UiT The Arctic University of Norway

Faculty of Health Sciences, Department of Clinical Medicine

The identification of cell type defining genes across human tissues and the functional study of the endothelial adhesion G protein-coupled receptor L4

Marthe Norreen-Thorsen

A dissertation for the degree of Philosophiae Doctor

June 2023

Contents

Acknowledgements

The work presented in this thesis was carried out during my employment at the Translational Vascular Research (TVR) group at the Department of Clinical Medicine, UiT-The Arctic University of Norway between February 2018 and May 2023. Funding as been through UiT – The Arctic University of Norway.

I would like to start by expressing my gratitude to my lovely and one-of-a-kind supervisor, Lynn Butler. You have been a beacon of light since the first time I met you and been by my side through my PhD journey with all the bumps, twists, and turns we have met on the road (which are not few), always with a smile on your face and with kind words of support. Thank you for taking a chance on me and letting me be a part of this amazing group that you lead. Under your supervision I have had the opportunity to grow, learn, make mistakes, and then rise above stronger than before. Thank you for all you have taught me, for being an amazing inspiration and for being a good friend. And last but not least, I just want to say, now I know what a good leader is all about!

To Philip Dusart, my amazing co-supervisor. Thank you so much for all your support, guidance, good advice, for all the knowledge you have shared, and for all the help I have received. My wet lab knowledge is now greatly expanded due to your terrific guidance, and I have really enjoyed the work and learning new skills. Thank you for being by my side in all the good and bad, for encouraging me and for being a good friend.

To everyone in TVR. Thank you for everything you have taught me, for all the support, and for being like a second family to me. I have been so lucky to end up in this group, as you really are wonderful people and a blessing to work with, which I do not take for granted. You are the reason why I might suffer through three more endless winters in the north, as I do not want to part with you, I could not wish for better colleagues. Thank you for all the laughs and the good times, I hope there will be more.

I would also like to acknowledge all the co-authors of the three articles included in this thesis, and to give a big thanks for all scientific contributions and good work.

To my beloved family, mum, dad, Ane and Christer. I can never say enough times how much I love and appreciate you. You are what keeps me going, day after day, and you give

my life so much meaning. When times are hard, I know it does not only affect me, as you are always following me in your mind, and have also through me, felt alt the pressure that I have had to manage on my journey. Therefore, here at the end, I want to thank you for picking me up when I am low, for tolerating me when I am unbearable, and for giving me an island of refuge in Sandefjord with love and support when I have needed it. You are the best and will forever be.

To my wonderful friends in Oslo. Thank you for all the moments we have had the last years, all our gatherings, for the support when I have competed in Oslo, for our girls' nights on Zoom and for our summer camps, it has brightened up my everyday life, and made all the workload feel lighter. Thank you for all the good conversations and wise words you have given me, it means a lot to share this journey with someone who is wearing the same shoes, who understands what it is all about. Now that I will finally get my life back, I hope to see you soon, and spend even more time with you in the future, and make many new memories.

Tromsø, May 2023

Marthe Norreen-Thorsen

List of papers

The following papers are the basis for this thesis:

I. A human adipose tissue cell-type transcriptome atlas.
Norreen-Thorsen M, Struck EC, Öling S, Zwahlen M, Von Feilitzen K, Odeberg J, Lindskog C, Pontén F, Uhlén M, Dusart PJ, Butler LM.
Cell Rep, 2022. 40(2): p. 111046.
Doi: 10.1016/j.celrep.2022.111046. PMID: 35830816.

II. A tissue centric atlas of cell type transcriptome enrichment signatures.
Dusart P, Öling S, Struck E, Norreen-Thorsen M, Zwahlen M, Von Feilitzen K, Oksvold P, Bosic M, Iglesias MJ, Renne T, Odeberg J, Pontén F, Lindskog C, Uhlén M, Butler LM.
bioRxiv, 2023.01.10.520698
DOI: 10.1101/2023.01.10.520698
Pre-print

III. Adhesion G protein-coupled receptor L4 modifies the endothelial cell response to tumour necrosis factor.
Norreen-Thorsen M, Gottlob L, Kral-Pointner JB, Öling S, Struck EC, Odeberg J, Naudin C, Dusart PJ, Butler LM.
Manuscript

Summary

All cells in the human body share a common genome, but cell type specific gene expression profiles underlie the differences in morphology, behaviour, and specialized function between cell types. Single cell RNA sequencing (scRNAseq) has made it possible to measure gene expression on an individual cell level, which has increased our understanding of the heterogeneity and complexity of gene expression. However, challenges remain with this technology, including limited read-depth, artefactual changes because of cell dissociation from the tissue microenvironment, difficulties in the analysis of fragile or morphologically complex cell types, and bias introduced from the analysis of a limited number of biological replicates.

One of the aims of this thesis was to use an integrative correlation analysis to define cell type enriched transcripts from bulk RNAseq, to circumvent some of the problematic aspects of scRNAseq. RNAseq data was sourced from the Genotype-Tissue Expression portal, where each sample contains transcripts from all tissue constituent cell types. Our method is based on the selection of panels of cell type specific reference transcripts that, through correlation analysis, can be used identify other transcripts with the same expression profile across samples, thus indicating a common cell type origin.

In **Paper I** we applied our method to profile all major constituent cell types in visceral (VAT) and subcutaneous (SAT) adipose tissue, including adipocytes, a cell type that is difficult to extract and process, and thus tend to be absent from scRNAseq databases. Other profiled cell types included adipocyte progenitors, macrophages, smooth muscle-, mast-, plasma-, T- and endothelial-cells. Additionally, cell types absent from SAT, mesothelial cells, and neutrophils, were profiled in VAT. We identified over 2300 cell type enriched coding and non-coding transcripts in VAT, and observed similar cell type enrichment profiles in SAT. We used our data to identify the cell type enrichment profiles of genes that were differentially expressed between VAT and SAT depots – revealing that mesothelial cells in VAT were largely driving these differences. We performed a sex subset analysis that uncovered a panel of male-only cell type-enriched genes, which were identified as Y-linked.

In paper II, we expanded our analysis to include the constituent cell types in 15 different human tissues to create a cell type enrichment prediction atlas for all protein coding genes. We classified 5979 transcripts as cell type enriched in a single tissue, of which 3141 were

expressed in testis. 8011 transcripts were enriched in two or more tissues, and of these, 741 were enriched in at least seven tissues, predominantly in cell types found in all tissues, e.g., endothelial cells. In depth comparisons revealed a large panel of co-enriched genes in related cell types, such as pancreatic alpha and beta cells. For germ cells in testis, spermatogenesis stage-specific enrichment signatures were identified, and temporal changes over the developmental trajectory profiled. Skin had most profiled cell types (18 in total) with keratinocytes having the highest proportion of cell type enriched transcripts, although the minority group of hair root cells were identified as the major source of skin specific transcripts. A cross-tissue cell type comparison identified shared enrichment signatures between cell types with related functions or features, such as those with motile cilia. We defined core identity profiles of cell types present in all or most tissue types, including endothelial cells (EC), which can vary in gene enrichment profiles across different vascular beds. Data from Paper I and II is available on the Human Protein Atlas (HPA) website (www.proteinatlas.org) in the Tissue Cell Type section (www.proteinatlas.org/humanproteome/tissue+cell+type). The HPA receives over 150,000 unique visits per month from all over the world, primarily from the life science research community,

EC play a major role in various biological processes, including the regulation of inflammatory responses and haemostasis. In **paper I and II** we identified the largely uncharacterised G protein-coupled receptor L4 (ADGRL4) as an EC enriched gene across tissue types. This suggests that ADGRL4 could have an important role in EC specific function and phenotype, which we sought to investigate in **paper III**. Mass spectrometry protein profiling revealed that under resting conditions ADGRL4 depletion caused an upregulation of EC proteins related to cell structure and morphogenesis and downregulation of cell cycle-associated proteins. To determine the potential role of ADGRL4 in inflammation, we treated EC with the cytokine tumour necrosis factor (TNF). ADGRL4 depletion was found to selectively augment TNF-induced tissue factor expression. This increase in tissue factor, which is the initiator of the intrinsic pathway of the coagulation cascade, led to enhanced EC thrombin generation from plasma and fibrin deposition from whole blood, confirming increased EC pro-coagulant activity. Concurrently, ADGRL4 depletion inhibited the expression of TNF-induced interferon response genes. Our study indicates that ADGRL4 has a currently unappreciated role in the EC function, with a potential role in the regulation of coagulation during inflammation.

Abbreviations

AC	Adipocyte
ADP	Adenosine diphosphate
aGPCR	Adhesion G protein-coupled receptor
AP	Adipocyte progenitor cell
ATP	Adenosine triphosphate
cDNA	Copy/complementary DNA
Corr.	Correlation
CRP	C-reactive protein
DNA	Deoxyribonucleic Acid
EC	Endothelial cell
FC	Fold change
FDR	False discovery rate
GPCR	G protein-coupled receptor
GO	Gene ontology
GTEx	Genotype-Tissue Expression (project)
HPA	The Human Protein Atlas
HUVEC	Human umbilical vein endothelial cells
ISG	Interferon-stimulated gene
LC-MS/MS	Liquid chromatography coupled to tandem mass spectrometry
LPS	(Gram-negative bacterial) lipopolysaccharide
MastC	Mast cell
MC	Macrophages
MesoC	Mesothelial cell
MFI	Mean fluorescence intensity
mRNA	Messenger RNA
MS	Mass spectrometry
NETs	Neutrophil extracellular traps
NP	Neutrophil
PC	Plasma cell
qPCR	Quantitative polymerase chain reaction
Ref.T.	Reference transcript

RNA	Ribonucleic acid
RNAseq	RNA sequencing
rRNA	Ribosomal RNA
SAT	Subcutaneous adipose tissue
scRNAseq	Single cell RNA sequencing
SD	Standard deviation
SEM	Standard error of the mean
SMC	Smooth muscle cell
siRNA	Small interfering RNA
snRNAseq	Single nuclei RNA sequencing
TC	T-cell
TNF	Tumour necrosis factor
TPM	Transcript per million
UMAP	Uniform Manifold Approximation and Projection
UTP	Uridine triphosphate
VAT	Visceral adipose tissue
WGCNA	Weighted network correlation analysis

1. Introduction

1.1 Cell specific gene expression and relation to cell phenotype and function

The human genome consists of approximately 20000 predicted protein coding genes [1, 2]. Despite sharing the same genome, cells in the different organs of the human body can differ in appearance, function and behaviour, due to their individual transcriptome and corresponding protein expression profile.

During the transition from pluripotent to differentiated cells, epigenetic changes take place, where chromatin regions in pluripotent cells that are found to have an open configuration, which causes genes to be responsive and accessible, becomes progressively and selectively closed during differentiation as a result of histone modifications and DNA methylation. Thus, accessible genes are fewer in differentiated cells relative to stem cells, where the specific pattern of gene silencing varies across cell lineages. Additionally, some genes that are inactive due to DNA methylation in early phases of differentiation, might become induced in a tissue specific manner in a later phase of differentiation and therefore only show expression in a subset of cell types [3]. This could partly explain why cell types with same developmental origin, or which represent a common cell lineage, can often be observed to have overlapping expression profiles and common functional attributes. In addition to the normal developmental epigenetic modulations that happens during cell differentiation, environmental factors can also cause an epigenetic imprint that affects the regulation of genes in germ cells and during embryogenesis, but also in later phases of life, and such factors might include various chemicals or toxins, diet habits, temperature exposure and maternal behaviour amongst others [4-7]. Transcriptional regulation affecting cell behaviour and morphology might also be a consequence of the structural properties of a tissue and the cell-cell and cell-matrix interactions through for example mechano-transduction, along with chemical and physical cues of the surrounding microenvironment, that can directly and indirectly regulate phenotype and prime cells for various behavioural responses [8, 9].

Many genes are tissue specific or enriched, which means their expression is unique to or higher in one (or few tissues) relative to others. These genes typically encode for proteins related to tissue specific biological functions. In cases where similar gene enrichment signatures extend to several tissues, it is often because the tissues share common functions and features [10, 11]. Highly specialized organs like testis and brain, are amongst the tissues with highest specific gene expression signatures, and tissues with related functions, such as

those belonging to the gastro-intestinal tract (e.g., stomach, colon and small intestine) share a common gene signatures that differ from other tissues [11]. These tissue specific genes are typically expressed in tissue specific cell types, with specialized functions in the tissue they are residing in. For example, bulk tissue gene expression profiling carried out by the Human Protein Atlas (HPA) [10] show that the surfactant protein A1 (*SFTPA1*) gene is lung enriched, while single cell data from HPA and the Tabula Sapiens human cell atlas [12, 13] identify the alveolar cell type 2, which are lung specific resident cells, as the main source of this protein. The protein is important for lung function as it is involved in surfactant homeostasis and pulmonary immunity [14, 15]. Similarly, the gene Myosin heavy chain 6 (*MYH6*) is mainly expressed in the heart, where it has a functional role in cardiac muscle contraction [16, 17] and is predominantly expressed in cardiomyocytes. Other genes can have a lower tissue specificity, but still be cell type specific, when that is found in all, or most, tissue types. For example, cadherin 5 (*CDH5*), is found in many different tissues, as it is enriched in endothelial cells across all vascular beds. Indeed, this protein has a key functional role in the establishment of tight contacts between neighbouring endothelial cells [18]. In some instances, gene enrichment might not be limited to one specific cell type, but rather a group of related cell types with similar function or behaviour. Selectin L (*SELL*) is found expressed across various lymphocytes [12, 13], as it is a key protein involved in diapedesis, which is needed for the circulating immune cells to cross the endothelial barrier and reach infected tissues [19].

Even though cell (or cell subset) specific functions are related to the expression of particular genes and proteins, these can be present at elevated levels there (enriched), rather than exclusively expressed in only one cell type, exemplified in a global protein profiling study carried out by HPA [20]. Here, relative levels of 4842 proteins were identified across 65 normal cell types from various tissues and organs using immunohistochemistry images produced from 5934 antibodies. Unsupervised cluster analysis of protein expression pattern across these cell types showed that cells with similar functions exhibited similar protein profiles and clustered together, even if they were not from the same tissue, and cluster similarity coincided with the developmental origin of the cells (endoderm, ectoderm, or mesoderm), with few exceptions where morphological differentiation seemed to override developmental origin. A large proportion of the proteins were expressed at a detectable level across most of the cell types in the study, while only 74 proteins were specifically expressed in only one cell type. Even though most of the proteins were not found to be exclusively

expressed in one cell type. Despite this, a highly differentiated global pattern was observed between cell types, indicating the extent of protein expression was the main driver behind cellular phenotype. Thus, relatively few genes are likely to have true cell type specificity.

1.2 Bulk and single cell sequencing

Bulk RNA sequencing and single cell RNA sequencing (scRNAseq) are both methods that obtain gene expressional information in cells. Bulk RNAseq measures transcripts from a mixed cell population, generating average expression values [21]. scRNAseq has the advantage of resolution at a cell level, where transcripts are measured on a cell-by-cell basis, revealing heterogeneity across cell populations and allowing the classification of cell types and subgroups [22]. It has proven valuable amongst others in studies of cellular development and in defining cellular composition and contributions in cancer [23]. Such data has furthermore been collated into various cell atlases, which are great public sources of information [12, 13, 24]. An overview of the established methods up to the sequencing point is described below for bulk and single cell sequencing, while the last section describing some important limitations in some of the processing steps that apply for scRNAseq relative, to bulk RNAseq.

1.2.1 Bulk sequencing methods

Next generation sequencing is recognized as a high throughput method to analyses the DNA or RNA content of many samples concurrently with high accuracy. Commonly used platforms for next generation sequencing are Illumina and Ion Torrent, known as “short-read” sequencing technologies, as the material being sequenced (DNA or RNA) is fragmented prior to analysis [25]. Short-read sequencing is popular for differential gene expression analysis of bulk samples. The alternative long-read sequencing, also known as third generation sequencing, can obtain information for longer sequencing stretches, and in the context of RNAseq, is useful to study alternative splicing, or for transcript variant discovery [26]. Common workflows and platforms for short-read sequencing is given in brief below.

Sample preparation normally begins with lysis of cells, and isolation of RNA. As total RNA contains a high amount of ribosomal RNA (rRNA), which can affect sequencing depth and detection of lower abundant transcripts, it is common to either carry out rRNA depletion by selective removal, or by selecting for polyadenylated RNAs through use of poly(T) oligos targeting the poly(A)-tail and removing the remaining RNA [21]. Subsequently, fragmentation

is carried out to achieve an optimal fragment size for the platform and application, which might be done by physical, enzymatic or chemical methods [27]. Reverse transcription, end repair and adaptor ligation then follow, where the adaptor contains a sequence that facilitates recognition and binding within the instrument used. It is common for the adaptor to also have a barcode, which during multiplexing can be used to identify reads originating from the same sample during data analysis. The final step of library preparation is normally PCR amplification. Short read sequencing is separated in two steps, clonal amplification, and the sequencing itself. For the Illumina technology, the template first binds to a flow cell through the adaptor and clonal amplification is done by bridging PCR, which causes a cluster of sequencing clones that can amplify the signal. Sequencing is based on the optical readout of fluorescent nucleotides that are incorporated into the growing nucleic acid chain and then imaged. Each nucleotide has a terminator attached that blocks incorporation of more than one nucleotide per cycle, which along with the fluorescent dye, is cleaved of before the next cycle to allow further nucleotide incorporation. For the Ion torrent platform, emulsion PCR is used for amplification, where the templates with adaptors are captured in a water-in-oil emulsion droplet (micelle) along with a bead covered with complementary adapters, nucleotides, polymerase, and primers. The micelles are then loaded onto a semiconductor chip, which consists of a flow chamber and complementary metal-oxide semiconductors (CMOS) pH sensor. The chip is flooded with unmodified nucleotides, whereby the incorporation of nucleotides results in the release of hydrogen ions that are detected by the CMOS pH sensor [25].

1.2.2 Single cell sequencing methods

The main steps of single cell sequencing involve cell separation, cell lysis, nuclei acid amplification, cell sequencing and data analysis [22, 28]. There are many methods and platforms available to achieve the various steps, and the choices are often related to the research question and resources available.

Cell isolation: A key factor to be able to study a cell on a single level, is the separation of cells found in tissues or body fluids, which is a processing step not required for bulk sequencing. Various methods can be used for separating cells, and they have different advantages and drawbacks. Manual cell picking (micromanipulation) uses an inverted microscope and micropipettes, which are movable through motorized mechanical stages. Cells in suspension can then be manually isolated by aspiration and transferred to a collection

vessel [29]. Advantages here are low costs, flexible and selective sampling, and a simple operation, but the throughput is low, and cells can sustain mechanical injury [30]. Laser capture microdissection (LCM) is used for isolation of cells from tissue, where cells are dissected by laser under visualization through a microscope. The cell is extracted by adhesive tapes, gravity extraction or by pressure catapulting with a defocused laser pulse [29]. Advantages are that cell morphology and structure can be maintained along with spatial location information, but the method has low throughput, high costs, can cause nuclear damage and RNA pollution, and requires a high skill set [30]. Fluorescence Activated Cell Sorting (FACS) is based on the principle of flow cytometry, where laser excitation is used to obtain cell property information which can be used for selection purposes. The cells can be automatically sorted by being assigned a charge and are thereafter guided by deflector plates into tubes or well plates. FACS can be very efficient in analysing and sorting cells and is widely used [29]. It is considered to show high specificity, accuracy and sensitivity, and high throughput, but have high costs, can cause mechanical injury and requires a large amount of input cells [30]. Microfluidics are considered as “lab on a chip” and have many different applications areas. They can be used for high throughput cell capture, where only very small volumes of sample and reagents are needed and can have flexible design to accommodate various needs. Multifunctional units can be integrated, to support successive preparation steps and automation. Droplet-based microfluidics are possibly the most relevant microfluidics method for sequencing and generate droplets by having one liquid phase breaking off another immiscible liquid, which can be used to encapsulate individual cells. They can then work as microreactors, which requires very low volumes and can be produced at high rates. The droplets can be manipulated, as in merged, split, re-loaded, incubated, and used for detection and sorting. The drawback here is that encapsulation is random, and droplets can be empty. Although the throughput is generally high for microfluidic systems, they can be costly [30].

Lysis steps: Lysis of single cells can be achieved in different ways, and might depend on the previous separation step, downstream analysis and the cell type studied. Physical lysis options include the use of mechanical force to break the membrane, heat induced denaturation of membrane and the use of voltage-based membrane disruption. Other options are the use of chemical lysis, which uses surfactants to disrupts the membrane and enzymatic cell lysis [30].

Amplification: Within a single cell, the quantity of RNA is very low, far less than what is required for sequencing, and amplification before sequencing is required. Before amplification, RNA needs to be reverse transcribed into cDNA, either by traditional PCR, modified PCR, T7-in vitro transcription (IVT) or Phi29 DNA polymerase-mediated RNA amplification [30]. The reverse transcription step normally involves the use of primers that targets the poly(A)-tail to achieve mRNA selection. These primers often also have a barcode specific for each cell prepared, which then makes multiplexing possible and can increase throughput. Thirdly, unique molecular identifiers (UMIs) can often be found along with primer and barcode to tag individual mRNAs, and can be used to distinguish between original molecules and amplification duplicates, in order to reduce amplification bias [31]. The conversion process into cDNA can be very inefficient, with as little as 10-20 % of the RNA being successfully converted, with some methods reported to perform better than others [30].

Single cell RNA sequencing and considerations: Many different methods can be used for scRNAseq. In general, strategies either focus on obtaining transcriptome full read coverage, where the full length of the cDNA is sequenced, and therefore is optimal when information on splicing patterns are necessary, or as shorter 5' and 3' end reads, which only provide sequence information for one end of the cDNA, which are mainly is useful for studies that focus on relative transcript abundance. The most appropriate strategy depends on the research question [30]. Another factor to consider is how the data output from single cell sequencing is affected by different factors related to the strategies used in each step, from cell isolation to sequencing. Methods can vary in accuracy (how well read quantification corresponds to the actual concentration of mRNAs), sensitivity (the probability to capture and convert a particular mRNA transcript present in a single cell into a cDNA molecule present in the library) and precision (the technical variation of the quantification). These factors, along with numbers of cells analysed, will all have an impact on power to detect relative differences in expression levels. A practical matter is also costs and resources available, which is why methodical considerations are important [31].

Challenges related to single cell sequencing: For scRNAseq analysis of cells from tissue samples, cells needs to be dissociated, while still maintaining cell viability and integrity [21]. Here, LCM can be used for dissecting cells, but cell contamination and damage is possible, and the method is labour intensive with low throughput [23]. Enzymes are useful to dissociate

cells, but the process is often carried out at 37°C, which can create artifacts, including gene expression changes. The use of enzymes that can work under conditions with low temperatures has been suggested to counteract this [32]. The processing time between when cells are harvested from tissue until they are lysed can also affect the expression profile of cells [33], meaning that sequencing results might be subjected to an artificial imprint and not fully represent the native expression pattern of a cell in its natural environment. In some cases, cells might be very fragile, large or morphologically complex, and therefore difficult to isolate in a viable and intact state, such as neuronal cells [28] and cardiomyocytes [21]. Such problems have consequences for cell representation in datasets, for example adipocytes, which are hard to isolate and process for scRNAseq [34], and are often are excluded from scRNAseq profiling datasets [12, 13, 35, 36]. There are other workarounds for this problem, such as single cell nuclei sequencing, but this sequencing excludes transcripts contained in the cytoplasm, thus producing an incomplete profile [37, 38]. Another limiting factor of scRNAseq, is the low detection rate of non-coding RNAs within scRNAseq datasets. This is a direct result of selective enrichment of polyadenylated RNAs in the reverse transcription step, which mostly enriches for mRNA, and to a lower extent non-coding RNAs [22]. Unlike bulk sequencing, which are presented with the opportunity to deplete the rRNA, most commercial methods so far depend on polyadenylated RNA enrichment for library preparation [31].

1.3 Endothelium

The endothelium is a monolayer of cells that makes up the inner lining of blood vessels. Endothelial cells (EC) are polarized, attached to the basement membrane on the basolateral side, and are in direct contact with blood on the apical side. They have a flat appearance, with nucleus and cell body alignment in direction of blood flow. EC are positioned as a barrier between blood and surrounding tissues, where they govern the exchange of fluids and various molecules, like nutrients and metabolites. They also regulate vascular tone and blood supply to tissues, are involved in angiogenesis, and leukocyte recruitment and transmigration during inflammation. They have a role in haemostasis, expressing both pro-and anti-coagulant factors [39].

1.3.1 Endothelial heterogeneity

The remarkable heterogeneity and plasticity of EC are prevalent already at early developmental stages. During vasculogenesis, cells of mesoderm origin develop into different angioblast pools with specific markers, expression patterns, and distinct signalling pathways, which affect their developmental potential. These different angioblast pools give rise to distinct vessel types in different organs. EC are heterogenous across tissue vascular beds, but also between vessel types within an organ, i.e., arteries, veins, capillaries, and lymphatic vessels. This is a result of both early developmental priming and adaption to tissue specific requirements driven by the surrounding microenvironment [40, 41].

1.3.2 Barrier function of endothelium

The endothelium acts as a selective barrier, regulating the exchange of fluids and molecules between blood and surrounding tissues. Barrier function is important for the maintenance of colloid osmotic pressure, a prerequisite for retaining water within the circulatory system, tissue perfusion and the supply of oxygen and nutrients. Generally, the endothelium allows passive diffusion of gases, ions and small solutes, but blocks movement of larger molecules like proteins [42]. Size selectivity and extent of movement depends largely on how tightly adjacent EC are connected by tight junctions (near the luminal surface) and adherens junctions (more basally located). Tight junctions are especially important for direct barrier function, but both tight and adherens junctions are important for integrity and stability of endothelium, and disruption or low junction density can lead to increased cellular gaps, and permeability. Several stimulatory factors can cause junctional changes and affect permeability dynamics, which can occur in various diseases, leading to increased leakage of proteins and fluids into tissues. The presence and density of junctions vary across the vascular tree and can also be organ dependent. Arteries, which are exposed to high rates of blood flow, have many tight junctions, and low permeability, as does the blood-brain barrier. Capillaries and post-capillary venules, which facilitate the exchange of gases and metabolites with nearby tissue, are leakier with a lower tight junction density [43]. Extracellular matrix proteins and the basement membrane might also play a part in permeability by supporting EC attachment and integrity. Glycocalyx acts as a repellent towards proteins, platelets and cells, to block inappropriate interactions [42]. Permeability can also refer to transcellular exchange, which often applies to larger molecules, and might include vesicular carriers like transendothelial channels and caveolae, the latter of which have high density in capillaries

[44]. Some organs perform specific tasks that require a high capillary permeability, like the filtration of blood, which depends on fenestra, a type of transcellular pore, abundantly found in liver, kidney, endocrine glands, and intestinal mucosa. Most fenestra have specialized structures called diaphragms that span across the cell and repel and block larger molecules like proteins but allow movement of fluids and small molecules [45]. Taken together, the permeability of the endothelium varies across the vascular tree and depends on the tissue function and physiological needs.

