the body, mostly within subcutaneous regions (beneath the skin), and visceral areas (around organs). BAT is predominantly localised in specific body regions such as the cervical, supraclavicular, axillary, periaortic, paravertebral, inguinal and perirenal regions surrounding organs. The two distinct main types, based on physical distribution, are visceral and subcutaneous adipose tissue. Visceral adipose tissue is located around organs, while subcutaneous adipose tissue is located under the skin (Figure adapted from BioRender).*

Adipose tissue is broadly characterised by physiological location into visceral adipose tissue (VAT), which is located in the abdomen and around organs, and subcutaneous

adipose tissue (SAT), which is located beneath the skin or by histology into beige, brown, and white adipose tissue (Figure 15) (121, 122, 123).

White adipose tissue (WAT) is the main type of adipose tissue, accounting for over 95% of the total adipose mass, while brown adipose tissue comprises of 1-2% of the overall body fat content. Beige adipose tissue comprises cells within WAT that have the capability to transform into brown-like adipocytes upon exposure to cold temperatures or body stress response (118).

There are notable distinctions in terms of secretory, morphological, and metabolic aspects between WAT and BAT: WAT insulates organs and stores energy, while BAT transforms energy into heat which dissipates into the body and acts as regulator of thermogenesis by burning calories. Therefore, BAT is highly vascularised and maintains a high number of mitochondria, as well as multilocular lipid droplets, displaying a high rate of fatty acid and glucose uptake and oxidation (122).

Excess VAT is involved in the pathogenesis of several diseases, including metabolic disorders, obesity, cardiovascular disease, and cancer (124). In addition to the occurrence of obesity, the distribution of body fat is related to variances in the risk of cardiometabolic disease and diabetes. Adipose tissue distribution is an important factor concerning the development of these co-morbidities (125).

This distribution is sex-linked: Women tend to have a higher occurrence of the "pear shape" fat distribution, while men have a higher prevalence of the "apple shape", characterised by a central fat distribution. The apple shape is associated with an increased risk of cardiometabolic disease (126).

1.3 The Endothelial-enriched Transcript KANK3
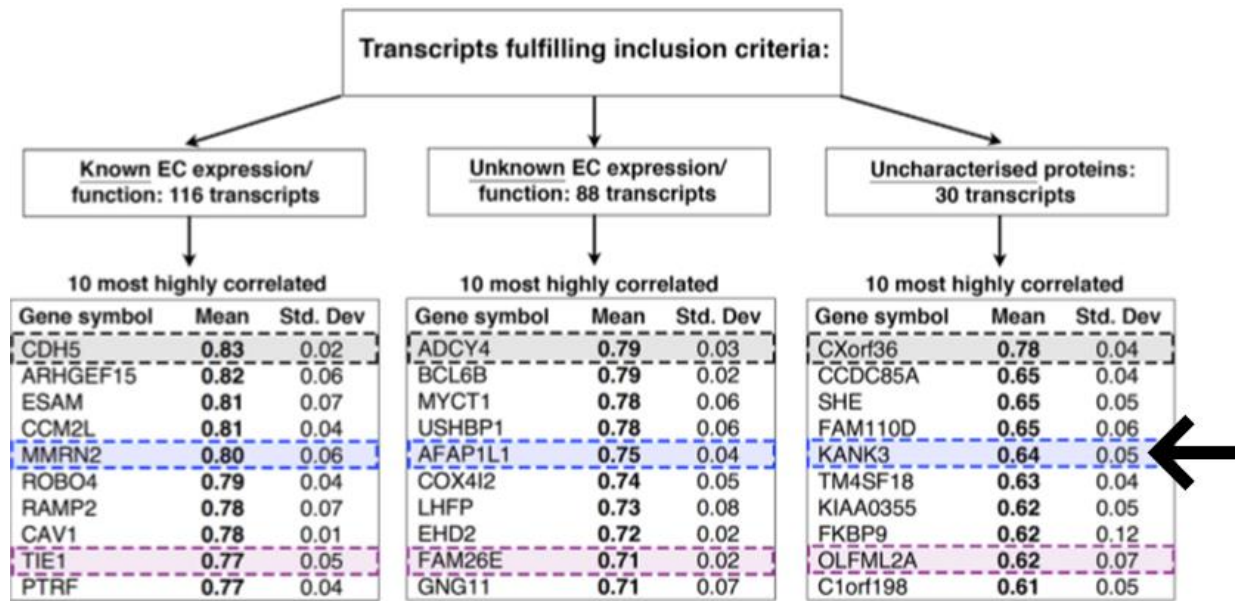
Figure 16: Endothelial-enriched genes defined by Butler et al. (2016): *Endothelial-enriched genes were sorted into three categories based on previous literature research. 116 Transcripts were defined with known EC expression, 88 transcripts with unknown EC expression or function and 30 were previously completely uncharacterised proteins. KANK3, the protein of interest of study 3, was found to be one of the highest uncharacterised proteins correlating to endothelial markers. Figure from Butler, Hallstrom (20) (2006).*

Butler et al. (2016) (20) identified the protein KANK3 as one of the most highly correlating endothelial-enriched transcripts, that had not been previously described (Figure 16). Transcript enrichment found for a tissue or cell type may suggest specialised function. Furthermore, since this initial study KANK3 has been consistently described as endothelial-enriched in multiple tissues by Dusart et al (2023) (107) in human brain (106), by Öling et al (2022) in stomach (127) and **by Norreen-Thorsen et al. (2022) in adipose tissue endothelium (105)**, indicating important endothelial and vascular specific functions.

KANK3 has been found present in vascular and lymphatic endothelial cells in human lung, pancreas, and testis (128). The human protein atlas profiled multiple immortalised cell lines. KANK3 is mainly expressed in the endothelial cell lines TIME and HUVEC/TERT2 (129). Numb-Binding Protein (NBP), a zebrafish homologue of KANK3, has furthermore been identified in vasculature and has been described as essential for epidermal integrity and embryonic transformation of the neural plate into the neural tube (130).

KANK3 is a member of the KN motif and ankyrin repeat domain-containing (KANK) family. The protein family of four members is evolutionary well conserved and defined by their common structural elements: The N-terminal KN-motif, the N-terminal and central coiled-coil domains and C-terminal ankyrin repeat units (131, 132) (Figure 17). While other members of the family have been previously described, the endothelial-enriched protein KANK3 remains largely unexplored.

KN-motif
(pos. 33-74)

Coiled coil domain
(pos 181-230, 367-404)

Ankyrin repeats
(pos. 622-652, 656-690,
695-724, 728-758, 762-795)
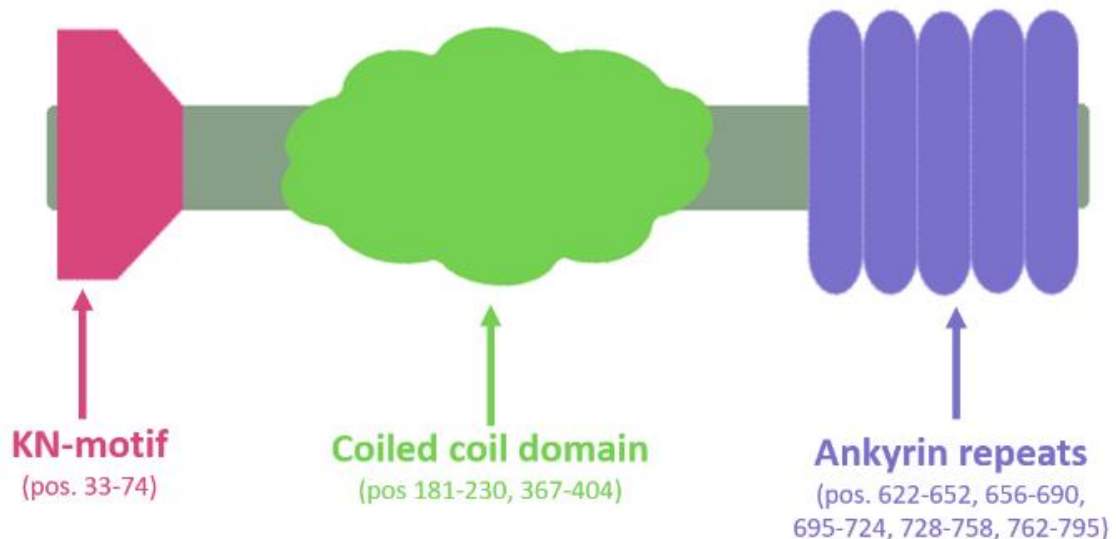
Figure 17: General structure of KANK proteins: The KANK family is evolutionary conserved with 4 paralogues in vertebrates. Their unique structure consists of three central elements: The KN-motif in the N-terminal site, which acts a as a talin binding domain, the coiled-coiled domains which binds to scaffolding proteins (liprins) and forms stabilisation complexes, and the ankyrin repeat units, which act as microtubule binding sites through KIF21A (adapted from Kakinuma et al. (2009) (131)).

The KN-motif has been shown to interact with talin, a focal adhesion protein. The coiled-coil domain interacts with liprins, a scaffolding protein mediating cell integrity. The ankyrin repeat units recognise the kinesin-like motor protein KIF21A, a protein associated with microtubules and intracellular transport (133, 134, 135). The N-terminal and central coiled-coil domains of KANK proteins have been shown to mediate interactions of KANK proteins with other proteins, such as liprin beta-1, in melanoma (136).

KANK1 and KANK2 have been recognised to be involved in cytoskeletal reorganisation mediated by binding of focal adhesions and stabilisation complexes. KANK1, 2 and 3 have been described to be involved in the regulation of cell migration, suppression of proliferation (137, 138, 139), as well as potential targets for tumour suppression, in renal cell carcinoma and lung adenocarcinoma. KANK3 expression is linked to lung

adenocarcinoma prognosis, with silencing of KANK3 leading to enhanced ability of cancer cells to migrate and proliferate (137, 140). Except for these cancer-focussed studies, the function of KANK3 remains unexplored in a vascular context.

Transcript ID	Name	bp	Protein
ENST00000330915.7	KANK3-201	2787	821aa
ENST00000593649.5	KANK3-203	2672	840aa
ENST00000595639.1	KANK3-204	506	146aa
ENST00000593331.1	KANK3-202	590	No protein

Figure 18: KANK3 splicing variants: KANK3 has four splicing variants. Three of them are proteinogenic and one is non-coding. Variant KANK3-201, give rise to 821 amino acids (aa) and is the most prevalent variant and therefore the one mainly explored in study 3 of this thesis. The protein product of variant KANK3-204 lacks most of the structural components of KANK family members, due to its short nature, while the protein product of variant KANK3-203 is structurally most similar to the KANK3-201 protein (data from ENSEMBL 110 (2023) (141)).

The molecular context of other family members, in particular KANK1 and 2, along with the discoveries in cancer tissue, suggests that KANK3 might play a significant role in cell migration, proliferation, and cytoskeletal organisation. Furthermore, the high endothelial enrichment observed across multiple tissues and cell types raises the possibility that KANK3 may also be involved in other vascular functions, such as haemostasis, inflammation, or mediation of vascular tone.

KANK3 has four known splicing variants (Figure 18). Three of them are protein coding, one is non-coding (KANK3-202). One variant (KANK3-204) is 146aa in length, and therefore missing most KANK specific domains and two are essentially similar in length and structure (KANK3-201, KANK3-203) (141).

1.3.1 Focal Adhesions

Focal adhesions (FA) are dynamic structures composed of more than hundred different proteins. They regulate cellular responses to mechanical signals from the extracellular matrix. FA serve as active sites for actin polymerisation and form complexes between actin bundles and integrins, which connect to the ECM (Figure 19). FA cluster within the cell and undergo continuous association and dissociation in response to signals. FA dynamically change in size and composition, adapting to mechanical stress (142).
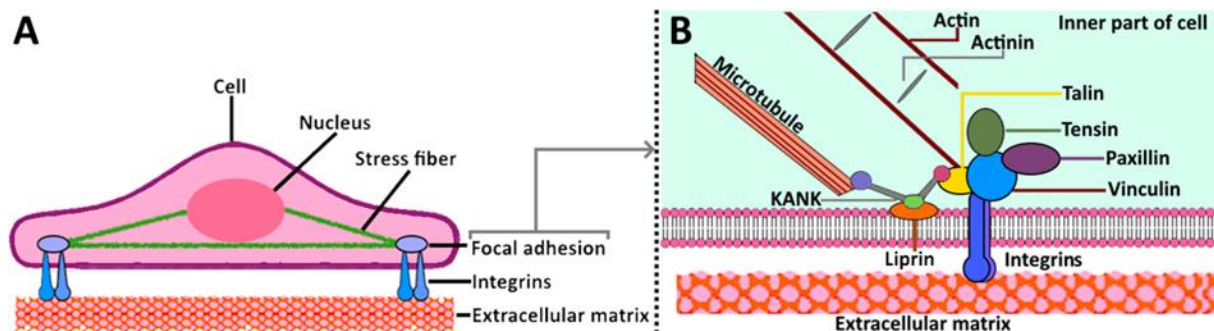
Figure 19: Localisation and structure of focal adhesions *(A) Focal adhesion are structures proximal to the extracellular matrix and intricately connected to integrins and stress fibres (143) (B) Integrin dimers are bound to the extracellular matrix, sense and transmit mechanical signals to the multi-protein focal adhesion complexes which in return are associated to the cytoskeleton. Notably, KANK proteins have binding sites for microtubules, liprins and talin. Figure partially adapted from Yoon (2019) (144) and Yu et al. (2022) (145).*

FA provide structural support to the cell, reorganise the cytoskeleton, transmit signals to the nucleus, and are controllers of cell adhesion, migration, and proliferation (142). The central proteins involved in focal adhesions include:

- **Integrins** are a principal component of FAs. They are transmembrane receptors that mediate cell adhesion to the extracellular matrix. Integrins consist of alpha (α) and beta (β) subunits. These subunits can combine in numerous ways to form integrin dimers and are able to bind to a variety of molecules, including glycoproteins, fibronectins, laminins, and collagen (146, 147).
- **Vinculin (VCL)** is a membrane-cytoskeletal protein that is enriched in focal adhesion and adherens junctions. It is a key player in the linkage of integrins to the actin cytoskeleton. VCL has binding sites for F-actin, paxillin, and lipids and is conformationally regulated and recruited to FA after activation (148).
- **Talin (TLN)** are focal adhesion proteins that bind to the integrin beta subunits and can activate integrins by coupling them to the cytoskeleton, and therefore regulate integrin signalling. Talin has a vinculin binding site (149).
- **Paxillin (PXN)** is a 68 kDa protein that gets recruited to focal adhesions during the initial stages of FA nascency. These proteins primarily serve as scaffold and adaptor proteins, functioning as linkers within the cytoskeleton. They act as targets for tyrosine kinases after to integrin signalling (150).

24

- **Focal adhesion kinases (FAK)** are a group cytoplasmic kinases that facilitate communication between integrins and intracellular proteins, influencing cell polarity, adhesion, migration, and invasion and strength of FA (146).

- **Zyxins (ZYX)** are a group of zinc-binding phosphoprotein which are enriched in focal adhesions. These mechanosensitive regulators of cytoskeletal dynamics bind alpha-actinins and target TES and ENA/VASP (151, 152).

- **ENA/Vasodilator-stimulated phosphoprotein (ENA/VASP)** is a group of focal adhesion proteins that binds to vinculin and zyxin. VASP recruits the G-actin binding profilin proteins and acts as a binding intermediate to focal adhesions and profilin. It is likely involved in actin filament assembly and cell motility (153)

- **α-actinins** are proteins that crosslink F-actin and anchor it to a multitude of intracellular structures, to create a scaffolding matrix and regulate cellular stability. Actinins regulate cell motility through the binding of myosin and actin. They play a key role in the development and maturation of focal adhesions, by linking actin to integrins and creating a cytoskeletal scaffold (154).

- **Proline-rich tyrosine kinase 2 (PYK2)** is a tyrosine kinase that regulates focal adhesions. After response to cell adhesion and ECM proteins, it is transported to focal adhesions. PYK2 promotes cell proliferation and formation of focal adhesions. PYK2 has been shown to compensate for loss of FAK (155).

- **Rho family GTPases** (including Rho, Rac, and Cdc42) are molecular switches that, among other functions, control signalling pathways that regulate the assembly or disassembly of actin and focal adhesions (155).

- **Phosphatidylinositol 3-kinase (PI3K)** is a mediator of integrin-mediated migration. This lipid kinase is involved in the restructuring of focal adhesion plaques by binding to α-actin (156, 157).

The cytoskeleton and focal adhesions play important roles in a multitude of vascular functions. These include upholding of the structural integrity to withstand shear forces imposed by the blood flow (158), regulation of cell motility in angiogenesis, (159), stabilisation and reinforcment of EC-junctions, regulation of vascular permeability (160), and in haemostasis (161) and inflammation (162).

1.4 Gene Expression Profiling

Gene expression profiling is the analysis of expression levels of a gene within a tissue or organism by measuring RNA transcripts or their final products, such as proteins. It allows to determine which genes are active or inactive in certain cell types or after activation, e.g., by a stimulus or during cell cycle, proliferation, etc.

1.4.1 The Central Dogma of Molecular Biology

The central dogma of molecular biology, defined by Francis Crick in 1958, describes the one-directional flow of biological information. It suggests that DNA contains all information to make proteins and is often shortened to "DNA → RNA → Protein". The dogma describes that DNA is transcribed into RNAs, which are then translated into an amino acid chain that is folded into functional proteins (163, 164) (Figure 20).
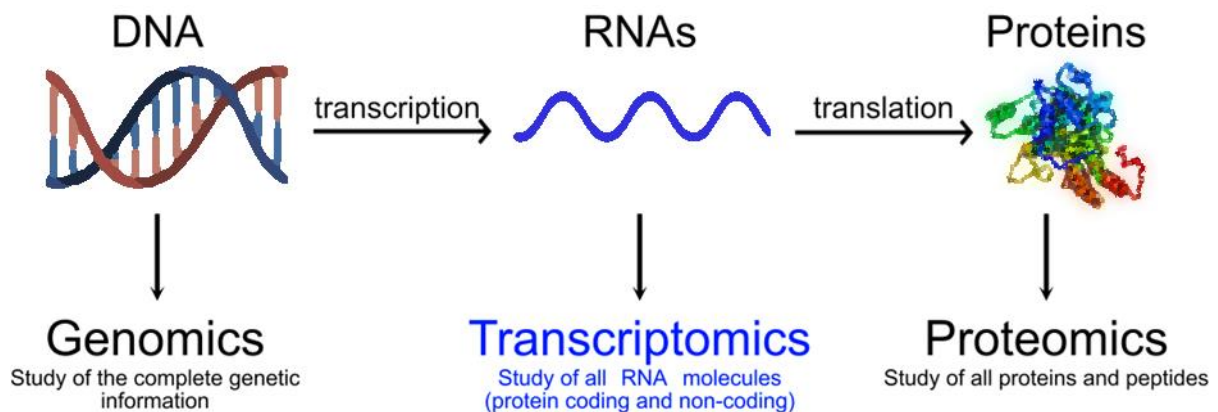
Figure 20: Overview of differences between genomics, transcriptomics, and proteomics: Genomics, transcriptomics, and proteomics are the three distinct disciplines of molecular biology. They centre on different facets of biological molecules. Genomics focus on the genetic information within DNA. Transcriptomics focus on the different RNA molecules which are transcribed and spliced from the DNA. Proteomics is the study all protein products and peptides.

Crick's simplified model is still mostly valid today, although it is now known that not all RNAs encode proteins. It also was not originally accounting for reverse transcription, the synthesis of complementary DNA (cDNA) from an RNA template. Earlier, non-protein-coding DNA elements were often referred to as "junk DNA" and thought to be non-functional (165, 166). Today it is well established that most RNA molecules (>80%) are non-protein coding gene regulatory elements, such as siRNAs, miRNAs, circular RNAs, pseudogenes or long non coding RNAs (167, 168, 169), while other RNA elements are mainly involved in protein biosynthesis (mRNA, tRNA, rRNA) (170).

1.4.2 The Fate of a Cell: From Genome to Transcriptome to Proteome

At the time of writing, the current human reference genome (GRCh38.p14) was released by the Genome Reference Consortium in May 2022. It has been expanded since its beginning around the year 2000 and contains nearly 20,000 protein coding genes, and 233,615 total splicing variants (86,245 of them protein coding) (171, 172).

The genome encompasses all DNA sequences in an organism, while the transcriptome refers to the complete set of transcripts in a cell at a certain time point. The genome is mostly stable, meaning each diploid cell contains the same DNA information (173). Even though all cells in a body share the same genome, their cellular phenotypes and functions differ depending on multiple factors, such as cell type, tissue, surrounding cells, intra-tissue location, as well as environmental factors, and sex. The cell identity is defined through a multitude of intra- and extra-cellular signals during development that aids in establishing a cell identity that is maintained throughout a cell life. Its role and morphology are defined by transcriptional regulation, which can be influenced by tissue structure, cell-cell and cell-matrix interactions, mechano-transduction, and chemical or physical cues from the microenvironment. These factors may be directly and indirectly impacting cellular phenotype and behavioural responses (174, 175).
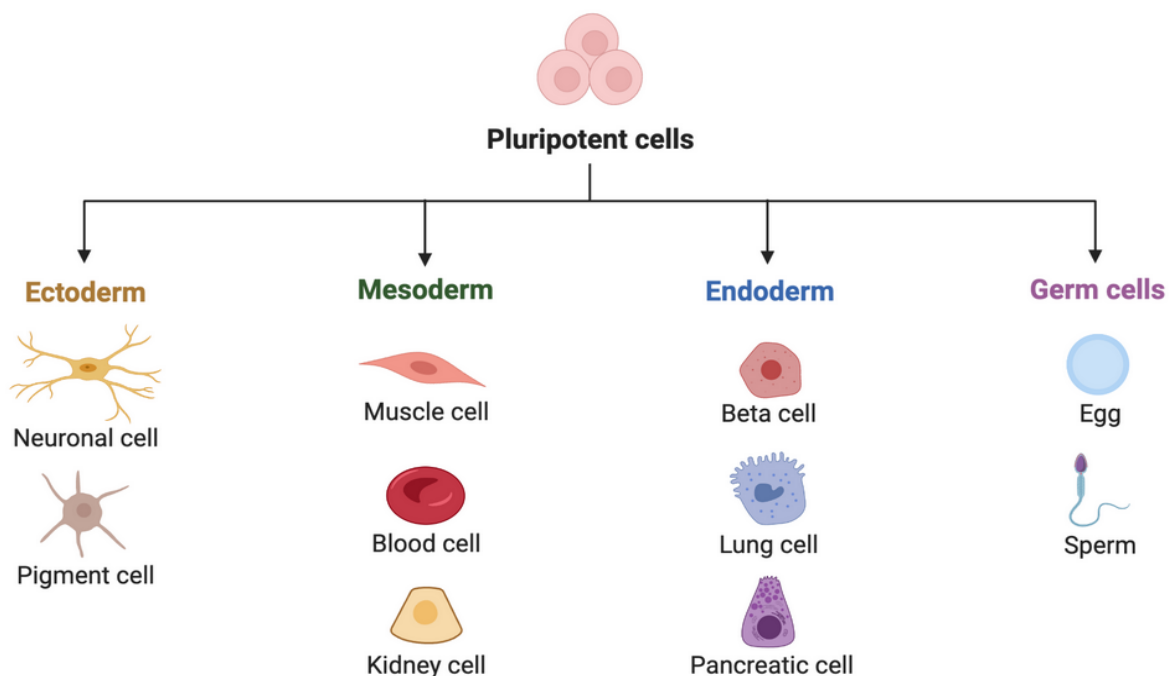
Figure 21: Derivation of ectoderm, mesoderm, endoderm, and germ cells from a common progenitor. Pluripotent cells generate diverse mature cell types, which further differentiate into tissue-specific subtypes crucial for organ and tissue development (from BioRender).

Cellular differentiation is the result of multiple factors. Pluripotent stem cells give rise to a multitude of mature cell types. These can branch into ectoderm (skin and nerves), mesoderm (bones and muscles), and endoderm (most of internal organs) before they differentiate further into tissue specific cell subtypes. They are important for the development of organs and tissues. Differentiation into a mature cell changes its type, function, and metabolic activity (Figure 21) (176, 177).

Cell types can be classified based on different factors, e.g., their function, source, location, shape, and transcriptome (106). Several transcripts exhibit tissue specificity or enrichment, indicating that their expression is distinct or elevated in certain tissues compared to others. These genes typically encode transcripts that are strongly associated with specific biological functions of their tissues. In instances where similar gene enrichment patterns are observed across multiple tissues, it is often due to shared functions among those tissues (178, 179, 180).

1.4.3 The Transcriptome

The transcriptome of a cell or organism encompasses the entirety of its RNA molecules, commonly known as transcripts. RNAs are less stable and more transient compared to other biomolecules. RNA is broadly categorised into coding or messenger RNA (mRNA) and non-coding RNA (ncRNA). Until recent times, it was common that text books referred to very few types of RNAs (rRNA, tRNA, mRNA) and did not acknowledge the others (181). There are different kinds of RNAs, which are involved in a multitude of processes, among others the protein coding ones (mRNA), the types involved in protein biosynthesis (mRNA, tRNA, rRNA), the types involved in nuclear RNA production (RNA precursors, snRNA, spliceosome), the types involved in RNA interference (siRNA, lncRNA, miRNA, piRNA) (170). Although the genome remains relatively stable, with variations such as mutations and epigenetic changes, the transcriptome exhibits high variability. Each cell type expresses a unique set of genes based on function, location, and tissue type, despite sharing the same genome. Furthermore, the cellular transcriptome undergoes additional variations influenced by factors such as the cell cycle stage, disease presence, exposure to drugs, and the aging process (182, 183, 184).

1.4.4 Transcript Types

All projects in this thesis rely on the analysis of gene transcripts, with many transcript types investigated in projects 1 and 2. Project 3 focuses on the analysis of transcripts of the KANK family. Transcript types shall therefore be explained in more detail here.

RNA can be broadly categorised as either protein coding or non-coding, serving various specialised roles depending on their subtype.

Protein coding RNAs are transcripts that encode for proteins or peptides.

- **Messenger RNA (mRNA)** is a single stranded type of RNA, which contains a genetic template for protein synthesis. mRNA is carried from the nucleus of a cell to the ribosomes in the cytoplasm. The mRNA is then translated to proteins in the ribosome by tRNAs. Although of critical biological function, mRNA makes up only about 1–5% of the RNA in a cell. mRNA are the product of pre-mRNAs post splicing, ligation of exons, capping and polyadenylation. (185, 186, 187).

- **Messenger RNA precursors (pre-mRNAs)** describes the intermediate product of mRNA before it is converted into the primary target of transcription. It contains non-coding regions (introns), which are removed, and coding regions (exons) which are reassembled and spliced together, followed by 5'-end capping, 3'-end cleavage, and polyadenylation. Pre-mRNA can be generated within seconds, while splicing takes longer (188, 189).

Non-coding RNAs (ncRNAs) are transcripts that do not encode for a protein. They can be regulatory elements and are directly involved in protein biosynthesis. Furthermore, in a recent pre-print, it has been suggested that ncRNAs are key players driving animal lifespan evolution (190).

- **Ribosomal RNA (rRNAs)** are involved in the formation and function of the ribosome. The large subunit of the ribosome is composed of 5S, 5.8S, and 28S RNA and complexed with numerous proteins. rRNA catalyses peptide bond formation between two amino acids, leading to the formation of proteins. rRNA is the most abundant type of RNA within the cell, making up 80% of RNA (170).

- **Transfer RNA (tRNA)** is a specialised molecule that helps translate the genetic information encoded in mRNA into proteins. It has a specific three-dimensional structure utilising anti-codons to recognise and pair with codons on the mRNA. The

flexibility of the pairing, guided by modifications in the tRNA, allows for accurate and efficient translation of the genetic code (170).

- **Small/short interfering RNAs (siRNA)** are short double stranded RNAs, 20-24 base pairs in length, which play roles in gene silencing through RNA interference (RNAi). They are produced through cleavage of long double-stranded RNA by the enzyme DICER, which cleaves long fragments into short siRNA. It is then complexed by the RNA-induced silencing complex (RISC), which uses the structure of the siRNA for recognition of a complementary target sequence to degrade (191).

- **MicroRNAs (miRNA)** are a well-studied RNA subtype, which has been shown as a regulatory RNA with many of its functions related to gene silencing. Similar to siRNAs they are processed by DICER to form a RISC. These small RNA molecules are 21-23 bp long and form negative regulatory networks (167, 170). They have been linked to several diseases, such as cardiovascular disease, immune diseases, respiratory disease, and diabetes, but have also shown protective effects, such as tumour suppressant (192, 193, 194, 195, 196).

- **Small nuclear RNAs (snRNAs)** have a significant function in the removal of introns from primary genomic transcripts through splicing. They are present in splicing speckles and Cajal bodies of the cell's nucleus and bind different proteins to form RNA-protein complexes, recognising them through base-pairing interactions or nucleotide-nucleotide contact between pre-mRNA and snRNA (197, 198, 199).

- **Small nucleolar RNAs (snoRNA)** are a group of small RNA molecules, found in the nucleolus. They are responsible for directing modifications, such as maturation, into various RNA molecules, particularly rRNA, tRNA and snRNA. C/D snoRNAs methylate target nucleotides, H/ACA snoRNAs modify uridine to pseudouridine (200).

- **circular RNAs (circRNA)** are single stranded RNAs, that form a closed loop that are covalently linked exons of a single pre-mRNA in a process called back-splicing. Their circular structure makes them resistant to degradation. circRNA perform various roles, including acting as decoys for microRNAs or RNA-binding proteins, thereby influencing gene expression or translation of proteins(201).

- **long-non-coding RNA (lncRNAs)** are RNAs that are longer than 200 nucleotides and transcribed from the non-coding regions of the genome. lncRNAs play significant roles as regulators of gene expression, operating at both the

transcriptional and post-transcriptional stages. Their functions are manifold, as they can modulate chromatin structure, gene expression. lncRNAs are able to interact with proteins, DNA and RNAs. They have been shown to be involved in multiple disease, such as neurodegenerative, cardiovascular, muscular, and immune disorders. Their potential as biomarkers has been well acknowledged (202, 203).

- **Piwi interacting RNAs (piRNA)** are recently discovered short non-coding RNAs, spanning 24 to 31 nucleotides. They interact with argonaut proteins from the piwi family and serve regulatory functions by modulating signalling pathways pre and post transcription. They are expressed across various human tissues (204).

- **Small activating RNAs (saRNA)** are a double-stranded RNAs of approximately 21 base pairs in length, similar in structure to siRNA. Unlike siRNA though, it enhances gene expression by an as yet unknown mechanism. Similar to siRNA it engages with argonaut proteins to establish RISC (205, 206).

- **Antisense RNA (asRNA)** is a single stranded RNA, which is complementary to mRNA and acts as suppressant by hybridising the complementary structure at multiple levels, such as replication, transcription, and translation (207).

- **Pseudogenes** are gene-like sequences that do not have a regulatory sequence needed for transcription, translation, or defective coding sequences. Previously labelled as junk-DNA. Four types are known: processed, duplicated, unitary, and retrotransposed. They have been shown to have some regulatory functions, with some of them processed into siRNAs (208, 209).

1.4.5 The Proteome

The proteome describes the complete set of proteins expressed by a cell, organ, or an organism at a given time point, including isoforms and post-translational variants. Similar to the transcriptome, the proteome is a functional unit of the genome. The proteome is a complex and dynamic unit, among other factors due to the presence of numerous proteoforms. Proteoforms are protein variations that lead to functional diversity. This happens through different processes, such as post-translational modifications, alternative splicing, and other modifications. Additionally, the expression levels of proteins exhibit significant dispersion to the transcriptome (210). On the other hand, only a small part of the transcriptome, the mRNA, gets translated to proteins while the other parts play regulatory functions. Therefore, neither the transcriptome nor the proteome are an accurate reflection of the other.

1.4.6 Advantages and Disadvantages of Transcriptomics and Proteomics

Transcriptomics and proteomics are the two central approaches to study biological data. Proteomics is a field within molecular biology focused on investigating proteins, encompassing their structure, function, interactions, and abundance within a specific biological system. Transcriptomics focusses on the transcriptome in a system at a certain time point and investigates all sets of RNAs.

Targeted approaches such as microarrays, qPCR, flow cytometry, western blot or immunostaining require prior knowledge on the potential target, while untargeted approaches such as mass spectrometry and RNAseq allow for a global profiling.

Transcriptomic approaches have advantages and disadvantages over proteomics:

- RNAsequencing is a sensitive and accurate tool for measuring transcript expression but mRNA expression is not a reliable indicator of protein levels and their activity. Therefore, it is hard to reliably predict the proteome from the transcriptome.
- Some regulatory elements, such as non-coding RNAs are missing in proteomic approaches.
- Transcriptomics reveal details about gene expression levels, but it does not offer insights into post-translational modifications (211).
- RNA sequencing methods are reliable and affordable for clinical settings, offering a high specificity and sequencing depth, while the depth of proteomics is much lower (212).
- RNAseq more cost-effective to mass-spectrometric approaches (213).

Proteomics and transcriptomic approaches can be used complementary to gain insight about the regulatory profile and the proteins of an organism. Analyses could gain substantial advantages through the employment of "cross-omics" strategies, the combination of different omics methods, such as proteomics, transcriptomics, lipidomics, glycomics, and metabolomics (214). When brought together, these distinct methods can complement each other to provide a better representation of the complex biological landscape. Notably, a recent pre-print with bacteria focus suggested that, both the proteome and the transcriptome modules consist of a similar roster of gene products and they used statistical modelling to infer the absolute allocation of the proteome from the transcriptome alone (215).

1.5 Web-Based Resources and Social Media in Biological Research

Biological research has been substantially changed by the adoption of web-based resources for biological data, which now openly facilitate the sharing of information. These resources enable the rapid and widespread availability of data to the scientific community, serving as a valuable tool for investigation and as a data repository.

Online resources include, among others:

- The Human Protein Atlas (HPA, https://www.proteinatlas.org/) encompasses a combination of imaging, transcriptomics and antibody-based proteomics approaches to map the full human proteome, as well as transcripts for tissue samples and cell lines (129, 179, 216), recently expanded using tissue specific transcriptomics data (105, 106, 107).
- The *tabula sapiens* is a human cell atlas across multiple organs, encompassing histology and single cell transcriptomics data (115).
- A similar approach for mouse organ data was created by *tabula muris* (114).
- GeneRanger offers access to processed data on gene and protein expression across normal human cells and tissues, serving as a valuable resource for target discovery (178).
- The Genotype-Tissue Expression (GTEx) project is a public resource to study tissue-specific gene expression, offering transcriptomics data for multiple organs from ~960 donors and over 30,000 samples (108).
- Important genomic databases include ENSEMBL, one of the biggest infrastructures to access gene data and annotation (141), and the human gene database GeneCards (217).

Study 1 of this thesis contributes to the multitude of web-based resources by offering a sex-matched, temporal transcriptional resource of the endothelial response to TNF and temporal baseline, available under http://www.endothelial-response.org (218). This resource allows for a multitude of analyses. Additionally, the results of study 2 are featured on the HPA, one of the world's major databases for biological research on https://www.proteinatlas.org/humanproteome/tissue+cell+type/adipose+tissue.

In addition to web-based resources, social networks such as X (formerly Twitter), Instagram, Facebook, Reddit, and ResearchGate play big roles in scientific communication, data availability, or sharing of presentations, scientific papers, and research data. Twitter, followed by Instagram and Facebook, is the most used social media platform to share and disseminate scientific research between peers and communicate it to the public (219, 220, 221), while ResearchGate serves the additional purposes of sharing scientific discoveries, asking for advice, discussing research, and fostering connections with peers (222). These platforms can help to facilitate world-wide collaborations and networking amongst researchers from different fields.

2 Aims of the Thesis

Global hypothesis: Cells from different tissues differ in their transcriptomic profile and tissue specific expression allows for cell or organ specific functions. Endothelial cells play a critical role in the response to various stimuli, and understanding the intricate mechanisms underlying their cellular responses is essential for advancing our knowledge of disease pathogenesis. Transcriptomics, with a focus on temporal resolution, response time, impact of sex and tissue specificity can be used to explore endothelial cell biology and tissue heterogeneity. Similar to the temporal transcriptomics in project 1, the tissue dependent transcriptomics in project 2 can expand the landscape of targets for diagnosis, disease prevention and treatment.

Project I: Global transcriptome of the endothelial response to tumour necrosis factor: The study aimed to investigate the temporal changes in gene expression patterns by subjecting endothelial cells to TNF stimulation over a time course ranging from 0.5 to 72 hours.

Project II: A human adipose tissue cell-type transcriptome atlas: The primary aim was to generate a comprehensive transcriptome atlas encompassing ten distinct cell types within human adipose tissue, utilizing publicly available bulk RNA sequencing data from subcutaneous and visceral fat samples. This was part of a larger series of related projects in the lab, with an overall aim to identify genes with endothelial enriched expression across tissue types.

Project III: Functional investigation of the previously undescribed endothelial-enriched transcript KANK3: The aim of this study was to investigate the function of the endothelial-enriched transcript KANK3 in a vascular context. Therefore, KANK3 function was assessed in multiple approaches testing multiple factors, ranging from wound healing, proliferation, haemostasis, and inflammation.

3 Methodology

The following sections aim to explain the theoretical framework of methods used in this this thesis, as well as the rationale and strength and weaknesses of the research approaches used in this thesis.

3.1 RNA Sequencing (RNAseq)

RNA sequencing (RNAseq) is a widely sensitive and highly accurate method to analyse the abundance and identity of gene transcripts. This method allows for detection of transcript isoforms and alternative splicing variants, and can indicate gene fusions, mutations, and single nucleotide polymorphisms (SNPs), coding- and non-coding RNAs and allow for sense or antisense strand determination (223, 224). It gives information over gene expression and is able to measure alternative splicing variants, post-transcriptional modifications, gene fusions and mutations/SNPs.

3.1.1 Bulk RNAseq
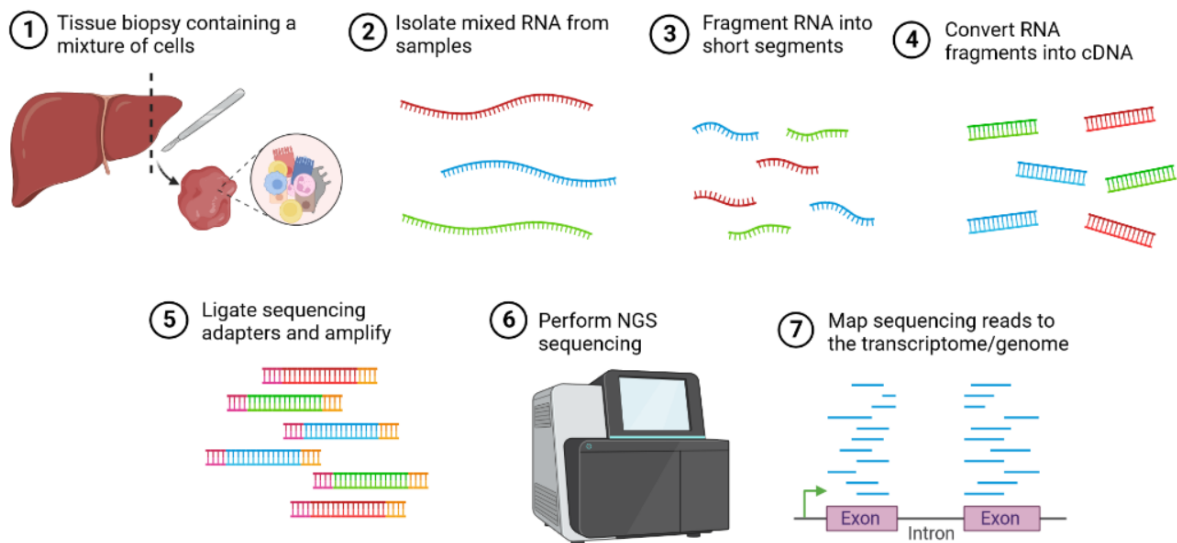
Figure 22: General principle of RNA sequencing: *General workflow of RNAseq, highlighting the crucial processes from sample collection to data analysis. (1) A tissue or cell samples are collected and (2) RNA molecules are isolated from the sample using various extraction methods, separating them from other cellular components (3) The isolated RNA is then fragmented into smaller pieces to facilitate downstream processing and (4) converted into cDNAs through reverse transcription, allowing for the amplification of the RNA sequence information. (5) Adapter sequences are ligated to the ends of the cDNA fragments. (6) The ligated cDNA fragments are sequenced using high-throughput technologies. This step generates a massive amount of short DNA sequences called "reads" or "read counts". (7) The reads are mapped and aligned to a reference transcriptome by matching them to known transcripts. Afterwards, the read counts are normalised to account for differences in sequencing depth and other technical factors (Created with BioRender.com.).*

Bulk RNAseq is a transcriptomic method that measures the average gene expression across a population of cells, which can be heterogenous (e.g. organ biopsies) or from a singular source (e.g. cultured cells) (225). RNAseq relies on libraries, the collections of RNA molecules that have been converted into complementary DNA (cDNA) fragments (Figure 22). It usually involves sequencing two types of libraries: mRNA-only library and whole transcriptome library that includes all RNA species except for rRNA. Two main approaches of RNAseq are "short-read" and "long-read" sequencing.

- **Short-read sequencing** is a common and powerful approach. It involves reading short fragments of RNA, usually ranging from 50 to 300 base pairs. Although it generates a high volume of reads, each individual read only covers a small section of the transcript. Single ended short-read sequencing is the most frequently used bulk RNAseq approach, often focused on differentially expressed genes.

- **Long-read sequencing** is especially useful for detecting complex RNA, as it captures much longer fragments of RNA, typically several thousand base pairs in length. Although it produces fewer reads compared to short-read sequencing, each read covers a significantly larger portion of the transcript. Thus, this technique can be used to detect alternative splicing variants. The main disadvantage of long-read sequencing are the higher costs and lower yields in read count compared to short-read sequencing. The low read count might also pose a challenge to conducting differential expression analyses (226, 227).

Bulk RNAseq includes a diverse set of different RNAs and is able to detect low abundance RNAs due to the high amount of sample used compared to single cell approaches (228). It relies on removal of the overly abundant rRNA (80-90% of total RNA in eukaryotic cells), to increase sequencing depth and detection limit. This can be performed by selectively removing rRNA, or selecting polyadenylated RNAs using poly(T) oligos that target the poly(A)-tail and remove all remaining RNAs, except for mRNA, lncRNA and pseudogenes (225).

Common workflow: Cells or tissues are lysed, and the RNA is isolated. RNA quality control is assessed through different methods, such as RNA integrity number (RIN) analysis, spectrophotometry, or electrophoresis. Enzymatic or chemical fragmentation of RNA is carried out. Reverse transcription and adaptor ligation for recognition follow. Complementary DNA (cDNA) is amplified in PCR, and the RNA is sequenced. The last

steps are data processing and analysis (229, 230). Most of these steps allow a multitude of different options. Common problems of bulk RNAseq include:

- **Limited resolution of cells:** Bulk RNAseq does not distinguish between cell types and their expression changes, as an averaged expression of the whole sample is measured (225).
- **Batch effects:** Technical variation between experiments and sequencing runs can lead to systematic differences that need to be adjusted for.
- **Low sensitivity for genes in lowly expressed cell types:** Bulk RNAseq could potentially not cover or underestimate genes that are only expressed by a small sub-population of cells, and could therefore be lost as "noise" (231). Furthermore, the amount of RNA is dependent on cell size. Bigger cells have been described to have more mRNA than smaller cells (232). Thus, differences in gene expression in smaller cells might be masked.

3.1.2 Single Cell RNAseq

Single cell sequencing is a commonly used technique in molecular biology, which enables the examination of gene transcript expression in individual cells.
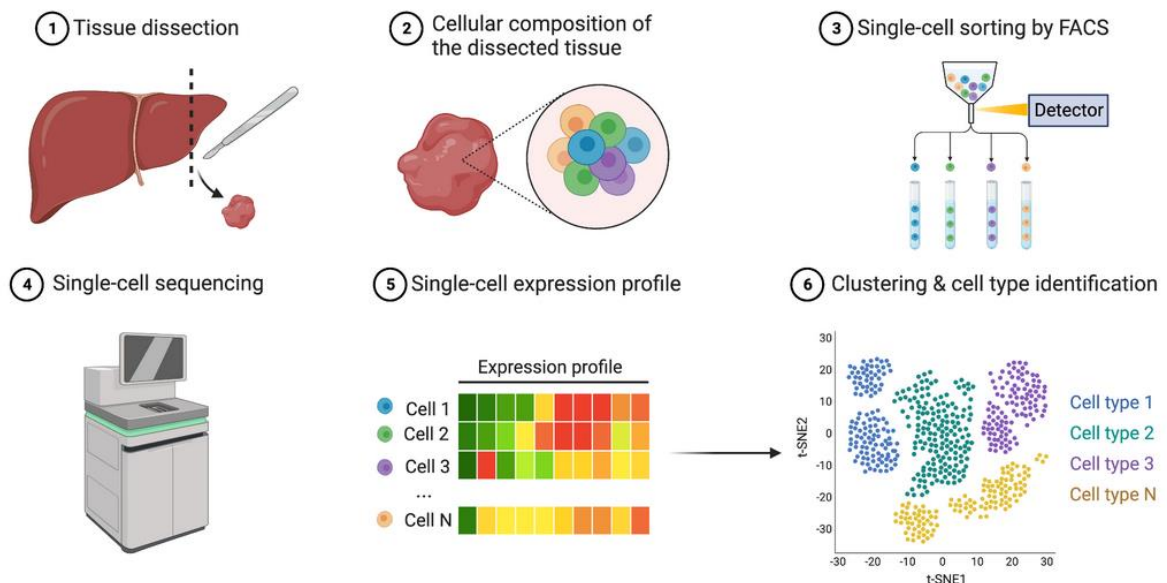
Figure 23: Principle of single cell RNAseq: *General workflow of RNAseq, highlighting important steps from sample collection to data analysis. (1) A tissue or cell samples are collected and (2) Cells are dissociated, extracted, and depending on isolation method marked, (3) and sorted. (4) Sequencing is performed after library preparation, followed by bioinformatic analyses (5) single cell expression profiling, (6) and cell clustering (Created with BioRender.com).*

Common workflow: The workflow is similar to bulk sequencing but differs in sample preparation, as cells must be separated to be analysed. Tissue is dissociated and single cells are suspended into single cell suspension droplets. Viability is usually assessed for quality control and dead cells are removed. Capture and barcoding: Cells are then captured and labelled using various methods, such encapsulating single cells in water-in-oil droplets along with microparticles carrying unique primers and lysis buffer. The mRNA content of each lysed cell is captured by a single primer's poly-A tail domain and labelled with cell-specific barcodes and unique molecular identifiers (UMIs), short random nucleotide sequences added to individual RNA molecules to distinguish them and correct for errors in the amplification (233). Most of these steps allow for a multitude of different methods to approach. From here the experimental approach is identical to bulk RNAseq and allows for multiple different techniques to be used. Demultiplexing methods, employing barcode information to sort samples, are performed before bioinformatics analysis. Common problems of scRNAseq include:

- **Sub-Poisson loading trade-offs**: Deviations from the goal of having exactly one barcoded cell in each droplet. This includes multiple cells or microparticles in a single droplet ("multiplets"/""doublets") or empty barcoded droplets (233).
- **Presence of 0 values:** Some genes may have no or low expression in certain cells or conditions, resulting in 0 read counts. This may be due to low cell numbers. The small amount of individual RNA per cell can result in low library complexity, and difficulty in detecting rare transcripts, especially in low abundant cell types.
- **Activation of live cells:** During the collection and dissociation process, live cells might be stimulated or activated, leading to changes in gene expression profiles that do not accurately represent their true physiological state.
- **Limited donors and small sample sizes:** scRNAseq is often limited in donor and sample size, often not reflecting a broader population.

3.1.3 Types of RNAseq Used in This Thesis

For projects 1 and 3 RNA sequencing was performed using poly-A selection based bulk sequencing for differential expression profiling (1,3) and splicing variant analysis (3). Project 2 and 3 used bulk RNAseq data from the Genotype-Tissue Expression (GTEx) project to calculate cellular enrichment scores for KANK3 (3) and the whole transcriptome of adipose tissue samples (2). Project 2 uses single cell RNAseq data from mice and humans as validation and comparison to bulk sequencing data.

3.2 Normalisation of RNAseq Data

Comparing RNAseq data requires some type of normalisation, to ensure that the data is uniformly scaled, remove technical bias and batch effects (234). Commonly used methods are, TPM, FPKM, RPKM, CPM/RPM, Limma, TMM and DESeq2. In the presented studies, DESeq2 normalisation for studies 1, TPM normalisation in study 1 and 2, and downloaded TPM normalised data from the GTEx portal in study 2 and 3. Studies 1 applied a temporal, sex-matched and gene-matched differential using DESeq2 and TPM normalised differential expression. While DESeq2 and TPM serve different purposes, in these projects they were used for complementary analyses.
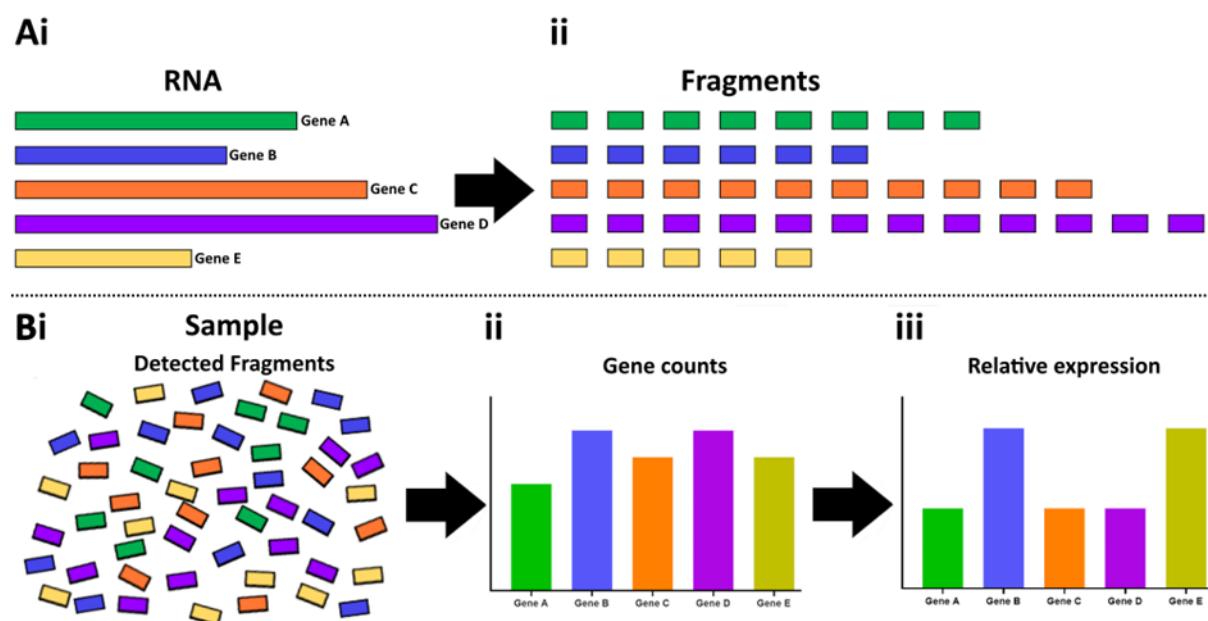
Figure 24: Rationale for normalisation in RNAseq: *RNAseq relies on normalisation techniques to enable comparisons of gene expression across samples. (Ai) Variability in RNA transcript lengths exists, and (ii) longer transcripts correspond to a greater number of fragments during sequencing compared to shorter ones. (Bi) This disparity introduces bias during the quantification process. RNAseq involves fragment detection and quantification, leading to the generation of (ii) raw counts that represent the detected fragments. (iii) If left unadjusted, these counts can show substantial differences due to transcript length effects.*

3.2.1 Gene Counts

Gene counts (also read counts or raw counts) refer to the number of reads mapped to a gene within a sample. Using gene counts alone, one cannot compare between genes and between samples and calculate differential gene expression, the gene counts need to be normalised. If a gene is being detected is dependent on several factors, the relative abundance (number of genes per sample), sequencing depth, and gene

length. To detect the relative abundance of a gene transcript, several methods rely on gene length (Figure 24) and sequencing depth (Figure 25) normalisation (235).
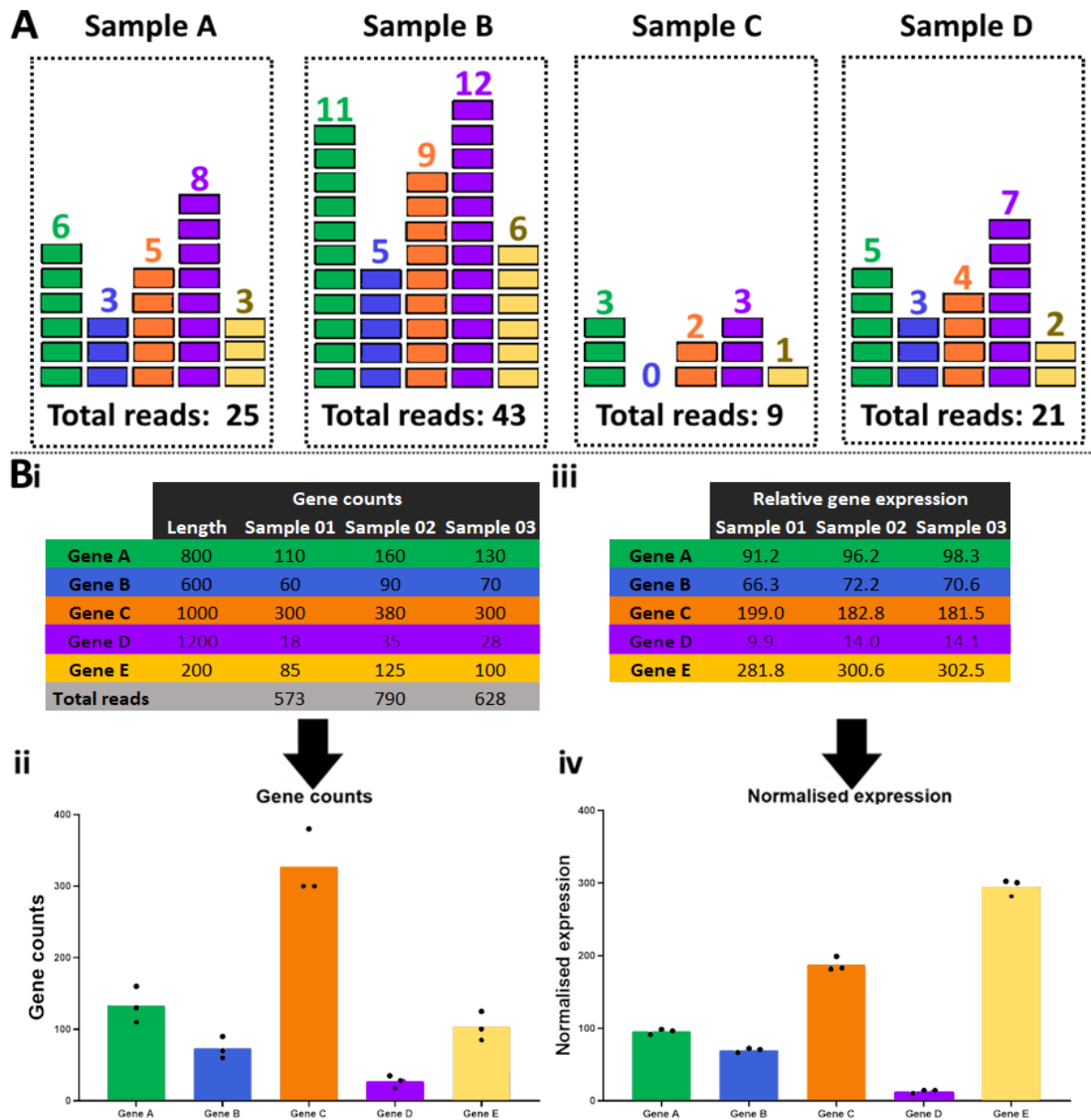
Figure 25: Sequencing depth in RNAseq: *The normalisation strategies adopted in this thesis incorporate gene counts, gene lengths, and total reads within each sample (read depth). (A) Disparities in read counts mapped to genes across samples can arise due to variances in sequencing read depth. Disregarding these differences might lead to an inaccurate perception of differential gene expression between distinct samples. (B) Read counts and read depth are considered during normalisation. Tables and graphs representing read counts and normalised expression. (i) To illustrate, a hypothetical read count profile is presented, demonstrating how (ii) the gene count alone might erroneously indicate Gene C as the most highly expressed on average. However, upon implementing normalisation techniques, the (iii and iv) normalised profiles highlight a comparable expression pattern among genes, with Gene E displaying the highest expression level on average.*

3.2.2 RPM/CPM, RPKM and FPKM

RPM (reads per million, also CPM (counts per million)), RPKM (Reads Per Kilobase Million), FPKM (Fragments Per Kilobase Million) were amongst the most popular methods for normalisation of RNAseq data, to measure transcript levels. These methods have been largely replaced by TPM (236, 237, 238, 239). They are not used within this thesis but are mentioned for the sake of completeness.

RPM is a normalisation method, which maps the reads of each gene ($reads_{GeneX}$) normalised to the total reads ($reads_{total}$), adjusted per million. RPKM is similar to RPM, but adjusted by the length in base pairs ($length_{GeneX}$) of the transcript and calculated per kilobase (x1000). The RPKM measure, although commonly used, is inconsistent in between samples (240). The formulas for RPM/RPKM are:

$$RPM_{GeneX} = \frac{reads_{GeneX}*1{,}000{,}000}{reads_{total}} \qquad\qquad RPKM_{GeneX} = \frac{RPM_{GeneX}*1000}{length_{GeneX}}$$

In FPKM analysis, the fragment counts per kilobase are normalised by length per million mapped fragments. It calculates the number of fragments of a certain gene ($Fragements_{GeneX}$) divided by the length of the gene in kilobases ($length_{GeneX}$) and the total reads of all transcripts ($reads_{total}$). The result is multiplied by a million. The method is prone to error, sensitive to outliers and variation, as it is affected by gene length and dependent on total mapped reads. It has been largely replaced by TPM, which fixes a number of those problems (240). The formula for FPKM is:

$$FPKM_{GeneX} = \frac{Fragments_{GeneX}}{length_{GeneX}*reads_{total}} *1{,}000{,}000$$

Differential expression is calculated in a similar fashion for all these normalisation types. By dividing the expression value of the Gene in two different conditions, the fold change is calculated. While RPKM is used in single end sequencing, FPKM is used in paired end sequencing.

3.2.3 Transcripts Per Million (TPM)

TPM (Transcripts Per Million) is a normalisation method used to quantify relative gene expression levels in RNAseq data. It is unit-less and therefore easy to compare. This technique allows the comparison of gene expression levels of all genes within a sample or across multiple samples. TPM normalises for the total number of obtained reads and the length of the transcripts.

The method adjusts for differences in gene length and sequencing depth by evaluating counts and gene lengths. TPM was developed as statistical modification of the commonly used model RPKM by switching the order of operations. It is similar to RPKM and scaling with it by sharing part of the formula, it eliminates statistical biases by applying adjustment for average invariance (240, 241, 242).

TPM values are calculated by dividing the number of exon reads of a gene "Gene X" ($reads_{GeneX}$) by the length in base pairs ($length_{GeneX}$) and the total number of exon reads in a sample by length Sum (A), multiplying by one million as scaling factor (= per million).

The formula for TPM is the following:

$$A = \frac{reads_{GeneX}}{length_{GeneX}} \qquad TPM_{GeneX} = 1,000,000 * \frac{A}{Sum(A)}$$

TPM based differential expression analysis is calculated by dividing the TPM value of one state by the TPM value of another state.

3.2.4 Linear Models for Microarray and RNAseq Data (Limma)

Limma (Linear Models for Microarray and RNAseq Data) is a software package used for analysis of gene expression data from RNAseq or micro arrays. It relies on previously normalised data to perform differential expression analysis (243). Differential expression between different conditions is estimated, using a linear model. The count data is normalised into log-counts per million (log-CPM) to stabilise the variance (236).

3.2.5 Differential Expression

$$DEG_{GeneX} = \frac{Gene_{X(condition1)}}{Gene_{X(condition2)}}$$

The data is often used to display logarithmic and decimal fold changes. A logarithmic model, usually base 2 or base 10, is used to scale the data around a central unit, such as 0 or 1, to assess effects in both directions in a comparable way.

3.2.6 Differential Expression Analysis for Sequencing 2 (DESeq2)

Another popular method to assess differential expression is DESeq2. It is a commonly used measure of normalisation. It estimates differential expression through gene-wise dispersion parameter of negative binomial distribution (244). It is a successor to its predecessor, DESeq. DESeq2 is applied to statistically compare gene expression levels between different conditions and identify differentially expressed genes.

DESeq2 calculation is more complex than traditional differential expression methods and requires a statistical programming language, such as Python or R. DESeq2 differs from Limma, as it does not need normalised datasets and a linear algorithm (236). The DESeq2 R-package offers a multitude of graphical outputs, including volcano plots and heatmaps (244). It can be used to create normalised and differential expression data.

DESeq2 applies a multi-step formula, starting with internal normalisation, followed by on estimation of size factors, estimation of dispersion, negative binomial regression, and fold change generation, adjusted by Benjamini-Hochberg (BH) statistical correction. DESeq2 can use gene counts directly. Previous normalisation of RNAseq data is not required for the analysis (243). The simplified multi-step formula:

- **Internal normalisation:** The count data is automatically normalised and corrected for differences in the library sizes. Gene length is not considered. DESeq2's objective involves creating a "virtual reference sample" by calculating the geometric mean of gene counts across all samples for each specific gene. DEseq2 counts have been described as more stable than TPM/RPKM data (242).

- **Estimate of size factors:** Size factor is calculated based on the median of ratios.

- **Estimate of Dispersion:** The dispersion estimate, defined as variability in expression levels among different samples per gene, is calculated.

- **Negative Binomial Regression:** A negative binomial model is applied to the count data, applying the estimate size factors and estimate of dispersion, as well as variation and noise.

- **Fold change generation:** The differential expression is estimated in the form of fold change. Statistical significance using null hypothesis (p-value) and false discovery rate (FDR) are calculated.

- **Statistical correction:** Statistical correction accounting for the possibility of false positives is achieved using Benjamini-Hochberg (BH) correction.

3.2.7 Other Methods

There is a multitude of other ways to calculate differential expression data and the choice of the right tool depends on the type of data generated or the desired analysis or type of data correction. Further methods of normalisation and differential expression analysis that are not discussed in this thesis include TMM, DESeq1, Two-Stage Poisson Model (TSPM), Differential gene expression in RNAseq data, edgeR, and Bayes estimation sequencing (baySeq) (237).

3.3 Statistical Methods

3.3.1 Calculation of Coefficient of Variation (CV)

The coefficient of variation (CV) is a value that is used to measure the relative variability of a dataset. CV was assessed in project 1, to assess the relative change in expression per gene. This was done using the normalised sets (TPM). These sets consider the gene length and library size, providing a measure of the proportion of transcripts for each gene in a sample.

CV can be affected by variation between samples. By performing CV analysis after TPM normalisation, less variation is generated by the data itself, but from the differential in between samples. The equation to calculate CV is the following:

$CV = (SD / \mu) * 100$ with **SD** = the standard deviation of dataset, **μ** = mean of dataset

3.3.2 Multiple Testing Correction

Multiple testing correction (p value adjustment) addresses an analysis for false positive values. With each individual test conducted within the same dataset the probability of false positives increases. Consequently, there is a need to adjust the p-value to address multiple testing (245).

When conducting a statistical test, a p-value cutoff of <0.05 indicates that less than 5% of the observed effects are attributable to random chance when the null hypothesis holds true. Using the example of study 1 in this thesis, the dataset comprises of 58884 genes, which is reduced to approximately 30,000 after cutoff for low expression. These genes are evaluated for differential expression with 30,000 distinct tests for identifying differential expression among the datasets. Using a p-value of 0.05 would therefore inaccurately label around 1,500 genes as differentially expressed solely due to stochastic variations (=random chance), thereby producing false positives.

To address this, the Benjamini-Hochberg (BH) (246) correction was employed to correct for false positives. This method adjusts for the false discovery rate (FDR), which represents the proportion of false positives among all significant results. BH correction effectively controls the FDR at the designated level (usually <= 0.05). BH correction assumes that the majority of null hypotheses hold true, with only a minor fraction of the dismissed null hypotheses representing false positives (247).

3.3.3 Weighted Gene Coexpression Network Analysis (WGCNA)

The Weighted Gene Coexpression Network Analysis (WGCNA), also known as weighted correlation network analysis, is a frequently utilised technique for examining transcript data. It allows exploration of connections or patterns of coexpression among gene sets. This method is commonly utilised for the detection of gene coexpression groups, referred to as modules. Therefore, it proves advantageous for analysing intricate datasets and operates in an "unsupervised" manner, by avoiding any pre-existing inputs that might influence the study's outcomes. Key applications of WGCNA include biomarker discovery, network-based functional prediction, and identification of key regulatory genes (248).
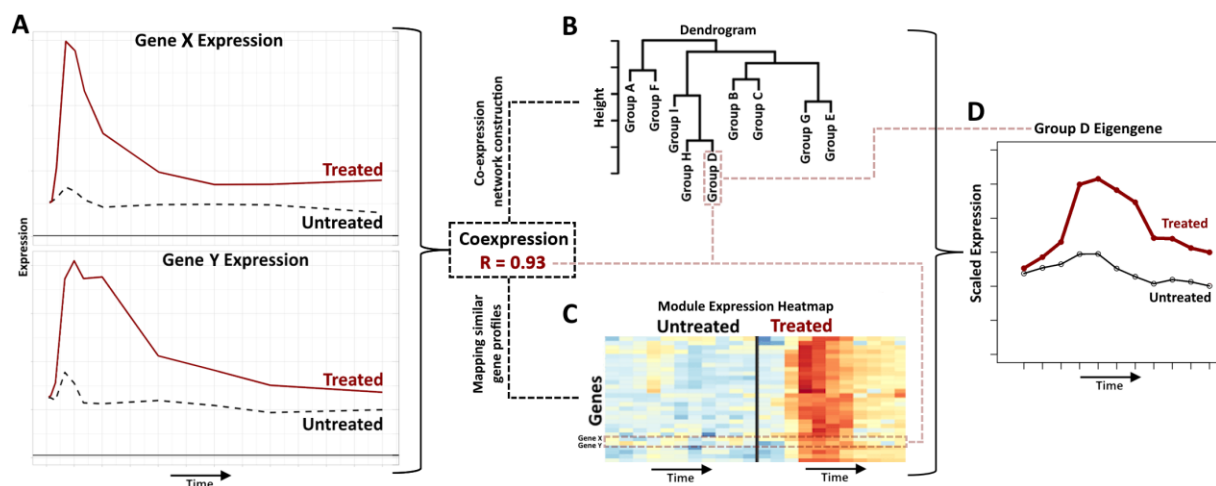
Figure 26: Premise of weighted gene coexpression analysis: WGCNA analyses the association between gene expression patterns. (A) Genes that behave similarly across sample sets, e.g., treatments and time points. (B) Coexpression is calculated between all genes and genes are grouped together based on their coexpression pattern in a hierarchical dendrogram. These groups are also called "modules" (C) Gene modules can be displayed on a heatmap to visualise the expression pattern of all involved genes. (D) The first principal component ("the module's average expression profile") can then be mapped to describe the module in an expression context in form of a general pattern for a panel of genes. Partially adapted from Langfelder and Horvath (2008) (248), generated from own data.

While other approaches, such as differential expression analysis, show which genes are regulated in a specific cellular state or in a cell type, they do not uncover the relationships among them. Furthermore, coexpression analyses illustrate the correlations among genes, providing insights into potential functionalities. These could manifest as clusters of genes linked to particular stimuli or disease states, as well as clusters associated with distinct cell types or pathways implicated in a specific disorder. The concept behind WGCNA is, that genes with correlated expression patterns across samples, also called co-expressed genes, are likely to be functionally related.

WGCNA creates an unbiased map of transcripts by calculating the similarity in expression patterns (Figure 26). This involves computing correlation coefficients among all transcripts and subsequently grouping them together based on their shared expression patterns. As this unsupervised methods groups genes based on their expression profiles, requiring the gene dataset to undergo normalisation and batch effect correction for WGCNA. The normalisation step ensures that the gene expression values are on a comparable scale, while the batch effect removes systematic bias. It is recommended to exclude genes with low counts and minimal variability, as the presence of noise may lead to incorrect clustering and overly large clusters of genes that are not regulated. Graphical outputs include heatmaps and network plots.

3.3.3.1 WGCNA in Study 1

Gene coexpression networks were constructed using the WGCNA package in R (248). Genes with low expression or variation (CV > 0.2) across all samples were filtered out. The goal of the correlation network in study 1 was to visualise how genes relate to one another based on their temporal and post stimulation expression patterns. "Soft thresholding" was used to make the connections more meaningful. Intricately connected gene croups were identified using topological overlap and the dissimilarity was calculated between them. The profiles of each module were visualised using a heatmap of scaled gene expression profiles and the module eigengene.

3.3.3.2 WGCNA in Study 2

WGCNA was used to create gene clusters. Those were compared with the previous lists containing reference transcripts to determine if the genes classified by cell type were predominantly represented within a single cluster, or if they were distributed or blended across cell types. The aim of was to reduce input bias of reference transcripts.

3.3.4 Gene ontology enrichment analysis

Gene Ontology (GO) enrichment analysis is a method used to detect the overrepresentation of genes in datasets corresponding to biological processes, cellular locations, and molecular functions within a given set of genes. The ontologies are clustered in a tree like hierarchical structure of terms (GO-terms). The gene annotations are defined from existing literature using automated methods and manual curation. It has to be noted, that a significant annotation bias exists, with a majority of annotations focused on a small fraction of human genes. (249).

The ontology tree is split into three main categories at the highest levels of the tree: Biological process, cellular component, and molecular function. These terms subcluster into a multitude of further terms, which get more and more precise with each level of the tree (represented in Figure 27). Full gene ontology trees can be accessed via https://amigo.geneontology.org (250).
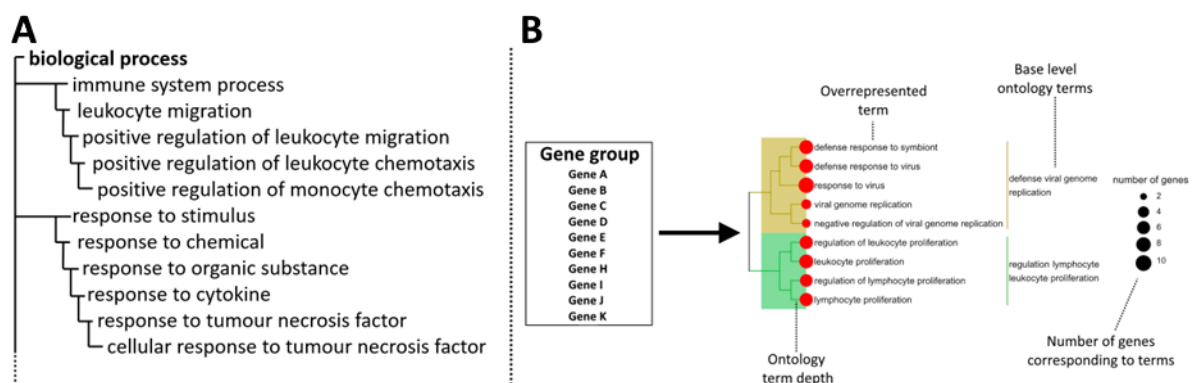
Figure 27: Illustration of Gene Ontology Analysis *(A) Hierarchical structure of the gene ontology tree, with each successive level providing a more detailed explanation of a term. (B) Implementation of GO analysis involves inputting a gene list into a GO tool or software. It reveals terms overrepresented in the gene list compared to a reference set. Statistical significance is evaluated, determining if GO term presence in the gene set exceeds chance expectations (term list from own data).*

There is a multitude of web tools and packages in programming languages to assess GO terms for a dataset, amongst them PantherDB (251), EasyGO (252), GOminer (253), AmiGO (250), GOstat (254), GOToolBox (255), topGO (256), GSEA (257), and DAVID (258), ShinyGO (259), and GOSeq (260). The R packages GOstats (261), clusterProfiler (262) and the PantherDB classification resource (251), were used to identify overrepresented terms in gene lists in the studies of this thesis. GO analysis is conducted by providing a gene list into a GO-tool or statistical software. GO enrichment

tests whether a GO term is overrepresented in a list of given genes compared to a reference list.