1.3.3 Leukocyte trafficking and inflammatory responses

Endothelial activation in response to inflammatory cytokines, like Interleukin-1 beta (IL1B) and tumour necrosis factor (TNF) [46, 47] , leads to the expression of chemokines that attract infection-fighting leukocytes to the site of inflammation [48-50]. Activation also increases the expression of adhesion receptors E-selectin (SELE), intercellular adhesion molecule 1 (ICAM1), and vascular cell adhesion protein 1 (VCAM1) [51, 52]. Concurrently, P-selectin (SELP), which is stored in Weibel-Palade bodies, is mobilized to the surface [53, 54]. These proteins are important for the process of leukocyte recruitment, which is a three-step process. EC initially capture leukocytes from flow via P-selectin and E-selectin binding-interactions with leukocyte L-selectin and P-selectin glycoprotein ligand 1 (SELPLG), which supports leukocyte rolling on the EC surface. Firm adhesion follows, through binding of leukocyte integrins to ICAM1 and VCAM1 on the EC apical surface. The leukocytes flatten and take on a polarized shape and crawl to the transmigration site, through integrin alpha-M (ITGAM) and ICAM1 interactions [55]. Then follows the extravasation from blood into tissue, via one of two different routes. For the paracellular route, the leukocyte crosses the endothelium by moving in between neighbouring cells, an action that requires disruption of adherens junctions, and depends on endothelial molecules like ICAM1, VCAM1, platelet endothelial cell adhesion molecule (PECAM1), CD99 antigen (CD99) and endothelial cell-selective adhesion molecule (ESAM). Alternatively, the transcellular route does not require junctional disruption, as the leukocyte migrates directly through EC by a channel formed from membrane fusion of the two cells, a process that requires many of the same protein interactions as for the paracellular route [55]. Several studies indicate that the route chosen for transmigration depends on the properties of the vascular bed, and how tight the junctions are. EC in the blood brain barrier have a high density of tight junctions, and here leukocytes preferentially use the transcellular route [56]; they seem to sense the stiffness and mechanical properties of the vascular bed

and take the route with least resistance [57]. In dermal microvascular endothelial cells, leukocytes preferably traverse by the transcellular route if caveolin-1 is abundant, while the paracellular route is favoured at low levels [58]. This shows that the inherent property of a vascular bed can affect leukocyte interactions. The presence of P-selectin, E-selectin and VCAM1, and their expression pattern in response to various stimuli, differs over time and in magnitude across tissues and vessels types, possibly reflecting a variation in capacity for leukocyte recruitment [59-61]. Furthermore, although leukocyte recruitment and diapedesis generally happen in post-capillary venules [58, 62, 63], there are exceptions, which can be organ dependent. For example, leukocyte recruitment in the lungs occurs in the alveolar capillaries, rather than post-capillary venules, which are often narrower than the leukocyte itself, making the selectin-dependent rolling-capture step observed in other tissues redundant [64]. Similarly, leukocyte recruitment in the liver is largely independent of selectins, occurring in the sinusoids, rather than post-sinusoidal venules [65]. Appreciation of such endothelial heterogeneity can expand our understanding of inflammatory diseases, and the potential contribution of subgroups of EC.

1.3.4 Regulation of haemostasis

The coagulation cascade, driven by various proteins referred to as coagulation factors (abbreviated F, and proteins are traditionally referred to with roman symbols, where ‘a’ refers to activated form), can be initiated by two different pathways and involves a stepwise cascade where zymogens are processed into active serine proteases that converge on the generation of thrombin from prothrombin (also known as coagulation factor II (F2)), leading to fibrin deposition [66]. The extrinsic pathway starts with tissue factor (also known as coagulation factor III (F3)), which, upon exposure by vessel wall injury, complexes with FVII (F7) to create activated FVIIa. The FVIIa and tissue factor complex, in presence of calcium, then activates factor FX (F10) to yield FXa [67, 68]. FXa is a central serine protease in coagulation, which along with the cofactor glycoprotein FVa (F5), converts prothrombin into active thrombin, and is referred to as the prothrombin complex. The assembly and activity of this complex normally requires calcium and association with negatively charged phospholipids like phosphatidylserine, mainly present on the surface of activated platelets [69, 70]. The intrinsic pathway involves an activation cascade of serine proteases, initiated by FXII (F12), which can be auto activated by contact with certain negatively charged surfaces [71, 72]. FXII can be activated by, and cause activation of, plasma kallikrein (KLKB1), resulting in a positive

feedback loop [73]. Subsequentially, FXIIa can convert FXI (F11) to FXIa, which then acts on FIX (F9) to give FIXa [74]. FIXa, along with cofactor VIIIa (F8) [75], in presence of calcium and phosphatidylserine [76], can efficiently convert FX to FXa. From FXa activation, the intrinsic and extrinsic pathway converge into the common pathway, resulting in prothrombin complex formation and thrombin generation [66]. Thrombin has several essential functions, it drives fibrin deposition by two-step processing of fibrinogen (composed of a fibrinogen alpha chain (FGA), beta chain (FGB) and gamma chain (FGG)), yielding fibrin aggregates, and creates a positive feedback loop for the coagulation cascade. The positive feedback loop involves processing of FV and FVIII into their respective active forms, FVa and FVIIIa, which can associate with FXa and FIX, respectively, and further amplify thrombin generation. Furthermore, thrombin converts FXIII (composed of two alpha chains (F13A1) and 2 beta chains (F13B)) into FXIIIa, which can then cross-link fibrin fibres to stabilize the clot [77].

Coagulation and clot formation does not solely depend on the coagulation cascade, but also on platelet activation. Activated platelets provide the negatively charged phospholipid surface that is required for activation of several coagulation factors and are important for creating a stable and dense clot. This platelet activation is initiated by weak affinity binding of platelets to von Willebrand factor (VWF) by platelet glycoprotein Ib alpha chain (GP1BA) and collagen through the platelet glycoprotein VI (GP6), which induce intracellular signalling and cellular changes, including exposure of high affinity receptors that causes firm adhesion of platelets to the exposed subendothelial matrix, and importantly, activate the high affinity integrin alpha-IIb and integrin beta-3 complex (ITGA2B and ITGB3) that binds fibrinogen, creating aggregates of platelets and fibrin that result in stable clot formation [78].

EC express the large multimeric glycoprotein VWF, which are compactly packed into storage organelles named Weibel Palade bodies (WPB) along with other proteins, including FVIII. Single WPB content can be released by basal secretion and might circulate in the blood. Upon endothelial stimulation or damage, WPB can fuse to large secretory pods before end-fusion with the plasma membrane, causing massive release of VWF-bundles with platelet adhesive properties [79]. The subsequent platelet adhesion is dependent on conformational changes caused by high shear rate, leading to exposure of platelet binding sites. The release of VWF causes the associated FVIII to be present in the blood stream, where it can participate in the coagulation cascade [80]. It is interesting to note that there is a heterogenous expression of both VWF and FVIII (protein and/or mRNA) across vascular beds, where FVIII

expression has mainly been identified in liver sinusoidal, renal glomerular and lymphatic endothelial cells, and in some high endothelial venules and lung endothelial cells, while VWF has been found in capillary endothelial cells, and endothelium of larger vessels including aorta, in addition to co-expression of both in postcapillary high endothelial venules [81-85].

Fibrinolysis is driven by the conversion of the plasma zymogenic protein plasminogen (PLG) into activated plasmin, a process catalysed by the serine protease tissue-type plasminogen activator (PLAT). This process requires plasminogen to bind to its substrate fibrin, which causes a conformational change that exposes the cleavage site to PLAT. Plasmin will then cleave fibrin into various degradation products that, depending on several factors like the extent of plasmin activity and clot composition, either will restrict clot growth or dissolve it entirely [86, 87]. The PLAT activity is directly blocked by binding to the specific plasminogen activator inhibitor 1 (SERPINE1), and the complex will then be quickly removed from the circulation. Furthermore, regulation can be achieved by plasmin binding to its direct inhibitor Alpha-2-antiplasmin (SERPINF2), or more indirectly, as Carboxypeptidase B2 (CPB2), also known as thrombin activatable fibrinolytic inhibitor, can remove exposed lysin residues from the fibrin surface, which blocks plasminogen docking and activation [87]. EC participate in these important processes by secreting PLAT [88, 89] and SERPINE1 [90-92]. According to the Tabula Sapiens human cell atlas [13], *PLAT*, and especially *SERPINE1*, show heterogeneous expression across subgroups of endothelial cells.

The anticoagulant protein C pathway plays an important role in limiting clot formation and procoagulant activity. Vitamin K-dependent protein C (PROC) circulates in plasma in an inactive form and can bind to a complex of thrombin and thrombomodulin (THBD), along with endothelial protein C receptor (PROCR), which induces its catalytic activation by thrombin. Activated protein C in complex with its cofactor Vitamin K-dependent protein S (PROS1), targets FVIIIa and FVa for cleavage and inactivation, thereby reducing generated thrombin [93]. The endothelium expresses both PROCR [94-97], and THBD [97-100], and the process of protein C activation occurs mainly on the endothelial surface, illustrating their important role in the regulation of coagulation [93]. Tissue factor pathway inhibitor (TFPI), which directly binds and blocks activity of the tissue factor-FVIIa complex, in addition to FXa-FVa [101], is expressed mainly by EC and megakaryocytes/platelets [102-104]. *TFPI* mRNA expression seems to vary depending on EC subtype and tissue location [13, 105]. Furthermore, antithrombin-III (SERPINC1) is an inhibitor of the serine protease coagulation factors, mainly

acting on thrombin, FXa and FIXa, the activity of which is greatly enhanced by endothelial heparin-like molecules [106]. Lastly, under resting normal conditions, the endothelium keeps platelets from undergoing unwanted activation by producing and secreting substances like nitric oxide (NO) and prostacyclin, in addition to expressing the membrane bound ectonucleoside triphosphate diphosphohydrolase 1 (ENTPD1) which converts the platelet activating molecules adenosine triphosphate (ATP) and adenosine diphosphate (ADP) into adenosine monophosphate [107]. Thus, EC play an extensive and intricate role in the regulation of haemostasis.

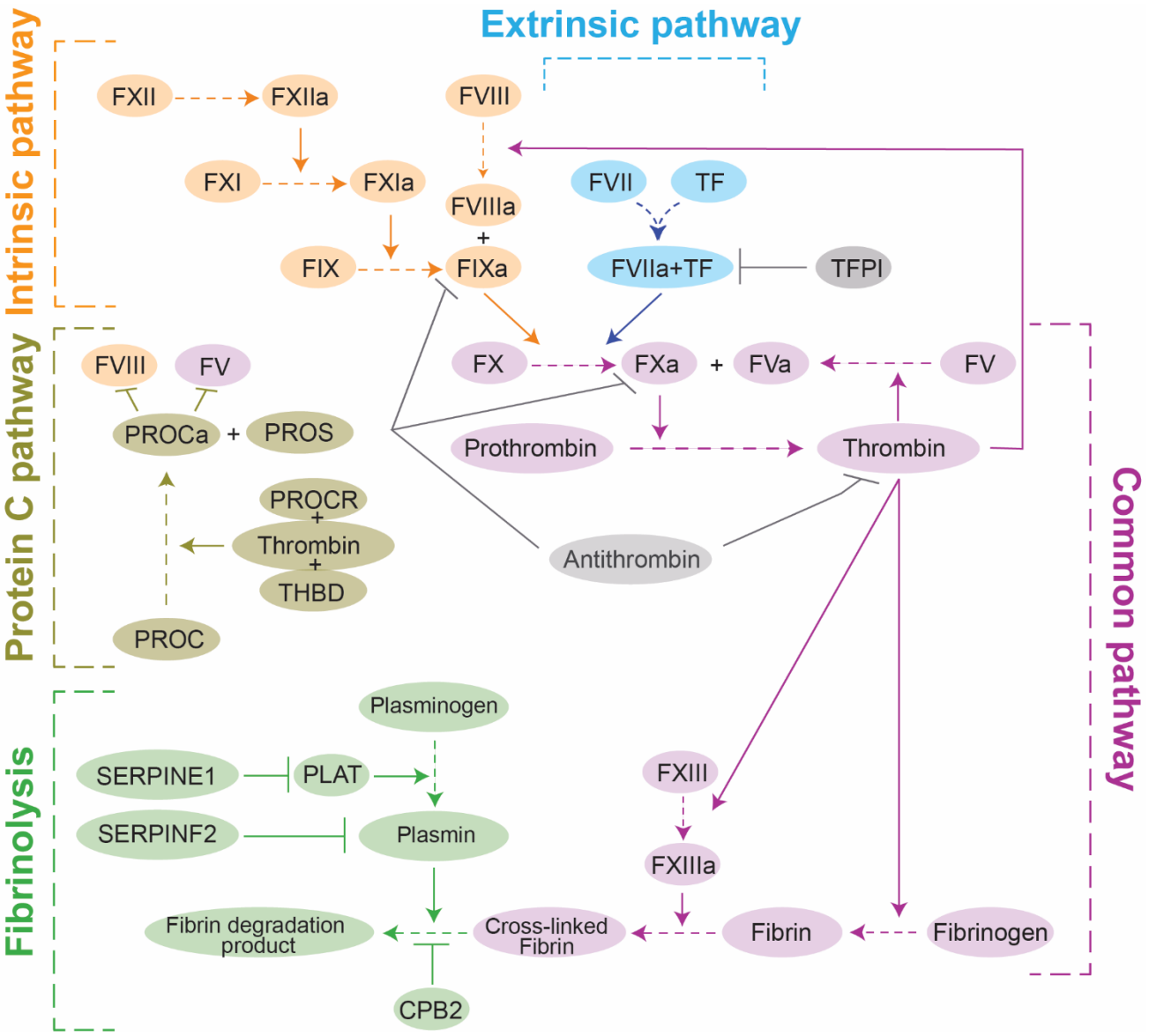


Figure 1: Regulation of haemostasis: Overview of the network of proteins that contribute to haemostasis elaborated on in section 1.3.4. The intrinsic pathway and the extrinsic pathway both converge upon the common pathway and activation of FX to generate thrombin, and subsequently the generation of a fibrin mesh. Fibrinolysis and the protein C pathway are important for controlling the extent of clotting, to shut down the activated coagulation factors when activity is no longer needed and for clot resolving in the aftermath.

1.3.5 Endothelium in thrombosis

There are three main contributing factors that can lead to thrombosis; changes in blood flow, blood coagulability and vascular dysfunction [108]. Thrombosis can be a heterogenous disease, in terms of both triggers and location. Different environmental and genetic factors can contribute to the development of a thrombus, which may vary between females and males or children and adults. Factors that can affect the location of thrombus include the nature and size of endothelial injury, the presence and binding of microparticles with pro-coagulant properties, the endothelial structure and inherent properties, inflammation, drug or toxin exposure, changes in flow dynamics, and an imbalance in prothrombotic and antithrombotic factors. Diseases, treatments and genetically inherited defects can all modify these risk factors [109]. EC show high heterogeneity and plasticity, with functional adaptions to organ site and with various inherent properties affecting their response to various stimuli and environmental conditions, which also extends to expression of pro- and anticoagulant molecules. [44]. Indeed, there are substantial differences between vessel types, where high shear stress exposed arteries generally express high amounts of PROCR, endothelial nitric oxide synthase (NOS3) and thrombomodulin, but lower amounts of VWF and PLAT, while low shear stress exposed veins express high amounts of PROCR, PLAT and VWF but lower amounts of thrombomodulin, relative to arteries. Capillaries predominantly express TFPI and thrombomodulin [110]. This endothelial heterogeneity likely contributes to the site of thrombus formation [44, 109].

The link between thrombosis and inflammation is becoming increasingly acknowledged, with the endothelium likely having a key role. Tissue injury results in the production of reactive oxygen species (ROS), cytokines, chemokines, and damage-associated molecular pattern (DAMP) signals from endothelial cells, platelets and leukocytes. These molecules induce a more pro-adhesive, pro-inflammatory and pro-coagulant endothelial phenotype. In turn, endothelial expression of adhesion molecules and chemokines, recruit more leukocytes to site of injury. Activated monocytes can express tissue factor and procoagulant microvesicles, potentially activating the coagulation cascade, while the presence of platelets with a negative phospholipid surface can further drive the cascade by activation of downstream factors. Neutrophils can support thrombosis by release of neutrophil extracellular traps (NETs). Additionally, inflamed endothelium can release WPB containing VWF and other procoagulants, which can cause platelet binding and activation. Platelets might then bind to

fibrin and endothelium and further support clot formation through the release of granules which can further recruit and activate platelets and neutrophils. Endothelial activation can increase permeability, which might cause leakage and exposure of the subendothelial layer. Tissue injury also causes the release of ADP and ATP, which are interpreted as danger signals by various cells, and can further exacerbate inflammatory and thrombotic pathway. These complex interactions between the various cells can amplify a thrombogenic environment. Endothelial activation upon inflammatory stimuli can therefore be an important driver of thrombosis [111].

In vitro studies show that EC can express tissue factor in response to TNF [112], Gram-negative bacterial lipopolysaccharide (LPS) [113], IL1B containing NETs [114], extracellular histones [115], uridine triphosphate (UTP) [116], disturbed flow [117, 118], thrombin [118] and C-reactive protein (CRP) [119]. Several of these factors could be present in sites of injury or inflammation, or under other pathophysiological conditions. Tissue factor can be expressed in an active form on the endothelial surface [114, 120], implying that endothelial tissue factor could contribute to thrombosis under various circumstances. Despite this, it has been debated if endothelial-originating tissue factor contributes to thrombosis [121-123]; the consensus is that EC do not contribute much to coagulation in a resting state, but the contribution in a pathological state is less clear. *In vitro* studies indicate little to no tissue factor expression in resting EC [112, 113, 120, 124]; indeed, exposure of tissue factor to blood would be detrimental under resting conditions. However, the endothelial contribution to thrombus formation has gained more attention [111, 125], and inflammation activated tissue factor expression may well contribute to the underlying disease pathophysiology.

2. Aims of the thesis

1. To perform an integrative correlation analysis of bulk RNA sequencing data of human visceral and subcutaneous adipose tissue and define the enriched transcriptomes of constituent cell types, including endothelial cells. To investigate possible differences between adipose depots and sexes by cross-comparison of cell type enrichment signatures.

2. To perform an integrative correlation analysis of bulk RNA sequencing data from 15 human tissue types and identify cell type enrichment profiles. To compare similarities and differences in cell profiles across tissue types, and to identify the core transcriptome signatures for cell types found in multiple tissue types, including endothelial cells.

3. To investigate the role of the endothelial enriched protein Adhesion G protein-coupled receptor (ADGRL4) under resting and inflammatory conditions.

3. Methods

3.1 Paper I and Paper II

In paper I and paper II a bioinformatics-based approach, developed in our group, was used to extract information on which genes are enriched in different cell types from bulk RNAseq data. Additionally, some common methodologic approaches for validation and sub-analysis will be presented in this section. Other details specific for each paper are found under the corresponding method section.

3.1.1 GTEX data

Bulk RNAseq data was retrieved from the Genotype-Tissue Expression (GTEx) project (release V8), through their portal (www.gtexportal.org) [126], which is an open access public resource with a comprehensive sample selection from various human non-diseased tissues.

3.1.2 Correlation analysis

The analysis method is based on the selection of specific markers for each cell, referred to as ´reference transcripts´ (Ref.T.), which are used in a correlation-based analysis to identify other cell specific transcripts. The analysis is based on the concept that transcripts originating from one cell type will be expressed at approximately the same ratio within this cell type across samples. The Spearman correlation coefficient between the Ref.T. and all other sequenced transcripts are calculated; those with high correlation with Ref.T from only a single cell type likely have cell type restricted expression (see Figure 2 for concept overview). Thus, the first aim was to define a panel of 3 Ref.T. for each constituent cell type and use these to identify a list of cell type enriched genes from bulk sequencing data. As the specificity of the Ref.T. lays the foundation for the identification of other cell enriched transcripts, it is important to find good and specific cell type representative transcripts. Therefore, a shortlist of candidates for tissue constituent cell types were selected based on literature searches, or the results of immunohistochemistry staining in the Human Protein Atlas. Spearman correlation coefficients were calculated between the expression levels of these candidates, and the rest of the transcriptome, and selection of cell type Ref.T. panels based on: (i) a high correlation between Ref.T. within each cell type panel, consistent with co-expression. (ii) a low correlation coefficient between Ref.T. in different cell type panels, consistent with high specificity and (iii) a normal distribution of Ref.T. expression across samples.

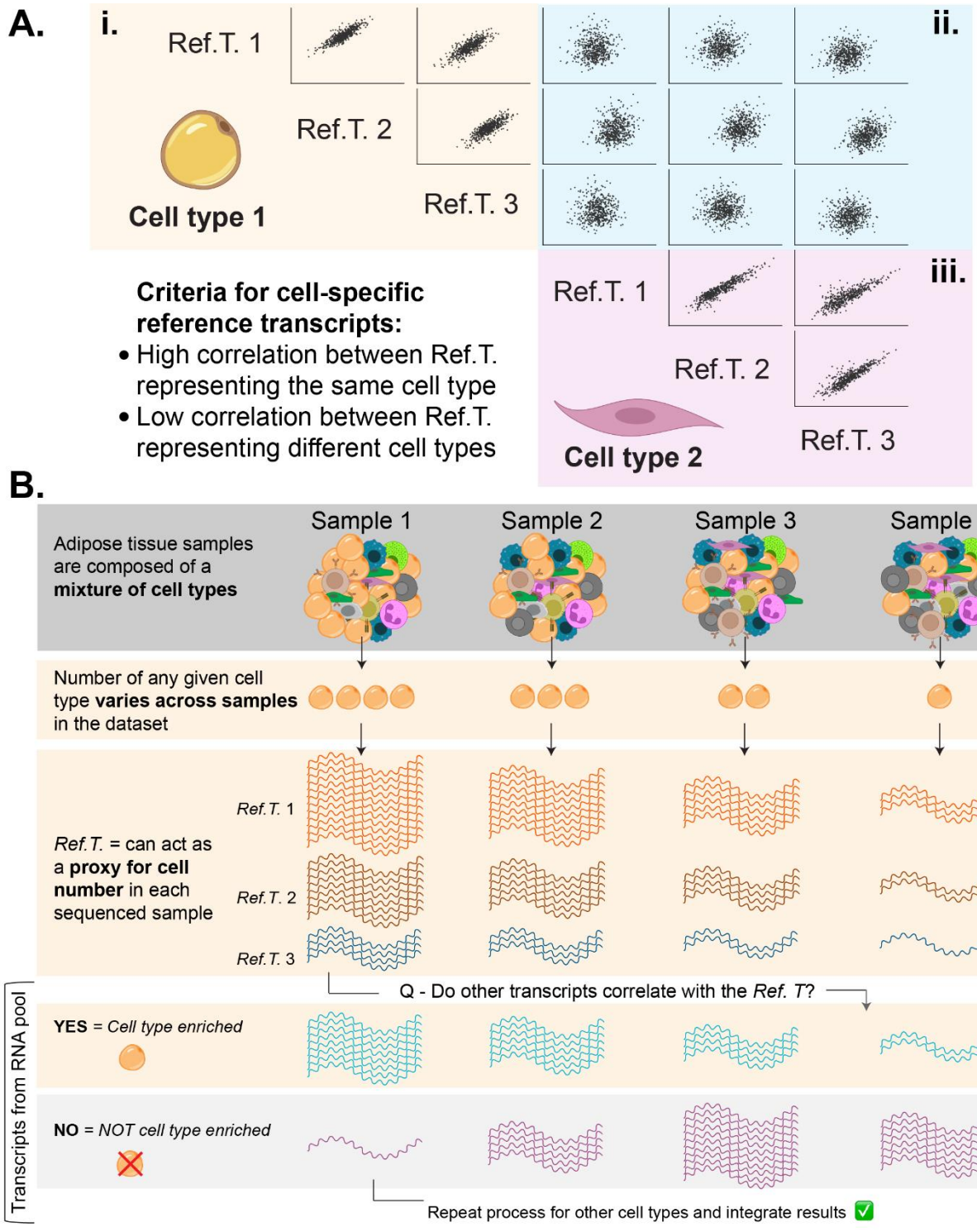

A.
i.
Ref.T. 1
Ref.T. 2
Ref.T. 3
Cell type 1
ii.
Criteria for cell-specific reference transcripts:
• High correlation between Ref.T. representing the same cell type
• Low correlation between Ref.T. representing different cell types
iii.
Ref.T. 1
Ref.T. 2
Ref.T. 3
Cell type 2
B.
Sample 1
Sample 2
Sample 3
Sample 4
Adipose tissue samples are composed of a mixture of cell types
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
Transcripts from RNA pool
Repeat process for other cell types and integrate results
C.
i.
Ref.T.1
Ref.T.2
Ref.T.3
Cell type enriched = High corr. with Ref.T. panel
Transcript 1 from RNA pool
ii.
Ref.T.1
Ref.T.2
Ref.T.3
Not cell type enriched = Low corr. with Ref.T.
Transcript 2 from RNA pool

Figure 2: Concept of integrative correlation analysis to identify cell-enriched transcripts: (A) (i and **iii)** Expression values for cell specific Ref.T. across samples have a high correlation, reflecting the variation in cell type proportion across samples and the consistent ratio between these transcripts (upper light orange or lower light pink plot). **(ii)** Unrelated transcripts will not have the same expression pattern across samples and will therefore not correlate (upper light blue plot). **(B)** RNA sequencing data from unfractionated tissue comprises a pooled transcriptome representative of a mixture of RNA from different constituent cell types, with adipose tissue used as an illustrative example. The proportion of each cell type within different samples (S1-S4) varies due to tissue composition and sampling variation, and cell type specific transcripts (Ref.T.) can act as a proxy for this variation. Other transcripts with the same expression pattern across samples are also likely to be cell enriched, unlike those expressed in one or several other cell types, which lack this relationship with the Ref.T. **(C)** Cell specific Ref.T. correlate **(i)** well with other cell type specifically expressed genes or **(ii)** poorly with those expressed in other, or multiple cell types.

3.1.3 Specifics for paper I

Here, we analysed cell type transcriptome profiles in visceral (VAT) [n=527] and subcutaneous (SAT) [n=646] adipose tissue (sourced from the GTEx portal (release V8, www.gtexportal.org) [126]). Correlation coefficients were calculated between the selected Ref.T. and all other sequenced transcripts. Subsequently, transcripts with a TPM value <0.1 in more than 50% of the samples were excluded (but are still included in data tables), prior to defining cell type enriched transcripts by using a correlation value threshold cut-off, which was either: (i) that above which >95% of transcripts reached this threshold with only that Ref.T. panel or (ii) ≥0.50; whichever was higher) (Figure 3A). False discovery rate (FDR) <0.0001 was required. To rule out potentially dual-enriched transcripts (where transcripts correlated with more than one cell-type reference panel), the 'differential correlation score' was calculated - defined as the difference between the highest mean correlation coefficient with a Ref.T. panel, and the next highest mean correlation coefficient with another Ref.T. panel. Transcripts with a differential correlation value >0.1 were classified as cell type enriched (principles are illustrated in Figure 3A). Various steps were taken to verify the results. Gene ontology (GO) biological process and reactome enrichment analysis was carried out for each cell type enriched list (Gene Ontology Consortium [127] and PANTHER classification resource [128, 129]), to establish if the significant over-represented terms were related to cell type function, which would act as support for the gene classification. For a selection of cell type enriched genes, the cell specificity of the corresponding protein was verified using immunohistochemistry [130, 131]. We also used weighted correlation network analysis (WGCNA) [132] to analyse the same dataset (Figure 3Bi). WGCNA is an unbiased method where pairwise correlation coefficients between all transcripts are calculated and transcripts clustered according to expression pattern similarity. As this analysis is not dependent on the manual selection of any input Ref.T., it can provide support for classification

of our cell type predicted genes if they cluster together. Furthermore, the cell type enrichment profiles were validated by comparison to single cell/nuclei RNAseq data (Figure 3Bii), in the following ways: (i) To demonstrate that genes of interest were assigned to the correct cell type, gene expression data was retrieved from the Tabula Sapiens human cell atlas [13] in order to generate UMAP plots with the R Seurat package [133], and our classification was verified if the cell type annotated cluster with highest expression was the same. (ii) Data from scRNAseq analysis of human SAT (Tabula Sapiens [13] and Hildreth et al. [35]), scRNAseq of mixed depot murine adipose tissue (Tabula Muris [24]) and snRNAseq of human SAT (Sun et al. [134]) were processed using the R Seurat package [133] and the FindAllMarkers function. Cell type enriched genes in this dataset were then classified based on the use of three different thresholds: an average log2 fold change (FC) expression of >0.2, >0.5 or >1 in each annotated cell type, relative to all other cell types within each study. Our cell type classifications were then compared to those found for each single cell/nucleus dataset, to identify numbers of overlapping enriched genes between this dataset and ours. (iii) A hypergeometric test was used to compare cell type enrichment signatures across all four previously mentioned studies and our data, to determine agreement between them.

Adipose tissue depot specific differences were investigated for cell types profiled in both VAT and SAT by calculating the ´depot differential correlation score' (difference between a transcript mean correlation with the Ref.T panel in VAT vs. SAT) for each transcript represented in the cell type respective VAT or SAT enrichment list (Figure 3Ci). A high positive or negative differential correlation score was indicative of enrichment in only one depot, and this was displayed by subsequent plotting against the corresponding enrichment ranking (defined as the position of a transcript in the respective depot enriched list(s), the highest correlating transcript = rank 1, second highest correlating transcript = rank 2…). Female and male differences were investigated within VAT and SAT after the separation of sample data in each depot according to sex, and subsequent generation of sex specific cell type enriched lists, which followed the same principal approach as for the whole VAT and SAT datasets (Figure 3Cii). Sex specific comparisons for a given cell type were performed by calculating the ´sex differential correlation score' (difference between a transcript mean correlation with the Ref.T panel in female vs. male dataset) for each transcript represented in the cell type respective female or male enrichment list, which was then plotted against the corresponding enrichment ranking.

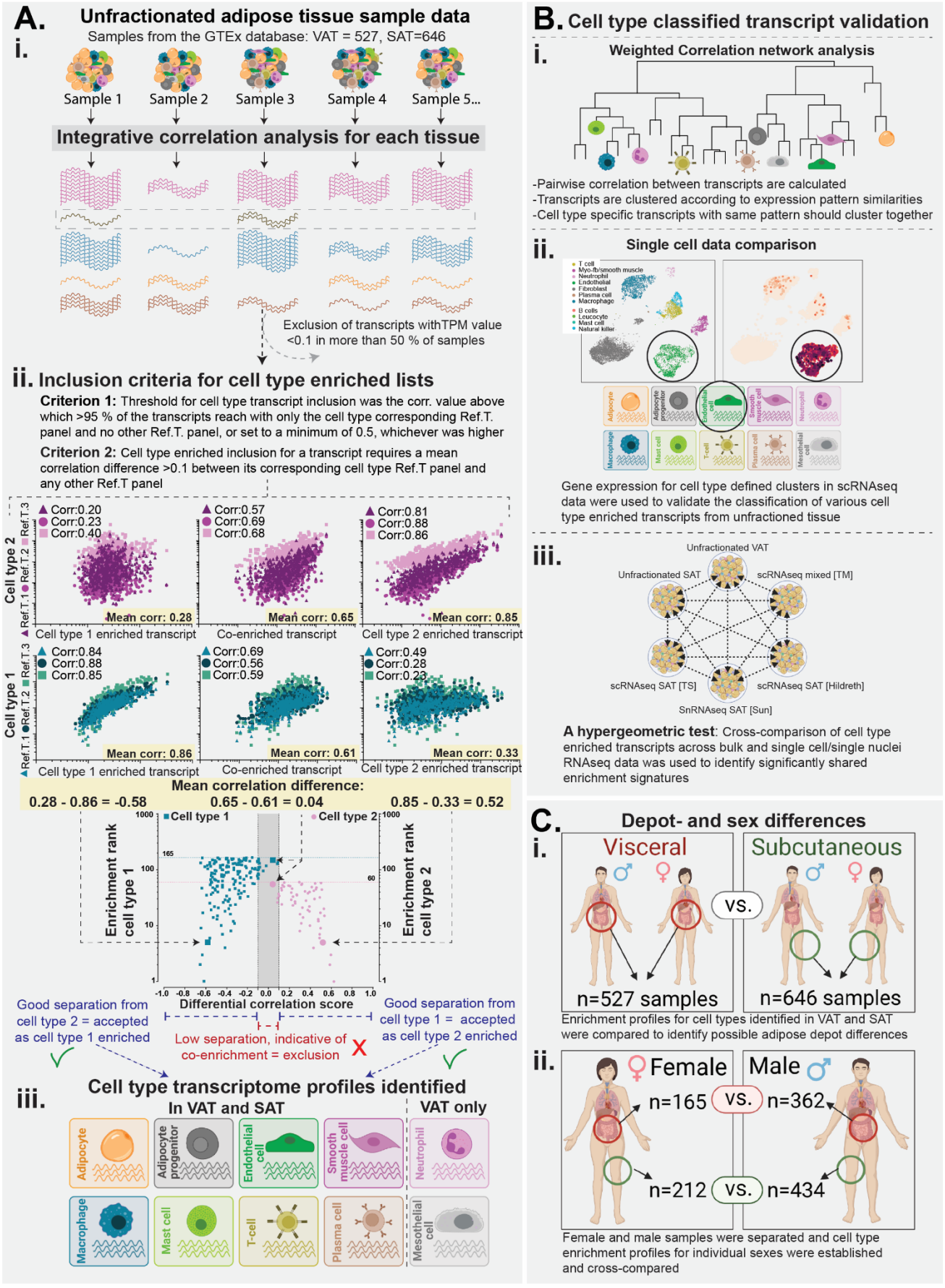

A.
Unfractionated adipose tissue sample data
i.
Samples from the GTEx database: VAT = 527, SAT=646
Sample 1
Sample 2
Sample 3
Sample 4
Sample 5...
Integrative correlation analysis for each tissue
Exclusion of transcripts withTPM value <0.1 in more than 50 % of samples
ii. Inclusion criteria for cell type enriched lists
Criterion 1: Threshold for cell type transcript inclusion was the corr. value above which >95 % of the transcripts reach with only the cell type corresponding Ref.T. panel and no other Ref.T. panel, or set to a minimum of 0.5, whichever was higher
Criterion 2: Cell type enriched inclusion for a transcript requires a mean correlation difference >0.1 between its corresponding cell type Ref.T panel and any other Ref.T panel
Cell type 2
Cell type 1
Ref.T.1
Ref.T.2
Ref.T.3
Corr:0.20
Corr:0.23
Corr:0.40
Corr:0.57
Corr:0.69
Corr:0.68
Corr:0.81
Corr:0.88
Corr:0.86
Mean corr: 0.28
Mean corr: 0.65
Mean corr: 0.85
Cell type 1 enriched transcript
Co-enriched transcript
Cell type 2 enriched transcript
Corr:0.84
Corr:0.88
Corr:0.85
Corr:0.69
Corr:0.56
Corr:0.59
Corr:0.49
Corr:0.28
Corr:0.23
Mean corr: 0.86
Mean corr: 0.61
Mean corr: 0.33
Cell type 1 enriched transcript
Co-enriched transcript
Cell type 2 enriched transcript
Mean correlation difference:
0.28 - 0.86 = -0.58
0.65 - 0.61 = 0.04
0.85 - 0.33 = 0.52
Cell type 1
Cell type 2
Enrichment rank cell type 1
Enrichment rank cell type 2
Differential correlation score
Good separation from cell type 2 = accepted as cell type 1 enriched
Low separation, indicative of co-enrichment = exclusion
Good separation from cell type 1 = accepted as cell type 2 enriched
iii.
Cell type transcriptome profiles identified
In VAT and SAT
VAT only
Adipocyte
Adipocyte progenitor
Endothelial cell
Smooth muscle cell
Neutrophil
Macrophage
Mast cell
T-cell
Plasma cell
Mesothelial cell
B.
Cell type classified transcript validation
i.
Weighted Correlation network analysis
-Pairwise correlation between transcripts are calculated
-Transcripts are clustered according to expression pattern similarities
-Cell type specific transcripts with same pattern should cluster together
ii.
Single cell data comparison
T cell
Myo-fb/smooth muscle
Neutrophil
Endothelial
Fibroblast
Plasma cell
Macrophage
B cells
Leucocyte
Mast cell
Natural killer
Gene expression for cell type defined clusters in scRNAseq data were used to validate the classification of various cell type enriched transcripts from unfractioned tissue
iii.
Unfractionated VAT
Unfractionated SAT
scRNAseq mixed [TM]
scRNAseq SAT [TS]
scRNAseq SAT [Hildreth]
SnRNAseq SAT [Sun]
A hypergeometric test: Cross-comparison of cell type enriched transcripts across bulk and single cell/single nuclei RNAseq data was used to identify significantly shared enrichment signatures
C.
Depot- and sex differences
i.
Visceral
Subcutaneous
vs.
n=527 samples
n=646 samples
Enrichment profiles for cell types identified in VAT and SAT were compared to identify possible adipose depot differences
ii.
Female
Male
n=165
vs.
n=362
n=212
vs.
n=434
Female and male samples were separated and cell type enrichment profiles for individual sexes were established and cross-compared

Figure 3: Concepts for paper 1. (A) Schematic overview of how cell type enriched transcripts were identified: Sequencing data for unfractionated VAT and SAT were retrieved from the GTEx portal, and Spearman correlation coefficients were calculated between the Ref.T. and other sequenced transcripts (I) and based on a given set of criteria for inclusion (ii), transcripts were included into a cell type enriched list (iii) (Differential correlation score = difference between a transcript mean correlation with a Ref.T panel of one cell type vs. another, enrichment rank = the position of a transcript in a cell type enriched list) (B) The classification of cell type enriched genes were validated by: (i) weighted correlation network analysis, (ii) comparison to human subcutaneous scRNAseq expression data (Tabula Sapiens [13]), or (iii) by a hypergeometric test that identify significantly shared transcriptome signatures when cross-comparing all adipose cell type enriched profiles identified in human VAT and SAT bulk RNAseq data in current study (unfractionated VAT, unfractionated SAT), or by single cell/single nuclei profiling of human SAT sourced from Tabula Sapiens (scRNAseq SAT [TS]), Hildreth et al. [35] (scRNAseq SAT [Hildreth]), Sun et al. [134] (snRNAseq SAT [Sun]), or in mixed murine adipose tissue from Tabula Muris [24] (scRNAseq mixed [TM]). (C) Enrichment differences between adipose depots were investigated by comparing enrichment profiles of cell types identified in both (i) VAT and SAT, while sex-specific enrichment differences were defined between (ii) male and female, following male-female subgroup-processing of samples to acquire sex specific enriched lists.

3.1.4 Specifics for paper II

Cell type enrichment profiling was carried out for a total of 15 tissues, including visceral and adipose tissue (tissue types and sample number summarized in Figure 4A) using the same method as for paper I, with some minor modifications to the inclusion criteria. Classification of transcripts into cell type enriched lists were based on following criteria: (i) the transcript had a mean correlation with a given Ref.T. panel ≥0.50 (FDR <0.0001), which was (ii) higher than the mean correlation with any other Ref.T. panel. As given here, the correlation threshold cut-off value was set to a minimum of 0.50 as default but was manually adjusted down if there was no correlation overlap with other cell type Ref.T. panels or adjusted up if there was a marked overlap. (iii) For a transcript to be included into a given cell type enriched list, the 'differential correlation score' - defined as the difference between the highest mean correlation coefficient with the corresponding Ref.T. panel, and the next highest mean correlation coefficient with another Ref.T. panel, was required to be >0.15. Exceptions were made when cell sub-types were profiled individually, where transcripts could be classified as selectively enriched in both, vs. other cell types e.g., pancreatic alpha and beta cells.

Selected enriched transcripts or cell profiles in various tissues were subjected to gene ontology analysis, tissue profiling, and WGCNA, similarly to paper I, to verify the cell type classification. Tabula Sapiens [13] and Human Testis Atlas [135] scRNAseq data was used to generate illustrative UMAP plots for a selection of cell type classified transcripts in the unfractionated datasets.

In addition to resolving the various cell profiles in tissues, this paper focused on comparison of cell type profiles within a tissue, or across tissues, to identify similarities and differences. For example, pancreatic alpha and beta cell enrichment profiles were compared by calculating and plotting the differential correlation scores against the enrichment ranking of transcripts in their respective enriched lists (Figure 4B). Spermatogenesis related enrichment profiles were identified using Ref.T. transcripts selected to represent the temporal order of development; a germ cell expressed panel [*MAGEB2, KDM1B, PIWIL4*] (spermatogonia), a meiotic cell cycle expressed panel [*ANKRD31*, *RBM44, TOP2A*] (spermatocytes), a spermatid structure-related panel [*CEP55, KPNA5*, *PBK*] (round/early elongating spermatids) and a nuclear condensation/protamine repackaging factors panel [*PRM1*, *PRM2*, *TNP1*] (late/elongated spermatids). Enrichment profiles for the specific developmental stages could then be established, along with temporal changes (Figure 4C). A comparison of cell type enrichment signatures across all cell types in all profiled tissues was performed using a hypergeometric test, which can identify significant signature overlaps for various cell types and can therefore be used to indicate common cellular features and functions. (Figure 4E). This analysis makes it possible to compare core cell type enrichment profiles from various tissues, including endothelial cells, which are present across tissue types, to identify those genes that are broadly enriched in a core cell type. Such genes can be of great interest as they might be cell type-defining, relative to genes that are specific to a core cell type subgroup, which might be more related to tissue-specific function.

RNAseq data from unfractionated human tissues can be used to identify genes with higher general expression in any given tissue, compared to others. We can use our our cell type enriched profiling data to predict the cell type from which the tissue enriched transcripts originate. We performed this type of analysis using lists of tissue enriched genes generated from the HPA [10] or GTEx material, the latter sourced from the Harmonizome database [136]. (see Figure 4D for details).

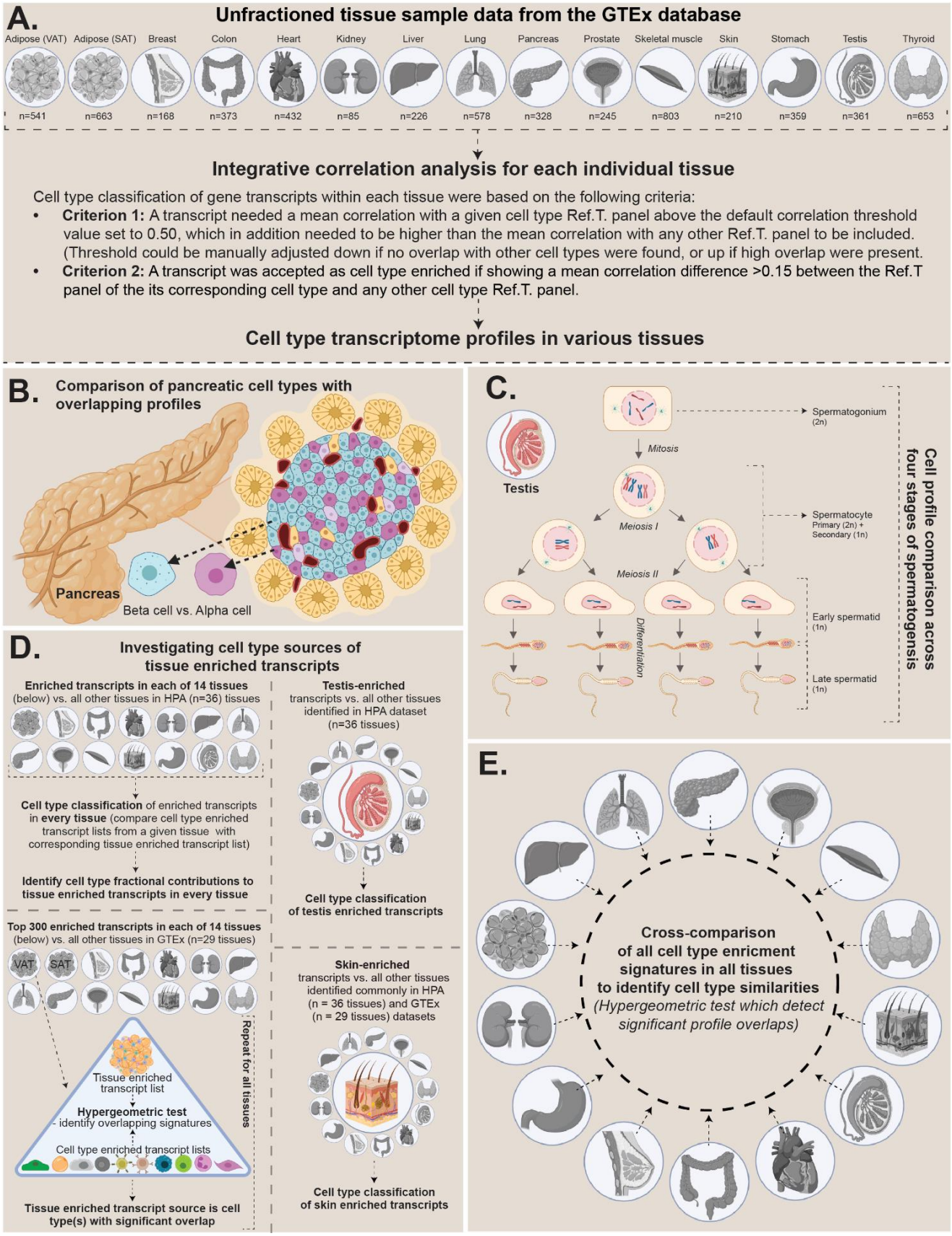

A.
Unfractioned tissue sample data from the GTEx database
Adipose (VAT)
Adipose (SAT)
Breast
Colon
Heart
Kidney
Liver
Lung
Pancreas
Prostate
Skeletal muscle
Skin
Stomach
Testis
Thyroid
n=541
n=663
n=168
n=373
n=432
n=85
n=226
n=578
n=328
n=245
n=803
n=210
n=359
n=361
n=653
Integrative correlation analysis for each individual tissue
Cell type classification of gene transcripts within each tissue were based on the following criteria:
• Criterion 1: A transcript needed a mean correlation with a given cell type Ref.T. panel above the default correlation threshold value set to 0.50, which in addition needed to be higher than the mean correlation with any other Ref.T. panel to be included. (Threshold could be manually adjusted down if no overlap with other cell types were found, or up if high overlap were present.
• Criterion 2: A transcript was accepted as cell type enriched if showing a mean correlation difference >0.15 between the Ref.T panel of the its corresponding cell type and any other cell type Ref.T. panel.
Cell type transcriptome profiles in various tissues
B.
Comparison of pancreatic cell types with overlapping profiles
Pancreas
Beta cell vs. Alpha cell
C.
Testis
Spermatogonium (2n)
Mitosis
Meiosis I
Spermatocyte Primary (2n) + Secondary (1n)
Meiosis II
Early spermatid (1n)
Differentiation
Late spermatid (1n)
Cell profile comparison across four stages of spermatogenesis
D.
Investigating cell type sources of tissue enriched transcripts
Enriched transcripts in each of 14 tissues (below) vs. all other tissues in HPA (n=36) tissues
Cell type classification of enriched transcripts in every tissue (compare cell type enriched transcript lists from a given tissue with corresponding tissue enriched transcript list)
Identify cell type fractional contributions to tissue enriched transcripts in every tissue
Top 300 enriched transcripts in each of 14 tissues (below) vs. all other tissues in GTEx (n=29 tissues)
VAT
SAT
Tissue enriched transcript list
Hypergeometric test - identify overlapping signatures
Cell type enriched transcript lists
Repeat for all tissues
Tissue enriched transcript source is cell type(s) with significant overlap
Testis-enriched transcripts vs. all other tissues identified in HPA dataset (n=36 tissues)
Cell type classification of testis enriched transcripts
Skin-enriched transcripts vs. all other tissues identified commonly in HPA (n = 36 tissues) and GTEx (n = 29 tissues) datasets
Cell type classification of skin enriched transcripts
E.
Cross-comparison of all cell type enricment signatures in all tissues to identify cell type similarities
(Hypergeometric test which detect significant profile overlaps)

Figure 4: Concepts for paper II. (A) Bulk sequencing data for 15 different tissues were retrieved from the GTEx portal, and Spearman correlation coefficients were generated between all protein coding genes and all cell type defined Ref.T.s within each tissue datasets and based on a given set of specific criteria, transcripts were included into a cell type enriched list. Some of the main investigate emphasis points in the paper were **(B)** the enrichment profile overlap between the pancreatic cell types alpha and beta cells, **(C)** identification of temporal enrichment signatures during the four stages of spermatogenesis, **(D**) annotation of the cell source of tissue enriched genes, and **(E)** use of a hypergeometric test to cross-compare all cell type enriched genes in all tissues to identify significant signature similarities.

3.2 Paper III

All specific methods for paper III are provided under the paper method section. This section gives background information related to these methods, and the reasoning behind these choices.

3.2.1 Background information

Paper III focuses the investigation on the adhesion G protein-coupled receptor L4 (ADGRL4). The ligand, interaction partners, activation and induced signalling pathways of ADGRL4 is unknown [137, 138], and there is limited knowledge on its function aside from a small number of reports of a role in angiogenesis and proliferation [138-140]. ADGRL4 expression is largely restricted to endothelial cells (EC) across vascular beds [12, 141-144], indicative of a central role in EC specific phenotype and function. We aimed to gain a wider understanding of ADGRL4 function in normal and inflamed conditions, using gene silencing in primary human umbilical vein endothelial cells (HUVEC) and an initial screening of resultant changes in the proteome profile using bottom-up mass spectrometry approach. For the mass spectrometry analysis, a total of 5 sample sets and 10 samples were used. Three different sample sets from three individual donors represented an unstimulated phenotype, referred to as sample sets 1-3 (SS1-3), while two sample sets from two different donors were exposed to the inflammatory cytokine TNF used for stimulating conditions, referred to as sample set 4-5 (SS4-5). All sample sets included one HUVEC sample receiving ADGRL4 siRNA (siRNA-*ADGRL4-EC*) and one sample with scrambled siRNA (siRNA-control-EC).

3.2.2 General overview of mass spectrometry analysis steps

The mass spectrometry method used, referred to as bottom-up, or shotgun proteomics, is a well-established method which generally relies on liquid chromatography, coupled to tandem mass spectrometry (LC-MS/MS), and can be used to identify thousands of proteins in complex samples like cell lysates. Compared to top-down proteomics, where full size proteins

are analysed, and fewer proteins are detected in a sample, bottom-up proteomics analyses peptides originating from a mixture of digested proteins. The peptides are relative to proteins, more unified in size, shape and charge, which confers preparation and data acquisition advantages and gives rise to simplified fragmentation patterns that are easier to interpretate during data analysis [145].

In order to acquire pure peptides from the HUVEC samples that could be used for mass spectrometry analysis, sample processing steps were performed [145]. Lysis of cells and centrifugation steps were done to separate proteins from non-protein content in the samples. Dithiothreitol was added to the samples, which reduces disulphide bonds, followed by alkylation of cysteine with iodoacetamide, which prevents re-formation of bonds and peptide branching, giving linearized proteins that facilitate downstream digestion, which was carried out using trypsin.

As the purpose of the project was not only to identify proteins, but also to quantify and compare protein content across samples to elucidate on affects from ADGRL4 depletion, isobaric labelling was applied after peptide digestion using the 10-plex tandem mass tag (TMT) isobaric labeling system (Thermo Scientific), which gives relative protein abundance across the samples [146]. The TMT molecules consist of a mass reporter, a cleavable linker region, a mass normalizer, and an amine reactive group. The 10-plex TMT system consists of 10 different tag variants that are created by incorporating heavy stable isotope variants of C12 and N13 in various configurations in the mass reporter and mass normalizer, where the same mass and biochemical properties are found for each resultant complete tag variant, but the mass reporters will upon cleavage show variation in mass. This system makes it possible to multiplex, where samples are chemically labelled with one out of the ten tag variants and then pulled together for combined analysis. As the biochemical properties of identical peptides with different tags are the same, they will show the same elution and ionizing pattern leading to more accurate detection and quantification across samples. The tags come into play during the second mass spectrometry measuring step where peptide ions are fragmented, causing release of mass reporters with unique mass that represents peptides of different samples, and their respective abundance can then be used for relative peptide quantitation.

Following isobaric labelling, the peptide solution was cleaned before the first fractionation method was applied. As the peptide mixture from cells are complex, and liquid chromatographic separation immediately prior to injection into the mass spectrometer is unlikely to separate the tens of thousands of peptides adequately, the peptides were first separated into 72 fractions by the aid of an immobilized pH gradient and isoelectric focusing (IPG-IEF) [147] on a strip, followed by peptide extraction. This results in better total peptide separation and increased likelihood of detecting various unique peptides. Each fraction was separated further by reverse phase liquid chromatography coupled to the mass spectrometer, which is a common step where peptides are separated and eluted in fractions and can be directly ionized and injected into the machine for data acquisition [145]. For tandem mass spectrometry, mass-to-charge ratio data for the precursor peptide ions are obtained first, followed by the fragmentation into smaller ions that are also analysed [145], and this was achieved by using the Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific), and peptide/protein identification was done by the database search algorithm SequestHT under the software platform Proteome Discoverer 1.4 (Thermo Scientific) against the Uniprot human protein database [148].