Statistical significance can be calculated by using different tools, often dependent on the GO enrichment tool used. Among the statistical tests are hypergeometric testing, binominal test, χ_2 test, or Fisher's exact test, which assesses whether the occurrence of GO terms in the gene set goes beyond chance. If the number of genes associated with a GO term within the set is greater than it would be by random chance, the term is categorised as overrepresented or enriched (261, 263). High enrichment scores suggest that a supplied gene list is biologically meaningful.

3.4 Human Umbilical Vein Endothelial Cells (HUVEC)

The *in vitro* study of endothelial function requires a stable cellular model. Human umbilical vein endothelial cells (HUVEC) are infant endothelial cells from the vein of the umbilical cord (Figure 28). They are the most used endothelial cell model in cell biology and are easy to isolate using collagenase (264, 265).
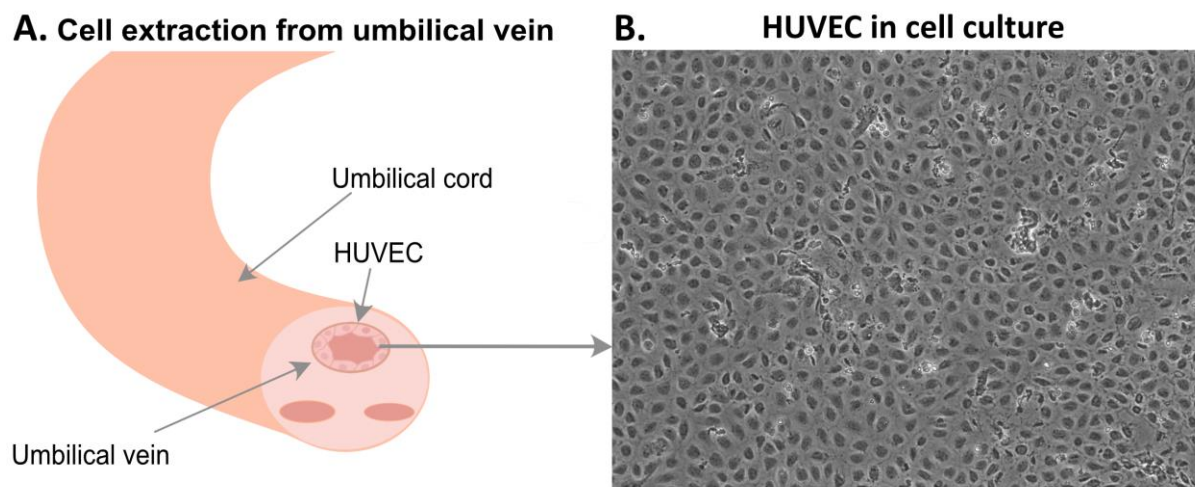
Figure 28: Extraction and culture of human umbilical vein endothelial cells: (A) The umbilical cord has two arteries and a vein. HUVEC are extracted from the vein of the umbilical cord, using a collagenase digest, breaking down the basement membrane and freeing the cells. (B) In in vitro culture, EC form a tight monolayer which has a "cobblestone-like" appearance in phase contrast microscopy.

HUVEC are primary cells, therefore they come with advantages and disadvantages compared to cell lines. An advantage is that they are isolated from an abundant tissue that does not require surgical isolation procedures, and would otherwise have been discarded as waste. Further, HUVEC can be passaged multiple times and are

commonly used for different analyses studying endothelial function and dysfunction. HUVEC express important endothelial regulators and markers and keep the native expression profile during early passages of cell culture (20) and can be used to study sex differences (266).

Of great importance for transcriptional studies is the donor variability of HUVEC; each donor expresses some degree of heterogeneity. In Study 1 and 2, we mitigated this problem by pooling multiple donors into a singular donor pool.

Limitations when using HUVEC is their short lifespan when compared to immortal cell lines, as HUVEC can only be passaged a certain number of times before they become senescent (267). Therefore, HUVEC are not suitable for long term experiments, ranging over the span of multiple weeks. Other disadvantages include slow growing times, lack of in-depth knowledge about the anonymised donors, and the fact that all donors come from foetal tissue rather than mature tissue.

Other Primary Endothelial Cell Models

In addition to HUVEC, we used different EC types as comparison in study 3. Among those are mouse liver sinusoidal endothelial cells (mLSEC), human pulmonary artery endothelial cells (HPAEC), human coronary artery endothelial cells (HCAEC), human dermal microvasculature cells (HDMEC), and human retina endothelial cells (HREC). All these cells are primary cells and therefore have the same disadvantages as described above. HDMEC, used in study 3, were isolated from the dermis of juvenile foreskin and are therefore always of male sex.

4 Methods

4.1 Endothelial Extraction, Culture and Stimulation

Projects 1 and 2 used HUVEC cells as a model. Ethical approval for endothelial cell isolation and subsequent experimentation was granted by Regionala Etikprövningsnämnden i Stockholm (diarienummer 2015/1294-31/2).

4.1.1 Extraction of endothelial cells

HUVEC were isolated from anonymised human umbilical cords, collected from Karolinska Hospital (Stockholm, Sweden) using a method previously described (268). In short, the umbilical cord was directly cut from the placenta and squeezed to remove any blood and blood clots. The cord was clamped 2 cm from each end, with the bottom end cut above the clamp, leaving the clamp intact, and the front end cut below the clamp, leaving the cord open. This measure helps prevent infection, as the ends are more susceptible to contamination.

Three blood vessels are found in umbilical cords: two thin elasticated arteries and one thicker, collapsible vein. The vein can be identified by its wider appearance and clearly visible walls and opened using a bent nosed clamp, which was inserted approximately 2-3 cm into it. A glass cannula was then inserted into the vein and secured using one or two cable ties. The cord-vein was filled with sterile PBS until it became taut, ensuring that any potential leaks were clamped to seal them.

To drain the PBS from the vein, the bottom clamp was opened, allowing the liquid to flow into a waste container. If the liquid was bloody, it was washed with additional sterile PBS until it became clear, after which the bottom clamp was reattached. Subsequently, the cord was filled with 0.1% collagenase in PBS until it became taut and placed in the incubator at 37-39 °C for 15-20 minutes. After the incubation period, the cord was removed from the incubator and massaged to facilitate the detachment of cells. The bottom end of each cord was placed into a falcon tube, and the liquid was forced out of the cords into the tube. Then, an equal volume of FCS in PBS was pushed through the cord to match the volume of collagenase in PBS for inactivation purposes.

The resulting cell suspension was centrifuged at 400xG to form a stable cell pellet. The supernatant was discarded into a waste bottle, and the pellet was resuspended in 6 ml of sterile Medium 199 with additives. The resulting cell suspension was plated in

cultivation flasks and placed in the incubator. If the medium appeared very bloody in any flask, it was washed with PBS and the medium was replaced after approximately four hours. Cells were cultivated in M199 with supplements for further experiments. For a subset of experiments in project 3, HUVEC and other EC types were purchased from Merck/Sigma Aldrich.

4.1.2 Inflammation Models

Inflammation models by stimulation with cytokines and LPS were used in projects 1 and 2.

Endothelial response to inflammation was measured by adding different cytokines to the cell culture medium. The activation of endothelial cells was subsequently analysed using transcriptomic or proteomic approaches. Cytokines were supplied in the following final concentrations: IFNα (6ng/ml), IFNγ, (60ng/ml), IL-1β (10ng/ml), IL-4 (10ng/ml), TNFα (10ng/ml).

4.1.3 RT-qPCR

Reverse transcription PCR (RT-PCR) followed by quantitative real time PCR (qPCR) was used in projects 1 and 2.

Reverse transcription followed by quantitative PCR used the TaqMan ™ Fast Cells-to-CT ™ Kit provided by ThermoFisher Scientific. In projects 1 and 2 it was used for sex-determination to select HUVEC donors for the target mRNA UTY (Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-Linked). In project 3, this method was employed to determine KANK3 knockdown efficiency, as well as mRNA expression of target proteins.

qPCR was performed using TaqMan Fast Universal PCR mix. Target primer conjugated to FAM-probes were used to assess target gene mRNA levels. 18s rRNA (4319413E conjugated to VIC probe, ThermoFisher) was used as endogenous control. qPCR was performed using a RealTime PCR LightCycler 96 ® system (Roche Life Sciences) or StepOnePlus™ from Applied Biosystems.

4.1.4 RNA isolation and sequencing

RNA isolation and sequencing was performed in projects 1 and 3. Project 2 and 3 used publicly available RNAseq data (GTEx, human protein atlas, and tabula sapiens).

RNA isolation was performed using the RNAeasy mini kit (Qiagen). The RNA concentration was measured using Nanodrop 2000 spectrophotometer and RNA

integrity number (RIN) determined using Agilent 2100 Bioanalyzer. RIN >9 were required for inclusion to sequencing.

Library preparation and RNA sequencing was performed by the National Genomics Infrastructure Sweden (NGI) using Illumina stranded TruSeq poly-A selection kit and Illumina NovaSeq6000S (4 lanes, 2x 150bp reads, incl 2Xp kits). The data was processed using demultiplexing.

Data storage and initial analyses were performed using server sided computation were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

Sequence annotation was carried out using the reference genome sequences Homo_sapiens.GRCh38 for genome assembly and the reference assembly Homo_sapiens.GRCh38.96 was used for sequence alignment and annotation.

Sequence alignment was carried out using STAR/2.5.3a. Gene mapping has been carried out using subread/1.5.2 and the module feature counts. Transcript mapping carried out using Salmon/0.9.1. Normalisation was performed from gene counts to TPM and Deseq2 as described before.

4.2 Project 1: 72h TNF-Stimulation Analysis

HUVEC from male and female donors were either treated with TNF or left untreated and bulk RNA sequencing was performed, as described above. Raw gene counts were normalised to TPM or differential DESEq2 counts. The study contained 132 samples: 6 donors (3 male, 3 female), with 11 time points (hours: 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48, 72), and two conditions ("Untreated", "TNF stimulated").

DESeq2 differential expression values were used to map gene count comparisons between samples across the same groups. Groups were defined by treatment type, sex, and time point to identify genes affected by TNF treatment, gene centric differential expression analysis was performed between TNF treated sample and the matched control. The analysis was executed by comparing sex-, time- and donor-matched TNF stimulated samples to their respective control and calculating fold change values. It should be noted that each sample had a time matched control. Average values were generated by merging same sex replicates for each timepoint or merging all replicate samples for respective time points.

Multidimensional scaling plots were generated to visualise the clustering of samples based on overall gene expression. These plots included both replicates and averaged expression across biological replicates. The DESeq2 time series design was employed to identify genes with different expression between treated and untreated samples for both female and male samples. The full model, represented by ~treatment + time + treatment:time, and the reduced model, represented by ~treatment + time, were used in the analysis. Genes with an adjusted P-value <0.05 and an absolute fold change of log2 >1 were classified as differentially expressed genes (DEGs).

4.2.1 Gene Centric and Module Centric Approaches
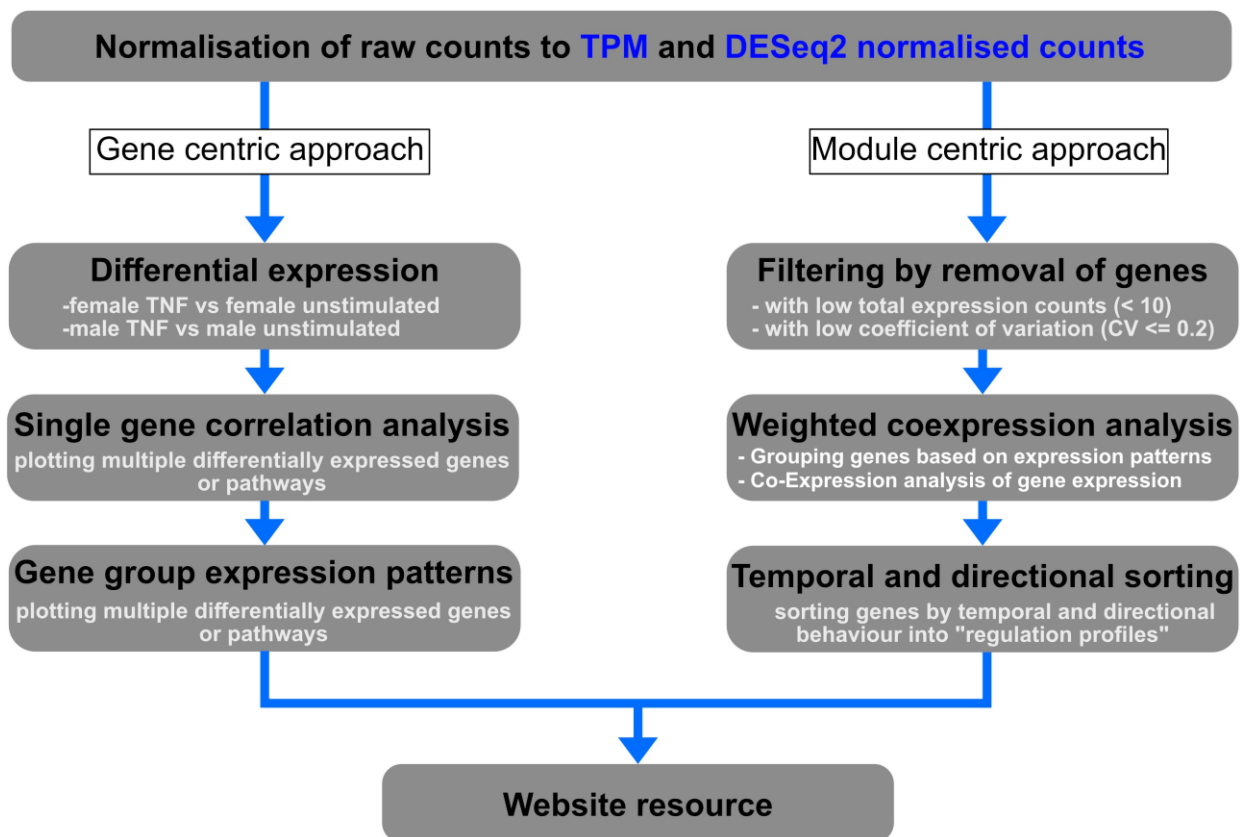
Figure 29: Two-sided approach for data analysis: *The gene centric approach encompasses the analysis of singular genes in both baseline behaviour and differential expression between analysed variables. The module centric approach uses an unsupervised clustering method, which generates gene groups with behavioural similarity throughout the whole study.*

The project employed two central approaches: The gene centric and the module centric approach. Both approaches rely on normalisation of the transcript data, followed up by several analysis steps (Figure 29).

The gene centric approach explores the baseline and differential expression of both individual gene transcripts and gene groups without removing any genes or applying any filters. This approach allows analysis of baseline and differential expression values for any gene.

- o Single gene correlation analysis allows plotting of individual genes to find gene groups that match the gene behaviour on baseline level or considering the differential profile.
- o Gene group expression analysis allows multiple genes to be checked simultaneously for comparison of differential and baseline profiles.

55

The **module centric approach** explores gene groups that have been clustered into modules by weighted coexpression analysis after filtering. Lowly expressed genes were filtered out and then clustered into groups with similar expression patterns using the WGCNA clustering algorithm.

- o Weighted coexpression analysis defined 48 modules with unique expression patterns, highlighting similarities and differences between baselines, differential profiles, and sexes.
- o Temporal and stimulation differential sorting was performed to identify differential regulation patterns after TNF treatment.
- o Furthermore, the principal component, the eigengene, of each module was defined to create a central pattern for behaviour of groups.

The website resource was then developed employing results from both approaches. Results generated from the website resource were subsequently merged into the results of this project.

4.2.2 Module Independent Definition of Regulatory Profiles

To further classify the DEGs, gene subgroups were created based on their reaction profile from the previous analysis, excluding genes with low expression. The DEGs were then clustered into positive or negative regulation profiles. Genes that fell into both categories at different time points and in different sexes were excluded from the analysis. For inclusion, DEGs had to be classified as temporally differentially expressed at two or more sequential time points in DESEq2.

Positive DEGs were classified using exclusion criteria, where genes with a decreasing baseline and stable TNF expression were excluded. Genes were excluded if they exhibited low variation within TNF-treated samples and high variation within the baseline samples compared to the start time (0.5 h).

Negative DEG profiles were defined as "downregulated" when TNF reduced gene expression from a stable baseline, or as "temporally delayed," "inhibited," or "other" when baseline changes could drive differential expression. DEGs that did not fit into any of these three categories were categorised as "other."

4.2.3 Website generation

A web resource available under http://www.endothelial-response.org has been created in the course of this project using R and shiny (269). The resources utilises normalised expression values (TPM) and the data generated in the WGCNA analysis to make the data available to the public. It allows to view differential expression of genes, profile gene groups and pathways, perform gene centric correlations to find clusters with similar behaviour to a target gene and to investigate the temporal activation profile, as well as the sex split differential between samples. The module central analysis allows for investigation of patterns for genes with similar behaviours, defined in study 1. All data generated in this study is freely available through this resource.

4.3 Adipose tissue enriched transcriptome

4.3.1 Genotype Expression Project Bulk RNAseq Data Set

Bulk RNAseq data for subcutaneous adipose tissue (SAT) [n = 646] and visceral adipose tissue (VAT) [n = 527] was downloaded from the Genotype-Tissue Expression (GTEx) project (GTEx Analysis V8 (dbGaP Accession phs000424.v8.p2), www.gtexportal.org) (270). The data was provided as TPM values corresponding to ENSEMBL IDs. The metadata contained the following variables: donor id, cause of death (categorical), sex, race, and age. Subcutaneous adipose tissue was taken from beneath the skin on the medial side of the leg. Visceral adipose tissue was taken from the greater omentum, a fibro-fatty tissue that spans the abdomen.

4.3.2 Integrative Correlation Based Analysis

Butler et al. (2016) developed a bioinformatics based approach using integrative correlation analysis of selected reference transcripts, to determine cell type enriched genes by the aids of selected cell type specific reference transcripts. While Butler et al. (2016) focused solely on EC enriched cell types, Dusart et al. (2019) significantly expanded the scope by introducing thresholds and conducting cell type comparisons, allowing for the examination of multiple cell types within the same tissue (brain) (20, 106) . Integrative correlation is a statistical method utilised to evaluate the consistency of gene expression data across a set of studies or samples. This highly reproducible gene-specific measure estimates the correlation of correlations among multiple genes. This type of analysis was developed for measure of cross-study reproducibility across gene expression data sets (271).

By evaluating the correlations of TPM values among numerous genes, in different samples or studies, the approach computes a gene-specific correlation score. This correlation score serves as an estimation of the extent to which a group of genes is likely to be enriched in a specific context. Integrative correlation can be performed through various statistical software, such as R or Python.

4.3.3 Integrative Correlation Based Analysis in Bulk Sequence Data Using Cell Type Specific Reference Transcripts
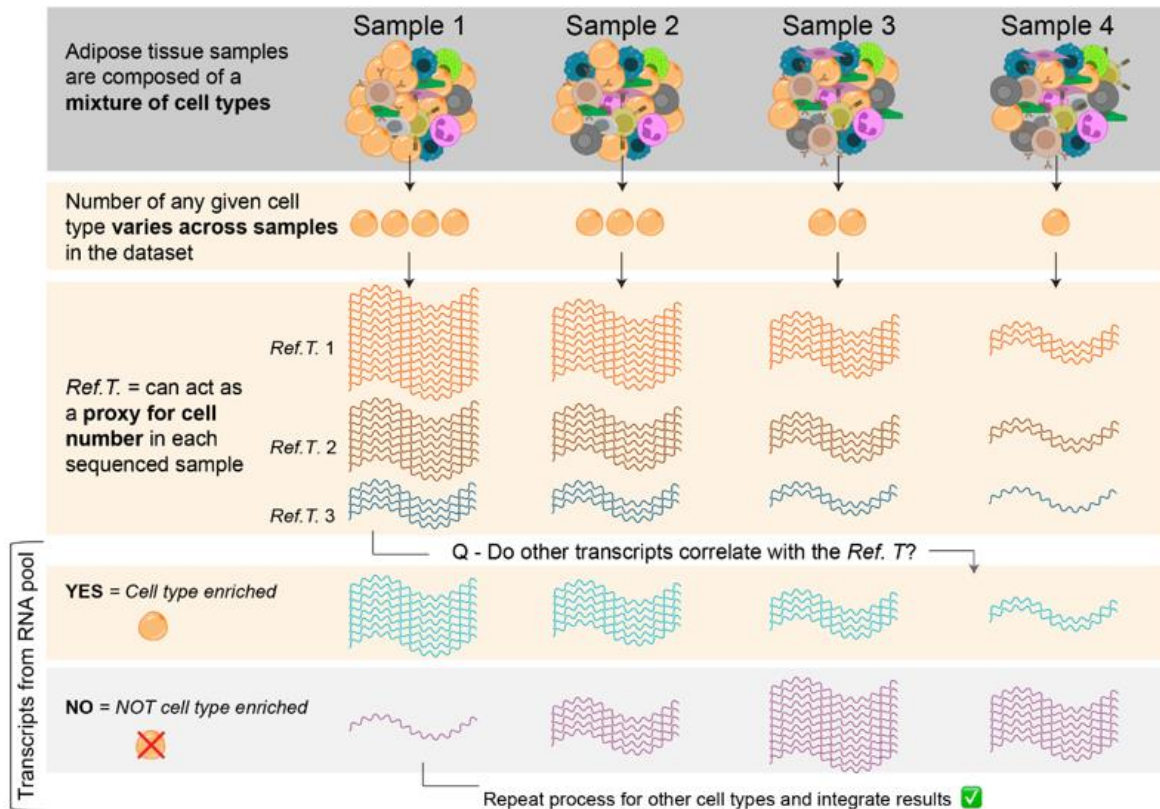
Figure 30: RNA sequencing data from adipose tissue: *Adipose tissue is composed of a mixture of cell types that are present at various numbers in each tissue sample of which bulk RNA sequencing data is based on. Reference transcripts, which are cell specific markers, are chosen based on a similar transcript ratio across samples, as they have same cell type origin. They can be used to identify other cell type enriched transcripts with a similar expression pattern. (Figure received, with courtesy, from Marthe Norreen-Thorsen).*

The GTEx samples were extracted from bulk sequencing of tissue samples. This means transcript data was not sorted by cell type but rather represents a mix of different cell types in different distributions. Specific markers for each cell type were selected using existing literature.

These markers are referred to as "reference transcripts". In this study three reference transcripts (Ref.T) were chosen per cell type. While each of these gene transcripts could serve as a representative marker for a particular cell type within a mixed population of cells, the combined three reference markers to establish that reference transcripts expression patterns reflect cell type proportion in tissue samples, and therefore transcripts from cell type are found with same ratio and correlate (Figure 30).

Expression values (TPM) of the Ref T. were assessed. To measure the strength of the relation, a correlation based analysis to the mapped transcriptome was performed using Spearman's ranked correlation (Figure 31). Spearman's rank correlation coefficient (ρ) is defined as follows:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \quad \text{with}$$

ρ = Spearman's correlation coefficient

d_i = difference between two ranks

n = number of data pairs

Spearman's rank correlation will generate values between -1 and 1, with the value of 1 being perfect correlation and -1 being a total inverse correlation. A value close to 0 indicates no relationship.

This analysis assumes that gene transcripts from the same cell type are expressed to a similar extent in each cell of the same type, but in different extent or not at all in other cell types. Housekeeping genes that appear in similar extents in diverse cell types or genes that correlate with multiple cells were corrected for by choosing multiple specific reference transcripts (Ref.T.) (Figure 31).

To find cell type enriched transcripts, three Ref.T. were defined per constituent cell type using literature research and cell type specific IHC staining results in the Human Protein Atlas. The following ten constituent cell types for adipose tissue were defined: adipocytes, adipocyte progenitor cells, endothelial cells, macrophages, mast cells, mesothelial cells, neutrophils, plasma cells, smooth muscle cells, and t-cells. The hypothesis was that transcripts with high correlation with Ref.T from only a single cell type likely have cell type specific expression.
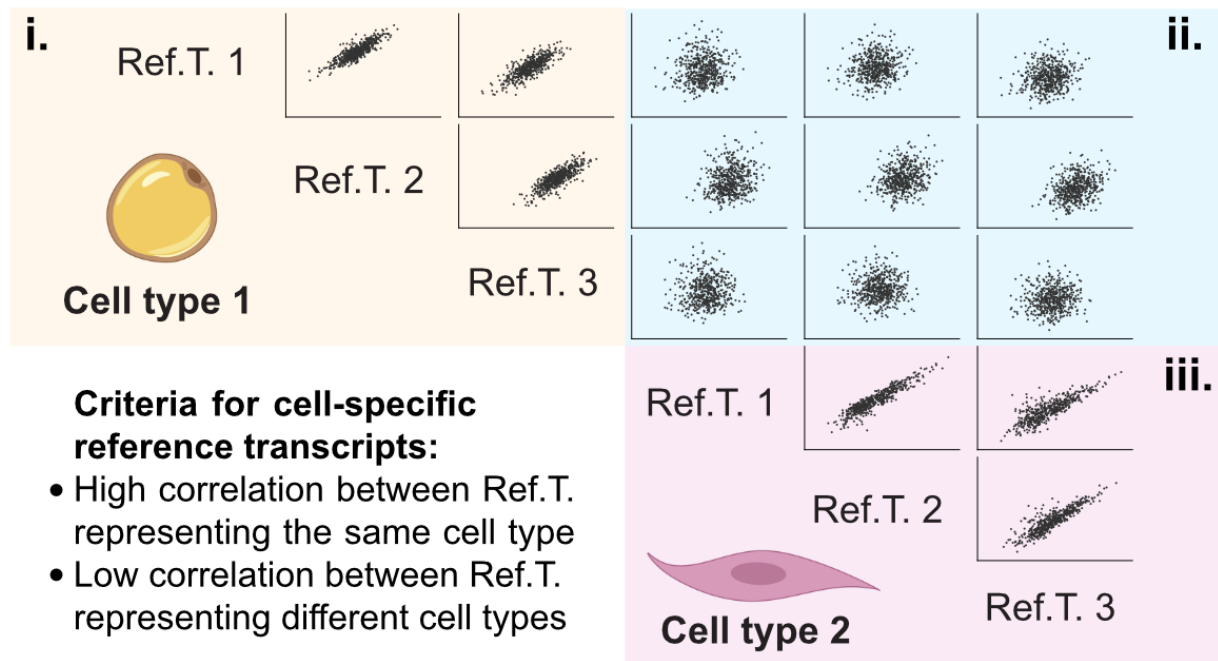
Figure 31: Expression of reference transcripts in two example cell types: (i, iii) Reference transcripts (Ref.T.) for the same cell types have a high correlation, when similar cells are expressing genes in a consistent manner (ii) but differ in correlation when the cell types are not related, as their expression patterns differ (Figure received, with courtesy, from Marthe Norreen-Thorsen).

To find suitable candidates for cell specific reference markers a shortlist of candidates was created, and Spearman correlation coefficients were calculated between these candidates. Candidates selected must have shown a high positive correlation coefficient with another and a low correlation coefficient with any other Ref.T.

4.3.3.1 Cell Type Enrichment

To reduce false positives, lowly expressed gene transcripts (TPM <0.1 in more than 50% of the samples) were regarded as noise and excluded from the analysis. Spearman's rank correlation coefficients were calculated between the three selected Ref.T. per cell type and all other sequenced transcripts. P-values below 0.05 and FDR < 0.0001 was required for inclusion. Subsequently, a threshold correlation value was chosen for inclusion and exclusion of gene transcripts:

(a) Correlation value above which >95% of all transcripts reached the threshold with only the Ref.T. panel.

(b) A minimum of ≥0.50, if higher than (a).

A differential correlation score (Figure 32) was used to identify and exclude multi-enriched transcripts: transcripts correlating with more than one cell-type. The differential score was defined as the difference between the highest mean correlation coefficient with a Ref.T. panel, and the second highest mean correlation coefficient with another Ref.T. panel. Transcripts with a differential correlation value >0.1 were classified as cell type enriched.
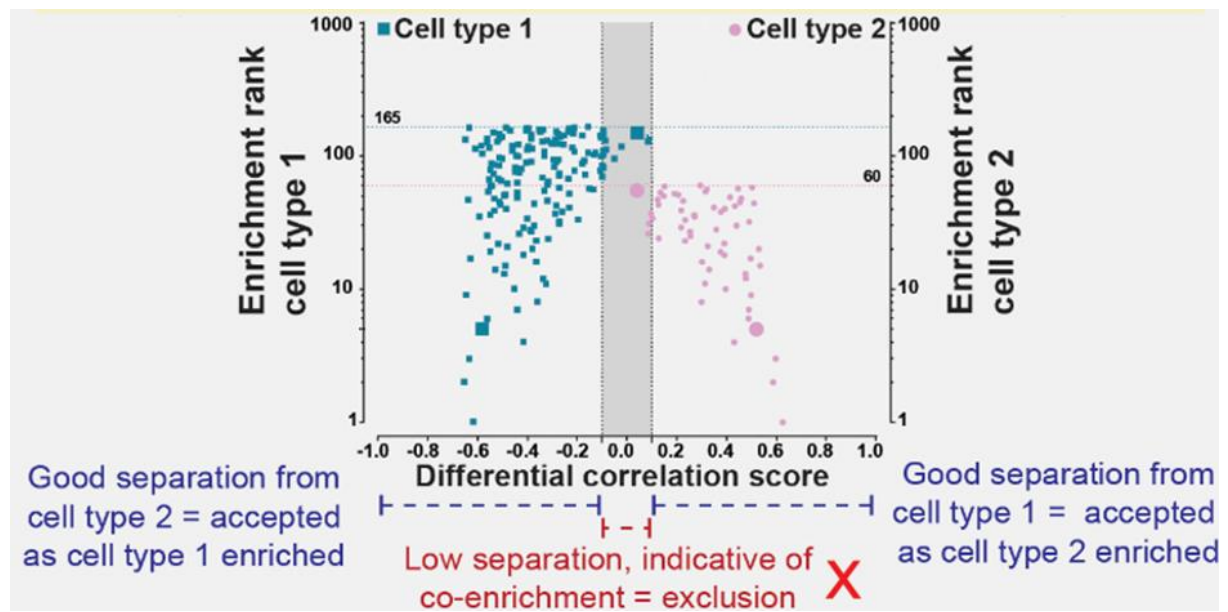
Figure 32: Differential correlation score: *The differential correlation score was defined as difference between a transcript mean correlation with a Ref.T panel of one cell type vs. another. The enrichment rank represents the position of a transcript in a cell type enriched list. Transcripts with a differential score below 0.1 were excluded from the cell type enriched list and classified as co-enriched (Graphic received from Marthe Norreen-Thorsen).*

4.3.4 Verification

Verification of the Ref.T. correlation results was performed in multiple ways:

- **Gene ontology (GO) analysis:** GO biological process enrichment analysis was carried out for each cell type enriched list using the Gene Ontology Consortium (272) and PANTHER classification resources (273, 274), to define significant over-represented terms. These terms could indicate involvement in cell specific processes. They thereby act as a reasonable indication for correctness of cell type classification and were used for validation of cell type specific processes.

- **WGCNA:** Calculation of pairwise correlation coefficients among all transcripts, followed by clustering based on their similarity in expression patterns, was performed as described in section 3.3.3. and was used as additional support to the reference-based classification, as it is not based on any pre-set input.

- **IHC staining:** For a selection of cell type enriched genes, the cell specificity of the corresponding protein was verified using IHC staining, provided by the Human Protein Atlas (179).
- **Single Cell sequencing resources:** We used single cell sequencing resources (*tabula sapiens* human cell atlas (115), *tabula muris* (114)) and created Uniform Manifold Approximation and Projection (UMAP) plots to compare the data generated in this study to the cell type ell type annotated cluster with highest expression. The four studies employed three thresholds: avg. log2 fold change of >0.2, >0.5 or >1 in each annotated cell type, relative to all other cell types within each study. Significance was assessed using hypergeometric testing.

4.4 Functional Characterisation of KANK3

4.4.1 Gene Knockdown and Recombinant KANK3 Protein Expression

KANK3 function was investigated using gene knockdown and overexpression strategies. For knockdown, three siRNA targeting KANK3 were used. siRNA was transfected into cells using lipofectamine and Opti-MEM. Knockdown efficiency was evaluated after 48-72h using RT-qPCR and Western Blot. Overexpression was achieved by transfecting plasmids into cells using Lipofectamine 3000 in HEK293 cells.

4.4.2 Confocal and Structured Illumination Microscopy

Confocal and structured illumination microscopy were used to determine the protein localisation and effects of gene expression loss. For confocal microscopy, cells were fixed in 4 % formaldehyde, permeabilised using 0.05% Triton X in phosphate buffered saline (PBS), and blocked using 3% bovine serum albumin (BSA) in Triton X. They were then incubated with primary and secondary antibodies and mounted with DAPI (4′,6-diamidino-2-phenylindole) nuclear stain. Images were captured using a confocal microscope and analysed using Fiji ImageJ2 software. Structured illumination microscopy was conducted on fixed cells grown on glass coverslips. Primary and secondary antibodies were applied, and DAPI staining was performed. Images were acquired using an OMX Blaze SIM microscope, and reconstruction and analysis were performed using software.

The cells were then imaged using a microscope, and gap size was measured over time using Fiji software.

4.4.3 Wound healing assay ("Scratch assay")

A.) Cultivate cells
B.) Scratch HUVECs with 100µl pipette tip to create gap
C.)
D.) Investigate gap closing for 72h under microscope
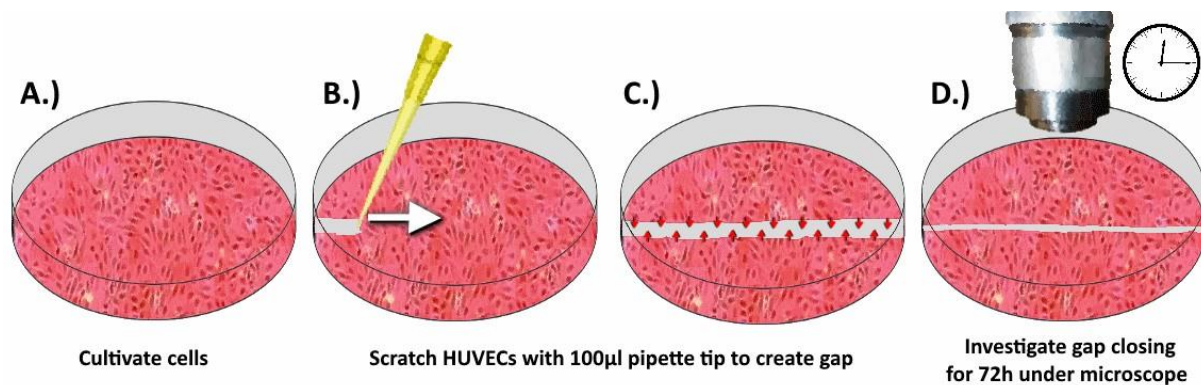
Figure 33: Principle of wound healing assay in endothelial cell culture. *(A) The EC layer is cultivated until confluence. (B) A gap using a pipette tip or culture inlets is created and the medium is washed to remove cells that have been lifted off (C) Surrounding cells will begin close the gap through proliferation and migration. (D) The gap size is captured over time.*

A wound healing assay, also known as scratch assay, is a commonly used technique to study cell migration, proliferation, and cell-cell- interactions. It is simple, low cost and easy to reproduce. Cells are grown to confluence and a wound is created by scratching the cell layer (Figure 33). The scratch is usually created with a pipette tip or another sharp object (275, 276), but several companies (e.g., Ibidi) sell inlets for cell culture creating a gap. The cells will usually start to close the gap, which can then analysed using time lapse microscopy (276).

An inherent drawback of the employed technique lies in the potential non-uniformity of the gaps generated while utilizing a pipette tip for gap creation. The cells close to the gap could also be damaged or activated. Meanwhile, the utilization of inlets on fully confluent endothelial cell monolayers resulted in inconsistent detachment of cells or cell segments for us. Manual gap measurements can lead to bias, which is why we randomised and anonymised the samples before analysis. Due to low image quality, an unbiased deep learning approach failed.

4.4.4 Calibrated Automated Thrombogram (CAT)

Thrombin generation is an essential step to induce and maintain haemostasis, directing the coagulation cascade to a fibrin clot, as explained in section 1.1.5. Thrombin cleaves soluble fibrinogen into insoluble fibrin, and thereby forming a blood clot. Measure of thrombin generation can be used as a tool to assess blood clotting ability (277).

A first thrombin generation test was developed by Macfarlane and Biggs (1953) and has been used in clinical and laboratory settings since then (278, 279). The assay has been constantly improved, by using different measurements and by employing computational approaches (280, 281).
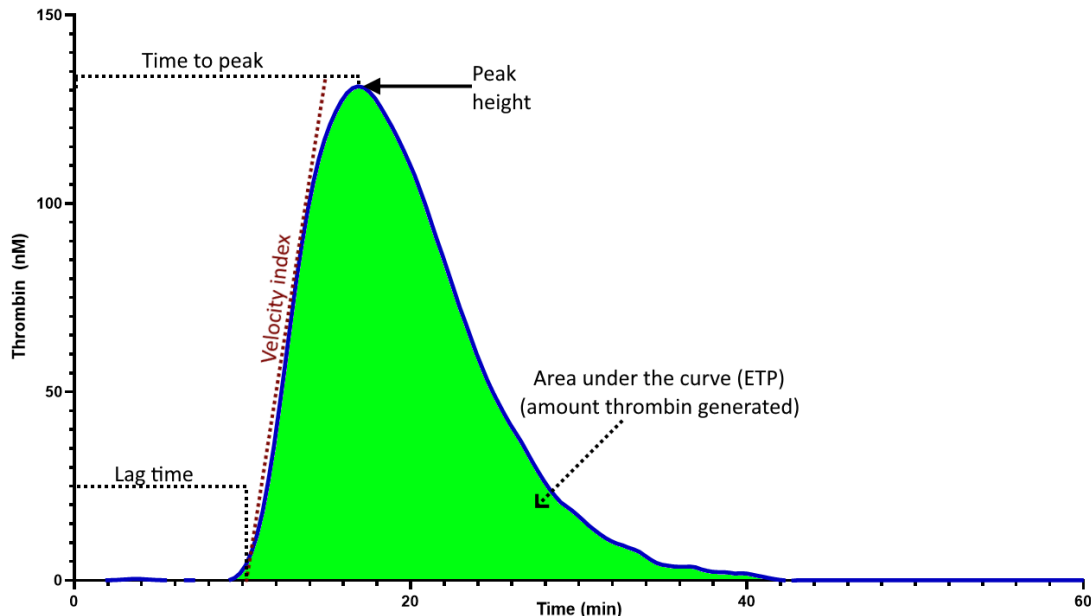
Figure 34: Thrombin generation analysis using calibrated automated thrombinoscope: *CAT determines thrombin activity over time in plasma by measuring released fluorescence due to specific substrate cleavage. The assay analyses the time until thrombin generation starts (lag time), time to peak (time to reach maximum thrombin concentration), peak height (maximum thrombin concentration), the area under the curve (representing total thrombin generation) is called endogenous thrombin potential (ETP).*

The calibrated automated thrombogram (CAT) assay encompasses several essential steps in evaluating thrombin generation (Figure 34).

- The initiation of the coagulation cascade is achieved through a triggering agent, activator, leading to thrombin production.
- A fluorescent substrate specific to thrombin is then introduced, responding to thrombin's proteolytic activity.
- This prompts thrombin to selectively break down the substrate, leading to the release of a fluorophore. The fluorescence signal is continuously monitored over time. This real-time signal corresponds to the amount of active thrombin present in the sample at any given moment.
- To extrapolate thrombin concentration from the fluorescence signal, the assay is "calibrated" by using known thrombin concentrations reagent as endogenous control, enabling precise quantification by comparing the generated thrombin activity with the reference calibrator.

64

- Thrombin generation is measured by assessment of several factors, such as lag time, endogenous thrombin potential (ETP), time to peak and peak height. The area under the curve represents the endogenous thrombin potential (ETP). It reflects the amount of thrombin generated. Lag time is the time until thrombin generation starts. Peak height shows the maximal thrombin concentration.

While CAT assays are usually performed with plasma samples which are activated by exogenous commercial activator reagent, we employed an in house assay where the activation is triggered by cultured cells in the flat bottom plate. Therefore, the assessment was indirectly measuring endothelial expression of coagulation factors and the potential of EC to induce the coagulation cascade. CAT was employed to investigate the impact of KANK3 on the coagulation pathway in ECs. HUVEC were stimulated with TNFα, and thrombin generation was initiated using a reaction mixture (a buffer containing a fluorescent substrate, calcium chloride, and phospholipids). Thrombin generation was quantified using Thrombinoscope software. Mouse monoclonal anti-tissue factor antibody or corn trypsin inhibitor were added 15 min prior adding fluorogenic substrate to control for extrinsic pathway or intrinsic pathway activation. Experiments were performed in triplicates.

4.4.5 Flow Cytometry

Flow cytometry was performed on HUVEC transfected with siRNA targeting KANK3. A subset of cells were stimulated with TNF and collected by trypsinisation after 24 hours. Flow cytometry analysis was conducted to identify tissue factor levels in in cells. HUVEC were collected and concentrated, and treated with PE-conjugated anti-CD142 Clone NY2 (30 μl/ml) and isotype-matched control mouse-IgG1 (6 μl/ml) followed by incubation on ice for 30min and centrifugation. The pellet was then resuspended in PBS. Flow cytometry was performed using the Beckman Coulter CytoFLEX Flow Cytometer (acquisition settings FSC 20 V, SSC 150 V, PE 130 V). Gating and data analysis was done using CytExpert for CytoFLEX Acquisition and Analysis Software and FlowJo™ v10.7. Gating was performed for live vs dead and singlets vs doublets. Dead cells and doublets were excluded, followed by gating for tissue factor positive and negative cells. Isotype control signal was subtracted from PE-conjugated anti-CD142 total fluorescent signal for each sample. Median fluorescence intensity and tissue factor positive cells were identified for each condition.

5 Paper summaries

5.1 Project I

Global transcriptome analysis reveals distinct phases of the endothelial response to tumour necrosis factor

Objectives: The objective of this study was to investigate the transcriptional changes in endothelial cells (EC) in response to TNF over an extended time period in both male and female HUVEC, to elucidate the overall dynamics of the EC response to TNF and identify regulated genes and coexpression profiles.

Methods: We conducted an extended time-course analysis of the EC response to TNF, ranging from 30 minutes to 72 hours. This study utilised gene expression analysis techniques to identify regulated genes and employed weighted gene correlation network analysis to decipher coexpression profiles. A website resource was created to analyse and download the data created in this study.

Results: The study revealed two distinct temporal phases of the EC response to TNF. The first phase, the acute response, occurred between 1-4 hours after TNF stimulation. During this phase, several previously uncharacterised genes were strongly regulated, among them lncRNAs and pseudogenes. The second phase, referred to as the later phase, occurred between 12-24 hours after TNF stimulation. In this phase, the majority of regulated genes were related to viral response, independent of *de novo* interferon production. Downregulation patterns were mostly not associated with inflammation GO related terms.

Conclusions: This study provides insights into the global dynamics of the endothelial cell transcriptional response to TNF. The analysis highlights distinct gene expression patterns during the acute and later phases of the response. Additionally, the findings suggest that TNF can regulate genes independent of interferon production. The identified genes and coexpression profiles may contribute to a better understanding of the role of endothelial cells in inflammation. Data for all analysed gene transcripts can be explored on the accompanying website http://www.endothelial-response.org/.

5.2 Project II

A human adipose tissue cell-type transcriptome atlas

Objectives: This human adipose tissue cell-type transcriptome atlas project uses an integrative correlation analysis to identify cell type enriched transcriptome profiles of adipose tissue constituent cell types from unfractionated tissue samples. In the context of this thesis, results from the endothelial profiling can be used to find candidate genes for further profiling.

Methods: An integrative correlation analysis was used to identify enriched genes in ten cell types using the adipose tissue data. Adipose tissue bulk RNAseq data was obtained from the GTEx web resource, and selected cell specific markers - reference transcripts, was used to identify other possible transcripts with similar expression pattern and ratio across the sample pool, of which could indicate enrichment within the corresponding cell type.

Results: Ten major constituent cell types were profiled in visceral human adipose tissue (VAT), based on 527 samples from male and female donors. More than 2,300 cell-type-enriched transcripts were identified. 8 constituent cell types were similarly profiled in subcutaneous adipose tissue (SAT) using and 646 samples. Comparative analysis revealed depot- and sex-specific cell-type-enriched genes. The variation between SAT and VAT gene expression profiles was primarily driven by mesothelial cells. Additionally, a panel of male-only cell-type-enriched genes was uncovered through sex-subset analysis. These were exclusively assigned to the Y-chromosome. KANK3 was found on the EC enriched list and was further investigated in study 3.

Conclusions: The generated transcriptome atlas serves as a roadmap for understanding the biology of adipose tissue, contributing to the identification of cell-type-specific genes, and is employed within the human protein atlas. Enriched genes can be targets for research in health and disease. The bulk sequencing approach overcomes technical challenges associated with single-cell RNA sequencing analysis of adipose tissue.

5.3 Project III

KANK3 is a shear stress regulated endothelial protein with a role in cell migration and tissue factor regulation

Objectives: The objective of this study was to functionally investigate the role of KANK3, a previously defined EC-enriched transcript, in endothelial cells. Considering that cell type specific genes have cell type specific functions, we investigated KANK3 in a vascular context, exploring coagulation, inflammation, and proliferation.

Methods: Human umbilical vein endothelial cells (HUVEC) were transfected with siRNAs targeting KANK3 and treated with or without the inflammatory cytokine TNFα. An in vitro scratch assay was performed to evaluate cell proliferation and migration. Quantitative polymerase chain, reaction (qPCR) was used to measure the expression of coagulation-related, inflammatory, and proliferation genes. Calibrated automated thrombinoscope (CAT) was used to evaluate changes thrombin generation. Protein localisation in HUVEC were analysed using fluorescence microscopy (SIM, confocal). Protein quantification was performed using western blot or flow cytometry.

Results: KANK3 knockdown was found to positively regulate the mRNA expression of coagulation factors tissue factor and the pro fibrinolytic factor PLAT. These findings were confirmed through the utilisation of a thrombin generation assay, wherein the reduction of KANK3 resulted in heightened thrombin production. In wound healing assays, the KANK3 siRNA-EC cells demonstrated a faster rate of gap closure, indicating enhanced cell migration, under both serum-starved and non-starved conditions, compared to the control group. KANK3 accumulation was observed in cell-cell interaction sites. KANK3 distribution and expression was regulated by flow.

Conclusions: The findings of this study suggest that KANK3 plays a role in the regulation of coagulation by modulating the expression of tissue factor in EC. Additionally, KANK3 appears to be important for the regulation of cell migration.

6 Discussion

6.1 Project I - TNF time course

In this study, we investigated the temporal response of endothelial cells (EC) to stimulation with TNF using RNAseq at 11 different time points up to 72 hours post-stimulation. To our knowledge, this is the first study in EC encompassing such a comprehensive temporal response to stimulus, along with a high sequencing depth. Existing studies focus on either specific gene(s) or phenotypic characteristics, e.g., (282, 283, 284), or global transcriptional response at a limited number of time points (285, 286).

We identified two distinct temporal phases of gene expression changes induced by TNF, as well as distinct types of downregulation profiles. The early phase of the response occurred between 1-4 hours after stimulation and included well-known genes involved in leukocyte adhesion, and other commonly described inflammatory markers. We identified several genes encoding uncharacterised proteins and uncharacterised non-coding transcripts, which showed similar expression dynamics during the initial phase, suggesting potential candidates for further investigation in the context of inflammation. While the early response included the expected TNF regulated genes, we also observed a delayed phase of response, that occurred between 12-24 hours post-stimulation, which primarily consisted of interferon-stimulated genes (ISGs). We also identified a panel of genes that were downregulated by TNF stimulation, including some that had not been previously reported in this context. The regulation of ISGs in EC by inflammatory cytokines is not well understood, but the findings of this study were consistent with a recent report suggesting the induction of a late-stage interferon response in EC following TNF stimulation (287). The expression of ISGs was independent of *de novo* production of interferon, contrary to previous reports that suggested that ISG expression is dependent on IFNβ (288, 289, 290).

We also found upregulation of genes involved in the cyclic GMP–AMP (cGAS) and STING pathway, which detect pathogenic DNA (291), as well as the upregulation of genes encoding various pattern recognition receptors, such as toll-like receptors and RIG-I-like receptors. These receptors are known to induce interferon production in response to bacterial or viral ligands, suggesting a potential mechanism for ISG

induction in EC (288, 292, 293, 294). This mechanism is likely a priming for antiviral response after infection. Additionally, we found several non-coding RNAs within the gene modules associated with ISGs, some of which have to-date unknown functions and could be interesting candidates for further study.

Furthermore, our study offers insight in the sex-based TNF response, comparing male and female cells. Studying sex differences, especially the female sex, is an often neglected variable in clinical and biological science. This gap in comprehending sex differences has led to healthcare inequality and life-threatening incidents. Approved drugs can have severe negative impacts on women due to male-centric biases in research across basic research the clinic (295). Sex even more neglected in *in vitro* studies, where the sex of primary cells is often not specified. A review of 295 articles from 2016 to 2017 published in ATVB, focusing on cells obtained from animal tissue, found that less than only 11-21 % of studies disclosed the sex, with only 3 articles comparing male and female cells (296). This study contrasts the response of male and female HUVEC after TNF stimulation. Our analysis revealed that the TNF-response was comparable between female and male HUVEC, but we saw differences on baseline and in magnitude of total response. Only a small set of genes was regulated between male and female and by TNF.

While we did not find significant differences in the EC response to TNF based on chromosomal sex, we found differences in baseline levels of gene expression. It must be noted that other factors, such as sex hormones, may influence vascular responses *in vivo*. It is also important to note that the study does not provide an overview of all aspects of the TNF response, such as immediate responders (e.g., secretion of P-selectin), protein phosphorylation or complex, and nuclear translocation, or any non-transcriptional aspect of TNF response.

We provide this data on a user-friendly website resource, the endothelial response database, which offers accessible information for researchers. The website is available under www.endothelial-response.org (Figure 35). It can serve as a valuable tool for analysing the dataset and exploring the extensive temporal coverage of the EC transcriptional response to TNF stimulation. Researchers can easily access the profiles of TNF-regulated genes or perform gene centric analyses. The resource also facilitates understanding of the global temporal context of transcriptomic changes through weighted network correlation analysis.
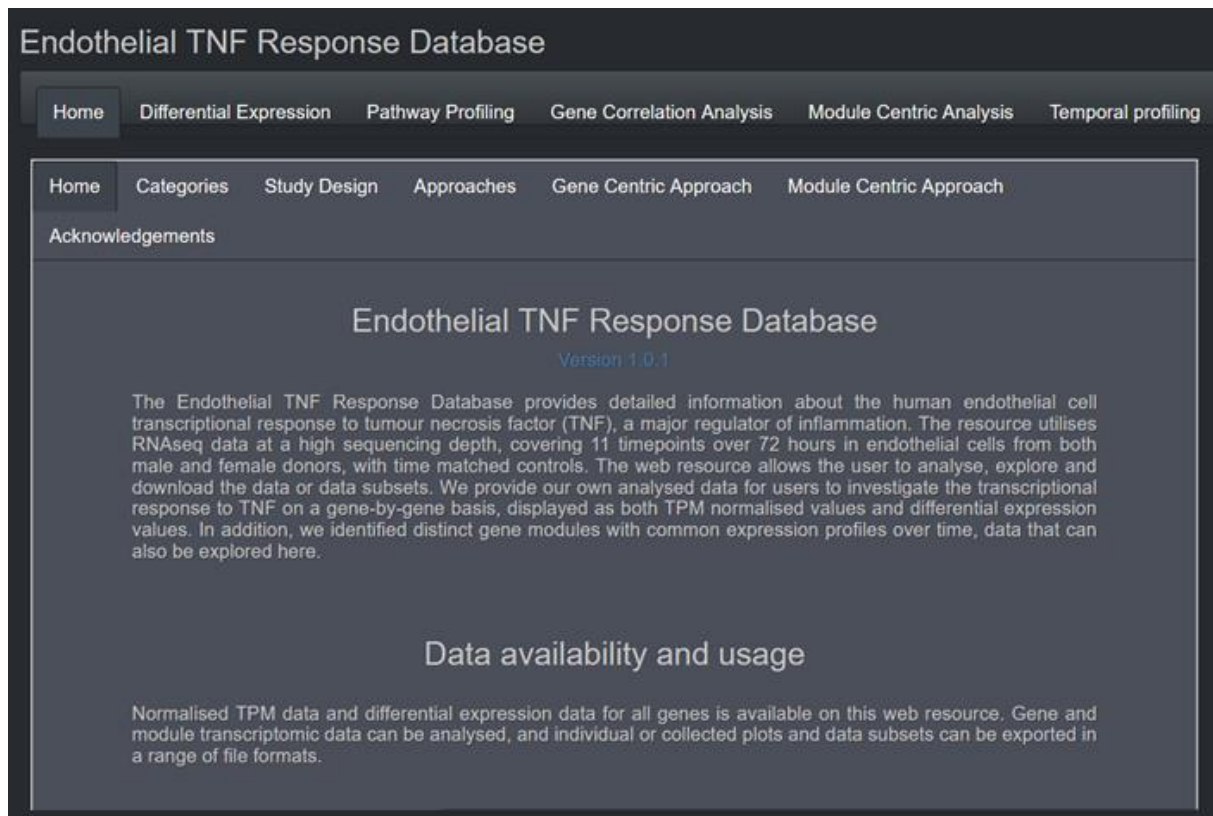
Figure 35: Endothelial-response website resource. *A key aim of study 1 was to provide a web resource for investigation of genes or gene groups in context of vascular inflammation. The resource expands the array of online resources by presenting a sex-matched transcriptional dataset of HUVEC response to TNF, along with a temporal baseline. It allows for different calculations without prior knowledge of molecular biology necessary.*

Limitations of the studies are among others, that EC is a foetal cell type, rather than adult. Discrepancies between adult and foetal cell types have been described before, among them variations in their transcriptional reaction to TNF (297). As described before, it should be mentioned that EC have organ-specific heterogeneity (298, 299). meaning that stimulation responses of other EC types or vascular bed might not be comparable to those we observed in HUVEC. Furthermore, the EC in this study were not cultured under flow or together with other cells of the same microenvironment, which could affect their transcriptome (300, 301, 302, 303) or the response to TNF (304). It should be noted that the proteome is not necessarily directly related to the transcriptome, and can also be temporarily delayed (305). While transcriptomics may reveal important regulatory elements, our study does not reveal changes in the proteome. Furthermore non-genetic regulatory elements, such as secreted proteins (e.g., P-selectin and Von Willebrand factor (306)) are not measured.

6.2 Adipose Tissue Transcriptome

This project aimed to study the gene expression profiles of different cell types within adipose tissue (AT) (105). AT serves various functions in the body, including energy storage, insulation, and regulation of metabolism (118). It consists primarily of white AT, categorised as visceral adipose tissue (VAT) located in the abdominal region and subcutaneous adipose tissue (SAT) located under the skin (121, 122, 123).

Previous studies have used single-cell RNA sequencing (scRNAseq) to analyse specific cell types within adipose tissue, but this approach has limitations, such as the need for fresh tissue and low read depth, as well as adipocytes themselves being difficult to profile (110, 113). This project aimed to overcome these challenges by analysing unfractionated tissue RNAseq data and identifying cell-type-enriched transcripts, using pre-existing data from GTEx (108). The substantial number of biological replicates allows for robust comparisons between different subgroups. We provide a comprehensive database of coding and non-coding gene expression profiles specific to both adipose tissue cell types.

We found several genes with enriched expression in adipocytes, including both known genes involved in adipocyte development and function, as well as previously uncharacterised genes. Several enriched non-coding RNAs, which have been increasingly recognised in biological regulation, were also identified. Sex specific differences were mainly driven by Y-chromosome genes. Gene expression profiles between VAT and SAT revealed differences at the cell-type level, with mesothelial cells being found to be a primary driver of these differences.

In silico approaches often come with limitations. Using our approach, several factors need to be taken into consideration.

- **Dependency on choice of reference transcript (Ref.T.):** The integrative correlation analysis relies on selecting appropriate reference transcripts that accurately represent the cell type under investigation. The choice of reference transcripts needs to be specific to the cell type they are intended to represent. It also needs to be a representative of its target cell population and needs to correspond to other reference transcripts. Otherwise, it is easy to miss out biological differences or use falsely annotated genes.

- **Limitation to known cell types:** The use of reference transcripts to identify cell type-enriched transcripts is limited to known cell types or even to well described ones.

- **Difficulty in profiling cell sub-types:** The method may face challenges in accurately profiling cell sub-types that do not have a distinct set of uniquely expressed genes. In cases where sub-types exhibit varying levels of the same genes, differentiation becomes more difficult using this method.

- **Statistical limitations:** Our threshold is not based on hard statistics, but rather on convenience factors, such as the arbitrary number of 0.5. A gene with a correlation of 0.50001 would be included, while ones with 0.49999 would be excluded. These limitations would reduce false positives. It was partially addressed by implementation of differentials. These differentials were intentionally set high, so that false negatives are more likely than false positives, as the aim was a gene list that is more positively enriched, rather than a larger list of more vaguely enriched genes.

- **Genes that are specific under stimulation:** Some genes that may not be expressed in resting state, but only in activated state, could be cell type enriched. These genes would not be found in our analysis.

This dataset can be used to optimise deconvolution algorithms used to determine the proportions of cell types in bulk RNAseq data. By employing an integrative correlation analysis, we successfully identified enrichment profiles of constituent cell types in 2 different adipose tissues.

The data is accessible through the "Tissue Cell Type" section of the Human Protein Atlas website: https://www.proteinatlas.org/humanproteome/tissue+cell+type.

6.3 KANK3

In this study, the functional characterisation of the uncharacterised, endothelial-enriched protein KANK3, was presented. It has been previously identified in several tissues as endothelial-enriched by our group (20, 105, 106, 107, 127) and as vasculature-specific for a homologue in zebrafish (130). In non-endothelial cells KANK3 is described as cancer enriched and influences cancer migration and proliferation (140). KANK3 belongs to an evolutionary well conserved protein family of four members with a unique structure: The N-terminal talin binding sites (KN motif), central coiled coil domains, which recognise liprin, and C-terminal ankyrin repeat units, which bind to KIF21A, a motor protein that binds to tubulin. Other family members have been described as focal adhesion proteins. While KANK1 and KANK2 are well described, KANK3 has remained understudied, especially in a vascular context, in vertebrates. Additionally, the HPA defines KANK3 as a cancer-associated protein that is highly expressed, although not exclusively, in rhabdoid cancer. It was found to be expressed in 28 out of 29 analysed cancer cell lines, with the exception of testicular cancer where its expression was not detected (129).

Our study of KANK3 focused primarily on vascular processes, such as inflammation, wound healing, and thrombosis, as well as cellular localisation and actin organisation. Based on the results in study 1, we found KANK3 expression enhanced after TNF stimulation (218) (Figure 36). This could indicate that KANK3 is directly or indirectly related to inflammatory processes, even though KANK3 knockdown does not influence the gene expression of inflammatory markers (SELE, ICAM1, VCAM1).
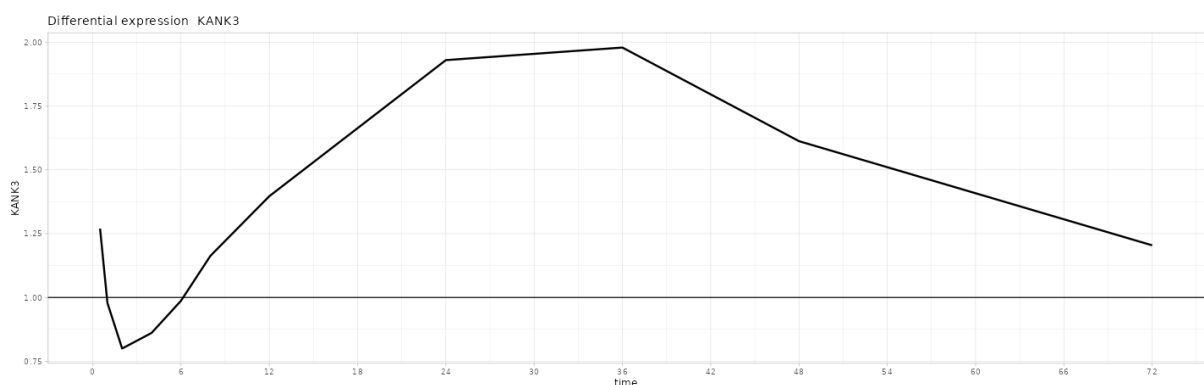

Figure 36: Temporal differential expression profile of KANK3 in HUVEC after TNF stimulation: *shows an enhanced differential profile. Employing the endothelial-response web resource from study 1, we investigated the temporal profile of KANK3, showing that its expression is enhanced compared to untreated control (218).*

74

Interestingly, however, we observed that knockdown of KANK3 enhances the expression of the prothrombotic coagulation factor tissue factor (F3) on mRNA and protein level, as well as profibrinolytic factor tPA. Our findings therefore suggest that KANK3 is a player in haemostatic balance via unknown mechanisms. A suppression of KANK3 in endothelial cells induced a shift towards a prothrombotic state in gene and protein expression. It induces a heightened procoagulant environment via mechanisms distinct from standard cellular activation, as evidenced by the absence of alteration in adhesion molecules. KANK3 could therefore be an interesting target for investigation in cardiovascular disease, such a thrombosis or atherosclerosis. A potentially prothrombotic phenotype *in vitro* is furthermore validated by functional experiments showing enhanced thrombin generation, following KANK3 inhibition, thereby providing additional evidence for these findings.

Previous studies have suggested that the KANK protein family plays a role in actin cytoskeletal organisation as a focal adhesion protein, as well as a migratory mediator in cancer cells. We conducted protein localisation studies of KANK3 in endothelial cells and found that KANK3 localises to focal adhesions, along the cytoskeleton, indicating its potential role in these cellular processes. Additionally, live-cell microscopy studies of endothelial cells with KANK3 knockdown showed increased cell motility without changes in cell proliferation. These results suggest that KANK3 has a role in modulating cellular migration, probably related to focal adhesion signalling.

In summary, the study sheds light on KANK3's potential significance in vascular biology and suggest a potential impact on thrombotic tendencies. These implications make KANK3 an interesting target for therapeutic interventions in cardiovascular disorders, such as atheroscleros and thrombosis. Further studies should be conducted to fully elucidate the intricate function and precise molecular mechanisms of KANK3 within the vasculature, including its precise localisation and its potential role as a therapeutic target for thrombotic disorders.

7 Conclusion

Study 1

Two distinct phases of EC gene expression patterns in response to TNF were identified through an extended time-course analysis (30 minutes to 72 hours): an acute response phase (1-4 hours) and a later phase (12-24 hours). Most other studies focus the analysis on a single time-point post stimulation, giving only a limited insight in the global response. The acute phase included previously studied genes associated with EC leukocyte adhesion receptors, as well as relevant uncharacterised genes. The later phase primarily consisted of TNF-induced expression of interferon-stimulated genes (ISGs), independent of de novo interferon production. Genes linked to pattern recognition receptors and non-coding RNAs within ISG-containing gene modules were also upregulated. Additionally, a novel panel of downregulated genes in response to TNF stimulation was identified. Furthermore, a web-based resource was created to analyse the defined genes and modules in this study.

Study 2

By employing integrative correlation analysis, enrichment profiles for various cell types in visceral and subcutaneous adipose tissue were predicted. The analysis revealed that these fat depots exhibited similar overall cell type expression profiles. The differences between depots were mainly driven by mesothelial cells. Only a small subset of sex-specific cell type enrichment differences were observed, all associated with Y-linked genes, enriched exclusively in males. The data from this study was implemented into the Human Protein Atlas. This study also provided candidate genes for further investigation, such as the endothelial enriched transcript KANK3.

Study 3

Building upon study 2, Study 3 delves into the functional significance of the predicted endothelial-enriched gene transcript KANK3. KANK3 is confirmed as an endothelial-enriched protein in multiple tissues. KANK3 expression is increased, and protein redistributed under flow. KANK3 inhibition is associated with higher cell motility, but not with higher proliferation rates. KANK3 has a potential role in the regulation of tissue factor and tissue plasminogen activator. Functional analyses showed that KANK3 knockdown increases thrombin generation. Thus, KANK3 might be involved in cell adhesion and intracellular signalling.

Collective conclusion

Collectively, these studies underscore the importance of multi-faceted approaches in endothelial research, ranging from comprehensive time-course analyses to integrative cell-type-specific investigations in tissues, and to functional analysis of specific proteins. Both project 1 and 2 can generate interesting targets for further investigation. When exploring different aspects of EC biology, interdisciplinary approaches can help and complement each other to advance our knowledge of complex biological systems. Furthermore, the development of web-based resources and the integration of data into broader platforms like the Human Protein Atlas can enhance the accessibility of these findings for the scientific community.

8 Future Perspectives

Project 1 investigated the temporal transcriptional response of HUVEC to TNF to identify regulated genes and coexpression profiles, as well as providing a web based resource. We identified two distinct response phases: an acute response phase and a later phase, involving IFN independent regulation of interferon stimulated genes, providing insights into the dynamics of endothelial cell transcriptional response to TNF. Future studies could include the functional analysis of previously uncharacterised genes in both acute and late responses, as well as a more comprehensive exploration of the secondary ISG regulation and uncover novel endothelial roles in inflammation and haemostasis. Further studies will deal with the expansion of the website resource and usage of different stimuli (such as interferon, shear stress and hypoxia), introduce new cell types (e.g., monocytes or different EC types), and investigate miRNAs, as well as investigation of interesting target genes, such as the late stage responders related to interferon, in the dataset. A comparative downstream-transcriptional map could be created, mapping gene-gene interactions using the data created in both study 1 and 2. Interesting targets could be adipose tissue specific EC transcripts defined in study 2. Future improvements could involve expanding and updating the web resource with additional data, integrating other relevant datasets, as well as incorporating advanced analysis tools and visualisations. Additionally, we are currently in the process of expanding the web resource by focussing on the untreated base line and shifts in gene expression over culture time.

Project 2 employed the analysis of bulk sequencing data from the GTEx portal to analyse adipose tissue. Follow-up and sister studies in our group have already been concluded and we defined cell type specific transcriptome profiles for, gastrointestinal tract, stomach, and the cell type enriched transcriptome atlas for 15 human tissues. Follow up projects could include the further investigation of non-coding transcripts or the analysis of the found tissue-enriched EC types in a clinical or *in vitro* context, as well as function or phenotype specific atlas projects, e.g., a body-wide fenestrated endothelial transcriptome.

Project 3 presented the endothelial-enriched transcript KANK3 and showed results in regard to regulation of tissue factor expression and plasminogen activator tissue type expression after knockdown. It also influences cell motility. Although highly interesting,

we could not establish if these were a direct result of the knockdown or a confounding effect by the influence of KANK3 on mechanosensing or actin assembly. Follow-up studies could explore this effect and help to widen our understanding of cardiovascular disease and wound healing. They could include investigating its response when cultured under oscillating flow conditions simulating disturbed flow observed in atherosclerosis and thrombosis. Moreover, exploring the interaction between focal adhesion kinase (FAK) and cadherin, a regulator of intercellular endothelial junctions, could provide insights into KANK3's role in LC transmigration speed and LC adhesion. Performing RNA sequencing after KANK3 knockdown could reveal potential gene expression changes associated with KANK3's absence, shedding light on its downstream effects. Conducting a vinculin knockdown to assess KANK3 levels and localization could uncover the relationship between KANK3 and vinculin, a main regulator of recruitment to focal adhesions. Similarly, the effect of inhibition of other focal adhesion regulators could be of interest. Considering KANK3's potential influence on actin assembly and therefore arterial stiffness, and cell spreading in the context of atherosclerosis might provide valuable insights into its role in regulating vascular health. Atomic Force Microscopy (AFM) could be employed to explore this aspect. The focus of the haemostatic system lied on tissue factor and thrombin generation, as we were not able to assess tPA at the time. Therefore, a reasonable follow-up could dive into the change in fibrin deposition and fibrinolysis.

Together, these projects and their follow-ups hold the potential for collaborative utilisation. Their insights can synergistically enhance each other in multiple ways: Projects 1 and 2 can be employed to uncover potential connections between tissue-specific endothelial cells and their temporal transcription profiles, thereby identifying additional EC specific targets for further investigation. Alternatively, the focused investigation of particular targets, as shown in project 3, within the context of tissue specificity or stimulation-induced dynamics could be of interest. Temporal enrichment profiles between different, EC and non-EC, cell types could help to create an endothelial enriched atlas of immune responders or in pathogenic states.

In these ways, the collaborative integration of findings from different types of projects can help to expand our understanding of endothelial cell biology.

9 References

1.   Milutinović A, Šuput D, Zorc-Pleskovič R. Pathogenesis of atherosclerosis in the tunica intima, media, and adventitia of coronary arteries: An updated review. Bosn J Basic Med Sci. 2020;20(1):21-30.

2.   Krüger-Genge A, Blocki A, Franke RP, Jung F. Vascular Endothelial Cell Biology: An Update. Int J Mol Sci. 2019;20(18).

3.   Brodsky SV, Goligorsky MS. Endothelium under stress: local and systemic messages. Semin Nephrol. 2012;32(2):192-8.

4.   Galley HF, Webster NR. Physiology of the endothelium. Br J Anaesth. 2004;93(1):105-13.

5.   Yau JW, Teoh H, Verma S. Endothelial cell control of thrombosis. BMC Cardiovasc Disord. 2015;15:130.

6.   Ley K, Reutershan J. Leucocyte-endothelial interactions in health and disease. Handb Exp Pharmacol. 2006(176 Pt 2):97-133.

7.   Sandoo A, van Zanten JJ, Metsios GS, Carroll D, Kitas GD. The endothelium and its role in regulating vascular tone. Open Cardiovasc Med J. 2010;4:302-12.

8.   Eilken HM, Adams RH. Dynamics of endothelial cell behavior in sprouting angiogenesis. Current Opinion in Cell Biology. 2010;22(5):617-25.

9.   Faulkner A. Trans-endothelial trafficking of metabolic substrates and its importance in cardio-metabolic disease. Biochemical Society Transactions. 2021;49(1):507-17.

10.   Rajendran P, Rengarajan T, Thangavel J, Nishigaki Y, Sakthisekaran D, Sethi G, et al. The vascular endothelium and human diseases. Int J Biol Sci. 2013;9(10):1057-69.

11.   Durand MJ, Gutterman DD. Diversity in mechanisms of endothelium-dependent vasodilation in health and disease. Microcirculation. 2013;20(3):239-47.

12.   Hadi HA, Carr CS, Al Suwaidi J. Endothelial dysfunction: cardiovascular risk factors, therapy, and outcome. Vasc Health Risk Manag. 2005;1(3):183-98.

13.   Poredos P, Jezovnik MK. Endothelial Dysfunction and Venous Thrombosis. Angiology. 2018;69(7):564-7.

14.   Aird WC. Phenotypic Heterogeneity of the Endothelium. Circulation Research. 2007;100(2):158-73.

15.   Atkins GB, Jain MK, Hamik A. Endothelial Differentiation. Arteriosclerosis, Thrombosis, and Vascular Biology. 2011;31(7):1476-84.

16.   Szafranska K, Kruse LD, Holte CF, McCourt P, Zapotoczny B. The wHole Story About Fenestrations in LSEC. Front Physiol. 2021;12:735573.

17.   Parab S, Quick RE, Matsuoka RL. Endothelial cell-type-specific molecular requirements for angiogenesis drive fenestrated vessel development in the brain. eLife. 2021;10:e64295.

18.   Lother A, Bergemann S, Deng L, Moser M, Bode C, Hein L. Cardiac Endothelial Cell Transcriptome. Arterioscler Thromb Vasc Biol. 2018;38(3):566-74.

19.   Hennigs JK, Matuszcak C, Trepel M, Körbelin J. Vascular Endothelial Cells: Heterogeneity and Targeting Approaches. Cells. 2021;10(10):2712.

20.   Butler LM, Hallstrom BM, Fagerberg L, Ponten F, Uhlen M, Renne T, et al. Analysis of Body-wide Unfractionated Tissue Data to Identify a Core Human Endothelial Transcriptome. Cell Syst. 2016;3(3):287-301 e3.

21.   Aird WC. Endothelial cell heterogeneity. Cold Spring Harb Perspect Med. 2012;2(1):a006429.

22.   Lee BK, Bhinge AA, Battenhouse A, McDaniell RM, Liu Z, Song L, et al. Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. Genome Res. 2012;22(1):9-24.