3.2.3 Identification of differently regulated proteins

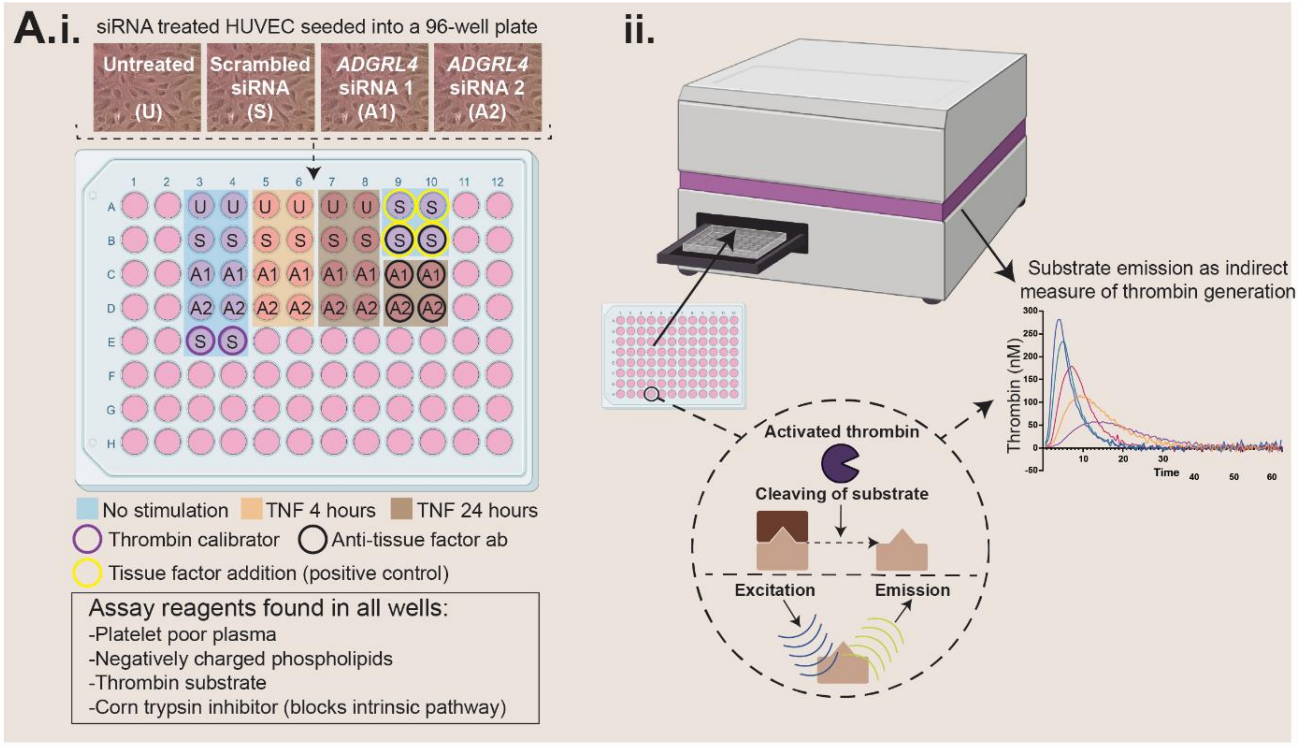
To identify proteins affected by ADGRL4 depletion in HUVEC, the fold change between siRNA-*ADGRL4-EC* and siRNA-control-EC in each sample set were calculated under unstimulated or TNF stimulated conditions, and criteria for upregulation and downregulation defined as fold change ≥ 1.3, and ≤ 0.7 in each of the sample sets, respectively. To identify changes related to ADGRL4 amongst TNF induced proteins, induction was defined as a fold change ≥2 in each of the TNF stimulated siRNA-controls-EC relative to average of the three non-stimulated siRNA-controls-EC.

3.2.4 Validation of data

To validate the findings of mass spectrometry, proteins of particular interest were selected for further analysis and verification, and alternative siRNA sequences were tested to rule out off-target effects. For all selected proteins, relative qPCR was done to determine if protein regulatory patterns were consistent with changes in mRNA or if changes could reflect post-translational modifications and mechanisms directly targeting protein stability and degradation. Western blotting was used as a second detection method to verify the mass

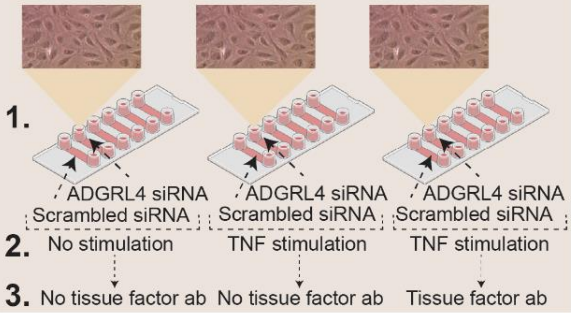
spectrometry results. Additionally, for detection of cell-surface expressed proteins, such as tissue factor and adhesion molecules, flow cytometry was used. This has additional advantages, such as showing if changes were general across the cell population or if they were driven by a subgroup, and it allowed for confirmation of protein external localisation, which could not be achieved by neither mass spectrometry nor western blot. The distinction between internal and external tissue factor is important, as internal tissue factor has little impact on coagulation, while that present on the surface facilitates procoagulant activity.

Tissue factor upregulation alone is not synonymous with an increase in procoagulant activity, even if displayed on the cell surface. Tissue factor can be present in an inactive form, and functional in-house adapted assays were therefore performed to verify activity. A real time thrombin generation assay, normally used to identify inherent procoagulant properties of plasma, was adapted to evaluate if HUVEC cell surface tissue factor could contribute to thrombin activation, and to establish the impact of ADGRL4 depletion on tissue factor activity. The assay was performed with HUVEC seeded into 96-well plates, where plasma, phospholipids and thrombin substrate were added in every well. Plasma is necessary as it contributes with coagulation factors that drive thrombin generation, while phospholipids act as a cofactor for thrombin activation, along with calcium found in the thrombin substrate buffer. The thrombin substrate is cleaved by thrombin and generates a fluorogenic product, used as an indirect measure for thrombin generated in each well. Two wells also received thrombin calibrator reagent with known concentration of thrombin, which was used for calibration of thrombin in all the other wells. Tissue factor was added to control wells, to establish thrombin generation potential, while thrombin generation in all other wells were driven by the presence of cellular surface expressed tissue factor with no additional tissue factor added. Corn trypsin inhibitor was used in all wells to block the intrinsic pathway and to rule out an alternative way for thrombin activation. Tissue factor blocking antibody was used for the purpose of blocking tissue factor activity, and hence blocking thrombin generation, to verify tissue factor as the main driver behind thrombin generation. By the aid of this set-up, thrombin generation under various conditions could be measured, and cell surface tissue factor contribution could be evaluated (overview of method in Figure 5A).



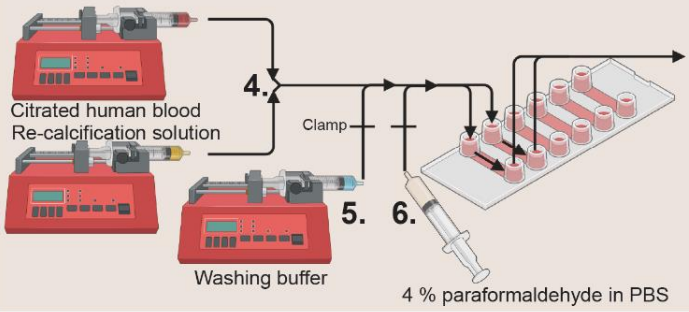
B.i. Cell preparations

1. HUVEC were transfected and then seeded and grown to confluency in Ibidi chamber slides
2. 48 hours post-transfection, cells either recieved a 24 hour TNF treatment, or remained untreated
3. Some TNF stimulated cells underwent treatment with tissue factor blocking antibody (tissue factor ab) before assay start



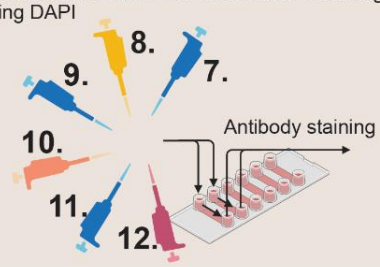
ii. Flow treatment

4. Channels were simultaneously infused with re-calcified citrated human blood at a continuous flow for 10 minutes (shear stress=1 dyn/cm²).
5. A washing step of 8 minutes were performed to remove remnants of blood contaminants (shear stress=0.5 dyn/cm²)
6. The channels were infused with 4% paraforomaldehyde in PBS, and cells were left to fixate for 10 minutes



iii. Antibody staining

7. The cells were blocked in 1% BSA in PBS until further use
8. Primary staining with an in-house produced anti-fibrin mouse monocolonal antibody was done overnight at 4°C
9. The cells were then washed with PBS three times
10. Secondary staining with Alexa Fluor 488 goat anti-mouse was done for 1 hour in room temperature
11. Cells were washed another three times with PBS
12. The channels were filled with Vectashield mounting medium containing DAPI



iv. Imaging

13. Z-stack immunofluorescence images were obtained by the aid of a Leica TCS SPE confocal microscope, equipped with a 63x oil immersion objective.
14. Fibrin surface area coverage (%) and fibrin depth were calculated using the image processing package Fiji/ImageJ

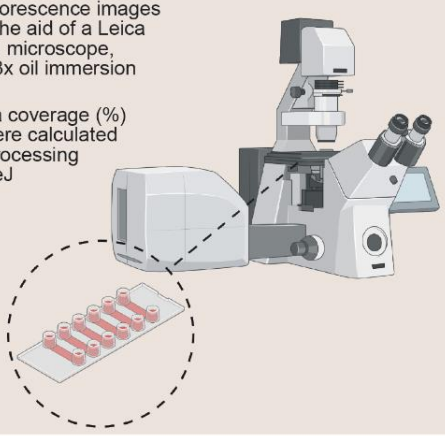


Figure 5: Overview of functional in-house adapted assays used in paper III to study tissue factor activity after ADGRL4 depletion in HUVEC. (A) For a thrombin generation assay, control or ADGRL4-depleted HUVEC were grown to confluence in a 96-well plate before treatment with or without the inflammatory cytokine TNF. Various reagents needed to support and measure thrombin generation was added in all wells, with the exception of tissue factor which was only added in control wells, and is indicated in the figure. CTI was used in order to block the intrinsic pathway and exclude this alternative possibility of thrombin activation, while tissue factor blocking antibody was added in a selection of wells to block tissue factor activity. Thrombin generation was indirectly measured across wells based on the signals from cleaved fluorogenic thrombin substrate. **(B)** Detailed description of a fibrin deposition assay subdivided in four main processes including **(i)** cell preparation, **(ii)** blood flow treatment, **(iii)** antibody staining and **(iv)** imaging.

Fibrin deposition was measured to evaluate effects of ADGRL4 depletion on tissue factor activity, and indirectly, the regulatory effects on tissue factor expression. Here, pre-transfected HUVEC were seeded and grown to confluence in Ibidi chamber slides, where they 48 hours post-transfection received either 24 hours of TNF treatment or no TNF treatment. Some samples also received tissue factor blocking antibody treatment, which was used to verify if tissue factor was the main driver behind fibrin deposition. Samples were exposed to re-calcified citrated human whole blood under continuous flow (shear stress approximately 1 dyn/cm²), followed by a washing step and fixation in 4% paraformaldehyde. For staining, an in-house produced anti-fibrin mouse monoclonal antibody and Alexa Fluor 488 goat anti-mouse were used, and the channels of the slides were in the end infused with Vectashield mounting medium containing DAPI. Immunofluorescence images of fibrin deposition were captured using a Leica TCS SPE confocal microscope, and images were analysed using the image processing package Fiji/ImageJ (a detailed overview is given in Figure 5B)

4. Results

4.1 Paper I: A human adipose tissue cell type transcriptome atlas

Based on an integrative correlation analysis of unfractionated human adipose tissue RNAseq data, the enriched transcriptome profiles of the constituent cell types in visceral (VAT) and subcutaneous (SAT) adipose tissue were defined. In VAT 10 cell types were profiled, and 2343 genes classified as cell type enriched, most of which were protein coding genes (87 %). Cell types with the most enriched genes were the tissue specific adipocytes (n=700), followed by mesothelial cells (n=532) and adipocyte progenitors (n=272), while the other less tissue specific cells; endothelial (n=157), smooth muscle (n=142), T-cells (n=200), plasma cells (n=114), macrophages (n=151), neutrophils (n=56) and mast cells (n=19), had fewer enriched genes. 307 non-coding transcripts, mostly long-non-coding RNAs, were classified as cell-type enriched in VAT, with the highest number found in adipocytes (n=144) and mesothelial cells (n=89). T-cell receptor and immunoglobin genes, were highly represented in T-cells (n=28) and plasma cells (n=99), respectively. For SAT, 8 of the 10 cell types profiled in VAT were also profiled, but low expression and Ref.T. correlations were found for mesothelial cells and neutrophils, indicating low numbers or absence of these cell types in SAT, so these were excluded. Cell type enriched genes in SAT cell types were similar to those in VAT, with the highest number of classified in adipocytes (n=715) and adipocyte progenitors (n=722), followed by smooth muscle (n=348), macrophages (n=256), endothelial (n=155), T-cells (n=116), plasma cells (n=75) and mast cells (n=31). Our classifications were supported by weighted network correlation analysis (WGCNA), where the majority of cell type classified genes clustered together. Furthermore, gene ontology (GO) and reactome analysis resulted in significant terms that were consistent with known cell functions. The enrichment profiles were also compared to three scRNAseq data sets including human subcutaneous [13, 35] and murine [36] adipose tissue in addition to a snRNAseq data set of human subcutaneous adipose tissue [134]. Even though none of these studies contained all the cell types we profiled, and adipocytes were only present in the snRNAseq data, our cell type enriched gene classifications were in good agreement with one or more of these studies. Sex specific differences in cell profiles were investigated using a subset analysis. Cell profiles were similar between sexes, and when genes were classified as male or female only, the difference was typically driven by a gene falling marginally below the designated threshold for classification as enriched in one sex. However, a small number of Y-linked male only cell

type-enriched transcripts were identified. A comparison of enrichment profiles was also carried out between cell types in both VAT and SAT, to identify potential depot specific differences. Also here, the expression profiles were generally similar. A small number of genes were cell type enriched in only one depot, which also in most cases corresponded with a higher expression value in that same depot. The main driver of cell specific depot difference was related to the presence of additional cell types in VAT, especially mesothelial cells.

4.2 Paper II: A tissue centric atlas of cell type transcriptome enrichment signatures

Based on bulk RNAseq data from more than 6000 samples, integrative correlation analysis was expanded to resolve the enriched protein coding transcriptome profile of the main constituent cell types in 15 different human tissues. The following tissues and number of cell types (n) were profiled: Subcutaneous (n=8) and visceral (n=10) adipose tissue, breast - mammary tissue (n=10), colon - sigmoid (n=12), heart - left ventricle (n=9), kidney - cortex (n=9), liver (n=10), lung (n=14), pancreas (n=9), prostate (n=9), muscle - skeletal (n=7), skin (n=18), stomach (n=11), testis (n=9) and thyroid (n=10). For cell type enriched lists in lung tissue, gene ontology analysis returned over-represented terms that corresponded with known cell type functions, and tissue profiling and scRNAseq data comparison [13] for selected genes, along with WGCNA, supported the cell type enriched classification. Across tissues, tissue specialized cells had the most enriched genes; of the total cell type classified transcripts in heart, kidney, liver, skin, stomach, and lung, cardiomyocytes represented 48%, proximal tubular cells 37%, hepatocytes 53%, keratinocytes 38.4%, gastric mucosal cells 28% and respiratory ciliated cells 28%, respectively. Of the 19,634 protein coding genes expressed in one or more tissues, 5644 (28.7%) had no cell type enrichment in any tissue, and GO analysis indicated housekeeping functions for this group of genes. 5979 (30.4%) were classified as cell type enriched in only a single tissue, with 3141 of these in the testis. 8011 transcripts were cell type enriched in two or more tissues, and of these, 741 (9.2%) were enriched in at least seven tissues. Alpha and beta cells in the pancreas had 131 co-expressed enriched genes, with 91 and 69 transcripts specific to only alpha or beta cells, respectively, and these classifications were well-supported by scRNAseq data for human pancreas [13] and tissue protein profiling. In testis, changes in cell-type transcriptomes during spermatogenesis were profiled over four developmental stages: spermatogonia (S1), spermatocytes (S2), early spermatids (S3) and late spermatids (S4). The individual Ref.T.s within each respective panel clustered well together in WGCNA analysis. 6179 transcripts

were identified as enriched in one or more of the four developmental stages, versus non-germ cells. It was observed that the germ cell enriched signatures could be separated into either stage-specific enrichment, or specific enrichment signatures common to two or more ensuing stages, and findings were again supported by scRNAseq data [135], GO analysis and tissue protein profiling. Germ cells were the main cell type source for genes identified as enriched in testis versus other tissues in the HPA data. Skin was the tissue type with most profiled cell types (n=18), and keratinocytes, separated into those found in the sub granular and granular layers, had the highest proportion of cell type enriched transcripts, 737 and 208 respectively. Melanocytes had the fewest enriched transcripts of those profiled (n= 17). A list of transcripts identified to be enriched in skin versus other tissues in the HPA tissue section [10] and the GTEX dataset [126], latter sourced from the Harmonizome database [136], were compared to our cell type enrichment classifications. We classified 63% of the skin specific transcripts as cell type enriched, mostly in hair root cell types and keratinocytes. Of the skin enriched genes not classified as cell type enriched, most were co-enriched amongst several cell types in the hair root, showing that hair root cells, a minority cell group in skin, express most of the genes defined as skin specific. A cross-tissue analysis, performed as a hypergeometric test, was done to investigate similarities through shared signatures between all cell types in the tissues profiled. Organ specific cell types with related functions in multiple tissues showed a high similarity in enriched gene profiles, e.g., endocrine cells from several tissues (alpha and beta cells from pancreas, enteroendocrine cells from the colon and stomach, and parafollicular cells from the thyroid). There was a significant signature overlap between adipocytes (visceral, subcutaneous and breast tissue), skin sebaceous gland cells, liver hepatocytes, and kidney proximal tubular cells, and these genes were linked to metabolic processes according to GO analysis, suggestive of processes they have shared participation in. Other genes uniquely enriched in only one of these cell types, are more likely to have tissue specialised functions. Late spermatids in the testis and respiratory ciliated cells in the lung shared a significant signature of as many as 441 transcripts, which was indicated by GO analysis to be related to cilia, which are present on both cell types. A total of 8 cell types, defined as core cell types, were profiled in all or most tissue types, and included endothelial cells (n=15 tissues), smooth muscle cells (n=14), fibroblasts (n=14), macrophages (n=14), neutrophils (n=8), mast cells (n=5), T-cells (n=13) and plasma cells (n=14). All over, there was a significant and specific shared gene enrichment signature for the same core cell type across tissues. Liver sinusoidal endothelial cells were an interesting outlier of the otherwise similarly grouped endothelial cells, as more than half of the enriched

gene list were not enriched in endothelial cells in other tissues, which could be explained by highly tissue specific functions of this endothelial subgroup. The identification of defining gene signatures for these core cell types were based on classification in at least half of the tissues profiled.

4.3 Paper III: ADGRL4 modifies endothelial cell response to TNF-induced pro-thrombotic phenotype

The adhesion G protein-coupled receptor L4 (ADGRL4) has previously been identified by our group as an endothelial enriched gene [12, 141-144], with potential importance for endothelial specific phenotype and function. ADGRL4 has a role in angiogenesis and proliferation [138-140], but has no known ligand, which makes investigation of protein activation and signal-transduction challenging [137, 138]. We carried out mass spectrometry protein profiling of ADGRL4 silenced endothelial cells (EC), with or without TNF stimulation, to gain insight into the ADGRL4 functional role in EC. Depletion of ADGRL4 in resting EC induced changes in the protein profile. 92 proteins, associated with cell structure and morphogenesis, were upregulated, while a cell-cycle related cluster of 177 proteins were downregulated. The inflammatory cytokine TNF, which causes EC activation [46, 47], induced a pro-inflammatory signature, with 70 proteins increased by at least two-fold. 5 proteins were identified as TNF-induced and further upregulated by ADGRL4-silencing, the most highly upregulated of which was tissue factor – coagulation factor III (F3). Several additional assays were performed to validate this observation, and similar changes were observed on the mRNA level by qPCR and on a protein level by western blot and flow cytometry. ADGRL4 silencing had most impact on mRNA and protein expression in the later phase of the TNF response, indicating that the resolution of tissue factor expression, rather than the induction, was primarily affected. In a functional study, we showed that thrombin generation was augmented by tissue factor upregulation as a result of ADGRL4 depletion, evident by a shortened thrombin generation lag-time, the thrombin peak time and height. Similarly, we observed an increase in fibrin deposition in an in-house assay, where EC were exposed to re-calcified citrated human whole blood under flow. Furthermore, the mass spectrometry data identified 16 proteins that were TNF-induced, but inhibited following ADGRL4 silencing. These proteins were mainly linked to the interferon signalling pathway. Changes in selected candidates, including VCAM1, interferon-induced GTP-binding protein Mx1 (MX1), Interferon-stimulated gene 20 kDa protein (ISG20) and interferon-induced protein with tetratricopeptide repeats 1 (IFIT1) were

validated by qPCR and western blot. Interferon regulatory factor 1 (IRF1), which controls the transcription of interferon-inducible genes [145], was not affected by ADGRL4 silencing. The mass spectrometry data did not indicate any strong and obvious candidates that could connect ADGRL4 with the regulatory outcome of the various proteins, but several proteins involved in signalling transduction, including dual specificity mitogen-activated protein kinase 1 and 6 (MAP2K1 and MAP2K6) and interferon-induced, double-stranded RNA-activated protein kinase (EIF2AK2), were identified as downregulated and could potentially be downstream effectors of ADGRL4. Collectively, these results indicate a currently unappreciated role for ADGRL4 in the EC response to inflammatory stimulation.

5. Discussion

5.1 Paper I and II

To be able to understand the complexity of the cell, and its contribution to tissue organization and function in health and disease, it is necessary to identify and characterize cell specifically expressed genes, as they control cellular phenotype. A problematic aspect of today's research is the tendency to focus on a limited set of already well studied proteins, while thousands remain poorly characterized. This trend can limit novel discoveries and progress in medical science, and there have been recent attempts to address this issue [149-152]. In this thesis, it is demonstrated that cell specific genes can be identified from existing material, and we show how one can use this information to select targets for functional studies. In paper I and paper II, we used a bioinformatics approach to analyse bulk RNAseq data to identify the cell type specific transcriptome profile for tissue constituent cell types, including our main cell type of interest - the vascular endothelial cell (EC). Within the EC enriched gene lists generated from adipose visceral and subcutaneous tissue (paper I), and across 13 other tissues (paper II), we observed genes that were consistently found as EC enriched, indicative of an important role in EC phenotype or function. These genes included ADGRL4, which was already identified in 2016 by our group as an uncharacterized EC enriched gene [141], using an earlier simplified version of the bioinformatics approach reported here. As ADGRL4 had high EC specificity and poorly described function, we investigate its potential role in EC, and identified new novel functions for this protein - of which results are presented in paper III.

In paper I and paper II, in addition to profiling individual cell type enriched transcriptomes, we compared cell profiles within and across tissues to identify similarities and differences between enrichment signatures. We verified our findings to single cell sequencing data where available, which supported our classifications. The use of unfractionated tissue data to identify cell type profiles has several advantages; the data is free and available for public use, which saves time and money, as there is no need for data acquirement. There is also an ethical aspect of re-using data; use of existing resources should be maximized by answering new research questions using existing material where possible. It is also important to remember that patients make the research possible, and by using these data in as many ways as possible, we are honouring their contribution. Another advantage is that public repositories, like the GTEx project, accumulate data from hundreds of samples over time,

which allows the type of analysis carried out in paper I and II, particularly for the detection and classification of lowly expressed genes. Through the analysis of bulk RNAseq data in paper I, many non-coding transcripts were identified as cell type enriched, but a considerable fraction were not identified in scRNAseq data, or they were expressed in too few cells to, with certainty, confirm cell type enrichment. In paper II, many protein coding transcripts with relatively low expression values across samples were classified as cell type enriched, but notably, many of these were lowly represented in scRNAseq data, implying that bulk RNAseq could be more sensitive for the classification of lowly detected gene transcripts. Indeed, single cell sequencing has more zero values with a necessary compromise between number of cells analysed and read depth [26, 153-156]. Enrichment analysis based on unfractionated tissue also offers the advantage of profiling cell types that are difficult to process for single cell sequencing, like adipocytes [34] and cardiomyocytes [21]. Our data could be a useful source of input for deconvolution methods that relays on gene expression matrices to estimate the cell proportions in tissue, such as CIBERSORT and BayesPrism [157, 158], as our data includes cell types that are often difficult to profile by scRNAseq, and might additionally offer a wider gene coverage. Cell type profiling by integrative correlation analysis makes it possible to identify common enrichment signatures for cell types identified across tissues of which can be cell phenotype defining, and to further dissect identity gene futures for cell type subgroups, exemplified with the panel of genes that specifically separated liver sinusoidal EC from EC in other tissues, which are likely relevant for tissue specific functions and not required by other endothelial cells. Information like this can be particularly useful when studying disease, where changes in gene or protein expression can be traced to a cell type source. This could increase understanding of disease mechanisms, and aid treatment developments.

Our integrative correlation analysis does have limitations: The performance and specificity of the analysis depends on the choice of Ref.T., which need to be specific for the cell type they are representing. The Ref.T. panel for each cell type were therefore carefully evaluated and selected for based on high intra-panel correlation and low inter-panel correlation. The use of reference transcripts to identify cell type enriched transcripts is limited to known cell types, and thus new cell types cannot be discovered. Profiling of cell sub-types, which may not have a large panel of uniquely expressed genes, but rather differing levels of the same genes, is difficult using this method. It should be further noticed that genes classified as enriched in our studies are based on thresholds that are guiding, and data for individual genes can give an indication to cell type enrichment profile, even in the absence of strict classification as such.

Genes that are expressed only in specific cell types, but are strongly regulated by environmental or hereditary factors, may not correlate with the corresponding Ref.T.s, thus representing a false negative. In summary, our integrative correlation analysis identified constituent cell type enrichment profiles across 15 different tissues, providing a resource to complement existing scRNAseq data, available through the Human Protein Atlas website in the ´Tissue Cell Type´ section (www.proteinatlas.org/humanproteome/tissue+cell+type).