23.   Zapotoczny B, Braet F, Kus E, Ginda-Mäkelä K, Klejevskaja B, Campagna R, et al. Actin-spectrin scaffold supports open fenestrae in liver sinusoidal endothelial cells. Traffic. 2019;20(12):932-42.

24.   Svistounov D, Warren A, McNerney GP, Owen DM, Zencak D, Zykova SN, et al. The Relationship between fenestrations, sieve plates and rafts in liver sinusoidal endothelial cells. PLoS One. 2012;7(9):e46134.

25.   Lüscher TF, Tanner FC. Endothelial regulation of vascular tone and growth. Am J Hypertens. 1993;6(7 Pt 2):283s-93s.

26.   Peiró C, Redondo J, Rodríguez-Martínez MA, Angulo J, Marín J, Sánchez-Ferrer CF. Influence of Endothelium on Cultured Vascular Smooth Muscle Cell Proliferation. Hypertension. 1995;25(4):748-51.

27.   Cooke JP, Tsao PS. Go With the Flow. Circulation. 2001;103(23):2773-5.

28.   Burnstock G. Endothelium-Derived Vasoconstriction by Purines and Pyrimidines. Circulation Research. 2008;103(10):1056-7.

29.   Hellenthal KEM, Brabenec L, Wagner NM. Regulation and Dysregulation of Endothelial Permeability during Systemic Inflammation. Cells. 2022;11(12).

30.   Mehta D, Malik AB. Signaling mechanisms regulating endothelial permeability. Physiol Rev. 2006;86(1):279-367.

31.   Sanchez-Cano F, Hernandez L, Ortega A. The Blood–Brain Barrier: Much More Than a Selective Access to the Brain. Neurotoxicity Research. 2021;39:1-21.

32.   Wallez Y, Huber P. Endothelial adherens and tight junctions in vascular homeostasis, inflammation and angiogenesis. Biochimica et Biophysica Acta (BBA) - Biomembranes. 2008;1778(3):794-809.

33.   Komarova YA, Kruse K, Mehta D, Malik AB. Protein Interactions at Endothelial Junctions and Signaling Mechanisms Regulating Endothelial Permeability. Circulation Research. 2017;120(1):179-206.

34.   Duong CN, Vestweber D. Mechanisms Ensuring Endothelial Junction Integrity Beyond VE-Cadherin. Frontiers in Physiology. 2020;11.

35.   Dejana E, Spagnuolo R, Bazzoni G. Interendothelial junctions and their role in the control of angiogenesis, vascular permeability and leukocyte transmigration. Thromb Haemost. 2001;86(1):308-15.

36.   Bazzoni G, Dejana E. Endothelial cell-to-cell junctions: molecular organization and role in vascular homeostasis. Physiol Rev. 2004;84(3):869-901.

37. Davis GE, Senger DR. Endothelial Extracellular Matrix. Circulation Research. 2005;97(11):1093-107.

38. Ambade AS, Hassoun PM, Damico RL. Basement Membrane Extracellular Matrix Proteins in Pulmonary Vascular and Right Ventricular Remodeling in Pulmonary Hypertension. Am J Respir Cell Mol Biol. 2021;65(3):245-58.

39. Jandl K, Marsh LM, Hoffmann J, Mutgan AC, Baum O, Bloch W, et al. Basement Membrane Remodeling Controls Endothelial Function in Idiopathic Pulmonary Arterial Hypertension. American Journal of Respiratory Cell and Molecular Biology. 2020;63(1):104-17.

40. Kong X, Kapustka A, Sullivan B, Schwarz GJ, Leckband DE. Extracellular matrix regulates force transduction at VE-cadherin junctions. Molecular Biology of the Cell. 2022;33(11):ar95.

41. Reitsma S, Slaaf D, Vink H, Zandvoort M, oude Egbrink M. The endothelial glycocalyx: composition, functions, and visualization. Pflügers Archiv : European journal of physiology. 2007;454:345-59.

42. Villalba N, Baby S, Yuan SY. The Endothelial Glycocalyx as a Double-Edged Sword in Microvascular Homeostasis and Pathogenesis. Front Cell Dev Biol. 2021;9:711003.

43. Hu Z, Cano I, D'Amore PA. Update on the Role of the Endothelial Glycocalyx in Angiogenesis and Vascular Inflammation. Front Cell Dev Biol. 2021;9:734276.

44. Weinbaum S, Tarbell JM, Damiano ER. The Structure and Function of the Endothelial Glycocalyx Layer. Annual Review of Biomedical Engineering. 2007;9(1):121-67.

45. Kolářová H, Víteček J, Černá A, Černík M, Přibyl J, Skládal P, et al. Myeloperoxidase mediated alteration of endothelial function is dependent on its cationic charge. Free Radic Biol Med. 2021;162:14-26.

46. Uchimido R, Schmidt EP, Shapiro NI. The glycocalyx: a novel diagnostic and therapeutic target in sepsis. Critical Care. 2019;23(1):16.

47. Chappell D, Hofmann-Kiefer K, Jacob M, Rehm M, Briegel J, Welsch U, et al. TNF-alpha induced shedding of the endothelial glycocalyx is prevented by hydrocortisone and antithrombin. Basic Res Cardiol. 2009;104(1):78-89.

48. Kadl A, Leitinger N. The role of endothelial cells in the resolution of acute inflammation. Antioxid Redox Signal. 2005;7(11-12):1744-54.

49. Pober JS, Sessa WC. Evolving functions of endothelial cells in inflammation. Nature Reviews Immunology. 2007;7(10):803-15.

50. Ley K, Laudanna C, Cybulsky MI, Nourshargh S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. Nat Rev Immunol. 2007;7(9):678-89.

51. Bevilacqua MP, Nelson RM, Mannori G, Cecconi O. Endothelial-leukocyte adhesion molecules in human disease. Annu Rev Med. 1994;45:361-78.

52. Cambien B, Wagner DD. A new role in hemostasis for the adhesion receptor P-selectin. Trends Mol Med. 2004;10(4):179-86.

53.	Carlos TM, Harlan JM. Leukocyte-Endothelial Adhesion Molecules. Blood. 1994;84(7):2068-101.

54.	Johnston B, Butcher EC. Chemokines in rapid leukocyte adhesion triggering and migration. Semin Immunol. 2002;14(2):83-92.

55.	Nourshargh S, Krombach F, Dejana E. The role of JAM-A and PECAM-1 in modulating leukocyte infiltration in inflamed and ischemic tissues. J Leukoc Biol. 2006;80(4):714-8.

56.	van Buul JD, van Alphen FP, Hordijk PL. The presence of alpha-catenin in the VE-cadherin complex is required for efficient transendothelial migration of leukocytes. International Journal of Biological Sciences. 2009;5(7):695-705.

57.	Bradfield PF, Nourshargh S, Aurrand-Lions M, Imhof BA. JAM Family and Related Proteins in Leukocyte Migration (Vestweber Series). Arteriosclerosis, Thrombosis, and Vascular Biology. 2007;27(10):2104-12.

58.	Shaw SK, Bamba PS, Perkins BN, Luscinskas FW. Real-Time Imaging of Vascular Endothelial-Cadherin During Leukocyte Transmigration Across Endothelium1. The Journal of Immunology. 2001;167(4):2323-30.

59.	Nourshargh S, Alon R. Leukocyte migration into inflamed tissues. Immunity. 2014;41(5):694-707.

60.	Mackay F, Loetscher H, Stueber D, Gehr G, Lesslauer W. Tumor necrosis factor alpha (TNF-alpha)-induced cell adhesion to human endothelial cells is under dominant control of one TNF receptor type, TNF-R55. J Exp Med. 1993;177(5):1277-86.

61.	Imaizumi T, Itaya H, Fujita K, Kudoh D, Kudoh S, Mori K, et al. Expression of Tumor Necrosis Factor-α in Cultured Human Endothelial Cells Stimulated With Lipopolysaccharide or Interleukin-1α. Arteriosclerosis, Thrombosis, and Vascular Biology. 2000;20(2):410-5.

62.	Pober JS. Endothelial activation: intracellular signaling pathways. Arthritis Research & Therapy. 2002;4(3):S109.

63.	Xia P, Gamble JR, Rye KA, Wang L, Hii CS, Cockerill P, et al. Tumor necrosis factor-alpha induces adhesion molecule expression through the sphingosine kinase pathway. Proc Natl Acad Sci U S A. 1998;95(24):14196-201.

64.	Swiecki M, Colonna M. Type I interferons: diversity of sources, production pathways and effects on immune responses. Curr Opin Virol. 2011;1(6):463-75.

65.	Grandvaux N, tenOever BR, Servant MJ, Hiscott J. The interferon antiviral response: from viral invasion to evasion. Curr Opin Infect Dis. 2002;15(3):259-67.

66.	Laschke MW, Gu Y, Menger MD. Replacement in angiogenesis research: Studying mechanisms of blood vessel development by animal-free in vitro, in vivo and in silico approaches. Frontiers in Physiology. 2022;13.

67.	Luzuriaga J, Irurzun J, Irastorza I, Unda F, Ibarretxe G, Pineda JR. Vasculogenesis from Human Dental Pulp Stem Cells Grown in Matrigel with Fully Defined Serum-Free Culture Media. Biomedicines. 2020;8(11).

68.	Siemerink MJ, Klaassen I, Van Noorden CJ, Schlingemann RO. Endothelial tip cells in ocular angiogenesis: potential target for anti-angiogenesis therapy. J Histochem Cytochem. 2013;61(2):101-15.

69.   Moccia F, Negri S, Shekha M, Faris P, Guerra G. Endothelial Ca(2+) Signaling, Angiogenesis and Vasculogenesis: just What It Takes to Make a Blood Vessel. Int J Mol Sci. 2019;20(16).

70.   Neubauer K, Zieger B. Endothelial cells and coagulation. Cell Tissue Res. 2022;387(3):391-8.

71.   Becker BF, Heindl B, Kupatt C, Zahler S. Endothelial function and hemostasis. Z Kardiol. 2000;89(3):160-7.

72.   Wagner DD. New Links Between Inflammation and Thrombosis. Arteriosclerosis, Thrombosis, and Vascular Biology. 2005;25(7):1321-4.

73.   Kohli S, Shahzad K, Jouppila A, Holthöfer H, Isermann B, Lassila R. Thrombosis and Inflammation-A Dynamic Interplay and the Role of Glycosaminoglycans and Activated Protein C. Front Cardiovasc Med. 2022;9:866751.

74.   Li Y-H, Kuo C-H, Shi G-Y, Wu H-L. The role of thrombomodulin lectin-like domain in inflammation. Journal of Biomedical Science. 2012;19(1):34.

75.   Mohan Rao LV, Esmon CT, Pendurthi UR. Endothelial cell protein C receptor: a multiliganded and multifunctional receptor. Blood. 2014;124(10):1553-62.

76.   Ducros E, Mirshahi S, Bermot C, Mirshahi M. Analysis of endothelial protein C receptor functionality on living cells'. Open Biochem J. 2009;3:49-54.

77.   Corral J, González-Conejero R, Hernández-Espinosa D, Vicente V. Protein Z/Z-dependent protease inhibitor (PZ/ZPI) anticoagulant system and thrombosis. Br J Haematol. 2007;137(2):99-108.

78.   Han X, Fiehler R, Broze GJ, Jr. Isolation of a protein Z-dependent plasma protease inhibitor. Proc Natl Acad Sci U S A. 1998;95(16):9250-5.

79.   Chan TK, Chan V. Antithrombin III, the major modulator of intravascular coagulation, is synthesized by human endothelial cells. Thromb Haemost. 1981;46(2):504-6.

80.   Sarrazin S, Lamanna WC, Esko JD. Heparan sulfate proteoglycans. Cold Spring Harb Perspect Biol. 2011;3(7).

81.   Mast AE. Tissue Factor Pathway Inhibitor. Arteriosclerosis, Thrombosis, and Vascular Biology. 2016;36(1):9-14.

82.   van Hinsbergh VW. Endothelium--role in regulation of coagulation and inflammation. Semin Immunopathol. 2012;34(1):93-106.

83.   Krejcy K, Schmetterer L, Kastner J, Nieszpaur-Los M, Monitzer B, Schütz W, et al. Role of Nitric Oxide in Hemostatic System Activation In Vivo in Humans. Arteriosclerosis, Thrombosis, and Vascular Biology. 1995;15(11):2063-7.

84.   Chirkov YY, Nguyen TH, Horowitz JD. Impairment of Anti-Aggregatory Responses to Nitric Oxide and Prostacyclin: Mechanisms and Clinical Implications in Cardiovascular Disease. International Journal of Molecular Sciences. 2022;23(3):1042.

85.   Oliveira-Paula GH, Lacchini R, Tanus-Santos JE. Endothelial nitric oxide synthase: From biochemistry and gene structure to clinical implications of NOS3 polymorphisms. Gene. 2016;575(2 Pt 3):584-99.

86.   Stafford NP, Pink AE, White AE, Glenn JR, Heptinstall S. Mechanisms Involved in Adenosine Triphosphate–Induced Platelet Aggregation in Whole Blood. Arteriosclerosis, Thrombosis, and Vascular Biology. 2003;23(10):1928-33.

87.   Collen D, Lijnen HR. The Tissue-Type Plasminogen Activator Story. Arteriosclerosis, Thrombosis, and Vascular Biology. 2009;29(8):1151-5.

88.   K. K. Wu MD, and, P. Thiagarajan MD. ROLE OF ENDOTHELIUM IN THROMBOSIS AND HEMOSTASIS. Annual Review of Medicine. 1996;47(1):315-31.

89.   Turner NA, Moake JL. Factor VIII Is Synthesized in Human Endothelial Cells, Packaged in Weibel-Palade Bodies and Secreted Bound to ULVWF Strings. PLoS One. 2015;10(10):e0140740.

90.   Bhopale GM, Nanda RK. Blood coagulation factor VIII: An overview. J Biosci. 2003;28(6):783-9.

91.   Mackman N. Role of Tissue Factor in Hemostasis, Thrombosis, and Vascular Development. Arteriosclerosis, Thrombosis, and Vascular Biology. 2004;24(6):1015-22.

92.   Cesari M, Pahor M, Incalzi RA. Plasminogen activator inhibitor-1 (PAI-1): a key factor linking fibrinolysis and age-related subclinical and clinical conditions. Cardiovasc Ther. 2010;28(5):e72-91.

93.   ADAMS RLC, BIRD RJ. Review article: Coagulation cascade and therapeutics update: Relevance to nephrology. Part 1: Overview of coagulation, thrombophilias and history of anticoagulants. Nephrology. 2009;14(5):462-70.

94.   Palta S, Saroa R, Palta A. Overview of the coagulation system. Indian J Anaesth. 2014;58(5):515-23.

95.   Gale AJ. Continuing education course #2: current understanding of hemostasis. Toxicol Pathol. 2011;39(1):273-80.

96.   Smith SA, Travers RJ, Morrissey JH. How it all starts: Initiation of the clotting cascade. Crit Rev Biochem Mol Biol. 2015;50(4):326-36.

97.   Macfarlane RG. AN ENZYME CASCADE IN THE BLOOD CLOTTING MECHANISM, AND ITS FUNCTION AS A BIOCHEMICAL AMPLIFIER. Nature. 1964;202:498-9.

98.   Renné T, Schmaier AH, Nickel KF, Blombäck M, Maas C. In vivo roles of factor XII. Blood. 2012;120(22):4296-303.

99.   Cimmino G, Ciccarelli G, Golino P. Role of Tissue Factor in the Coagulation Network. Semin Thromb Hemost. 2015;41(7):708-17.

100.   Reyes Gil M. Chapter 91 - Overview of the Coagulation System. In: Shaz BH, Hillyer CD, Reyes Gil M, editors. Transfusion Medicine and Hemostasis (Third Edition): Elsevier; 2019. p. 559-64.

101.   Cesarman-Maus G, Hajjar KA. Molecular mechanisms of fibrinolysis. British Journal of Haematology. 2005;129(3):307-21.

102.   Favresse J, Lippi G, Roy P-M, Chatelain B, Jacqmin H, ten Cate H, et al. D-dimer: Preanalytical, analytical, postanalytical variables, and clinical applications. Critical Reviews in Clinical Laboratory Sciences. 2018;55(8):548-77.

103. Meltzer ME, Doggen CJM, de Groot PG, Rosendaal FR, Lisman T. The Impact of the Fibrinolytic System on the Risk of Venous and Arterial Thrombosis. Semin Thromb Hemost. 2009;35(05):468-77.

104. Saito H, Hamilton SM, Tavill AS, Louis L, Ratnoff OD. Production and release of plasminogen by isolated perfused rat liver. Proc Natl Acad Sci U S A. 1980;77(11):6837-40.

105. Norreen-Thorsen M, Struck EC, Öling S, Zwahlen M, Von Feilitzen K, Odeberg J, et al. A human adipose tissue cell-type transcriptome atlas. Cell Rep. 2022;40(2):111046.

106. Dusart P, Hallström BM, Renné T, Odeberg J, Uhlén M, Butler LM. A Systems-Based Map of Human Brain Cell-Type Enriched Genes and Malignancy-Associated Endothelial Changes. Cell Rep. 2019;29(6):1690-706.e4.

107. Dusart P, Öling S, Struck E, Norreen-Thorsen M, Zwahlen M, Feilitzen Kv, et al. A tissue centric atlas of cell type transcriptome enrichment signatures. bioRxiv. 2023:2023.01.10.520698.

108. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nature Genetics. 2013;45(6):580-5.

109. Lu J, Zhao J, Meng H, Zhang X. Adipose Tissue-Resident Immune Cells in Obesity and Type 2 Diabetes. Frontiers in Immunology. 2019;10.

110. Acosta JR, Joost S, Karlsson K, Ehrlund A, Li X, Aouadi M, et al. Single cell transcriptomics suggest that human adipocyte progenitor cells constitute a homogeneous cell population. Stem Cell Research & Therapy. 2017;8(1):250.

111. Vijay J, Gauthier M-F, Biswell RL, Louiselle DA, Johnston JJ, Cheung WA, et al. Single-cell analysis of human adipose tissue identifies depot- and disease-specific cell types. Nature Metabolism. 2020;2(1):97-109.

112. Estève D, Boulet N, Belles C, Zakaroff-Girard A, Decaunes P, Briot A, et al. Lobular architecture of human adipose tissue defines the niche and fate of progenitor cells. Nature Communications. 2019;10(1):2549.

113. Viswanadha S, Londos C. Optimized conditions for measuring lipolysis in murine primary adipocytes. Journal of Lipid Research. 2006;47(8):1859-64.

114. Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature. 2018;562(7727):367-72.

115. Consortium* TTS, Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, et al. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science. 2022;376(6594):eabl4896.

116. Hildreth AD, Ma F, Wong YY, Sun R, Pellegrini M, O'Sullivan TE. Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity. Nature Immunology. 2021;22(5):639-53.

117. Kim N, Kang H, Jo A, Yoo SA, Lee HO. Perspectives on single-nucleus RNA sequencing in different cell types and tissues. J Pathol Transl Med. 2023;57(1):52-9.

118. Kahn CR, Wang G, Lee KY. Altered adipose tissue and adipocyte function in the pathogenesis of metabolic syndrome. The Journal of Clinical Investigation. 2019;129(10):3990-4000.

119.   Cuthbertson DP, Tompsett SL. The degree of unsaturation of the fats of human adipose tissue in relation to depth from skin surface. Biochem J. 1933;27(4):1103-6.

120.   Richard AJ WU, Elks CM, et al. Adipose Tissue: Physiology to Metabolic Dysfunction. In: Feingold KR AB, Blackman MR, et al., editors, editor. Endotext South Dartmouth (MA)2020.

121.   Mittal B. Subcutaneous adipose tissue & visceral adipose tissue. Indian J Med Res. 2019;149(5):571-3.

122.   Rosell M, Kaforou M, Frontini A, Okolo A, Chan YW, Nikolopoulou E, et al. Brown and white adipose tissues: intrinsic differences in gene expression and response to cold exposure in mice. Am J Physiol Endocrinol Metab. 2014;306(8):E945-64.

123.   Cypess AM, Kahn CR. The role and importance of brown adipose tissue in energy homeostasis. Curr Opin Pediatr. 2010;22(4):478-84.

124.   Britton KA, Massaro JM, Murabito JM, Kreger BE, Hoffmann U, Fox CS. Body Fat Distribution, Incident Cardiovascular Disease, Cancer, and All-Cause Mortality. Journal of the American College of Cardiology. 2013;62(10):921-5.

125.   Chait A, den Hartigh LJ. Adipose Tissue Distribution, Inflammation and Its Metabolic Consequences, Including Diabetes and Cardiovascular Disease. Frontiers in Cardiovascular Medicine. 2020;7.

126.   Lumish HS, O'Reilly M, Reilly MP. Sex Differences in Genomic Drivers of Adipose Distribution and Related Cardiometabolic Disorders. Arteriosclerosis, Thrombosis, and Vascular Biology. 2020;40(1):45-60.

127.   Öling S, Struck E, Thorsen M, Zwahlen M, Feilitzen Kv, Odeberg J, et al. A human stomach cell type transcriptome atlas. bioRxiv. 2023:2023.01.10.520700.

128.   Guo SS, Seiwert A, Szeto IYY, Fässler R. Tissue distribution and subcellular localization of the family of Kidney Ankyrin Repeat Domain (KANK) proteins. Exp Cell Res. 2021;398(1):112391.

129.   Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.

130.   Boggetti B, Jasik J, Takamiya M, Strähle U, Reugels AM, Campos-Ortega JA. NBP, a zebrafish homolog of human Kank3, is a novel Numb interactor essential for epidermal integrity and neurulation. Dev Biol. 2012;365(1):164-74.

131.   Kakinuma N, Zhu Y, Wang Y, Roy BC, Kiyama R. Kank proteins: structure, functions and diseases. Cell Mol Life Sci. 2009;66(16):2651-9.

132.   Zhu Y, Kakinuma N, Wang Y, Kiyama R. Kank proteins: a new family of ankyrin-repeat domain-containing proteins. Biochim Biophys Acta. 2008;1780(2):128-33.

133.   Weng Z, Shang Y, Yao D, Zhu J, Zhang R. Structural analyses of key features in the KANK1·KIF21A complex yield mechanistic insights into the cross-talk between microtubules and the cell cortex. Journal of Biological Chemistry. 2018;293(1):215-25.

134.   Bouchet BP, Gough RE, Ammon Y-C, van de Willige D, Post H, Jacquemet G, et al. Talin-KANK1 interaction controls the recruitment of cortical microtubule stabilizing complexes to focal adhesions. eLife. 2016;5:e18124.

135.   Guo Q, Liao S, Zhu Z, Li Y, Li F, Xu C. Structural basis for the recognition of kinesin family member 21A (KIF21A) by the ankyrin domains of KANK1 and KANK2 proteins. J Biol Chem. 2018;293(2):557-66.

136.   Luo M, Mengos AE, Mandarino LJ, Sekulic A. Association of liprin β-1 with kank proteins in melanoma. Exp Dermatol. 2016;25(4):321-3.

137.   Sarkar S, Roy BC, Hatano N, Aoyagi T, Gohji K, Kiyama R. A novel ankyrin repeat-containing gene (Kank) located at 9p24 is a growth suppressor of renal cell carcinoma. J Biol Chem. 2002;277(39):36585-91.

138.   Kakinuma N, Roy BC, Zhu Y, Wang Y, Kiyama R. Kank regulates RhoA-dependent formation of actin stress fibers and cell migration via 14-3-3 in PI3K-Akt signaling. J Cell Biol. 2008;181(3):537-49.

139.   Roy BC, Kakinuma N, Kiyama R. Kank attenuates actin remodeling by preventing interaction between IRSp53 and Rac1. J Cell Biol. 2009;184(2):253-67.

140.   Dai Z, Xie B, Yang B, Chen X, Hu C, Chen Q. KANK3 mediates the p38 MAPK pathway to regulate the proliferation and invasion of lung adenocarcinoma cells. Tissue Cell. 2023;80:101974.

141.   Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode M R, Armean Irina M, et al. Ensembl 2022. Nucleic Acids Research. 2021;50(D1):D988-D95.

142.   Burridge K. Focal adhesions: a personal perspective on a half century of progress. The FEBS Journal. 2017;284(20):3355-61.

143.   Hoffmann M, Schwarz U. A kinetic model for RNA-interference of focal adhesions. BMC systems biology. 2013;7:2.

144.   Yoon J-Y. Focal Adhesion. Tissue Engineering: A Primer with Laboratory Demonstrations. Cham: Springer International Publishing; 2022. p. 123-35.

145.   Yu M, Le S, Ammon Y-C, Goult BT, Akhmanova A, Yan J. Force-Dependent Regulation of Talin–KANK1 Complex at Focal Adhesions. Nano Letters. 2019;19(9):5982-90.

146.   Rosen GD, Dube DS. ADHESION, CELL–MATRIX | Focal Contacts and Signaling. In: Laurent GJ, Shapiro SD, editors. Encyclopedia of Respiratory Medicine. Oxford: Academic Press; 2006. p. 41-7.

147.   Schwartz MA. Integrins and extracellular matrix in mechanotransduction. Cold Spring Harb Perspect Biol. 2010;2(12):a005066.

148.   Peng X, Nelson ES, Maiers JL, DeMali KA. Chapter five - New Insights into Vinculin Function and Regulation. In: Jeon KW, editor. International Review of Cell and Molecular Biology. 287: Academic Press; 2011. p. 191-231.

149.   Goult BT, Yan J, Schwartz MA. Talin as a mechanosensitive signaling hub. J Cell Biol. 2018;217(11):3776-84.

150.   Turner CE. Paxillin and focal adhesion signalling. Nature Cell Biology. 2000;2(12):E231-E6.

151.   Yamamura M, Noguchi K, Nakano Y, Segawa E, Zushi Y, Takaoka K, et al. Functional analysis of Zyxin in cell migration and invasive potential of oral squamous cell carcinoma cells Corrigendum in /10.3892/ijo.2016.3702. Int J Oncol. 2013;42(3):873-80.

152.	Drees B, Friederich E, Fradelizi J, Louvard D, Beckerle MC, Golsteyn RM. Characterization of the interaction between zyxin and members of the Ena/vasodilator-stimulated phosphoprotein family of proteins. J Biol Chem. 2000;275(29):22503-11.

153.	Holt MR, Critchley DR, Brindle NP. The focal adhesion phosphoprotein, VASP. Int J Biochem Cell Biol. 1998;30(3):307-11.

154.	Blanchard A, Ohanian V, Critchley D. The structure and function of alpha-actinin. J Muscle Res Cell Motil. 1989;10(4):280-9.

155.	Lim Y, Lim S-T, Tomar A, Gardel M, Bernard-Trifilo JA, Chen XL, et al. PyK2 and FAK connections to p190Rho guanine nucleotide exchange factor regulate RhoA activity, focal adhesion formation, and cell motility. Journal of Cell Biology. 2008;180(1):187-203.

156.	Greenwood JA, Theibert AB, Prestwich GD, Murphy-Ullrich JE. Restructuring of Focal Adhesion Plaques by Pi 3-Kinase: Regulation by Ptdins (3,4,5-P)3 Binding to α-Actinin. Journal of Cell Biology. 2000;150(3):627-42.

157.	Zeng Z-Z, Jia Y, Hahn NJ, Markwart SM, Rockwood KF, Livant DL. Role of Focal Adhesion Kinase and Phosphatidylinositol 3′-Kinase in Integrin Fibronectin Receptor-Mediated, Matrix Metalloproteinase-1‑Dependent Invasion by Metastatic Prostate Cancer Cells. Cancer Research. 2006;66(16):8091-9.

158.	Katoh K, Kano Y, Ookawara S. Role of stress fibers and focal adhesions as a mediator for mechano-signal transduction in endothelial cells in situ. Vasc Health Risk Manag. 2008;4(6):1273-82.

159.	Ramjaun AR, Hodivala-Dilke K. The role of cell adhesion pathways in angiogenesis. Int J Biochem Cell Biol. 2009;41(3):521-30.

160.	Wu MH. Endothelial focal adhesions and barrier function. J Physiol. 2005;569(Pt 2):359-66.

161.	Vouret-Craviari V, Boquet P, Pouyssegur J, Van Obberghen-Schilling E. Regulation of the actin cytoskeleton by thrombin in human endothelial cells: role of Rho proteins in endothelial barrier function. Mol Biol Cell. 1998;9(9):2639-53.

162.	Lee J, Song KH, Kim T, Doh J. Endothelial Cell Focal Adhesion Regulates Transendothelial Migration and Subendothelial Crawling of T Cells. Front Immunol. 2018;9:48.

163.	Crick FH. On protein synthesis. Symp Soc Exp Biol. 1958;12:138-63.

164.	Cobb M. 60 years ago, Francis Crick changed the logic of biology. PLoS Biol. 2017;15(9):e2003243.

165.	Ohno S, editor So much" junk" DNA in our genome. In" Evolution of Genetic Systems". Brookhaven Symposium in Biology; 1972.

166.	Palazzo AF, Gregory TR. The case for junk DNA. PLoS Genet. 2014;10(5):e1004351.

167.	Arraiano CM. Regulatory noncoding RNAs: functions and applications in health and disease. The FEBS Journal. 2021;288(22):6308-9.

168.	Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011;12(12):861-74.

169. Hombach S, Kretz M. Non-coding RNAs: Classification, Biology and Functioning. Adv Exp Med Biol. 2016;937:3-17.

170. Sasso JM, Ambrose BJB, Tenchov R, Datta RS, Basel MT, DeLong RK, et al. The Progress and Promise of RNA Medicine—An Arsenal of Targeted Treatments. Journal of Medicinal Chemistry. 2022;65(10):6975-7015.

171. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, et al. The complete sequence of a human genome. Science. 2022;376(6588):44-53.

172. Arang R, Sergey N, Monika C, Savannah JH, Dylan JT, Nicolas A, et al. The complete sequence of a human Y chromosome. bioRxiv. 2022:2022.12.01.518724.

173. Casamassimi A, Ciccodicola A. Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation. Int J Mol Sci. 2019;20(6).

174. Bloom AB, Zaman MH. Influence of the microenvironment on cell fate determination and migration. Physiological Genomics. 2014;46(9):309-14.

175. Bhat R, Bissell MJ. Of plasticity and specificity: dialectics of the micro- and macro-environment and the organ phenotype. Wiley Interdiscip Rev Membr Transp Signal. 2014;3(2):147-63.

176. Zipori D. The nature of stem cells: state rather than entity. Nat Rev Genet. 2004;5(11):873-8.

177. Jiang Y, Jahagirdar BN, Reinhardt RL, Schwartz RE, Keene CD, Ortiz-Gonzalez XR, et al. Pluripotency of mesenchymal stem cells derived from adult marrow. Nature. 2002;418(6893):41-9.

178. Marino Giacomo B, Ngai M, Clarke Daniel JB, Fleishman Reid H, Deng Eden Z, Xie Z, et al. GeneRanger and TargetRanger: processed gene and protein expression levels across cells and tissues for target discovery. Nucleic Acids Research. 2023.

179. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.

180. Yang RY, Quan J, Sodaei R, Aguet F, Segrè AV, Allen JA, et al. A systematic survey of human tissue-specific gene expression and splicing reveals new opportunities for therapeutic target identification and evaluation. bioRxiv. 2018:311563.

181. Eddy SR. Non–coding RNA genes and the modern RNA world. Nature Reviews Genetics. 2001;2(12):919-29.

182. LaRossa RA. Transcriptome. In: Maloy S, Hughes K, editors. Brenner's Encyclopedia of Genetics (Second Edition). San Diego: Academic Press; 2013. p. 101-3.

183. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015;348(6235):660-5.

184. Wolff CA, Gutierrez-Monreal MA, Meng L, Zhang X, Douma LG, Costello HM, et al. Defining the age-dependent and tissue-specific circadian transcriptome in male mice. Cell Rep. 2023;42(1):111982.

185.   Cooper G, Adams K. The cell: a molecular approach: Oxford University Press; 2022.

186.   Johnson A, Lewis J, ALBERTS B. Molecular biology of the cell. 2002.

187.   Liljas A. Messenger RNA. In: Maloy S, Hughes K, editors. Brenner's Encyclopedia of Genetics (Second Edition). San Diego: Academic Press; 2013. p. 369-70.

188.   Darnell JE, Jr. Reflections on the history of pre-mRNA processing and highlights of current knowledge: a unified picture. Rna. 2013;19(4):443-60.

189.   Jurado AR, Tan D, Jiao X, Kiledjian M, Tong L. Structure and Function of Pre-mRNA 5′-End Capping Quality Control and 3′-End Processing. Biochemistry. 2014;53(12):1882-98.

190.   Wang A. Noncoding RNAs evolutionarily extend animal lifespan. bioRxiv. 2023:2023.06.09.544283.

191.   Neumeier J, Meister G. siRNA Specificity: RNAi Mechanisms and Strategies to Reduce Off-Target Effects. Front Plant Sci. 2020;11:526455.

192.   Romaine SPR, Tomaszewski M, Condorelli G, Samani NJ. MicroRNAs in cardiovascular disease: an introduction for clinicians. Heart. 2015;101(12):921-8.

193.   Li X, Fu Q, Li H, Zhu L, Chen W, Ruan T, et al. MicroRNA-520c-3p functions as a novel tumor suppressor in lung adenocarcinoma. Febs j. 2019;286(14):2737-52.

194.   Zhang L, Lv Z, Xu J, Chen C, Ge Q, Li P, et al. MicroRNA-134 inhibits osteosarcoma angiogenesis and proliferation by targeting the VEGFA/VEGFR1 pathway. Febs j. 2018;285(7):1359-71.

195.   Li Y, Kowdley KV. MicroRNAs in common human diseases. Genomics Proteomics Bioinformatics. 2012;10(5):246-53.

196.   Rupani H, Sanchez-Elsner T, Howarth P. MicroRNAs and respiratory diseases. European Respiratory Journal. 2013;41(3):695-705.

197.   Morais P, Adachi H, Yu YT. Spliceosomal snRNA Epitranscriptomics. Front Genet. 2021;12:652129.

198.   Valadkhan S, Gunawardane LS. Role of small nuclear RNAs in eukaryotic gene expression. Essays Biochem. 2013;54:79-90.

199.   Matera AG, Wang Z. A day in the life of the spliceosome. Nat Rev Mol Cell Biol. 2014;15(2):108-21.

200.   Bachellerie J-P, Cavaillé J, Hüttenhofer A. The expanding snoRNA world. Biochimie. 2002;84(8):775-90.

201.   Yu C-Y, Kuo H-C. The emerging roles and functions of circular RNAs and their generation. Journal of Biomedical Science. 2019;26(1):29.

202.   Borkiewicz L, Kalafut J, Dudziak K, Przybyszewska-Podstawka A, Telejko I. Decoding LncRNAs. Cancers (Basel). 2021;13(11).

203.   Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nature Reviews Molecular Cell Biology. 2021;22(2):96-118.

204.   Liu Y, Dou M, Song X, Dong Y, Liu S, Liu H, et al. The emerging role of the piRNA/piwi complex in cancer. Molecular Cancer. 2019;18(1):123.

205.   Li L-C, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, et al. Small dsRNAs induce transcriptional activation in human cells. Proceedings of the National Academy of Sciences. 2006;103(46):17337-42.

206.   Kwok A, Raulf N, Habib N. Developing small activating RNA as a therapeutic: current challenges and promises. Ther Deliv. 2019;10(3):151-64.

207.   Xu JZ, Zhang JL, Zhang WG. Antisense RNA: the new favorite in genetic research. J Zhejiang Univ Sci B. 2018;19(10):739-49.

208.   Tutar Y. Pseudogenes. Comp Funct Genomics. 2012;2012:424526.

209.   Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Carter DR. Pseudogenes: pseudo-functional or key regulators in health and disease? Rna. 2011;17(5):792-8.

210.   Harper JW, Bennett EJ. Proteome complexity and the forces that drive proteome imbalance. Nature. 2016;537(7620):328-38.

211.   Bugyi F, Szabó D, Szabó G, Révész Á, Pape VFS, Soltész-Katona E, et al. Influence of Post-Translational Modifications on Protein Identification in Database Searches. ACS Omega. 2021;6(11):7469-77.

212.   Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. Nature Reviews Genetics. 2019;20(11):631-56.

213.   Evans TG. Considerations for the use of transcriptomics in identifying the 'genes that matter' for environmental adaptation. J Exp Biol. 2015;218(Pt 12):1925-35.

214.   Kerkhofs M, Haijes HA, Willemsen AM, van Gassen KLI, van der Ham M, Gerrits J, et al. Cross-Omics: Integrating Genomics with Metabolomics in Clinical Diagnostics. Metabolites. 2020;10(5).

215.   Patel A, McGrosso D, Hefner Y, Campeau A, Sastry AV, Maurya S, et al. Proteome allocation is linked to transcriptional regulation through a modularized transcriptome. bioRxiv. 2023.

216.   Digre A, Lindskog C. The human protein atlas-Integrated omics for single cell mapping of the human proteome. Protein Sci. 2023;32(2):e4562.

217.   Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Current Protocols in Bioinformatics. 2016;54(1):1.30.1-1..3.

218.   Struck E, Belova T, Hsieh P-H, Odeberg J, Kuijjer ML, Dusart P, et al. Temporal transcriptome analysis of the endothelial response to tumour necrosis factor. bioRxiv. 2023:2023.06.04.543378.

219.   Cormier M, Cushman M. Innovation via social media – The importance of Twitter to science. Research and Practice in Thrombosis and Haemostasis. 2021;5.

220.   Reyna N, Pruett C, Morrison M, Fowler J, Pandey S, Hensley L. Twitter: More than Tweets for Undergraduate Student Researchers. Journal of Microbiology & Biology Education. 2022;23.

221.   S E, Shiffman D, Côté I, Drew J. The role of Twitter in the life cycle of a scientific publication. Ideas in Ecology and Evolution. 2013;6.

222. Abraham G. The Importance of Science Communication. Metallography, Microstructure, and Analysis. 2020;9.

223. Hrdlickova R, Toloue M, Tian B. RNA-Seq methods for transcriptome analysis. Wiley Interdiscip Rev RNA. 2017;8(1).

224. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. Cold Spring Harb Protoc. 2015;2015(11):951-69.

225. Hegenbarth J-C, Lezzoche G, De Windt LJ, Stoll M. Perspectives on Bulk-Tissue RNA Sequencing and Single-Cell RNA Sequencing for Cardiac Transcriptomics. Frontiers in Molecular Medicine. 2022;2.

226. Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. Frontiers in Genetics. 2019;10.

227. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biology. 2020;21(1):30.

228. Li X, Wang C-Y. From bulk, single-cell to spatial RNA sequencing. International Journal of Oral Science. 2021;13(1):36.

229. Hu T, Chitnis N, Monos D, Dinh A. Next-generation sequencing technologies: An overview. Human Immunology. 2021;82(11):801-11.

230. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: Overviews and challenges. BioTechniques. 2014;56(2):61-77.

231. Thind AS, Monga I, Thakur PK, Kumari P, Dindhoria K, Krzak M, et al. Demystifying emerging bulk RNA-Seq applications: the application and utility of bioinformatic methodology. Briefings in Bioinformatics. 2021;22(6).

232. Marguerat S, Bähler J. Coordinating genome expression with cell size. Trends in Genetics. 2012;28(11):560-5.

233. Slovin S, Carissimo A, Panariello F, Grimaldi A, Bouché V, Gambardella G, et al. Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview. Methods Mol Biol. 2021;2284:343-65.

234. Abrams ZB, Johnson TS, Huang K, Payne PRO, Coombes K. A protocol to evaluate RNA sequencing normalization methods. BMC Bioinformatics. 2019;20(24):679.

235. Koch CM, Chiu SF, Akbarpour M, Bharat A, Ridge KM, Bartom ET, et al. A Beginner's Guide to Analysis of RNA Sequencing Data. Am J Respir Cell Mol Biol. 2018;59(2):145-57.

236. Tong Y. The comparison of limma and DESeq2 in gene analysis. E3S Web Conf. 2021;271:03058.

237. Guo Y, Li C-I, Ye F, Shyr Y. Evaluation of read count based RNAseq analysis methods. BMC Genomics. 2013;14(8):S2.

238. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511-5.

239. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621-8.

240. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 2012;131(4):281-5.

241. Zhao S, Ye Z, Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. Rna. 2020;26(8):903-9.

242. Zhao Y, Li M-C, Konaté MM, Chen L, Das B, Karlovich C, et al. TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. Journal of Translational Medicine. 2021;19(1):269.

243. Liu S, Wang Z, Zhu R, Wang F, Cheng Y, Liu Y. Three Differential Expression Analysis Methods for RNA Sequencing: limma, EdgeR, DESeq2. J Vis Exp. 2021(175).

244. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.

245. Bender R, Lange S. Adjusting for multiple testing--when and how? J Clin Epidemiol. 2001;54(4):343-9.

246. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.

247. Liu G, Shiraito Y. Multiple Hypothesis Testing in Conjoint Analysis. Political Analysis. 2023;31:380 - 95.

248. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.

249. Tomczak A, Mortensen JM, Winnenburg R, Liu C, Alessi DT, Swamy V, et al. Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. Scientific Reports. 2018;8(1):5115.

250. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics. 2008;25(2):288-9.

251. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 2003;13(9):2129-41.

252. Zhou X, Su Z. EasyGO: Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species. BMC Genomics. 2007;8:246.

253. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol. 2003;4(4):R28.

254. Beissbarth T, Speed TP. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. Bioinformatics. 2004;20(9):1464-5.

255. Martin D, Brun C, Remy E, Mouren P, Thieffry D, Jacq B. GOToolBox: functional analysis of gene datasets based on Gene Ontology. Genome Biol. 2004;5(12):R101.

256.   Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22(13):1600-7.

257.   Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

258.   Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.

259.   Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2019;36(8):2628-9.

260.   Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biology. 2010;11(2):R14.

261.   Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. Bioinformatics. 2006;23(2):257-8.

262.   Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3):100141.

263.   Rivals I, Personnaz L, Taing L, Potier M-C. Enrichment or depletion of a GO category within a class of genes: which test? Bioinformatics. 2006;23(4):401-7.

264.   Medina-Leyte DJ, Domínguez-Pérez M, Mercado I, Villarreal-Molina MT, Jacobo-Albavera L. Use of Human Umbilical Vein Endothelial Cells (HUVEC) as a Model to Study Cardiovascular Disease: A Review. Applied Sciences. 2020;10(3):938.

265.   Baudin B, Bruneel A, Bosselut N, Vaubourdolle M. A protocol for isolation and culture of human umbilical vein endothelial cells. Nat Protoc. 2007;2(3):481-5.

266.   Addis R, Campesi I, Fois M, Capobianco G, Dessole S, Fenu G, et al. Human umbilical endothelial cells (HUVECs) have a sex: characterisation of the phenotype of male and female cells. Biol Sex Differ. 2014;5(1):18.

267.   Seidl K, Solis NV, Bayer AS, Hady WA, Ellison S, Klashman MC, et al. Divergent Responses of Different Endothelial Cell Types to Infection with Candida albicans and Staphylococcus aureus. PLOS ONE. 2012;7(6):e39633.

268.   Cooke BM, Usami S, Perry I, Nash GB. A simplified method for culture of endothelial cells and analysis of adhesion of blood cells under conditions of flow. Microvasc Res. 1993;45(1):33-45.

269.   Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, et al. Shiny: Web Application Framework for R. 2022.

270.   Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348(6235):648-60.

271.   Cope L, Naiman DQ, Parmigiani G. Integrative correlation: Properties and relation to canonical correlations. J Multivar Anal. 2014;123:270-80.

272.   Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-9.

273. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nat Protoc. 2013;8(8):1551-66.

274. Mi H, Poudel S, Muruganujan A, Casagrande JT, Thomas PD. PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. 2016;44(D1):D336-42.

275. Liang C-C, Park AY, Guan J-L. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. Nature Protocols. 2007;2(2):329-33.

276. Cory G. Scratch-wound assay. Methods Mol Biol. 2011;769:25-30.

277. van Paridon PCS, Panova-Noeva M, van Oerle R, Schultz A, Hermanns IM, Prochaska JH, et al. Thrombin generation in cardiovascular disease and mortality - results from the Gutenberg Health Study. Haematologica. 2020;105(9):2327-34.

278. Macfarlane RG, Biggs R. A thrombin generation test; the application in haemophilia and thrombocytopenia. J Clin Pathol. 1953;6(1):3-8.

279. Tripodi A. Thrombin generation: a global coagulation procedure to investigate hypo- and hyper-coagulability. Haematologica. 2020;105(9):2196-9.

280. Hemker HC, Willems GM, Béguin S. A computer assisted method to obtain the prothrombin activation velocity in whole plasma independent of thrombin decay processes. Thromb Haemost. 1986;56(1):9-17.

281. Hemker HC, Giesen P, Al Dieri R, Regnault V, de Smedt E, Wagenvoord R, et al. Calibrated automated thrombin generation measurement in clotting plasma. Pathophysiol Haemost Thromb. 2003;33(1):4-15.

282. Ulfhammer E, Larsson P, Karlsson L, Hrafnkelsdottir T, Bokarewa M, Tarkowski A, et al. TNF-alpha mediated suppression of tissue type plasminogen activator expression in vascular endothelial cells is NF-kappaB- and p38 MAPK-dependent. J Thromb Haemost. 2006;4(8):1781-9.

283. Jung HS, Shimizu-Albergine M, Shen X, Kramer F, Shao D, Vivekanandan-Giri A, et al. TNF-alpha induces acyl-CoA synthetase 3 to promote lipid droplet formation in human endothelial cells. J Lipid Res. 2020;61(1):33-44.

284. Brandt M, Gerke V, Betz T. Human endothelial cells display a rapid tensional stress increase in response to tumor necrosis factor-alpha. PLoS One. 2022;17(6):e0270197.

285. Ryan FJ, Ma Y, Ashander LM, Kvopka M, Appukuttan B, Lynn DJ, et al. Transcriptomic Responses of Human Retinal Vascular Endothelial Cells to Inflammatory Cytokines. Transl Vis Sci Technol. 2022;11(8):27.

286. Rastogi S, Rizwani W, Joshi B, Kunigal S, Chellappan SP. TNF-alpha response of vascular endothelial and vascular smooth muscle cells involve differential utilization of ASK1 kinase and p73. Cell Death Differ. 2012;19(2):274-83.

287. Valenzuela NM. Late phase endothelial cell inflammation is characterized by interferon response genes and driven by JAK/STAT, not NFκB. Vascul Pharmacol. 2022;146:107090.

288. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. Nat Rev Immunol. 2014;14(1):36-49.

289.   Sheikh F, Dickensheets H, Gamero AM, Vogel SN, Donnelly RP. An essential role for IFN-β in the induction of IFN-stimulated gene expression by LPS in macrophages. J Leukoc Biol. 2014;96(4):591-600.

290.   Honda K, Takaoka A, Taniguchi T. Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors. Immunity. 2006;25(3):349-60.

291.   Decout A, Katz JD, Venkatraman S, Ablasser A. The cGAS–STING pathway as a therapeutic target in inflammatory diseases. Nature Reviews Immunology. 2021;21(9):548-69.

292.   Sun Z, Hornung V. cGAS-STING signaling. Curr Biol. 2022;32(13):R730-r4.

293.   Rehwinkel J, Gack MU. RIG-I-like receptors: their regulation and roles in RNA sensing. Nature Reviews Immunology. 2020;20(9):537-51.

294.   Willemsen J, Neuhoff M-T, Hoyler T, Noir E, Tessier C, Sarret S, et al. TNF leads to mtDNA release and cGAS/STING-dependent interferon responses that support inflammatory arthritis. Cell Reports. 2021;37(6):109977.

295.   Lee SK. Sex as an important biological variable in biomedical research. BMB Rep. 2018;51(4):167-73.

296.   Lu HS, Schmidt AM, Hegele RA, Mackman N, Rader DJ, Weber C, et al. Reporting Sex and Sex Differences in Preclinical Studies. Arterioscler Thromb Vasc Biol. 2018;38(10):e171-e84.

297.   Viemann D, Goebeler M, Schmid S, Nordhues U, Klimmek K, Sorg C, et al. TNF induces distinct gene expression programs in microvascular and macrovascular human endothelial cells. J Leukoc Biol. 2006;80(1):174-85.

298.   Kalucka J, de Rooij L, Goveia J, Rohlenova K, Dumas SJ, Meta E, et al. Single-Cell Transcriptome Atlas of Murine Endothelial Cells. Cell. 2020;180(4):764-79.e20.

299.   Tabula Sapiens C, Jones RC, Karkanias J, Krasnow MA, Pisco AO, Quake SR, et al. The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science. 2022;376(6594):eabl4896.

300.   Nakajima H, Mochizuki N. Flow pattern-dependent endothelial cell responses through transcriptional regulation. Cell Cycle. 2017;16(20):1893-901.

301.   Helle E, Ampuja M, Dainis A, Antola L, Temmes E, Tolvanen E, et al. HiPS-Endothelial Cells Acquire Cardiac Endothelial Phenotype in Co-culture With hiPS-Cardiomyocytes. Frontiers in Cell and Developmental Biology. 2021;9.

302.   Heydarkhan-Hagvall S, Helenius G, Johansson BR, Li JY, Mattsson E, Risberg B. Co-culture of endothelial cells and smooth muscle cells affects gene expression of angiogenic factors. J Cell Biochem. 2003;89(6):1250-9.

303.   Afshar Y, Ma F, Quach A, Jeong A, Sunshine HL, Freitas V, et al. Transcriptional drifts associated with environmental changes in endothelial cells. Elife. 2023;12.

304.   Sheikh S, Rainger GE, Gale Z, Rahman M, Nash GB. Exposure to fluid shear stress modulates the ability of endothelial cells to recruit neutrophils in response to tumor necrosis factor-alpha: a basis for local variations in vascular sensitivity to inflammation. Blood. 2003;102(8):2828-34.

305.	Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016;165(3):535-50.

306.	Metcalf DJ, Nightingale TD, Zenner HL, Lui-Roberts WW, Cutler DF. Formation and function of Weibel-Palade bodies. Journal of Cell Science. 2008;121(1):19-27.

PAPER 1

Global transcriptome analysis reveals distinct phases of the endothelial response to tumour necrosis factor

(resubmitted to Journal of Immunology, in revision)

Global transcriptome analysis reveals distinct phases of the endothelial response to tumour necrosis factor

Running title: *A systems level analysis of the endothelial response to TNF*

Eike C Struck [1], Tatiana Belova[2], Ping-Han Hsieh [2], Jacob O Odeberg [1, 3, 4, 5], Marieke L Kuijjer [2,7,8], Philip J Dusart [3,6 #], Lynn M Butler [1, 3, 6, 9,#]

[1] Department of Clinical Medicine, The Arctic University of Norway, Tromsø, Norway

[2] Centre for Molecular Medicine Norway, Nordic EMBL Partnership, University of Oslo, Norway.

[3] Science for Life Laboratory, Department of Protein Science, Royal Institute of Technology (KTH), Stockholm, Sweden

[4] The University Hospital of North Norway (UNN), Tromsø, Norway

[5] Coagulation Unit, Department of Haematology, Karolinska University Hospital, Stockholm, Sweden

[6] Clinical Chemistry and Blood Coagulation Research, Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

[7] Department of Pathology, Leiden University Medical Center, Leiden, the Netherlands

[8] Leiden Center for Computational Oncology, Leiden University Medical Center, Leiden, the Netherlands

[9] Clinical Chemistry, Karolinska University Laboratory, Karolinska University Hospital, Stockholm, Sweden

# Equal contribution

Correspondence information:

Dr. L.M Butler

Clinical Chemistry and Blood Coagulation Research,

Department of Molecular Medicine and Surgery,

Karolinska Institute,

SE-171 76 Stockholm, Sweden

Email: Lynn.butler@ki.se

KEY WORDS: Endothelium, inflammation, transcriptome analysis, tumour necrosis factor

ABSTRACT

The vascular endothelium acts as a dynamic interface between blood and tissue. Tumour necrosis factor-α (TNF), a major regulator of inflammation, induces endothelial cell (EC) transcriptional changes, the overall response dynamics of which have not been fully elucidated. Here, we conducted an extended time-course analysis of the human EC response to TNF, from 30 minutes to 72 hours. We identified regulated genes and used weighted gene correlation network analysis to decipher co-expression profiles, uncovering two distinct temporal phases - an acute response (between 1-4 hours), and a later phase (between 12-24 hours). Sex-based subset analysis revealed that the response was comparable between female and male cells. Several previously uncharacterised genes were strongly regulated during the acute phase, while the majority in the later phase were interferon-stimulated genes (ISGs). A lack of interferon transcription indicated that this ISG expression was independent of *de novo* interferon production. We also observed two groups of genes whose transcription was inhibited by TNF - those that resolved towards baseline levels, and those that did not. Our study provides insights into the global dynamics of the EC transcriptional response to TNF, highlighting distinct gene expression patterns during the acute and later phases. Data for all coding and non-coding genes is provided on the website (http://www.endothelial-response.org/). These findings may be useful in understanding the role of EC in inflammation and in developing TNF signalling-targeting therapies.

KEY POINTS

1. This study investigates TNF-induced endothelial transcriptional changes over 72 hours

2. Global analysis revealed discrete acute and delayed temporal gene expression phases

3. Interferon stimulated gene expression was independent of interferon production

2

INTRODUCTION

The vascular endothelium is a dynamic interface between blood and tissue that has a role in the regulation of coagulation, blood pressure, solute movement, and inflammation. The resting endothelium is an anti-inflammatory and anti-thrombotic surface, which is unreceptive to interactions with circulating blood cells (Ley and Reutershan 2006, Yau, Teoh et al. 2015). The cytokine tumour necrosis factor-α (TNF) is a key driver of acute and chronic inflammation (Webster and Vucic 2020, Jang, Lee et al. 2021) and can bind to endothelial cells (EC) via TNF-receptors 1 and 2. This interaction induces signalling cascades that regulate the activity of several transcription factors, including NF-kappaB (NFκB) and activator protein-1 and 2 (Vandenabeele, Declercq et al. 1995, Baud and Karin 2001), leading to various cellular responses, such as the expression of adhesion molecules and chemokines that facilitate leukocyte recruitment into tissue (Liao 2013). TNF can also induce the EC expression of interferon regulatory factors and interferon-stimulated genes (ISGs) (Venkatesh, Ernandez et al. 2013, Yan, van Meurs et al. 2017). Indeed, the concept that EC are multifaceted conditional innate immune cells has gained traction in recent years (Shao, Saredy et al. 2020, Lu, Sun et al. 2022). However, the global dynamics of the EC response to TNF is not well understood, with existing studies tending to focus on specific gene(s) or phenotypic characteristics, e.g., (Ulfhammer, Larsson et al. 2006, Jung, Shimizu-Albergine et al. 2020, Brandt, Gerke et al. 2022), or global transcriptional changes at a single, or small number, of time points (Rastogi, Rizwani et al. 2012, Ryan, Ma et al. 2022). The same is true of studies of the TNF response in other cell types, such as the immortalised embryonic kidney cell line HEK293 (Bouwmeester, Bauch et al. 2004, Ma, Li et al. 2009) or fibroblasts (Hao and Baltimore 2013, Paulsen, Veloso et al. 2013). Existing studies also tend to neglect the influence of chromosomal sex on cell behaviour (Lu, Schmidt et al. 2018), despite reported differences between male and female EC, e.g., in preeclampsia (Zhou, Yan et al. 2019), and following exposure to shear stress (James and Allen 2021), X-ray induced damage (Campesi, Brunetti et al. 2022) and hyperoxia (Zhang and Lingappan 2017).

Here, we profiled the EC transcriptional response to TNF over an extended time, including 11 time points ranging from 30 minutes to 72 hours post stimulation. We identified TNF up regulated or down regulated genes and used weighted gene network correlation

3

analysis to decipher co-expression profiles, revealing two distinct temporal phases; an acute response initiated between 1-4 hours, and a subsequent later one between 12-24 hours. Sex-based analysis revealed a high similarity in response profile between female and male EC. Several completely uncharacterised genes were strongly regulated during the acute phase of the response, while the majority of those in the latter phase were interferon stimulated genes (ISG). Our data indicated that this ISG expression was independent of *de novo* interferon production and subsequent autocrine signalling. All data is available on (http://www.endothelial-response.org/), which users can view on a gene centric or regulation-profile basis.

4

RESULTS

Existing studies of EC responses to inflammatory stimuli, such as TNF, focus primarily on early signalling and associated transcriptional changes. Here, we profiled global changes in the EC transcriptome following TNF stimulation, over an extended time course. Human umbilical vein endothelial cells were extracted and pooled by donor sex, before treatment with or without TNF for 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48 or 72 hours, and subsequent analysis by RNA sequencing (Figure S1 A).

TNF induces EC transcriptome modifications over an extended time course

Average DESeq2 expression values for each treated sample (biological replicates: n=3 male and n=2 female) were normalised to the sex- and time-matched untreated controls, to identify differentially expressed genes (DEG). 2099 positive and 2162 negative DEG were identified (fold change vs. untreated control FC log2 > 1 and FC log2 < -1, respectively [in both male and female sample sets], TPM>1 at least one sample time point, raw counts >10 in all samples, adjusted by p-value) (Figure S1 B and methods for analysis details) (Table S1 B). Of these, a total of 918 genes were further classified as up regulated in response to TNF (Figure 1 A.i), and 210 as down regulated (Figure 1B.i), based on a TNF-induced change from a relatively stable baseline expression in untreated (control) EC over time (coefficient of variation [CV] between TPM values < 0.3) (Figure S1 C.i and D.i). Genes classified as DEG due to changes in expression in untreated control EC over time (Figure S1 C.ii and S1 D.iii), or a TNF-induced lag in such changes (Figure S1 D.ii), were excluded, as such changes may be linked to *in vitro* culture conditions and/or the influence of TNF on other inflammation-independent temporal processes.

The majority of up regulated genes were classified as protein coding (823/918; 89.7 %), followed by long non-coding RNAs (lncRNA) (61/918; 6.6%), pseudogenes (30/918; 3.8%) and ´others´ (3 TEC, 1 sRNA) (Figure 1A.i). A similar profile was observed for down regulated transcripts, with the majority classified as protein coding (193/210; 91.9 %), followed by lncRNA (14/210; 6.9%), pseudogenes (2/210; 1%) and ´others´ (1 TEC, 1 sRNA) (Figure 1B.i). We performed gene ontology analysis (Ashburner, Ball et al. 2000) to identify over-represented groups in genes classified as regulated by TNF; significant enrichment terms included ´*immune system processes*´ (FDR 9.5 x10^{-44})´ and ´*response to cytokine*´ (FDR 8.2 x10^{-39}) (Table S2 A) and (overall summary in Figure 1C).
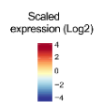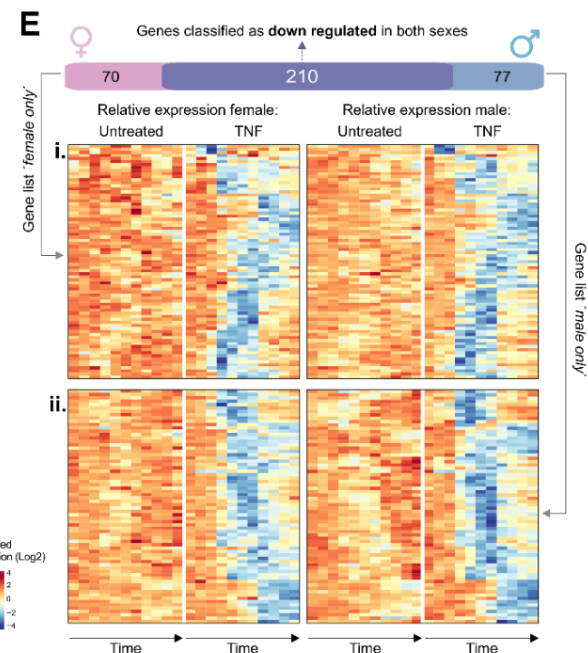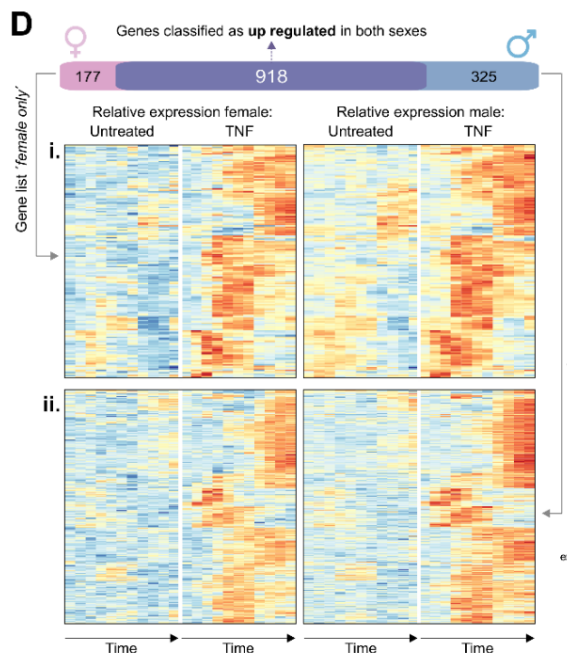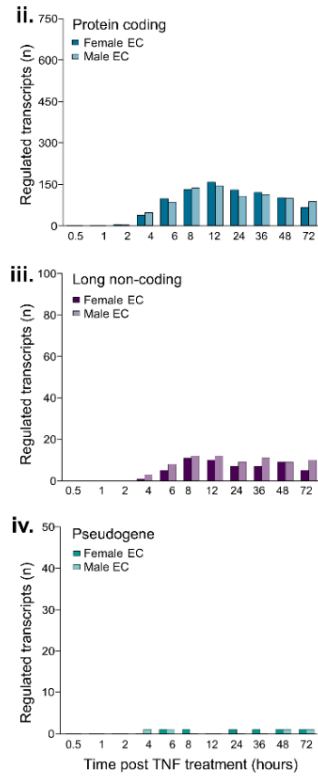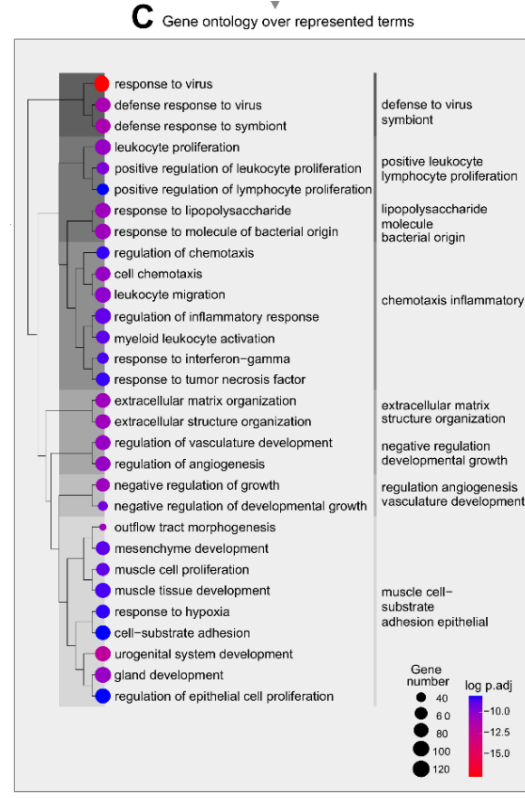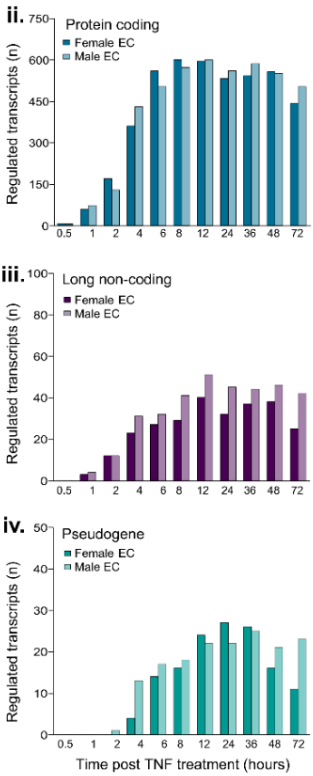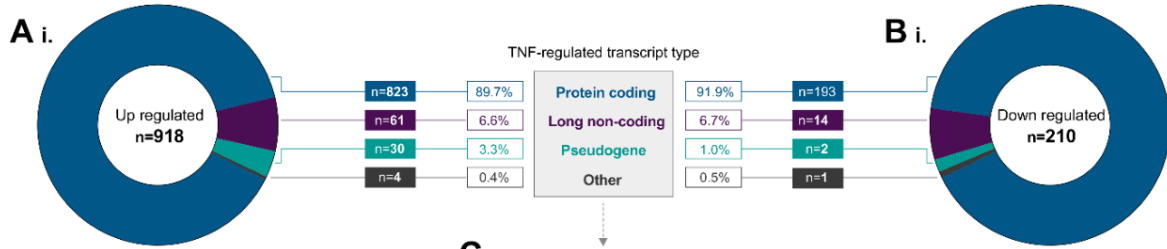
5

1

6

Figure 1. Overview of tumour necrosis factor (TNF) regulated genes in human endothelial cells. Human umbilical vein endothelial cells (EC, n=5) were treated with or without tumour necrosis factor alpha (TNF) and harvested at 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48 or 72 hrs, before RNAseq analysis. Genes were classified as (**A**) up regulated or (**B**) down regulated by TNF (fold change vs. untreated control >2 or <0.5, respectively). Plots show: (i) total TNF regulated genes (in both sexes) and corresponding biotype, and number of (ii) protein coding, (iii) lncRNA or (iv) pseudogenes regulated at each time point, by sex. (**C**) Gene ontology analysis showing over-represented terms for all TNF regulated transcripts. (**D**) and (**E**) show the number of genes classified as either up regulated or down regulated, respectively, in females or males only, or in both sexes. Heatmaps show relative expression in control or TNF treated female (left) or male (right) samples for genes reaching the threshold for classification as TNF-regulated in (i) males or (ii) females only.

7

To investigate the temporal profile of genes up or down regulated in response to TNF, we determined the number of transcripts classified as such at each time point, in each sex (Figure 1 A.ii-iv and B.ii-iv). A limited number of protein coding transcripts (n=10) were classified as up regulated already by 30 minutes post-stimulation, in both sexes (Figure 1 A.ii), including those encoding for components of the NFκB-signalling pathway (*NFKBIA*, *NFKBIZ*), chemokines regulated by this pathway (*CXCL2*, *CXCL3*, *IL6*) and *FAM167A*, which was recently described as an activator of the non-canonical NFκB-pathway in chronic myeloid leukaemia (Yang, Sim et al. 2022) (Table S2 B).

A small panel of lncRNAs were classified as up regulated by 1-hour post-stimulation, in both sexes (n=10) (Figure 1 A.iii); including those known to be expressed in response to inflammation in EC, e.g., *MIR155HG* (Barros Ferreira, Ashander et al. 2022), but also those not previously described in this context, e.g., wound and keratinocyte migration-associated long noncoding RNA 2 (*WAKMAR2*), which restricts inflammatory chemokine production in keratinocytes, and enhances their migratory capacity (Herter, Li et al. 2019). TNF-responsive pseudogenes tended to be classified as up regulated slightly later in the time course than protein coding or long non-coding genes. From a total of 30 pseudogenes classified as up regulated by TNF (Figure 1 A.iv), 10/30 [33%] were members of the ferritin gene family (*FTH1P2, 7, 8, 10, 11, 12, 15, 16, 20, 23*), some of which have known regulatory functions (Di Sanzo, Quaresima et al. 2020). Ferritin has a role in EC angiogenesis (Tesfay, Huhn et al. 2012) and chemokine signalling (Li, Luo et al. 2006); a potential role of this family in the EC response to inflammatory stimulation remains to be explored. Overall, the total number of genes that were up regulated in response to TNF was highest between 12- and 24-hours post-stimulation (Figure 1 A.ii, iii and iv).

In contrast to TNF-induced up regulated genes, no protein coding genes were classified as down regulated at the earliest time point (Figure 1 B.ii). Those classified as such within the first four hours (n=54), included *HOXA9*, which inhibits NFκB dependent EC activation (Trivedi, Patel et al. 2007). The total number of down regulated protein coding transcripts peaked around 12 hours (Figure 1 B.ii). lncRNAs classified as down regulated in response to TNF were also classified as such later than those that were up regulated (Figure 1B.iii),

8

and only a small number of pseudogenes were consistently down regulated across the time course (Figure 1 B.iv).

Thus, the global primary EC response to TNF stimulation predominantly consists of the induction of protein coding and, to a lesser extent, long non-coding gene expression.

TNF-induced changes are comparable between male and female endothelial cells

Differences in inflammatory response have been previously described in female and male EC (Addis, Campesi et al. 2014, Zhou, Yan et al. 2019). In our analysis described above, 177 genes were classified as up regulated only in female samples and 325 only in male samples (Figure 1 D), and 70 genes were classified as down regulated only in female samples and 77 only in male samples (Figure 1 E) (Table S1 C). Heatmap plots of genes classified as up or down regulated only in female EC (Figure 1 D.i and E.i, respectively) or only in male EC (Figure 1 D.ii and E.ii, respectively), revealed similar patterns of regulation over the time course in the other respective sex (relative expression in female EC on the left and male EC on the right). Thus, these differences in classification were likely due to the strict thresholding criteria we initially applied to identify the most consistently and strongly regulated genes, rather than a fundamental difference between the responsiveness of EC from each sex. Indeed, when we applied a threshold of up- or down regulation in one or more samples of one sex (FC log2 abs >1), versus no regulation in any samples of the other sex (FC abs <1.1 [FC log2 <0.1375]) no genes were classified as sex-specifically regulated by TNF. Multidimensional scaling revealed high global similarity in TNF-induced transcriptome modifications over the time course in female (Figure S2 A.i) and male (Figure S2 A.ii) EC. Thus, the chromosomal composition of the EC does not appear to markedly effect the global transcriptional response to TNF stimulation.

To identify baseline (unstimulated) differences in gene expression in female versus male EC, we performed differential expression analysis between the control unstimulated samples of each sex. 99 genes were classified as differentially expressed between the sexes at every time point (FC log2 abs >1, adjusted p<0.05), with 58 genes being higher in females and 41 higher in males (Figure S2 B). As expected, Y chromosome genes represented the most significantly differentially expressed genes in male EC (Figure S2 B.i) and the long noncoding RNA *XIST*, a regulator for X-inactivation (Loda and Heard

9

2019), was the most significantly differentially expressed gene in female EC (Table S2 C). Of the non-sex chromosome linked genes that were differentially expressed between male and female EC at baseline (with mean expression [in either sex] TPM>1), 14 were also classified as TNF-regulated in both sexes (Figure S2 B.ii and Table S2 C). Of these, nine were more highly expressed in female vs. male cells (*PLLP*, *TNFSF10*, *CBLN3*, *MMP12*, *ANGPT2* [up regulated by TNF] and *PLXDC2*, *ABCA8*, *MS4A2*, *AC104083.1* [down regulated by TNF]) (Figure S2 B.ii), and five were more highly expressed in male vs. female cells (*FOSB*, *IL27RA*, *CDH2*, *PHLDA2* [up regulated by TNF] and *PITPNM3* [down regulated by TNF]) (Figure S2 B.ii). Although such genes had similar TNF regulatory profiles in female and male cells (for examples see Figure S2 B.iii and iv), the differences in absolute expression levels could indicate sex specific transcriptional regulation.

Weighted gene network analysis reveals TNF-induced gene signatures

To explore the potential relationship between TNF-regulated genes, in terms of expression dynamics over time, we performed a weighted co-expression gene network analysis (WGNCA) (Langfelder and Horvath 2008), where correlation coefficients between all transcripts across the sample set (male and female samples were handled together) were calculated and subsequently clustered into 48 related modules (Figure 2 A.i), based on expression profile similarity. In addition to the identification of co-regulated genes, this analysis could potentially highlight genes with a currently unknown roles in specific stages of the EC response to TNF.

TNF up regulated genes have two main regulation profiles

The majority of genes we earlier classified as up regulated in response to TNF stimulation fell into two main module regions on the WGNCA dendrogram (Figure 2 A.i), occupying neighbouring leaves on common clades: annotated as group 1 [*orange*, *purple*, *black* and *royal blue*] and group 2 [*tan* and *turquoise*] modules (Figure 2 A.ii) (see shaded boxes).