5.2 Paper III

We selected the adhesion G protein coupled receptor ADGRL4 for functional exploration. This was based on its identification as a largely uncharacterized endothelial enriched gene, in a previous study from our group [141], with further more extensive confirmation in paper I and paper II in this thesis [142-144]. ADGRL4 function is largely undescribed, and thus we investigated its role in HUVEC, primary endothelial cells, which originate from human umbilical cords that do not require invasive procedures to be extracted, and are a commonly used endothelial model [159]. We used HUVEC at an early passage, to minimize loss of their phenotypic characteristics [160]. It should here be mentioned, that HUVEC are found to have different expression patterns relative to other EC of adult origin [161], but this is likely also true when comparing EC from various vascular beds, and the results extracted from the use of any type of EC, cannot automatically be inferred as general for all EC.

A limiting factor for the mass spectrometry analysis, was the number of samples that could be analysed concurrently. The TMT 10-plex allowed for multiplexing of 10 samples, the addition of any more, would have required a second batch for analysis, introducing batch (run-to-run) variability, and also likely introduced missing values (the same precursor ions might not be selected for in the different runs) [162, 163]. Such analysis would also have required an internal reference standard for the normalization of values across batches and would have been more costly and resource demanding. Thus, we limited the analysis to 10 samples in one batch. With three replicates of ADGRL4 and scrambled siRNA treated samples for unstimulated conditions and only two for stimulated conditions, statistical calculations would have had little power, and was therefore not applied. This is a weakness, but the aim of this part of the study was explorative in nature, designed to identify possible targets to study further.

Our study identified several biological processes and pathways that potentially related to ADGRL4 function, but further work is required to decipher the underlying mechanisms, such as key signalling pathways and transcription factors. Further investigation of the role of IRF1, which is known to transcribe several proteins identified to be downregulated by ADGRL4, such as ISG20, IFIT1, MX1 and VCAM1 [164-166] is warranted. Even though IRF1 expression was not modified by ADGRL4 depletion, it is possible that other mechanisms, such as post translational modifications, could affect IRF1 nuclear localization or transcriptional capacity [167-171] and subsequent expression of interferon pathway related proteins. IRF1 localization could be studied using confocal microscopy imaging, or by measuring protein content in separated cell compartments (nucleus and cytoplasm). MAP2K1 and MAP2K6, kinases involved in signalling transduction [172-177], were downregulated by ADGRL4 deletion, along with the kinase EIF2AK2. EIF2AK2 has a role in viral responses, where it inhibits general translation [178, 179], in signalling responses through phosphorylation and activation of MAP2K6 and p38-MAPKs [180, 181] and in the modulation of interferon responses [182, 183] and NFKB activity [184-186]. Considering the potentially wide array of targets these proteins might have, and the substantial impact on cellular functions and responses they might elicit, they can be worth investigating further. For example, if knock down by siRNA for any of these proteins recapitulate any of the ADGRL4-related findings, they could possibly be indicated as downstream effectors of ADGRL4. In the end, ADGRL4 has been identified as a regulator of various biological processes in EC which has not previously been identified, and although there are questions left to be answered, this information can provide useful information that might complement research done in the future or inspire to further investigation.

The work carried out in this thesis shows how cell specific genes can be identified from existing datasets – highlighting the potential applications of publicly available resources to generate new knowledge. The data generated here are available on the Human Protein Atlas website, providing a further resource to complement existing atlas style datasets. Many genes in our identified cell type enriched panels were uncharacterised or not well understood, presenting interesting candidates for functional investigation. Thus this thesis also illustrates how the generation of such data can be used as a rational for the selection of proteins for functional characterisation in a specific context.

6. Conclusions

1. Integrative correlation analysis identified enrichment profiles for constituent cell types in visceral and subcutaneous adipose tissue, and established that these depots broadly were similar, with mesothelial cells as the main driver of depot differences. Sex-specific differences, which were few, were related to Y-linked genes only enriched in male.

2. Enrichment profiles of constituent cell types in 15 tissues were identified. Common and specific gene signatures for cell types across tissues were furthermore established, and similarities were found across cell types with related functions. Core signatures for cell types profiled in many tissues were defined and included endothelial cells, smooth muscle cells, fibroblasts, macrophages, neutrophils, mast cells, T-cells and plasma cells. The generated enrichment data are publicly available through the Human Protein Atlas.

3. ADGRL4 depletion in endothelial cells caused significant changes in the protein profile of endothelial cells, which after TNF-induction was recognized by augmentation of the procoagulant activity in endothelial cells caused by upregulation of tissue factor. A targeted downregulation of interferon pathway proteins was observed under the same conditions.

7. References

1. Nurk, S., et al., *The complete sequence of a human genome.* Science, 2022. **376**(6588): p. 44-53.
2. Arang, R., et al., *The complete sequence of a human Y chromosome.* bioRxiv, 2022: p. 2022.12.01.518724.
3. Barrero, M.J., S. Boué, and J.C. Izpisúa Belmonte, *Epigenetic mechanisms that regulate cell identity.* Cell Stem Cell, 2010. **7**(5): p. 565-70.
4. Marsit, C.J., *Influence of environmental exposure on human epigenetic regulation.* Journal of Experimental Biology, 2015. **218**(1): p. 71-79.
5. del Blanco, B. and A. Barco, *Impact of environmental conditions and chemicals on the neuronal epigenome.* Current Opinion in Chemical Biology, 2018. **45**: p. 157-165.
6. Pacchierotti, F. and M. Spanò, *Environmental Impact on DNA Methylation in the Germline: State of the Art and Gaps of Knowledge.* BioMed Research International, 2015. **2015**: p. 123484.
7. Nilsson, E., M. Ben Maamar, and M.K. Skinner, *Environmental impacts on sperm and oocyte epigenetics affect embryo cell epigenetics and transcription to promote the epigenetic inheritance of pathology and phenotypic variation.* Reproduction, Fertility and Development, 2021. **33**(2): p. 102-107.
8. Bloom, A.B. and M.H. Zaman, *Influence of the microenvironment on cell fate determination and migration.* Physiological Genomics, 2014. **46**(9): p. 309-314.
9. Bhat, R. and M.J. Bissell, *Of plasticity and specificity: dialectics of the micro- and macro-environment and the organ phenotype.* Wiley Interdiscip Rev Membr Transp Signal, 2014. **3**(2): p. 147-163.
10. Uhlén, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
11. Robert, Y.Y., et al., *A systematic survey of human tissue-specific gene expression and splicing reveals new opportunities for therapeutic target identification and evaluation.* bioRxiv, 2018: p. 311563.
12. Karlsson, M., et al., *A single-cell type transcriptomics map of human tissues.* Sci Adv, 2021. **7**(31).
13. Jones, R.C., et al., *The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans.* Science, 2022. **376**(6594): p. eabl4896.
14. Floros, J., et al., *Human Surfactant Protein SP-A1 and SP-A2 Variants Differentially Affect the Alveolar Microenvironment, Surfactant Structure, Regulation and Function of the Alveolar Macrophage, and Animal and Human Survival Under Various Conditions.* Frontiers in Immunology, 2021. **12**.
15. Kishore, U., et al., *Surfactant proteins SP-A and SP-D: structure, function and receptors.* Mol Immunol, 2006. **43**(9): p. 1293-315.
16. Wessels, A., et al., *The Ontogenesis of Myosin Heavy Chain Isoforms in the Developing Human Hearta.* Annals of the New York Academy of Sciences, 1990. **588**(1): p. 461-464.
17. Reiser, P.J., et al., *Human cardiac myosin heavy chain isoforms in fetal and failing adult atria and ventricles.* American Journal of Physiology-Heart and Circulatory Physiology, 2001. **280**(4): p. H1814-H1820.
18. Komarova, Y.A., et al., *Protein Interactions at Endothelial Junctions and Signaling Mechanisms Regulating Endothelial Permeability.* Circ Res, 2017. **120**(1): p. 179-206.
19. Ivetic, A., H.L. Hoskins Green, and S.J. Hart, *L-selectin: A Major Regulator of Leukocyte Adhesion, Migration and Signaling.* Front Immunol, 2019. **10**: p. 1068.
20. Pontén, F., et al., *A global view of protein expression in human cells, tissues, and organs.* Molecular Systems Biology, 2009. **5**(1): p. 337.
21. Hegenbarth, J.-C., et al., *Perspectives on Bulk-Tissue RNA Sequencing and Single-Cell RNA Sequencing for Cardiac Transcriptomics.* Frontiers in Molecular Medicine, 2022. **2**.
22. Hedlund, E. and Q. Deng, *Single-cell RNA sequencing: Technical advancements and biological applications.* Mol Aspects Med, 2018. **59**: p. 36-46.
23. Potter, S.S., *Single-cell RNA sequencing for the study of development, physiology and disease.* Nature Reviews Nephrology, 2018. **14**(8): p. 479-492.
24. Tabula Muris, C., et al., *Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris.* Nature, 2018. **562**(7727): p. 367-372.

25. Hu, T., et al., *Next-generation sequencing technologies: An overview.* Human Immunology, 2021. **82**(11): p. 801-811.
26. Stark, R., M. Grzelak, and J. Hadfield, *RNA sequencing: the teenage years.* Nat Rev Genet, 2019. **20**(11): p. 631-656.
27. Head, S.R., et al., *Library construction for next-generation sequencing: Overviews and challenges.* BioTechniques, 2014. **56**(2): p. 61-77.
28. Kulkarni, A., et al., *Beyond bulk: a review of single cell transcriptomics methodologies and applications.* Curr Opin Biotechnol, 2019. **58**: p. 129-136.
29. Gross, A., et al., *Technologies for Single-Cell Isolation.* Int J Mol Sci, 2015. **16**(8): p. 16897-919.
30. Zhou, W.-m., et al., *Microfluidics applications for high-throughput single cell sequencing.* Journal of Nanobiotechnology, 2021. **19**(1): p. 312.
31. Ziegenhain, C., et al., *Comparative Analysis of Single-Cell RNA Sequencing Methods.* Molecular Cell, 2017. **65**(4): p. 631-643.e4.
32. Adam, M., A.S. Potter, and S.S. Potter, *Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development.* Development, 2017. **144**(19): p. 3625-3632.
33. Massoni-Badosa, R., et al., *Sampling time-dependent artifacts in single-cell genomics studies.* Genome Biology, 2020. **21**(1): p. 112.
34. Deutsch, A., et al., *The Impact of Single-Cell Genomics on Adipose Tissue Research.* Int J Mol Sci, 2020. **21**(13).
35. Hildreth, A.D., et al., *Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity.* Nat Immunol, 2021. **22**(5): p. 639-653.
36. *Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris.* Nature, 2018. **562**(7727): p. 367-372.
37. Bakken, T.E., et al., *Single-nucleus and single-cell transcriptomes compared in matched cortical cell types.* PLoS One, 2018. **13**(12): p. e0209648.
38. Thrupp, N., et al., *Single-Nucleus RNA-Seq Is Not Suitable for Detection of Microglial Activation Genes in Humans.* Cell Rep, 2020. **32**(13): p. 108189.
39. Krüger-Genge, A., et al., *Vascular Endothelial Cell Biology: An Update.* Int J Mol Sci, 2019. **20**(18).
40. Gurevich, D.B., et al., *Endothelial Heterogeneity in Development and Wound Healing.* Cells, 2021. **10**(9).
41. Qiu, J. and K.K. Hirschi, *Endothelial Cell Development and Its Application to Regenerative Medicine.* Circ Res, 2019. **125**(4): p. 489-501.
42. Hellenthal, K.E.M., L. Brabenec, and N.M. Wagner, *Regulation and Dysregulation of Endothelial Permeability during Systemic Inflammation.* Cells, 2022. **11**(12).
43. Claesson-Welsh, L., E. Dejana, and D.M. McDonald, *Permeability of the Endothelial Barrier: Identifying and Reconciling Controversies.* Trends Mol Med, 2021. **27**(4): p. 314-331.
44. Aird, W.C., *Phenotypic heterogeneity of the endothelium: I. Structure, function, and mechanisms.* Circ Res, 2007. **100**(2): p. 158-73.
45. Stan, R.V., et al., *The diaphragms of fenestrated endothelia: gatekeepers of vascular permeability and blood composition.* Dev Cell, 2012. **23**(6): p. 1203-18.
46. Pober, J.S., *Endothelial activation: intracellular signaling pathways.* Arthritis Res, 2002. **4 Suppl 3**(Suppl 3): p. S109-16.
47. Cotran, R.S. and J.S. Pober, *Cytokine-endothelial interactions in inflammation, immunity, and vascular injury.* J Am Soc Nephrol, 1990. **1**(3): p. 225-35.
48. Øynebråten, I., et al., *Rapid chemokine secretion from endothelial cells originates from 2 distinct compartments.* Blood, 2004. **104**(2): p. 314-20.
49. Hillyer, P., et al., *Chemokines, chemokine receptors and adhesion molecules on different human endothelia: discriminating the tissue-specific functions that affect leucocyte migration.* Clin Exp Immunol, 2003. **134**(3): p. 431-41.

50. Barzilai, S., et al., *M-sec regulates polarized secretion of inflammatory endothelial chemokines and facilitates CCL2-mediated lymphocyte transendothelial migration.* Journal of Leukocyte Biology, 2016. **99**(6): p. 1045-1055.
51. Haraldsen, G., et al., *Cytokine-regulated expression of E-selectin, intercellular adhesion molecule-1 (ICAM-1), and vascular cell adhesion molecule-1 (VCAM-1) in human microvascular endothelial cells.* J Immunol, 1996. **156**(7): p. 2558-65.
52. Fries, J.W., et al., *Expression of VCAM-1 and E-selectin in an in vivo model of endothelial activation.* Am J Pathol, 1993. **143**(3): p. 725-37.
53. Nightingale, T.D., et al., *Tuning the endothelial response: differential release of exocytic cargos from Weibel-Palade bodies.* Journal of Thrombosis and Haemostasis, 2018. **16**(9): p. 1873-1886.
54. Gotsch, U., et al., *Expression of P-selectin on endothelial cells is upregulated by LPS and TNF-alpha in vivo.* Cell Adhes Commun, 1994. **2**(1): p. 7-14.
55. Filippi, M.D., *Mechanism of Diapedesis: Importance of the Transcellular Route.* Adv Immunol, 2016. **129**: p. 25-53.
56. Wolburg, H., K. Wolburg-Buchholz, and B. Engelhardt, *Diapedesis of mononuclear cells across cerebral venules during experimental autoimmune encephalomyelitis leaves tight junctions intact.* Acta Neuropathol, 2005. **109**(2): p. 181-90.
57. Martinelli, R., et al., *Probing the biomechanical contribution of the endothelium to lymphocyte migration: diapedesis by the path of least resistance.* J Cell Sci, 2014. **127**(Pt 17): p. 3720-34.
58. Marmon, S., et al., *Caveolin-1 expression determines the route of neutrophil extravasation through skin microvasculature.* Am J Pathol, 2009. **174**(2): p. 684-92.
59. Eppihimer, M.J., et al., *Heterogeneity of expression of E- and P-selectins in vivo.* Circ Res, 1996. **79**(3): p. 560-9.
60. Petzelbauer, P., et al., *Heterogeneity of dermal microvascular endothelial cell antigen expression and cytokine responsiveness in situ and in cell culture.* J Immunol, 1993. **151**(9): p. 5062-72.
61. Yano, K., et al., *Vascular endothelial growth factor is an important determinant of sepsis morbidity and mortality.* J Exp Med, 2006. **203**(6): p. 1447-58.
62. Ley, K. and P. Gaehtgens, *Endothelial, not hemodynamic, differences are responsible for preferential leukocyte rolling in rat mesenteric venules.* Circ Res, 1991. **69**(4): p. 1034-41.
63. Feng, D., et al., *Neutrophils emigrate from venules by a transendothelial cell pathway in response to FMLP.* J Exp Med, 1998. **187**(6): p. 903-15.
64. Burns, A.R., C.W. Smith, and D.C. Walker, *Unique structural features that influence neutrophil emigration into the lung.* Physiol Rev, 2003. **83**(2): p. 309-36.
65. Wong, J., et al., *A minimal role for selectins in the recruitment of leukocytes into the inflamed liver microvasculature.* J Clin Invest, 1997. **99**(11): p. 2782-90.
66. Palta, S., R. Saroa, and A. Palta, *Overview of the coagulation system.* Indian J Anaesth, 2014. **58**(5): p. 515-23.
67. Mackman, N., R.E. Tilley, and N.S. Key, *Role of the extrinsic pathway of blood coagulation in hemostasis and thrombosis.* Arterioscler Thromb Vasc Biol, 2007. **27**(8): p. 1687-93.
68. Grover, S.P. and N. Mackman, *Tissue Factor: An Essential Mediator of Hemostasis and Trigger of Thrombosis.* Arterioscler Thromb Vasc Biol, 2018. **38**(4): p. 709-725.
69. Camire, R.M., *Blood coagulation factor X: molecular biology, inherited disease, and engineered therapeutics.* J Thromb Thrombolysis, 2021. **52**(2): p. 383-390.
70. Steen, M., *Factor Va-factor Xa interactions: molecular sites involved in enzyme:cofactor assembly.* Scand J Clin Lab Invest Suppl, 2002. **237**: p. 5-11.
71. Maas, C., C. Oschatz, and T. Renné, *The plasma contact system 2.0.* Semin Thromb Hemost, 2011. **37**(4): p. 375-81.
72. Naudin, C., et al., *Factor XII Contact Activation.* Semin Thromb Hemost, 2017. **43**(8): p. 814-826.
73. Weidmann, H., et al., *The plasma contact system, a protease cascade at the nexus of inflammation, coagulation and immunity.* Biochim Biophys Acta Mol Cell Res, 2017. **1864**(11 Pt B): p. 2118-2127.

74. Mohammed, B.M., et al., *An update on factor XI structure and function.* Thromb Res, 2018. **161**: p. 94-105.
75. Mazurkiewicz-Pisarek, A., et al., *The factor VIII protein and its function.* Acta Biochim Pol, 2016. **63**(1): p. 11-16.
76. Majumder, R., *Phosphatidylserine Regulation of Coagulation Proteins Factor IXa and Factor VIIIa.* The Journal of Membrane Biology, 2022. **255**(6): p. 733-737.
77. Crawley, J.T., et al., *The central role of thrombin in hemostasis.* J Thromb Haemost, 2007. **5 Suppl 1**: p. 95-101.
78. Sang, Y., et al., *Interplay between platelets and coagulation.* Blood Rev, 2021. **46**: p. 100733.
79. Lenting, P.J., O.D. Christophe, and C.V. Denis, *von Willebrand factor biosynthesis, secretion, and clearance: connecting the far ends.* Blood, 2015. **125**(13): p. 2019-28.
80. Kiouptsi, K. and C. Reinhardt, *Physiological Roles of the von Willebrand Factor-Factor VIII Interaction.* Subcell Biochem, 2020. **94**: p. 437-464.
81. Do, H., et al., *Expression of factor VIII by murine liver sinusoidal endothelial cells.* J Biol Chem, 1999. **274**(28): p. 19587-92.
82. Shahani, T., et al., *Human liver sinusoidal endothelial cells but not hepatocytes contain factor VIII.* J Thromb Haemost, 2014. **12**(1): p. 36-42.
83. Pan, J., et al., *Patterns of expression of factor VIII and von Willebrand factor by endothelial cell subsets in vivo.* Blood, 2016. **128**(1): p. 104-9.
84. Hollestelle, M.J., et al., *Tissue distribution of factor VIII gene expression in vivo--a closer look.* Thromb Haemost, 2001. **86**(3): p. 855-61.
85. Jacquemin, M., et al., *FVIII production by human lung microvascular endothelial cells.* Blood, 2006. **108**(2): p. 515-7.
86. Keragala, C.B. and R.L. Medcalf, *Plasminogen: an enigmatic zymogen.* Blood, 2021. **137**(21): p. 2881-2889.
87. Chapin, J.C. and K.A. Hajjar, *Fibrinolysis and the control of blood coagulation.* Blood Rev, 2015. **29**(1): p. 17-24.
88. Kooistra, T., et al., *Regulation of endothelial cell t-PA synthesis and release.* Int J Hematol, 1994. **59**(4): p. 233-55.
89. Huber, D., et al., *Tissue-type plasminogen activator (t-PA) is stored in Weibel-Palade bodies in human endothelial cells both in vitro and in vivo.* Blood, 2002. **99**(10): p. 3637-45.
90. Puy, C., et al., *Endothelial PAI-1 (Plasminogen Activator Inhibitor-1) Blocks the Intrinsic Pathway of Coagulation, Inducing the Clearance and Degradation of FXIa (Activated Factor XI).* Arterioscler Thromb Vasc Biol, 2019. **39**(7): p. 1390-1401.
91. Schleef, R.R. and D.J. Loskutoff, *Fibrinolytic system of vascular endothelial cells. Role of plasminogen activator inhibitors.* Haemostasis, 1988. **18**(4-6): p. 328-41.
92. Marchand, A., et al., *miR-421 and miR-30c inhibit SERPINE 1 gene expression in human endothelial cells.* PLoS One, 2012. **7**(8): p. e44532.
93. Dahlbäck, B. and B.O. Villoutreix, *Regulation of blood coagulation by the protein C anticoagulant pathway: novel insights into structure-function relationships and molecular recognition.* Arterioscler Thromb Vasc Biol, 2005. **25**(7): p. 1311-20.
94. Shaydakov, M.E., et al., *Targeted gene expression analysis of human deep veins.* J Vasc Surg Venous Lymphat Disord, 2021. **9**(3): p. 770-780.e7.
95. Medina, P., et al., *Functional analysis of two haplotypes of the human endothelial protein C receptor gene.* Arterioscler Thromb Vasc Biol, 2014. **34**(3): p. 684-90.
96. Fukudome, K. and C.T. Esmon, *Identification, cloning, and regulation of a novel endothelial cell protein C/activated protein C receptor.* J Biol Chem, 1994. **269**(42): p. 26486-91.
97. Laszik, Z., et al., *Human protein C receptor is present primarily on endothelium of large blood vessels: implications for the control of the protein C pathway.* Circulation, 1997. **96**(10): p. 3633-40.
98. Giri, H., et al., *Thrombomodulin is essential for maintaining quiescence in vascular endothelial cells.* Proc Natl Acad Sci U S A, 2021. **118**(11).

99. Araujo, T.L.S., et al., *Cell-surface HSP70 associates with thrombomodulin in endothelial cells.* Cell Stress Chaperones, 2019. **24**(1): p. 273-282.
100. Faioni, E.M., et al., *Expression of endothelial protein C receptor and thrombomodulin in the intestinal tissue of patients with inflammatory bowel disease.* Crit Care Med, 2004. **32**(5 Suppl): p. S266-70.
101. Mast, A.E., *Tissue Factor Pathway Inhibitor: Multiple Anticoagulant Activities for a Single Protein.* Arterioscler Thromb Vasc Biol, 2016. **36**(1): p. 9-14.
102. Bajaj, M.S., et al., *Cultured normal human hepatocytes do not synthesize lipoprotein-associated coagulation inhibitor: evidence that endothelium is the principal site of its synthesis.* Proc Natl Acad Sci U S A, 1990. **87**(22): p. 8869-73.
103. Maroney, S.A. and A.E. Mast, *Expression of tissue factor pathway inhibitor by endothelial cells and platelets.* Transfus Apher Sci, 2008. **38**(1): p. 9-14.
104. Grabowski, E.F., et al., *Shear stress decreases endothelial cell tissue factor activity by augmenting secretion of tissue factor pathway inhibitor.* Arterioscler Thromb Vasc Biol, 2001. **21**(1): p. 157-62.
105. Bajaj, M.S., et al., *Transcriptional expression of tissue factor pathway inhibitor, thrombomodulin and von Willebrand factor in normal human tissues.* Thromb Haemost, 1999. **82**(3): p. 1047-52.
106. Rezaie, A.R. and H. Giri, *Anticoagulant and signaling functions of antithrombin.* J Thromb Haemost, 2020. **18**(12): p. 3142-3153.
107. Neubauer, K. and B. Zieger, *Endothelial cells and coagulation.* Cell Tissue Res, 2022. **387**(3): p. 391-398.
108. Wolberg, A.S., et al., *Procoagulant activity in hemostasis and thrombosis: Virchow's triad revisited.* Anesth Analg, 2012. **114**(2): p. 275-85.
109. Kwaan, H.C. and M.M. Samama, *The significance of endothelial heterogeneity in thrombosis and hemostasis.* Semin Thromb Hemost, 2010. **36**(3): p. 286-300.
110. Bochenek, M.L. and K. Schäfer, *Role of Endothelial Cells in Acute and Chronic Thrombosis.* Hamostaseologie, 2019. **39**(2): p. 128-139.
111. Pilard, M., et al., *Endothelial Cell Phenotype, a Major Determinant of Venous Thrombo-Inflammation.* Front Cardiovasc Med, 2022. **9**: p. 864735.
112. Eisenreich, A., et al., *Cdc2-like kinases and DNA topoisomerase I regulate alternative splicing of tissue factor in human endothelial cells.* Circ Res, 2009. **104**(5): p. 589-99.
113. Krikun, G., et al., *Lipopolysaccharide appears to activate human endometrial endothelial cells through TLR-4-dependent and TLR-4-independent mechanisms.* Am J Reprod Immunol, 2012. **68**(3): p. 233-7.
114. Folco, E.J., et al., *Neutrophil Extracellular Traps Induce Endothelial Cell Activation and Tissue Factor Production Through Interleukin-1α and Cathepsin G.* Arterioscler Thromb Vasc Biol, 2018. **38**(8): p. 1901-1912.
115. Yang, X., et al., *Extracellular histones induce tissue factor expression in vascular endothelial cells via TLR and activation of NF-κB and AP-1.* Thromb Res, 2016. **137**: p. 211-218.
116. Liu, Y., et al., *Purinergic P2Y2 Receptor Control of Tissue Factor Transcription in Human Coronary Artery Endothelial Cells: NEW AP-1 TRANSCRIPTION FACTOR SITE AND NEGATIVE REGULATOR.* J Biol Chem, 2016. **291**(4): p. 1553-1563.
117. Moriguchi, T. and B.E. Sumpio, *PECAM-1 phosphorylation and tissue factor expression in HUVECs exposed to uniform and disturbed pulsatile flow and chemical stimuli.* J Vasc Surg, 2015. **61**(2): p. 481-8.
118. Abe, R., et al., *Varying effects of hemodynamic forces on tissue factor RNA expression in human endothelial cells.* J Surg Res, 2011. **170**(1): p. 150-6.
119. Chen, Y., et al., *CRP regulates the expression and activity of tissue factor as well as tissue factor pathway inhibitor via NF-kappaB and ERK 1/2 MAPK pathway.* FEBS Lett, 2009. **583**(17): p. 2811-8.
120. Kothari, H., U.R. Pendurthi, and L.V. Rao, *Analysis of tissue factor expression in various cell model systems: cryptic vs. active.* J Thromb Haemost, 2013. **11**(7): p. 1353-63.
121. Antoniak, S. and N. Mackman, *Editorial Commentary: Tissue factor expression by the endothelium: Coagulation or inflammation?* Trends Cardiovasc Med, 2016. **26**(4): p. 304-5.