***Up regulated group 1- Early induction (1-4 hours post stimulation)*:**

Group 1 modules contained genes that were up regulated by 1 hour [*orange*] (Figure 2 B.i-ii), 2 hours [*purple*, *black*] (Figure 2 Ci-ii and Di-ii) or 4 hours [*royal blue*] (Figure 2 E.i - ii) post-TNF stimulation. Genes in the *orange* (Figure 2 B.i-ii), *purple* (Figure 2 C.i-ii) and *black* (Figure 2 D.-ii) modules reached a peak between 4 and 8 hours, after which the differential expression vs. control gradually declined. Several genes with well-established

10

roles in the initial stages of inflammation were in the ´earliest responder´ *orange* module (see Figure 2B.iii for the top 50 up regulated genes with highest correlation to the eigengene), including those encoding for components of the NFκB-signalling pathway e.g., *NFKBIA, NFKBIZ*, leukocyte adhesion receptors e.g., *VCAM1, SELE*, and various chemokines or cytokines, e.g., *CXCL8*, *CX3CL1*, *CCL2* and *CCL20*. A PubMed search, retrieving the number of publications containing the search terms [gene] and [TNF], revealed most of the *orange* module top 50 genes had been previously referred to in the context of the TNF response (Figure 2 B.iii). However, some, e.g., *FAM177B*, an - uncharacterised gene with little to no expression under baseline conditions, have no previous reported link with the EC response to TNF, or indeed inflammation processes in general. This was also true for some genes in the other modules, e.g., *purple* module long non-coding gene *MIR1915HG*, which currently has no assigned function, and *black* module gene *ALOXE3*, which encodes for a member of the lipoxygenase family that was recently reported as expressed in EC and up regulated by shear stress exposure (Sabbir, Wigle et al. 2022). One could speculate that such genes encode for proteins with currently unknown important roles in the initial stages of inflammation, and thus represent interesting candidates for functional investigation. Gene ontology analysis of TNF up regulated genes in the *orange* (Figure 2 B.iv), *purple* (Figure 2 C.iv) and *black* (Figure 2 D.iv), revealed over representation of similar inflammation-related terms, such as *inflammatory response* and *response to cytokine* (Table S2 D). Apoptosis-related GO groups were overrepresented in the orange module only (*apoptotic processes* [FDR 1.18 x10^{-5}], Table S2 D). Upregulated genes in these groups included those encoding for proteins that inhibit apoptosis, such as *TNFAIP3, SOD2, NUAK2* and *BCL2A1,* as well as those with a likely role in the induction of cell death, e.g., *CYLD* and *PMAIP1,* and those with potential roles in both, depending on context, e.g., *CD40*. Outside of the orange module, other genes with anti-apoptotic and pro-survival functions, e.g., *XIAP*, *BCL2A1*, *BIRC3* and *CFLAR* were also up regulated by TNF (Table S1 B). Together, this suggests that TNF induces both pro- and anti-apoptotic factors with complex interactions, consistent with the concept that death is not the default cellular response to TNF, and that various protectives brakes, or cell death checkpoints protect against apoptosis (van Loo and Bertrand 2023).
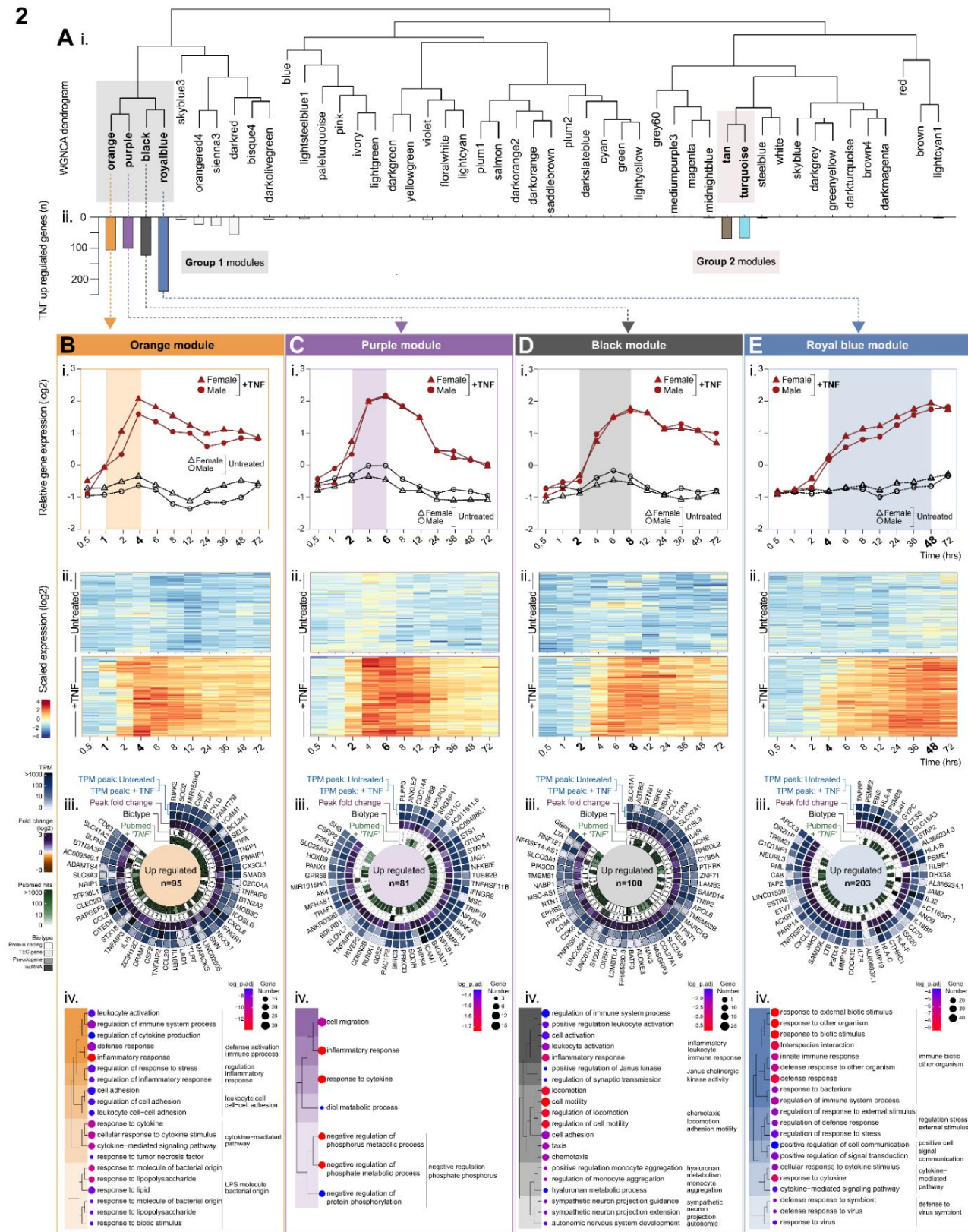
11

2

Figure 2. Weighted network correlation analysis (WGNCA) reveals temporal relationships between TNF up regulated genes: Group 1 ´early induced´. Human umbilical vein endothelial cells (EC, n=5) were treated with or without tumour necrosis factor alpha (TNF) and harvested at 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48 or 72 hrs, before RNAseq analysis. Weighted correlation network analysis (WGCNA) was used to cluster genes into modules, based on expression pattern similarity across sample sets. (**A**) (i) Dendrogram showing WGCNA modules and (ii) corresponding distribution of genes previously classified as TNF up regulated. For modules (**B**) *orange*, (**C**) *purple*, (**D**) *black*, (**E**) *royal blue*: (i) relative gene expression plots displaying the module eigengenes and (ii) heatmaps showing the temporal expression profile for all genes in the module with eigengene >0.8 (p<0.05). (iii) Circle plots for the top 50 genes classified as TNF-upregulated within each module: showing expression values in control and stimulated EC, the peak fold change, the biotype and number of PubMed hits for '*gene name*' + 'TNF', and (iv) over-represented terms by gene-ontology analysis (biological processes).

13

In contrast to the other group 1 modules, TNF up regulated genes in the *royal blue* module remained elevated across the 72-hour time course (Figure 2 E.i-ii) and contained genes encoding for several types of pattern recognition receptors (PRR), e.g., toll-like- (*TLR2, TLR5)*, RIG-I-like- *(DDX58*, *DHX58*, *IFIH1)* and NOD-like- (*NOD2*, *NLRC5*) receptors, and cyclic guanosine monophosphate adenosine monophosphate synthase (*cGAS*). This module also contained *ISG20,* an interferon stimulated gene (ISG) that encodes a nuclease enzyme that can cleave viral RNA. Gene ontology analysis of TNF up regulated genes in this module revealed that overrepresented terms included those linked to viral defence, such as *response to virus* (Figure 2 E.iv) (Table S2 D).

Up regulated group 2 - Delayed induction (12-24 hours post stimulation):

Group 2 modules [*turquoise* and *tan*] (Figure 3 A) contained genes that were up regulated between 12-24 hours post-TNF treatment, with the highest differential expression vs. control at 72-hours (Figure 3 Bi-ii and Ci-ii). Genes that were up regulated following TNF treatment in the *turquoise* module (Figure 3 B) included those encoding for a panel of interferon-induced cytokine ligands for the antigen presenting cell receptor CXCR3: *CXCL9*, *CXCL10* and *CXCL11* (Figure 3 B.iii), the expression of which is linked with viral infection, progression and replication control (Yin, Wang et al. 2019, Karin 2020, Callahan, Hawks et al. 2021). Other ISGs in this module included IFIT2, IFIT5 and IFIT35, that, to our knowledge, have not previously been reported as being modified following TNF stimulation. Gene ontology analysis of up regulated genes in the *turquoise* module revealed over representation of terms associated with *defence to virus symbiont* (Figure 3 B.iv), in addition to general inflammation terms, such as *response to cytokine* (Table S2 D). Genes that were up regulated following TNF treatment in the *tan* module also included a large panel of ISGs e.g., among others *IFIT1*, *IFI6*, *IFI27*, *OAS1/2/3/L*, *MX1/2* and *IFITM1* (Figure 3 C.iii) and, correspondingly, gene ontology analysis revealed a highly significant over representation of terms related to response to interferon and viral infection (Figure 3 C.iv) (Table S2 D). Thus, this latter stage of the response was dominated by the induction of interferon/anti-viral related gene transcription.
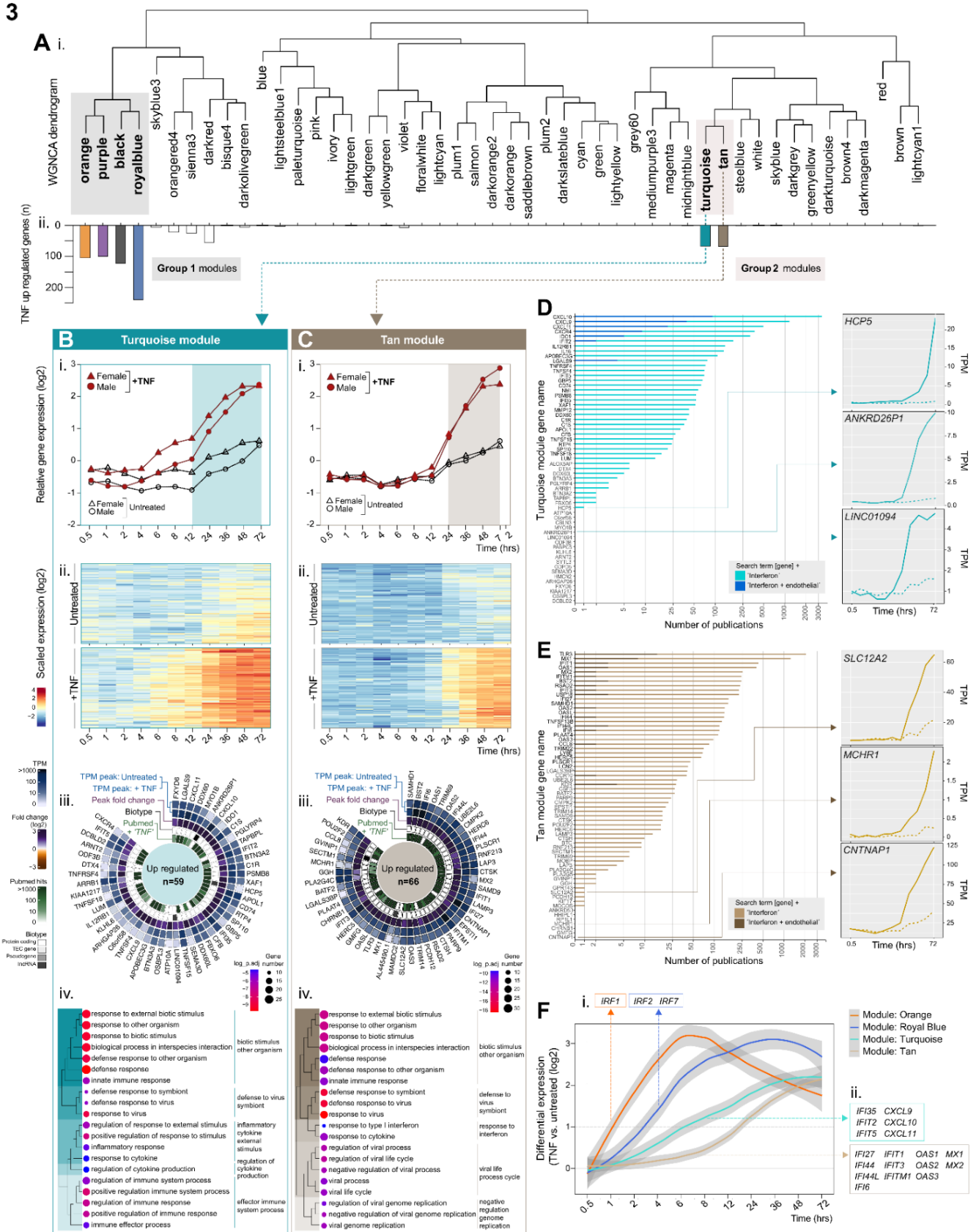
14

15

Figure 3. Weighted network correlation analysis (WGNCA) reveals temporal relationships between TNF up regulated genes: Group 2 ´delayed induced´. Human umbilical vein endothelial cells (EC, n=5) were treated with or without tumour necrosis factor alpha (TNF) and harvested at 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48 or 72 hrs, before RNAseq analysis. Weighted correlation network analysis (WGCNA) was used to cluster genes into modules, based on expression pattern similarity across sample sets. (**A**) (i) Dendrogram showing WGCNA modules and (ii) corresponding distribution of genes previously classified as TNF down regulated. For modules (**B**) *tan* and (**C**) *turquoise*: (i) relative gene expression plots displaying the module eigengenes and (ii) heatmaps showing the temporal expression profile for all genes in the module with correlation to the eigengene >0.8 (p<0.05). (iii) Circle plots for the top 50 genes classified as TNF-up regulated within each module: showing expression values in control and stimulated EC, the peak fold change, the biotype and number of PubMed hits for '*gene name*' + 'TNF', and (iv) over-represented terms by gene-ontology analysis (biological processes). (**D**) and (**E**) show the number of hits returned for TNF up regulated genes in the *tan* or *turquoise* modules, respectively, in a PubMed search for '*gene name*' + 'interferon' and '*gene name*' + 'interferon' + 'endothelial', with temporal expression plots for selected examples (created using the website tool provided as part of this study). (**F**) Temporal distribution of interferon-related genes upregulated by TNF across *orange*, *royal blue*, *tan*, and *turquoise* modules.

Indeed, a PubMed search for studies citing genes that were upregulated following TNF treatment in the *turquoise* (Figure 3D) and *tan* (Figure 3E) modules, together with the term 'interferon', revealed most had been previously reported in this context (40/59 [68%] and 57/66 [88%], respectively). Additional searches using related terms ([virus], [viral], [antiviral], [interferon stimulated gene], or [immune response]) showed a similar distribution of hits (Table S2 F). Whilst not all hits containing two terms necessarily imply a meaningful functional association, such an analysis can offer a perspective as to the likelihood of the absence or presence of a link. Only a small proportion of the genes had PubMed hits that included the terms ´interferon´ and ´endothelial´ (Figure 3 D and E, dark shaded bars) or ´viral´ and ´endothelial´ (Table S2 F), indicating that these pathways are less well understood in this cellular context. Genes in the *turquoise* module with only one, or no hits linking them to interferon (Figure 3D), included several non-coding genes, which are typically less well studied than protein coding genes, e.g., *HCP5,* which has polymorphisms linked to HIV viral load (Thorner, Erikstrup et al. 2016) and is a susceptibility locus for Kawasaki disease, a systemic vasculitis of infants and children (Kim, Yun et al. 2017), the uncharacterised pseudogene *ANKRD26P1,* and *LINC01094* (Figure 3 D, right panels). Thus, it is possible that such genes have currently unknown roles in EC interferon-related signalling, based on the similarity of their temporal expression profile with others in the module.

11 genes in the *tan* module had only one, or no hits linking them to interferon, including the solute transported *SLC12A2*, the G-protein coupled receptor *MCHR1* and the paranodal junction component *CNTNAP1* (Figure 3 E, right panel). The role of these genes in inflammation, and any possible connection to EC interferon and/or viral response signalling remains to be established.

ISG expression following TNF treatment is not driven by *de novo* interferon production

Although the mechanisms of ISG expression following TNF treatment of EC are not well understood, previous studies have reported that it is driven by *de novo* production and subsequent autocrine signalling of type I interferon (IFNβ) (Venkatesh, Ernandez et al. 2013). We found that three of the nine members of the interferon regulatory factor family (IRF1-9), which are critical for the induction of type I interferon (McNab, Mayer-Barber et

17

al. 2015), were up regulated following TNF stimulation, and all were found in the group 1 ´early responding´ modules *orange* (*IRF1*) and *royal blue* (*IRF2* and *IRF7*) (Figure 3 F.i), thus temporally preceding the induction of the majority of the ISGs, which fall in modules *turquoise* and *tan* (Figure 3 F.ii). However, only a modest induction of *IFNB1* was observed at later time points (max. any sample, any time point = 1.02 TPM [24h]) (mean values for all interferon proteins in Table S2 E); importantly, expression of many ISGs preceded the time point at which *IFNB1* was expressed at detectable levels e.g., *ISG20* (Figure 4 A.i), *IFIT3* (Figure 4 A.ii), *IFI35* (Figure 4 A.iii), *CXCL10* (Figure 4 A.iv) *CXCL11* (Figure 4 A.v), and *MX1* (Figure 4 A.vi). The same was also observed for various pattern recognition receptors, e.g., *DDX58* (Figure 4 A.vii), *TLR2* (Figure 4 A.viii), *NOD2* (Figure 4 A.ix), and *IFIH1* (Figure 4 A.x). Thus, in this system, the transcriptional induction of such genes was not driven by autocrine IFNβ signalling. Changes in the expression of genes encoding for key components of the interferon/anti-viral response pathways (JAK/STAT, NFκB-IRF1, TLR, RLR and cGAS/STING) following TNF treatment are summarised in Figure 4 B.
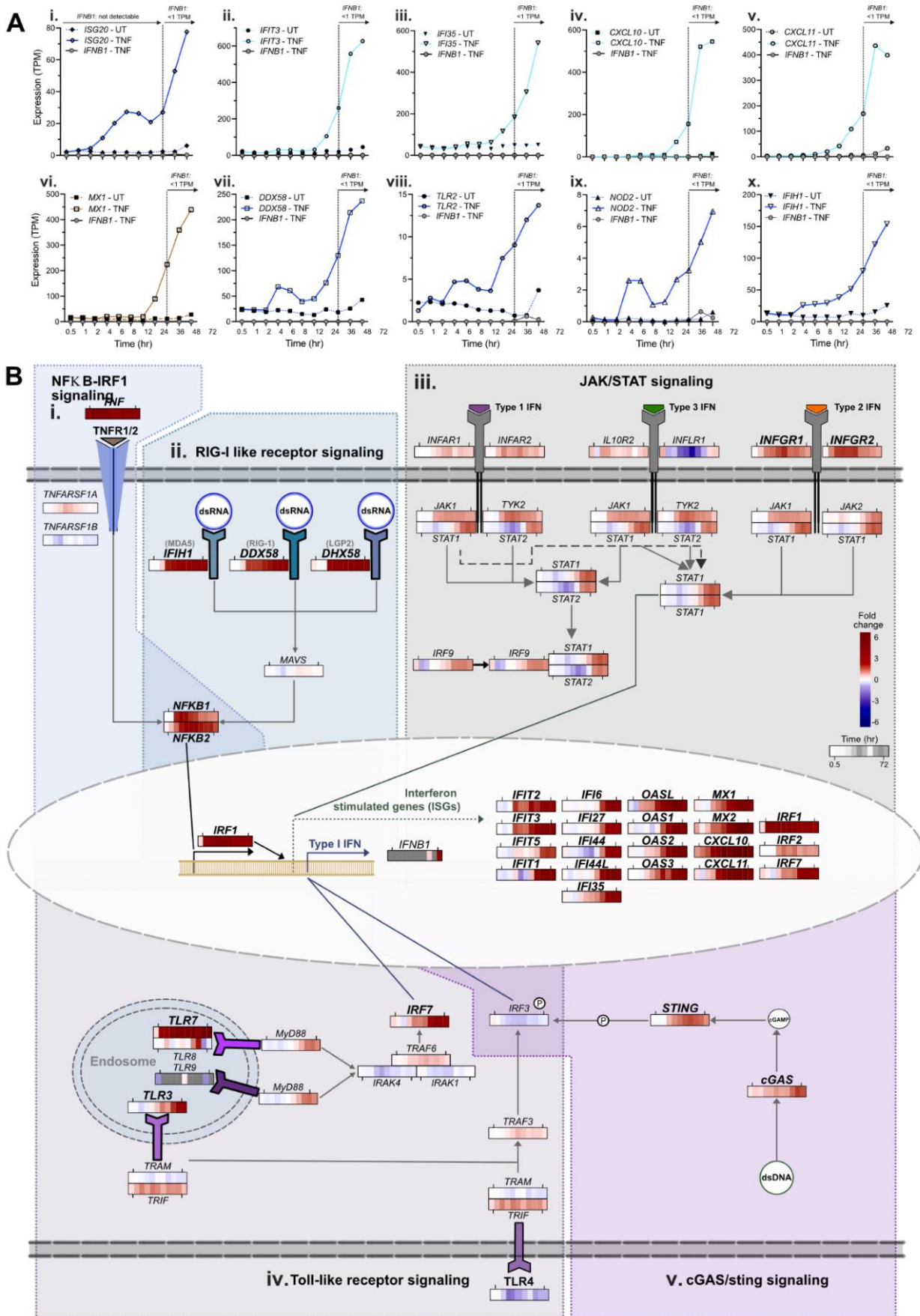
18

Figure 4. TNF up regulation of interferon-stimulated and pattern recognition receptor gene expression is independent of *de novo* interferon production. (**A**) Temporal expression profiles in unstimulated control or tumour necrosis factor alpha (TNF) stimulated EC for *IFNB1* and (i) *ISG20*, (ii) *IFIT3*, (iii) *IFI35*, (iv) *CXCL10*, (v) *CXCL11*, (vi) *MX1*, (vii) *DDX58*, (viii) *TRL2*, (ix) *NOD2* and (x) *IFIH1* (sample set F). (**B**) Summary of key genes and pathways linked to interferon stimulated gene expression: (i) NFKB-IRF1 signalling (adapted from Feng et al., 2021), (ii) RIG-I like receptor signalling (adapted from Rehwinkel and Gack., 2020), (iii) JAK-STAT signalling (adapted from Schneider et al., 2014), (iv) TLR-signalling (adapted from Duan et al., 2022), and (v) CGAS-Sting signalling (adapted from Feng et al., 2021). Heatmaps show the differential gene expression between unstimulated control and TNF stimulated EC for the adjacent gene. Grey squares in the heatmap are the result of zero TPM values, thus differential expression is not calculated. Bold gene symbols denote those that were classified as TNF upregulated. Heatmaps were created using the website tool provided as part of this study.

TNF down regulated genes have two main regulation profiles

The majority of genes classified as down regulated in response to TNF stimulation fell into two main regions on the WGNCA dendrogram (Figure 5 A.i), occupying neighbouring leaves on common clades: annotated as group 1 [*dark orange, saddle brown*] and group 2 [*green* and *light yellow*] modules (Figure 5 A.ii) (highlighted with shaded boxes). Modules in both groups contained genes that were down regulated by TNF between 2- and 4-hours post-stimulation (Figure 5 Bi-ii, Ci-ii, Di-ii, and Ei-ii).

Down regulated group 1 – Inhibition maintained over time:

Genes in the group 1 modules *dark orange* (Figure 5 B.i-ii) and *saddle brown* (Figure 5 C.i-ii) reached maximum down regulation 4- and 12-hours post stimulation, respectively, an effect that was maintained across the remainder of the time course. There were no significantly enriched gene ontology terms in lists of down regulated genes appearing in the group 1 down regulated modules, possibly due to the low numbers, or lack of previous reports of gene function.

Down regulated group 2 - Inhibition resolved over time:

In contrast to group 1, TNF down regulated genes in group 2 modules, *green* (Figure 5 D.i-ii) and *light yellow* (Figure 5 E.i-ii), trended back towards baseline level after reaching maximum down regulation at 6- and 12-hours post stimulation, respectively. Again, there were no significantly enriched gene ontology terms in any of lists of down regulated genes appearing in the highlighted modules; indeed, an automated PubMed search for the search terms [down regulated gene] and [TNF], retrieved markedly fewer hits than the equivalent search for genes we previously classified as up regulated by TNF (Figure 5 B-E.iii). Despite a lack of significant enrichment terms, some links between constituent genes could be observed e.g., four out of five TNF down regulated homeobox family transcription factor genes were classified into module *green* (*HOXA6*, *HOXA10, HOXA11* and *HOXD8*) (Figure 5 D.iii). Whilst *HOX10* has been reported as an activator of canonical NF-κB signalling in pancreatic cancer cells (Li, Chang et al. 2022), and the antisense to *HOXA11* (*HOXA11-AS*) linked to protection of EC barrier function following injury (Yuan, Yuan et al. 2022), insight into the potential role of these genes in the EC TNF response is currently lacking.
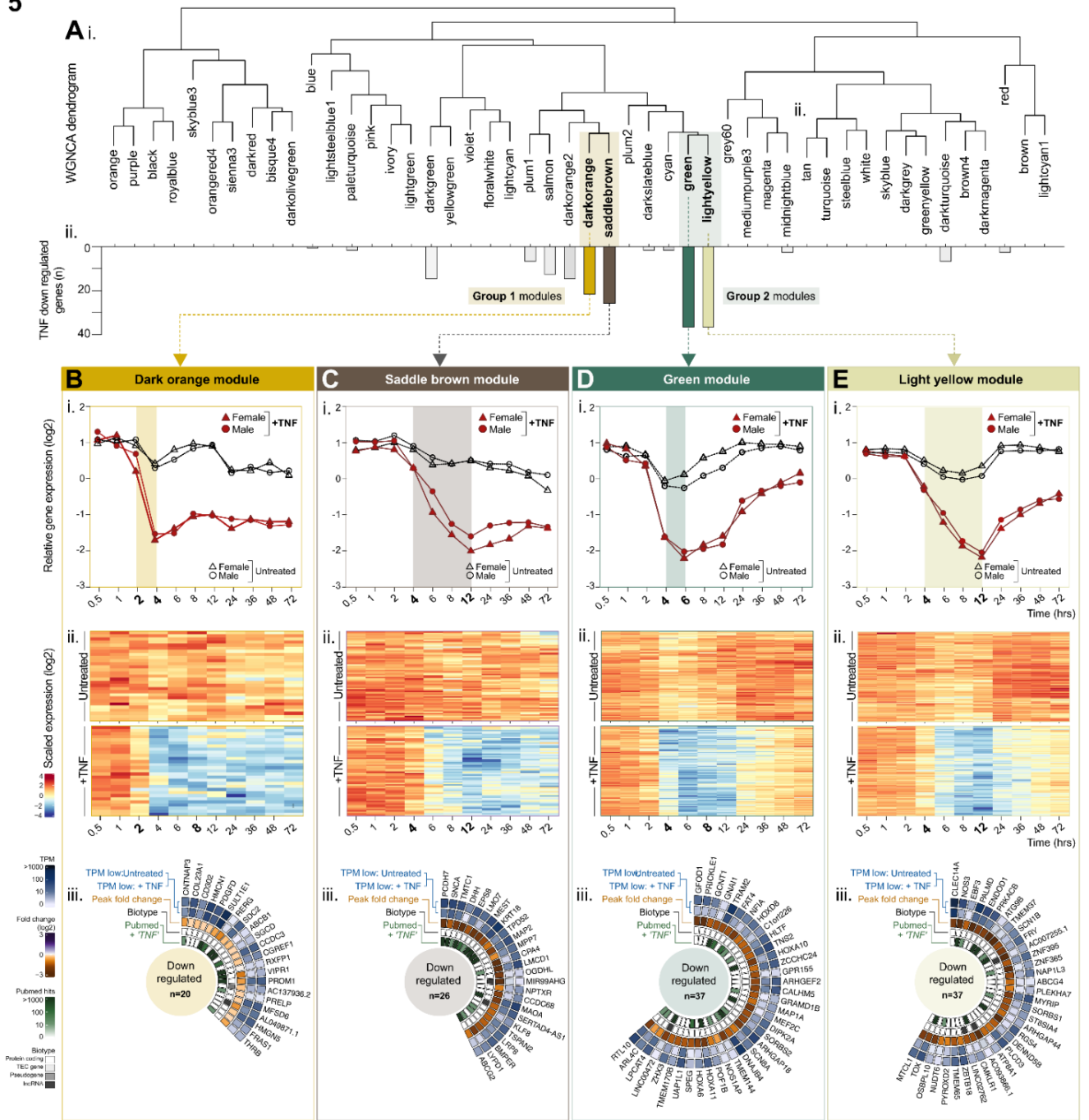
5

Figure 5. Weighted network correlation analysis (WGNCA) reveals temporal relationships between TNF down regulated genes. Human umbilical vein endothelial cells (EC, n=5) were treated with or without tumour necrosis factor alpha (TNF) and harvested at 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48 or 72 hrs, before RNAseq analysis. Weighted correlation network analysis (WGCNA) was used to cluster genes into modules, based on expression pattern similarity across sample sets. (**A**) (i) Dendrogram showing WGCNA modules and (ii) corresponding distribution of genes previously classified as TNF down regulated. For modules (**B**) dark orange, (**C**) saddle brown, (**D**) green, (**E)** light yellow: (i) relative gene expression plots displaying the module eigengenes and (ii) heatmaps showing the temporal expression profile for all genes in the module with correlation to the eigengene >0.8 (p<0.05). (iii) Circle plots for genes classified as TNF-down regulated within each module: showing expression values in control and stimulated EC, the peak fold change, the biotype and number of PubMed hits for '*gene name'* + 'TNF'.

23

In both TNF down regulated modules group 1 and 2, gene baseline expression levels and TNF-response profiles were similar between male and female samples (Figure 5 B.i-E.i and Figure S2 C).

Visualisation of temporal gene regulation pathways using the website tool

We have created a website resource that allows users to perform both gene centric or module-based lookup of our endothelial TNF time course data. Key features include a data viewer, to observe the transcriptional responses of specific genes (Figure S3 A), or the WGCNA module into which they were classified (Figure S3 C), and the generation of vector-image downloadable expression plots for both predefined and custom gene lists, e.g., TNF regulated CXCL- and CCL-chemokines (Figure S3 B). The dataset can also be analysed to identify genes with high correlation across conditions with any given input gene (Figure S3 D).

24

DISCUSSION

Here, we used RNA sequencing to measure the temporal response of EC to TNF stimulation, incorporating 11 time points up to 72h. Following the identification of TNF regulated genes, we used weighted network correlation analysis to understand the global temporal context of these transcriptomic changes. To our knowledge, this is the first study to map the EC TNF response at a transcriptome-wide level in such temporal detail.

We identified two main profiles into which TNF-induced genes could be classified - those with an ´early´ or ´delayed´ induction. Early induced genes (~1-4 hours post stimulation) included many previously well studied in this context, such as those encoding for EC leukocyte adhesion receptors (e.g., *SELE*, *VCAM1* and *ICAM1*) (Pober 2002), but we also identified several genes with similar expression dynamics that encoded for completely uncharacterised proteins (e.g.. *FAM177B*), which could be interesting candidates for future study in the context of inflammation. TNF up regulated genes with a delayed induction (12-24 hours post stimulation) were primarily ISGs, whose regulation by inflammatory cytokines in EC is generally not well understood, but our observations were consistent with one recent study which reported the induction of a late stage interferon response in EC, following TNF stimulation (Valenzuela 2022). ISG expression is primarily considered to be driven by the production and subsequent signalling of interferon, via canonical (JAK-STAT) or non-canonical pathways (Mazewski, Perez et al. 2020).

We found that three out of the nine members of the interferon regulatory factor family (*IRF1*, *2* and *7*), which are critical for the induction of interferon (McNab, Mayer-Barber et al. 2015), were up regulated by TNF in EC at time points that preceded ISG expression. Of these, *IRF1* and *IRF7* have been implicated as positive regulators of type I interferon production (Honda, Takaoka et al. 2006) and previous reports have shown that TNF-induced expression of ISGs, such as *CXCL9* and *CXCL10*, in murine EC was dependent on *IRF1*-induced *de novo* production of IFNβ (*IFNB1*), and its subsequent autocrine signalling through STAT1 (Venkatesh, Ernandez et al. 2013). However, our data indicated that EC expression of ISGs following TNF treatment was independent of *de novo* production of interferon, as we did not observe its transcription prior to ISG expression.

We observed an up regulation of genes encoding for cyclic GMP–AMP (CGAS) and the cyclic GMP–AMP receptor stimulator of interferon genes (STING), a system which detects

pathogenic DNA (Hopfner and Hornung 2020). A recent study showed that the expression of various ISGs that were induced by TNF in fibroblasts, including *CXCL10*, *IFIT1* and *IFIT44* (all of which were also up regulated by TNF in the current study) was markedly reduced in CGAS and STING knockout cells (Willemsen, Neuhoff et al. 2021). TNF-dependent mitochondrial damage and mtDNA leakage was shown to underlie this response; one could speculate a similar mechanism contributes occurs in EC.

We observed an up regulation of genes encoding for other pattern recognition receptors, including toll-like- (*TLR2, TLR5)*, RIG-I-like- *(DDX58*, *DHX58*, *IFIH1)* and NOD-like- (*NOD2*, *NLRC5*) receptors. Whilst these receptors are known to induce production of interferon and subsequent expression of ISG in response to bacterial or viral ligands, (Uematsu and Akira 2007, Opitz, Eitel et al. 2009), whether or not they have a role in the induction of ISG in EC following TNF production, similar to that reported for CGAS and STING (Willemsen, Neuhoff et al. 2021), remains to be explored.

We identified several non-coding RNAs within the gene modules that otherwise predominantly contained ISGs, including ENSG00000225886 (antisense to *IFI6*), *NRIR*, a negative regulator of SARS-CoV-2 infection (Enguita, Leitao et al. 2022) and *LINC02056*, an interferon-inducible transcript with a proposed role in IRF3 nuclear translocation (Xu, Yu et al. 2021). As is often the case with non-coding genes, functional annotation of others was lacking e.g., *LINC02051* and *LINC02068*; these are potentially interesting candidates to study in the context of the EC interferon response. Overall, deciphering the relative contributions of various pathways in the TNF-induced expression of ISGs is complex, with potential differences between cell types and species.

We identified a panel of 210 gene transcripts that were at lower levels following TNF-stimulation. Whilst studies of the TNF response tend to focus on genes whose expression in increased in response to stimulation, several of those we identified as down regulated had been previously reported as such, e.g., *NOS3*, the mRNA stability of which is inhibited by TNF (Yan, You et al. 2008), *DHH*, which prevents EC activation (Chapouly, Hollier et al. 2020), and *RGS4*, which regulates the secretion of VWF (Patella and Cutler 2020). However, many had not been previously reported in this context, e.g., *CNR1*, *SLC7A8* and *CLEC14A*, which were amongst the most down regulated by TNF.

Sex differences have been reported in several inflammatory conditions of the vasculature, such as cardiovascular disease (Gao, Chen et al. 2019) and thrombosis (Nordstrom and Weiss 2008). Whilst our data indicates that chromosomal composition alone does not markedly affect the EC response to TNF, a multitude of other factors influence vascular responses *in vivo*, such as sex hormones, which may drive sex linked inflammatory differences (Rathod, Kapil et al. 2017, Pabbidi, Kuppusamy et al. 2018).

***Study strengths and limitations*:** One of the main strengths of our study is size of the dataset generated; we analysed 130 samples, incorporating 35 biological replicates. The global EC transcriptome was analysed at 11 different time points post-TNF treatment, and the inclusion of matched control samples for each sample set, at every time point, allowed us to control for baseline transcriptional changes, such as those due to changes in the microenvironment (Majewska, Wilkus et al. 2021) or cell density (Hamada, Osaka et al. 2014), which could otherwise be incorrectly annotated as TNF driven. To our knowledge, the website resource we provide (http://www.endothelial-response.org/) is the most extensive of its type; all data is accessible without the need for bioinformatic expertise.

The EC we used in our study were isolated from human umbilical veins (HUVEC), from which we could generate a large amount of fresh primary EC. Thus, we avoided the need to passage, freeze/thaw, or culture the EC for a prolonged period prior to treatment and analysis, factors that could affect behaviour and response to cytokines (Liao, He et al. 2014, Ohori, Nakayama et al. 2021). However, it should be acknowledged that this EC type is foetal, rather than adult, and differences have been reported between the two, including the transcriptional response to TNF (Viemann, Goebeler et al. 2006). Comparisons of microarray data for human dermal microvascular EC (HMEC1) and HUVEC have shown that around half of the TNF induced changes were specific for only one or other of these EC types (Viemann, Goebeler et al. 2006). However, key genes that were highlighted as only up regulated by TNF in HMEC1, but not HUVEC (e.g., *IL1B*, *DUSP6*, *OAS1*, *CLDN1*, *CD70*, *MMP12*), were classified as up regulated in our HUVEC dataset - potentially indicating that other factors, such as the sensitivity of EC types to passage in culture or freeze thaw cycles could influence response, as opposed to core characteristics of the EC types *per se*. Indeed, other studies show more similar characteristics between EC types, such as the response to shear stress exposure, which

27

was largely comparable between human adult aortic cells and HUVEC (Maurya, Gupta et al. 2021). Furthermore, EC have organ-specific heterogeneity (Kalucka, de Rooij et al. 2020, Tabula Sapiens, Jones et al. 2022), meaning that vascular bed specific responses to TNF might not be comparable to those we observed in HUVEC.

It should also be noted that EC in our study are not cultured under flow, or together with other cell types found in the normal microenvironment, both factors that can affect *in vitro* gene expression (Heydarkhan-Hagvall, Helenius et al. 2003, Nakajima and Mochizuki 2017, Helle, Ampuja et al. 2021, Afshar, Ma et al. 2023). Exposure to different levels of flow *in vitro* can modify the EC response to TNF (Sheikh, Rainger et al. 2003) and thus, our data may be more representative of EC responses in low, rather than high, shear exposed vessels. Finally, although under steady-state conditions protein expression is highly dependent on mRNA level (Liu, Beyer et al. 2016), the relationship between the two after state transitions, such as those induced by TNF, is subject to time dependent processes, such as maturation, export and translation of mRNA (Liu, Beyer et al. 2016). Thus, there will be a delay between transcriptional changes and the associated protein level increase or decrease. Although extensive, our dataset does not provide a comprehensive overview of all aspects of the TNF response; the majority of microRNAs are not profiled, due to the RNA isolation method used, and the release of stored and secreted immediate responders, e.g., P-selectin and Von Willebrand factor (Metcalf, Nightingale et al. 2008), and processes such as protein phosphorylation and nuclear translocation are not measured.

MATERIALS AND METHODS

LEAD CONTACT

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact: Dr. Lynn Marie Butler. Email: Lynn.butler@ki.se

MATERIALS AVAILABILITY

This study did not generate new unique reagents.

DATA AND CODE AVAILABILITY

The data generated by this study is publicly available. Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Isolation, culture, and sex-determination of human umbilical vein endothelial cells

Human umbilical vein endothelial cells (EC) were isolated from human umbilical cords, collected from Karolinska University Hospital, Stockholm, Sweden, as described(Cooke, Usami et al. 1993). Ethical approval was granted by *Regionala etikprövningsnämnden i Stockholm* (2015/1294-31/2). EC were cultured in Medium M199, supplemented with 10% foetal bovine serum, 10 ml/l penicillin-Streptomycin, 2.5 mg/l Amphotericin B (all ThermoFisher, Gibco), 1 mg/l hydrocortisone and 1 µg/l and human epidermal growth factor (hEGF) (both Merck). To determine EC sex, transcripts encoding *Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-Linked* (*UTY*) were measured by qPCR. Cell lysis and cDNA generation was performed using the 2-Step Fast-Cells-to-CT-Kit (Invitrogen, ThermoFisher) according to their protocols. qPCR was performed using TaqMan Fast Universal PCR mix and target (UTY) primer conjugated to FAM-probe (Hs01076483, ThermoFisher) with 18s rRNA primer (4319413E conjugated to VIC probe, ThermoFisher) as endogenous control. qPCR was performed using a RealTime PCR LightCycler 96 ® system (Roche Life Sciences). EC positive for *UTY* expression were classified male, and those negative classified female. 5-6 sex-matched biological replicates were pooled to create each sample set (33 donors in total). Sample sets were A, B, C (annotated male) and D, E, F (annotated female). Following sequencing, a low level of Y-linked transcripts were detected in sample set F, indicating an incorrect

29

annotation of one of the constituent donors as female. Thus, this sample was excluded from any subsequent sex-based analysis but included in non sex-split analyses.

EC treatment, RNA isolation and sequencing

Pooled HUVEC sample sets were grown to confluence before treatment with or without recombinant tumour necrosis factor alpha (TNF; 10 ng/mL) (ThermoFisher) in cell culture medium. HUVEC treated with a concentration of 10 ng/mL TNF express key markers of inflammation, such as *SELE*, *VCAM1* and *ICAM1* (Mackay, Loetscher et al. 1993, Majewska, Paleolog et al. 1997) and can support leukocyte recruitment (Butler, Rainger et al. 2005), whilst maintaining viability over a prolonged culture period (Khan, Awad et al. 2017). HUVEC maintained morphological features and confluence over the time course. Following the initial treatment with or without TNF, the cell culture medium was not changed for the remainder of the time course, to minimise cell exposure to external shear forces and/or temperature fluctuations, factors associated with changes in gene expression in EC (Braddock, Schwachtgen et al. 1998, Horioka, Tanaka et al. 2020). EC were lysed at 0.5, 1, 2, 3, 4, 6, 8, 12, 24, 36, 48 or 72 hours post-stimulation, using RLT lysis buffer from the RNAeasy mini kit (Qiagen).

RNA isolation and purification was performed using the RNAeasy mini kit (Qiagen). RNA concentration was measured using Nanodrop 2000 spectrophotometer and RNA integrity number (RIN) determined using Agilent 2100 Bioanalyzer (RIN>9 required for inclusion). Library preparation and RNA sequencing was performed by the National Genomics Infrastructure Sweden (NGI) using Illumina stranded TruSeq poly-A selection kit and Illumina NovaSeq6000S (4 lanes, 2x 150bp reads, incl 2Xp kits). The data was processed using demultiplexing. Data storage and initial analyses were performed using server sided computation supplied by the Swedish National Infrastructure for Computing (SNIC). Genome assembly used for sequence alignment: Homo_sapiens.GRCh38.dna.primary_assembly.fa and annotation performed using: Homo_sapiens.GRCh38.96.gtf. Sequence alignment was carried out using STAR/2.5.3a. Gene mapping has been carried out using subread/1.5.2 and the module feature counts. Transcript mapping carried out using Salmon/0.9.1.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data normalisation and differential gene expression analysis

We used the "DESeq2" package in R to normalise raw gene expression counts (Love, Huber et al. 2014), which were log2 transformed and averaged across biological replicates. Genes with total DESeq count <10 across all samples were excluded and differentially expressed genes (DEG) between untreated control and TNF-treated EC, at each time point, were identified using the DESeq2 time series design. In this case, the full model is represented by ~treatment + time+ treatment:time and the reduced model by ~treatment+time. To correct for multiple sampling for DEG p values were adjusted using Benjamini-Hochberg (BH) correction. DEG were defined as those with adjusted P-value <0.05 and an absolute fold change between untreated control and TNF treated EC of log2 >1. Female and male samples were analysed separately, thus excluding a possible influence of sex-based discordant baseline expression levels, and data was later merged, where stated. In total, comparisons for 11 time points (hours: 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48, 72) were performed. Classifications as DEG defined by DESeq2 (without application of additional criteria as described below) and associated statistical information is provided in Table S3 B.

To see the clustering of samples, multidimensional scaling (MDS) plots were generated to visualise the clustering patterns among samples. Pairwise Euclidean distances between samples were calculated and scaled using classical MDS transformation. The resulting MDS coordinates were used to create a two-dimensional scatter plot using the ggplot package.

Thresholding and classification of genes as TNF-regulated

To identify genes that were regulated by TNF we applied several additional criteria. Raw gene counts (Table S3 A) were normalised to TPM (transcripts per kilobase per million reads mapped) and genes with low expression (max TPM <1 at any time point or condition) were excluded. Of those remaining, 4262 genes were classified as DEG by DEseq2 (Figure S1 B.i). When regulated at multiple time points, DEG had a consistent effect direction (i.e., showing only positive or negative regulation in response to TNF). Genes that were not differentially expressed at two or more sequential time points were excluded from further categorisation (Figure S2 B.ii). Genes were classified as TNF up regulated when the following additional criteria were fulfilled (Figure S2 C.i): (i) a low variation of expression over time in untreated EC (coefficient of variation [CV] of mean

31

expression across time points <0.3) (ii) high variation of expression over time in TNF treated samples (CV >0.3) and (iii) a high minimal fold change (log2 FC>0.7) vs. the initial timepoint (0.5 h). Positive DEG that did not fulfil these criteria were excluded (Figure S1 C.ii), as the differential expression between untreated and TNF stimulated EC was driven by expression changes in untreated EC over time (Figure S1 C.ii). Genes were classified as TNF down regulated when the following criteria were fulfilled: (i) a low variation of expression over time in untreated EC (coefficient of variation [CV] < 0.3), (ii) a high variation of expression over time in the TNF-treated samples (CV >0.3) and (iii) a minimum log2 absolute fold change of >0.7 in TNF treated EC vs. control untreated EC (Figure S1 D.i). Negative DEG that did not fulfil these criteria were excluded (Figure S1 D ii-iii), as the differential expression was primarily driven by either a TNF-induced delay (Figure S1 D.ii) or inhibition (Figure S2.iii) of expression in control untreated EC over time.

We applied these criteria to highlight genes with clear, distinct TNF-regulated expression profiles; our classifications should thus be viewed as an illustrative guide, rather than a comprehensive categorisation. Indeed, individual regulation profiles for genes of interest are best considered on a gene-by-gene basis.

Gene co-expression network construction

A gene co-expression network was constructed using the WGCNA package in R (Langfelder and Horvath 2008). After filtering of low expressed genes (total DESeq2 counts over the time course <10) and genes with low variation across all samples (CV >0.2), 12 428 genes were remained for further analysis. The appropriate soft-thresholding power was selected by applying "pickSoftThreshold" function with parameter "networkType" set to signed hybrid. Then the correlation network adjacency matrix was calculated using selected soft thresholding of 12 and with parameter "networkType" set to signed hybrid. The adjacency matrix was turned into topological overlap (TOM) and the corresponding dissimilarity (dissTOM) was calculated. Finally, average linkage hierarchical clustering was applied according to the dissTOM value and gene modules were identified using dynamic tree cut algorithm with the minimum module size of 30 genes. Module eigengenes (MEs) representing the first principal component of each expression module were calculated for each module. Genes with correlation to the

eigengene >0.80 (p-value <0.05) were featured in the figures and associated subsequent analysis. Heatmaps show log2 transformed scaled expression, as calculated by DSeq2.

Module visualisation

The profiles of each module were visualised using a heatmap of scaled gene expression profiles. Additionally, plots displaying the expression of each gene within module and the expression of its eigengene were produced (Langfelder and Horvath 2008), which returns the gene in each module with the highest connectivity.

Functional enrichment of co-expression modules and DEGs

The Gene Ontology Consortium (Ashburner, Ball et al. 2000) and PANTHER classification resource (Mi, 2019) were used to identify overrepresented terms in gene lists using the GO databases (release date 2023-07-05). Plots of GO terms were created using the R package clusterProfiler (Wu, Hu et al. 2021).

Additional statistical analyses and website development

Graphs were created using the packages ggplot2 (Wickham 2016) and base R(Team 2022) and GraphPad Prism. Temporal graphs for Figures 3D, 3E, 4, S2 and S3 were created using our website tool. Circle plots were created by using the R package circlize (Gu, Gu et al. 2014) and PubMed lookups were performed using the R package easyPubMed by Damiano Fantini (Fantini 2019) (date of lookup 01.03.2023). Figures were assembled using Affinity Designer and Adobe Illustrator.

The following additional R packages were used for mapping, normalisation, clustering, and display of the data: readR, dplyr, data. table, matrixStats, NormExpression, edgeR, viridis, RcolorBrewer.

Further statistical analyses and website implementation were performed in RStudio (R version 4.0.3) using shinyapps. The following packages were used: shiny (Chang, Cheng et al. 2022), shinyjs, gplot, ggplot2, DT, plotly and ggthemes.

The website is available here: http://www.endothelial-response.org/

Website and data availability resource

Average Gene TPM and DEG values (available as log2, log10 or decimal) for each timepoint are available in Table S1 A and B. Raw gene counts are available in Table S3 A or from https://github.com/PhilipDusart/TNF_timecourse. TPM, DESeq2 and DEG values for each individual donor pool, and data for modules can be downloaded from

http://www.endothelial-response.org/ All gene expression profile plots and data for all genes can be downloaded directly from the website. The data is available as TPM or DEG fold change values and can be split by sex and filtered by observation timeframes of interest.

AUTHOR CONTRIBUTIONS

Conceptualisation: LMB, ECS, PD. Methodology: LMB, ECS, PD. Formal analysis: ECS, TB, PD, PH, MLK. Investigation: LMB, ECS, PD, TB, PH, MLK. Writing – Original Draft: ECS, LMB, PD. Writing – Review & Editing: All. Website development: ECS, PD. Visualisation: ECS, TB, LMB, PD. Supervision: LMB, PD, MLK. Resources: LMB, JO, MLK. Funding Acquisition: LMB, JO, MLK.

ACKNOWLEDGEMENTS

DECLARATION OF INTERESTS

The authors declare no competing interests.

34

S1

Figure S1. Experimental design and gene classifications. (**A**) Overview of the experimental procedure. (i) Endothelial cells were extracted from human umbilical veins and cultured to confluency before (ii) qPCR for Y-chromosome gene *UTY* was used to identify male vs. female donors. (iii) Cells were pooled into sex-matched sample sets, grown to confluency and (iv) stimulated with tumour necrosis factor alpha (TNF; 10 ng/mL), before RNA extraction at 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48 or 72 hours post stimulation, followed by (v) RNA sequencing analysis. (**B**) Sankey plots displaying (i) total number of genes detected and classified as differentially expressed genes (DEG) (ii) numbers of positive and negative DEG, and the subsequent classification as (**C**) Positive DEGs as: (i) *up regulated* by TNF from a stable baseline expression in control EC, or (ii) *stable* on the background of reduced baseline expression in control EC over time. (**D**) Negative DEGs as: (i) *down regulated* by TNF from a stable baseline expression in control EC, or (ii) *delayed* or (iii) *inhibited* by TNF, where baseline gene expression in control EC increases over time, but this change either occurs later or is not observed in TNF treated EC, respectively. Genes that did not fall into these groups were categorised as ´other´.
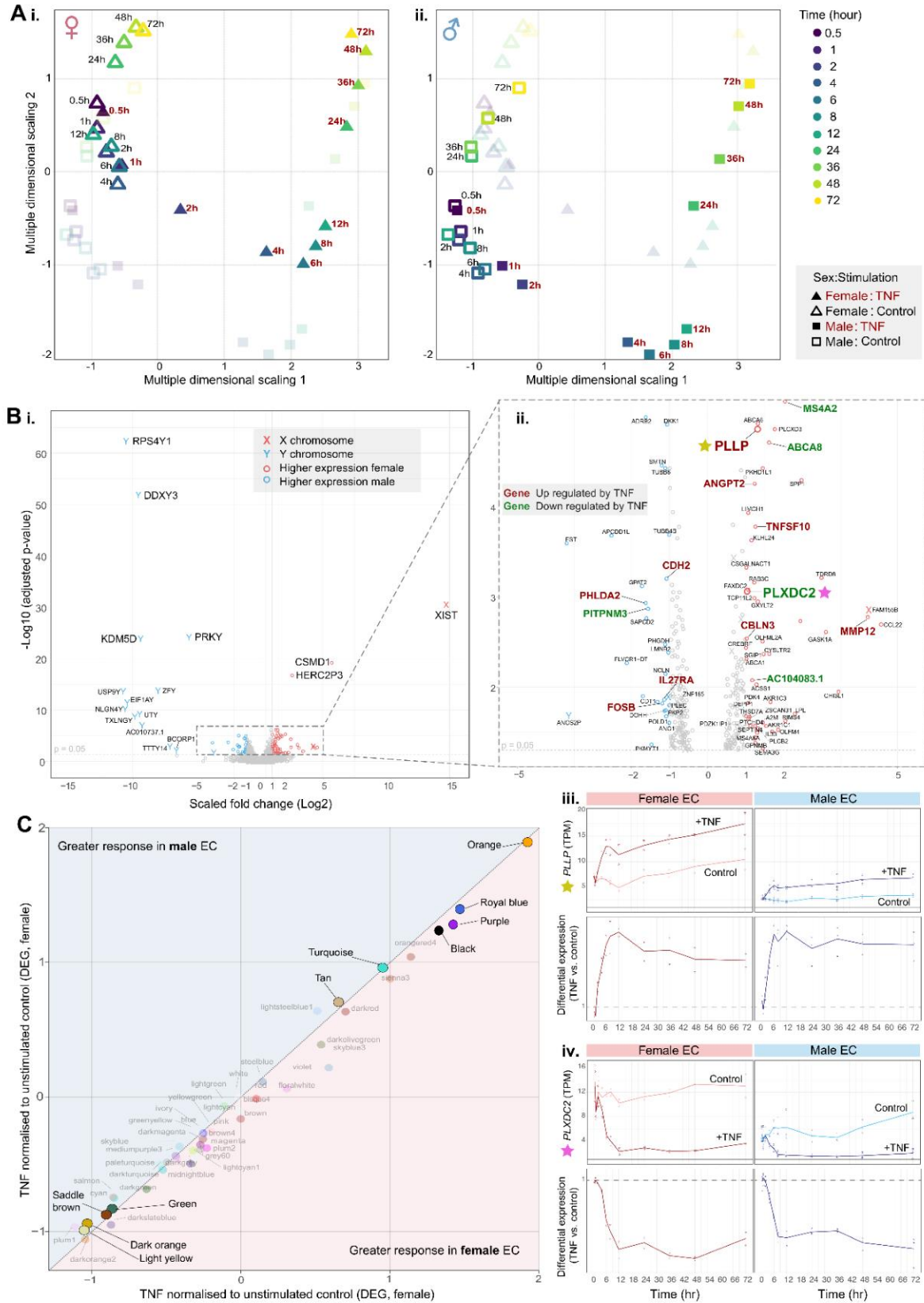
36

S2

Figure S2. Sex-based comparison of global gene expression profiles. Human umbilical vein endothelial cells (EC, male n=3, female n=2) were treated with or without tumour necrosis factor alpha (TNF) and harvested at 0.5, 1, 2, 4, 6, 8, 12, 24, 36, 48 or 72 hrs, before RNAseq analysis. (**A**) Multidimensional scaling plot for control or TNF treated (i) female or (ii) male EC, at all analysed time points. (**B**) (i) Volcano plot displaying differentially expressed genes between male and female EC under baseline (unstimulated control) conditions, with X- and Y-chromosomal genes annotated with ´Y´ and ´X´, respectively (ii) panel showing all transcripts classified as sex-differentially expressed at baseline (mean expression [in either sex] TPM>1), with those also classified as up or down regulated by TNF highlighted in larger red or green text, respectively. Expression data is provided for example genes annotated with star symbols: (iii) *PLLP* and (iv) *PLXDC2*. (**C**) Scatter plot comparing normalised average expression values between male and female samples for each WGCNA module.
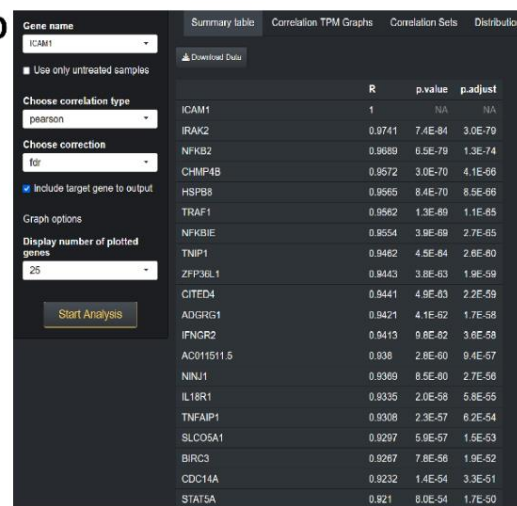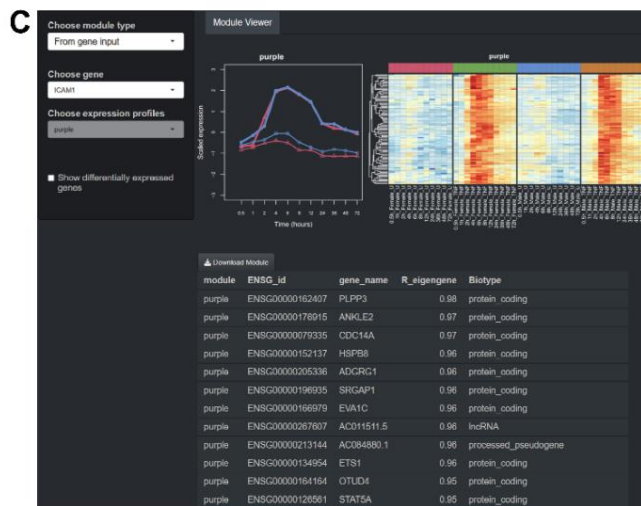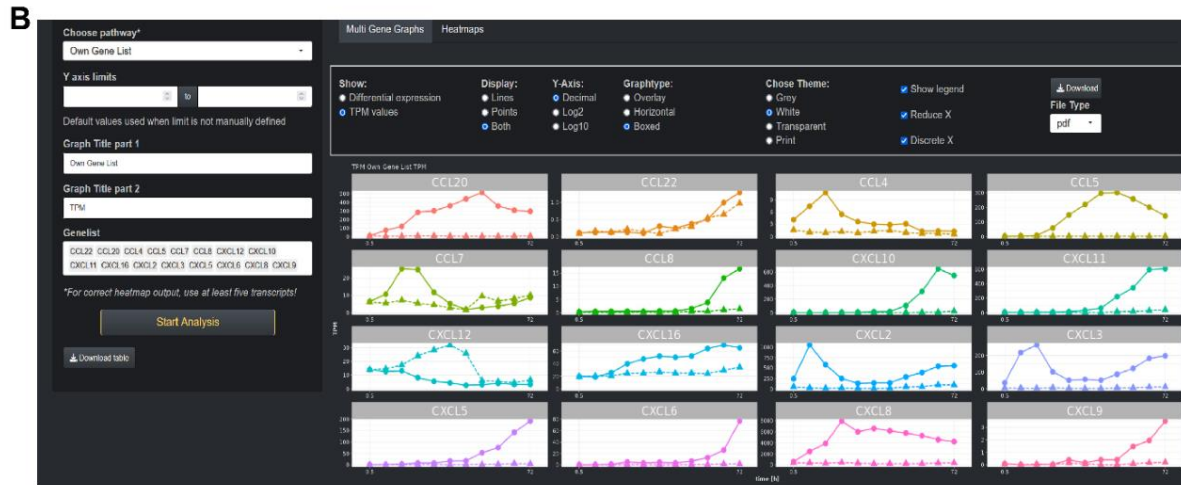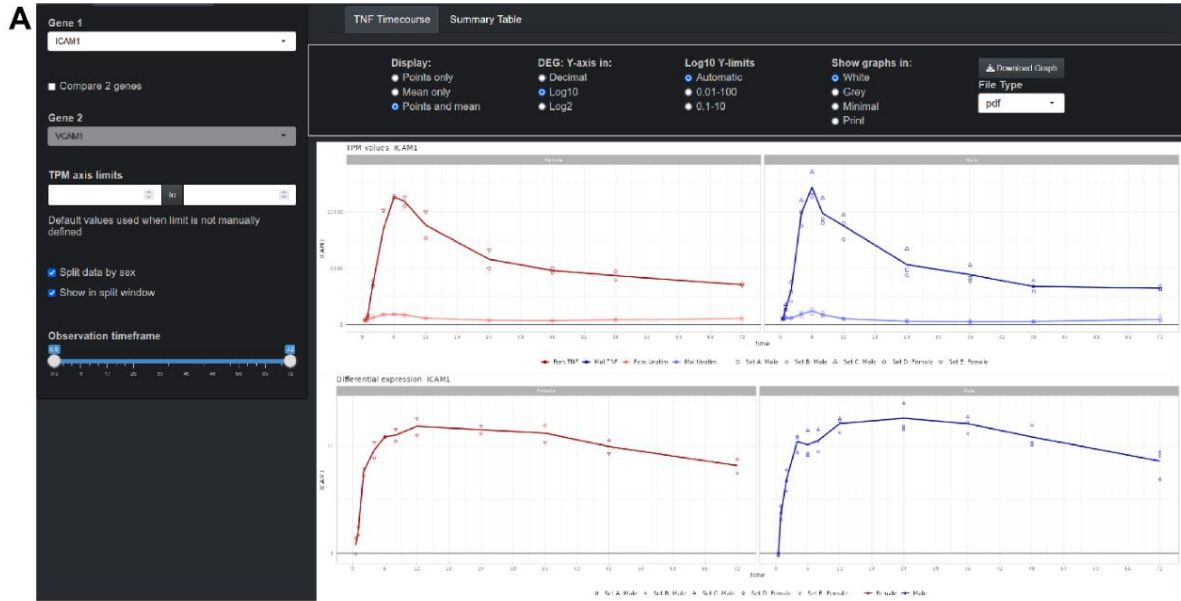
38

S3

Figure S3. Website resource examples. All data generated in this study is available on http://www.endothelial-response.org/. Selected features include: (**A**) data viewer for temporal gene expression over time (displayed as absolute values and relative differential expression), (**B**) data plot generator to be used either with user defined gene lists, or predefined gene categories, selected from the dropdown menu (e.g., ʹleukocyte recruitmentʹ), (**C**) weighted network correlation analysis data section and module look up tool for any given input gene, (**D**) Expression similarity tool that can identify genes with the highest correlation to any given input gene across the dataset. Data plots can be downloaded as vector-images

40

REFERENCES

Addis, R., I. Campesi, M. Fois, G. Capobianco, S. Dessole, G. Fenu, A. Montella, M. G. Cattaneo, L. M. Vicentini and F. Franconi (2014). "Human umbilical endothelial cells (HUVECs) have a sex: characterisation of the phenotype of male and female cells." Biology of Sex Differences **5**(1).

Afshar, Y., F. Ma, A. Quach, A. Jeong, H. L. Sunshine, V. Freitas, Y. Jami-Alahmadi, R. Helaers, X. Li, M. Pellegrini, J. A. Wohlschlegel, C. E. Romanoski, M. Vikkula and M. L. Iruela-Arispe (2023). "Transcriptional drifts associated with environmental changes in endothelial cells." Elife **12**.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene Ontology: tool for the unification of biology." Nature Genetics **25**(1): 25-29.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.

Barros Ferreira, L., L. M. Ashander, B. Appukuttan, Y. Ma, K. A. Williams and J. R. Smith (2022). "Expression of Long Non-Coding RNAs in Activated Human Retinal Vascular Endothelial Cells." Ocul Immunol Inflamm: 1-6.

Baud, V. and M. Karin (2001). "Signal transduction by tumor necrosis factor and its relatives." Trends in Cell Biology **11**(9): 372-377.

Bouwmeester, T., A. Bauch, H. Ruffner, P. O. Angrand, G. Bergamini, K. Croughton, C. Cruciat, D. Eberhard, J. Gagneur, S. Ghidelli, C. Hopf, B. Huhse, R. Mangano, A. M. Michon, M. Schirle, J. Schlegl, M. Schwab, M. A. Stein, A. Bauer, G. Casari, G. Drewes, A. C. Gavin, D. B. Jackson, G. Joberty, G. Neubauer, J. Rick, B. Kuster and G. Superti-Furga (2004). "A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway." Nat Cell Biol **6**(2): 97-105.

Braddock, M., J. L. Schwachtgen, P. Houston, M. C. Dickson, M. J. Lee and C. J. Campbell (1998). "Fluid Shear Stress Modulation of Gene Expression in Endothelial Cells." News Physiol Sci **13**: 241-246.

Brandt, M., V. Gerke and T. Betz (2022). "Human endothelial cells display a rapid tensional stress increase in response to tumor necrosis factor-alpha." PLoS One **17**(6): e0270197.

Butler, L. M., G. E. Rainger, M. Rahman and G. B. Nash (2005). "Prolonged culture of endothelial cells and deposition of basement membrane modify the recruitment of neutrophils." Exp Cell Res **310**(1): 22-32.

Callahan, V., S. Hawks, M. A. Crawford, C. W. Lehman, H. A. Morrison, H. M. Ivester, I. Akhrymuk, N. Boghdeh, R. Flor, C. V. Finkielstein, I. C. Allen, J. Weger-Lucarelli, N. Duggal, M. A. Hughes and K. Kehn-Hall (2021). "The Pro-Inflammatory Chemokines CXCL9, CXCL10 and CXCL11 Are Upregulated Following SARS-CoV-2 Infection in an AKT-Dependent Manner." Viruses **13**(6).

Campesi, I., A. Brunetti, G. Capobianco, A. Galistu, A. Montella, F. Ieri and F. Franconi (2022). "Sex Differences in X-ray-Induced Endothelial Damage: Effect of Taurine and N-Acetylcysteine." Antioxidants (Basel) **12**(1).

Chang, W., J. Cheng, J. Allaire, C. Sievert, B. Schloerke, Y. Xie, J. Allen, J. McPherson, A. Dipert and B. Borges (2022). Shiny: Web Application Framework for R.

Chapouly, C., P. L. Hollier, S. Guimbal, L. Cornuault, A. P. Gadeau and M. A. Renault (2020). "Desert Hedgehog-Driven Endothelium Integrity Is Enhanced by Gas1 (Growth Arrest-Specific 1) but Negatively Regulated by Cdon (Cell Adhesion Molecule-Related/Downregulated by Oncogenes)." Arterioscler Thromb Vasc Biol **40**(12): e336-e349.

Cooke, B. M., S. Usami, I. Perry and G. B. Nash (1993). "A simplified method for culture of endothelial cells and analysis of adhesion of blood cells under conditions of flow." Microvasc Res **45**(1): 33-45.

Di Sanzo, M., B. Quaresima, F. Biamonte, C. Palmieri and M. C. Faniello (2020). "FTH1 Pseudogenes in Cancer and Cell Metabolism." Cells **9**(12).

Enguita, F. J., A. L. Leitao, J. T. McDonald, V. Zaksas, S. Das, D. Galeano, D. Taylor, E. S. Wurtele, A. Saravia-Butler, S. B. Baylin, R. Meller, D. M. Porterfield, D. C. Wallace, J. C. Schisler, C. E. Mason and A. Beheshti (2022). "The interplay between lncRNAs, RNA-binding proteins and viral genome during SARS-CoV-2 infection reveals strong connections with regulatory events involved in RNA metabolism and immune response." Theranostics **12**(8): 3946-3962.

Fantini, D. (2019). "easyPubMed: Search and Retrieve Scientific Publication Records from PubMed." from https://cran.r-project.org/web/packages/easyPubMed/index.html.

Gao, Z., Z. Chen, A. Sun and X. Deng (2019). "Gender differences in cardiovascular disease." Medicine in Novel Technology and Devices **4**: 100025.

Gu, Z., L. Gu, R. Eils, M. Schlesner and B. Brors (2014). "circlize Implements and enhances circular visualization in R." Bioinformatics **30**(19): 2811-2812.

Hamada, K., M. Osaka and M. Yoshida (2014). "Cell density impacts epigenetic regulation of cytokine-induced E-selectin gene expression in vascular endothelium." PLoS One **9**(4): e90502.

Hao, S. and D. Baltimore (2013). "RNA splicing regulates the temporal order of TNF-induced gene expression." Proc Natl Acad Sci U S A **110**(29): 11934-11939.

Helle, E., M. Ampuja, A. Dainis, L. Antola, E. Temmes, E. Tolvanen, E. Mervaala and R. Kivelä (2021). "HiPS-Endothelial Cells Acquire Cardiac Endothelial Phenotype in Co-culture With hiPS-Cardiomyocytes." Frontiers in Cell and Developmental Biology **9**.

Herter, E. K., D. Li, M. A. Toma, M. Vij, X. Li, D. Visscher, A. Wang, T. Chu, P. Sommar, L. Blomqvist, D. Berglund, M. Stahle, J. D. Wikstrom and N. Xu Landen (2019). "WAKMAR2, a Long Noncoding RNA Downregulated in Human Chronic Wounds, Modulates Keratinocyte Motility and Production of Inflammatory Chemokines." J Invest Dermatol **139**(6): 1373-1384.

Heydarkhan-Hagvall, S., G. Helenius, B. R. Johansson, J. Y. Li, E. Mattsson and B. Risberg (2003). "Co-culture of endothelial cells and smooth muscle cells affects gene expression of angiogenic factors." J Cell Biochem **89**(6): 1250-1259.

Honda, K., A. Takaoka and T. Taniguchi (2006). "Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors." Immunity **25**(3): 349-360.

Hopfner, K. P. and V. Hornung (2020). "Molecular mechanisms and cellular functions of cGAS-STING signalling." Nat Rev Mol Cell Biol **21**(9): 501-521.

Horioka, K., H. Tanaka, S. Isozaki, H. Konishi, L. Addo, S. Takauji and H. Druid (2020). "Low temperature induces von-willebrand factor expression via increased early growth response 1 transcriptional activity in splenic sinusoidal endothelial cells." Biochem Biophys Res Commun **526**(1): 239-245.

James, B. D. and J. B. Allen (2021). "Sex-Specific Response to Combinations of Shear Stress and Substrate Stiffness by Endothelial Cells In Vitro." Adv Healthc Mater **10**(18): e2100735.

Jang, D. I., A. H. Lee, H. Y. Shin, H. R. Song, J. H. Park, T. B. Kang, S. R. Lee and S. H. Yang (2021). "The Role of Tumor Necrosis Factor Alpha (TNF-alpha) in Autoimmune Disease and Current TNF-alpha Inhibitors in Therapeutics." Int J Mol Sci **22**(5).

Jung, H. S., M. Shimizu-Albergine, X. Shen, F. Kramer, D. Shao, A. Vivekanandan-Giri, S. Pennathur, R. Tian, J. E. Kanter and K. E. Bornfeldt (2020). "TNF-alpha induces acyl-CoA synthetase 3 to promote lipid droplet formation in human endothelial cells." J Lipid Res **61**(1): 33-44.

Kalucka, J., L. de Rooij, J. Goveia, K. Rohlenova, S. J. Dumas, E. Meta, N. V. Conchinha, F. Taverna, L. A. Teuwen, K. Veys, M. García-Caballero, S. Khan, V. Geldhof, L. Sokol, R. Chen, L. Treps, M. Borri, P. de Zeeuw, C. Dubois, T. K. Karakach, K. D. Falkenberg, M. Parys, X. Yin, S. Vinckier, Y. Du, R. A. Fenton, L. Schoonjans, M. Dewerchin, G. Eelen, B. Thienpont, L. Lin, L. Bolund, X. Li, Y. Luo and P. Carmeliet (2020). "Single-Cell Transcriptome Atlas of Murine Endothelial Cells." Cell **180**(4): 764-779.e720.

Karin, N. (2020). "CXCR3 Ligands in Cancer and Autoimmunity, Chemoattraction of Effector T Cells, and Beyond." Front Immunol **11**: 976.

Khan, S. Y., E. M. Awad, A. Oszwald, M. Mayr, X. Yin, B. Waltenberger, H. Stuppner, M. Lipovac, P. Uhrin and J. M. Breuss (2017). "Premature senescence of endothelial cells upon chronic exposure to TNFalpha can be prevented by N-acetyl cysteine and plumericin." Sci Rep **7**: 39501.