122. Witkowski, M. and U. Rauch, *Letter to the Editor: Tissue factor of endothelial origin: Just another brick in the wall?* Trends Cardiovasc Med, 2017. **27**(2): p. 155-156.
123. Witkowski, M., U. Landmesser, and U. Rauch, *Tissue factor as a link between inflammation and coagulation.* Trends Cardiovasc Med, 2016. **26**(4): p. 297-303.
124. Bavendiek, U., et al., *Induction of Tissue Factor Expression in Human Endothelial Cells by CD40 Ligand Is Mediated via Activator Protein 1, Nuclear Factor κB, and Egr-1 *.* Journal of Biological Chemistry, 2002. **277**(28): p. 25032-25039.
125. Borgel, D., et al., *Inflammation in deep vein thrombosis: a therapeutic target?* Hematology, 2019. **24**(1): p. 742-750.
126. Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.
127. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
128. Mi, H., et al., *Large-scale gene function analysis with the PANTHER classification system.* Nat Protoc, 2013. **8**(8): p. 1551-66.
129. Mi, H., et al., *PANTHER version 10: expanded protein families and functions, and analysis tools.* Nucleic Acids Res, 2016. **44**(D1): p. D336-42.
130. Ponten, F., K. Jirstrom, and M. Uhlen, *The Human Protein Atlas - a tool for pathology.* Journal of Pathology, 2008. **216**(4): p. 387-393.
131. Uhlen, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.
132. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis.* BMC Bioinformatics, 2008. **9**: p. 559.
133. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587.e29.
134. Sun, W., et al., *snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis.* Nature, 2020. **587**(7832): p. 98-102.
135. Guo, J., et al., *The adult human testis transcriptional cell atlas.* Cell Res, 2018. **28**(12): p. 1141-1157.
136. Rouillard, A.D., et al., *The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins.* Database (Oxford), 2016. **2016**.
137. Hamann, J., et al., *International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G protein-coupled receptors.* Pharmacol Rev, 2015. **67**(2): p. 338-67.
138. Favara, D.M., et al., *Elevated expression of the adhesion GPCR ADGRL4/ELTD1 promotes endothelial sprouting angiogenesis without activating canonical GPCR signalling.* Scientific Reports, 2021. **11**(1): p. 8870.
139. Favara, D.M., et al., *ADGRL4/ELTD1 Silencing in Endothelial Cells Induces ACLY and SLC25A1 and Alters the Cellular Metabolic Profile.* Metabolites, 2019. **9**(12).
140. Masiero, M., et al., *A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis.* Cancer Cell, 2013. **24**(2): p. 229-41.
141. Butler, Lynn M., et al., *Analysis of Body-wide Unfractionated Tissue Data to Identify a Core Human Endothelial Transcriptome.* Cell Systems, 2016. **3**(3): p. 287-301.e3.
142. 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.
143. Norreen-Thorsen, M., et al., *A human adipose tissue cell-type transcriptome atlas.* Cell Rep, 2022. **40**(2): p. 111046.
144. Dusart, P., et al., *A tissue centric atlas of cell type transcriptome enrichment signatures.* bioRxiv, 2023: p. 2023.01.10.520698.
145. Gao, Y. and J.R. Yates Iii, *Protein Analysis by Shotgun Proteomics*, in *Mass Spectrometry-Based Chemical Proteomics*. 2019. p. 1-38.
146. Zhang, L. and J.E. Elias, *Relative Protein Quantification Using Tandem Mass Tag Mass Spectrometry.* Methods Mol Biol, 2017. **1550**: p. 185-198.

147. Pergande, M.R. and S.M. Cologna, *Isoelectric Point Separations of Peptides and Proteins.* Proteomes, 2017. **5**(1).
148. *UniProt: the Universal Protein Knowledgebase in 2023.* Nucleic Acids Res, 2023. **51**(D1): p. D523-d531.
149. Kustatscher, G., et al., *Understudied proteins: opportunities and challenges for functional proteomics.* Nature Methods, 2022. **19**(7): p. 774-779.
150. Sinha, S., et al., *Darkness in the Human Gene and Protein Function Space: Widely Modest or Absent Illumination by the Life Science Literature and the Trend for Fewer Protein Function Discoveries Since 2000.* PROTEOMICS, 2018. **18**(21-22): p. 1800093.
151. Stoeger, T., et al., *Large-scale investigation of the reasons why potentially important genes are ignored.* PLoS Biol, 2018. **16**(9): p. e2006643.
152. Kustatscher, G., et al., *An open invitation to the Understudied Proteins Initiative.* Nature Biotechnology, 2022. **40**(6): p. 815-817.
153. Haque, A., et al., *A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications.* Genome Med, 2017. **9**(1): p. 75.
154. Hicks, S.C., et al., *Missing data and technical variability in single-cell RNA-sequencing experiments.* Biostatistics, 2018. **19**(4): p. 562-578.
155. Zheng, Y., et al., *SCC: an accurate imputation method for scRNA-seq dropouts based on a mixture model.* BMC Bioinformatics, 2021. **22**(1): p. 5.
156. Jiang, R., et al., *Statistics or biology: the zero-inflation controversy about scRNA-seq data.* Genome Biology, 2022. **23**(1): p. 31.
157. Glastonbury, C.A., et al., *Cell-Type Heterogeneity in Adipose Tissue Is Associated with Complex Traits and Reveals Disease-Relevant Cell-Specific eQTLs.* The American Journal of Human Genetics, 2019. **104**(6): p. 1013-1024.
158. Chu, T., et al., *Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology.* Nature Cancer, 2022. **3**(4): p. 505-517.
159. Medina-Leyte, D.J., et al., *Use of Human Umbilical Vein Endothelial Cells (HUVEC) as a Model to Study Cardiovascular Disease: A Review.* Applied Sciences, 2020. **10**(3): p. 938.
160. Chang, M.W., et al., *Comparison of early passage, senescent and hTERT immortalized endothelial cells.* Exp Cell Res, 2005. **309**(1): p. 121-36.
161. Viemann, D., et al., *TNF induces distinct gene expression programs in microvascular and macrovascular human endothelial cells.* J Leukoc Biol, 2006. **80**(1): p. 174-85.
162. Brenes, A., et al., *Multibatch TMT Reveals False Positives, Batch Effects and Missing Values.* Mol Cell Proteomics, 2019. **18**(10): p. 1967-1980.
163. Ohori, M., et al., *Gene regulatory network analysis defines transcriptome landscape with alternative splicing of human umbilical vein endothelial cells during replicative senescence.* BMC Genomics, 2021. **22**(1): p. 869.
164. Feng, H., et al., *Interferon regulatory factor 1 (IRF1) and anti-pathogen innate immune responses.* PLoS Pathog, 2021. **17**(1): p. e1009220.
165. Li, Y., et al., *Ataxin-10 Inhibits TNF-α-Induced Endothelial Inflammation via Suppressing Interferon Regulatory Factor-1.* Mediators Inflamm, 2021. **2021**: p. 7042148.
166. Lechleitner, S., et al., *Interferon enhances tumor necrosis factor-induced vascular cell adhesion molecule 1 (CD106) expression in human endothelial cells by an interferon-related factor 1-dependent pathway.* J Exp Med, 1998. **187**(12): p. 2023-30.
167. Kautz, B., et al., *SHP1 protein-tyrosine phosphatase inhibits gp91PHOX and p67PHOX expression by inhibiting interaction of PU.1, IRF1, interferon consensus sequence-binding protein, and CREB-binding protein with homologous Cis elements in the CYBB and NCF2 genes.* J Biol Chem, 2001. **276**(41): p. 37868-78.
168. Park, J., et al., *Elevated level of SUMOylated IRF-1 in tumor cells interferes with IRF-1-mediated apoptosis.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 17028-33.

169. Garvin, A.J., et al., *GSK3β-SCFFBXW7α mediated phosphorylation and ubiquitination of IRF1 are required for its transcription-dependent turnover.* Nucleic Acids Res, 2019. **47**(9): p. 4476-4494.
170. Sgarbanti, M., et al., *IκB kinase ε targets interferon regulatory factor 1 in activated T lymphocytes.* Mol Cell Biol, 2014. **34**(6): p. 1054-65.
171. Lin, R. and J. Hiscott, *A role for casein kinase II phosphorylation in the regulation of IRF-1 transcriptional activity.* Mol Cell Biochem, 1999. **191**(1-2): p. 169-80.
172. Guo, Y.J., et al., *ERK/MAPK signalling pathway and tumorigenesis.* Exp Ther Med, 2020. **19**(3): p. 1997-2007.
173. Smits, P.J., et al., *Endothelial MAP2K1 mutations in arteriovenous malformation activate the RAS/MAPK pathway.* Biochem Biophys Res Commun, 2020. **529**(2): p. 450-454.
174. Lu, N. and C.J. Malemud, *Extracellular Signal-Regulated Kinase: A Regulator of Cell Growth, Inflammation, Chondrocyte and Bone Cell Receptor-Mediated Gene Expression.* Int J Mol Sci, 2019. **20**(15).
175. Goebeler, M., et al., *The MKK6/p38 stress kinase cascade is critical for tumor necrosis factor-alpha-induced expression of monocyte-chemoattractant protein-1 in endothelial cells.* Blood, 1999. **93**(3): p. 857-65.
176. Goedert, M., et al., *Activation of the novel stress-activated protein kinase SAPK4 by cytokines and cellular stresses is mediated by SKK3 (MKK6); comparison of its substrate specificity with that of other SAP kinases.* Embo j, 1997. **16**(12): p. 3563-71.
177. Raingeaud, J., et al., *MKK3- and MKK6-regulated gene expression is mediated by the p38 mitogen-activated protein kinase signal transduction pathway.* Mol Cell Biol, 1996. **16**(3): p. 1247-55.
178. Kang, J.I., et al., *PKR protein kinase is activated by hepatitis C virus and inhibits viral replication through translational control.* Virus Res, 2009. **142**(1-2): p. 51-6.
179. Samuel, C.E., *The eIF-2 alpha protein kinases, regulators of translation in eukaryotes from yeasts to humans.* J Biol Chem, 1993. **268**(11): p. 7603-6.
180. Goh, K.C., M.J. deVeer, and B.R. Williams, *The protein kinase PKR is required for p38 MAPK activation and the innate immune response to bacterial endotoxin.* Embo j, 2000. **19**(16): p. 4292-7.
181. Silva, A.M., et al., *Protein kinase R (PKR) interacts with and activates mitogen-activated protein kinase kinase 6 (MKK6) in response to double-stranded RNA stimulation.* J Biol Chem, 2004. **279**(36): p. 37670-6.
182. Taghavi, N. and C.E. Samuel, *Protein kinase PKR catalytic activity is required for the PKR-dependent activation of mitogen-activated protein kinases and amplification of interferon beta induction following virus infection.* Virology, 2012. **427**(2): p. 208-16.
183. Pfaller, C.K., et al., *Protein kinase PKR and RNA adenosine deaminase ADAR1: new roles for old players as modulators of the interferon response.* Curr Opin Immunol, 2011. **23**(5): p. 573-82.
184. McAllister, C.S., N. Taghavi, and C.E. Samuel, *Protein kinase PKR amplification of interferon β induction occurs through initiation factor eIF-2α-mediated translational control.* J Biol Chem, 2012. **287**(43): p. 36384-92.
185. Takada, Y., et al., *Genetic deletion of PKR abrogates TNF-induced activation of IkappaBalpha kinase, JNK, Akt and cell proliferation but potentiates p44/p42 MAPK and p38 MAPK activation.* Oncogene, 2007. **26**(8): p. 1201-12.
186. Bonnet, M.C., et al., *PKR stimulates NF-kappaB irrespective of its kinase function by interacting with the IkappaB kinase complex.* Mol Cell Biol, 2000. **20**(13): p. 4532-42.

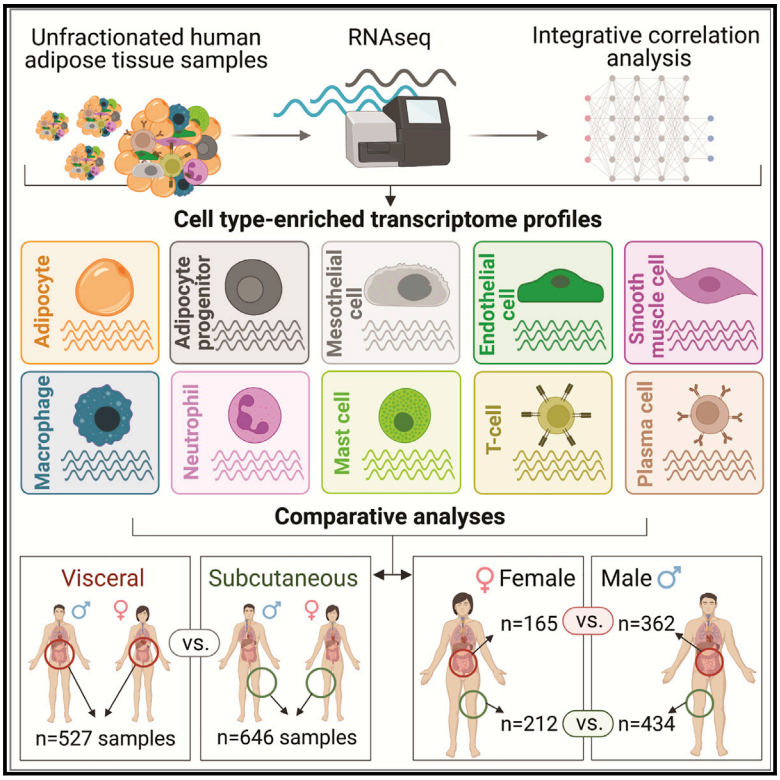
Paper I

Cell Reports

Resource

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

Norreen-Thorsen et al., 2022, Cell Reports *40*, 111046
July 12, 2022 © 2022 The Authors.
https://doi.org/10.1016/j.celrep.2022.111046

CellPress

Cell Reports

CellPress
OPEN ACCESS

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

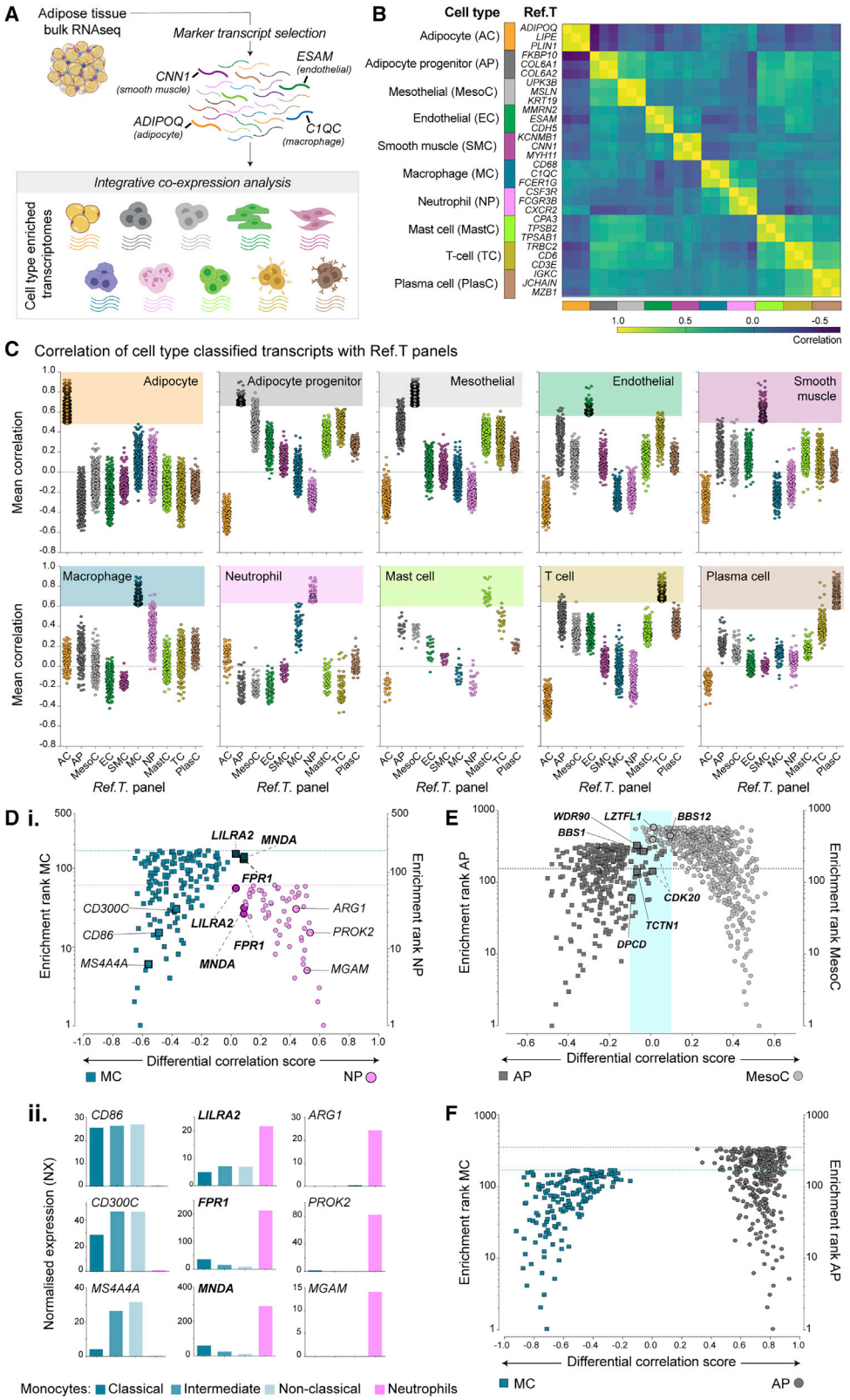
The importance of defining cell-type-specific genes is well acknowledged. Technological advances facilitate high-resolution sequencing of single cells, but practical challenges remain. Adipose tissue is composed primarily of adipocytes, large buoyant cells requiring extensive, artefact-generating processing 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

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).



(legend on next page)

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 $p < 4.0 \times 10^{-99}$); (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.

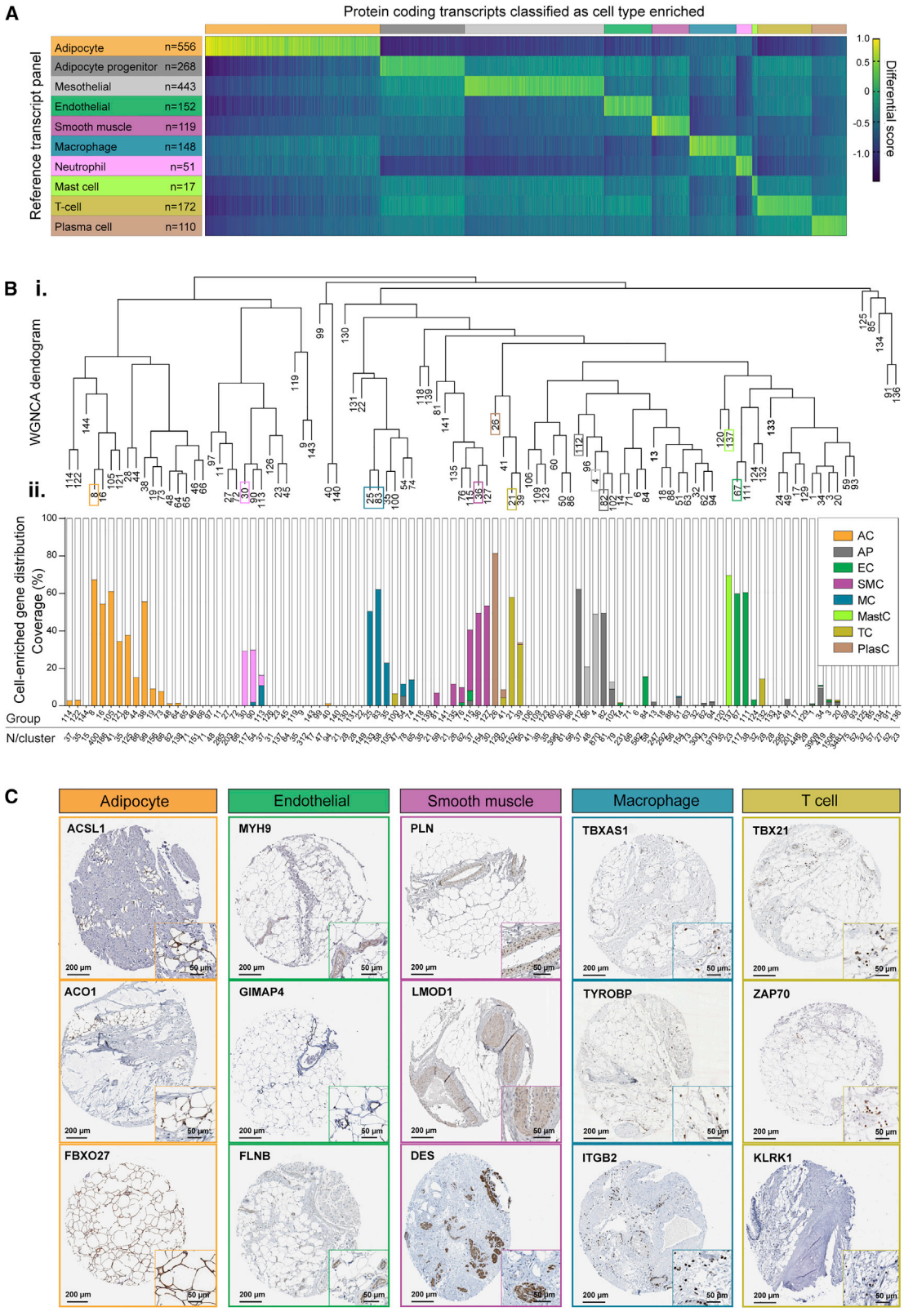
(B) Heatmap of Spearman correlation (corr.) values between *Ref.T.* panels selected for VAT cell types: adipocyte (AC), adipocyte progenitor (AP), mesothelial (MesoC), endothelial (EC), smooth muscle (SMC), macrophage (MC), neutrophil (NP), mast cell (MastC), T cell (TC), and plasma cell (PlasC).

(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)

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) $\geq$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

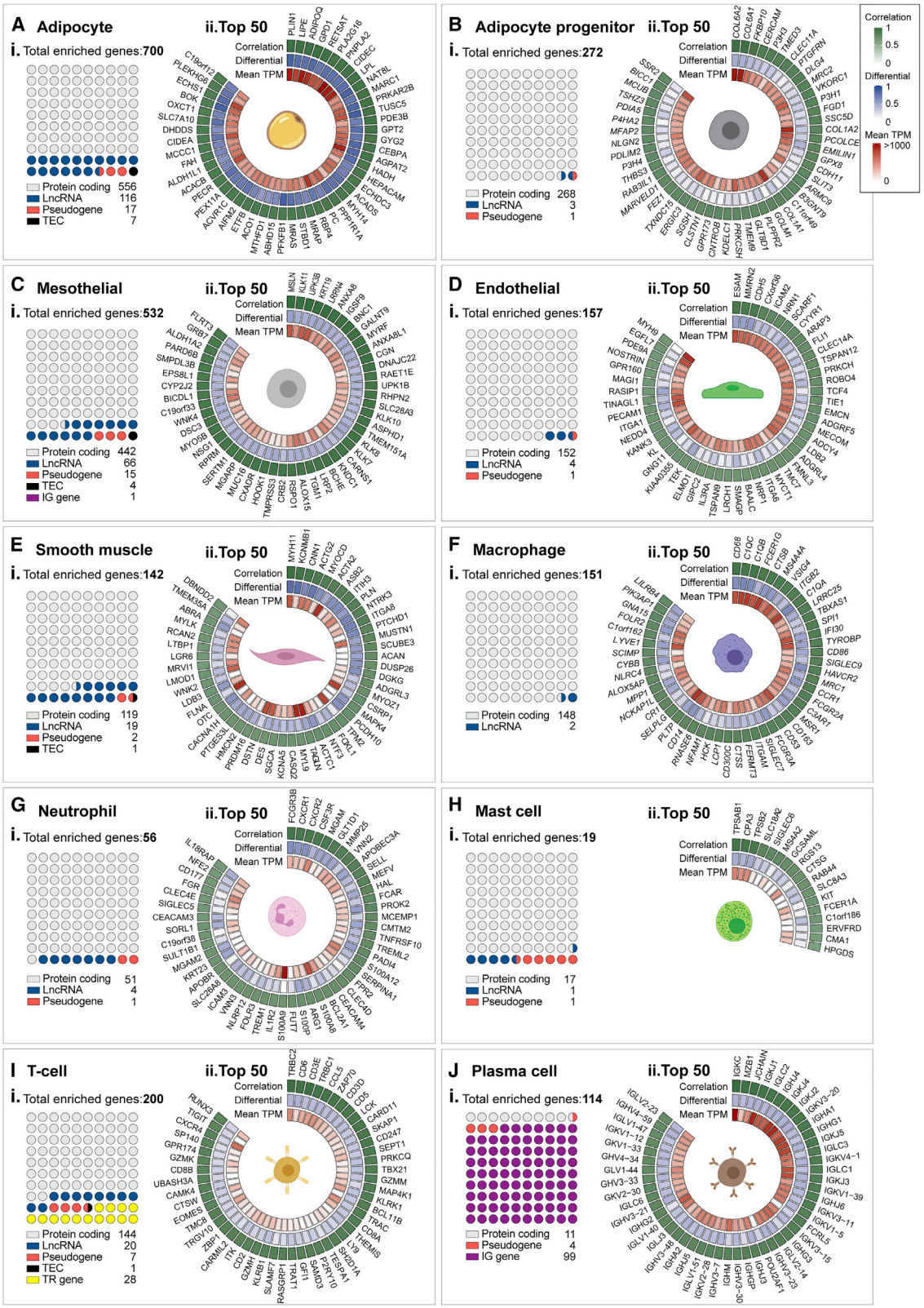
Figure 2. Integrative co-expression analysis of unfractionated RNA-seq reveals enriched protein-coding transcriptomes of human visceral adipose tissue (VAT) cell types