Kim, J. J., S. W. Yun, J. J. Yu, K. L. Yoon, K. Y. Lee, H. R. Kil, G. B. Kim, M. K. Han, M. S. Song, H. D. Lee, K. S. Ha, S. Sohn, T. A. Johnson, A. Takahashi, M. Kubo, T. Tsunoda, K. Ito, Y. Onouchi, Y. M. Hong, G. Y. Jang, J. K. Lee and C. Korean Kawasaki Disease Genetics (2017). "A genome-wide association analysis identifies NMNAT2 and HCP5 as susceptibility loci for Kawasaki disease." J Hum Genet **62**(12): 1023-1029.

Langfelder, P. and S. Horvath (2008). "WGCNA: an R package for weighted correlation network analysis." BMC Bioinformatics **9**: 559.

Ley, K. and J. Reutershan (2006). "Leucocyte-endothelial interactions in health and disease." Handb Exp Pharmacol(176 Pt 2): 97-133.

Li, J., J. Chang, J. Wang, D. Xu, M. Yang, Y. Jiang, J. Zhang, X. Jiang and Y. Sun (2022). "HOXA10 promote pancreatic cancer progression via directly activating canonical NF-kappaB signaling pathway." Carcinogenesis **43**(8): 787-796.

Li, R., C. Luo, M. Mines, J. Zhang and G. H. Fan (2006). "Chemokine CXCL12 induces binding of ferritin heavy chain to the chemokine receptor CXCR4, alters CXCR4 signaling, and induces phosphorylation and nuclear translocation of ferritin heavy chain." J Biol Chem **281**(49): 37616-37627.

Liao, H., H. He, Y. Chen, F. Zeng, J. Huang, L. Wu and Y. Chen (2014). "Effects of long-term serial cell passaging on cell spreading, migration, and cell-surface ultrastructures of cultured vascular endothelial cells." Cytotechnology **66**(2): 229-238.

Liao, J. K. (2013). "Linking endothelial dysfunction with endothelial cell activation." J Clin Invest **123**(2): 540-541.

Liu, Y., A. Beyer and R. Aebersold (2016). "On the Dependency of Cellular Protein Levels on mRNA Abundance." Cell **165**(3): 535-550.

Loda, A. and E. Heard (2019). "Xist RNA in action: Past, present, and future." PLoS Genet **15**(9): e1008333.

Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology **15**(12): 550.

Lu, H. S., A. M. Schmidt, R. A. Hegele, N. Mackman, D. J. Rader, C. Weber and A. Daugherty (2018). "Reporting Sex and Sex Differences in Preclinical Studies." Arterioscler Thromb Vasc Biol **38**(10): e171-e184.

Lu, Y., Y. Sun, K. Xu, Y. Shao, F. Saaoud, N. W. Snyder, L. Yang, J. Yu, S. Wu, W. Hu, J. Sun, H. Wang and X. Yang (2022). "Editorial: Endothelial cells as innate immune cells." Front Immunol **13**: 1035497.

Ma, D. J., S. J. Li, L. S. Wang, J. Dai, S. L. Zhao and R. Zeng (2009). "Temporal and spatial profiling of nuclei-associated proteins upon TNF-alpha/NF-kappaB signaling." Cell Res **19**(5): 651-664.

Mackay, F., H. Loetscher, D. Stueber, G. Gehr and W. Lesslauer (1993). "Tumor necrosis factor alpha (TNF-alpha)-induced cell adhesion to human endothelial cells is under dominant control of one TNF receptor type, TNF-R55." J Exp Med **177**(5): 1277-1286.

Majewska, A., K. Wilkus, K. Brodaczewska and C. Kieda (2021). "Endothelial Cells as Tools to Model Tissue Microenvironment in Hypoxia-Dependent Pathologies." Int J Mol Sci **22**(2).

Majewska, E., E. Paleolog, Z. Baj, U. Kralisz, M. Feldmann and H. Tchórzewski (1997). "Role of tyrosine kinase enzymes in TNF-alpha and IL-1 induced expression of ICAM-1 and VCAM-1 on human umbilical vein endothelial cells." Scand J Immunol **45**(4): 385-392.

Maurya, M. R., S. Gupta, J. Y. Li, N. E. Ajami, Z. B. Chen, J. Y. Shyy, S. Chien and S. Subramaniam (2021). "Longitudinal shear stress response in human endothelial cells to atheroprone and atheroprotective conditions." Proc Natl Acad Sci U S A **118**(4).

Mazewski, C., R. E. Perez, E. N. Fish and L. C. Platanias (2020). "Type I Interferon (IFN)-Regulated Activation of Canonical and Non-Canonical Signaling Pathways." Front Immunol **11**: 606456.

McNab, F., K. Mayer-Barber, A. Sher, A. Wack and A. O'Garra (2015). "Type I interferons in infectious disease." Nat Rev Immunol **15**(2): 87-103.

Metcalf, D. J., T. D. Nightingale, H. L. Zenner, W. W. Lui-Roberts and D. F. Cutler (2008). "Formation and function of Weibel-Palade bodies." Journal of Cell Science **121**(1): 19-27.

Nakajima, H. and N. Mochizuki (2017). "Flow pattern-dependent endothelial cell responses through transcriptional regulation." Cell Cycle **16**(20): 1893-1901.

Nordstrom, S. M. and E. J. Weiss (2008). "Sex differences in thrombosis." Expert Review of Hematology **1**(1): 3-8.

Ohori, M., Y. Nakayama, M. Ogasawara-Shimizu, H. Toyoshiba, A. Nakanishi, S. Aparicio and S. Araki (2021). "Gene regulatory network analysis defines transcriptome landscape with alternative splicing of human umbilical vein endothelial cells during replicative senescence." BMC Genomics **22**(1): 869.

Opitz, B., J. Eitel, K. Meixenberger and N. Suttorp (2009). "Role of Toll-like receptors, NOD-like receptors and RIG-I-like receptors in endothelial cells and systemic infections." Thromb Haemost **102**(6): 1103-1109.

Pabbidi, M. R., M. Kuppusamy, S. P. Didion, P. Sanapureddy, J. T. Reed and S. P. Sontakke (2018). "Sex differences in the vascular function and related mechanisms: role of 17beta-estradiol." Am J Physiol Heart Circ Physiol **315**(6): H1499-H1518.

Patella, F. and D. F. Cutler (2020). "RGS4 controls secretion of von Willebrand factor to the subendothelial matrix." J Cell Sci **133**(14).

Paulsen, M. T., A. Veloso, J. Prasad, K. Bedi, E. A. Ljungman, Y. C. Tsan, C. W. Chang, B. Tarrier, J. G. Washburn, R. Lyons, D. R. Robinson, C. Kumar-Sinha, T. E. Wilson and M. Ljungman (2013). "Coordinated regulation of synthesis and stability of RNA during the acute TNF-induced proinflammatory response." Proc Natl Acad Sci U S A **110**(6): 2240-2245.

Pober, J. S. (2002). "Endothelial activation: intracellular signaling pathways." Arthritis Res **4 Suppl 3**(Suppl 3): S109-116.

Rastogi, S., W. Rizwani, B. Joshi, S. Kunigal and S. P. Chellappan (2012). "TNF-alpha response of vascular endothelial and vascular smooth muscle cells involve differential utilization of ASK1 kinase and p73." Cell Death Differ **19**(2): 274-283.

Rathod, K. S., V. Kapil, S. Velmurugan, R. S. Khambata, U. Siddique, S. Khan, S. Van Eijl, L. C. Gee, J. Bansal, K. Pitrola, C. Shaw, F. D'Acquisto, R. A. Colas, F. Marelli-Berg, J. Dalli and A. Ahluwalia (2017). "Accelerated resolution of inflammation underlies sex differences in inflammatory responses in humans." J Clin Invest **127**(1): 169-182.

Ryan, F. J., Y. Ma, L. M. Ashander, M. Kvopka, B. Appukuttan, D. J. Lynn and J. R. Smith (2022). "Transcriptomic Responses of Human Retinal Vascular Endothelial Cells to Inflammatory Cytokines." Transl Vis Sci Technol **11**(8): 27.

Sabbir, M. G., J. T. Wigle, C. G. Taylor and P. Zahradka (2022). "Growth State-Dependent Expression of Arachidonate Lipoxygenases in the Human Endothelial Cell Line EA.hy926." Cells **11**(16).

45

Shao, Y., J. Saredy, W. Y. Yang, Y. Sun, Y. Lu, F. Saaoud, C. t. Drummer, C. Johnson, K. Xu, X. Jiang, H. Wang and X. Yang (2020). "Vascular Endothelial Cells and Innate Immunity." Arterioscler Thromb Vasc Biol **40**(6): e138-e152.

Sheikh, S., G. E. Rainger, Z. Gale, M. Rahman and G. B. Nash (2003). "Exposure to fluid shear stress modulates the ability of endothelial cells to recruit neutrophils in response to tumor necrosis factor-alpha: a basis for local variations in vascular sensitivity to inflammation." Blood **102**(8): 2828-2834.

Tabula Sapiens, C., R. C. Jones, J. Karkanias, M. A. Krasnow, A. O. Pisco, S. R. Quake, J. Salzman, N. Yosef, B. Bulthaup, P. Brown, W. Harper, M. Hemenez, R. Ponnusamy, A. Salehi, B. A. Sanagavarapu, E. Spallino, K. A. Aaron, W. Concepcion, J. M. Gardner, B. Kelly, N. Neidlinger, Z. Wang, S. Crasta, S. Kolluru, M. Morri, A. O. Pisco, S. Y. Tan, K. J. Travaglini, C. Xu, M. Alcantara-Hernandez, N. Almanzar, J. Antony, B. Beyersdorf, D. Burhan, K. Calcuttawala, M. M. Carter, C. K. F. Chan, C. A. Chang, S. Chang, A. Colville, S. Crasta, R. N. Culver, I. Cvijovic, G. D'Amato, C. Ezran, F. X. Galdos, A. Gillich, W. R. Goodyer, Y. Hang, A. Hayashi, S. Houshdaran, X. Huang, J. C. Irwin, S. Jang, J. V. Juanico, A. M. Kershner, S. Kim, B. Kiss, S. Kolluru, W. Kong, M. E. Kumar, A. H. Kuo, R. Leylek, B. Li, G. B. Loeb, W. J. Lu, S. Mantri, M. Markovic, P. L. McAlpine, A. de Morree, M. Morri, K. Mrouj, S. Mukherjee, T. Muser, P. Neuhofer, T. D. Nguyen, K. Perez, R. Phansalkar, A. O. Pisco, N. Puluca, Z. Qi, P. Rao, H. Raquer-McKay, N. Schaum, B. Scott, B. Seddighzadeh, J. Segal, S. Sen, S. Sikandar, S. P. Spencer, L. C. Steffes, V. R. Subramaniam, A. Swarup, M. Swift, K. J. Travaglini, W. Van Treuren, E. Trimm, S. Veizades, S. Vijayakumar, K. C. Vo, S. K. Vorperian, W. Wang, H. N. W. Weinstein, J. Winkler, T. T. H. Wu, J. Xie, A. R. Yung, Y. Zhang, A. M. Detweiler, H. Mekonen, N. F. Neff, R. V. Sit, M. Tan, J. Yan, G. R. Bean, V. Charu, E. Forgo, B. A. Martin, M. G. Ozawa, O. Silva, S. Y. Tan, A. Toland, V. N. P. Vemuri, S. Afik, K. Awayan, O. B. Botvinnik, A. Byrne, M. Chen, R. Dehghannasiri, A. M. Detweiler, A. Gayoso, A. A. Granados, Q. Li, G. Mahmoudabadi, A. McGeever, A. de Morree, J. E. Olivieri, M. Park, A. O. Pisco, N. Ravikumar, J. Salzman, G. Stanley, M. Swift, M. Tan, W. Tan, A. J. Tarashansky, R. Vanheusden, S. K. Vorperian, P. Wang, S. Wang, G. Xing, C. Xu, N. Yosef, M. Alcantara-Hernandez, J. Antony, C. K. F. Chan, C. A. Chang, A. Colville, S. Crasta, R. Culver, L. Dethlefsen, C. Ezran, A. Gillich, Y. Hang, P. Y. Ho, J. C. Irwin, S. Jang, A. M. Kershner, W. Kong, M. E. Kumar, A. H. Kuo, R. Leylek, S. Liu, G. B. Loeb, W. J. Lu, J. S. Maltzman, R. J. Metzger, A. de Morree, P. Neuhofer, K. Perez, R. Phansalkar, Z. Qi, P. Rao, H. Raquer-McKay, K. Sasagawa, B. Scott, R. Sinha, H. Song, S. P. Spencer, A. Swarup, M. Swift, K. J. Travaglini, E. Trimm, S. Veizades, S. Vijayakumar, B. Wang, W. Wang, J. Winkler, J. Xie, A. R. Yung, S. E. Artandi, P. A. Beachy, M. F. Clarke, L. C. Giudice, F. W. Huang, K. C. Huang, J. Idoyaga, S. K. Kim, M. Krasnow, C. S. Kuo, P. Nguyen, S. R. Quake, T. A. Rando, K. Red-Horse, J. Reiter, D. A. Relman, J. L. Sonnenburg, B. Wang, A. Wu, S. M. Wu and T. Wyss-Coray (2022). "The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans." Science **376**(6594): eabl4896.

Team, R. C. (2022). R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computin.

Tesfay, L., A. J. Huhn, H. Hatcher, F. M. Torti and S. V. Torti (2012). "Ferritin blocks inhibitory effects of two-chain high molecular weight kininogen (HKa) on adhesion and survival signaling in endothelial cells." PLoS One **7**(7): e40030.

Thorner, L. W., C. Erikstrup, L. H. Harritshoj, M. H. Larsen, G. Kronborg, C. Pedersen, C. S. Larsen, G. Pedersen, J. Gerstoft, N. Obel and H. Ullum (2016). "Impact of polymorphisms in the HCP5 and HLA-C, and ZNRD1 genes on HIV viral load." Infect Genet Evol **41**: 185-190.

Trivedi, C. M., R. C. Patel and C. V. Patel (2007). "Homeobox gene HOXA9 inhibits nuclear factor-kappa B dependent activation of endothelium." Atherosclerosis **195**(2): e50-60.

Uematsu, S. and S. Akira (2007). "Toll-like receptors and Type I interferons." J Biol Chem **282**(21): 15319-15323.

Ulfhammer, E., P. Larsson, L. Karlsson, T. Hrafnkelsdottir, M. Bokarewa, A. Tarkowski and S. Jern (2006). "TNF-alpha mediated suppression of tissue type plasminogen activator expression in vascular endothelial cells is NF-kappaB- and p38 MAPK-dependent." J Thromb Haemost **4**(8): 1781-1789.

Valenzuela, N. M. (2022). "Late phase endothelial cell inflammation is characterized by interferon response genes and driven by JAK/STAT, not NFkappaB." Vascul Pharmacol **146**: 107090.

van Loo, G. and M. J. M. Bertrand (2023). "Death by TNF: a road to inflammation." Nat Rev Immunol **23**(5): 289-303.

Vandenabeele, P., W. Declercq, R. Beyaert and W. Fiers (1995). "Two tumour necrosis factor receptors: structure and function." Trends in Cell Biology **5**(10): 392-399.

Venkatesh, D., T. Ernandez, F. Rosetti, I. Batal, X. Cullere, F. W. Luscinskas, Y. Zhang, G. Stavrakis, G. Garcia-Cardena, B. H. Horwitz and T. N. Mayadas (2013). "Endothelial TNF receptor 2 induces IRF1 transcription factor-dependent interferon-beta autocrine signaling to promote monocyte recruitment." Immunity **38**(5): 1025-1037.

Viemann, D., M. Goebeler, S. Schmid, U. Nordhues, K. Klimmek, C. Sorg and J. Roth (2006). "TNF induces distinct gene expression programs in microvascular and macrovascular human endothelial cells." J Leukoc Biol **80**(1): 174-185.

Webster, J. D. and D. Vucic (2020). "The Balance of TNF Mediated Pathways Regulates Inflammatory Cell Death Signaling in Healthy and Diseased Tissues." Front Cell Dev Biol **8**: 365.

Wickham, H. (2016). Data Analysis. ggplot2: Elegant Graphics for Data Analysis. H. Wickham. Cham, Springer International Publishing: 189-201.

Willemsen, J., M.-T. Neuhoff, T. Hoyler, E. Noir, C. Tessier, S. Sarret, T. N. Thorsen, A. Littlewood-Evans, J. Zhang, M. Hasan, J. S. Rush, D. Guerini and R. M. Siegel (2021). "TNF leads to mtDNA release and cGAS/STING-dependent interferon responses that support inflammatory arthritis." Cell Reports **37**(6): 109977.

Wu, T., E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo and G. Yu (2021). "clusterProfiler 4.0: A universal enrichment tool for interpreting omics data." The Innovation **2**(3): 100141.

Xu, R., S. S. Yu, R. R. Yao, R. C. Tang, J. W. Liang, X. Pang and J. Zhang (2021). "Interferon-Inducible LINC02605 Promotes Antiviral Innate Responses by Strengthening IRF3 Nuclear Translocation." Front Immunol **12**: 755512.

Yan, G., B. You, S. P. Chen, J. K. Liao and J. Sun (2008). "Tumor necrosis factor-alpha downregulates endothelial nitric oxide synthase mRNA stability via translation elongation factor 1-alpha 1." Circ Res **103**(6): 591-597.

Yan, R., M. van Meurs, E. R. Popa, R. M. Jongman, P. J. Zwiers, A. E. Niemarkt, T. Kuiper, J. A. Kamps, P. Heeringa, J. G. Zijlstra, G. Molema and J. Moser (2017). "Endothelial Interferon Regulatory Factor 1 Regulates Lipopolysaccharide-Induced VCAM-1 Expression Independent of NFkappaB." J Innate Immun **9**(6): 546-560.

Yang, T., K. Y. Sim, G. H. Ko, J. S. Ahn, H. J. Kim and S. G. Park (2022). "FAM167A is a key molecule to induce BCR-ABL-independent TKI resistance in CML via noncanonical NF-kappaB signaling activation." J Exp Clin Cancer Res **41**(1): 82.

Yau, J. W., H. Teoh and S. Verma (2015). "Endothelial cell control of thrombosis." BMC Cardiovasc Disord **15**: 130.

Yin, X., Z. Wang, T. Wu, M. Ma, Z. Zhang, Z. Chu, Q. Hu, H. Ding, X. Han, J. Xu, H. Shang and Y. Jiang (2019). "The combination of CXCL9, CXCL10 and CXCL11 levels during primary HIV infection predicts HIV disease progression." J Transl Med **17**(1): 417.

Yuan, S., X. Yuan and L. Li (2022). "Long non-coding RNA HOXA11-AS protects the barrier function of corneal endothelial cells by sponging microRNA-155 to alleviate corneal endothelial injury." Am J Transl Res **14**(12): 8489-8503.

Zhang, Y. and K. Lingappan (2017). "Differential sex-specific effects of oxygen toxicity in human umbilical vein endothelial cells." Biochem Biophys Res Commun **486**(2): 431-437.

Zhou, C., Q. Yan, Q. Y. Zou, X. Q. Zhong, C. T. Tyler, R. R. Magness, I. M. Bird and J. Zheng (2019). "Sexual Dimorphisms of Preeclampsia-Dysregulated Transcriptomic Profiles and Cell Function in Fetal Endothelial Cells." Hypertension **74**(1): 154-163.
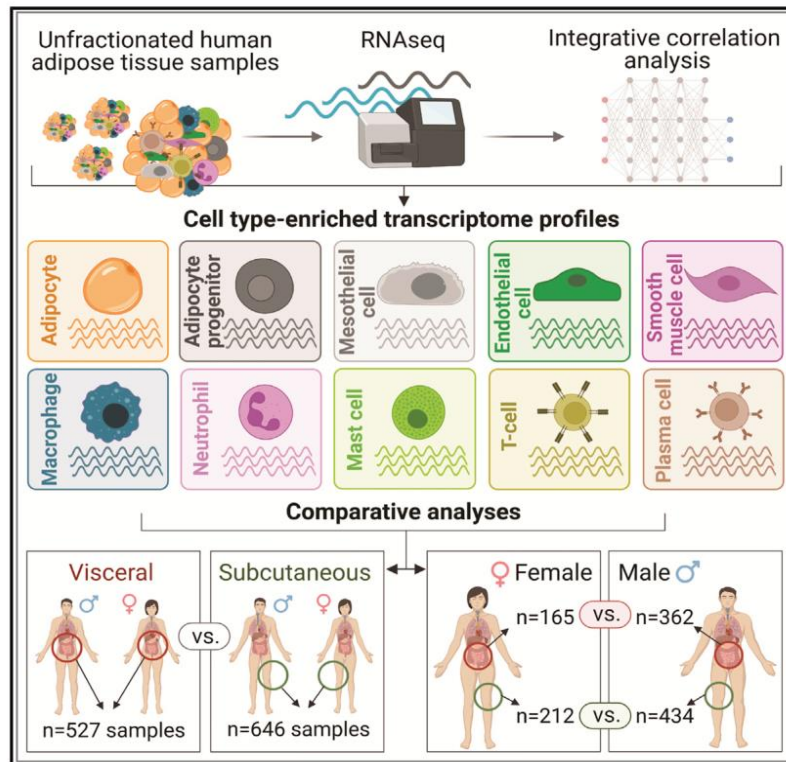
PAPER 2

A human adipose tissue cell-type transcriptome atlas

A human adipose tissue cell-type transcriptome atlas

Graphical abstract

Authors

Marthe Norreen-Thorsen,
Eike Christopher Struck, Sofia Öling, ...,
Mathias Uhlén, Philip James Dusart,
Lynn Marie Butler

Correspondence

lynn.butler@ki.se, lynn.m.butler@uit.no

In brief

Norreen-Thorsen et al. use an integrative correlation analysis of human adipose tissue RNA-seq data, identifying >2,000 cell-type-enriched coding and non-coding transcripts. Comparative analyses highlight transcripts with visceral and subcutaneous depot-specific and/or sex-specific cell-type enrichment. The method allows profiling of adipocytes, whose physical characteristics make analysis with other methods challenging.

Highlights

- Uses publicly available adipose tissue bulk RNA-seq data from two human fat depots

- Enriched genes in 10 cell types profiled using an integrative correlation analysis

- Comparative analysis identifies depot- and sex-specific cell-type-enriched genes

- Method circumvents technical challenges with adipose tissue scRNA-seq analysis

CellPress

151

Resource

A human adipose tissue cell-type transcriptome atlas

Marthe Norreen-Thorsen,[1] Eike Christopher Struck,[1] Sofia Öling,[1] Martin Zwahlen,[2] Kalle Von Feilitzen,[2] Jacob Odeberg,[1,2,3,4] Cecilia Lindskog,[5] Fredrik Pontén,[5] Mathias Uhlén,[2] Philip James Dusart,[1,2,6,7] and Lynn Marie Butler[1,2,6,7,8,*]

[1]Translational Vascular Research, Department of Clinical Medicine, The Arctic University of Norway, 9019 Tromsø, Norway
[2]Science for Life Laboratory, Department of Protein Science, Royal Institute of Technology (KTH), 171 21 Stockholm, Sweden
[3]The University Hospital of North Norway (UNN), 9019 Tromsø, Norway
[4]Coagulation Unit, Department of Hematology, Karolinska University Hospital, 171 76 Stockholm, Sweden
[5]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, 752 37 Uppsala, Sweden
[6]Clinical Chemistry and Blood Coagulation Research, Department of Molecular Medicine and Surgery, Karolinska Institute, 171 76 Stockholm, Sweden
[7]Clinical Chemistry, Karolinska University Laboratory, Karolinska University Hospital, 171 76 Stockholm, Sweden
[8]Lead contact
*Correspondence: lynn.butler@ki.se or lynn.m.butler@uit.no
https://doi.org/10.1016/j.celrep.2022.111046

SUMMARY

The importance of defining cell-type-specific genes is well acknowledged. Technological advances facilitate high-resolution sequencing of single cells, but practical challenges remain. Adipose tissue is composed primarily of adipocytes, large buoyant cells requiring extensive, artefact-generating processing for separation and analysis. Thus, adipocyte data are frequently absent from single-cell RNA sequencing (scRNA-seq) datasets, despite being the primary functional cell type. Here, we decipher cell-type-enriched transcriptomes from unfractionated human adipose tissue RNA-seq data. We profile all major constituent cell types, using 527 visceral adipose tissue (VAT) or 646 subcutaneous adipose tissue (SAT) samples, identifying over 2,300 cell-type-enriched transcripts. Sex-subset analysis uncovers a panel of male-only cell-type-enriched genes. By resolving expression profiles of genes differentially expressed between SAT and VAT, we identify mesothelial cells as the primary driver of this variation. This study provides an accessible method to profile cell-type-enriched transcriptomes using bulk RNA-seq, generating a roadmap for adipose tissue biology.
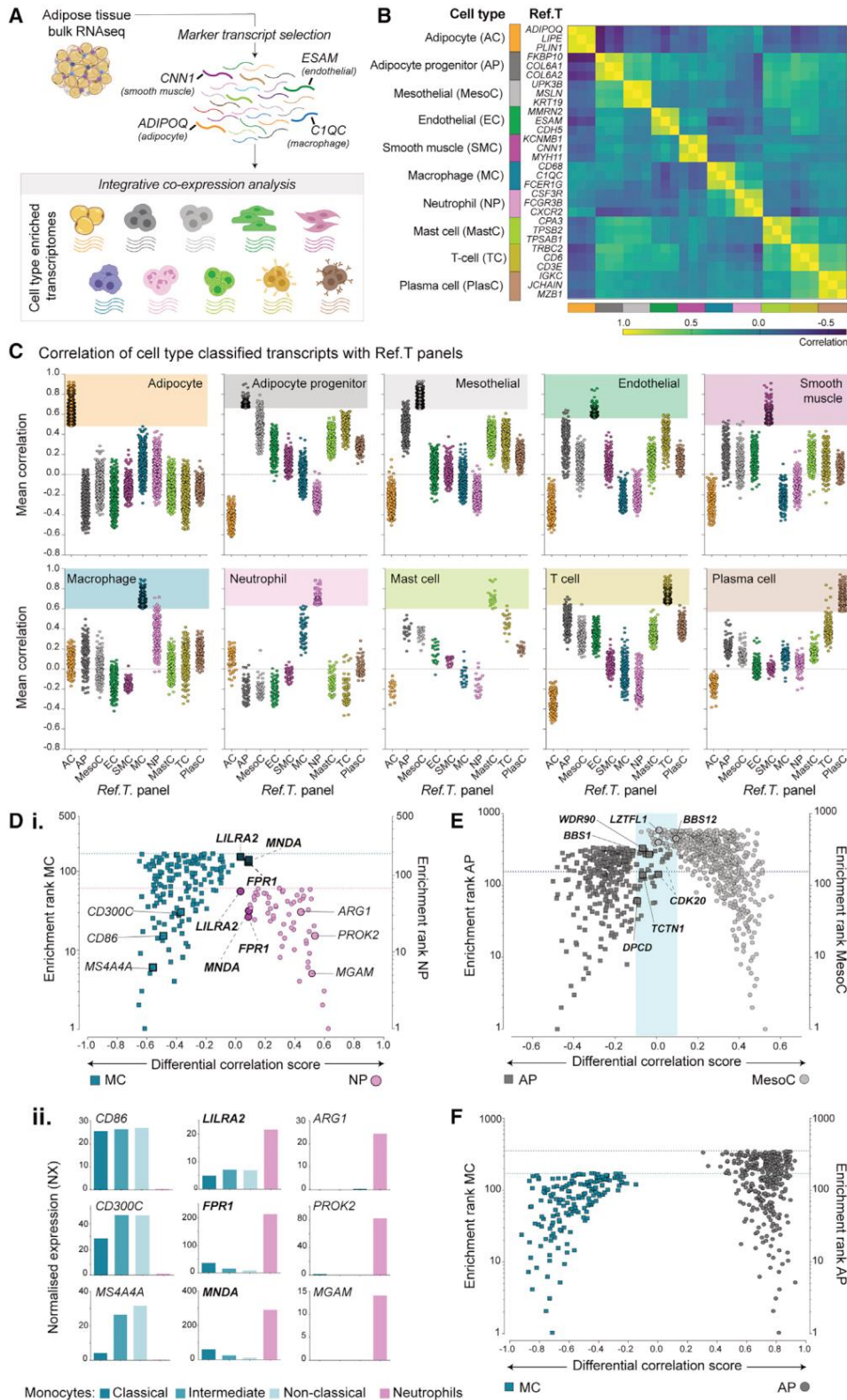
INTRODUCTION

Adipose tissue acts as an energy depot, provides insulation, and is an important endocrine organ that communicates with other tissues to regulate systemic metabolism (Kahn et al., 2019). Most adipose tissue in adults is white adipose tissue, broadly categorized as visceral adipose tissue (VAT), located deep in the abdomen and around internal organs, or subcutaneous adipose tissue (SAT), located under the skin. Excess VAT is associated with metabolic disorders, e.g., diabetes and cardiovascular disease (Britton et al., 2013; Chait and den Hartigh, 2020; Oikonomou and Antoniades, 2019), while SAT is associated with reduced risk, possibly even protection (Lumish et al., 2020). Recent studies have profiled differences in gene expression between adipose depots using bulk RNA sequencing (RNA-seq) (Bradford et al., 2019; Schleinitz et al., 2020), but the relative contribution of specific cell types to the observed differences is not known.

Adipose tissue contains adipocytes, adipocyte progenitor cells, endothelial cells, smooth muscle cells, stromal cells, and immune cell types, including macrophages and T cells (Lu

et al., 2019). Single-cell RNA-seq (scRNA-seq) has been used to profile macrophages, endothelial cells, fibroblasts, and adipocyte progenitors from human VAT or SAT (Acosta et al., 2017; Esteve et al., 2019; Vijay et al., 2020). Such studies provide high resolution of different cell (sub)types but are limited by the requirement for fresh tissue, low number of biological replicates, and compromised read depth (Beliakova-Bethell et al., 2014; Rizzetto et al., 2017; Saliba et al., 2014; Ziegenhain et al., 2017). Furthermore, the analysis of adipocytes, the major functional cell type in adipose tissue, is challenging; with high buoyancy and large size, they require extensive, artefact-generating proteolytic digestion for tissue separation (Rondini and Granneman, 2020; Viswanadha and Londos, 2006), and thus, adipocyte data are absent from many scRNA-seq datasets (e.g., Hildreth et al., 2021; Karlsson et al., 2021; Tabula Muris et al., 2018; Tabula Sapiens et al., 2022; Vijay et al., 2020). Progenitor cells isolated from human adipose tissue have been used to generate adipocytes in culture for analysis (Min et al., 2019), but potential phenotype modifications, due to induced differentiation in the absence of the native microenvironment, are a limitation of this model. Transgenic labeling of cell-type-specific

A Adipose tissue bulk RNAseq — *Marker transcript selection*

Integrative co-expression analysis

Cell type enriched transcriptomes

B Cell type / Ref.T

C Correlation of cell type classified transcripts with Ref.T panels

D i.

ii.

Monocytes: Classical / Intermediate / Non-classical / Neutrophils

E

F

(legend on next page)

153

mRNA (Roh et al., 2017) has been used to overcome these technical problems for analysis of murine adipocytes, but this cannot be applied to human tissue. Adipocytes have been analyzed using single-nuclei RNA-seq, circumventing some scRNA-seq limitations, but transcript expression can differ between nuclei and cytoplasm (Thrupp et al., 2020). Non-coding RNA is emerging as a novel, important class of molecules in adipose biology (Squillaro et al., 2020; Xu and Sun, 2020), but to date, there is no description of adipose-cell-type-specific non-coding RNAs.

Here, using an analysis approach to circumvent technical challenges associated with profiling individual cell types in adipose tissue, we identified over 2,300 transcripts with cell-type-enriched expression. Of all cell types profiled, adipocytes had the highest number of enriched transcripts and the greatest proportion of non-coding. Comparisons between female and male samples revealed a panel of cell-type-enriched Y-linked transcripts, of which three were adipocyte enriched in both depots. Finally, we resolve the overall differences in gene expression between VAT and SAT to a cell-type level, uncovering the primary driver to be cell-type composition, specifically the presence of mesothelial cells in VAT, but not SAT. Data are available through the Human Protein Atlas portal (www.proteinatlas.org/humanproteome/tissue+cell+type/adipose+tissue).

RESULTS

### Identification of cell-type transcriptome profiles in visceral adipose tissue
Cell-type reference transcripts correlate across unfractionated adipose RNA-seq data

VAT is linked to the development of metabolic dysfunction and associated disorders (Chait and den Hartigh, 2020). To identify cell-type-enriched transcriptome profiles, we performed an analysis based on our previously reported method (Butler et al., 2016; Dusart et al., 2019) (concept summary, Figures 1A and S1A), using VAT RNA-seq data (n = 527) from Genotype-Tissue Expression (GTEx) portal v.8 (www.gtexportal.org) (Consortium, 2015). For each cell type, a panel of three marker genes were selected ("reference transcripts" [Ref.T.s]). Correlation coefficients (corr.) between the expression levels of the Ref.T.s and all other sequenced transcripts were calculated across samples; those that highly and selectively correlated with the Ref.T. panel

were classified as enriched in the corresponding cell type (Figure S1A). We shortlisted candidate Ref.T.s for all main constituent cell types, including (1) well-established markers identified in older "none-omics" studies (e.g., Hu et al., 1996); (2) markers identified by scRNA-seq of mouse (Tabula Muris et al., 2018) or human (Han et al., 2020) adipose tissue; (3) marker lists from large databases containing data from multiple sources, e.g., Cell Marker (Zhang et al., 2019) and PanglaoDB (Franzen et al., 2019); and (4) commercial marker panels (e.g., https://www.rndsystems.com/). VAT RNA-seq data were processed to generate Spearman correlation coefficients (corr.s) between all candidate Ref.T.s and a panel selected to represent each cell type, based on the following criteria: (1) a high corr. (minimum 0.70) between Ref.T.s within each cell type panel (Figure 1B; Table S1, tab 1, table A), consistent with cell type co-expression; adipocyte panel (ADIPOQ, LIPE, and PLIN1; mean corr. ± SD 0.91 ± 0.002), adipocyte progenitor panel (FKBP10, COL6A1, and COL6A2; 0.86 ± 0.06), mesothelial panel (UPK3B, MSLN, and KRT19; 0.92 ± 0.02), endothelial panel (MMRN2, ESAM, and CDH5; 0.80 ± 0.03), smooth muscle panel (KCNMB1, CNN1, and MYH11; 0.80 ± 0.06), macrophage panel (CD68, C1QC, and FCER1G; 0.83 ± 0.03), neutrophil panel (CSF3R, FCGR3B, and CXCR2; 0.81 ± 0.04), mast cell panel (CPA3, TPSB2, and TPSAB1; 0.83 ± 0.03), T cell panel (TRBC2, CD6, and CD3E; 0.89 ± 0.02), and plasma cell panel (IGKC, JCHAIN, and MZB1; 0.89 ± 0.04; all p < 4.0 x 10^{-99}); (2) a low corr. between Ref.T.s in different cell type panels (Figure 1B; Table S1, tab 1, table A), consistent with high specificity of each panel (mean inter-panel corr. ± SD 0.05 ± 0.25); and (3) a normal distribution of Ref.T. expression across samples (Figure S1B). Candidate B cell Ref.T.s were lowly expressed in VAT, with low intra-panel corr. (Table S1, tab 2, table A). In a comparative dataset, human spleen RNA-seq, GTEx v.8 (n = 241), selected due to high B cell content, the candidate B cell Ref.T.s were highly expressed and strongly correlated with each other (Table S1, tab 2, table B). Thus, B cells were excluded from subsequent profiling of VAT, due to presumed low numbers or absence from a large proportion of VAT samples. Candidates within the panels selected as potential Ref.T.s for pericytes, lymphatic endothelial cells, and dendritic cells did not correlate as well as those selected to represent other cell types (Table S1, tab 2, tables C, D, and E, respectively), consistent with previous reports that these cell types lack multiple highly

Figure 1. Integrative co-expression analysis of unfractionated human visceral adipose tissue (VAT) RNA-seq can resolve constituent cell-type identities

(A) Overview of analysis concept; human VAT RNA-seq data (n = 527 individuals) were retrieved from GTEx portal v.8 and constituent cells "virtually tagged" using cell-type-specific reference transcripts (Ref.T.s). Integrative co-expression analysis was used to identify transcripts with comparable profiles.

(B) Heatmap of Spearman correlation (corr.) values between Ref.T. panels selected for VAT cell types: adipocyte (AC), adipocyte progenitor (AP), mesothelial (MesoC), endothelial (EC), smooth muscle (SMC), macrophage (MC), neutrophil (NP), mast cell (MastC), T cell (TC), and plasma cell (PlasC).

(C) Mean corr. values between genes above designated threshold (see results section for criteria) and all Ref.T. panels.

(D) (i) For transcripts above the designated corr. threshold with macrophage (squares, MC) or neutrophil (circles, NP) Ref.T. panels, the "differential corr. score" (difference between mean corr. with MC and NP Ref.T.s) was plotted versus "enrichment ranking" (position in each respective list; highest corr. = rank 1). Corresponding colored lines indicate numbers above the designated threshold. Bold text annotations show transcripts in both MC and NP lists (circular and square symbol on the same x axis dimension). (ii) scRNA-seq data from the Human Protein Blood Atlas (Uhlen et al., 2019) showing gene expression in classical, intermediate, and non-classical monocytes and neutrophils from whole blood.

(E and F) Comparative plots for transcripts are classified as (E) adipocyte progenitor (AP) or mesothelial (MesoC) enriched (shaded blue box indicates co-enriched genes) and (F) MC or AP enriched.

See also Figures S1 and S2 and Tables S1, tabs 1, 4a, and 4b, and S2, tab 1.

(legend on next page)

155
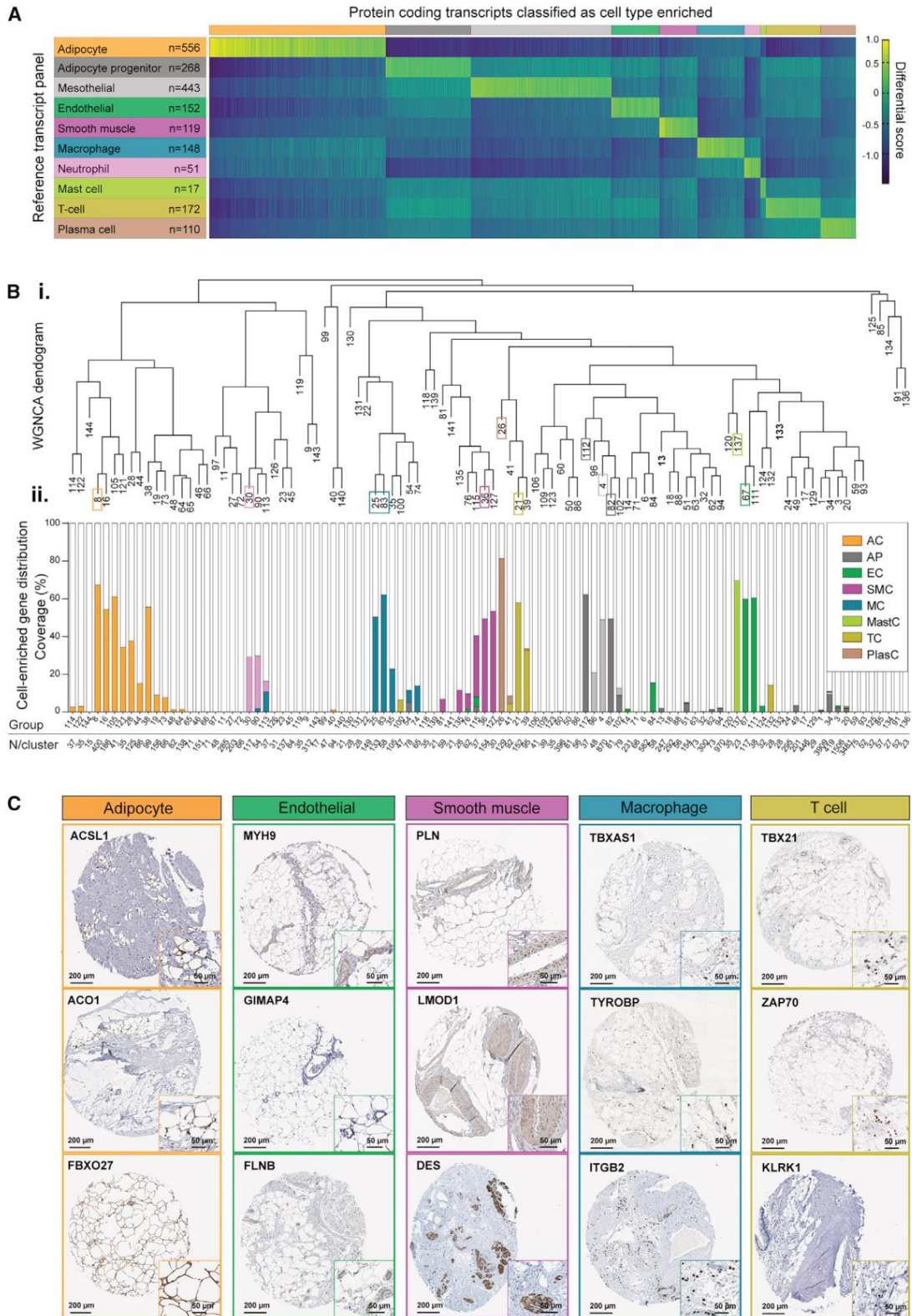
specific pan-markers (Armulik et al., 2011; Sichien et al., 2017; Takeda et al., 2019). Thus, they were not included in the subsequent analysis.

Reference transcripts analysis can identify distinct cell-type-associated transcripts

We generated corr.s between each *Ref.T.* and all sequenced transcripts (~53,625) across VAT samples. The proportion of constituent cell types between samples vary, due to both sampling and heritability factors (Glastonbury et al., 2019), but ratios between cell-specific co-expressed genes should remain constant. Thus, a high corr. of a given transcript with all *Ref.T.*s in any one panel is consistent with expression in the corresponding cell type. For each *Ref.T.* panel, a list of such transcripts was generated using a corr. value threshold cutoff, which was either (1) that above which >95% of transcripts reached this threshold with *only* that *Ref.T.* panel or (2) ≥0.50, whichever was higher (for thresholds, see Table S1, Tab 1, Table B). Resultant transcripts were generally well separated (Figure 1C), but some overlap was observed between closely related cell types, e.g., macrophages [MCs] and neutrophils [NPs]; Figure 1C, row 2). To compare specific transcripts in two cell-type classified groups, e.g., MC and NP enriched (Figure 1D.i), the following was calculated for each transcript: (1) the "differential correlation score," defined as the difference between the mean corr. with the two sets of *Ref.T.*s, i.e., MC panel (*CD68*, *C1QC*, and *FCER1G*) and NP panel (*CSF3R*, *FCGR3B*, and *CXCR2*), and (2) the "enrichment ranking," based on the mean corr. value with the *Ref.T.* panel (rank 1 = highest corr.). Three transcripts were provisionally classified as both MC and NP enriched: *LILRA2*, *MNDA*, and *FPR1* (Figure 1D.i; gene IDs in bold text). *LILRA2* had comparable corr. with both MC and NP *Ref.T.* panels (mean corr. ± SD: 0.61 ± 0.07 and 0.65 ± 0.07, respectively), while *MNDA* and FPR1, despite reaching the threshold for both, were more highly correlated with the NP *Ref.T.* panel than the MC panel (*MNDA*: NP 0.71 ± 0.04 versus MC 0.62 ± 0.08 and *FPR1*: NP 0.72 ± 0.07 versus MC 0.63 ± 0.14). We extracted expression data for these genes in monocytes (MonoC) and neutrophils in blood (Figure 1D.ii) from scRNA-seq generated as part of our Human Protein Atlas (HPA) blood atlas (Uhlen et al., 2019). In all three cases, these transcripts were expressed in both MonoC and NP (Figure 1D.ii, central column). In contrast, transcripts we classified as enriched *only* in MC (*CD300C*, *CD86*, and *MS4A4A*) or *only* in NP (*ARG1*, *PROK2*, and *MGAM*) were predominantly expressed in MonoC or NP in blood, respectively (Figure 1D.ii). Although MonoC and MC are not directly comparable, the majority of the monocyte transcriptional profile is maintained during differentiation (Martinez et al., 2006), and so these

data support our annotated classifications. These annotations were also consistent with data from scRNA-seq of macrophages and neutrophils from human SAT (Tabula Sapiens et al., 2022; Figure S2B). Thus, to exclude potentially dual-enriched transcripts from cell-type classification, we excluded transcripts with a differential corr. value < 0.1 versus any other *Ref.T.* panel. The highest number of transcripts excluded for this reason were those that correlated with both adipocyte progenitor cell and mesothelial cell *Ref.T.* panels; 84 transcripts were excluded from cell-type classification due to likely co-expression (Figure 1E; Table S1, tab 4a). Gene ontology (GO) and reactome analysis of this gene list revealed over-representation of terms related to "plasma membrane bounded cell projection organization" (false discovery rate [FDR] 1.26×10^{-2}) and "BBSome-mediated cargo targeting to cilium" (FDR 8.09×10^{-3}), respectively (Table S1, tab 4b, with selected examples highlighted in Figure 1E), suggesting a possible link to the importance of primary cilia in the regulation of adipose tissue expansion (Hilgendorf, 2021; Ritter et al., 2018). In most other cases, transcripts were well separated between cell types, e.g., MC classified versus adipocyte progenitor (AP) classified (Figure 1F). We classified 2,343 transcripts as cell-type enriched in VAT (Tables S1, tab 1, table B, and S2, tab 1), the majority of which (2,036 [87%]) were protein coding (Figure 2A).

Independent methods and datasets support cell-type classifications

Unsupervised weighted network correlation analysis (WGNCA) is consistent with Ref.T. analysis

As our analysis method is based on manually selected *Ref.T.* panels, cell-type classification is subject to an input bias. As a comparison, we analyzed the same dataset using an unbiased WGCNA (Langfelder and Horvath, 2008). Corr.s between all transcripts were calculated, and they were subsequently clustered into related groups, based on expression similarity (Figure 2B.i). Members of the same *Ref.T.* panels clustered into the same WGCNA group, e.g., adipocyte *Ref.T.*s (*ADIPOQ*, *LIPE*, and *PLIN1*; cluster 8, orange box) and mesothelial *Ref.T.*s (*UPK3B*, *MSLN*, and *KRT19*; cluster 4, light gray box, Figure 2B.i), or into adjacent leaves on the same clade, e.g., macrophage *Ref.T.*s (*CD68*, *C1QC*, and *FCER1G*; group 25 and 83, blue box, Figure 2B.i). The locations of all other *Ref.T.*s are indicated by the respective colored boxes. Thus, WGCNA results were consistent with intra-panel *Ref.T.*s having shared expression profiles (i.e., in a common cell type). Protein-coding transcripts classified as cell-type enriched (Figure 2A) predominantly clustered into the same WGCNA group(s) as the corresponding

---

Figure 2. Integrative co-expression analysis of unfractionated RNA-seq reveals enriched protein-coding transcriptomes of human visceral adipose tissue (VAT) cell types

(A) Heatmap of protein-coding transcripts classified as cell type enriched (indicated by horizontal colored bars), showing differential score between mean corr. with the corresponding *Ref.T.* panel versus highest mean corr. coefficient among the other *Ref.T.* panels.

(B) Human VAT RNA-seq data (n = 527 individuals) were subject to weighted correlation network analysis (WGCNA). (i) Colored squares indicate *Ref.T.* location on dendrogram (colors correspond to cell types as annotated in A). (ii) Distribution of protein-coding transcripts classified as cell type enriched across dendrogram groups.

(C) Human adipose tissue sections were stained using antibodies targeting proteins encoded by transcripts classified as adipocyte, endothelial, smooth muscle, macrophage, or T cell enriched. Scale bar, 200 μm; inset, 50 μm.

See also Table S2, tab 1, and Figure S3.
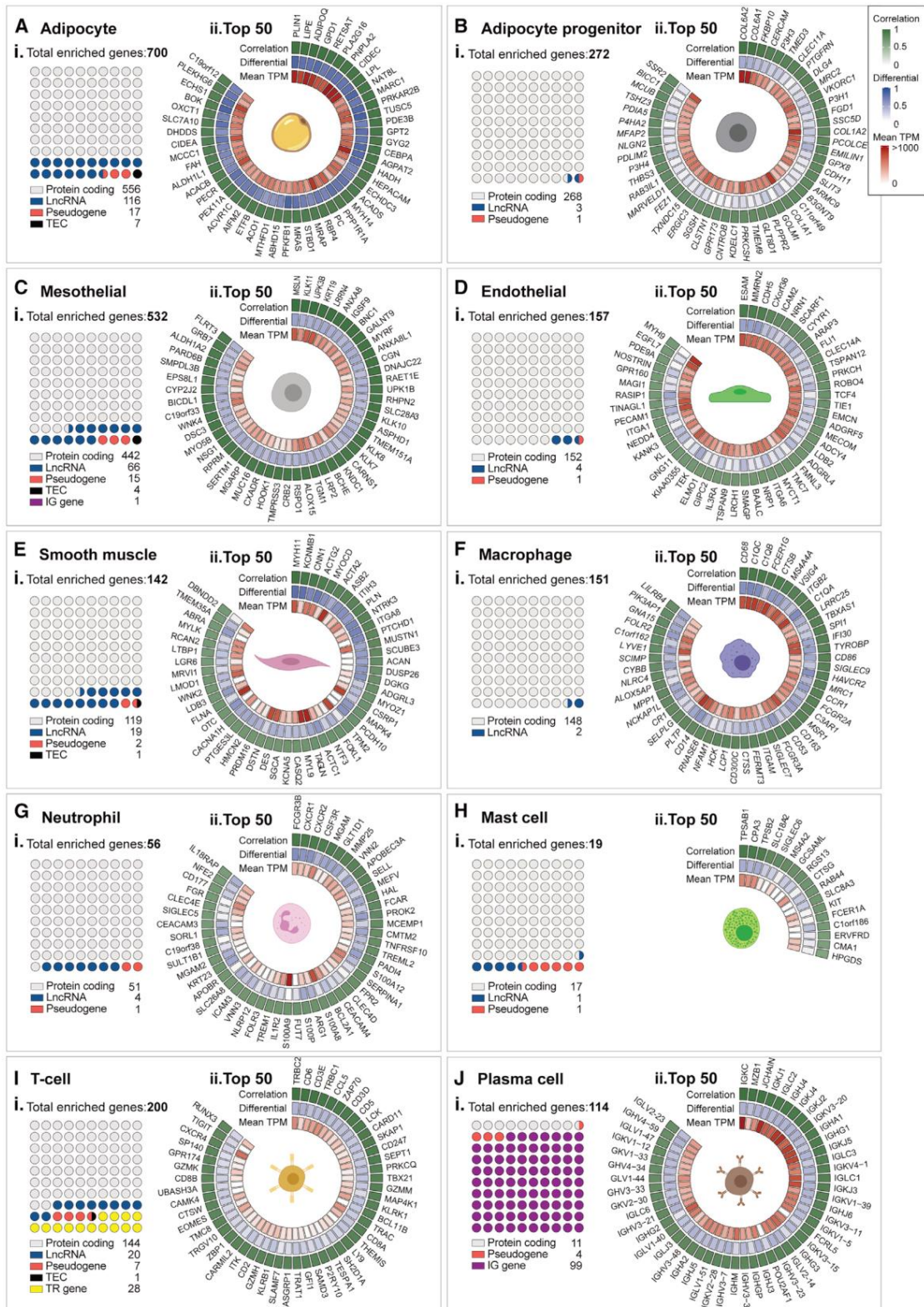
(legend on next page)

*Ref.T.*s, e.g., mast cell enriched (lime green bar, group 137) and macrophage enriched (blue bars, groups 25 and 83;, (Figure 2B.ii) and, in some cases, different leaves on a common clade, e.g., adipocyte enriched (orange bars, groups 8, 16, 105, 121, 28, 44, and 38) and smooth muscle cells (purple bars, groups 36, 127, and 115) Figure 2B.ii). Thus, protein-coding transcripts classified as cell-type enriched also clustered together in an unbiased WGCNA, consistent with cell type co-expression. Protein profiling of human adipose tissue is provided for selected adipocyte-, endothelial-, smooth-muscle-cell-, macrophage- and T-cell-enriched transcripts (Figures 2C and S3A).

Cell-type-enriched transcripts in visceral adipose tissue

Adipocytes had the most enriched transcripts (n = 700), of which the greatest proportion were protein coding (Figure 3A.i, light gray circles), followed by long non-coding RNA (lncRNA) and pseudogenes (Figure 3A.i; Vertebrate Genome Annotation [VEGA] database; Harrow et al., 2014). Mesothelial cells (Figure 3C.i), and adipocyte progenitor cells (Figure 3B.i) also had a relatively high number of enriched transcripts (n = 532 and 272, respectively), while immune cells, and other non-tissue-specific cells, had fewer: endothelial (n = 157) (Figure 3D.i), smooth muscle (n = 142) (Figure 3E.i), macrophages (n = 151) (Figure 3F.i), neutrophils (n = 56) (Figure 3G.i), mast cells (n = 19) (Figure 3H.i), T cells (n = 200) (Figure 3I.i), and plasma cells (n = 114) (Figure 3J.i; Table S2, tab 1). In most cell types, lncRNAs represented the majority of transcripts outside those classified as protein coding (Figures 3A.i–3J.i), with the notable exception of T cells and plasma cells, where the majority were T cell receptor (TR) (Figure 3I.i) or immunoglobulin (IG) genes (Figure 3J.i), respectively. GO and reactome analysis (Ashburner et al., 2000; Gene Ontology, 2021) was performed to identify over-represented classes and pathways in each list of enriched transcripts (Table S2, tabs 2–11). Results were consistent with known cell functions. For example, for adipocyte-enriched transcripts, most significant GO terms included "small molecule metabolic process" (FDR 2.5×10^{-60}) and "carboxylic acid metabolic process" (FDR 1.4×10^{-57}) and reactome pathways included "metabolism" (FDR 2.9×10^{-70}); for smooth muscle-enriched transcripts, GO terms included "muscle system processes" (FDR 5.9×10^{-14}) and reactome pathways included "muscle contraction" (FDR 1.9×10^{-9}); for endothelial-cell-enriched transcripts, GO terms included "blood vessel development" (FDR 8.8×10^{-10}) and "angiogenesis" (FDR 1.1×10^{-9}); and for neutrophil-enriched transcripts, GO terms included "neutrophil activation" (FDR 1.9×10^{-18}) and reactome pathways included "neutrophil degranulation" (FDR 5.6×10^{-19}) (for all cell types, see Table S2, tabs 2–11, tables Ai and Aii). We visualized the top 50 enriched protein-coding transcripts for each cell type (Figures 3A.ii–3J.ii), ranked by highest mean corr. with the *Ref.T.* panel, to compare differential corr.

values (corr. with corresponding cell type *Ref.T.* panel minus max corr. with any other *Ref.T.* panel) and expression. Overall, expression values for enriched genes were highest for adipocytes (Figure 3A.ii), adipocyte progenitor cells (Figure 3B.ii), endothelial cells (Figure 3D.ii), macrophages (Figure 3F.ii), and plasma cells (Figure 3J.ii) and lowest for neutrophils, mast cells, and T cells (Figures 3G.ii–3I.ii). However, cell-type-enriched transcripts had a range of expression values, indicating variation in regulatory mechanisms, transcript stability, or presence of cell subtypes.

Cell-type-enriched non-coding transcripts in VAT

We classified 307 non-coding transcripts as cell-type enriched in VAT, the highest number of which were in adipocytes (n = 144), followed by mesothelial cells (n = 89) (Figure 4A; Table S2, tab 1). Cell-enriched non-coding transcripts were typically expressed at lower levels than cell-enriched protein-coding transcripts (mean transcripts per million [TPM] ± SD, protein coding: 45.4 ± 118.5 versus non-coding: 3.06 ± 5.2), with a higher frequency of samples with low or no expression (mean % samples with expression >0.1 TPM ± SD, coding 2.7 ± 7.8 versus non-coding 12.4 ± 14.9). Cell-type-enriched non-coding transcripts predominantly clustered into the same WGCNA group(s)/clades as the corresponding protein-coding *Ref.T.*s (Figures 4A.ii and 4A.iii, *Ref T.* location marked by colored boxes), e.g., adipocyte enriched (orange ovals, groups 8, 16, 105, 121, 28, 44, and 38; Figure 4A.iii), consistent with cell type co-expression. We visualized up to the top 50 enriched non-coding transcripts for the four cell types with the highest number (Figures 4B.i–4B.iv), ranked by highest mean corr. with the *Ref.T.* panel, to compare differential corr. values (corr. with corresponding cell type *Ref.T.* panel minus max corr. with any other *Ref.T.* panel) and expression. Overall, expression values for non-coding enriched genes were highest for adipocytes (Figure 4B.i). Although there is no comparable existing dataset to validate these results, we used scRNA-seq data from the analysis of human SAT (Tabula Sapiens et al., 2022; Figure 4C.i) to make some comparisons. Although this dataset does not include adipocytes or mesothelial cells (those with the highest number of predicted enriched non-coding transcripts) and lacks data for many non-coding transcripts, it provides supportive evidence for our classifications in the other cell types (Figures 4C.ii–4C.viii). All enrichment scores for non-coding transcripts can be searched via the web portal https://cell-enrichment.shinyapps.io/noncoding/.

Identification of cell-type transcriptome profiles in subcutaneous adipose tissue (SAT)

White adipose tissue is broadly classified by location; VAT is intra-abdominal, adjacent to internal organs, while SAT lies underneath the skin. The proportion of VAT to SAT increases in obesity and is linked to metabolic dysregulation (Chait and den Hartig, 2020). In order to compare these two depots, SAT-cell-type-enriched profiles were determined as described for VAT, using human SAT RNA-seq data (n = 646) from GTEx portal

---

Figure 3. Core transcriptional identities of human VAT cell types

Cell-type-enriched transcripts in (A) adipocytes, (B) adipocyte progenitor cells, (C) mesothelial cells, (D) endothelial cells, (E) smooth muscle cells, (F) macrophages, (G) neutrophils, (H) mast cells, (I) T cells, and (J) plasma cells, displayed to show (i) proportional representation of transcript types (absolute numbers below) and (ii) the top-50 protein-coding genes ranked by corr. score, with differential expression scores (corr. with corresponding cell type *Ref.T.* panel minus max corr. with any other *Ref.T.* panel) and mean TPM expression. TEC, to be experimentally confirmed. See also Table S2, tab 1, and Figures S4 and S5.

(legend on next page)

159
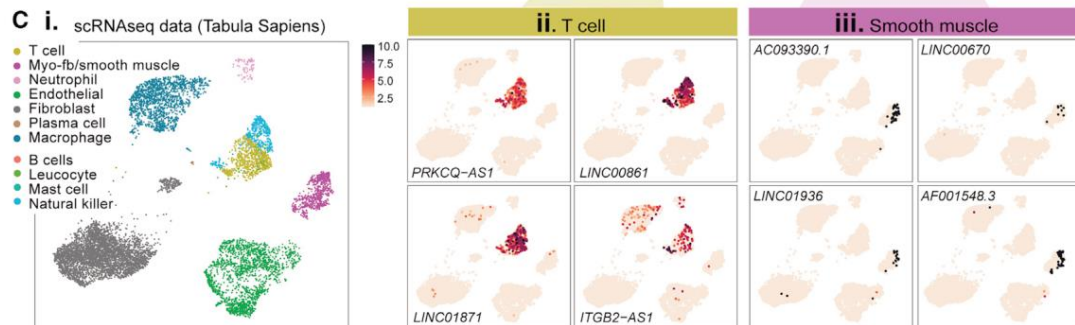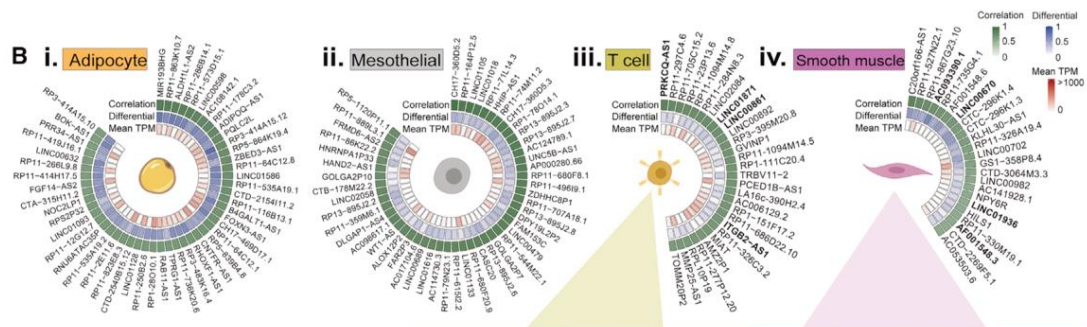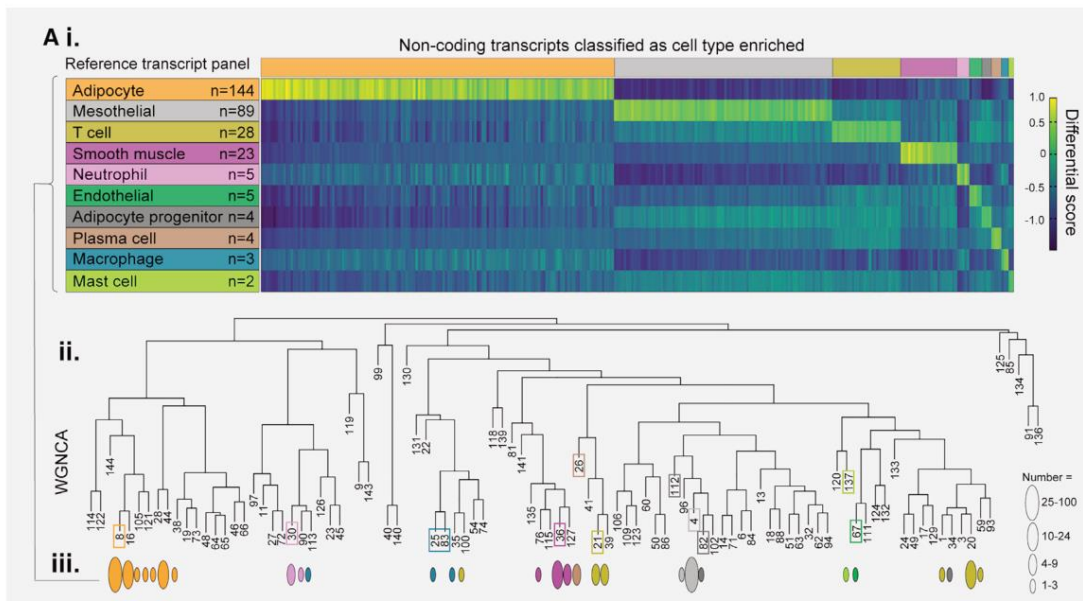
v.8 (www.gtexportal.org; Consortium, 2015). Adipocyte, adipocyte progenitor, endothelial, smooth muscle, macrophage, mast cell, T cell and plasma cell *Ref.T.*s had high intra-panel corr. (all >0.72; $p < 9.0 \times 10^{-98}$) with low inter-panel corr. (Table S1, tab 3). However, the *Ref.T.*s selected for the mesothelial cell and neutrophil panel in VAT did not correlate well with each other in SAT (mesothelial *Ref.T.* panel [mean corr. ± SD]: SAT 0.15 ± 0.04 versus VAT 0.92 ± 0.02; neutrophil *Ref.T.* panel: SAT 0.62 ± 0.08 versus VAT 0.81 ± 0.04). Expression of these genes was also much lower in SAT than VAT (mesothelial: *UPK3B* [SAT versus VAT TPM] 1.4 versus 125.8, *MSLN* 0.3 versus 144.8, *KRT19* 14.1 versus 153.4; neutrophil: *CSF3R* 0.81 versus 37.8, *FCGR3B* 2.1 versus 8.8, *CXCR2* 0.9 versus 4.6), indicating a low number, or absence, of these cell types in SAT, consistent with reports that mesothelial cells are absent (Esteve et al., 2019) and neutrophils preferentially infiltrate VAT rather than SAT (Elgazar-Carmon et al., 2008). Thus, these cell types were excluded from subsequent profiling of SAT. As for VAT, SAT cell-type-enriched transcripts were well separated by designated *Ref.T.* panels (Figure S3B) and clustered into related groups in WGCNA (Figure S3C), and terms identified by GO and reactome analysis were consistent with cell identity (Table S2, tabs 2–11, tables Bi and Bii).

Adipose tissue scRNA-seq is consistent with Ref.T. analysis

We performed a comparison between our results and scRNA-seq or small nuclear RNA-seq (snRNA-seq) data of human SAT or murine adipose tissue generated by Sun et al. (2020) (snRNA-seq), Hildreth et al. (2021), Tabula Sapiens et al. (2022), and Tabula Muris et al. (2018) (all scRNA-seq) (Table S1, tab 5). None of these studies contained all cell types we profiled; adipocytes were only in the snRNA-seq study from Sun et al. (2020) and plasma cells only in the Tabula Sapiens et al. (2022) dataset. For some cell types, e.g., progenitors, classification and/or terminology varied between studies, as is typical (Wang et al., 2021), and so comparisons were made between closely related cell types with common marker genes, e.g., "adipocyte progenitor," "pre-adipocytes," "fibroblasts," or "mesenchymal stem cells" (Table S1, tab 5 [row 2 states cell-type annotation]). For cell types represented in all, or most, of the independent studies, a high proportion of our predicted cell-type-enriched genes were elevated in the corresponding cell type in at least one (average Log2 fold change >1.0, >0.5, or >0.2 versus all other cell types [p < 0.01]): adipocyte progen-

itor (81%), endothelial cell (98%), smooth muscle cell (69%), macrophage (87%), neutrophil (96%), T cell (83%), and plasma cell (81%) enriched (Table S1, tab 5; Figures S4A and S4B). For adipocyte-enriched genes, independent validation was lowest of all cell types at 30%, which could be due to the limited coverage given by comparison with only a single study (Sun et al., 2020) or differences between the sensitivity of snRNA-seq versus bulk RNA-seq (Pimpalwar et al., 2020). Gene ontology and reactome analysis of the predicted adipocyte-enriched genes that were *not* consistent with data from Sun et al. (2020), revealed significant enrichment of terms associated with adipocyte function, e.g., "small molecule metabolic process" (adjusted FDR 8.7×10^{-29}) and "metabolism" (adjusted FDR 4.3×10^{-29}). To compare cell profiles across all datasets, we calculated the significance of overlap using a hypergeometric test (Figure S5). Genes predicted as cell-type enriched in our study were over-represented in the enriched genes in the corresponding cell types in the scRNA-seq and snRNA-seq studies (defined as those ≥0.5 Log2 fold change in expression versus all other cell types in the same study [p < 0.01]) (Figure S5), and this overlap was comparable to, or more significant than, that between the scRNA-seq and snRNA-seq studies themselves.

### Ref.T. analysis can predict source of adipose-tissue-enriched genes

RNA-seq data from unfractionated human or murine tissues can be used to identify genes with enriched expression in adipose tissue versus other tissues. Adipocytes make up the majority of adipose tissue, with the most specialized function. We extracted lists of the top 200 human-adipose-tissue-enriched genes from HPA (Uhlen et al., 2015) and GTEx (Consortium, 2015), collated in the Harminozome database (Rouillard et al., 2016; Figure S4C.i). Of those genes classified as adipose tissue enriched in both datasets (n = 86), our analysis classified 66 (76.7%) as adipocyte enriched and one (1.2%) as endothelial enriched (*ARHGEF15*) (Figure S4C.ii). Thus, our analysis indicates that the majority of adipose-tissue-enriched genes are selectively expressed in adipocytes.

Sex- and depot-specific differences in adipose-cell-type transcriptome profiles

There are sex- and depot-specific differences in accumulation, distribution, endocrine, and metabolic function of adipose tissue (Blaak, 2001; Chait and den Hartigh, 2020; Lumish et al.,

---

Figure 4. Integrative co-expression analysis of unfractionated RNA-seq reveals enriched non-coding transcriptomes of human VAT cell types

(A) (i) Heatmap of all non-coding transcripts classified as cell type enriched (indicated by horizontal colored bars), showing differential score between mean corr. with the corresponding *Ref.T.* panel versus highest mean corr. coefficient among the other *Ref.T.* panels. (ii) Human VAT RNA-seq data (n = 527) for all sequenced transcripts were subject to WGCNA. Colored squares indicate *Ref.T.* location on resultant dendrogram (colors correspond to cell types as annotated in A.i). (iii) Colored ovals indicate distribution of non-coding transcripts classified as cell type enriched across dendrogram groups.

(B) Cell-type-enriched non-coding transcripts in (i) adipocytes, (ii) mesothelial cells, (iii) T cells, and (iv) smooth muscle cells displayed to show up to the top 50 ranked by corr. score, with differential expression scores (corr. with corresponding cell type *Ref.T.* panel minus max corr. with any other *Ref.T.* panel) and mean TPM expression.

(C) scRNA-seq data from analysis of cell types from human subcutaneous adipose tissue were sourced from Tabula Sapiens (Tabula Sapiens et al., 2022) and used to generate uniform manifold approximation and projection (UMAP) plots showing (i) scRNA-seq cell-type annotations and the expression of examples of non-coding genes we predicted as being (ii) T cell, (iii) smooth muscle cell, (iv) neutrophil, (v) endothelial, (vi) adipocyte progenitor, (vii) plasma cell, or (viii) macrophage enriched.

See also Table S2, tab 1.

(legend on next page)

161
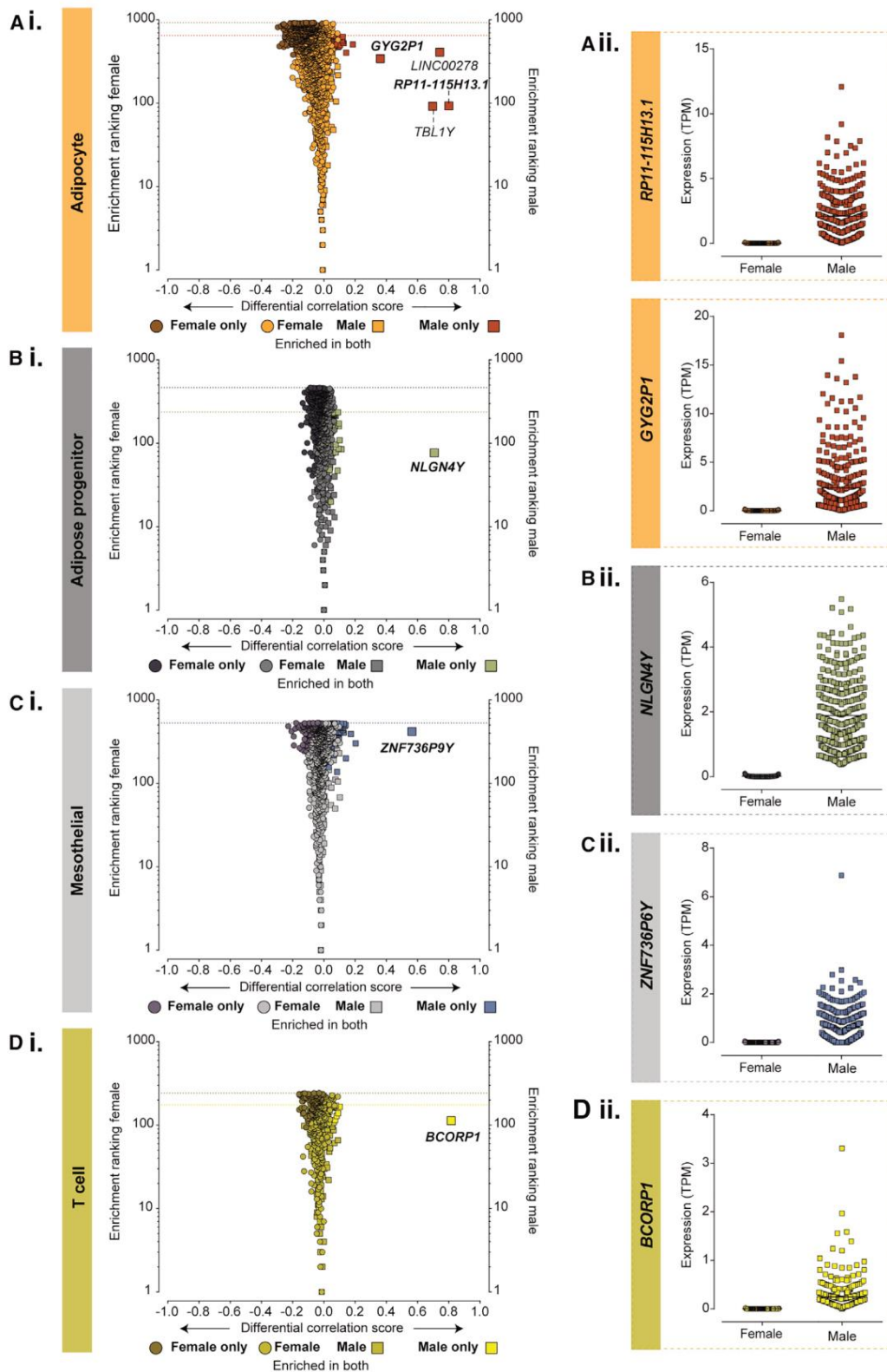
2020; Valencak et al., 2017), but to our knowledge, there are no studies describing sex- and depot-specific differences in adipose-cell-type-specific transcriptome profiles. Therefore, we profiled SAT-cell-type-enriched transcriptomes and performed a comparative sex subset analysis in VAT and SAT, and we did a comparison between cell types found in both depots.