(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

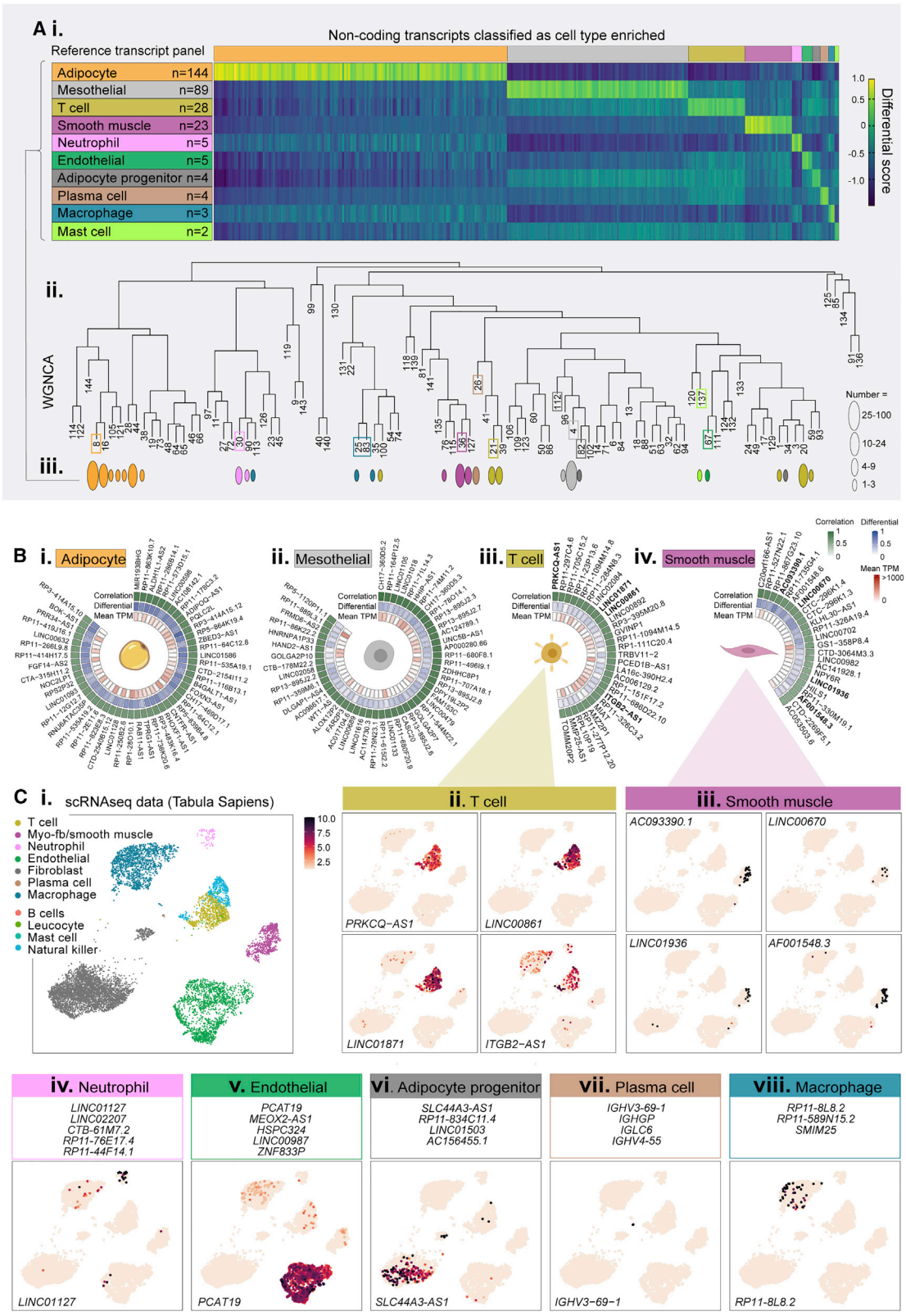
We classified 307 non-coding transcripts as cell-type enriched in VAT, the highest number of which were in adipocytes (n = 144), followed by mesothelial cells (n = 89) (Figure 4A; Table S2, tab 1). Cell-enriched non-coding transcripts were typically expressed at lower levels than cell-enriched protein-coding 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 Hartigh, 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)

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 progenitor (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.,

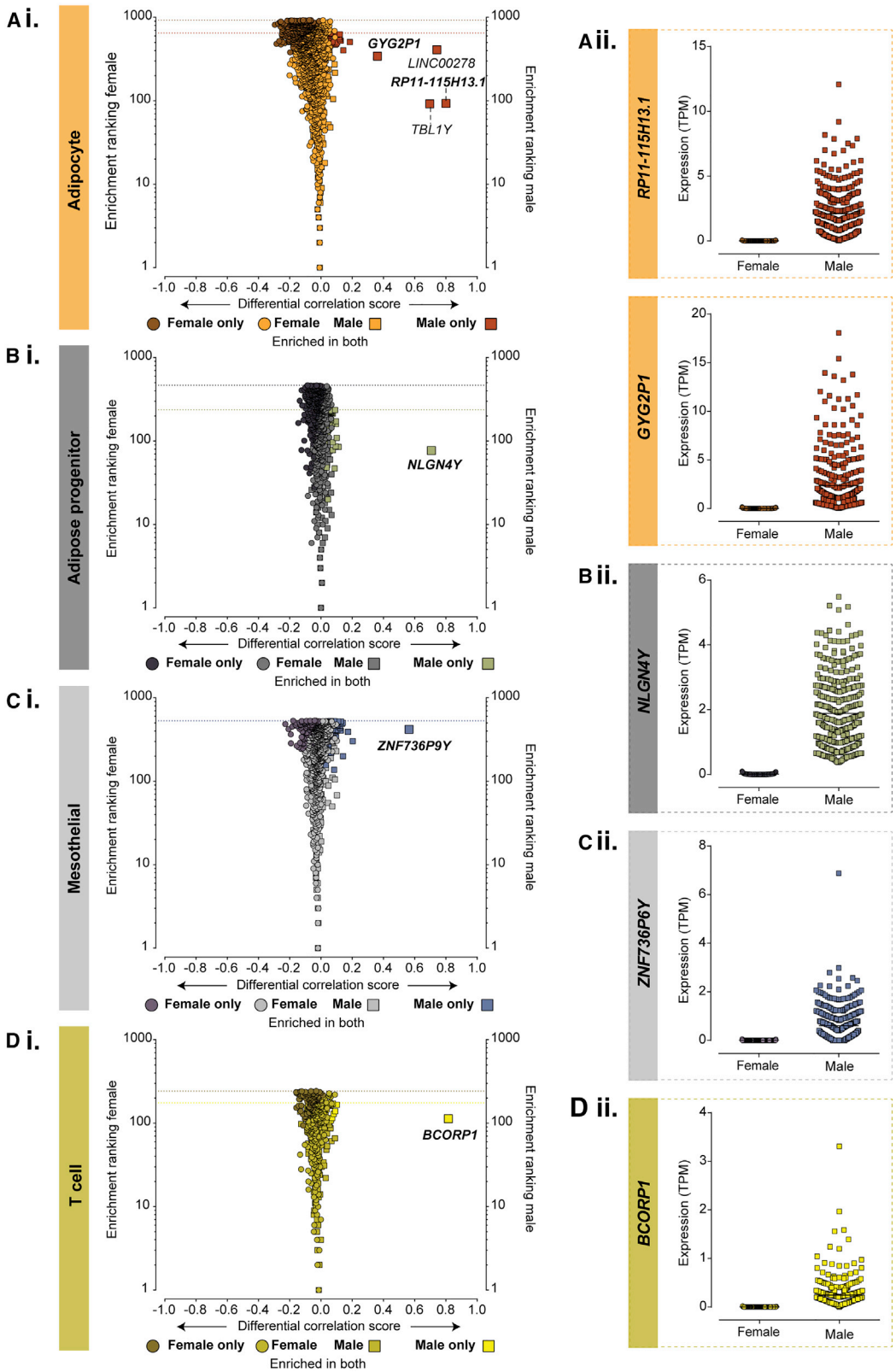
Figure 4. Integrative co-expression analysis of unfractionated RNA-seq reveals enriched non-coding transcriptomes of human VAT cell types

(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)

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; $p < 1.0 \times 10^{-33}$) (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

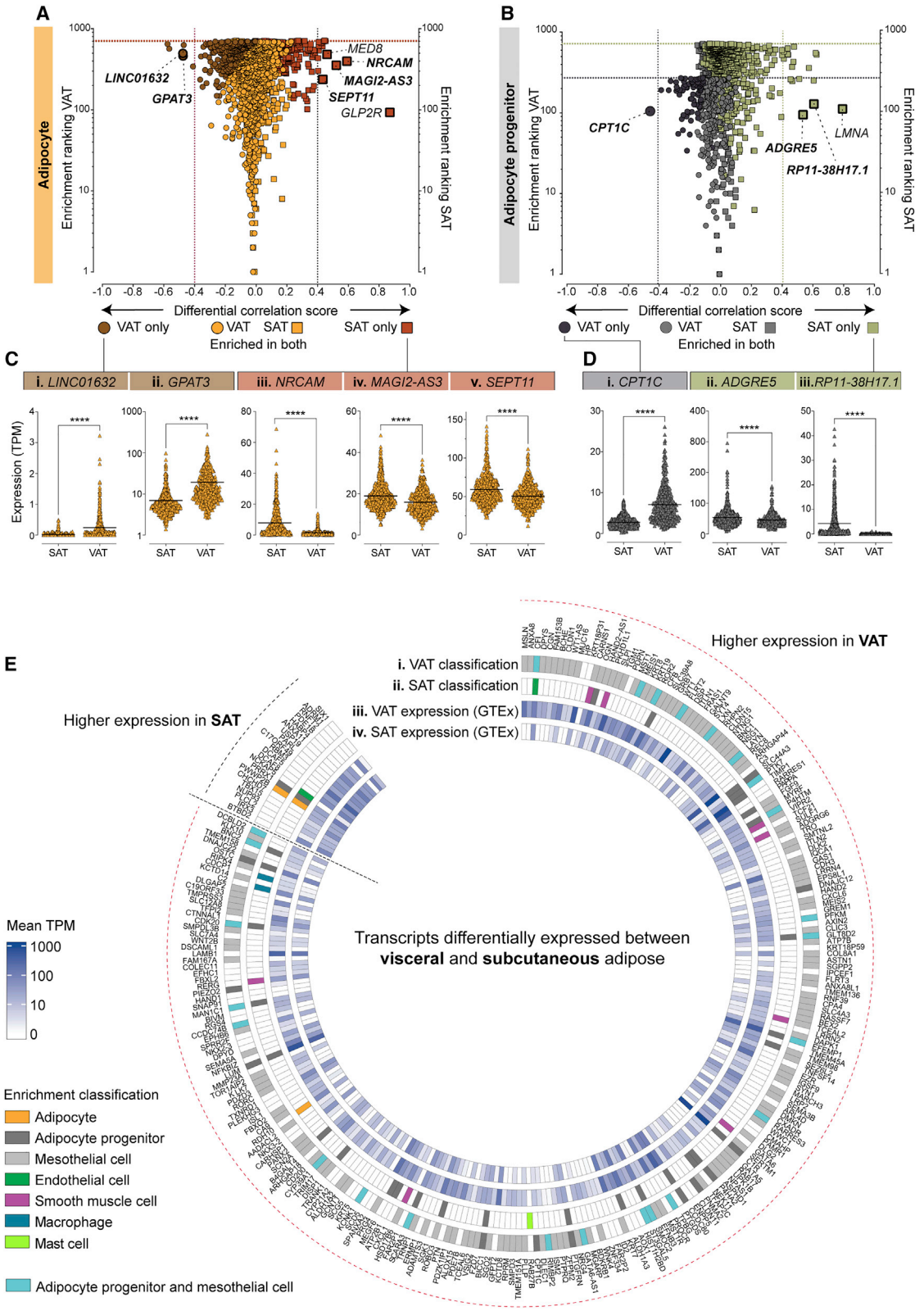
To establish whether these sex-specific differences also existed in SAT, we performed an equivalent subset analysis of the SAT RNA-seq dataset (female n = 212; male n = 434). As in the full dataset, intra-panel cell type *Ref.T*.s correlated well in female- and male-sample subsets (all >0.71; $p < 14.0 \times 10^{-31}$) (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)

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 preprocessing, 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-AS1*. 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.

ACKNOWLEDGMENTS

Funding was granted to L.M.B. from Hjärt Lungfonden (20170759, 20170537, and 20200544) and Swedish Research Council (2019–01493) and to J.O. from Stockholm County Council (SLL 2017–0842). The Human Protein Atlas is funded by The Knut and Alice Wallenberg Foundation. We used data from Genotype-Tissue Expression (GTEx) Project (gtexportal.org) (Consortium, 2015), supported by the Office of the Director of the National Institutes of Health and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

AUTHOR CONTRIBUTIONS

Conceptualization, L.M.B.; methodology, M.N.-T., E.C.S., S.Ö., and P.J.D.; formal analysis, M.N.-T.; investigation, M.N.-T., P.J.D., L.M.B., and C.L.; resources, M.U., F.P., J.O., L.M.B., and C.L.; writing – original draft, M.N.-T. and L.M.B.; writing – review & editing, all; visualization, M.N.-T., L.M.B., P.J.D., M.Z., and K.V.F.; supervision, L.M.B. and P.J.D.; funding acquisition, L.M.B. and J.O.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 30, 2021
Revised: April 29, 2022
Accepted: June 13, 2022
Published: July 12, 2022

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.

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 non-coding 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 lincNORS. 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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-ACSL1 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA011316; RRID:AB_1844536
Anti-ACO1 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA019371; RRID:AB_1844519
Anti-FBXO27 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA046800; RRID:AB_2679813
Anti-MYH9 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA064783; RRID:AB_2732721
Anti-GIMAP4 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA019135; RRID:AB_1849670
Anti-FLNB polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA004886; RRID:AB_1848600
Anti-PLN polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA026900; RRID:AB_1855314
Anti-LMOD1 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA028435; RRID:AB_10602180
Anti-DES polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA018803; RRID:AB_1847616
Anti-TBXAS1 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA031257; RRID:AB_2673812
Anti-ITGB2 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA016894; RRID:AB_1846257
Anti-NKG2D (KLRK1) monoclonal antibody produced in mouse	Merck (www.merckmillipore.com)	Cat#05-945; Clone 3.1.1.1 (HPA:CAB021896)
Anti-TYROBP polyclonal antibody produced in rabbit	Santa Cruz Biotechnology (www.scbt.com)	Cat#sc-20783; RRID:AB_638987 (HPA:CAB009493)
Anti- TBX21 polyclonal antibody produced in rabbit	Santa Cruz Biotechnology (www.scbt.com)	Cat#sc-21003; RRID:AB_2200557 (HPA:CAB009524)
Anti- ZAP70 monoclonal antibody produced in mouse	Thermo Fisher Scientific (www.thermofisher.com)	Cat#MS-1911 Clone: 2F3.2 (HPA:CAB002625)
Anti-PRKAR2B polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA008421, RRID:AB_1855421
Anti-C19orf12 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA046930, RRID:AB_10962836
Anti-CDH13 monoclonal antibody produced in rat	R&D Systems (www.rndsystems.com)	Cat#MAB3264 Clone: 392411 (HPA:CAB025863)
Anti-SHANK3 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA003446; RRID:AB_1079958
Anti-CASQ2 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA027285; RRID:AB_1845933
Anti-SLC30A3 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA060505; RRID:AB_2684296
Anti-IFI30 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA026650; RRID:AB_10602237

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Anti-LCP1 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA019493; RRID:AB_1855457
Anti-SP140 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA006162; RRID:AB_1857403
Anti-CD247 polyclonal antibody produced in rabbit	Human Protein Atlas (www.proteinatlas.org)/ Atlas Antibodies (www.atlasantibodies.com)	Cat#HPA008750; RRID:AB_1857863
Biological samples		
Adipose tissue samples	Uppsala biobank Uhelen et al. (2015) Proteomics. Tissue-based map of the human proteome . Science *347*, 1260419.	https://www.proteinatlas.org/
Deposited data		
All generated data (also contained in manuscript Tables S1, S2, and S3)	This paper	Butler, Lynn (2022), "A human adipose tissue cell type transcriptome atlas", Mendeley Data, V1, https://doi.org/10.17632/6wmbw2nt4x.1
Software and algorithms		
R	R Core Team (2022). R: A language and environment for statistical computing.	https://www.R-project.org
RStudio	RStudio Team (2021). RStudio: Integrated Development Environment for R.	http://www.rstudio.com/
Code for cell type enrichment analysis	This paper	https://github.com/PhilipDusart/cell-enrichment
WCGNA clustering analysis	Langfelder and Horvath, (2008) WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559	https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/
Seurat single cell RNAseq analysis	Hao et al., (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573-3587 e3529	https://satijalab.org/seurat/
Circlize package, for creation of circular plots	Gu et al., (2014) Circlize Implements and enhances circular visualization in R. Bioinformatics *30*, 2811-2812.	https://jokergoo.github.io/circlize_book/book/
GraphPad Prism 6	GraphPad	www.graphpad.com/
Other		
Adipose-visceral (omentum) RNAseq data Adipose-subcutaneous RNAseq data	Genotype-Tissue Expression (GTEx) Project	gtexportal.org; dbGaP Accession phs000424.v8.p2
GO Biological process and Reactome pathways analysis	Ashburner et al. (2000) Gene ontology: tool for the unification of biology. Nat Genet. May 25(1):25-9. The Gene Ontology resource: enriching a GOld mine. (2021) Nucleic Acids Res. 49(D1):D325-D334.	Gene ontology resource and PANTHER classification resource; http://geneontology.org/ GO Ontology database https://doi.org/10.5281/zenodo.4081749 Released 2020-10-09 Reactome version 65. Released 2020-11-17
scRNAseq human subcutaneous adipose tissue	Hildreth et al. (2021). Single-cell sequencing of human white adipose tissue identifies new cell states in health and obesity. Nat Immunol *22*, 639-653.	https://www.nature.com/articles/s41590-021-00922-4 (https://doi.org/10.1038/s41590-021-00922-4;) PMID: **33907320**
scRNAseq human subcutaneous adipose tissue	Sapiens, T. (2021). The Tabula Sapiens: a multiple organ single cell transcriptomic atlas of humans. bioRxiv preprint	https://www.biorxiv.org/content/10.1101/2021.07.19.452956v3 (https://doi.org/10.1101/2021.07.19.452956)

(*Continued on next page*)

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).

Cell Reports, Volume *40*

Supplemental information

A human adipose tissue cell-type transcriptome atlas

Marthe Norreen-Thorsen, Eike Christopher Struck, Sofia Öling, Martin Zwahlen, Kalle Von Feilitzen, Jacob Odeberg, Cecilia Lindskog, Fredrik Pontén, Mathias Uhlén, Philip James Dusart, and Lynn Marie Butler

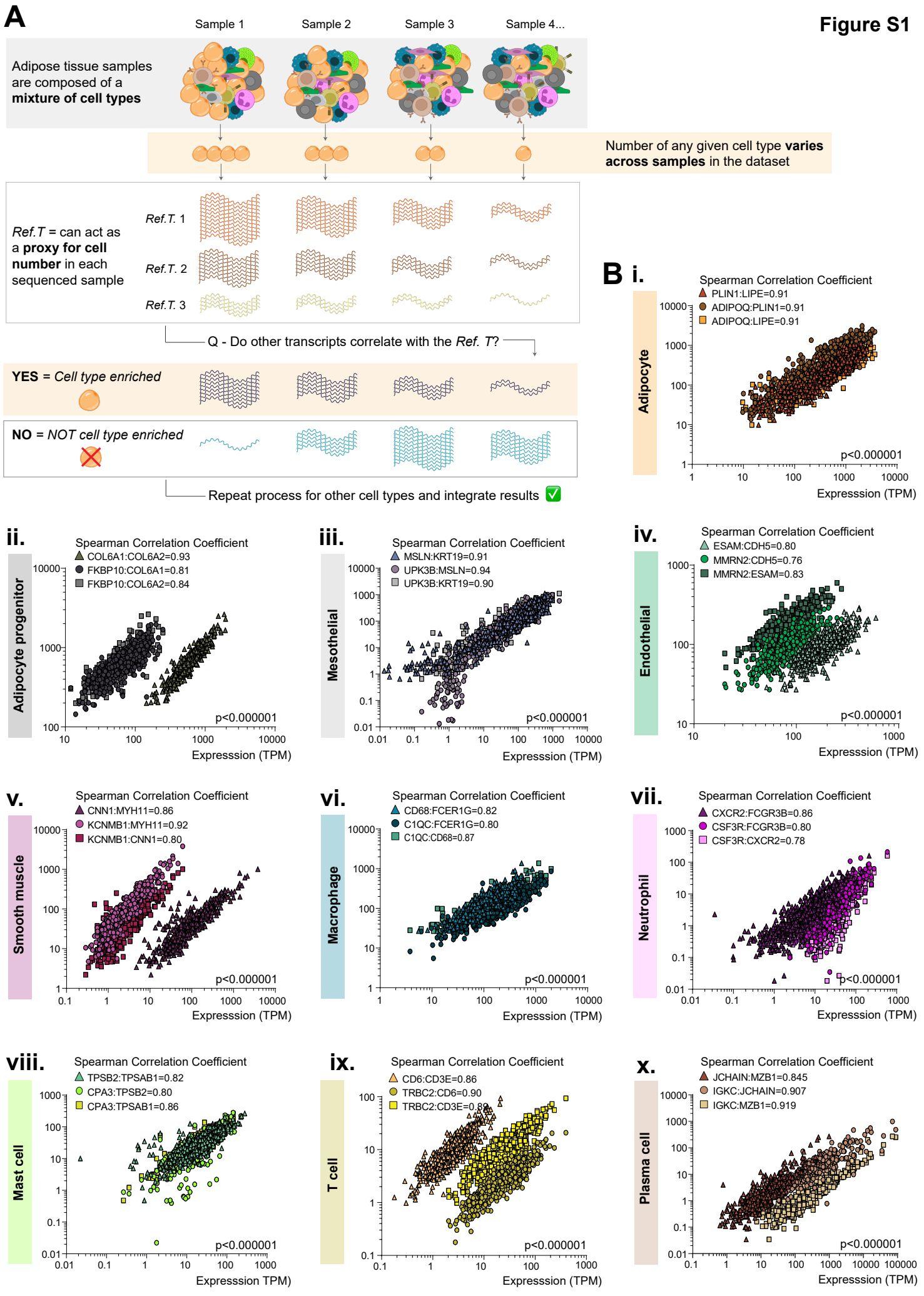
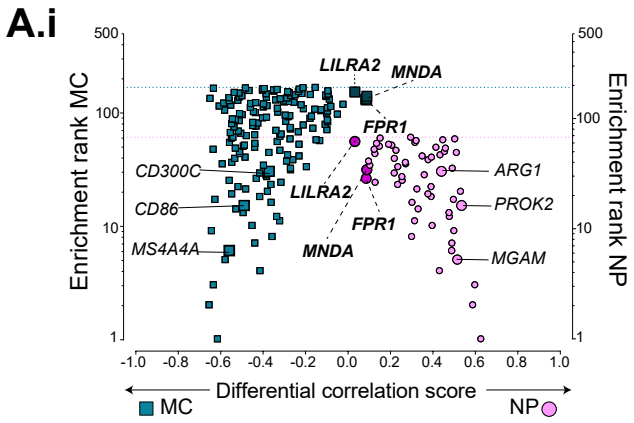
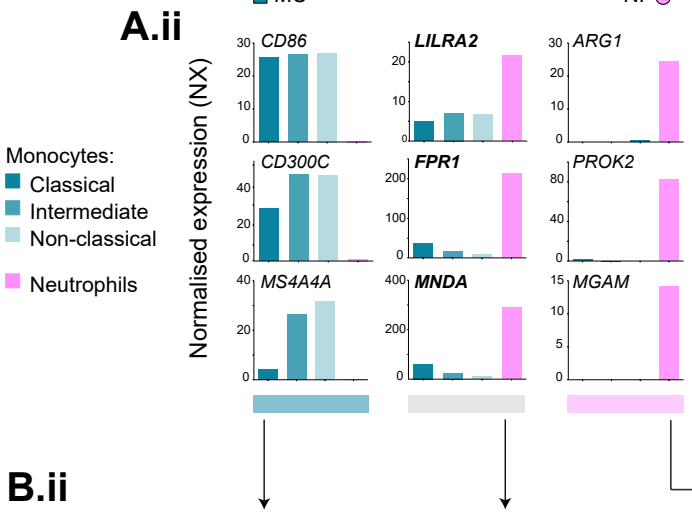
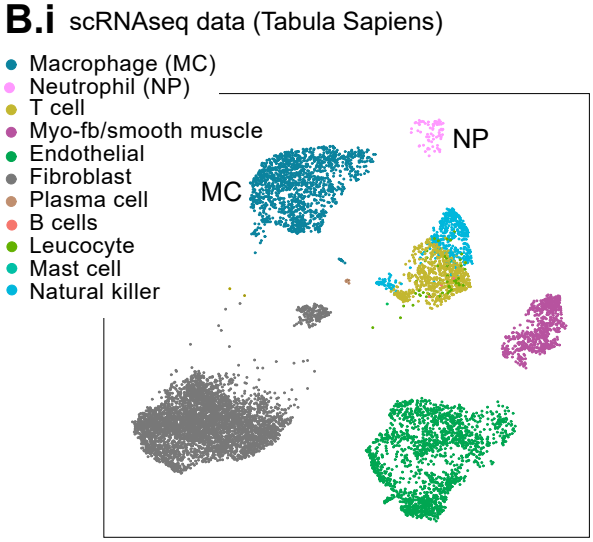

A
Figure S1
Sample 1
Sample 2
Sample 3
Sample 4...
Adipose tissue samples are composed of a **mixture of cell types**
Number of any given cell type **varies across samples** in the dataset
Ref.T = can act as a **proxy for cell number** in each sequenced sample
Ref.T. 1
Ref.T. 2
Ref.T. 3
Q - Do other 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=0.88
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
Enrichment rank MC
Enrichment rank NP
500
100
10
1
LILRA2
MNDA
FPR1
CD300C
LILRA2
ARG1
CD86
FPR1
PROK2
MS4A4A
MNDA
MGAM
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Differential correlation score
MC
NP


A.ii
Normalised expression (NX)
CD86
LILRA2
ARG1
CD300C
FPR1
PROK2
MS4A4A
MNDA
MGAM
Monocytes:
Classical
Intermediate
Non-classical
Neutrophils