Prediction of sex-specific differences in adipose-cell-type-enriched transcripts

We performed a subset analysis of the VAT RNA-seq dataset (female n = 165; male n = 362) to identify sex-specific, cell-type-enriched transcriptome profiles. As in the full dataset, intra-panel cell type *Ref.T*.s correlated well in female- and male-sample subsets (all >0.83; p < 1.0 × 10^{-33}) (Table S3, tab 1, tables A and B). We compared transcripts classified as male or female cell type enriched (Figure 5; Table S3, tab 2). Cell profiles were largely comparable between sexes (Figures 5 and S6; transcripts enriched in *both* males and females represented by common colored circle and square symbols, respectively). Some transcripts were classified as enriched *only* in males or females (Figures 5 and S6; represented by differently colored circle and square symbols, respectively); however, most had differential corr. scores close to zero, indicating that they fell marginally below the designated threshold for classification as enriched in the other sex. A small number of markedly male-only cell-type-enriched transcripts were identified in adipocytes (Figure 5A.i; *TBL1Y*, *RP11-115H13.1*, LINC00278, and *GYG2P1*), adipocyte progenitor cells (Figure 5B.i; *NLGN4Y*), mesothelial cells (Figure 5C.i; *ZNF736P9Y*), and T cells (Figure 5D.i; *BCORP1*). In all cases, transcripts were Y linked, and mRNA expression was only detected above background levels in male VAT samples (Figures 5A.ii–5D.ii). There were no clear sex-specific differences in the other cell-enriched transcriptome profiles (Figure S6).

Comparison of predicted sex-specific VAT- and SAT-cell-type-enriched transcripts

To establish whether these sex-specific differences also existed in SAT, we performed an equivalent subset analysis of the SAT RNA-seq dataset (female n = 212; male n = 434). As in the full dataset, intra-panel cell type *Ref.T*.s correlated well in female- and male-sample subsets (all >0.71; p < 14.0 × 10^{-31}) (Table S3, tab 3, tables A and B). We compared transcripts classified as male or female cell type enriched (Figures 5 and S7; Table S3, tab 4). Three out of the four transcripts we identified as adipocyte enriched in male VAT, but not female VAT, had the same profile in SAT (*TBL1Y*, *RP11-115H13.1*, and *GYG2P1*) (Figure S7A), showing consistency between adipose depot type. The single

transcript we identified as adipocyte progenitor enriched in male, but not female, VAT (*NLGN4Y*) (Figure 5B.i) was not classified as such in SAT (Figure S7B). However, the corr. value between *NLGN4Y* and the adipocyte progenitor *Ref.T*. panel fell marginally below the threshold for definition as enriched in SAT, and a clear male-female differential corr. existed (SAT male corr. 0.46 versus SAT female corr. 0.10) (Table S3, tab 4). The transcript we identified as mesothelial enriched in male, but not female, VAT (*ZNF736P9Y*) (Figure 5C.i) was not expressed in SAT, consistent with the absence of mesothelial cells in this depot (Esteve et al., 2019). The male-only T cell enriched in VAT (*BCORP1*) was excluded from analysis in SAT, due to low expression in the majority of samples (>50% with TPM < 0.1). There were no clear sex-specific differences in the other SAT-cell-type-enriched transcriptome profiles (Figures S7B–S7H).

Prediction of depot-specific differences in adipose-cell-type-enriched transcripts

Previous studies have reported differential gene expression profiles between VAT and SAT depots, using bulk sequencing data (Bradford et al., 2019; Schleinitz et al., 2020), but reports on differences at the cell-type level are lacking. Here, we compared transcripts classified as cell type enriched in VAT or SAT. As we profiled two additional cell types in VAT, compared with SAT (mesothelial cells and neutrophils), prior to comparison, we excluded any SAT-cell-type-enriched transcripts that were predicted as primarily enriched in neutrophils or mesothelial cells in VAT (see Table S1, tab 6). This exclusion revealed that 79 genes were predominantly enriched in different cell types in VAT and SAT, e.g., *IL18* was classified as macrophage enriched in SAT but mesothelial-enriched in VAT (Table S1, tab 6, line 74), where its expression was higher (mean TMP ± SD; VAT 44.0 ± 41.1 versus SAT 11.8 ± 10.2). Adipocyte-enriched profiles were similar between depots, with around 500 transcripts classified as such in both VAT and SAT (Figure 6A; represented by common colored circle and square symbols, respectively) (Table S2, tab 1). *LINC01632* and *GPAT3* were classified as adipocyte enriched in VAT only (Figure 6A), and both were expressed at higher levels in VAT than SAT (Figures 6C.i and 6C.ii). Conversely, *NRCAM*, *MAGI2-AS3*, and *SEPT11* were classified as adipocyte enriched in SAT only (Figure 6A) and were all expressed at higher levels in SAT than VAT (Figures 6C.iii–6C.v). These data are consistent with these transcripts having both an adipocyte-restricted and depot-restricted profile. Glucagon-like peptide-2 receptor (*GLP2R*) was classified as adipocyte enriched in SAT only (Figure 6A), but it was expressed at higher levels in VAT (mean TMP ± STD; VAT 11.2 ± 11.28 versus SAT 1.37 ± 0.85). This gene could be strongly

---

Figure 5. Identification of sex-specific, cell-enriched transcripts in human VAT

VAT RNA-seq data (n = 527 individuals) were divided into female and male subgroups (female n = 165; male n = 362) before classification of cell-type-enriched transcripts. For transcripts classified as (A.i) adipocyte, (B.i) adipocyte progenitor, (C.i) mesothelial, or (D.i) T cell enriched, in either sex, the "sex differential corr. score" (difference between mean corr. with the *Ref.T*. panel in females versus males) was plotted versus "enrichment ranking" (position in each respective enriched list; highest corr. = rank 1). On each plot, transcripts enriched in *both* females and males are represented by common colored circle and square symbols, respectively, and transcripts classified as enriched *only* in females or males are represented by differently colored circle and square symbols, respectively. Expression in female and male samples for transcripts identified as male-only enriched in (A.ii) adipocytes, (B.ii) adipocyte progenitor, (C.ii) mesothelial, or (D.ii) T cells. See also Figure S6 and Table S3.

(legend on next page)

163
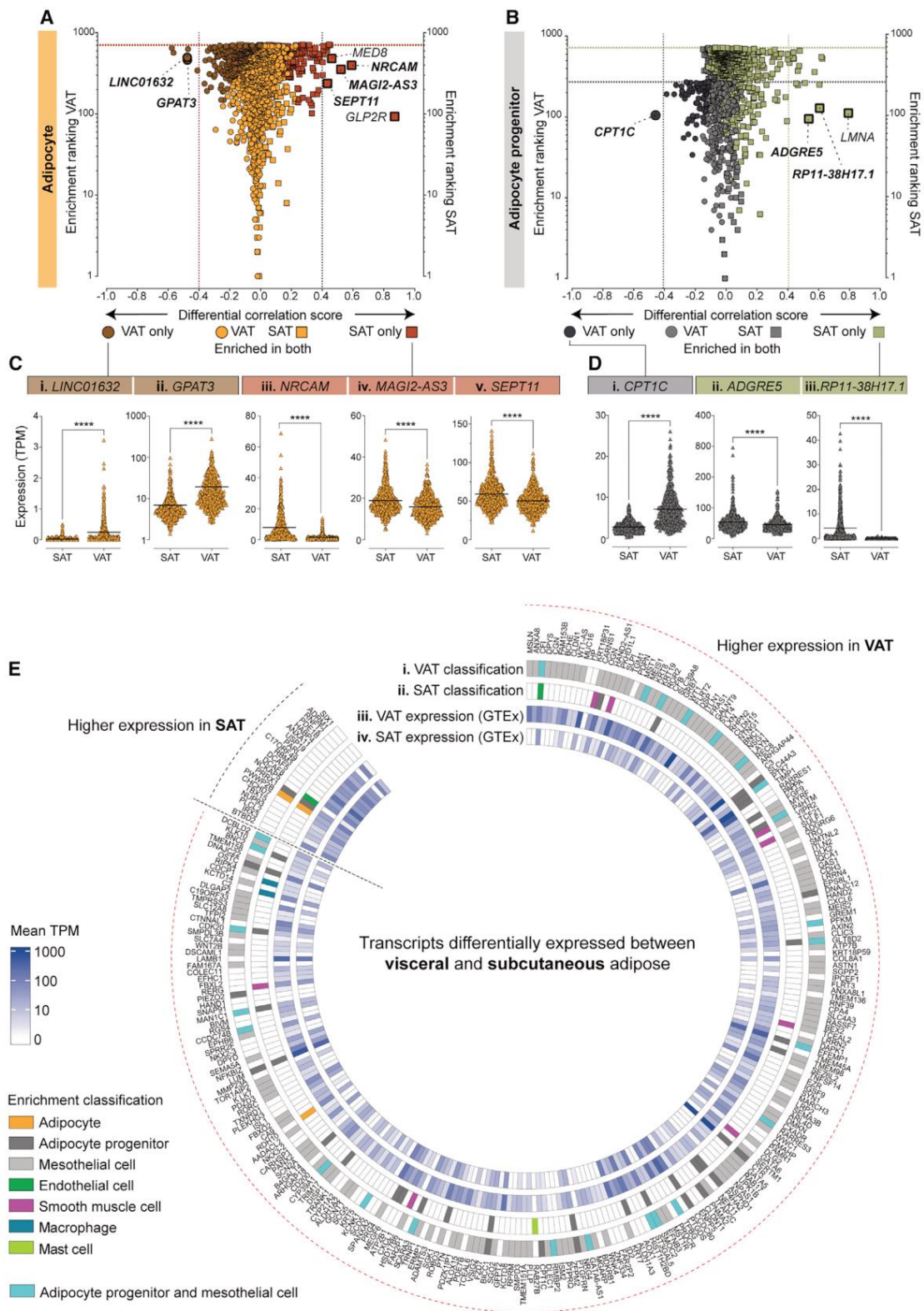
transcriptionally regulated by environmental factors, analogous to regulation of the related glucagon receptor gene in response to glucose (Svoboda et al., 1999), in VAT only and thus here does not consistently correlate with the stably expressed adipocyte *Ref. T*. Like adipocytes, adipocyte-progenitor-enriched profiles were similar between depots, with 186 transcripts classified as such in both VAT and SAT (Figure 6B; represented by common colored circle and square symbols, respectively) (Table S2, tab 1). *CPT1C* was classified as adipocyte progenitor enriched in VAT only (Figure 6B) and expressed at higher levels in VAT than SAT (Figure 6D.i). Conversely, *ADGRE5* and *RP11-38H17.1* were classified as adipocyte progenitor enriched in SAT only (Figure 6B) and were expressed at higher levels in SAT than VAT (Figures 6D.ii and 6D.iii). These data are consistent with these transcripts having both an adipocyte-progenitor-restricted and depot-restricted profile. There were no clear depot-specific differences in the other cell-type-enriched transcriptome profiles (Figures S8A–S8F).

As our analysis indicated that cell-type-enrichment profiles did not differ substantially in VAT and SAT, we investigated the cell-type-expression profile of genes identified as most differentially expressed between depots. We extracted data from a study (Schleinitz et al., 2020) where the authors analyzed samples from 15 human-fat depots and generated a list of most differentially expressed genes between SAT and VAT. Of the 298 transcripts identified by the authors, data for 272 were available in our analysis. We performed lookups in our dataset to determine whether these genes were classified cell type enriched and whether this expression profile differed between VAT and SAT (Figure 6E). For transcripts with a higher expression in VAT, compared with SAT (Schleinitz et al., 2020; Figure 6E, indicated by red dashed external line), the majority were classified as mesothelial cell enriched in VAT in our analysis (Figure 6E.i), a cell type not found in SAT (Esteve et al., 2019). A further 25 of these transcripts were found in the list of 84 we excluded from cell-type classification, due to likely co-expression in both VAT mesothelial and adipocyte progenitor cells (Figure 6E.i; Table S1, tab 4a). For those transcripts with a higher expression in SAT versus VAT (Schleinitz et al., 2020; Figure 6E, indicated by black dashed line), most were not annotated as cell type enriched, but those that were had similar expression profiles between depots. Classification as depot enriched in Schleinitz et al. (2020) was broadly consistent with corresponding TPMs in the GTEx data (Figures 6E.iii and 6E.iv). The application of our data in this way demonstrates its usefulness for extracting cell-type information from whole-tissue data, allowing further understanding of observations made in other studies, with broad applicability across datasets and analysis platforms.

DISCUSSION

Here, we present a method to resolve unfractionated tissue RNA-seq data, providing an alternative to scRNA-seq for the identification of cell-type-enriched transcripts. Our approach circumvents some challenges associated with scRNA-seq, e.g., requirement for fresh tissue, artefact-generating sample pre-processing, and limited read depth (Beliakova-Bethell et al., 2014; Rizzetto et al., 2017; Saliba et al., 2014; Ziegenhain et al., 2017). By analyzing a high number of biological replicates, this approach allows for well-powered subgroup comparisons, e.g., female versus male. Public repositories contain thousands of bulk RNA-seq datasets; our method can utilize these resources to profile cell types for which little or no information currently exists.

To our knowledge, this study provides the most comprehensive publicly accessible database of adipose-tissue-cell-type coding and non-coding gene-expression-enrichment profiles, searchable on a gene-by-gene basis. Our dataset could also be a useful tool for the optimization of deconvolution algorithms used to determine proportions of constituent cell types in adipose tissue bulk RNA-seq, e.g., CIBERSORT (Glastonbury et al., 2019; Newman et al., 2015). Such analyses typically use input expression matrices generated from transcriptome analysis of isolated cell types to identify cell-type reference genes. Various factors can reduce the accuracy of input matrices, including contaminating cell types in input datasets, technical artefacts due to cell extraction and processing, and limited input data availability for some cell types or for cells sourced from adipose tissue. Cross checking input matrices against our dataset could identify the most likely highly enriched genes *in vivo*.

Genes classified as adipocyte enriched in VAT or SAT included those with established roles in adipocyte development or function, e.g., *GPD1*, *AQP7*, *LPL* (Rotondo et al., 2017), *CIDEC* (Keller et al., 2008), *GYG2*, *TUSC5*, and *PPP1R1A* (Ambele et al., 2016), but others had no known function, e.g., *HEPACAM*, *PECR*, *C19orf12*, and AL845331.1. *HEPACAM* encodes an adhesion molecule studied mainly in brain glial cells (Barrallo-Gimeno and Estevez, 2014), but it was identified as a key driver in a regulatory gene network associated with BMI and cholesterol in VAT from patients with coronary artery disease (Franzen et al., 2016). *HEPACAM* was one of 47 genes differentially expressed in SAT

Figure 6. Identification of depot-specific, cell-enriched transcripts in human adipose tissue

Human VAT (n = 527 individuals) or SAT (n = 646 individuals) RNA-seq data were used for classification of cell-type-enriched transcripts (see results for criteria).

(A and B) For transcripts classified as (A) adipocyte or (B) adipocyte progenitor enriched, in either VAT or SAT, the "depot differential corr. score" (difference between mean corr. with the *Ref.T.* panel in VAT versus SAT) was plotted versus enrichment ranking (position in each respective enriched list; highest corr. = rank 1). On each plot, transcripts enriched in *both* VAT and SAT are represented by common colored circle and square symbols, respectively, and transcripts classified as enriched *only* in VAT or SAT are represented by differently colored circle or square symbols, respectively. Correspondingly colored threshold lines denote ranking below which transcripts were classified as VAT or SAT enriched.

(C and D) Expression levels in SAT and VAT of (C) transcripts classified as adipocyte enriched in (i and ii) VAT only or (iii–v) SAT only; (D) adipocyte progenitor-enriched classified transcripts in (i) VAT only or (ii and iii) SAT only.

(E) Transcripts identified as differentially expressed between VAT and SAT depots by Schleinitz et al. (2020) are displayed with cell-type-enrichment classification in our analysis of (i) VAT and (ii) SAT. Corresponding expression levels in the GTEx datasets are displayed for (iii) VAT and (iv) SAT.

See also Figure S8 and Tables S1, tab 6, and S2.

from twin pairs with high and low BMI and was associated with adipocyte diameter (Kaartinen et al., 2020). *PECR* is involved in chain elongation of fatty acids (Gloerich et al., 2006) and is a candidate gene influencing fat mass in mice (Karst et al., 2011), intramuscular fat deposition in cows (Sadkowski et al., 2014), and pig weight (Stuczynska et al., 2018). Mutations in *C19orf12* cause neurodegeneration with brain iron accumulation (Gagliardi et al., 2015), and *C19orf12* could have a role in lipid homeostasis (Hartig et al., 2011), due to high expression in adipose tissue and co-regulation with genes involved in fatty-acid metabolism. *AL845331.1* has been re-classified from non-coding to novel protein coding, on the basis of its similarity to *AQP7*, a gene we also classified as adipocyte enriched, as have others (Rotondo et al., 2017).

Non-coding RNAs are increasingly recognized as important in adipose biology (Squillaro et al., 2020; Statello et al., 2021; Xu and Sun, 2020), but descriptions of adipose-cell-type expression profiles are lacking. In our analysis, adipocytes had the most enriched non-coding genes, including antisense transcripts to adipocyte-enriched protein-coding genes, e.g., *ALDH1L1-AS2*, *ADIPOQ-AS1*, *LIPE-AS1*, and *CNTFR-AS1*. Other adipocyte-enriched non-coding genes included *RP11-863K10.7*, an antisense transcript to *ERLIN2*, a gene with a role in the accumulation of cytosolic lipid droplets (Wang et al., 2012); *MIRLET7BHG*, which is important for adipocyte differentiation in mice (McGregor and Choi, 2011; Sun et al., 2009); and *MIR193BHG*, which was characterized as a cellular steroid biosynthesis pathway modulator in MCF7 cells (Wu et al., 2020). Mesothelial-cell-enriched non-coding genes included antisense transcripts to mesothelial-enriched protein-coding genes, e.g., *SEMA3B-AS*, *DPP10-AS1*, *FAM83H-AS1*, and *WT1-AS*. Other mesothelial-enriched non-coding genes included *LINC01133*, reported as having a role in the Wnt signaling pathway (Yang et al., 2021), which is associated with metabolic disease development, with adipose depot-specific roles (Chen and Wang, 2018). Most non-coding transcripts classified as T cell or plasma cell enriched in our analysis were TR genes or IG genes, respectively. Other non-coding transcripts classified as T cell enriched included *PRKCQ-AS1*, which was postulated to have a role in T cell function in a study of lncRNAs in vaccine response (de Lima et al., 2019) and targets the protein-coding gene *PRKCQ*, which was also classified as T cell enriched in our analysis. However, the majority of non-coding transcripts we identified as cell type enriched are uncharacterized.

Sex differences in the accumulation, distribution, and endocrine and metabolic function of adipose tissue is well acknowledged (Blaak, 2001; Lumish et al., 2020; Valencak et al., 2017), although studies on underlying cell-type gene-expression differences are limited. Cell-type profiles were similar between sexes, but we did identify a small panel of transcripts with sex-dependent enrichment profiles, including *TBL1Y*, *GYG2P1*, and *RP11-115H13.1*, which were adipocyte enriched only in male VAT and SAT. *TBL1Y*, a Y-linked gene similar to its gonosomal homologue *TBL1X*, is one of 27 genes that encode distinct male-specific Y proteins (Jeffery et al., 2013). *TBL1Y* has a role in hereditary hearing loss (Di Stazio et al., 2019) and cardiac developmental regulation (Meyfour et al., 2017), the latter of which has been suggested to contribute to the sexual

dimorphism of cardiac diseases. Although previously reported as expressed in adipose tissue (Jeffery et al., 2013), there are no reports of *TBL1Y* being adipocyte specific or its function there. *GYG2P1* is a Y-linked pseudogene of *GYG2* (Meyfour et al., 2017). *GYG2* was classified as adipocyte enriched in both males and females in our analysis; although its function in adipocytes has not been studied, its expression coincides with adipocyte maturation of adipose-derived stromal cells (Ambele et al., 2016). Although pseudogenes are often assumed to lack function, they are increasingly found to have key roles (Cheetham et al., 2020) functioning as antisense, endogenous small-interference or competing endogenous transcripts (Singh et al., 2020). There are no reports of *GYG2P1* function, but it was downregulated in SAT from children with obesity versus those without (Liu et al., 2018). A recent study used bulk RNA-seq to analyze SAT from females and males to identify genes with differential expression (Anderson et al., 2020). According to our data, the identified genes had similar cell-type enrichment profiles between sexes (majority adipocyte enriched). Thus, sex differences in SAT are likely driven by variable gene expression in a common cell type or differences in the proportion of this cell type.

Previous studies have identified differences in cellular composition, adipocyte size, activity, and capacity for fat uptake between VAT and SAT (Ibrahim, 2010). Although RNA-seq has been used to determine differences in gene expression between VAT and SAT (Bradford et al., 2019; Schleinitz et al., 2020), to our knowledge, there have been no studies comparing cell-type-enriched transcriptome profiles. We found that *NRCAM*, a neuronal cell adhesion molecule of the immunoglobulin superfamily, mainly studied in a neuronal development (Sakurai, 2012), was adipocyte enriched in SAT, but not VAT. *NRCAM* was one of 32 genes upregulated throughout the differentiation of human-adipose-derived stromal cells isolated from SAT (Ambele et al., 2016), and in a study of SAT from siblings with high and low BMI, *NRCAM* was identified as part of an obesity-related transcript network (Walley et al., 2012). *NRCAM* was expressed in SAT, but not VAT, from individuals with extreme obesity (Gerhard et al., 2014). However, *NRCAM* function is unknown. In SAT, but not VAT, *CALB2* and *PKP2* were adipocyte enriched and *IL18* was macrophage enriched. In all three cases, overall expression was higher in VAT than SAT, and these genes were predominantly expressed in mesothelial cells in VAT, consistent with previous reports for *IL18* (Darimont et al., 2008) and *CALB2* (Barberis et al., 1997). Indeed, our data show that mesothelial cells in VAT drive differences in global gene expression between depots (Bradford et al., 2019; Schleinitz et al., 2020).

In summary, our method circumvents some challenges associated with the analysis of adipose tissue to provide an atlas of constituent cell type defining transcriptional profiles. The data can be used to further interpretate existing observations and to identify candidates for functional studies to expand our knowledge of adipose tissue in health and disease.

Limitations of the study

There are limitations to our study. We do not profile specific cell subtypes; while it may be possible to resolve the data further in this way, there is a lack of consensus regarding cell-subtype

identity, e.g., multiple adipocyte progenitor cells subtypes have been reported (Raajendiran et al., 2019), but others claim this population is homogeneous (Acosta et al., 2017). Thus, selection of subtype *Ref.T*.s required for input into our analysis model, or interpretation of WGNCA, is challenging. Thus, our classification informs about cell-type restricted expression but does not discriminate between transcripts expressed uniformly across all cells of a given type and those expressed in a sub-population. Some cell types are not profiled in our analysis, due to difficulties in the identification of cell-type-specific markers as suitable *Ref.T*.s. Thus, some genes classified as cell type enriched in our analysis may also be expressed in other (non-profiled) cell types, a limitation that applies to existing scRNA-seq and snRNA-seq adipose tissue datasets, which all lack data for some constituent cell types (e.g., Hildreth et al., 2021; Tabula Muris et al., 2018; Tabula Sapiens et al., 2022; Vijay et al., 2020). Expression of some genes in adipose tissue can be modified by genetic, epigenetic, or environmental factors (Sun et al., 2019). Such genes may not correlate with the *Ref.T*.s, due to a variation in expression that is independent of cell-type proportions. Thus, such genes could be false negatives in our analysis. We have used high thresholds for classification of genes as cell type enriched, likely leading to the incorrect exclusion of some. For example, *EPAS1*, *SHROOM4*, and *GPR4* are endothelial-enriched transcripts across tissue beds (Butler et al., 2016), but they fall just below the threshold for classification as endothelial enriched here. However, in these cases, the enrichment score clearly indicates a cell-type-restricted expression; thus, our classifications are intended only as a guide, and the reader should consider the data on a transcript-by-transcript basis.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Tissue profiling: human tissue sections
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Reference transcript-based correlation analysis
  - Weighted correlation network (WGCNA) analysis
  - Gene ontology and reactome analysis
  - Processing of data from adipose tissue scRNA-seq and snRNA-seq datasets
  - Visualization
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111046.

REFERENCES

Acosta, J.R., Joost, S., Karlsson, K., Ehrlund, A., Li, X., Aouadi, M., Kasper, M., Arner, P., Rydén, M., and Laurencikiene, J. (2017). Single cell transcriptomics suggest that human adipocyte progenitor cells constitute a homogeneous cell population. Stem Cell Res. Ther. *8*, 250. https://doi.org/10.1186/s13287-017-0701-4.

Ambele, M.A., Dessels, C., Durandt, C., and Pepper, M.S. (2016). Genome-wide analysis of gene expression during adipogenesis in human adipose-derived stromal cells reveals novel patterns of gene expression during adipocyte differentiation. Stem Cell Res. *16*, 725–734. https://doi.org/10.1016/j.scr.2016.04.011.

Anderson, W.D., Soh, J.Y., Innis, S.E., Dimanche, A., Ma, L., Langefeld, C.D., Comeau, M.E., Das, S.K., Schadt, E.E., Björkegren, J.L., and Civelek, M. (2020). Sex differences in human adipose tissue gene expression and genetic regulation involve adipogenesis. Genome Res. *30*, 1379–1392. https://doi.org/10.1101/gr.264614.120.

Armulik, A., Genové, G., and Betsholtz, C. (2011). Pericytes: developmental, physiological, and pathological perspectives, problems, and promises. Dev. Cell *21*, 193–215. https://doi.org/10.1016/j.devcel.2011.07.001.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene Ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29. https://doi.org/10.1038/75556.

Barberis, M.C., Faleri, M., Veronese, S., Casadio, C., and Viale, G. (1997). Calretinin. Acta Cytol. *41*, 1757–1761. https://doi.org/10.1159/000333181.

Barrallo-Gimeno, A., and Estevez, R. (2014). GlialCAM, a glial cell adhesion molecule implicated in neurological disease. Adv. Neurobiol. *8*, 47–59. https://doi.org/10.1007/978-1-4614-8090-7_3.

Beliakova-Bethell, N., Massanella, M., White, C., Lada, S.M., Du, P., Vaida, F., Blanco, J., Spina, C.A., and Woelk, C.H. (2014). The effect of cell subset isolation method on gene expression in leukocytes. Cytometry *85*, 94–104. https://doi.org/10.1002/cyto.a.22352.

Blaak, E. (2001). Gender differences in fat metabolism. Curr. Opin. Clin. Nutr. Metab. Care *4*, 499–502. https://doi.org/10.1097/00075197-200111000-00006.

166

Bradford, S.T., Nair, S.S., Statham, A.L., van Dijk, S.J., Peters, T.J., Anwar, F., French, H.J., von Martels, J.Z.H., Sutcliffe, B., Maddugoda, M.P., et al. (2019). Methylome and transcriptome maps of human visceral and subcutaneous adipocytes reveal key epigenetic differences at developmental genes. Sci. Rep. 9, 9511. https://doi.org/10.1038/s41598-019-45777-w.

Britton, K.A., Massaro, J.M., Murabito, J.M., Kreger, B.E., Hoffmann, U., and Fox, C.S. (2013). Body fat distribution, incident cardiovascular disease, cancer, and all-cause mortality. J. Am. Coll. Cardiol. 62, 921–925. https://doi.org/10.1016/j.jacc.2013.06.027.

Butler, L.M., Hallström, B., Fagerberg, L., Pontén, F., Ponten, F., Uhlén, M., Uhlen, M., Renné, T., and Odeberg, J. (2016). Analysis of body-wide unfractionated tissue data to identify a core human endothelial transcriptome. Cell Syst. 3, 287–301.e3. https://doi.org/10.1016/j.cels.2016.08.001.

Chait, A., and den Hartigh, L.J. (2020). Adipose tissue distribution, inflammation and its metabolic consequences, including diabetes and cardiovascular disease. Front. Cardiovasc. Med 7, 22. https://doi.org/10.3389/fcvm.2020.00022.

Cheetham, S.W., Faulkner, G.J., and Dinger, M.E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. Nat. Rev. Genet. 21, 191–201. https://doi.org/10.1038/s41576-019-0196-1.

Chen, N., and Wang, J. (2018). Wnt/β-Catenin signaling and obesity. Front. Physiol. 9, 792. https://doi.org/10.3389/fphys.2018.00792.

Consortium, G.T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660. https://doi.org/10.1126/science.1262110.

Darimont, C., Avanti, O., Blancher, F., Wagniere, S., Mansourian, R., Zbinden, I., Leone-Vautravers, P., Fuerholz, A., Giusti, V., Giusti, V., et al. (2008). Contribution of mesothelial cells in the expression of inflammatory-related factors in omental adipose tissue of obese subjects. Int. J. Obes. 32, 112–120. https://doi.org/10.1038/sj.ijo.0803688.

de Lima, D.S., Cardozo, L.E., Maracaja-Coutinho, V., Suhrbier, A., Mane, K., Jeffries, D., Silveira, E.L.V., Amaral, P.P., Rappuoli, R., de Silva, T.I., and Nakaya, H.I. (2019). Long noncoding RNAs are involved in multiple immunological pathways in response to vaccination. Proc. Natl. Acad. Sci. USA 116, 17121–17126. https://doi.org/10.1073/pnas.1822046116.

Di Stazio, M., Collesi, C., Vozzi, D., Liu, W., Myers, M., Morgan, A., D'Adamo, P.A., Girotto, G., Rubinato, E., Giacca, M., and Gasparini, P. (2019). TBL1Y: a new gene involved in syndromic hearing loss. Eur. J. Hum. Genet. 27, 466–474. https://doi.org/10.1038/s41431-018-0282-4.

Dusart, P., Hallström, B.M., Renné, T., Odeberg, J., Uhlén, M., and Butler, L.M. (2019). A systems-based map of human brain cell-type enriched genes and malignancy-associated endothelial changes. Cell Rep. 29, 1690–1706.e4. https://doi.org/10.1016/j.celrep.2019.09.088.

Elgazar-Carmon, V., Rudich, A., Hadad, N., and Levy, R. (2008). Neutrophils transiently infiltrate intra-abdominal fat early in the course of high-fat feeding. J. Lipid Res. 49, 1894–1903. https://doi.org/10.1194/jlr.m800132-jlr200.

Estève, D., Boulet, N., Belles, C., Zakaroff-Girard, A., Decaunes, P., Briot, A., Veeranagouda, Y., Didier, M., Remaury, A., Guillemot, J.C., et al. (2019). Lobular architecture of human adipose tissue defines the niche and fate of progenitor cells. Nat. Commun. 10, 2549. https://doi.org/10.1038/s41467-019-09992-3.

Franzén, O., Ermel, R., Cohain, A., Akers, N.K., Di Narzo, A., Talukdar, H.A., Foroughi-Asl, H., Giambartolomei, C., Fullard, J.F., Sukhavasi, K., et al. (2016). Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. Science 353, 827–830. https://doi.org/10.1126/science.aad6970.

Franzen, O., Gan, L.M., and Bjorkegren, J.L.M. (2019). PanglaoDB: A Web Server for Exploration of Mouse and Human Single-Cell RNA Sequencing Data (Database).

Gagliardi, M., Annesi, G., Lesca, G., Broussolle, E., Iannello, G., Vaiti, V., Gambardella, A., and Quattrone, A. (2015). C19orf12 gene mutations in patients with neurodegeneration with brain iron accumulation. Park. Relat. Disord. 21, 813–816. https://doi.org/10.1016/j.parkreldis.2015.04.009.

Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49, D325–D334. https://doi.org/10.1093/nar/gkaa1113.

Gerhard, G.S., Styer, A.M., Strodel, W.E., Roesch, S.L., Yavorek, A., Carey, D.J., Wood, G.C., Petrick, A.T., Gabrielsen, J., Gabrielsen, J., et al. (2014). Gene expression profiling in subcutaneous, visceral and epigastric adipose tissues of patients with extreme obesity. Int. J. Obes. 38, 371–378. https://doi.org/10.1038/ijo.2013.152.

Glastonbury, C.A., Couto Alves, A., El-Sayed Moustafa, J.S., and Small, K.S. (2019). Cell-type heterogeneity in adipose tissue is associated with complex traits and reveals disease-relevant cell-specific eQTLs. Am. J. Hum. Genet. 104, 1013–1024. https://doi.org/10.1016/j.ajhg.2019.03.025.

Gloerich, J., Ruiter, J.P., van den Brink, D.M., Ofman, R., Ferdinandusse, S., and Wanders, R.J. (2006). Peroxisomal trans-2-enoyl-CoA reductase is involved in phytol degradation. FEBS Lett. 580, 2092–2096. https://doi.org/10.1016/j.febslet.2006.03.011.

Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). Circlize Implements and enhances circular visualization in R. Bioinformatics 30, 2811–2812. https://doi.org/10.1093/bioinformatics/btu393.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., et al. (2020). Construction of a human cell landscape at single-cell level. Nature 581, 303–309. https://doi.org/10.1038/s41586-020-2157-4.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

Harrow, J.L., Steward, C.A., Frankish, A., Gilbert, J.G., Gonzalez, J.M., Loveland, J.E., Mudge, J., Sheppard, D., Thomas, M., Trevanion, S., and Wilming, L.G. (2014). The vertebrate genome annotation browser 10 years on. Nucleic Acids Res. 42, D771–D779. https://doi.org/10.1093/nar/gkt1241.

Hartig, M.B., Iuso, A., Haack, T., Kmiec, T., Jurkiewicz, E., Heim, K., Roeber, S., Tarabin, V., Dusi, S., Krajewska-Walasek, M., et al. (2011). Absence of an orphan mitochondrial protein, c19orf12, causes a distinct clinical subtype of neurodegeneration with brain iron accumulation. Am. J. Hum. Genet. 89, 543–550. https://doi.org/10.1016/j.ajhg.2011.09.007.

Hildreth, A.D., Ma, F., Wong, Y.Y., Sun, R., Pellegrini, M., and O'Sullivan, T.E. (2021). Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity. Nat. Immunol. 22, 639–653. https://doi.org/10.1038/s41590-021-00922-4.

Hilgendorf, K.I. (2021). Primary cilia are critical regulators of white adipose tissue expansion. Front. Physiol. 12, 769367. https://doi.org/10.3389/fphys.2021.769367.

Hu, E., Liang, P., and Spiegelman, B.M. (1996). AdipoQ is a novel adipose-specific gene dysregulated in obesity. J. Biol. Chem. 271, 10697–10703. https://doi.org/10.1074/jbc.271.18.10697.

Ibrahim, M.M. (2010). Subcutaneous and visceral adipose tissue: structural and functional differences. Obes. Rev. 11, 11–18. https://doi.org/10.1111/j.1467-789x.2009.00623.x.

Jeffery, H.C., Söderberg-Naucler, C., and Butler, L.M. (2013). Human cytomegalovirus induces a biphasic inflammatory response in primary endothelial cells. J. Virol. 87, 6530–6535. https://doi.org/10.1128/jvi.00265-13.

Kaartinen, M.T., Arora, M., Heinonen, S., Rissanen, A., Kaprio, J., and Pietiläinen, K.H. (2020). Transglutaminases and obesity in humans: association of F13A1 to adipocyte hypertrophy and adipose tissue immune response. Int. J. Mol. Sci. 21, 8289. https://doi.org/10.3390/ijms21218289.

Kahn, C.R., Wang, G., and Lee, K.Y. (2019). Altered adipose tissue and adipocyte function in the pathogenesis of metabolic syndrome. J. Clin. Invest. 129, 3990–4000. https://doi.org/10.1172/jci129187.

Karlsson, M., Zhang, C., Méar, L., Zhong, W., Digre, A., Katona, B., Sjöstedt, E., Butler, L., Odeberg, J., Dusart, P., et al. (2021). A single–cell type transcriptomics map of human tissues. Sci. Adv. 7, eabh2169. https://doi.org/10.1126/sciadv.abh2169.

Kärst, S., Cheng, R., Schmitt, A.O., Yang, H., de Villena, F.P.M., Palmer, A.A., and Brockmann, G.A. (2011). Genetic determinants for intramuscular fat content and water-holding capacity in mice selected for high muscle mass. Mamm. Genome *22*, 530–543. https://doi.org/10.1007/s00335-011-9342-6.

Keller, P., Petrie, J.T., De Rose, P., Gerin, I., Wright, W.S., Chiang, S.H., Nielsen, A.R., Fischer, C.P., Pedersen, B.K., and MacDougald, O.A. (2008). Fat-specific protein 27 regulates storage of triacylglycerol. J. Biol. Chem. *283*, 14355–14365. https://doi.org/10.1074/jbc.m708323200.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

Liu, Y., Ji, Y., Li, M., Wang, M., Yi, X., Yin, C., Wang, S., Zhang, M., Zhao, Z., and Xiao, Y. (2018). Integrated analysis of long noncoding RNA and mRNA expression profile in children with obesity by microarray analysis. Sci. Rep. *8*, 8750. https://doi.org/10.1038/s41598-018-27113-w.

Lu, J., Zhao, J., Meng, H., and Zhang, X. (2019). Adipose tissue-resident immune cells in obesity and type 2 diabetes. Front. Immunol. *10*, 1173. https://doi.org/10.3389/fimmu.2019.01173.

Lumish, H.S., O'Reilly, M., and Reilly, M.P. (2020). Sex differences in genomic drivers of adipose distribution and related cardiometabolic disorders: opportunities for precision medicine. Arterioscler. Thromb. Vasc. Biol. *40*, 45–60. https://doi.org/10.1161/atvbaha.119.313154.

Martinez, F.O., Gordon, S., Locati, M., and Mantovani, A. (2006). Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. J. Immunol. *177*, 7303–7311. https://doi.org/10.4049/jimmunol.177.10.7303.

A McGregor, R., and S Choi, M. (2011). microRNAs in the regulation of adipogenesis and obesity. Curr. Mol. Med. *11*, 304–316. https://doi.org/10.2174/156652411795677990.

Meyfour, A., Ansari, H., Pahlavan, S., Mirshahvaladi, S., Rezaei-Tavirani, M., Gourabi, H., Baharvand, H., and Salekdeh, G.H. (2017). Y chromosome missing protein, TBL1Y, may play an important role in cardiac differentiation. J. Proteome Res. *16*, 4391–4402. https://doi.org/10.1021/acs.jproteome.7b00391.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. *8*, 1551–1566. https://doi.org/10.1038/nprot.2013.092.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. Nucleic Acids Res. *44*, D336–D342. https://doi.org/10.1093/nar/gkv1194.

Min, S.Y., Desai, A., Yang, Z., Sharma, A., DeSouza, T., Genga, R.M.J., Kucukural, A., Lifshitz, L.M., Nielsen, S., Scheele, C., et al. (2019). Diverse repertoire of human adipocyte subtypes develops from transcriptionally distinct mesenchymal progenitor cells. Proc. Natl. Acad. Sci. USA *116*, 17970–17979. https://doi.org/10.1073/pnas.1906512116.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457. https://doi.org/10.1038/nmeth.3337.

Oikonomou, E.K., and Antoniades, C. (2019). The role of adipose tissue in cardiovascular health and disease. Nat. Rev. Cardiol. *16*, 83–99. https://doi.org/10.1038/s41569-018-0097-6.

Pimpalwar, N., Czuba, T., Smith, M.L., Nilsson, J., Gidlöf, O., and Smith, J.G. (2020). Methods for isolation and transcriptional profiling of individual cells from the human heart. Heliyon *6*, e05810. https://doi.org/10.1016/j.heliyon.2020.e05810.

Pontén, F., Jirström, K., and Uhlen, M. (2008). The human protein atlas - a tool for pathology. J. Pathol. *216*, 387–393. https://doi.org/10.1002/path.2440.

Raajendiran, A., Ooi, G., Bayliss, J., O'Brien, P.E., Schittenhelm, R.B., Clark, A.K., Taylor, R.A., Rodeheffer, M.S., Burton, P.R., and Watt, M.J. (2019). Identification of metabolically distinct adipocyte progenitor cells in human adipose

tissues. Cell Rep. *27*, 1528–1540.e7. https://doi.org/10.1016/j.celrep.2019.04.010.

Ritter, A., Friemel, A., Kreis, N.N., Hoock, S.C., Roth, S., Kielland-Kaisen, U., Brüggmann, D., Solbach, C., Louwen, F., and Yuan, J. (2018). Primary cilia are dysfunctional in obese adipose-derived mesenchymal stem cells. Stem Cell Rep. *10*, 583–599. https://doi.org/10.1016/j.stemcr.2017.12.022.

Rizzetto, S., Eltahla, A.A., Lin, P., Bull, R., Lloyd, A.R., Ho, J.W.K., Venturi, V., and Luciani, F. (2017). Impact of sequencing depth and read length on single cell RNA sequencing data of T cells. Sci. Rep. Uk *7*, 12781. https://doi.org/10.1038/s41598-017-12989-x.

Roh, H.C., Tsai, L.T.Y., Lyubetskaya, A., Tenen, D., Kumari, M., and Rosen, E.D. (2017). Simultaneous transcriptional and epigenomic profiling from specific cell types within heterogeneous tissues *in vivo*. Cell Rep. *18*, 1048–1061. https://doi.org/10.1016/j.celrep.2016.12.087.

Rondini, E.A., and Granneman, J.G. (2020). Single cell approaches to address adipose tissue stromal cell heterogeneity. Biochem. J. *477*, 583–600. https://doi.org/10.1042/bcj20190467.

Rotondo, F., Ho-Palma, A.C., Remesar, X., Fernández-López, J.A., Romero, M.D.M., and Alemany, M. (2017). Glycerol is synthesized and secreted by adipocytes to dispose of excess glucose, via glycerogenesis and increased acyl-glycerol turnover. Sci. Rep. *7*, 8983. https://doi.org/10.1038/s41598-017-09450-4.

Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The Harmonizome: A Collection of Processed Datasets Gathered to Serve and Mine Knowledge about Genes and Proteins (Database).

Sadkowski, T., Ciecierska, A., Majewska, A., Oprządek, J., Dasiewicz, K., Ollik, M., Wicik, Z., and Motyl, T. (2014). Transcriptional background of beef marbling - novel genes implicated in intramuscular fat deposition. Meat Sci. *97*, 32–41. https://doi.org/10.1016/j.meatsci.2013.12.017.

Sakurai, T. (2012). The role of NrCAM in neural development and disorders–beyond a simple glue in the brain. Mol. Cell. Neurosci. *49*, 351–363. https://doi.org/10.1016/j.mcn.2011.12.002.

Saliba, A.-E., Westermann, A.J., Gorski, S.A., and Vogel, J. (2014). Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res. *42*, 8845–8860. https://doi.org/10.1093/nar/gku555.

Schleinitz, D., Krause, K., Wohland, T., Gebhardt, C., Linder, N., Stumvoll, M., Blüher, M., Bechmann, I., Kovacs, P., Gericke, M., and Tönjes, A. (2020). Identification of distinct transcriptome signatures of human adipose tissue from fifteen depots. Eur. J. Hum. Genet. *28*, 1714–1725. https://doi.org/10.1038/s41431-020-0681-1.

Sichien, D., Lambrecht, B.N., Guilliams, M., and Scott, C.L. (2017). Development of conventional dendritic cells: from common bone marrow progenitors to multiple subsets in peripheral tissues. Mucosal Immunol. *10*, 831–844. https://doi.org/10.1038/mi.2017.8.

Singh, R.K., Singh, D., Yadava, A., and Srivastava, A.K. (2020). Molecular fossils "pseudogenes" as functional signature in biological system. Genes Genomics *42*, 619–630. https://doi.org/10.1007/s13258-020-00935-7.

Squillaro, T., Peluso, G., Galderisi, U., and Di Bernardo, G. (2020). Long noncoding RNAs in regulation of adipogenesis and adipose tissue function. Elife *9*, e59053. https://doi.org/10.7554/elife.59053.

Statello, L., Guo, C.J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol., 96–118. https://doi.org/10.1038/s41580-020-00315-9.

Stuczyńska, A., Piórkowska, K., Tyra, M., and Zukowski, K. (2018). The effect of QTL-rich region polymorphisms identified by targeted DNA-seq on pig production traits. Mol. Biol. Rep. *45*, 361–371. https://doi.org/10.1007/s11033-018-4170-3.

Sun, T., Fu, M., Bookout, A.L., Kliewer, S.A., and Mangelsdorf, D.J. (2009). MicroRNA let-7 regulates 3T3-L1 adipogenesis. Mol. Endocrinol. *23*, 925–931. https://doi.org/10.1210/me.2008-0298.

Sun, W., Dong, H., Balaz, M., Slyper, M., Drokhlyansky, E., Colleluori, G., Giordano, A., Kovanicova, Z., Stefanicka, P., Balazova, L., et al. (2020). snRNA-seq

reveals a subpopulation of adipocytes that regulates thermogenesis. Nature *587*, 98–102. https://doi.org/10.1038/s41586-020-2856-x.

Sun, W., von Meyenn, F., Peleg-Raibstein, D., and Wolfrum, C. (2019). Environmental and nutritional effects regulating adipose tissue function and metabolism across generations. Adv. Sci. *6*, 1900275. https://doi.org/10.1002/advs.201900275.

Svoboda, M., Portois, L., and Malaisse, W.J. (1999). Glucose regulation of the expression of the glucagon receptor gene. Mol. Genet. Metab. *68*, 258–267. https://doi.org/10.1006/mgme.1999.2913.

Tabula Muris Consortium; Overall coordination; Logistical coordination; Organ collection and processing; Library preparation and sequencing; Computational data analysis; Cell type annotation; Writing group; Supplemental text writing group; Principal investigators (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367–372. https://doi.org/10.1038/s41586-018-0590-4.

Tabula Sapiens Consortium; Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., Hemenez, M., et al. (2022). The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. https://doi.org/10.1126/science.abl4896.

Takeda, A., Hollmén, M., Dermadi, D., Pan, J., Brulois, K.F., Kaukonen, R., Lönnberg, T., Boström, P., Koskivuo, I., Irjala, H., et al. (2019). Single-cell survey of human lymphatics unveils marked endothelial cell heterogeneity and mechanisms of homing for neutrophils. Immunity *51*, 561–572.e5. https://doi.org/10.1016/j.immuni.2019.06.027.

Thrupp, N., Sala Frigerio, C., Wolfs, L., Skene, N.G., Fattorelli, N., Poovathingal, S., Fourne, Y., Matthews, P.M., Theys, T., Mancuso, R., et al. (2020). Single-nucleus RNA-seq is not suitable for detection of microglial activation genes in humans. Cell Rep. *32*, 108189. https://doi.org/10.1016/j.celrep.2020.108189.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science *347*, 1260419. https://doi.org/10.1126/science.1260419.

Uhlen, M., Karlsson, M.J., Zhong, W., Tebani, A., Pou, C., Mikes, J., Lakshmikanth, T., Forsström, B., Edfors, F., Odeberg, J., et al. (2019). A genome-wide transcriptomic analysis of protein-coding genes in human blood cells. Science *366*, eaax9198. https://doi.org/10.1126/science.aax9198.

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., Benfeitas, R., Arif, M., Liu, Z., Edfors, F., et al. (2017). A pathology atlas of the human cancer transcriptome. Science *357*, eaan2507. https://doi.org/10.1126/science.aan2507.

Valencak, T.G., Osterrieder, A., and Schulz, T.J. (2017). Sex matters: the effects of biological sex on adipose tissue biology and energy metabolism. Redox Biol. *12*, 806–813. https://doi.org/10.1016/j.redox.2017.04.012.

Vijay, J., Gauthier, M.F., Biswell, R.L., Louiselle, D.A., Johnston, J.J., Cheung, W.A., Belden, B., Pramatarova, A., Biertho, L., Gibson, M., et al. (2020). Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. Nat. Metab. *2*, 97–109. https://doi.org/10.1038/s42255-019-0152-6.

Viswanadha, S., and Londos, C. (2006). Optimized conditions for measuring lipolysis in murine primary adipocytes. J. Lipid Res. *47*, 1859–1864. https://doi.org/10.1194/jlr.d600005-jlr200.

Walley, A.J., Jacobson, P., Falchi, M., Bottolo, L., Andersson, J.C., Petretto, E., Bonnefond, A., Vaillant, E., Lecoeur, C., Vatin, V., et al. (2012). Differential coexpression analysis of obesity-associated networks in human subcutaneous adipose tissue. Int. J. Obes. *36*, 137–147. https://doi.org/10.1038/ijo.2011.22.

Wang, G., Zhang, X., Lee, J.S., Wang, X., Yang, Z.Q., and Zhang, K. (2012). Endoplasmic reticulum factor ERLIN2 regulates cytosolic lipid content in cancer cells. Biochem. J. *446*, 415–425. https://doi.org/10.1042/bj20112050.

Wang, S., Pisco, A.O., McGeever, A., Brbic, M., Zitnik, M., Darmanis, S., Leskovec, J., Karkanias, J., and Altman, R.B. (2021). Leveraging the Cell Ontology to classify unseen cell types. Nat. Commun. *12*, 5556. https://doi.org/10.1038/s41467-021-25725-x.

Wu, X., Niculite, C.M., Preda, M.B., Rossi, A., Tebaldi, T., Butoi, E., White, M.K., Tudoran, O.M., Petrusca, D.N., Jannasch, A.S., et al. (2020). Regulation of cellular sterol homeostasis by the oxygen responsive noncoding RNA linc-NORS. Nat. Commun. *11*, 4755. https://doi.org/10.1038/s41467-020-18411-x.

Xu, D., and Sun, L. (2020). A functional non-conserved long non-coding RNA in human adipose tissue. Nat. Metab. *2*, 385–386. https://doi.org/10.1038/s42255-020-0208-7.

Yang, H., Qu, H., Huang, H., Mu, Z., Mao, M., Xie, Q., Wang, K., and Hu, B. (2021). Exosomes-mediated transfer of long noncoding RNA LINC01133 represses bladder cancer progression via regulating the Wnt signaling pathway. Cell Biol. Int. *45*, 1510–1522. https://doi.org/10.1002/cbin.11590.

Zhang, X., Lan, Y., Xu, J., Quan, F., Zhao, E., Deng, C., Luo, T., Xu, L., Liao, G., Yan, M., et al. (2019). CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res. *47*, D721–D728. https://doi.org/10.1093/nar/gky900.

Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. Mol. Cell *65*, 631–643.e4. https://doi.org/10.1016/j.molcel.2017.01.023.

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
scRNA murine adipose tissue data	Tabula Muris et al. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367-372.	https://www.nature.com/articles/s41586-018-0590-4 (https://doi.org/10.1038/s41586-018-0590-4) PMID: **30283141**
snRNAseq human subcutaneous adipose tissue	Sun et al. (2020). snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis. Nature *587*, 98-102.	https://www.nature.com/articles/s41586-020-2856-x (https://doi.org/10.1038/s41586-020-2856-x) PMID: **33116305**
Website resource for protein coding gene enrichment	This paper	Human Protein Atlas https://www.proteinatlas.org/humanproteome/tissue+cell+type/adipose+tissue
Website resource for non-protein coding gene enrichment	This paper	https://cell-enrichment.shinyapps.io/noncoding/

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact: Dr. Lynn Marie Butler. Email: Lynn.butler@ki.se.

Materials availability
This study did not generate new unique reagents.

Data and code availability
- This paper analyses existing, publicly available data. The accession number for the datasets are listed in the key resources table.
- All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bulk RNA-seq data analyzed in this study was obtained from the Genotype-Tissue Expression (GTEx) Project (gtexportal.org) (Consortium, 2015) accessed on 2019.11.29 (dbGaP Accession phs000424.v8.p2). Sample IDs of visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) samples used in the analysis can be found in Table S1. Human tissue protein profiling was performed in house as part of the Human Protein Atlas (HPA) project (Ponten et al., 2008; Uhlen et al., 2015, 2017) (www.proteinatlas.org). Adipose tissue samples were obtained from the Department of Pathology, Uppsala University Hospital, Uppsala, Sweden, as part of the Uppsala Biobank. Samples were handled in accordance with Swedish laws and regulations, with approval from the Uppsala Ethical Review Board (Uhlen et al., 2015).

METHOD DETAILS

Tissue profiling: human tissue sections
Adipose tissue sections were stained, as previously described (Ponten et al., 2008; Uhlen et al., 2015). Briefly, formalin fixed and paraffin embedded tissue samples were sectioned, de-paraffinized in xylene, hydrated in graded alcohols and blocked for endogenous peroxidase in 0.3% hydrogen peroxide diluted in 95% ethanol. For antigen retrieval, a Decloaking chamber® (Biocare Medical, CA) was used. Slides were boiled in Citrate buffer®, pH6 (Lab Vision, CA). Primary antibodies and a dextran polymer visualization system (UltraVision LP HRP polymer®, Lab Vision) were incubated for 30 min each at room temperature and slides were developed for 10 min using Diaminobenzidine (Lab Vision) as the chromogen. Slides were counterstained in Mayers hematoxylin (Histolab) and scanned using Scanscope XT (Aperio). Primary antibodies, source, target and identifier are as follows: Atlas Antibodies: ACSL1 (Cat#HPA011316; RRID:AB_1844536), ACO1 (Cat#HPA019371; RRID:AB_1844519), FBXO27 (Cat#HPA046800; RRID:AB_2679813), MYH9 (Cat#HPA064783; RRID:AB_2732721), GIMAP4 (Cat#HPA019135; RRID:AB_1849670), FLNB

(Cat#HPA004886; RRID:AB_1848600), PLN (Cat#HPA026900; RRID:AB_1855314), LMOD1 (Cat#HPA028435; RRID:AB_10602180), DES (Cat#HPA018803; RRID:AB_1847616), TBXAS1 (Cat#HPA031257; RRID:AB_2673812), ITGB2 (Cat#HPA016894; RRID:AB_1846257), PRKAR2B (Cat#HPA008421, RRID:AB_1855421), C19orf12 (Cat#HPA046930, RRID:AB_10962836), SHANK3 (Cat#HPA003446; RRID:AB_1079958), CASQ2 (Cat#HPA027285; RRID:AB_1845933), SLC30A3 (Cat#HPA060505; RRID:AB_2684296), LCP1 (Cat#HPA019493; RRID:AB_1855457), IFI30 (Cat#HPA026650; RRID:AB_10602237), SP140 (Cat#HPA006162; RRID:AB_1857403), CD247 (Cat#HPA008750; RRID:AB_1857863). Santa Cruz: TYROBP (Cat#sc-20783; RRID:AB_638987), TBX21 (Cat#sc-21003; RRID:AB_2200557), Thermo Fisher Scientific: ZAP70 (Cat#MS-1911), Merck: KLRK1 (Cat#05-945;), R&D Systems: CDH13 (Cat#MAB3264). All IHC images are available on the HPA website (https://www.proteinatlas.org/).

QUANTIFICATION AND STATISTICAL ANALYSIS

Reference transcript-based correlation analysis

This method was adapted and expanded from that previously developed to determine the cross-tissue pan-EC-enriched transcriptome (Butler et al., 2016) and human brain cell enriched genes (Dusart et al., 2019). Pairwise Spearman correlation coefficients were calculated between reference transcripts selected as proxy markers for: adipocytes [*ADIPOQ*, *LIPE*, *PLIN1*], adipocyte progenitor cells [*FKBP10*, *COL6A1*, *COL6A2*], mesothelial cells [*UPK3B*, *MSLN*, *KRT19*], endothelial cells [*MMRN2*, *ESAM*, *CDH5*], smooth muscle cells [*KCNMB1*, *CNN1*, *MYH11*], macrophages [*CD68*, *C1QC*, *FCER1G*], neutrophils [*CSF3R*, *FCGR3B*, *CXCR2*], mast cells [*CPA3*, *TPSB2*, *TPSAB1*], T cells [*TRBC2*, *CD6*, *CD3E*] and plasma cells [*IGKC*, *JCHAIN*, *MZB1*] and all other sequenced transcripts. Transcripts with a TPM value < 0.1 in more than 50% of samples were excluded from analysis (but are still included in data tables). See results section for full criteria required for transcript classification of transcripts as cell-type enriched (also Table S1, tab 1, table B). Correlation coefficients were calculated in R using the *corr.test* function from the *psych* package (v 1.8.4). In addition to correlation coefficients False Discovery Rate (FDR) adjusted p-values (using Bonferroni correction) and raw p-values were calculated. FDR <0.0001 for correlation was required for inclusion as cell type enriched, but no transcripts in either VAT or SAT required exclusion due to this criterion.

Weighted correlation network (WGCNA) analysis

The R package WGCNA (Langfelder and Horvath, 2008) was used to perform co-expression network analysis for gene clustering, on log2 expression TPM values. The analysis was performed according to recommendations in the WGCNA manual. Transcripts with too many missing values were excluded using the goodSamplesGenes() function. The remaining genes were used to cluster the samples, and obvious outlier samples were excluded.

Gene ontology and reactome analysis

The Gene Ontology Consortium (Ashburner et al., 2000) and PANTHER classification resource (Mi et al., 2013, 2016) were used to identify over represented terms (biological processes) in the panel of identified cell-type-enriched transcripts from the GO ontology (release date 2021-10-09) or reactome (release date 2021-11-17) databases.

Processing of data from adipose tissue scRNA-seq and snRNA-seq datasets

Data from scRNA-seq analysis of human SAT (Hildreth et al., 2021; Tabula Sapiens et al., 2022), scRNA-seq of murine adipose tissue (mixed depot) (Tabula Muris et al., 2018) and snRNA-seq of human SAT (Sun et al., 2020) was downloaded or received from the authors upon request. Cell type clustering and categorization was performed as originally described, but immune cell subtypes in (Hildreth et al., 2021) were merged, and myofibroblasts and smooth muscle cells in (Tabula Sapiens et al., 2022) were handled together. The R Seurat package (Hao et al., 2021) and the FindAllMarkers function was used to determine the Log2 fold change values for each gene in all cell types versus all others within each study, and to generate illustrative UMAP plots when required. The statistical significance of overlap between cell-type enriched genes in each study was calculated using a hypergeometric test (Figure S5). Criteria used for comparison of our cell type-enriched datasets with expression profiles in the independent studies are given in the relevant results sections and associated tables or figure legends.

Visualization

Circular graphs (Figures 3, 4B, and 6E) were constructed using the R package *circlize* (Gu et al., 2014). Some figure sections were created with BioRender.com.

ADDITIONAL RESOURCES

Analyzed data for all protein coding genes is provided on the Human Protein Atlas website: (www.proteinatlas.org/humanproteome/tissue+cell+type/adipose+tissue). Analyzed data for non-coding transcripts is provided on: https://cell-enrichment.shinyapps.io/noncoding/. The published article includes all datasets generated during this study, including depot- and sex-subset analysis (Tables S1, S2, and S3).
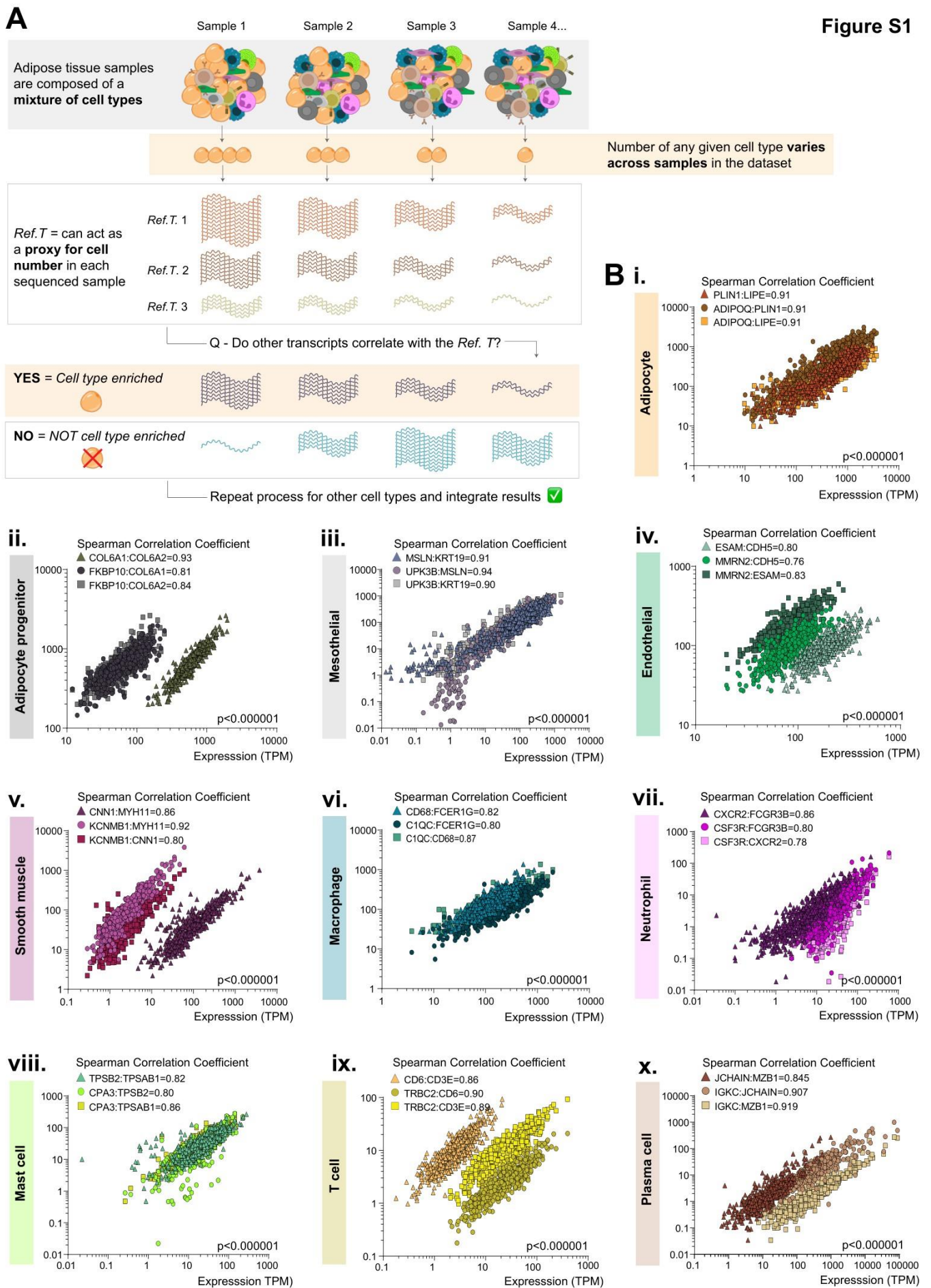
Figure S1

A

Adipose tissue samples are composed of a **mixture of cell types**

Sample 1 Sample 2 Sample 3 Sample 4...

Number of any given cell type **varies across samples** in the dataset

Ref.T = can act as a **proxy for cell number** in each sequenced sample

Ref.T. 1

Ref.T. 2

Ref.T. 3

Q - Do other transcripts correlate with the *Ref. T*?

YES = *Cell type enriched*

NO = *NOT cell type enriched*

Repeat process for other cell types and integrate results ✅

B i.

Adipocyte

Spearman Correlation Coefficient
PLIN1:LIPE=0.91
ADIPOQ:PLIN1=0.91
ADIPOQ:LIPE=0.91

p<0.000001

Expresssion (TPM)

ii.

Adipocyte progenitor

Spearman Correlation Coefficient
COL6A1:COL6A2=0.93
FKBP10:COL6A1=0.81
FKBP10:COL6A2=0.84

p<0.000001

Expresssion (TPM)

iii.

Mesothelial

Spearman Correlation Coefficient
MSLN:KRT19=0.91
UPK3B:MSLN=0.94
UPK3B:KRT19=0.90

p<0.000001

Expresssion (TPM)

iv.

Endothelial

Spearman Correlation Coefficient
ESAM:CDH5=0.80
MMRN2:CDH5=0.76
MMRN2:ESAM=0.83

p<0.000001

Expresssion (TPM)

v.

Smooth muscle

Spearman Correlation Coefficient
CNN1:MYH11=0.86
KCNMB1:MYH11=0.92
KCNMB1:CNN1=0.80

p<0.000001

Expresssion (TPM)

vi.

Macrophage

Spearman Correlation Coefficient
CD68:FCER1G=0.82
C1QC:FCER1G=0.80
C1QC:CD68=0.87

p<0.000001

Expresssion (TPM)

vii.

Neutrophil

Spearman Correlation Coefficient
CXCR2:FCGR3B=0.86
CSF3R:FCGR3B=0.80
CSF3R:CXCR2=0.78

p<0.000001

Expresssion (TPM)

viii.

Mast cell

Spearman Correlation Coefficient
TPSB2:TPSAB1=0.82
CPA3:TPSB2=0.80
CPA3:TPSAB1=0.86

p<0.000001

Expresssion TPM

ix.

T cell

Spearman Correlation Coefficient
CD6:CD3E=0.86
TRBC2:CD6=0.90
TRBC2:CD3E=0.89

p<0.000001

Expresssion (TPM)

x.

Plasma cell

Spearman Correlation Coefficient
JCHAIN:MZB1=0.845
IGKC:JCHAIN=0.907
IGKC:MZB1=0.919

p<0.000001

Expresssion (TPM)

Figure S1. Methodological summary and expression distribution and correlations between human visceral adipose tissue (VAT) cell type reference transcripts. Related to Figure 1 and Table S1, Tab 1. (**A**) Schematic of analysis concept. (**B**) Expression of *Ref.T* selected to represent: (**i**) adipocytes, (**ii**) adipocyte progenitors, (**iii**) mesothelial cells, (**iv**) endothelial cells, (**v**) smooth muscle cells, (**vi**) macrophages, (**vii**) neutrophils, (**viii**) mast cells, (**ix**) T-cells and (**x**) plasma cells.

Figure S2

A.i

A.ii

B.i scRNAseq data (Tabula Sapiens)
- Macrophage (MC)
- Neutrophil (NP)
- T cell
- Myo-fb/smooth muscle
- Endothelial
- Fibroblast
- Plasma cell
- B cells
- Leucocyte
- Mast cell
- Natural killer

B.ii

Predicted: Macrophage (MC) enriched

Predicted: MC and NP enriched

Predicted: Neutrophil (NP) enriched

174
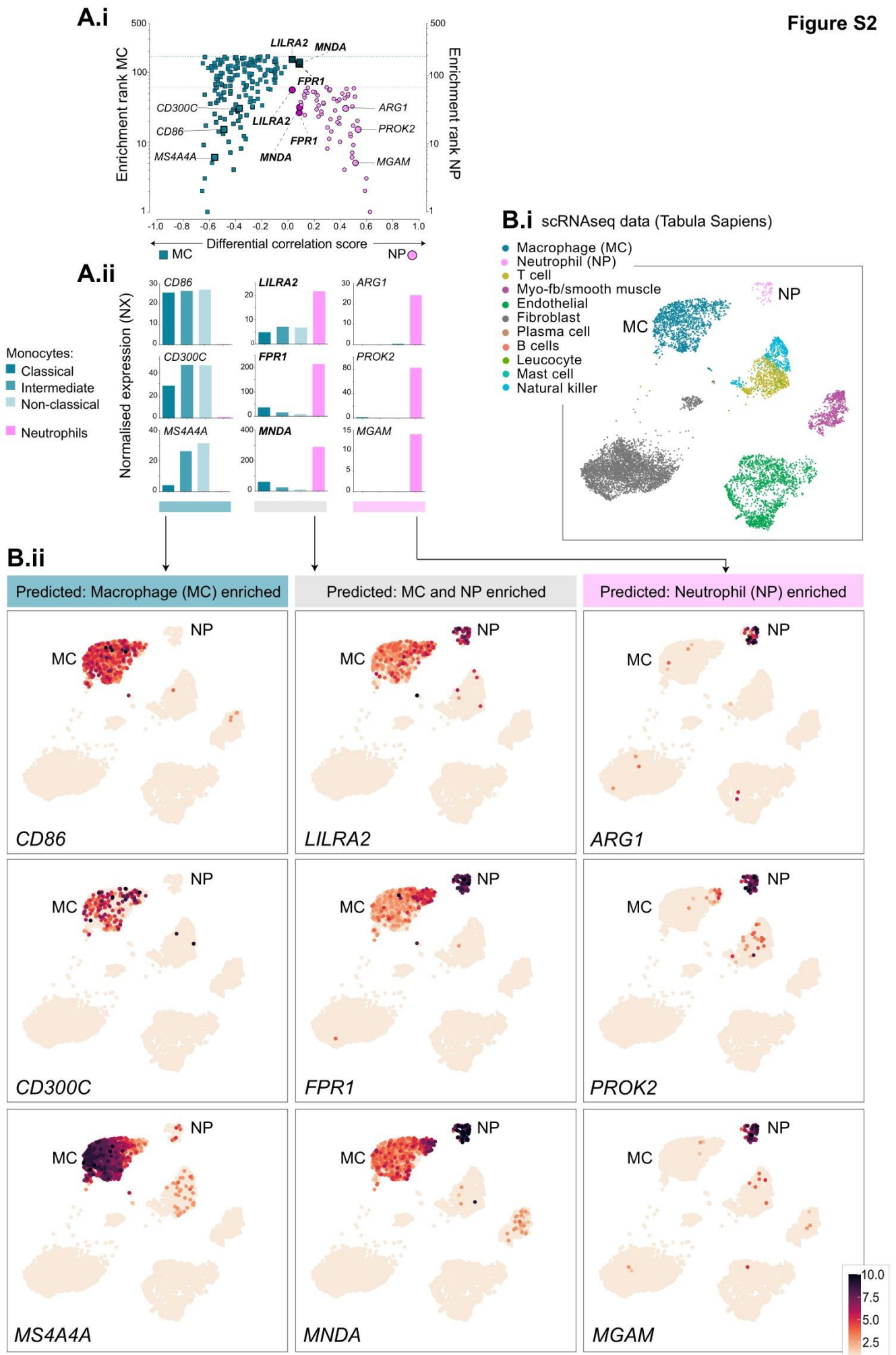
Figure S2. scRNAseq of subcutaneous adipose tissue provides supportive evidence for cell type enrichment predictions from integrative co-expression analysis of unfractionated visceral adipose tissue (VAT). Related to Figure 1D. (**A**) (**i**) For transcripts above the designated correlation threshold with the macrophage (squares, MC) or neutrophil (circles, NP) *Ref.T.* panels, the '*differential correlation score*' (difference between mean corr. with MC and NP *Ref.T.*) was plotted vs. 'enrichment ranking'. Bold text annotations show transcripts appearing in *both* MC- and NP lists (circular *and* square symbol, on the same X-axis dimension). (**ii**) scRNAseq data from the Human Protein Blood Atlas (Uhlen et al., 2019) showing gene expression in classical, intermediate, and non-classical monocytes, and neutrophils from whole blood. (**B**) scRNAseq data from analysis of cell types in human subcutaneous adipose tissue was sourced from Tabula Sapiens (Tabula Sapiens., 2021), and used to generate UMAP plots showing (**i**) scRNAseq cell type annotations, and (**ii**) expression profiles of genes we predicted as macrophage (MC)-enriched [*CD86*, *CD300C*, *MS4A4A*] (blue bar), co-enriched in both MC and neutrophils (NP) [*LILRA2*, *FPR1*, *MNDA*] (grey bar) or predominantly NP-enriched [*ARG1*, *PROK2*, *MGAM*] (pink bar).
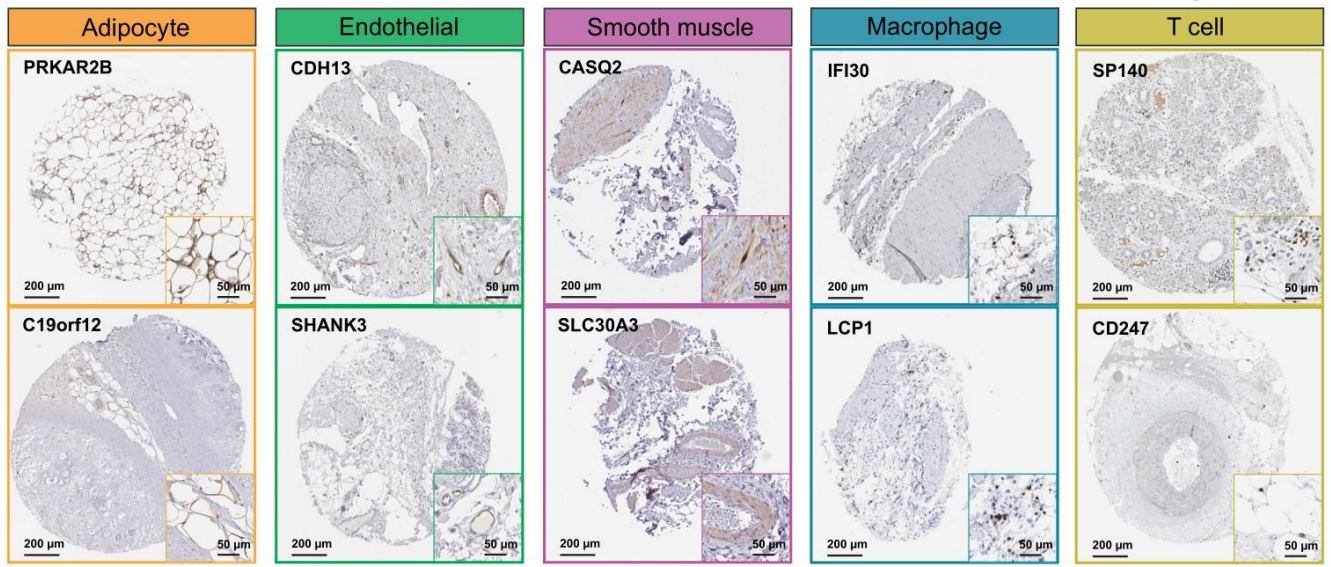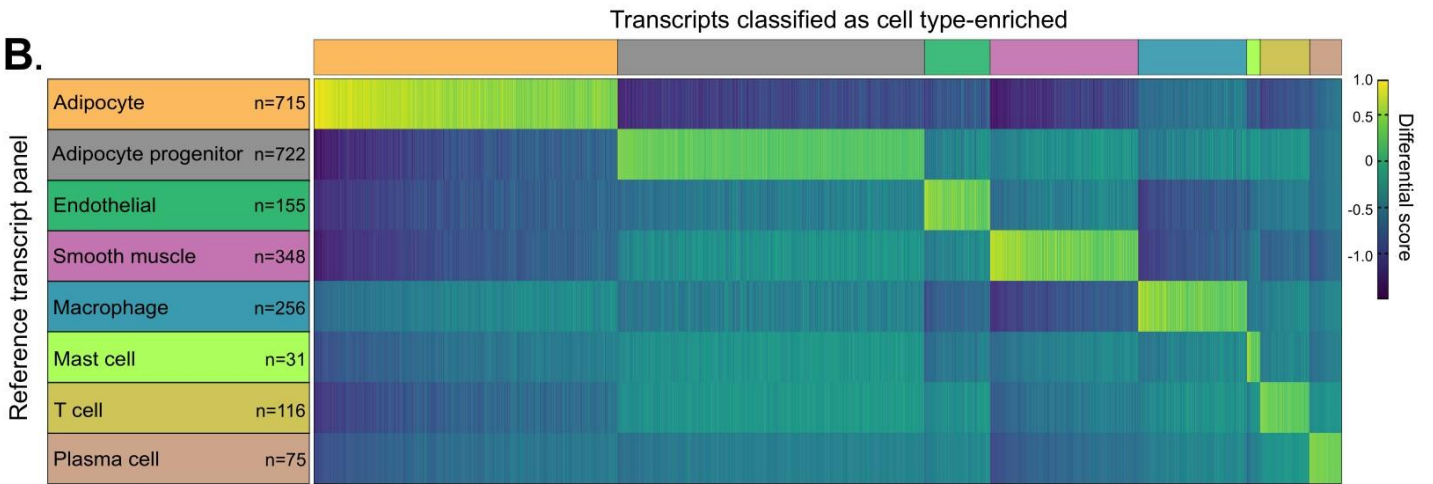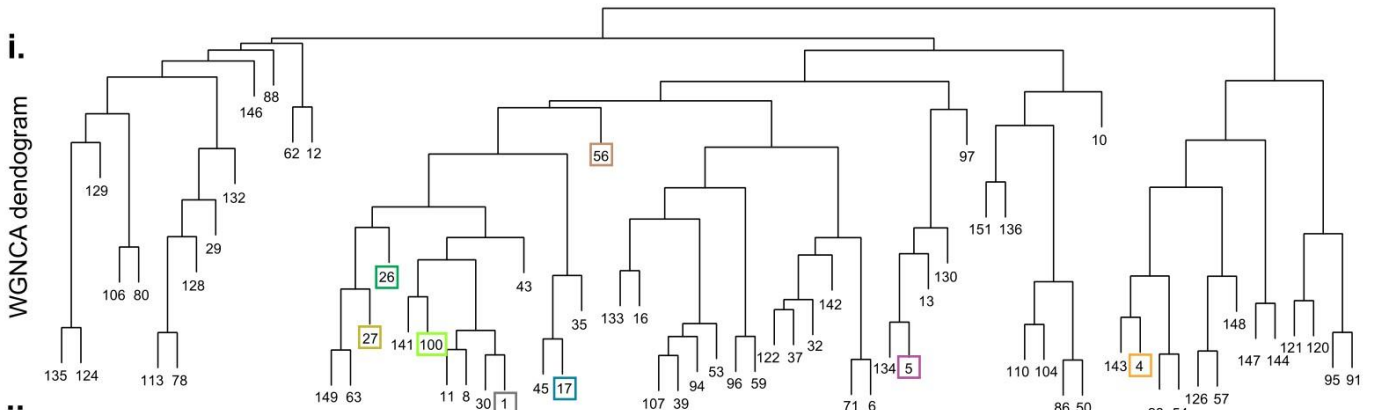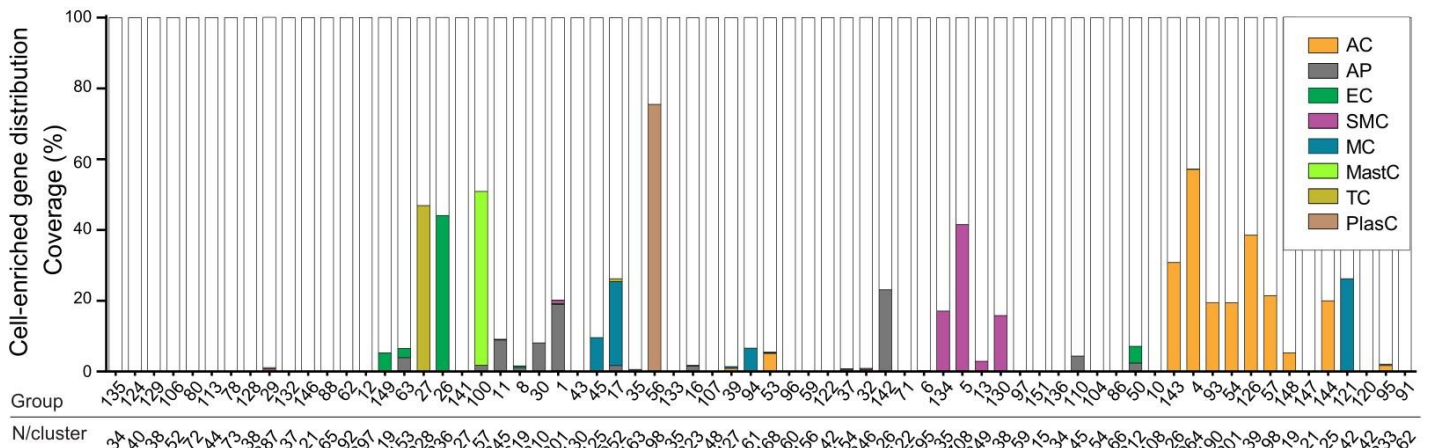
Figure S3

Figure S3. (A) Protein profiling of transcripts identified as cell-enriched in VAT. Related to Figure 2. Human adipose tissue sections were stained using primary antibodies targeting proteins encoded by transcripts classified as adipocyte-, endothelial-, smooth muscle-, macrophage- or T-cell-enriched. Scale bar 200μm, inset 50μm. See also Table S2, Tab 1.

(B-C) Integrative co-expression analysis of unfractionated RNAseq reveals enriched transcriptomes of subcutaneous adipose tissue (SAT) cell types. Related to Figure 6, Figure S7, Figure S8. Human SAT RNAseq data (n=646), retrieved from Genotype-Tissue Expression (GTEx) portal V8, was used to determine correlation coefficients (corr.) between selected adipose cell type *Ref.T* and all other sequenced transcripts. (**A**) Heat map plot of transcripts classified as cell type-enriched (indicated by horizontal-coloured bars), showing differential score between mean correlation coefficient with the corresponding *Ref.T.* panel vs. highest mean correlation coefficient amongst the other *Ref.T.* panels. (**B**) SAT RNAseq data was subject to weighted correlation network analysis (WGCNA). (**i**) Coloured squares indicate *Ref.T.* location on resultant dendrogram. (**ii**) Distribution of transcripts classified as cell type-enriched across dendrogram groups. See also Table S2, Tab 1.

Figure S4

A
i. scRNAseq SAT [**TS**]
FB · EC · TC

ii. snRNAseq SAT [**Sun**]
Pre-AC · EC · CD4+ TC

iii. scRNAseq SAT [**Hildreth**]
AP/Pre-AC · EC · TC

iv. scRNAseq mixed [**TM**]
MSC · EC · TC

Genes logFC >0.2 in at least 1 dataset
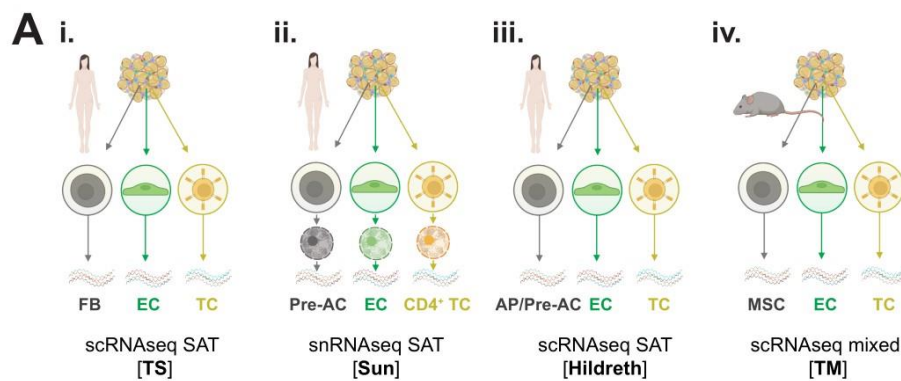
B
i. Adipocyte progenitor (AP)
Total genes = 186
Avg log2FC:
>1.0
>0.5
>0.2
<0.2
Not detected/in dataset
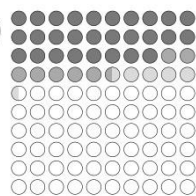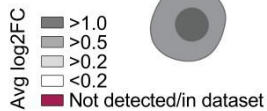
ii. Endothelial cell (EC)
Total genes = 93
Avg log2FC:
>1.0
>0.5
>0.2
<0.2
Not detected/in dataset
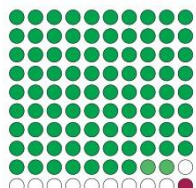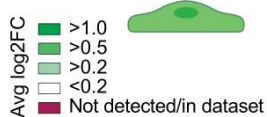
iii. T-cell (TC)
Total genes = 98
Avg log2FC:
>1.0
>0.5
>0.2
<0.2
Not detected/in dataset
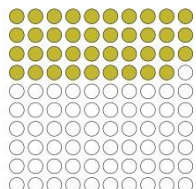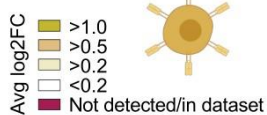
C
i.
GTEx Portal (29 tissues)
THE HUMAN PROTEIN ATLAS (31 tissues)
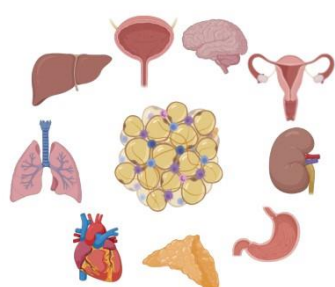
Top 200 → **86** common **adipose-enriched** genes ← Top 200
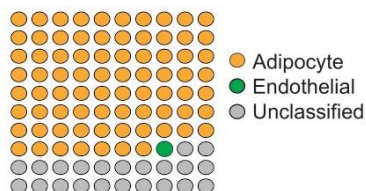
Adipose enriched genes identified from bulk RNAseq comparison

Adipose enriched genes identified from bulk RNAseq comparison

ii. *Cell type classification*
Adipocyte
Endothelial
Unclassified

178

Figure S4. Comparison of predicted human visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) cell type enriched transcriptomes with scRNAseq or snRNAseq of human SAT or murine adipose tissue. Related to Figure 3, Table S1 Tab 5, and Table S2, Tab 1. (A) Data generated by single cell (scRNAseq) or single nuclear (snRNAseq) profiling of human SAT or murine adipose tissue was sourced from (**i**) Tabula Sapiens (Tabula Sapiens., 2021) (scRNAseq SAT [TS]), (**ii**) Sun *et al.* (Sun et al., 2020) (snRNAseq SAT [Sun]), (**iii**) Hildreth *et al.* (Hildreth et al., 2021) (ssRNAseq SAT [Hildreth]) and (**iv**) Tabula Muris (Tabula Muris et al., 2018) (scRNAseq mixed [TM]). (**B**) Genes predicted as enriched in (**i**) adipocyte progenitor cells, (**ii**) endothelial cells or (**iii**) T-cells, in both VAT and SAT, were cross checked with the independent studies. Colour coding indicates proportion of genes that have average Log2 fold change >1.0, >0.5 or >0.2 [p<0.01] in the corresponding cell type vs. all other cell types profiled in the independent study. **FB**: fibroblast, **AP**: adipocyte progenitor, **Pre-AC**: pre-adipocyte, **MSC**: mesenchymal stem cell, **EC**: endothelial cell, **TC**: T-cell. (**C.i**) The top 200 human adipose enriched genes in Human Protein Atlas and GTEx datasets were sourced from Harminozome database (Rouillard et al., 2016) and (**C.ii**) classification as cell type-enriched in our analysis of VAT determined.
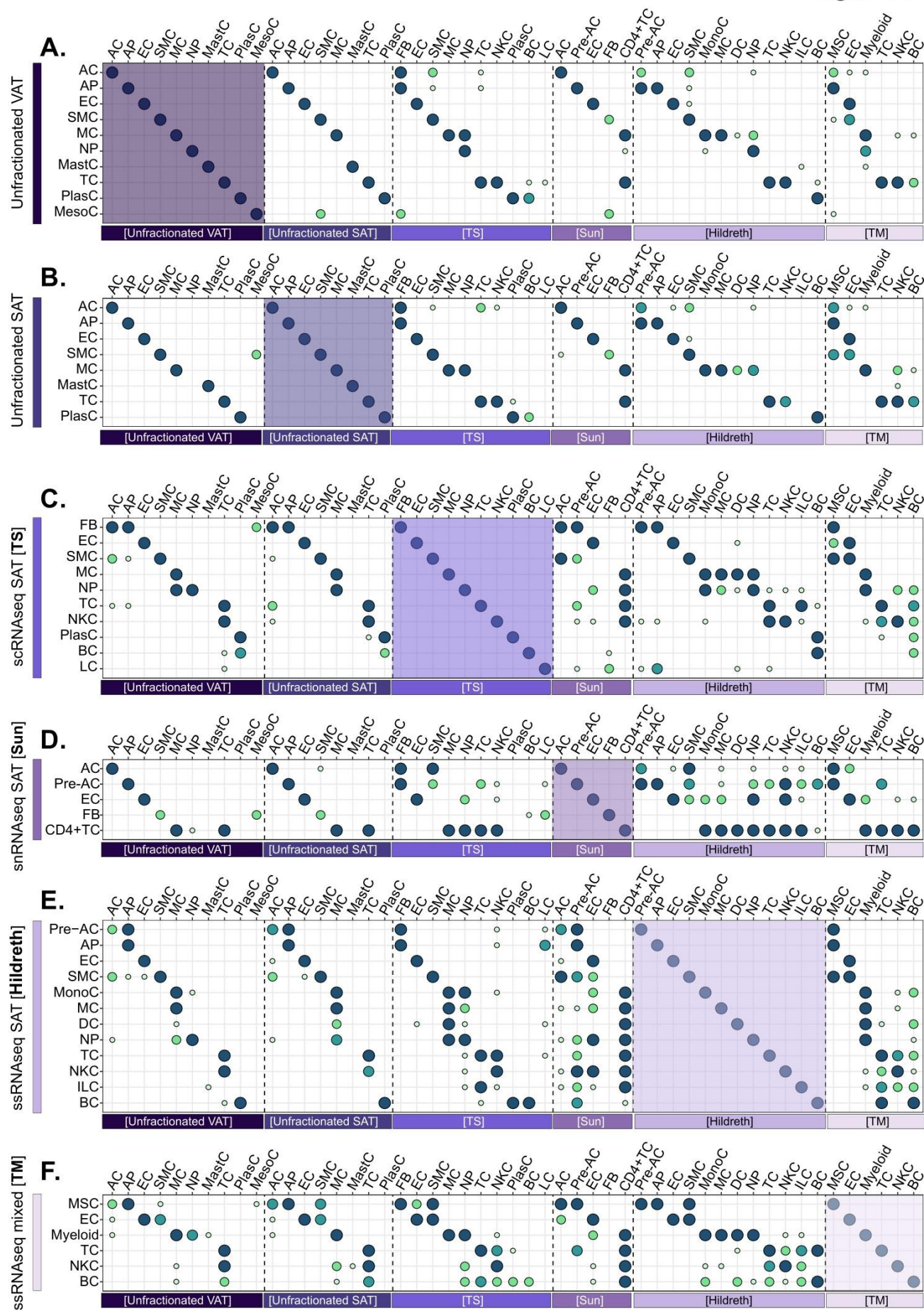
Figure S5. A comparison of cell-type enriched genes identified in different adipose tissue cell type profiling studies; Related to Figure 3 and S4. Bubble heatmap showing the significance (indicated by dot size and colour) of shared enriched genes between adipose tissue cell types, as identified in the current study using integrative correlation analysis of bulk RNAseq of (**A**) human visceral adipose tissue [VAT] (Unfractionated VAT) and (**B**) human subcutaneous adipose tissue [SAT] (Unfractionated SAT), or by single cell/single nuclear profiling of human SAT, sourced from (**C**) Tabula Sapiens (Tabula Sapiens., 2021) (scRNAseq SAT [TS]) (**D**) Sun *et al.* (Sun et al., 2020) (snRNAseq SAT [Sun]) and (**E**) Hildreth et al. (Hildreth et al., 2021) (ssRNAseq SAT [Hildreth]), or in (**F**) murine adipose tissue from Tabula Muris (Tabula Muris et al., 2018) (scRNAseq mixed [TM]) (enriched genes defined as those ≥0.5 Log2 fold change in expression vs. all other cell types in the same study [p<0.01]). Cell type-enriched genes were compared across all studies (indicated by different coloured blocks on x-axis), N.B. not all cell types were represented in every study. When overlap of enriched genes was not statistically significant (hypergeometric test, $P > 0.05$), no dot is displayed. **AC**: adipocyte, **AP**: adipocyte progenitor, **EC**: endothelial cell, **SMC**: smooth muscle cell, **MC**: macrophage, **NP**: neutrophil, **MastC**: mast cell, **TC**: T-cell, **PlasC**: plasma cell, **MesoC**: mesothelial cell, **FB**: fibroblast, **NKC**: natural killer cell, **BC**: B-cell, **LC**: leukocyte, **Pre-AC**: pre-adipocyte, **MonoC**: monocyte, **DC**: dendritic cell, **ILC**: innate lymphoid cell, **MSC**: mesenchymal stem cell.
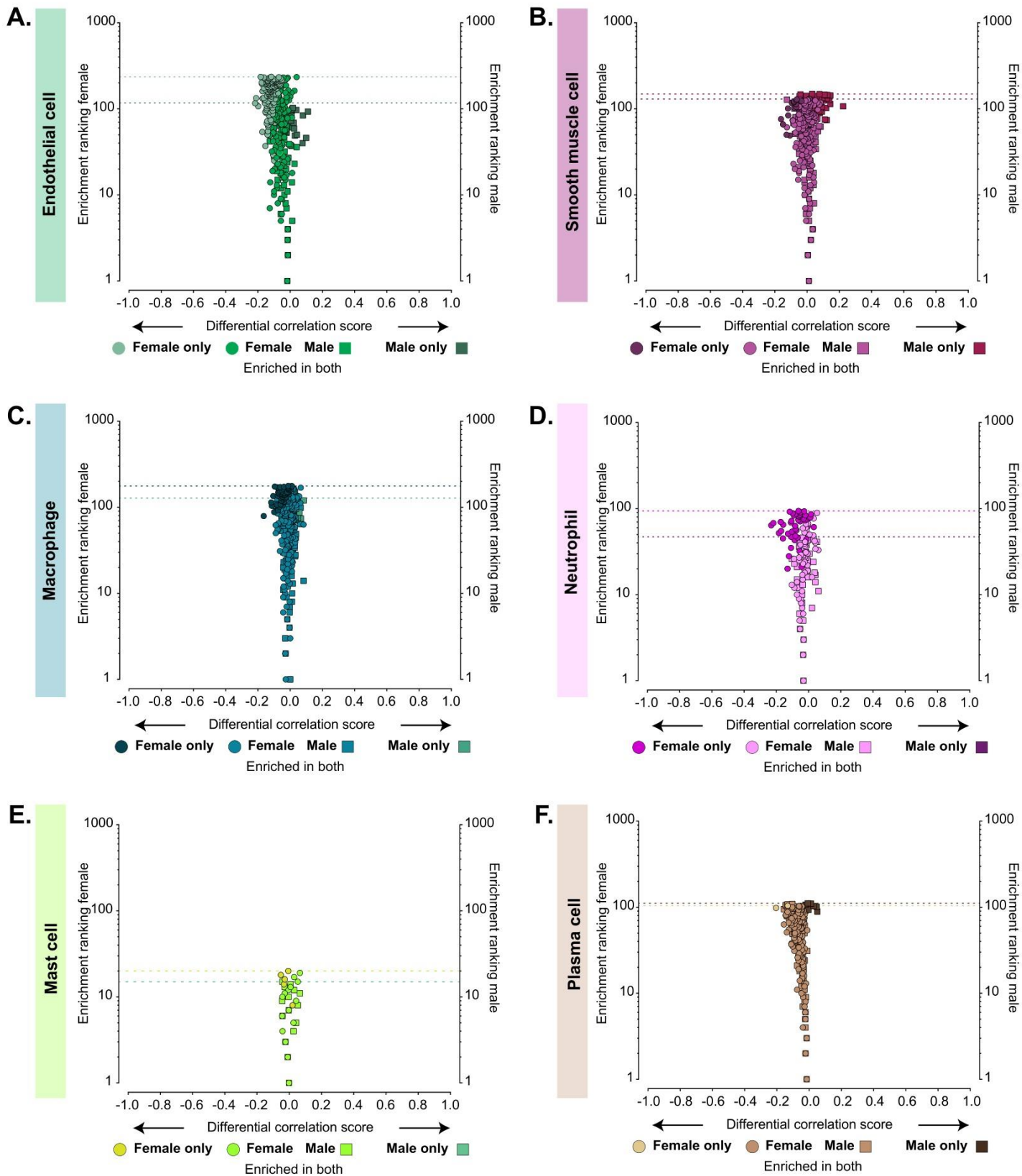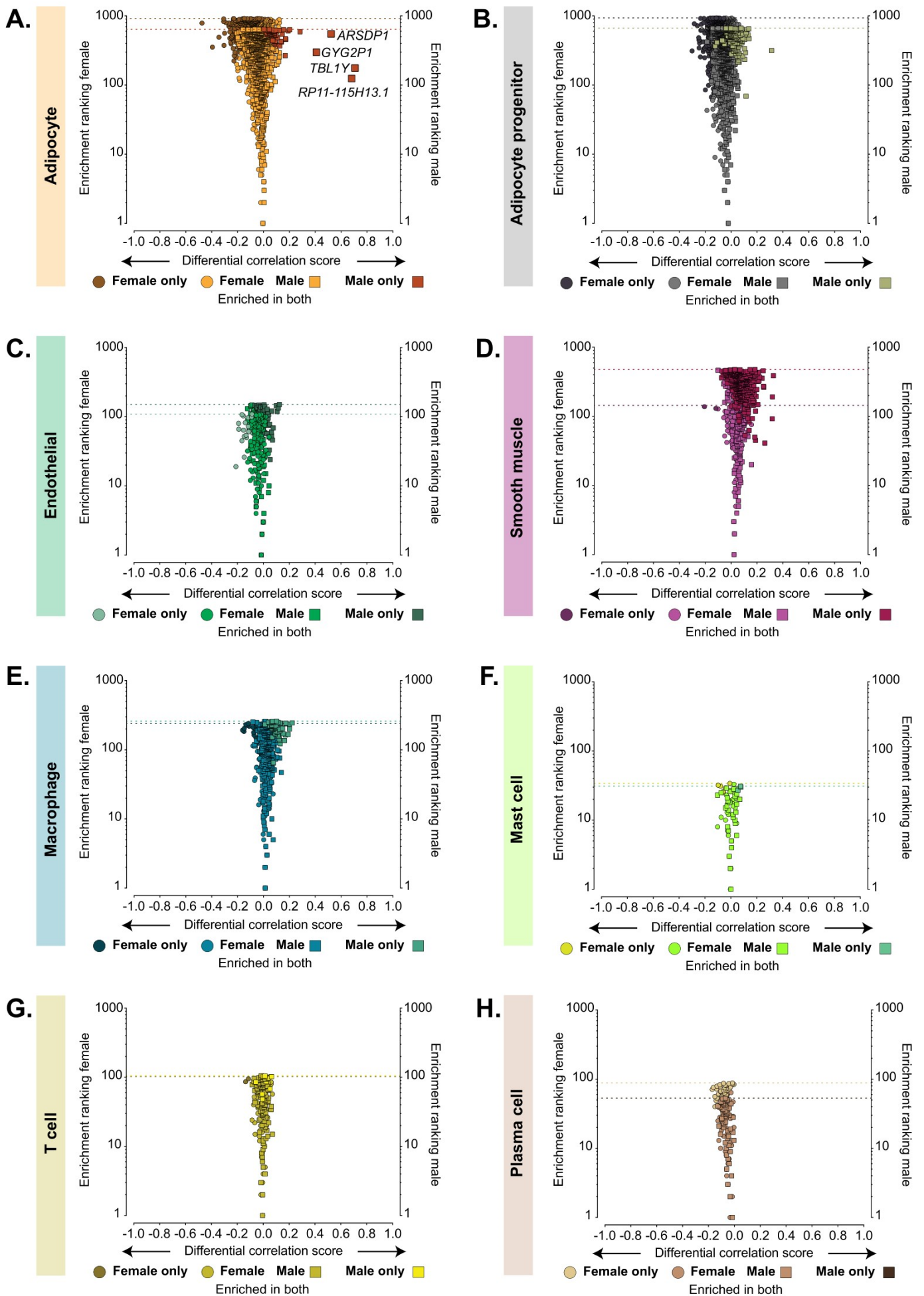
Figure S6. Identification of sex-specific cell type-enriched transcripts in human visceral adipose tissue (VAT); Related to Figure 5. VAT RNAseq data (n=527), retrieved from Genotype-Tissue Expression (GTEx) portal V8, was divided into female and male sample subgroups (female n=165, male n=362) before classification of cell type-enriched transcripts (see results section for criteria). For transcripts classified as: (**A**) endothelial, (**B**) smooth muscle, (**C**) macrophage, (**D**) neutrophil, (**E**) mast cell, or (**F**) plasma cell enriched, in either female or male subsets, the ´*sex differential correlation score'* (difference between mean corr. with the *Ref.T* panel in females vs. males) was plotted vs. 'enrichment ranking' (position in each respective enriched list, highest corr. = rank 1). On each plot, transcripts classified as enriched in *both* females and males are represented by common coloured circle and square symbols, respectively, and transcripts classified as enriched *only* in females or males are represented by differently coloured circle or square symbols, respectively. Correspondingly coloured threshold lines denote ranking below which transcripts were classified as female or male enriched. See also Table S3, Tab 2.

Figure S7. Identification of sex-specific cell-enriched transcripts in human subcutaneous adipose tissue (SAT); Related to Figure 5. Human SAT RNAseq data (n=646), retrieved from Genotype-Tissue Expression (GTEx) portal V8, was divided into female and male sample subgroups (female n=212, male n=434) before classification of cell type-enriched transcripts (see results section for criteria). For transcripts classified as: (**A**) adipocyte, (**B**) adipocyte progenitor, (**C**) endothelial, (**D**) smooth muscle, (**E**) macrophage (**F**) mast cell, (**G**) T-cell, or (**H**) plasma cell enriched, in either female or male subsets, the ´*sex differential correlation score'* (difference between mean corr. with the *Ref.T* panel in females vs. males) was plotted vs. 'enrichment ranking' (position in each respective enriched list, highest corr. = rank 1). On each plot, transcripts enriched in *both* females and males are represented by common coloured circle and square symbols, respectively, and transcripts classified as enriched *only* in females or males are represented by differently coloured circle or square symbols, respectively. Correspondingly coloured threshold lines denote ranking below which transcripts were classified as female or male enriched. Transcripts with differential corr. score >0.40 are labelled with identifiers. See also Table S3, Tab 4.

Figure S8

186
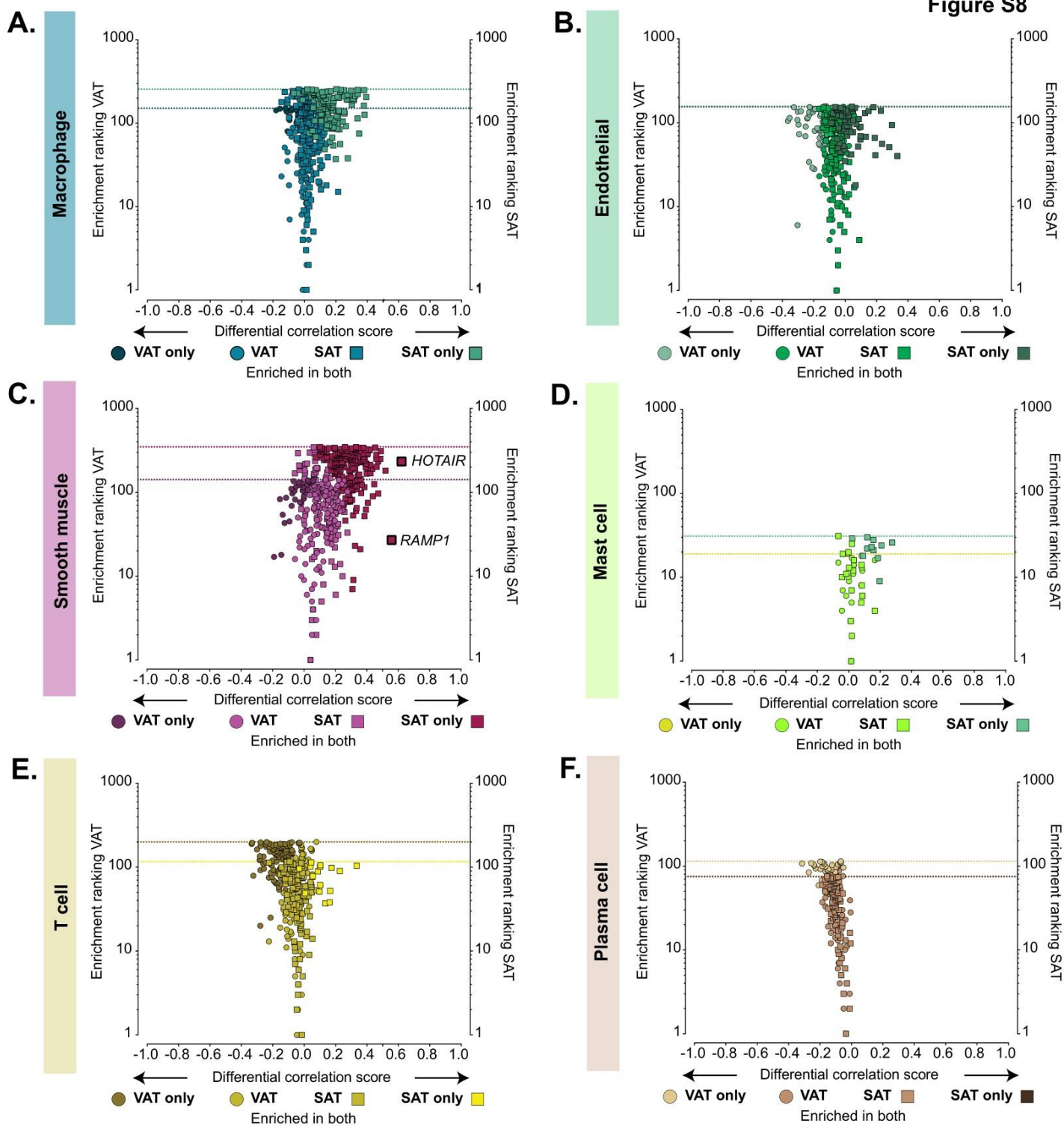
Figure S8. Identification of depot-specific cell-enriched transcripts in human adipose tissue; Related to Figure 6. Human visceral adipose tissue (VAT, n=527) or subcutaneous adipose tissue (SAT, n=646) RNAseq data, retrieved from Genotype-Tissue Expression (GTEx) portal V8, was used for classification of cell type-enriched transcripts (see results section for criteria). For transcripts classified as: (**A**) macrophage, (**B**) endothelial, (**C**) smooth muscle, (**D**) mast cell, (**E**) T-cell, or (**F**) plasma cell enriched, in either VAT or SAT, the *´depot differential correlation score'* (difference between mean corr. with the *Ref.T* panel in VAT vs. SAT) was plotted vs. 'enrichment ranking' (position in each respective enriched list, highest corr. = rank 1). On each plot, transcripts enriched in *both* VAT and SAT are represented by common coloured circle and square symbols, respectively, and transcripts classified as enriched *only* in VAT or SAT are represented by differently coloured circle or square symbols, respectively. Correspondingly coloured threshold lines denote ranking below which transcripts were classified as VAT or SAT enriched. Selected transcripts with differential corr. score >0.40 are labelled with identifiers.

PAPER 3

KANK3 is a shear stress regulated endothelial protein with a role in cell migration and tissue factor regulation

(manuscript)

KANK3 is a shear stress regulated endothelial protein with a role in cell migration and tissue factor regulation

Eike Christopher Struck[1], Sofia Maria Öling[1], Philip James Dusart[2], Marthe Norreen-Thorsen[1], Julian Connor Eckel, Larissa Dorothea Kruse[3], Casper Ullsten-Wahlund[1], Jacob Odeberg[1, 2, 4], Clément Naudin[1,2,] Lynn Marie Butler[1, 2, 5]

[1]Department of Clinical Medicine, The Arctic University of Norway, N-9037, Tromsø, Norway

[2]Science for Life Laboratory, Department of Protein Science, School of Engineering Sciences, Stockholm, Sweden

[3]Department of Medical Biology, The Arctic University of Norway, N-9037, Tromsø, Norway

[4]The University Hospital of North Norway (UNN), PB100, 9038 Tromsø, Norway

[5]Clinical Chemistry and Blood Coagulation Research, Department of Molecular Medicine and Surgery, Karolinska Institute, SE-171 76 Stockholm, Sweden

Correspondence information:

Dr. L.M Butler, PhD

Department of Clinical Medicine,

The Arctic University of Norway,

N-9037, Tromsø,

Norway

Email: Lynn.m.butler@uit.no

KEY WORDS: Endothelium, KANK3, focal adhesions, cytoskeleton, cell migration, coagulation

1

ABSTRACT

The endothelium is the innermost layer of all blood vessels. Endothelial cells (EC) play a central role in the regulation of vascular processes, such as coagulation, inflammation, and angiogenesis. Proteins with EC restricted expression tend to be critical for such cell type specific functions. In a previously published bioinformatic based analysis of RNAseq, we predicted that *KANK3,* which encodes an uncharacterised protein, had body wide enriched expression in human EC. Here, we verify that KANK3 is a body-wide endothelial-enriched protein at the transcript and protein level. We characterise its subcellular distribution in primary EC and uncover that its expression is strongly induced in response to shear stress exposure. When KANK3 protein was depleted using siRNA, the distribution of the EC intermediate filament vimentin was disrupted in both static and shear stress exposed cultures, indicating a direct or indirect interaction between these proteins. Correspondingly, in a wound healing model, depletion of KANK3 increased EC migratory capacity, but did not increase proliferative capacity. Furthermore, we observed an increase in the expression of the pro-coagulant protein tissue factor in KANK3 depleted EC, indicating that it could have further regulatory roles, beyond those associated with cytoskeletal modification and motility.

2

INTRODUCTION

The vascular endothelium lines the inside of all blood and lymphatic vessels and has numerous functions, including in the regulation of inflammation, haemostasis, and blood pressure [1, 2]. Proteins specifically expressed in endothelial cells (EC) tend to have central roles in cell specialised functions, e.g., cadherin-5 (CDH5), claudin-5 (CLDN5) and endothelial cell-selective adhesion molecule (ESAM), play established roles in EC integrity, polarity and shape, vessel permeability and signalling [3-6], and the vascular endothelial growth factor receptor 1 (FLT1) and 2 (KDR) are central to angiogenesis [7]. In earlier work, based on bioinformatic analysis of bulk RNAseq, we predicted that the gene encoding the uncharacterised protein KN Motif And Ankyrin Repeat Domains 3 (KANK3) had body wide enriched expression in human EC [8].

The KANK family consist of four members (KANK 1-4), which arose through gene duplication and diversification, with strong conservation across the evolutionary tree [9, 10]. They are defined by their unique structure, consisting of a variable number of coiled-coil motifs in the central N-terminal regions, five ankyrin repeats in the C-terminal region and a talin-binding KN-motif domain at the N-terminus [10, 11]. The interaction between KANK1 and talin regulates the recruitment of complexes that stabilize cortical microtubules to focal adhesions [11]. KANK2 promotes the creation of central adhesions by triggering talin activation and is responsible for the reduction of force transduction across integrins [12]. KANK1 and 2 are involved in cell migration and adhesion, via interactions with kinesin family member 21A (KIF21A), and the regulation of its activity through its coiled-coil domain [10, 13].

The cytoskeleton and focal adhesions are crucial for various EC specialised functions, such as the maintenance of the structural integrity required to withstand the mechanical forces exerted by the blood flow [14], to control movement and migration during processes such as angiogenesis [15], to stabilise junctional connections and control vascular permeability [16] and in processes such as coagulation [17] and inflammation [18].

3

While KANK1 and KANK2 are well relatively studied, KANK3 is comparably poorly described; currently, it has no reported function in a vascular context in vertebrates. A homologue of KANK3 has been described in vascular EC in of zebrafish embryos, where it was essential for embryonic development and survival, with a potential role in cell adhesion and tissue integrity [19, 20]. Over expression of KANK3 in NIH3T3 cells revealed a possible role in actin stress fibre formation [21], and other studies have indicated a role in the regulation of cell migration in hepatocellular carcinoma [22] and lung adenocarcinoma [23].

In this study, we verify that KANK3 is a body-wide endothelial-enriched protein. We characterise its subcellular distribution in primary EC, and report that it is shear stress-induced gene. We show that KANK3 depletion modifies the subcellular distribution of the EC intermediate filament vimentin and increases EC motility in a gap closing assay. We observed an increase in the expression of the pro-coagulant protein tissue factor in KANK3 depleted EC, particularly under inflammatory conditions, together with an increased capacity to induce thrombin generation in plasma. Thus, we demonstrate a role for the EC enriched protein KANK3 in EC specialised functions.

4

RESULTS

KANK3 IS AN ENDOTHELIAL ENRICHED PROTEIN IN HUMAN

KANK3 mRNA expression correlates with endothelial cell genes in human tissues

Proteins expressed specifically by EC tend to be critical for EC specialised functions. Previously, using mixed tissue bulk RNAseq, we found that *KANK3* expression was strongly correlated with EC marker genes, indicating EC specificity [8]. More recently, using a similar approach, we profiled gene enrichment signatures for cell types in 15 individual tissue datasets [24-27] (data is displayed on the Human Protein Atlas [www.proteinatlas.org/humanproteome/tissue+cell+type]). Here, *KANK3* was predicted to be EC enriched in multiple vascular beds (Figure S1 A.i), whilst other KANK family members, *KANK1*, *2* or *4* were not (Figure S1 B-D.i).

To further explore this potential relationship using a *KANK3*-centric approach in an expanded dataset, we retrieved bulk RNAseq datasets for 36 human tissue types from Genotype-Tissue Expression (GTEx) V8 (www.gtexportal.org)[28] (mean samples/tissue =377, range 85-803) (Table S1A). For each dataset, we calculated Pearson correlation coefficient values between *KANK3* and all other mapped genes (Table S1B). The top 100 most highly correlating genes with *KANK3* (all correlation coefficient [corr.] >0.5, p-value<0.0001) in each tissue type (Table S1C) were cross-compared, to identify 67 genes that were highly correlated with *KANK3* in 10 or more tissue types (Figure 1A) (Table S1D). These genes included *EGFL7* (32 tissues; mean corr. =0.76), *ROBO4* (29 tissues; mean corr. =0.74), *ESAM* (29 tissues; mean corr. =0.76) and *CDH5* (29 tissues; mean corr. =0.72); all of which have key roles in EC specific functions [29-32].

Gene ontology (GO) analysis [33] was performed to identify over represented groups within this list of 67 genes (Figure 1B and Table S1E). Over-represented terms were related to vascular or EC function and included ´*vasculature development*´ (p=1.1 x 10^{13}), ´*angiogenesis*´ (p=1.8 x 10^{11}) and ´*establishment of endothelial barrier*´ (p=3.2 x 10^{7}) (Figure 1 B) (Table S1E). As high correlation values between genes within tissue can indicate co-expression in a common cell-type, these results are consistent with our prediction that *KANK3* is an EC enriched gene.

5
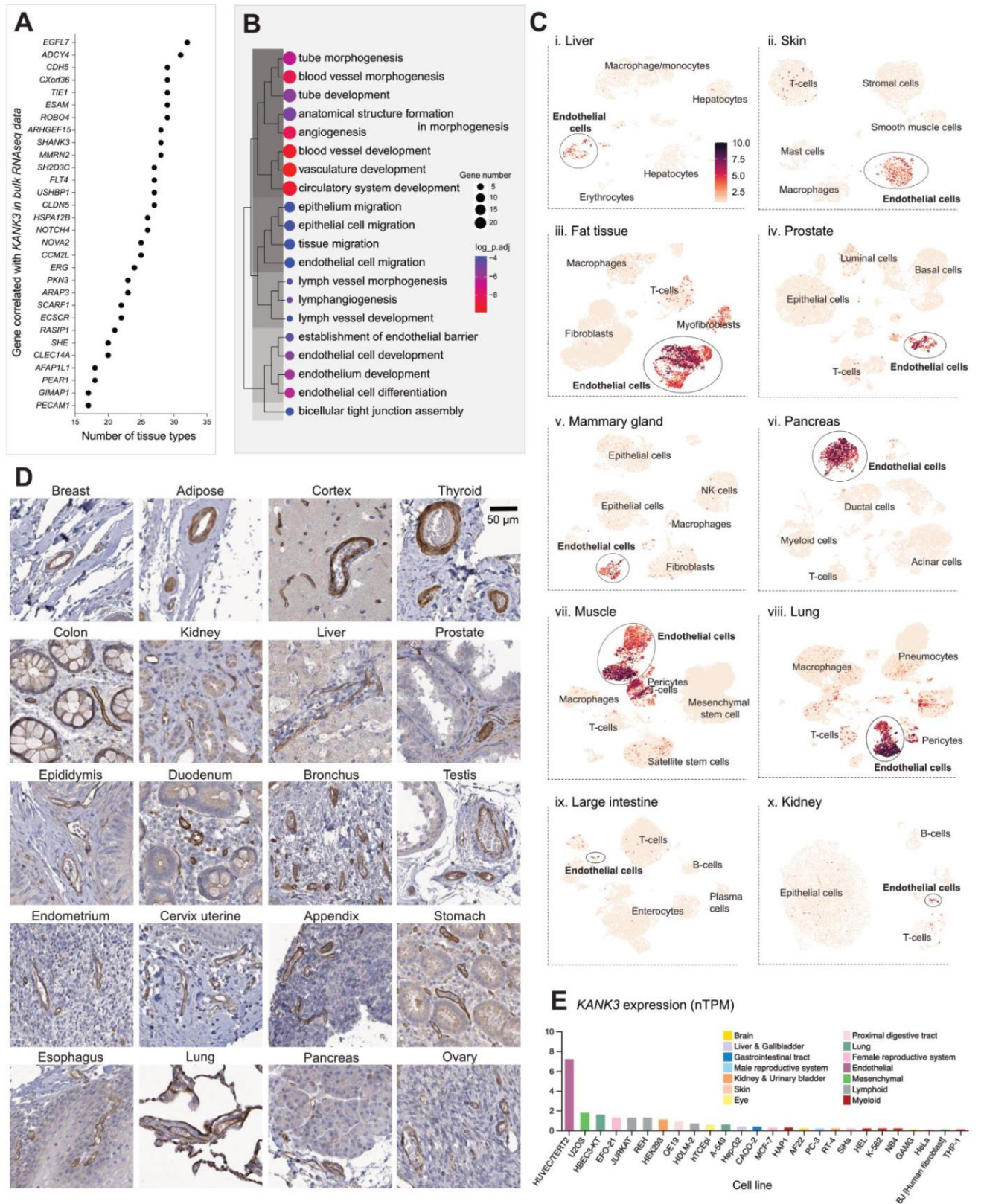
Figure 1. KANK3 is an endothelial cell enriched protein in the human. RNAseq data from 36 human tissue types was sourced from Genotype-Tissue Expression (www.gtexportal.org) [28]. Pearson correlation coefficient values between *KANK3* and all other mapped genes were calculated for each. (**A**) Genes most frequently among the top 100 most highly *KANK3* correlated genes (all >0.50, p<0.0001) across tissue types (**B**) Gene Ontology over-represented terms associated with genes among the top 100 most highly *KANK3* correlated genes in ≥10 tissues. (**C**) Data was downloaded from Tabula Sapiens [34] and used to generate Uniform manifold approximation and projection (UMAP) visualizations for *KANK3* expression in human: (i) liver, (ii) skin, (iii) fat, (iv) prostate, (v) mammary gland, (vi) pancreas, (vii) muscle, (viii) lung, (ix) large intestine and (x) kidney. (**D**) Protein profiling for KANK3 across human tissue types. (**D**) Immortalised human cell lines in the panel tested that had with the highest expression of *KANK3*, generated as part of the Human Protein Atlas project (ref).

7

KANK3 endothelial enriched expression can be verified by scRNAseq and protein profiling

Single cell RNAseq data from the Tabula Sapiens [34] was used to explore expression profiles of KANK3 in skin, liver, fat, prostate, mammary gland, pancreas, muscle, lung, large intestine, and kidney (Figure 1 C.i-x). In all cases, KANK3 was predominantly expressed within clusters annotated as EC. Low levels of *KANK3* were detected in myofibroblasts in fat (Figure S1 C.iii), and pericytes in muscle and lung (Figure 1C.vii and viii). There was little or no *KANK3* expression in tissue specific cell types, e.g., hepatocytes in the liver (Figure 1 C.i) or pneumocytes in the lung (Figure 1 C.viii). Protein profiling confirmed EC expression of KANK3 in multiple tissues, including colon, kidney, liver, breast, adipose, cortex, prostate, skeletal muscle, thyroid (Figure 1D) and others (Figure S1B) [35].

KANK3 expression is enriched in cell lines of endothelial origin

To determine if *KANK3* expression is maintained in cells of EC origin following immortalisation, we examined its expression in RNA-sequencing data from different 41 cell lines from the HPA [35]. *KANK3* was not detectable, or detectable only at very low levels (nTPM ≤ 0.5) in 30/41 (73%) of the cell lines tested. The highest expression was detected in human umbilical vein endothelial cells (HUVEC), HUVEC/TERT2 cells (7.2 nTPM), followed by the mesenchymal cell line U2OS cells (1.8 nTPM) (Figure 1E) (data for all shown in Figure S1). In comparison, other members of the KANK family (KANK1, 2 and 4), which we have previously predicted to lack EC specificity across human tissue types [36], show no EC specificity in immortalised cell lines (Figure S2) or in scRNAseq data from Tabula Sapiens (Figure S2). Together, these data support our prediction [8, 24, 27, 36] that *KANK3*, but not *KANK1*, *KANK2* or *KANK4*, is a human endothelial enriched gene.
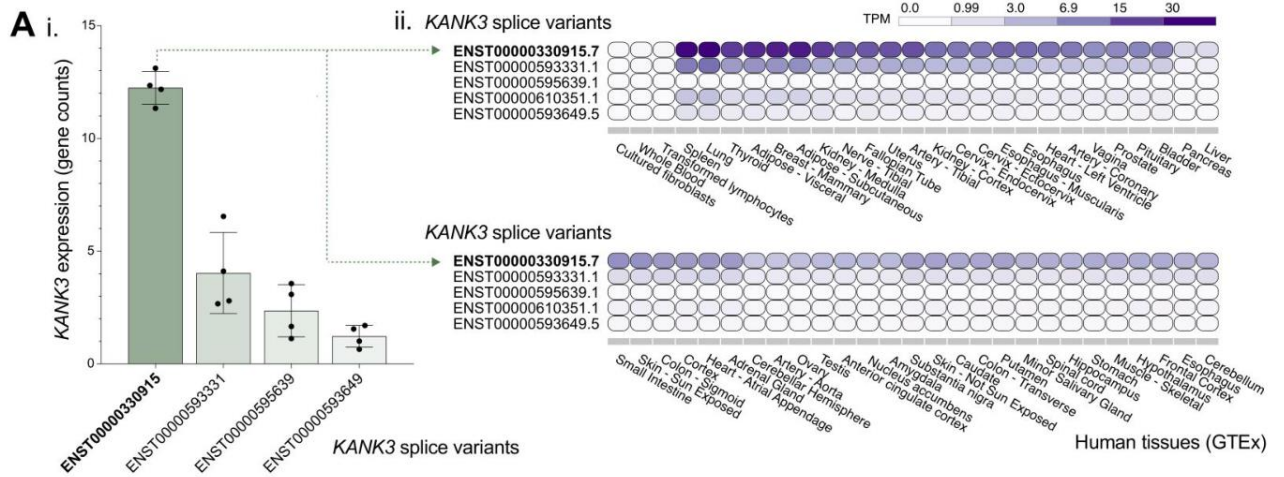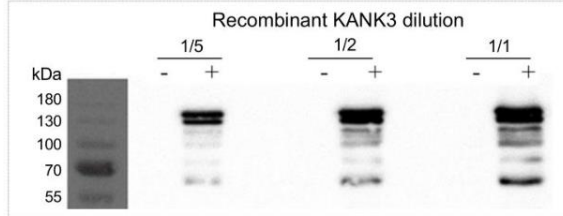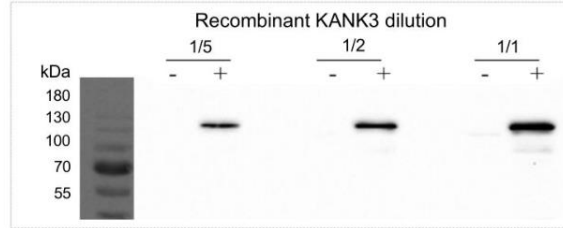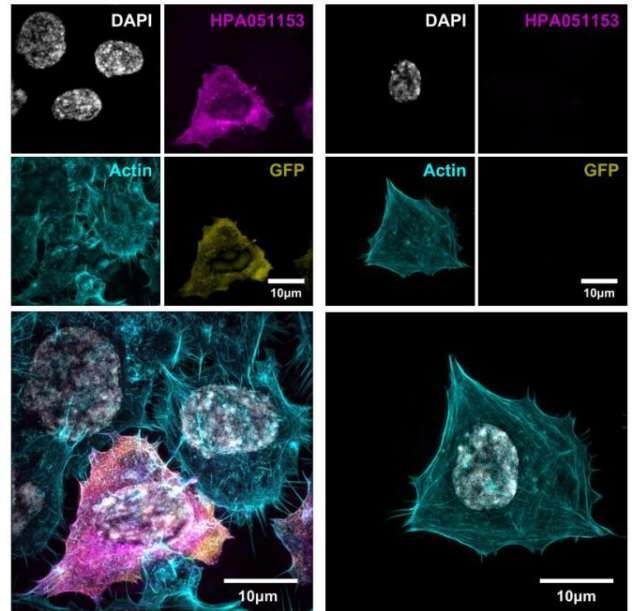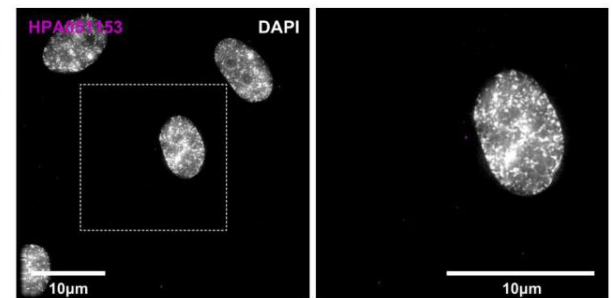
8

197
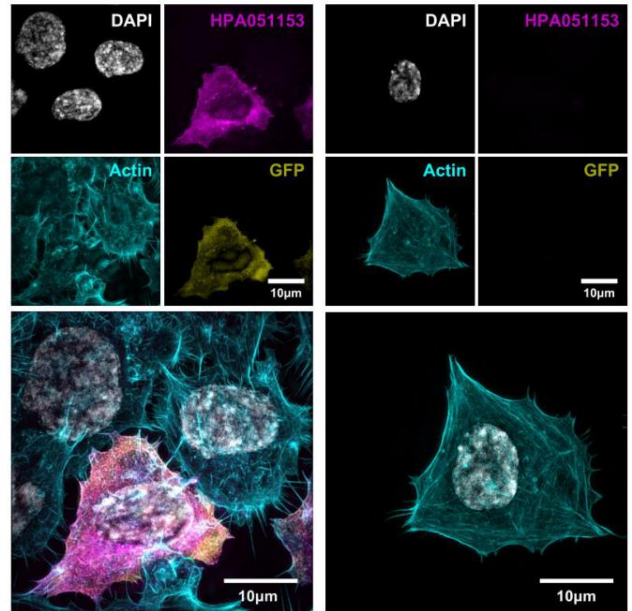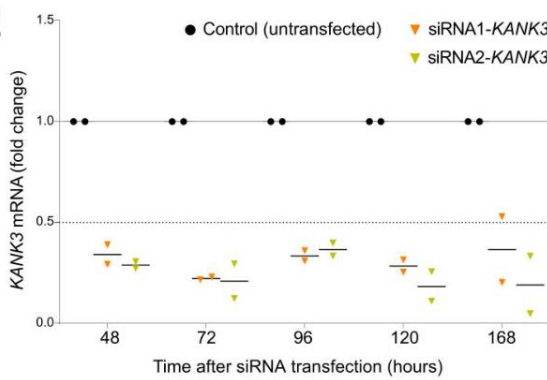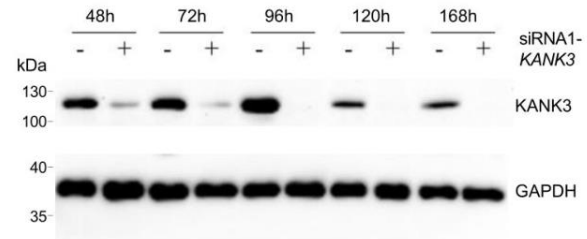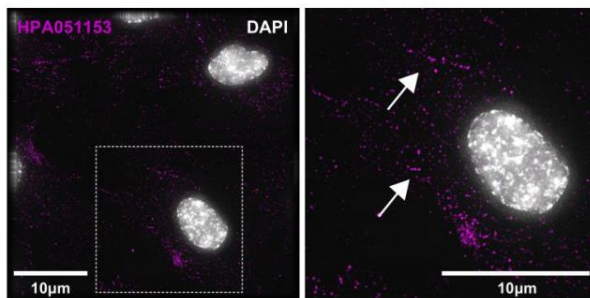
Figure 2. Validation of tools for the study of KANK3 function. (**A**) Detection of KANK3 splice variants by RNAseq in: (i) primary HUVEC and (ii) unfractionated human tissue from Genotype‑Tissue Expression (www.gtexportal.org). HEK293 cells were untreated or transfected with KANK3-eGFP expressing plasmids and used to generate (**B**) cell lysates for Western Blot analysis using primary antibodies: (i) HPA051153, a rabbit polyclonal antibody raised to target KANK3, or (ii) an anti-eGFP antibody, or (**C**) immunocytochemistry staining to show signals from eGFP (yellow), HPA051153 (Magenta), Actin (Phalloidin-647; yellow) or DAPI (Nuclear staining, Gray), in (i) KANK3 transfected or (ii) untreated cells. HUVEC were transfected with siRNAs targeting KANK3 and cultured for between 48 and 168h before cell lysis and (**E**) measurement of KANK3 mRNA expression using real time qPCR and (**F**) Western blot analysis using primary antibody HPA051153 and an anti-GAPDH as loading control, or (G) immunocytochemistry using HPA051153 (Magenta), and DAPI (Nuclear staining, grey) (72 hours post transfection). All immunocytochemistry images were captured using captured using structured illumination microscopy (SIM).

10

GENERATION AND VERIFICATION OF TOOLS FOR THE STUDY OF KANK3

The search term KANK3 returns 11 hits on PubMed (https://pubmed.ncbi.nlm.nih.gov/), confirming this protein is not well studied. One reason why under studied proteins are unattractive targets for functional analysis can be a lack, or unknown reliability, of research tools with which to investigate them [37, 38]. Thus, we tested our model system and generated or purchased reagents prior to study of KANK3 functional role in EC.

KANK3 splice variant expression profile in HUVEC reflects that found in vivo

Having confirmed that KANK3 is an EC enriched protein across tissue types, we went on to verify its expression profile in freshly isolated primary HUVEC, which we planned to use as an experimental model. RNA sequencing of *in vitro* cultured HUVEC (n=5) revealed that *KANK3* was reasonably highly expressed (mean 26.37 TPM ± std dev 2.25). Other KANK family members were expressed at similar levels: *KANK1* (39.5 TPM ± std dev 3.31), *KANK2* (27.3 TPM ± std dev 2.62), with the exception of *KANK4*, which was very lowly expressed (0.04 TPM ± std dev 0.03). *KANK3* splice variant ENST00000330915 was the most common isoform in HUVEC (61.6%; length: 821 aa), whilst the other variants were expressed at lower levels: ENST00000593331 (20.3%; non-protein coding), ENST00000595639 (11.9%; length: 146 aa) and ENST00000593649 (6.2%; length: 840 aa) (Figure 2A.i). Data from bulk sequencing of unfractionated human tissues in GTEx revealed that, similar to HUVEC, ENST00000330915 was the most highly expressed *KANK3* splice variant, followed by the non-protein coding variant ENST00000593331 (Figure 2 A.ii). The transcript ENST00000610351.1 in the GTEx data (Figure 1 A.ii) was retired after ENSEMBL version 104 and is not part of the current ENSEMBL gene set (V110), which our sequencing data was mapped against (hence its absence from Figure 1 A.i). It can be assumed, based on the verification of KANK3 as an endothelial enriched protein (Figure 1), that the expression site of the *KANK3* isoforms within GTEx tissues is largely EC restricted. Thus, the relative expression profile of *KANK3* variants in HUVEC reflects that found *in vivo*.
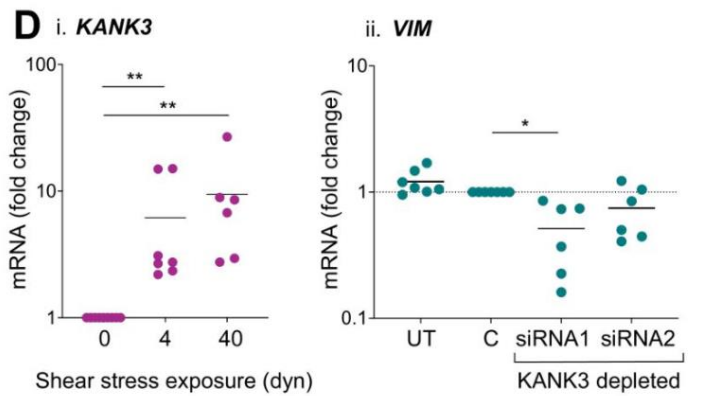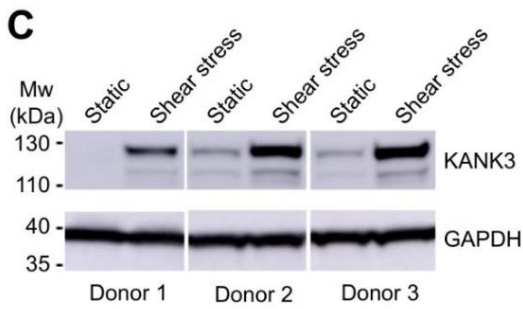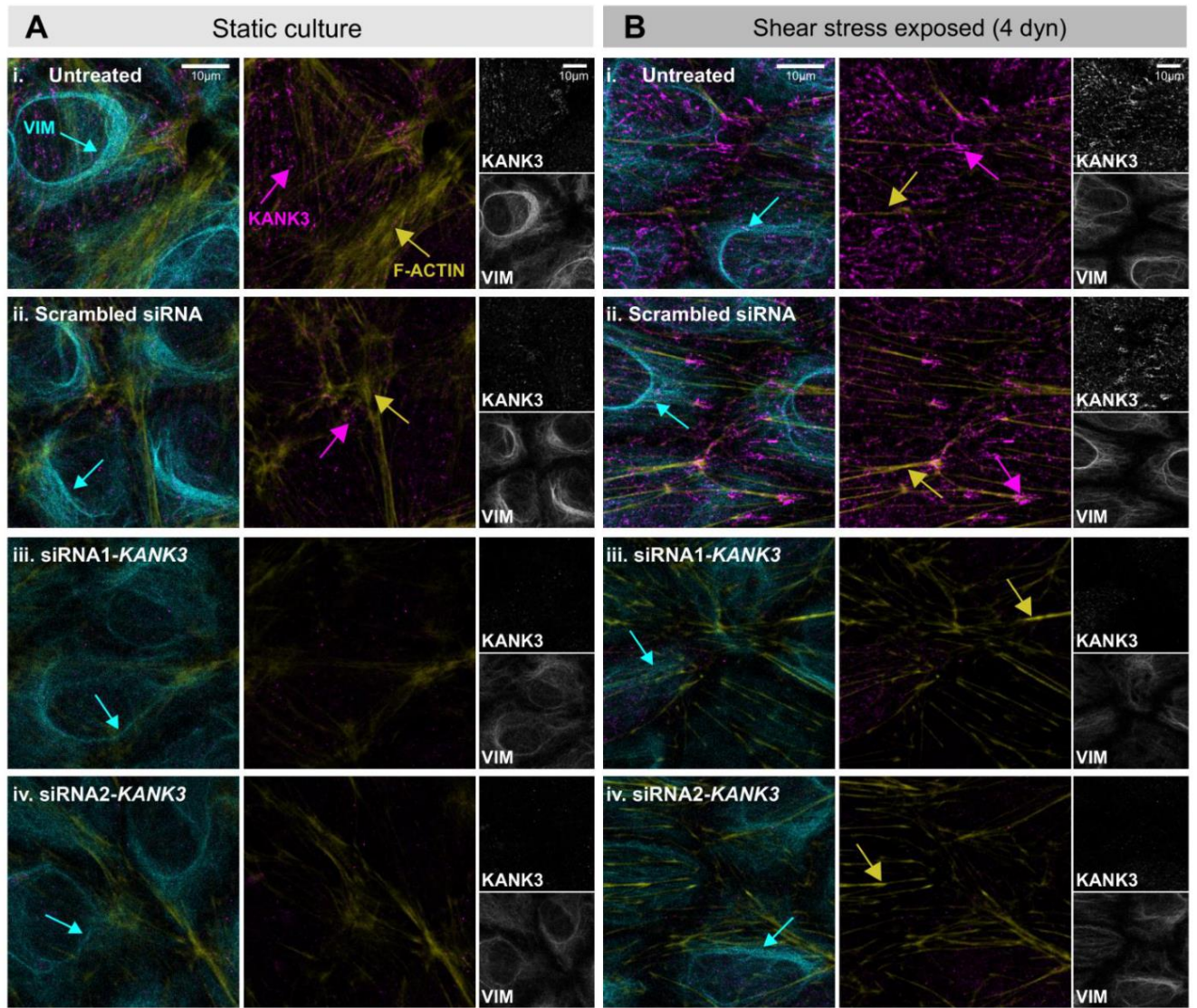
11

Figure 3: KANK3 is a shear stress regulated protein. HUVEC were cultured under (**A**) static or (**B**) shear stress exposed conditions (4 or 40 dyn/cm^2) for 48 hours after they were (i) untreated, or transfected with (ii) scrambled control siRNA, (iii) siRNA1-*KANK3* or (iv) siRNA2-*KANK3*. Immuno-cytochemistry was performed using primary antibodies targeting KANK3 (magenta), F-actin (yellow) or vimentin (VIM; cyan). (**C**) Untreated HUVEC from 3 donors were cultured under static or shear stress exposed conditions (4 dyn/cm^2) for 48 hours (´flow´) before analysis of KANK3 protein expression by Western blot, with GAPDH as a loading control. (**D**) Expression of (i) *KANK3* mRNA in HUVEC following static or shear stress exposure or (ii) *VIM* in HUVEC following transfection with or without siRNAs targeting *KANK3*. ** p<0.001, * p<0.01

13

Verification of KANK3 antibody specificity and siRNA knockdown efficiency

Antibody reliability can be a problematic issue and a major source of research waste; it has been suggested that up to half of all commercially available antibodies have significant issues with sensitivity and specificity [39, 40]. Furthermore, the availability and testing of such reagents targeting understudied proteins is limited [41] and thus, the validation of antibody reagents is important.

We obtained an in house generated rabbit polyclonal antibody targeting KANK3 (HPA051153) [35]. To verify its binding specificity, we took a twofold approach. Firstly, we designed a plasmid vector coding for *KANK3* with an eGFP tag, based on the sequence of the most common KANK3 isoform (ENST00000330915). This was used to express recombinant KANK3 in HEK293 cells, which do not express endogenous *KANK3*. Western Blot analysis of cell lysates with antibody HPA051153 detected bands corresponding to the size of the KANK3 protein (130kDA) (Figure 2 B.i), and staining with an anti-eGFP antibody gave similar results to HPA051153, over a range of dilutions (Figure 2B.ii).

Immunofluorescence staining of KANK3-eGFP transfected HEK293 cells with antibody HPA051153 (Figure 2 C.i) showed selective binding (pink) to cells expressing recombinant KANK3-eGFP (yellow) (Figure 1 C.i, large panel). HPA051153 did not bind KANK3-eGFP negative HEK293 cells within the transfected culture (Figure 1C.i, large panel), or to untreated HEK293 cells (Figure 2 C.ii). Secondly, to test antibody HPA051153 specificity for HUVEC KANK3, we used siRNA to deplete the protein. Two different siRNAs (siRNA1-*KANK3* and siRNA2-*KANK3*) effectively depleted HUVEC *KANK3* mRNA expression over an extended time course (fold change at 48h [mean ± std dev]: siRNA1 0.22 ± 0.006, siRNA2 0.21 ± 0.086) (Figure 2E).

Correspondingly, subsequent Western Blot analysis with HPA051153 showed that bands of a size corresponding to KANK3 protein (130kDA) were smaller, or absent, in cell lysates from siRNA transfected HUVEC (Figure 2F); an inhibition that was maintained over several days. Having verified efficient *KANK3* knockdown in HUVEC using siRNA, we performed immunofluorescence

14

staining using HPA051153 as a primary antibody (Figure 2G). HPA051153 showed clear punctate staining in untreated HUVEC (Figure 2 G.i), which was absent when cells were transfected with siRNA targeting *KANK3* (Figure 2 G.ii). Thus, we can have high confidence that HPA051153 selectively binds KANK3 protein in both Western Blot and immunofluorescence staining applications. Furthermore, HUVEC appear to be a suitable model system for the functional investigation of KANK3, and siRNA-mediated depletion induced a robust knockdown of KANK3 protein for several days after transfection.

KANK3 localizes within the cytoplasm and accumulates in cell-cell interaction sites

A recent study found KANK3 to be expressed at the plasma membrane of mouse EC in dermal and lymphatic vessels (S. S. Guo et al. 2021), with more diffuse staining in kidney, lung brain and oesophagus EC. Immunofluorescence staining of native KANK3 expression in endothelial cells such as HUVEC and mouse LSEC however, shows punctate distribution of KANK3 in the cytoplasm and accumulation in cell-cell interaction sites (FIGURE S2B) as well as partial colocalization to the cytoskeleton (Figure S2A).

Expression of KANK3 is enhanced under flow versus static conditions

Whilst KANK3 is poorly studied in a functional context, it shares structural homology with other members of the KANK family, which have been shown to have a role in cytoskeletal organisation, and focal adhesion formation [21, 42]. As such processes are key in the EC response to shear stress [14, 43], we investigated KANK3 expression and distribution in this context.

Untreated HUVEC, or those transfected with siRNA1-*KANK3*, siRNA2-*KANK3* or a scrambled siRNA control were cultured under static or shear stress exposed conditions (4 or 40 dyne/cm^2) for 48 hours. Cell were fixed and stained for KANK3, vimentin - the major endothelial intermediate filament (IF) that is a key regulator of focal contact size and cell-matrix adhesions in EC subjected to shear stress [44, 45] and the actin cytoskeleton which has a vital role in cell-cell adhesions [46]. Under static conditions, in both untreated and control HUVEC, KANK3 had a diffuse punctate distribution (Figure 3 A.i and ii, magenta arrows) and its expression was markedly up regulated

15

following shear stress exposure (Figure 3 B.i and ii, magenta arrows). Staining patterns revealed distinct areas of dense KANK3 expression. Western blotting confirmed that a significant up regulation of KANK3 protein was induced in HUVEC that had been exposed to shear stress, in 3 biological replicates (Figure 3C). Measurement of *KANK3* mRNA showed significantly elevated levels following exposure to shear stress of both 4 and 40 dyne/cm^2 (fold change vs. static ± std dev: 4 dyne/cm2 6.15 ± 5.6 p=0.004, 40 dyne/cm2 9.5 ± 8.1 p=0.002) (Figure 3D).

Under static conditions, in both untreated and control HUVEC, vimentin was located to the endogenous IF network (Figure 3 A.i and ii, turquoise arrows) with notable directional redistribution following shear stress exposure (Figure 3 B.i and ii, turquoise arrows), as previously described [47]. In KANK3 depleted HUVEC, cultured under static conditions, vimentin expression was markedly reduced compared to untreated or control HUVEC (Figure 3 A.iii and iv, turquoise arrows, Figure S2C & D). Shear stress exposure failed to induce a recovery of vimentin expression or a more typical redistribution pattern in KANK3 depleted cells (Figure 3 B.iii and iv, turquoise arrows).

In EC cultured under static conditions, *VIM* mRNA expression was lower in siRNA1-*KANK3* treated HUVEC, compared to untreated or control HUVEC, but not in siRNA2-*KANK3* treated HUVEC (fold change control EC vs. siRNA1-*KANK3* ± std dev 0.51 ± 0.27 p=0.03, siRNA2-*KANK3* 0.74 ± 0.25 p=0.10) (Figure 3 D.ii). Thus, the effects of EC *KANK3* depletion on vimentin expression/cellular distribution is unlikely to be driven by changes at the transcriptional level. Actin was diffusely expressed under static conditions, in both untreated and control HUVEC (Figure 3 A.i and ii, yellow arrows). Actin was redistributed to align with flow direction following shear stress exposure, as previously described [48] (Figure 3 B.i and ii, yellow arrows). Unlike vimentin, actin redistribution in response to shear stress was not markedly modified by KANK3 depletion (Figure 3 B.iii and iv, yellow arrows).

Thus, our results show that KANK3 levels are increased in response to EC and that KANK3 has a previously unreported direct or indirect link to vimentin distribution.
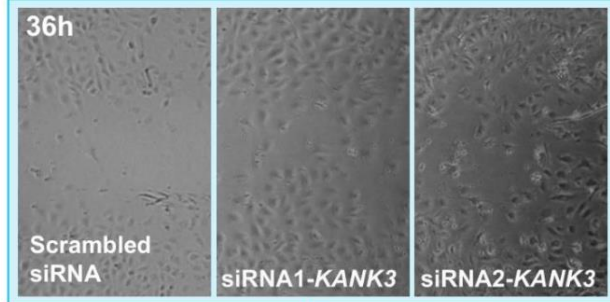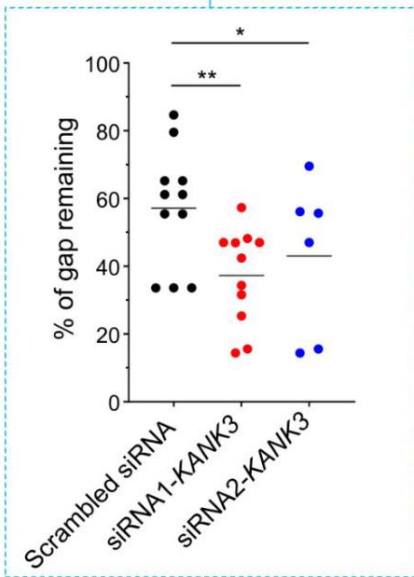
16
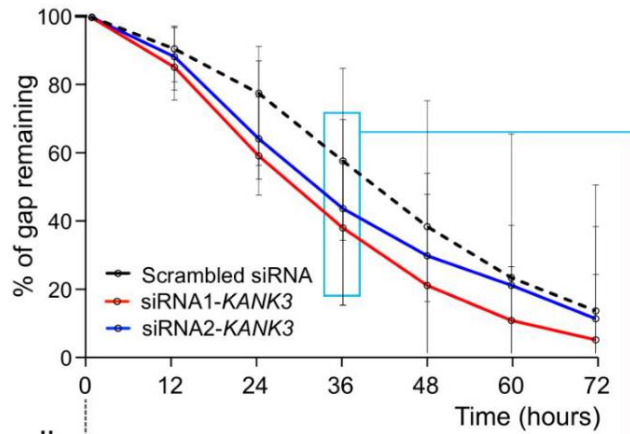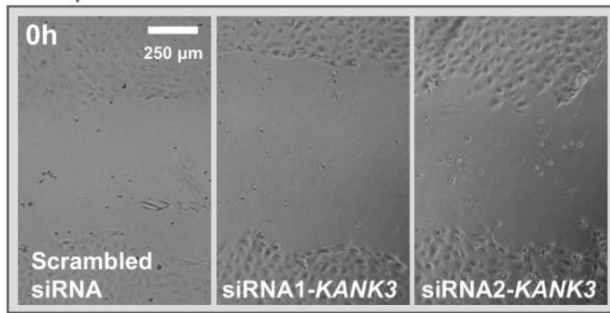
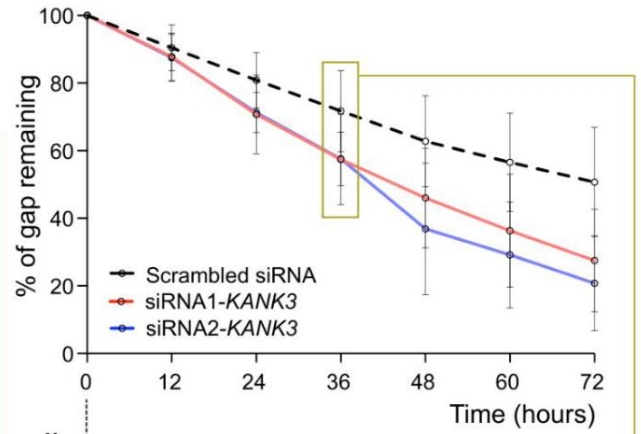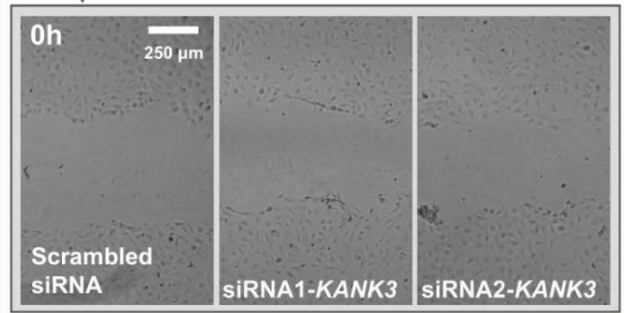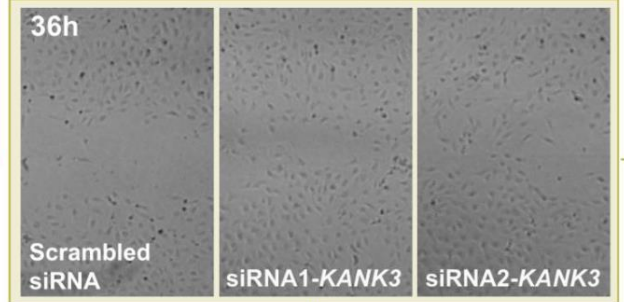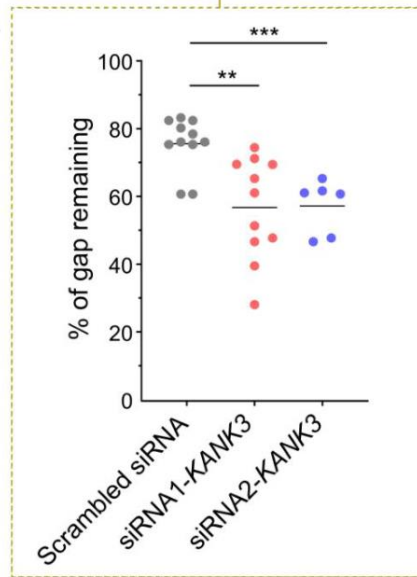**Figure 4. KANK3 depletion increases EC migration *in vitro*.** HUVEC were cultured to confluence in (**A**) standard or (**B**) low serum (0.5%) culture medium, following transfection with scrambled control siRNA, siRNA1-*KANK3* or siRNA2-*KANK3*. A ´wound´ was created in the monolayer, using a pipette tip, and (i) gap closure was monitored over 72 hours. (ii) Representative phase contrast images and (iii) corresponding data points from individual experiments, from the 36-hour time point. *** p<0.001, ** p<0.01, *p<0.05

18

KANK3 has a role in endothelial cell migration

KANK3 has not been studied in the context of EC motility, but it has been reported to have a role in the regulation of cell motility in cancer cells [22, 23]. Furthermore, we observed that KANK3 depletion modified EC vimentin distribution, a protein with a key role in EC migration[49]. Therefore, we analysed the influence of KANK3 depletion in an EC gap closing assay. EC were transfected with one of 2 siRNAs targeting KANK3, or a scrambled siRNA control and cultured to confluence before a gap was created in the monolayer. Gap closure was monitored in real time, using phase contrast microscopy, and the wound area was measured at 24h, 36h, 48h and 72h. *KANK3* siRNA treated EC tended to close the gap faster than those treated with the scrambled control EC (p for trend C vs siRNA 1 p for trend 0.0142, C vs siRNA 2 p for trend: 0.0505) (Figure 4 A.i). Representative phase contrast images show the gap size at 0 and 36 hours (Figure 4 A.ii), with corresponding data points for replicate experiments (Figure 4 A.iii) (Change in gap closure normalised to control [%] ± std dev: siRNA1-*KANK3*: +30.9 ± 24.9, p=0.004, siRNA2-*KANK3*: +36.0 ± 24.6, p=0.028). Accelerated gap closure could be either due to increased migratory capacity of cells in which KANK3 has been depleted, or an increase in cell proliferation.

To assess the relative role of each, we performed the same experiment in low serum culture medium as we have previously showed that proliferation is inhibited in this condition [50]. As expected, closure of the gap was inhibited in low serum medium (Figure 2 A.ii) (% of gap remaining [control HUVEC] standard vs. low serum ± std dev: 36h 57.1 ± 16.8 vs. 75.9 ± 7.5 and 72h 10.3 ± 11.8 vs. 50.3 ± 13.7).

As observed for standard medium, *KANK3* siRNA treated EC tended to close the gap faster than control EC (p for trend C vs siRNA 1 p for trend 0.0307, C vs siRNA 2 p for trend: 0.0402) (Figure 4 B.i), but no gap closed completely within the 72-hour time frame. Representative phase contrast images of the gap size at 0 and 36 hours (Figure 4 B.ii) and corresponding data points for replicate experiments (Figure 4B A.iii) (Change in gap closure normalised to control [%] ± std dev: siRNA1-*KANK3*: 23.3 ± 21.0, p=0.018, siRNA2-*KANK3*: +23.5 ± 6.0, p=0.0004), showed similar results as

19

those obtained in standard culture medium. Furthermore, measurement of PCNA mRNA revealed no difference in expression between control of KANK3 depleted cells (fold change control EC vs. siRNA1-*KANK3* ± std dev 0.71 ± 0.11, siRNA2-*KANK3* 1.08 ± 0.07). Therefore, the increased rate of gap closing observed in KANK3 depleted HUVEC appears to be primarily driven by an increased migratory capacity, as opposed to an increased rate of proliferation.
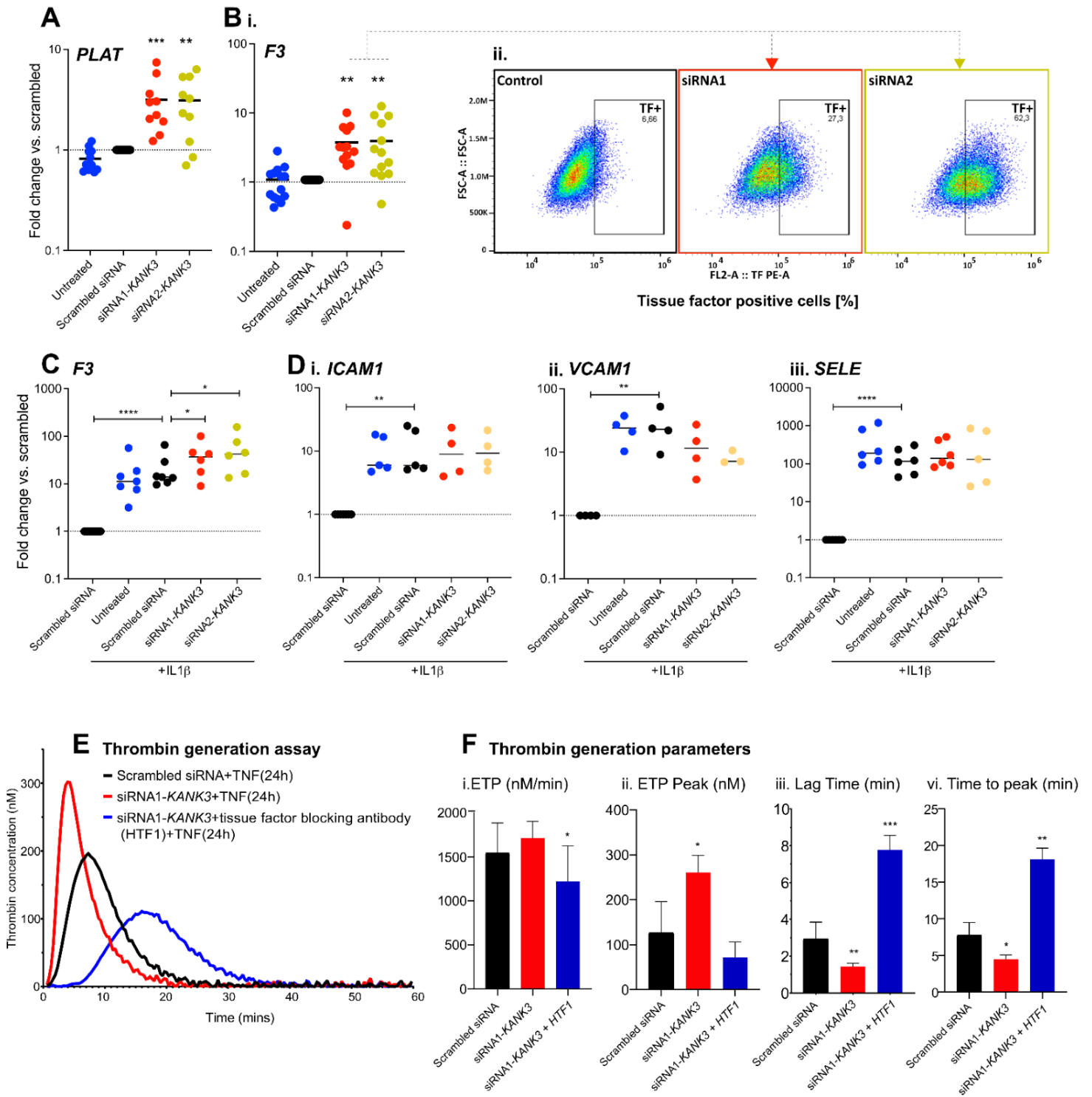
20

Figure 5: Effects of endothelial KANK3 knockdown on coagulation related proteins. HUVEC were untreated, or transfected with scrambled control siRNA, siRNA1-*KANK3* or siRNA2-*KANK3*. Measurement of: (**A**) *PLAT* or (**B**) (i) *F3* mRNA expression by qPCR, or (ii) cell surface tissue factor protein by flow cytometry. HUVEC were treated with or without IL1β (10 ng/ml) before measurement of mRNA encoding for (**C**) F3, or (**D**) the cytokine responsive adhesion molecules (i) *ICAM1*, (ii) *VCAM1*, (iii) *SELE*. (**E**) Calibrated automated thrombogram (CAT) assay was used to assess the thrombin generation potential of HUVEC treated with TNF (10 ng/ml) for 24 hours with or without pre-incubation with a function blocking anti-tissue factor antibody (HTF1). (F) Bar plots show the (i) total endogenous thrombin potential (ii) maximum endogenous thrombin potential (iii) lag time until the beginning of thrombin production (iv) time until peak thrombin production. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ vs. scrambled control.

KANK3 effects tissue factor and tissue type plasminogen activator expression

As other members of the KANK family have potential functions beyond cytoskeletal regulation, e.g. KANK4 can regulate VEGFR2 signalling via its interaction with talin [51], we screened untreated, scrambled siRNA transfected and KANK3 depleted EC for expression profiles of the following gene panels, which are related to EC specialised functions: (i) *Coagulation related*: Factor 8 (*F8*), protein C receptor (*PROCR*), protein S (*PROS*), tissue factor (*F3*), tissue factor pathway inhibitor (*TFPI*), tissue plasminogen activator (*PLAT*) and von Willebrand Factor (*VWF*) (ii) *Inflammation related*: intracellular adhesion molecule 1 (*ICAM1*), vascular cell adhesion molecule 1 (*VCAM1*), E-selectin (*SELE*) and (iii) *angiogenesis related*: angiopoietin-1 receptor (TEK), angiopoietin-2 (ANGPT2), vascular endothelial growth factor A (VEGFA) and vascular endothelial growth factor receptor 1 (FLT1), and kinase insert domain receptor (KDR, also known as vascular endothelial growth factor receptor 2).

Two genes in the coagulation related panel, *PLAT* and *F3,* were expressed at higher levels in both siRNA1-*KANK3* and siRNA2-*KANK3* treated EC, compared to the scrambled control (mean fold change ± std dev *F3:* 4.1 ± 2.8, *PLAT*: 3.1 ± 1.9, both $p<0.05$) (Figure 5 A and B.i). No other genes tested were consistently elevated, or reduced, in KANK3 depleted EC (Figure S3). As the *F3* gene encodes for the protein tissue factor, which is the key initiator of the extrinsic coagulation cascade, we investigated its relationship with KANK3 further. In line with the changes observed at the transcript level, flow cytometry confirmed an increase in the cell surface expression of tissue factor (TF) protein on KANK3 depleted HUVEC, compared to the scrambled control (Figure 5 B.ii) (scrambled control: MFI 29489; 6.66% cells TF positive, siRNA1-*KANK3*: MFI 47548; 27.3% cells TF positive, siRNA2-*KANK3:* MFI 93347; 62.3% cells TF positive).

F3 expression is relatively low on resting EC, but strong induced by inflammatory cytokines [52], so we tested if KANK3 depletion would modify this response. HUVEC transfected with scrambled siRNA, siRNA1-*KANK3*, or siRNA2-*KANK3* were treated with the inflammatory cytokine IL-1β for

23

24h. IL-1β strongly induced *F3* mRNA expression in scrambled siRNA HUVEC (scrambled siRNA vs. scrambled siRNA +IL-1β [mean fold change ± std dev] 22.4 ± 18.7), an increase that was exacerbated further by KANK3 depletion (scrambled siRNA +IL-1β vs. siRNA1-*KANK3*+IL-1β [mean fold change ± std dev]: 2.1 ± 1.0, siRNA2-*KANK3*+IL-1β: 3.2 ± 1.6) (Figure 5C).

To test if this was a consequence of a general enhancement of IL-1B signalling, we measured the expression of EC adhesion molecules. As expected, all were induced by IL-1β treatment (Figure 5 D i-iii), but we did not observe any further increase in expression in KANK3 depleted EC (Figure 5 D i-iii). Thus, KANK3 depletion does not appear to enhance cytokine signalling *per se.*

F3 is also induced by the inflammatory cytokine TNF, which has many signalling pathways in common with IL-1β [53]. To test if there was a functional consequence of this enhanced expression of *F3* in KANK depleted EC, thrombin generation potential was measured using the calibrated automated thrombograph (CAT) (Figure 5E). Relative to scrambled siRNA control, the depletion of KANK3 resulted in enhanced thrombin generation, as shown by a representative curve (Figure 5E). Whilst no statistically significant difference in endogenous thrombin potential (ETP) was observed (Figure 5 D.i), response time (scrambled siRNA vs. siRNA1-*KANK3* [difference [min;%], ± std dev]: -1.5 min, -52.2 %), time to peak (scrambled siRNA vs. siRNA1-*KANK3* [difference [min;%], ± std dev]: -1.5 min, -52.2 % -3.3min; -42.2 %), and peak thrombin generation levels (scrambled siRNA vs. siRNA1-*KANK3* [difference [levels; %], ± std dev]: +134.04; +105.1%) were all enhanced in KANK3 depleted EC (Figure 5 F.i-iv). Similar results were observed with siRNA2-*KANK3* (Figure S3 B and C). When KANK3 depleted EC were pre-treated with tissue factor function blocking antibodies, these effects were largely abolished (Figure 5 E and F ii-vi), consistent with tissue factor being the driver behind the increased thrombin generation.

24

DISCUSSION

Previously, we generated the first prediction that *KANK3* was an EC enriched protein in the human [8]. Here we confirm that KANK3 has a body wide EC enrichment at both the transcript and protein level. Recent work in mice also showed KANK3 expression exclusively in the vasculature, whilst KANK1 was expressed at the basal side of epithelial cells in various tissues, and KANK2 was observed predominantly at the plasma membrane and/or in the cytoplasm of mesenchymal cells [54]. Thus, the strictly restricted cell type expression profile of KANK3 is unique among the KANK family, the other members of which are also expressed more broadly in the human [26]. The EC type specific profile of KANK3 could underlie the lack of studies on its function, as indeed a several studies regarding KANK3 primarily center on cancer and overlook EC types in experimental design [9, 22, 23, 55], presumably due to the fact that EC are a minority cell type within any given tissue [8]. In the absence of understanding the likely context (cell type) in which a protein functions, it can be challenging to functionally characterise it. Understudied proteins can be unattractive targets for functional analysis due to a lack, or unknown reliability, of research tools with which to investigate them [37, 38]. Here, we validated our in house generated anti-KANK3 antibody, using both over expression and knockdown systems, as antibody specificity and reliability can be a problematic issue [39, 40, 56]. A study from 2008 showed that this problem is systemic. Of 6000 tested antibodies, fewer than 3000 were able to bind their target correctly. Due to the increased use of commercial antibodies, without validation, this might mean that an entire project could be based on artifacts. Due to the widespread utilisation of research antibodies, this is potentially a billion-dollar problem, with approximately 1.7 billion USD that have been wasted to antibodies in 2019. Hence, validation of any antibody used in a project is an essential step in research [57, 58]. The KANK family have a unique shared structure, consisting of a small N-terminal motif ("KN-motif"), C-terminal coiled-coil domains and ankyrin repeats (Zhu et al. 2008; Kakinuma et al. 2009).

25

The protein structure of KANK3 would indicate a function within cytoskeletal organization and in focal adhesions, due to the presence of a talin binding domain. It consists of a liprin binding domain and a KIF21A binding domain, which links them to the cytoskeleton [11, 59]. Immunofluorescence staining of endogenous KANK3 expression in EC, including HUVEC and mouse LSEC, reveals a distinctive punctate pattern within the cytoplasm. Notably, there is an accumulation of KANK3 at sites of cell-cell interactions in murine LSEC. This supports the role of KANK3 being linked to cell adhesions and to the basement membrane, which are important for integrity of the endothelial layer under stressful conditions such as flow or inflammation [60, 61]. Crosstalk between focal adhesion proteins and cell junction proteins has been previously described in the regulation of endothelial barrier function [62]. Considering the critical role of EC barrier function and the established involvement of the cytoskeleton and focal adhesions in the maintenance of cellular junctions [16], it could be speculated that KANK3 has a role in anchoring EC to the extracellular matrix and neighbouring cells. Furthermore, KANK3 might play a role in cell junction processes in some vascular beds.

EC shear forces have been shown to modulate various biochemical processes, in addition to cellular morphology and reorganisation of the cytoskeleton [63]. Shear stress has previously been shown to induce focal assembly, recruit signalling complexes to FA, and induce redistributions in stress fibres in EC [14]. We observe a significant upregulation of KANK3 gene and protein expression following EC exposure to shear stress. Similar mechanoregulation has been previously reported for different focal adhesion proteins through vinculin-vinculin regulation [64]. Laminar shear stress has been shown to induce integrin expression [65], α-actinin recruitment [66] and to mediate redistribution of intracellular stress fibers. Consequently, under the influence of this shear stress, there is a force-dependent alteration in the dynamics of FAs due to enhancement in actin fibers. The enhancement in actin fibers results in the sustenance or growth of the FAs connected to them [67]. FA have been shown to be involved in matrix-adhesion and mechanosensing in EC

26

[14], and vascular resistance in SMC [68]. Due to its vascular specificity, one could speculate, that KANK3 is involved in similar vascular functions.

We also observed that KANK3 depletion had a marked effect on the distribution of vimentin. This did not appear to be driven by a modification in VIM transcription, and could therefore be driven by a direct or indirect KANK3-vimentin protein-protein interaction, such as is observed with other cytoskeletal components, e.g., actin and myosin [69] or intermediate filaments and microtubules [70]. Vimentin is a major intermediate filament in EC [44], which has roles in cell migration and polarity, cell structure and integrity, response to mechanical stress and EC differentiation [71-74] and has been described as integral to cell adhesion and EC sprouting [44]. Although there is a lack of research regarding the interactions between KANK proteins and vimentin, it's worth noting that vimentin plays a role in cell migration. Previous research has indicated its involvement in determining cellular polarity, regulating the formation of cell contacts, and organizing and transporting signalling proteins that contribute to cell motility [75]. Vimentin increases cell stiffness and promotes cell migration when cells are densely cultured. However, its impact on the migration of cells plated sparsely is minimal or negligible [76].

Here we show that KANK3 depletion increases EC migratory capacity. Previously, talin has been identified as a regulator of cell motility [77] and the stability of talin rods has been shown to control cell migration [78]. It could be reasonably assumed that KANK3 migratory control is driven through KANK3-talin interactions. Although to our knowledge, KANK3 has not been studied in the context of EC motility, previous reports showed it had a role in the regulation of oxygen dependent suppression of cell motility in hepatocellular carcinoma cells [22], and the inhibition of invasion and migration of lung adenocarcinoma [23] and was therefore considered a valuable target in cancer research. The increase of cell motility in KANK3 depleted cells could suggest a protective effect against shear forces and increase in matrix adhesion, similar to the contribution of vimentin networks to the stiffening of cells, which allows them to withstand mechanical forces [79].

27

Focal adhesions play a pivotal role in relaying mechanical forces and external signals to intracellular pathways. Disturbed multidirectional shear stress of the vasculature is recognised for its role in triggering the activation of atherogenic and thrombogenic genes in EC and SMC [80]. Hence, it is plausible that KANK3 mediated modifications in the cytoskeleton might initiate signalling pathways linked to the regulation of genes associated with coagulation. FA activation in vascular smooth muscle cells regulates arterial stiffness and procoagulant properties of the vessel wall. Their findings revealed a decrease in thrombin generation potential of vascular smooth muscle cells (VSMCs) as the matrix stiffness increases On a rigid matrix, the presence of αvb3 integrin within the FA complex diminishes the accessibility of binding sites for prothrombin. As a consequence, this leads to a reduction in the generation of thrombin on VSMCs. Conversely, it could be hypothesized that this outcome is reversed when dealing with a less rigid matrix [81]. This connection underscores the intricate interplay between FA signalling, vascular mechanics, and thrombotic potential.

In summary, our study provides insight into the function of KANK3 in the vascular compartment. Our findings are consistent with it having EC specific functions, in line with its enriched expression profile in this cell type.

28

METHODS

Tissue profiling

Protein profiling was performed as part of the Human Protein Atlas (HPA) project. Tissue sections from breast, adipose tissue, cortex, thyroid gland, colon, kidney, liver, prostate, epididymis, duodenum, bronchus, testis, endometrium, cervix, appendix, stomach, oesophagus, lung, pancreas and ovary were generated and stained, as previously described (Pontén, Jirström, and Uhlen 2008; Uhlen et al. 2015). Briefly, formalin fixed, and paraffin embedded tissue samples were sectioned, de-paraffinized in xylene, hydrated in graded alcohols and blocked for endogenous peroxidase in 0.3% hydrogen peroxide diluted in 95% ethanol. For antigen retrieval, a Decloaking chamber® (Biocare Medical, CA) was used. Slides were boiled in Citrate buffer®, pH6 (Lab Vision, CA). Primary antibody against KANK3 (HPA051153) and a dextran polymer visualization system (UltraVision LP HRP polymer®, Lab Vision) were incubated for 30 min each at room temperature and slides were developed for 10 min using Diaminobenzidine (Lab Vision) as the chromogen. Slides were counterstained in Mayers haematoxylin (Histolab) and scanned using Scanscope XT (Aperio).

Isolation and culture of human umbilical vein endothelial cells

Ethical approval for endothelial cell isolation and subsequent experimentation was granted by *Regionala etikprövningsnämnden i Stockholm* (diarienummer 2015/1294-31/2). Human umbilical vein endothelial cells (HUVEC) were isolated from human umbilical cords, collected from Karolinska Hospital (Stockholm, Sweden) and from the University Hospital of Northern Norway (UNN; Tromsø, Norway), as previously described (Cooke et al. 1993). HUVEC were cultured in Medium M199, supplemented with 10% fetal bovine serum (FBS) (or 0.5% FBS in some experiments), 10ml/l Penicillin-Streptomycin, 2.5mg/l Amphotericin B (all ThermoFisher, Gibco), 1mg/l Hydrocortisone 1µg/l and human Epidermal Growth Factor (hEGF) (both Merck). In some

29

experiments, EC were cultured under laminar shear stress (4dyn or 40dyn) for 48 hours in flow chamber slides (μ-slide VI 0.4, Ibidi), integrated into an Ibidi flow pump system.

HEK293 cells were obtained in frozen vials from ATCC (HEK-293 CRL-1573) and cultured in DMEM Cell culture medium supplemented with 10ml/l Penicillin-Streptomycin and 10% fetal bovine serum (FBS).

Mouse liver sinusoidal endothelial cells (mLSEC) were gifted from *Vascular Biology Research Group (VBRG) at UiT The Arctic University of Norway*, isolated and cultured in RPMI 1640 supplemented with L-Glutamine (300mg/l), 10ml/l Penicillin-Streptomycin, as previously described [82].

siRNA transfection

HUVECs and HEK239 cells were transfected with siRNA sequences targeting KANK3 (silencer select siRNA s230059, s230061, ThermoFisher) or Silencer Select negative control siRNA (ThermoFisher: 4390843). Transfection was performed using Lipofectamine RNAiMAX transfection reagent (Invitrogen), according to manufacturer instructions, at 60-80% confluency in Opti-MEM reduced serum medium (ThermoFisher, Gibco) without additives for 4h. Medium was changed to standard cell culture medium. Knockdown efficiency was accessed after 48h by qPCR, Western blot, or immunofluorescence staining.

Recombinant KANK3-eGFP protein expression

Transient transfection of vector coding for KANK3_eGFP (GenScript) into HEK293 (ATCC: CRL-3216) cells was done using Lipofectamine 3000 (ThermofisherScientific), according to the manufacturer's instructions. Plasmid transfection was performed at cell confluency of 60-80% in Opti-MEM reduced serum medium without additives for 5h using Lipofectamine 3000 transfection reagent (Invitrogen), according to manufacturer instructions. 48 h after transfection, transfected cells were lysed with RIPA cell Lysis Buffer and sample was frozen, or cells were fixed for further analyses.

30

RNA sequencing

RNA isolation and purification was performed using the RNAeasy mini kit (Qiagen). RNA concentration was measured using Nanodrop 2000 spectrophotometer and RNA integrity number (RIN) determined using Agilent 2100 Bioanalyzer (RIN>9 required for inclusion). Library preparation and RNA sequencing was performed by the National Genomics Infrastructure Sweden (NGI) using Illumina stranded TruSeq poly-A selection kit and Illumina NovaSeq6000S (4 lanes, 2x 150bp reads, incl 2Xp kits). The data was processed using demultiplexing. Data storage and initial analyses were performed using server sided computation supplied by the Swedish National Infrastructure for Computing (SNIC). Genome assembly used for sequence alignment: Homo_sapiens.GRCh38.dna.primary_assembly.fa and annotation performed using: Homo_sapiens.GRCh38.96.gtf. Sequence alignment was carried out using STAR/2.5.3a. Gene mapping has been carried out using subread/1.5.2 and the module feature counts. Transcript mapping carried out using Salmon/0.9.1.

Gap closing ("scratch") assay

A ´gap´ was created in a confluence EC monolayer using a 100 μl pipette tip. Gap size was monitored with an Olympus IXplore Live microscope in phase/contrast mode in 10x magnification, with cells in a 37°C, 5% CO_2 on stage incubator chamber. Gaps were imaged every 30min for 96h. Gap size was measured every 6h in Fiji using ImageJ2 graphics procession software.

Shear stress exposure

Endothelial cells were cultured in flow chamber slides (μ-slide VI 0.4, Ibidi) until confluence. The slide was connected to an Ibidi flow pump system and cultured under laminar shear stress (4dyn or 40dyn) for 48 hours. The cells were then lysed for qPCR, Western blot, and fixed for confocal microscopy.

qPCR

Cell lysis and cDNA creation were performed using the 2-Step Fast-Cells-to-CT-Kit (Invitrogen, ThermoFisher) according to their protocols. qPCR was performed using TaqMan Fast Universal

31

PCR mix. Target primer conjugated to FAM-probe (4448892, ThermoFisher) was used to access KANK3 levels. 18s rRNA (4319413E conjugated to VIC probe, ThermoFisher) was used as endogenous control. qPCR was performed using a RealTime PCR LightCycler 96 ® system (Roche Life Sciences).

SDS-PAGE and Western blot

Lysate was analysed for KANK3 expression by western blotting with rabbit polyclonal anti-KANK3 antibody (1:250, HPA051153) and horseradish peroxidase (HRP)-coupled goat anti-rabbit antibody (1:2000, Dako). After chemiluminescence detection, membrane was washed, incubated in stripping buffer, and analysed for GAPDH housekeeping gene expression. Additionally, mouse monoclonal anti-eGFP antibody (1:1000) with secondary HRP-coupled goat anti-mouse antibody (1:2000, Dako), were used.

Flow cytometry

Endothelial cells were cultured in 6 well plates and transfected with 2 different siRNAs targeting KANK3 or a scrambled siRNA control 48h prior 2 harvesting cells. 4h prior to harvest cells were stimulated with 10ng/ml TNF. Cells were harvested by trypsin digest followed by centrifugation (x350g, 7min) and separation, decantation of supernatant and resuspension of EC in ice-cold PBS (Gibco, ThermoFisher). Cells were split into two tubes and treated with PE-conjugated anti-CD142 Clone NY2 (30 µl/ml) and isotype-matched control mouse-IgG1 (6 µl/ml) and incubated on ice for 30min, followed by centrifugation (x350g, 7min), decantation of supernatant and resuspension of cells in PBS. Flow cytometry was performed in Beckman Coulter CytoFLEX Flow Cytometer (acquisition settings FSC 20V, SSC 150V, PE 130V). Gating and data analysis was performed using CytExpert for CytoFLEX Acquisition and Analysis Software and FlowJo™ v10.7. Gating was performed for live vs dead cells and singlets vs doublets. Dead cells and doublets were excluded, followed by gating for TF positive and TF negative cells. Isotype control signal was subtracted from full stain for each sample and median fluorescence intensity (MFI) and TF positive cells (%) were identified for each condition.

32

Confocal microscopy

Cells were fixed in 4 % paraformaldehyde in PBS, permeabilised in 0.5% triton X-100 and blocked using 5% BSA. Primary antibodies against KANK3 (HPA051153, Atlas antibodies) and vimentin (OMA1-06001, Invitrogen) were incubated on cells for 20 minutes, followed by FITC-conjugated anti-rabbit antibody (F-9887, Sigma), Alexa 555-conjugated anti mouse IgG, and either TRITC-conjugated (P1951, Sigma) or Atto-647N-conjugated (65906, Sigma) phalloidin, depending on experiment, and were then coated in mounting medium (VectaShield) containing DAPI nuclear stain for storage and imaging. Images were taken using a Leica TP5 SP5 confocal microscope and image analysis was performed in Fiji ImageJ2 graphics procession software.

Structured Illumination Microscopy and Deconvolution imaging

48h after siRNA or plasmid transfection, cells were plated on fibronectin coated (1 µg/cm^2) #1.5 glass coverslips (Zeiss) subconfluently (30-50k cells per cm^2) and cultivated for 1h (HEK293 cells) or 4h (HUVECs). Afterwards, cells were fixed in 4% paraformaldehyde (Merck, Sigma) in PBS (Dulbecco, Sigma) for 20 min, washed with PBS (Dulbecco, Sigma) and left in PBS until further analysis. Samples were permeabilised in 0.05% Triton X-100 (Sigma) in PBS and blocked in 3% BSA in PBS. Primary antibody from Rabbit against KANK3 (HPA051153, Atlas antibodies) was prepared in blocking buffer, followed by incubation with phalloidin conjugated to atto-647, secondary anti rabbit IgG conjugated to Alexa 555 and anti-mouse IgG conjugated to Alexa 488 for 30 minutes at RT, nuclear stain was performed with DAPI for 20 minutes at RT in the dark. Samples were mounted using hardset antifade mounting medium (VectaShield). Images were taken in an OMX Blaze SIM microscope using a 60X 1.42NA oil-immersion objective (GE Healthcare; Olympus). 3D-SIM images stacks of up to 3 µm were acquired every 125 nm in five phases and three angles, resulting in 15 raw images per z-plane and total of 24 focal planes. Reconstruction used SoftWoRx software (GE Healthcare). Image analysis was performed in Fiji ImageJ2 graphics procession software.

33

Calibrated automated thrombinoscope (CAT) assay

HUVECs were cultivated until confluency in medium M199 and, if stated, knocked down as described above. Cells were transferred to flat-bottom 96 well plates (VWR) treated with tumour necrosis factor alpha (TNF; 10 ng/mL) (ThermoFisher) for 24h before thrombin generation assay. After washing the cells with PBS, thrombin formation was initiated in 120 µL reaction mixtures containing human citrated plasma, 4 µM phospholipids (Thrombinoscope BV), 16.6 mM Ca2 + and 2.5 mM fluorogenic substrate (Z-Gly-Gly-Arg-AMC, Thrombinoscope BV). As controls, Tissue factor (1 pM, Dade Innovin), mouse monoclonal anti-TF antibody (12.5 µg/ml, HTF-1, BD Pharmingen) or corn trypsin inhibitor were added 15 min before adding the substrate. All real time thrombin formation experiments were run in triplicates. Thrombin generation was quantified using the Thrombinoscope software package (Version 5.0.0.742) that reported means ± SD.

Gene ontology analysis

The Gene Ontology Consortium [33] and PANTHER classification resource (Mi, 2019) were used to identify overrepresented terms in gene lists using the GO databases (release date 2023-07-05). Plots of GO terms were created using the R package clusterProfiler [83].

Data usage and analysis

Human tissue RNAseq data was retrieved from Genotype-Tissue Expression (GTEx) portal V8 (www.gtexportal.org) [84]. Statistical analyses were done in RStudio (R V 4.0.3), using the corr.test function from the additional package psych (V 2.0.12) (Pearson correlation coefficient). Single Cell sequencing data was sourced from data collected into the Tabula Sapiens [85].

Software: Image analyses

Image analysis was performed using ImageJ2 Fiji using ImageJ2 graphics procession software [86].

**Software: Graphs, Figures and Tables**

Graphs and calculation tables were created using GraphPad Prism (V. 8.4.3) and Microsoft Excel 2019 (Office 365). Figures were created in Photoshop.
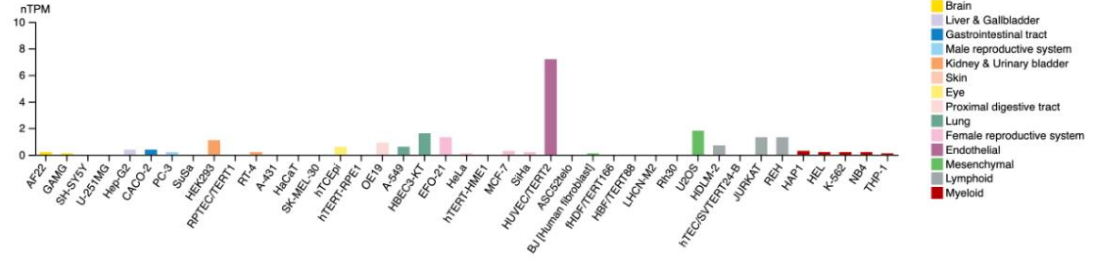
34

Software: Statistical analyses

Statistical analyses were performed in RStudio (R version 4.0.3) using the following additional packages: psych, readr dplyr, data.table and tools.
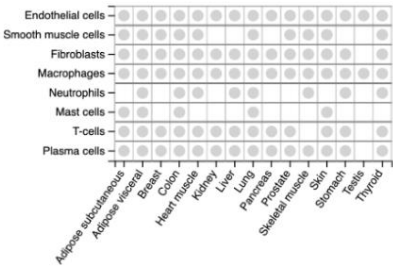
35

A KANK3
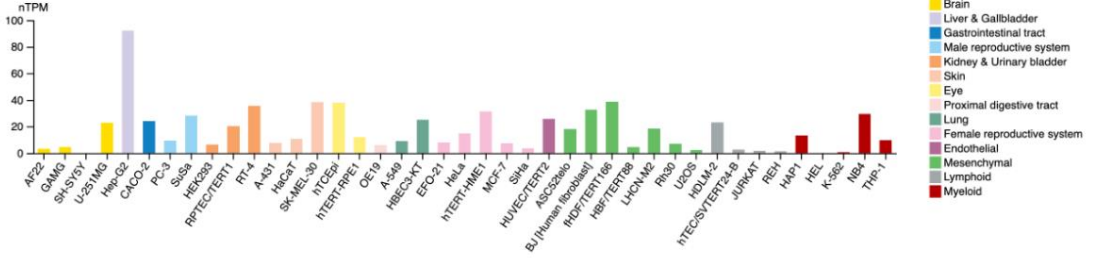
i. Cell type enrichment
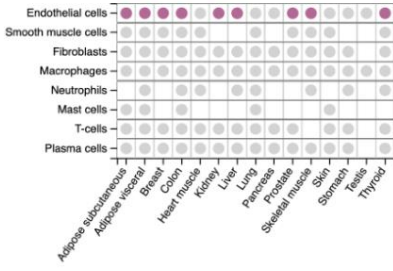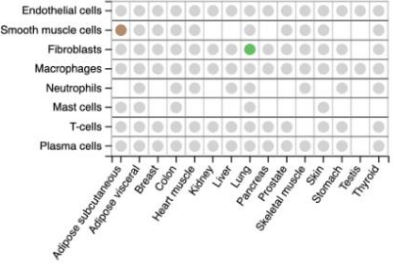
ii. Cell line expression

B KANK1

i. Cell type enrichment
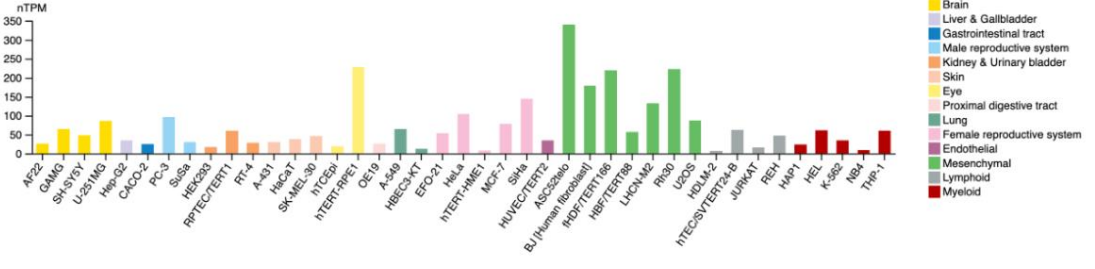
ii. Cell line expression

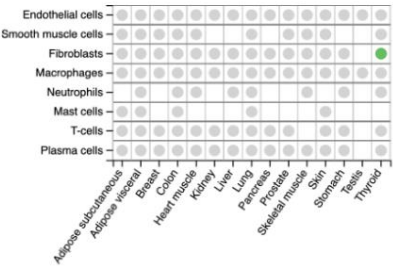# C KANK2

i. Cell type enrichment

ii. Cell line expression

D KANK4

i. Cell type enrichment
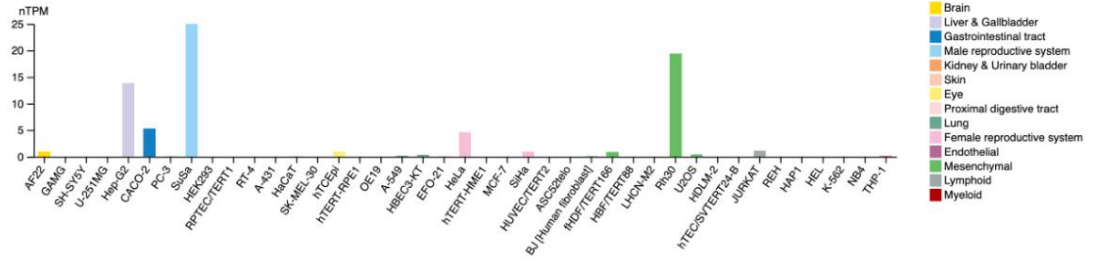
ii. Cell line expression

Figure S1. Enrichment of KANK family members in human tissues and immortalized cell lines. Expression profiles for (**A**) *KANK3*, (**B**) *KANK1*, (**C**) *KANK2* and (**D**) *KANK4* in: (i) human cell types profiled using bioinformatic based analysis of bulk RNAseq data [26], or in (ii) immortalised cell lines [35].
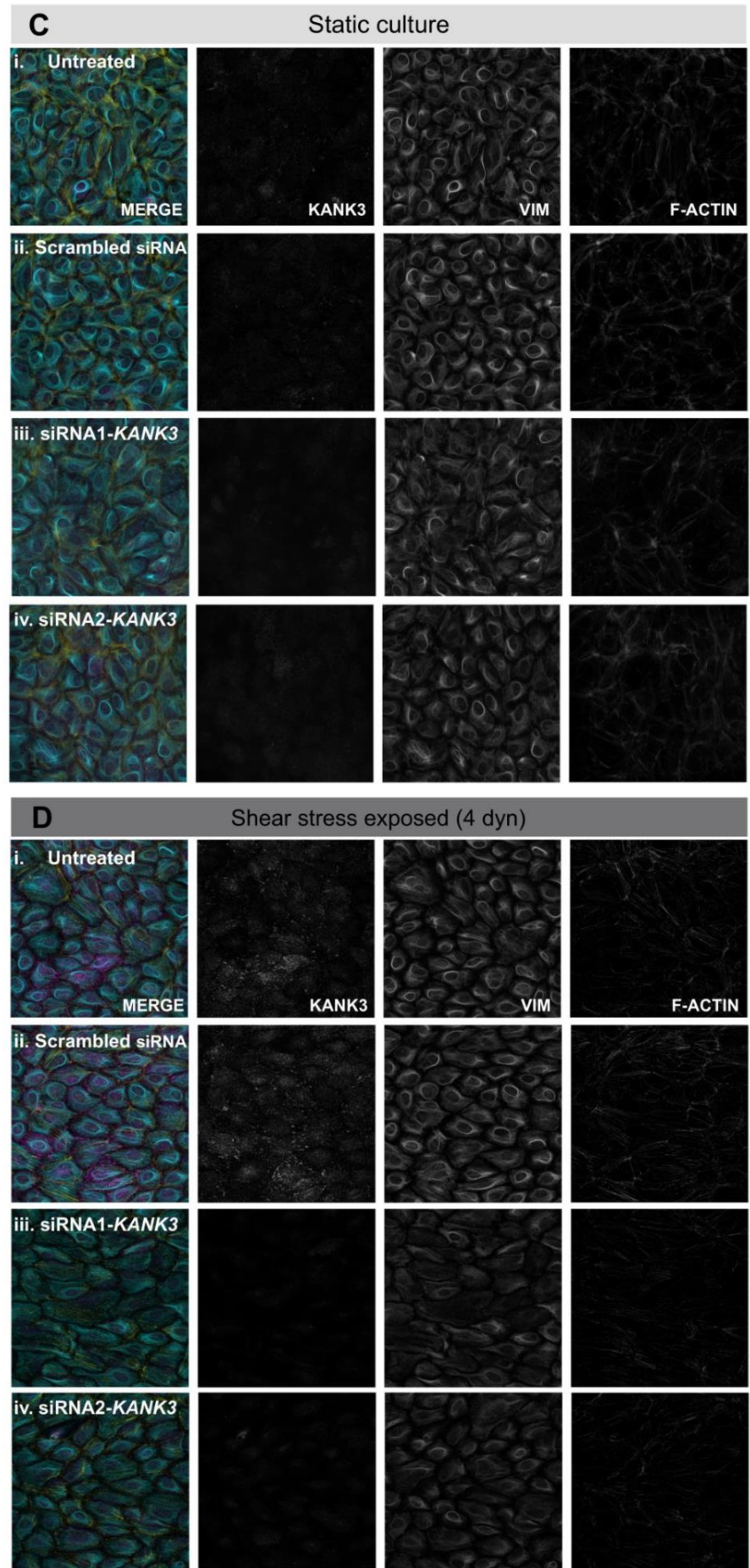
37

A HUVEC Z-stack

MERGE Z-STACK

F-ACTIN

KANK3

B mLSEC

MERGE

DAPI

KANK3

F-ACTIN

C Static culture

i. Untreated

MERGE

KANK3

VIM

F-ACTIN

ii. Scrambled siRNA

iii. siRNA1-*KANK3*

iv. siRNA2-*KANK3*

D Shear stress exposed (4 dyn)

i. Untreated

MERGE

KANK3

VIM

F-ACTIN

ii. Scrambled siRNA

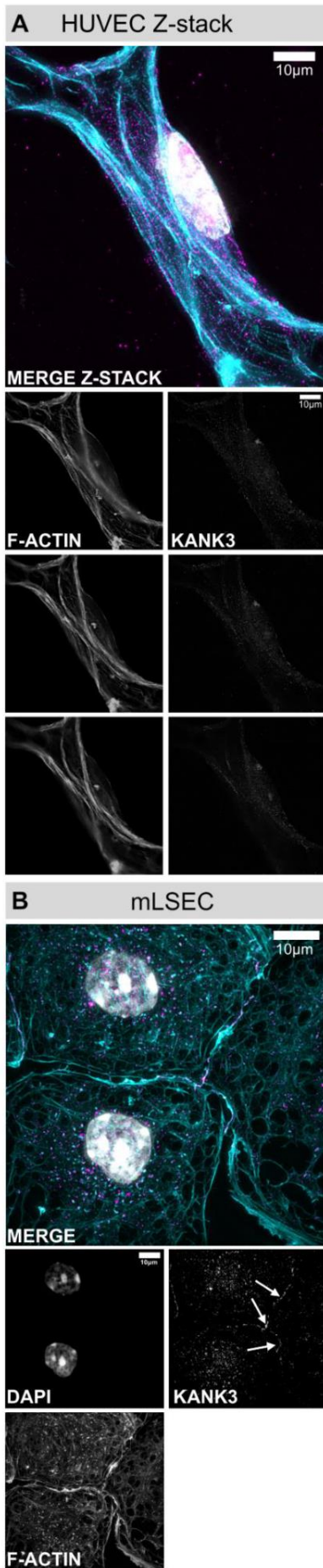iii. siRNA1-*KANK3*

iv. siRNA2-*KANK3*

Figure S2. Subcellular location of KANK3 in static and flow culture: Immunocytochemistry staining using anti-KANK3 antibody (magenta), DAPI nuclear stain (grey), and phalloidin F-actin stain (cyan) for (A) HUVEC Z stack average (B) mLSEC. (C,D) Immunocytochemistry staining using anti-KANK3 antibody (magenta), anti-vimentin antibody (yellow), DAPI nuclear stain (grey), and phalloidin F-actin stain (cyan) in (C) static cultured and (D) shear stress exposed HUVEC.
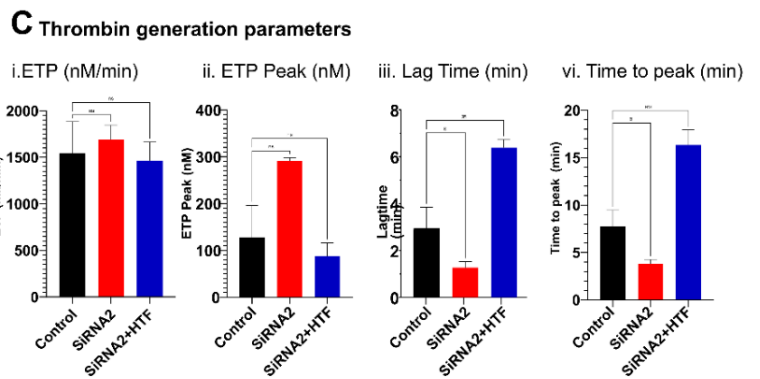
39

A

i. *PROCR*
ii. *PROS*
iii. *TFPI*
iv. *ICAM1* *

v. *VCAM1*
vi. *SELE*
vii. *VWF*
viii. *TEK*

ix. *ANGPT2*
x. *VEGFA*
xi. *FLT1*
xii. *KDR*

B Thrombin generation assay

C Thrombin generation parameters

i. ETP (nM/min)
ii. ETP Peak (nM)
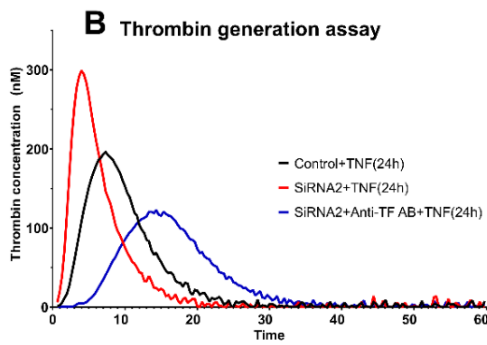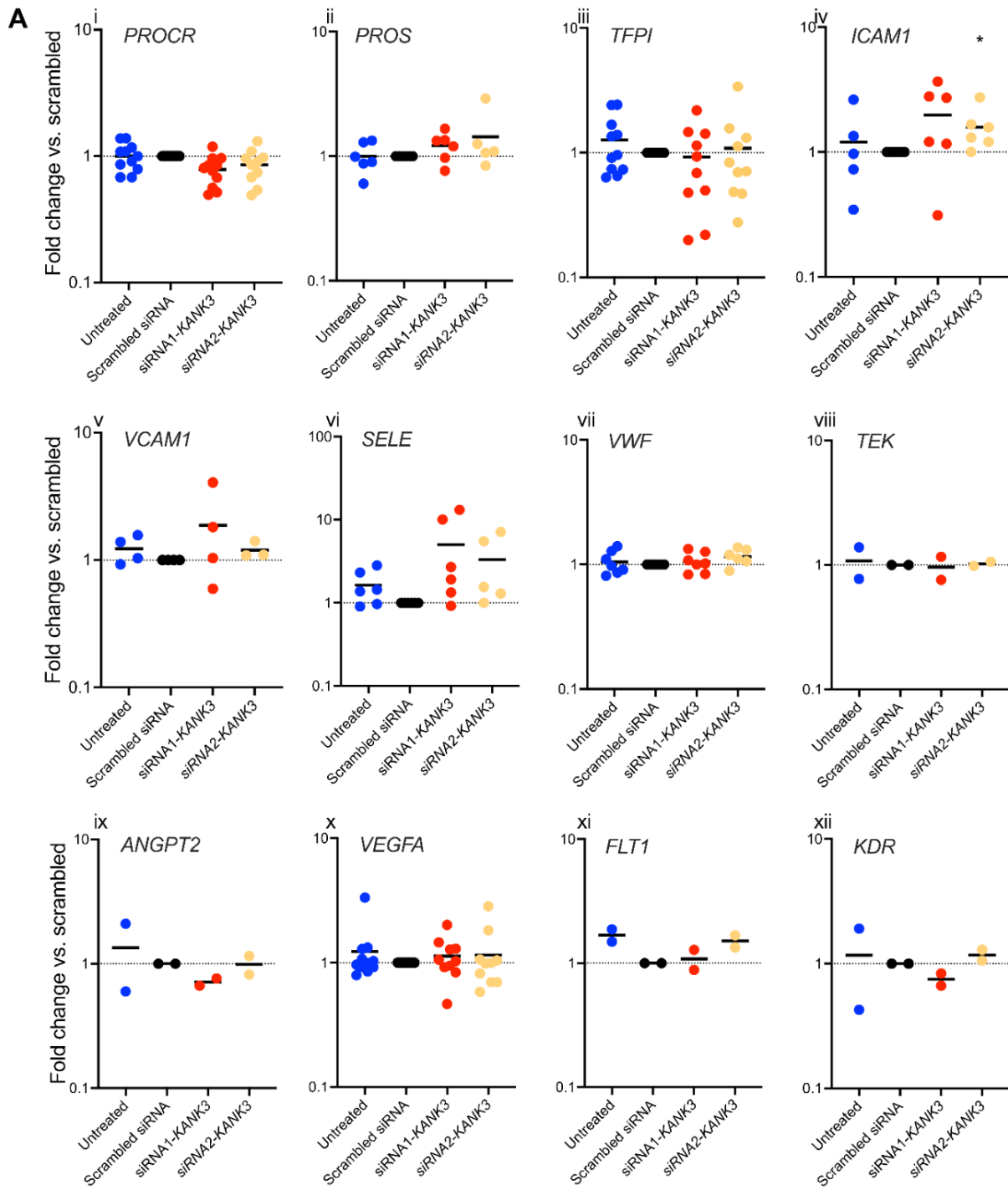iii. Lag Time (min)
vi. Time to peak (min)

Figure S3: Effect of KANK3 depletion on gene expression. (A) HUVEC were untreated, or transfected with scrambled control siRNA, siRNA1-*KANK3* or siRNA2-*KANK3* before measurement of mRNA level of genes indicated by qPCR. *p<0.05 vs. scrambled control. (**B**) Calibrated automated thrombogram (CAT) assay was used to assess the thrombin generation potential of HUVEC treated with TNF (10 ng/ml) for 24 hours with or without pre-incubation with a function blocking anti-tissue factor antibody (HTF1). (C) Bar plots show the (i) total endogenous thrombin potential (ii) maximum endogenous thrombin potential (iii) lag time until the beginning of thrombin production (iv) time until peak thrombin production. *** p<0.001, ** p<0.01, *p<0.05 vs. scrambled control.

41

REFERENCES

1. Galley, H.F. and N.R. Webster, *Physiology of the endothelium.* Br J Anaesth, 2004. **93**(1): p. 105-13.
2. Pober, J.S. and W.C. Sessa, *Evolving functions of endothelial cells in inflammation.* Nat Rev Immunol, 2007. **7**(10): p. 803-15.
3. Dejana, E., *Endothelial cell–cell junctions: happy together.* Nature Reviews Molecular Cell Biology, 2004. **5**(4): p. 261-270.
4. Gavard, J., *Endothelial permeability and VE-cadherin: a wacky comradeship.* Cell Adh Migr, 2014. **8**(2): p. 158-64.
5. Scalise, A.A., et al., *The blood-brain and gut-vascular barriers: from the perspective of claudins.* Tissue Barriers, 2021. **9**(3): p. 1926190.
6. Hirata, K., et al., *Cloning of an immunoglobulin family adhesion molecule selectively expressed by endothelial cells.* J Biol Chem, 2001. **276**(19): p. 16223-31.
7. Melincovici, C.S., et al., *Vascular endothelial growth factor (VEGF) - key factor in normal and pathological angiogenesis.* Rom J Morphol Embryol, 2018. **59**(2): p. 455-467.
8. Butler, L.M., et al., *Analysis of Body-wide Unfractionated Tissue Data to Identify a Core Human Endothelial Transcriptome.* Cell Syst, 2016. **3**(3): p. 287-301 e3.
9. Tadijan, A., et al., *KANK family proteins in cancer.* Int J Biochem Cell Biol, 2021. **131**: p. 105903.
10. Kakinuma, N., et al., *Kank proteins: structure, functions and diseases.* Cell Mol Life Sci, 2009. **66**(16): p. 2651-9.
11. Bouchet, B.P., et al., *Talin-KANK1 interaction controls the recruitment of cortical microtubule stabilizing complexes to focal adhesions.* Elife, 2016. **5**.
12. Sun, Z., et al., *Kank2 activates talin, reduces force transduction across integrins and induces central adhesion formation.* Nature Cell Biology, 2016. **18**(9): p. 941-953.
13. Guo, Q., et al., *Structural basis for the recognition of kinesin family member 21A (KIF21A) by the ankyrin domains of KANK1 and KANK2 proteins.* J Biol Chem, 2018. **293**(2): p. 557-566.
14. Katoh, K., Y. Kano, and S. Ookawara, *Role of stress fibers and focal adhesions as a mediator for mechano-signal transduction in endothelial cells in situ.* Vasc Health Risk Manag, 2008. **4**(6): p. 1273-82.
15. Ramjaun, A.R. and K. Hodivala-Dilke, *The role of cell adhesion pathways in angiogenesis.* Int J Biochem Cell Biol, 2009. **41**(3): p. 521-30.
16. Wu, M.H., *Endothelial focal adhesions and barrier function.* J Physiol, 2005. **569**(Pt 2): p. 359-66.
17. Vouret-Craviari, V., et al., *Regulation of the actin cytoskeleton by thrombin in human endothelial cells: role of Rho proteins in endothelial barrier function.* Mol Biol Cell, 1998. **9**(9): p. 2639-53.
18. Lee, J., et al., *Endothelial Cell Focal Adhesion Regulates Transendothelial Migration and Subendothelial Crawling of T Cells.* Front Immunol, 2018. **9**: p. 48.
19. Boggetti, B., et al., *NBP, a zebrafish homolog of human Kank3, is a novel Numb interactor essential for epidermal integrity and neurulation.* Dev Biol, 2012. **365**(1): p. 164-74.
20. Hensley, M.R., et al., *Evolutionary and developmental analysis reveals KANK genes were co-opted for vertebrate vascular development.* Sci Rep, 2016. **6**: p. 27816.
21. Zhu, Y., et al., *Kank proteins: a new family of ankyrin-repeat domain-containing proteins.* Biochim Biophys Acta, 2008. **1780**(2): p. 128-33.
22. Kim, I., et al., *A novel HIF1AN substrate KANK3 plays a tumor-suppressive role in hepatocellular carcinoma.* Cell Biol Int, 2018. **42**(3): p. 303-312.

42

23. Dai, Z., et al., *KANK3 mediates the p38 MAPK pathway to regulate the proliferation and invasion of lung adenocarcinoma cells.* Tissue Cell, 2023. **80**: p. 101974.

24. Norreen-Thorsen, M., et al., *A human adipose tissue cell-type transcriptome atlas.* Cell Rep, 2022. **40**(2): p. 111046.

25. Dusart, P., et al., *A Systems-Based Map of Human Brain Cell-Type Enriched Genes and Malignancy-Associated Endothelial Changes.* Cell Rep, 2019. **29**(6): p. 1690-1706.e4.

26. Dusart, P., et al., *A tissue centric atlas of cell type transcriptome enrichment signatures.* bioRxiv, 2023: p. 2023.01.10.520698.

27. Öling, S., et al., *A human stomach cell type transcriptome atlas.* bioRxiv, 2023: p. 2023.01.10.520700.

28. Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.

29. Usuba, R., et al., *EGFL7 regulates sprouting angiogenesis and endothelial integrity in a human blood vessel model.* Biomaterials, 2019. **197**: p. 305-316.

30. Dai, C., et al., *Regulatory mechanisms of Robo4 and their effects on angiogenesis.* Biosci Rep, 2019. **39**(7).

31. Ishida, T., et al., *Targeted disruption of endothelial cell-selective adhesion molecule inhibits angiogenic processes in vitro and in vivo.* J Biol Chem, 2003. **278**(36): p. 34598-604.

32. Sauteur, L., et al., *Cdh5/VE-cadherin promotes endothelial cell interface elongation via cortical actin polymerization during angiogenic sprouting.* Cell Rep, 2014. **9**(2): p. 504-13.

33. Ashburner, M., et al., *Gene Ontology: tool for the unification of biology.* Nature Genetics, 2000. **25**(1): p. 25-29.

34. Tabula Sapiens, C., et al., *The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans.* Science, 2022. **376**(6594): p. eabl4896.

35. Uhlén, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.

36. Dusart, P., et al., *A Systems-Based Map of Human Brain Cell-Type Enriched Genes and Malignancy-Associated Endothelial Changes.* Cell Rep, 2019. **29**(6): p. 1690-1706 e4.

37. Kustatscher, G., et al., *An open invitation to the Understudied Proteins Initiative.* Nat Biotechnol, 2022. **40**(6): p. 815-817.

38. Kustatscher, G., et al., *Understudied proteins: opportunities and challenges for functional proteomics.* Nat Methods, 2022. **19**(7): p. 774-779.

39. Baker, M., *Reproducibility crisis: Blame it on the antibodies.* Nature, 2015. **521**(7552): p. 274-6.

40. Weller, M.G., *Quality Issues of Research Antibodies.* Anal Chem Insights, 2016. **11**: p. 21-7.

41. Stoeger, T., et al., *Large-scale investigation of the reasons why potentially important genes are ignored.* PLoS Biol, 2018. **16**(9): p. e2006643.

42. Rafiq, N.B.M., et al., *A mechano-signalling network linking microtubules, myosin IIA filaments and integrin-based adhesions.* Nat Mater, 2019. **18**(6): p. 638-649.

43. Tzima, E., et al., *Activation of Rac1 by shear stress in endothelial cells mediates both cytoskeletal reorganization and effects on gene expression.* EMBO J, 2002. **21**(24): p. 6791-800.

44. Dave, J.M. and K.J. Bayless, *Vimentin as an integral regulator of cell adhesion and endothelial sprouting.* Microcirculation, 2014. **21**(4): p. 333-44.

45. Tsuruta, D. and J.C. Jones, *The vimentin cytoskeleton regulates focal contact size and adhesion of endothelial cells subjected to shear stress.* J Cell Sci, 2003. **116**(Pt 24): p. 4977-84.

46. Hoelzle, M.K. and T. Svitkina, *The cytoskeletal mechanisms of cell-cell junction formation in endothelial cells.* Mol Biol Cell, 2012. **23**(2): p. 310-23.

43

47.   Helmke, B.P., R.D. Goldman, and P.F. Davies, *Rapid displacement of vimentin intermediate filaments in living endothelial cells exposed to flow.* Circ Res, 2000. **86**(7): p. 745-52.

48.   Wechezak, A.R., R.F. Viggers, and L.R. Sauvage, *Fibronectin and F-actin redistribution in cultured endothelial cells exposed to shear stress.* Lab Invest, 1985. **53**(6): p. 639-47.

49.   Battaglia, R.A., et al., *Vimentin on the move: new developments in cell migration.* F1000Res, 2018. **7**.

50.   Dusart, P., et al., *A systems-approach reveals human nestin is an endothelial-enriched, angiogenesis-independent intermediate filament protein.* Sci Rep, 2018. **8**(1): p. 14668.

51.   Zhang, C., et al., *KANK4 Promotes Arteriogenesis by Potentiating VEGFR2 Signaling in a TALIN-1-Dependent Manner.* Arterioscler Thromb Vasc Biol, 2022. **42**(6): p. 772-788.

52.   Witkowski, M., U. Landmesser, and U. Rauch, *Tissue factor as a link between inflammation and coagulation.* Trends in Cardiovascular Medicine, 2016. **26**(4): p. 297-303.

53.   Ott, L.W., et al., *Tumor Necrosis Factor-alpha- and interleukin-1-induced cellular responses: coupling proteomic and genomic information.* J Proteome Res, 2007. **6**(6): p. 2176-85.

54.   Guo, S.S., et al., *Tissue distribution and subcellular localization of the family of Kidney Ankyrin Repeat Domain (KANK) proteins.* Exp Cell Res, 2021. **398**(1): p. 112391.

55.   Zhu, Y., et al., *Identification of three immune subtypes characterized by distinct tumor immune microenvironment and therapeutic response in stomach adenocarcinoma.* Gene, 2022. **818**: p. 146177.

56.   Fredolini, C., et al., *Systematic assessment of antibody selectivity in plasma based on a resource of enrichment profiles.* Sci Rep, 2019. **9**(1): p. 8324.

57.   Bradbury, A. and A. Plückthun, *Reproducibility: Standardize antibodies used in research.* Nature, 2015. **518**(7537): p. 27-29.

58.   Schonbrunn, A., *Editorial: Antibody can get it right: confronting problems of antibody specificity and irreproducibility.* Mol Endocrinol, 2014. **28**(9): p. 1403-7.

59.   Sun, Z., et al., *Kank2 activates talin, reduces force transduction across integrins and induces central adhesion formation.* Nat Cell Biol, 2016. **18**(9): p. 941-53.

60.   Davis, G.E. and D.R. Senger, *Endothelial Extracellular Matrix.* Circulation Research, 2005. **97**(11): p. 1093-1107.

61.   Thomsen, M.S., L.J. Routhe, and T. Moos, *The vascular basement membrane in the healthy and pathological brain.* J Cereb Blood Flow Metab, 2017. **37**(10): p. 3300-3317.

62.   Quadri, S.K., *Cross talk between focal adhesion kinase and cadherins: role in regulating endothelial barrier function.* Microvasc Res, 2012. **83**(1): p. 3-11.

63.   Mott, R.E. and B.P. Helmke, *Mapping the dynamics of shear stress-induced structural changes in endothelial cells.* Am J Physiol Cell Physiol, 2007. **293**(5): p. C1616-26.

64.   Carisey, A., et al., *Vinculin regulates the recruitment and release of core focal adhesion proteins in a force-dependent manner.* Curr Biol, 2013. **23**(4): p. 271-81.

65.   Urbich, C., et al., *Laminar Shear Stress Upregulates Integrin Expression.* Circulation Research, 2000. **87**(8): p. 683-689.

66.   Ye, N., et al., *Direct observation of α-actinin tension and recruitment at focal adhesions during contact growth.* Exp Cell Res, 2014. **327**(1): p. 57-67.

67.   Verma, D., et al., *Flow-induced focal adhesion remodeling mediated by local cytoskeletal stresses and reorganization.* Cell Adh Migr, 2015. **9**(6): p. 432-40.

68.   Sun, Z., et al., *Extracellular matrix-specific focal adhesions in vascular smooth muscle produce mechanically active adhesion sites.* Am J Physiol Cell Physiol, 2008. **295**(1): p. C268-78.

69.   Fletcher, D.A. and R.D. Mullins, *Cell mechanics and the cytoskeleton.* Nature, 2010. **463**(7280): p. 485-92.

44

70.   Powell , K., *Actin and microtubules interact via MAPs.* Journal of Cell Biology, 2005. **170**(5): p. 701-701.

71.   Pogoda, K., et al., *Unique Role of Vimentin Networks in Compression Stiffening of Cells and Protection of Nuclei from Compressive Stress.* Nano Letters, 2022. **22**(12): p. 4725-4732.

72.   Satelli, A. and S. Li, *Vimentin in cancer and its potential as a molecular target for cancer therapy.* Cell Mol Life Sci, 2011. **68**(18): p. 3033-46.

73.   Ridge, K.M., et al., *Roles of vimentin in health and disease.* Genes Dev, 2022. **36**(7-8): p. 391-407.

74.   Boraas, L.C. and T. Ahsan, *Lack of vimentin impairs endothelial differentiation of embryonic stem cells.* Scientific Reports, 2016. **6**(1): p. 30814.

75.   Chernoivanenko, I.S., A.A. Minin, and A.A. Minin, *[Role of vimentin in cell migration].* Ontogenez, 2013. **44**(3): p. 186-202.

76.   Messica, Y., et al., *The role of Vimentin in Regulating Cell Invasive Migration in Dense Cultures of Breast Carcinoma Cells.* Nano Letters, 2017. **17**(11): p. 6941-6948.

77.   Lawson, C., et al., *FAK promotes recruitment of talin to nascent adhesions to control cell motility.* J Cell Biol, 2012. **196**(2): p. 223-32.

78.   Rahikainen, R., et al., *Mechanical stability of talin rod controls cell migration and substrate sensing.* Sci Rep, 2017. **7**(1): p. 3571.

79.   Pogoda, K., et al., *Unique Role of Vimentin Networks in Compression Stiffening of Cells and Protection of Nuclei from Compressive Stress.* Nano Lett, 2022. **22**(12): p. 4725-4732.

80.   Peng, Z., et al., *Endothelial Response to Pathophysiological Stress.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2019. **39**(11): p. e233-e243.

81.   Raoul, A., et al., *Role of focal adhesions of vascular smooth muscle cells in thrombin generation.* Archives of Cardiovascular Diseases Supplements, 2020. **12**(1): p. 145.

82.   Elvevold, K., I. Kyrrestad, and B. Smedsrød, *Protocol for Isolation and Culture of Mouse Hepatocytes (HCs), Kupffer Cells (KCs), and Liver Sinusoidal Endothelial Cells (LSECs) in Analyses of Hepatic Drug Distribution*, in *Antisense RNA Design, Delivery, and Analysis*, V. Arechavala-Gomeza and A. Garanto, Editors. 2022, Springer US: New York, NY. p. 385-402.

83.   Wu, T., et al., *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* The Innovation, 2021. **2**(3): p. 100141.

84.   Lonsdale, J., et al., *The Genotype-Tissue Expression (GTEx) project.* Nature Genetics, 2013. **45**(6): p. 580-585.

85.   Jones, R.C., et al., *The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans.* Science, 2022. **376**(6594): p. eabl4896.

86.   Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nat Methods, 2012. **9**(7): p. 676-82.

45

ACKNOWLEDGEMENTS

46

Afterword

As I reflect on the last years, I consider myself to be exceptionally lucky to have been part of these very different approaches of scientific investigation with a great research group

In concluding this doctoral journey, these studies do not only contribute to the expanding body of knowledge in the field of endothelail biology, but also underscore the ongoing quest for innovative solutions and underline the importance of multi-faceted approaches in research. The three projects presented in this thesis are merely a starting point for deeper investigation to broaden our understanding. I hope, that as the pages of this thesis close, they open new chapters of inquiry, inspiring future scholars to explore these projects in even greater depths.

As scientists, we bear the responsibility of pushing the boundaries of knowledge and uncovering the hidden truths that shape our world. This noble obligation urges us to persevere tirelessly, facing challenges head-on and refusing to yield. Hence, I would like to conclude this thesis with the following quote:

"When the beating of your heart echoes the beating of the drums. There is a life about to start when tomorrow comes."

Do you hear the people sing - Les Misérables
Claude-Michel Schönberg, Alain Boublil, Jean-Marc Nate, Victor Hugo

Thank you for reading!