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
MC
NP

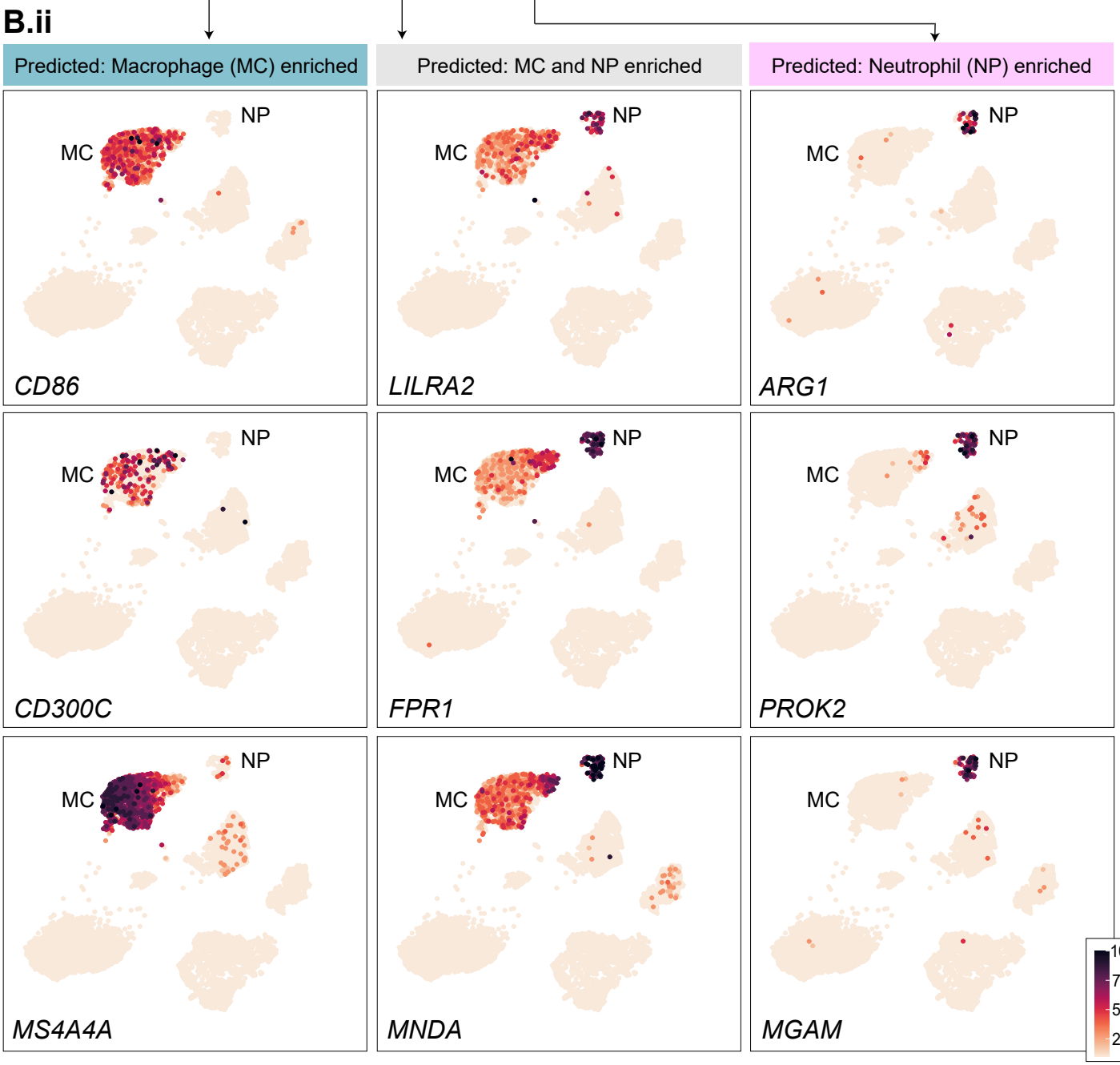
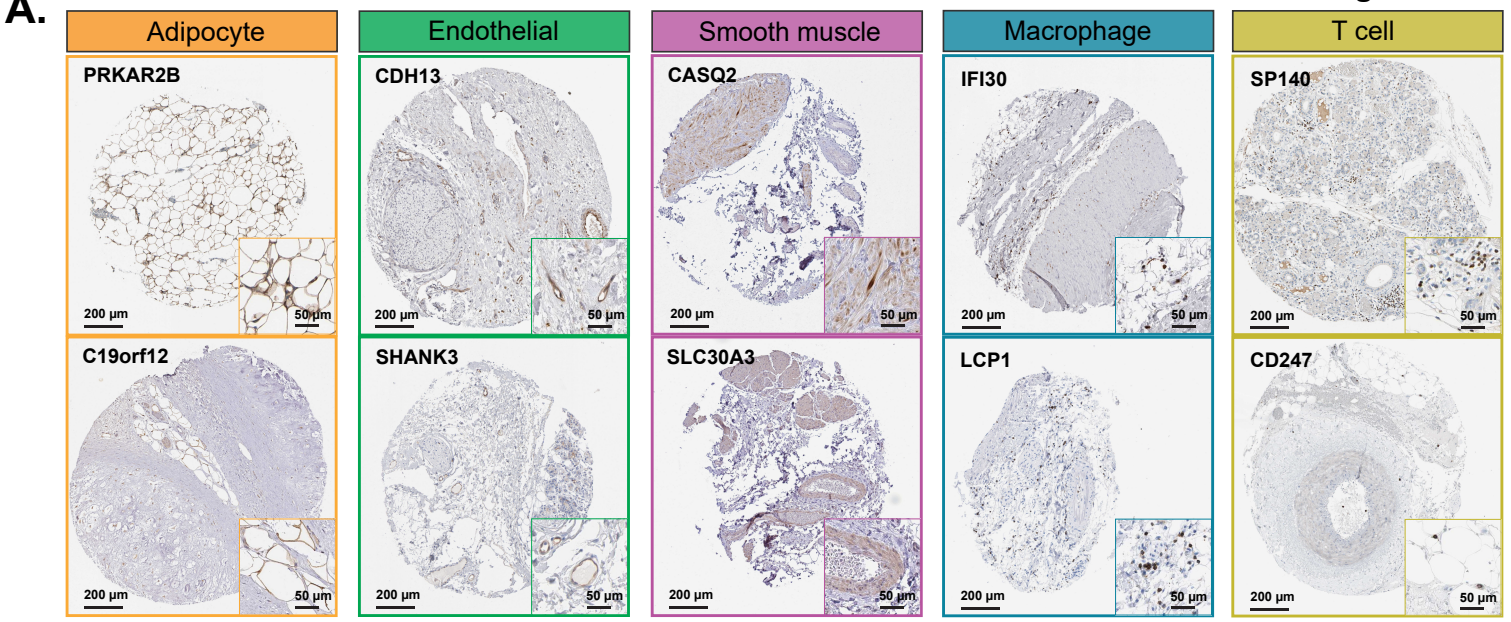
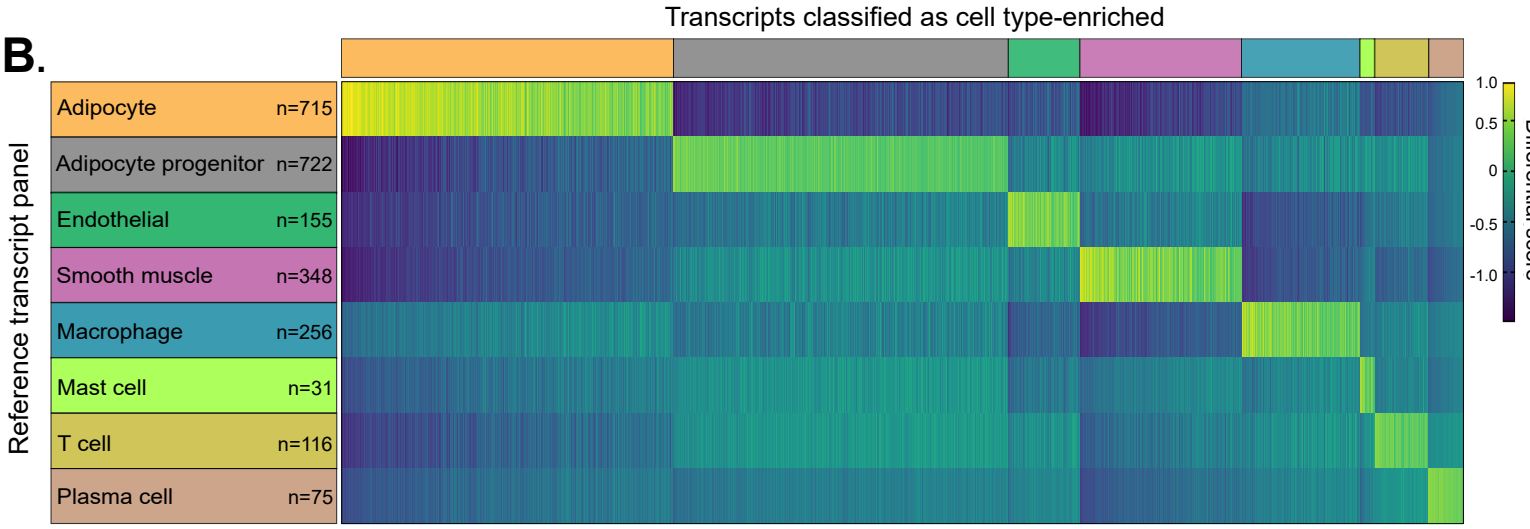
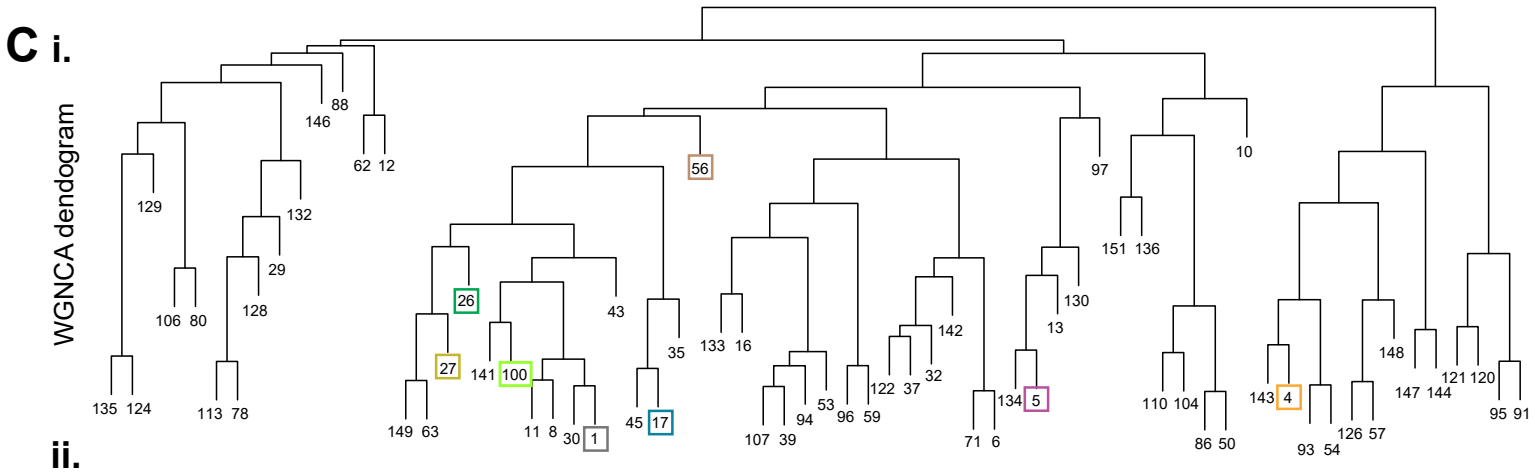

B.ii
Predicted: Macrophage (MC) enriched
Predicted: MC and NP enriched
Predicted: Neutrophil (NP) enriched
MC
NP
CD86
LILRA2
ARG1
CD300C
FPR1
PROK2
MS4A4A
MNDA
MGAM
10.0
7.5
5.0
2.5

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


A.
Adipocyte
Endothelial
Smooth muscle
Macrophage
T cell
PRKAR2B
CDH13
CASQ2
IFI30
SP140
C19orf12
SHANK3
SLC30A3
LCP1
CD247
200 µm
50 µm


B.
Transcripts classified as cell type-enriched
Reference transcript panel
Adipocyte n=715
Adipocyte progenitor n=722
Endothelial n=155
Smooth muscle n=348
Macrophage n=256
Mast cell n=31
T cell n=116
Plasma cell n=75
Differential score
1.0
0.5
0
-0.5
-1.0


C i.
WGNCA dendogram

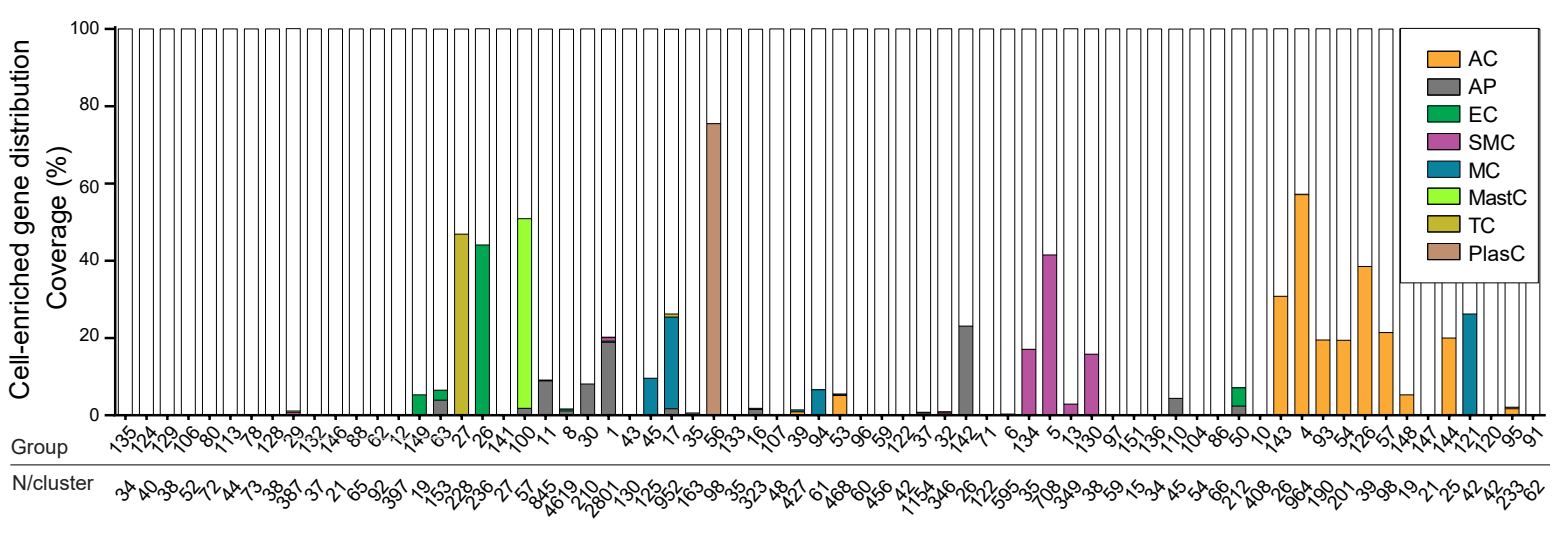
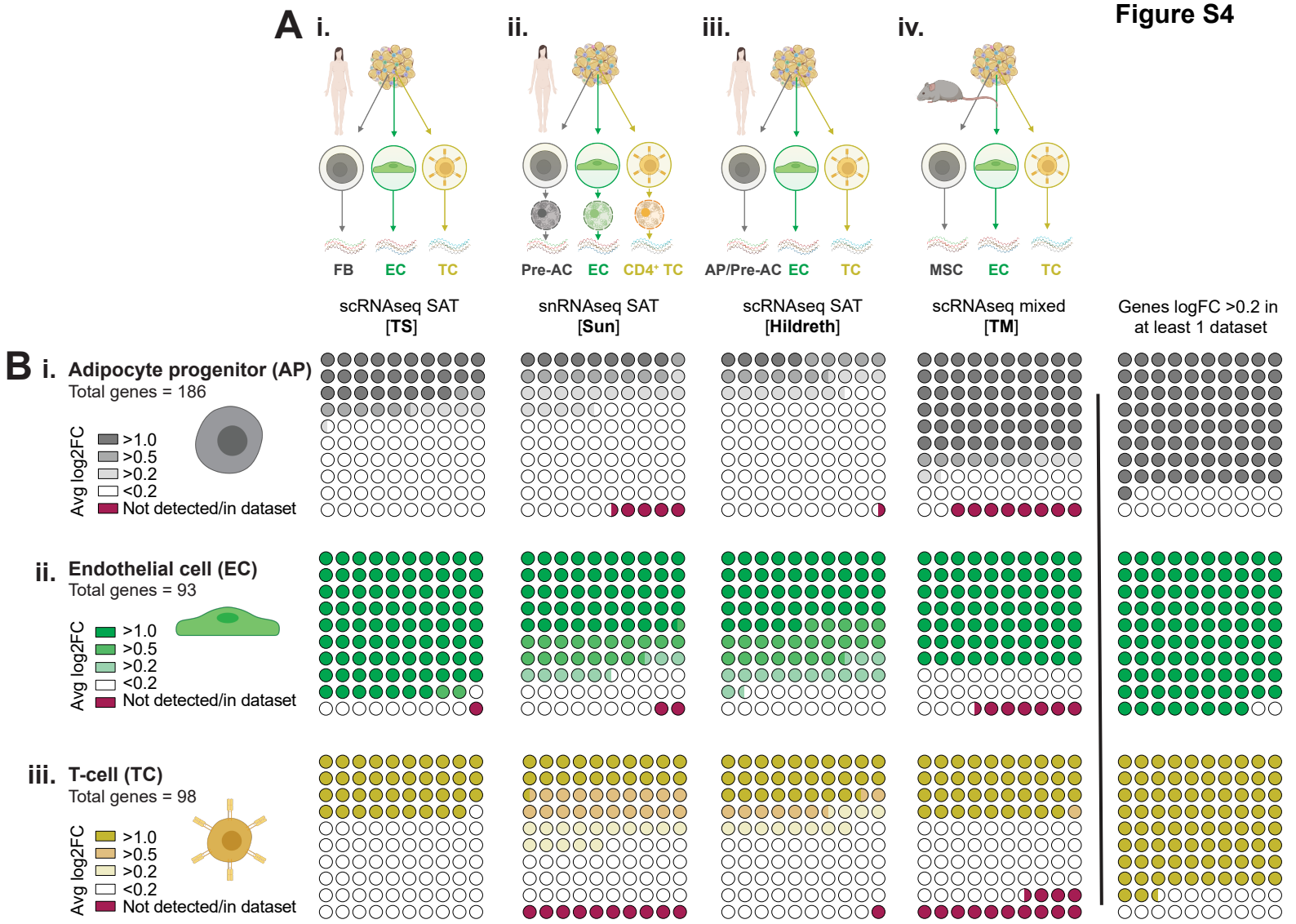

ii.
Cell-enriched gene distribution
Coverage (%)
AC
AP
EC
SMC
MC
MastC
TC
PlasC
Group
N/cluster

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.
ii.
iii.
iv.
FB EC TC
Pre-AC EC CD4+ TC
AP/Pre-AC EC TC
MSC EC TC
scRNAseq SAT [TS]
snRNAseq SAT [Sun]
scRNAseq SAT [Hildreth]
scRNAseq mixed [TM]
Genes logFC >0.2 in at least 1 dataset
B
i. Adipocyte progenitor (AP)
Total genes = 186
ii. Endothelial cell (EC)
Total genes = 93
iii. T-cell (TC)
Total genes = 98
Avg log2FC
>1.0
>0.5
>0.2
<0.2
Not detected/in dataset

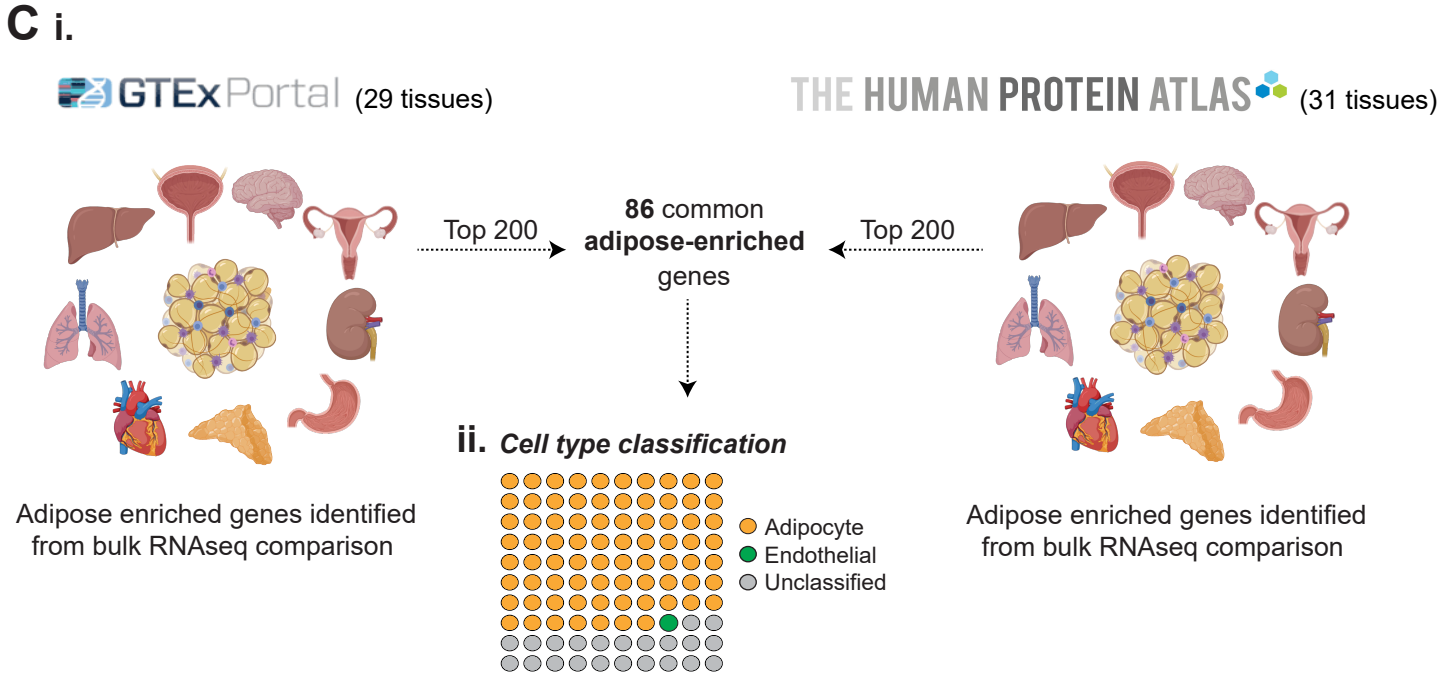

C i.
GTEx Portal (29 tissues)
THE HUMAN PROTEIN ATLAS (31 tissues)
Top 200
86 common adipose-enriched genes
Top 200
Adipose enriched genes identified from bulk RNAseq comparison
Adipose enriched genes identified from bulk RNAseq comparison
ii. Cell type classification
Adipocyte
Endothelial
Unclassified

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

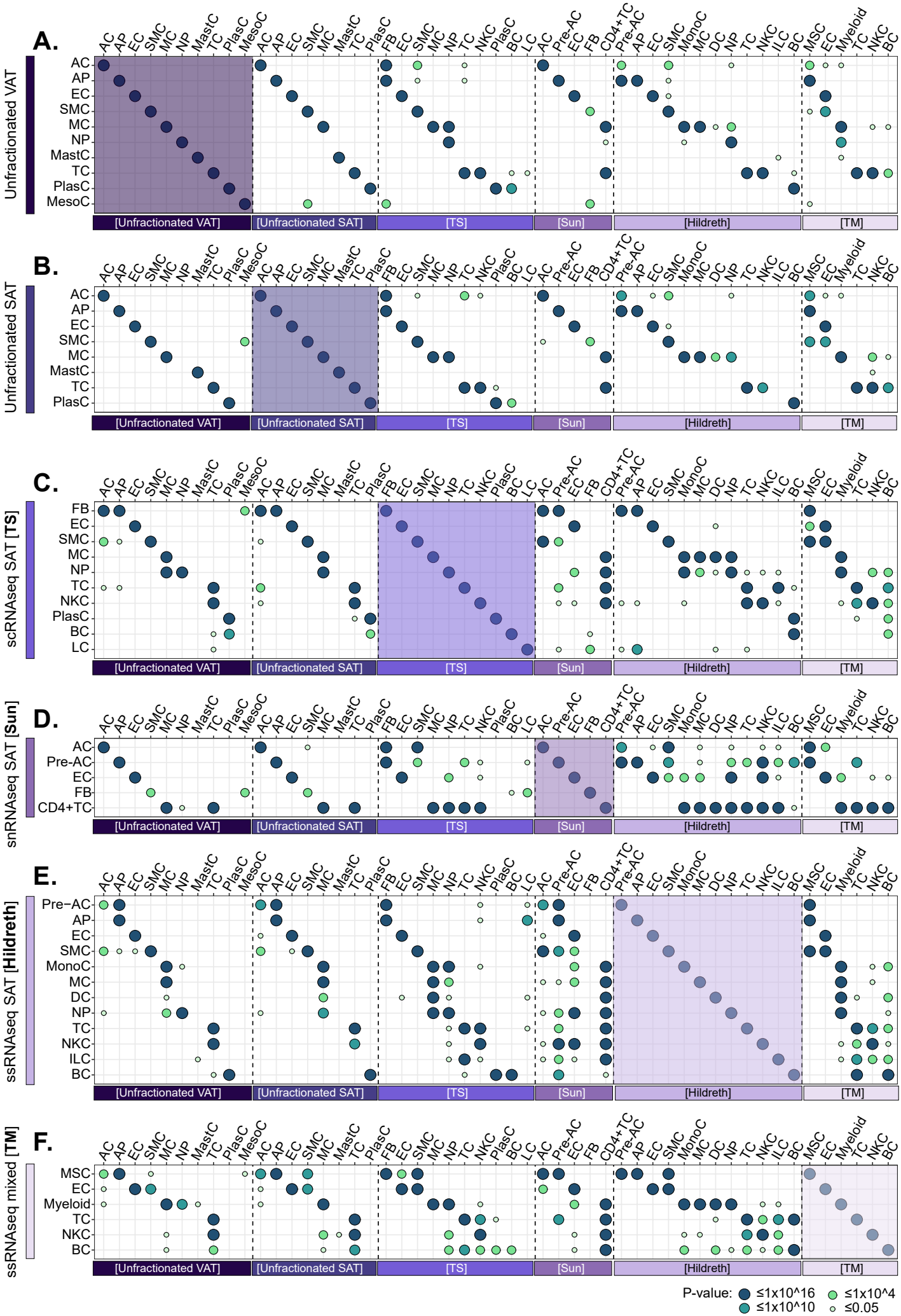

A.
Unfractionated VAT
AC AP EC SMC MC NP MastC TC PlasC MesoC
B.
Unfractionated SAT
AC AP EC SMC MC MastC TC PlasC
C.
scRNAseq SAT [TS]
FB EC SMC MC NP TC NKC PlasC BC LC
D.
snRNAseq SAT [Sun]
AC Pre-AC EC FB CD4+TC
E.
ssRNAseq SAT [Hildreth]
Pre−AC AP EC SMC MonoC MC DC NP TC NKC ILC BC
F.
ssRNAseq mixed [TM]
MSC EC Myeloid TC NKC BC
AC AP EC SMC MC NP MastC TC PlasC MesoC AC AP EC SMC MC MastC TC PlasC FB EC SMC MC NP TC NKC PlasC BC LC AC Pre-AC EC FB CD4+TC Pre-AC AP EC SMC MonoC MC DC NP TC NKC ILC BC MSC EC Myeloid TC NKC BC
[Unfractionated VAT] [Unfractionated SAT] [TS] [Sun] [Hildreth] [TM]
P-value: ≤1x10^16 ≤1x10^10 ≤1x10^4 ≤0.05

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

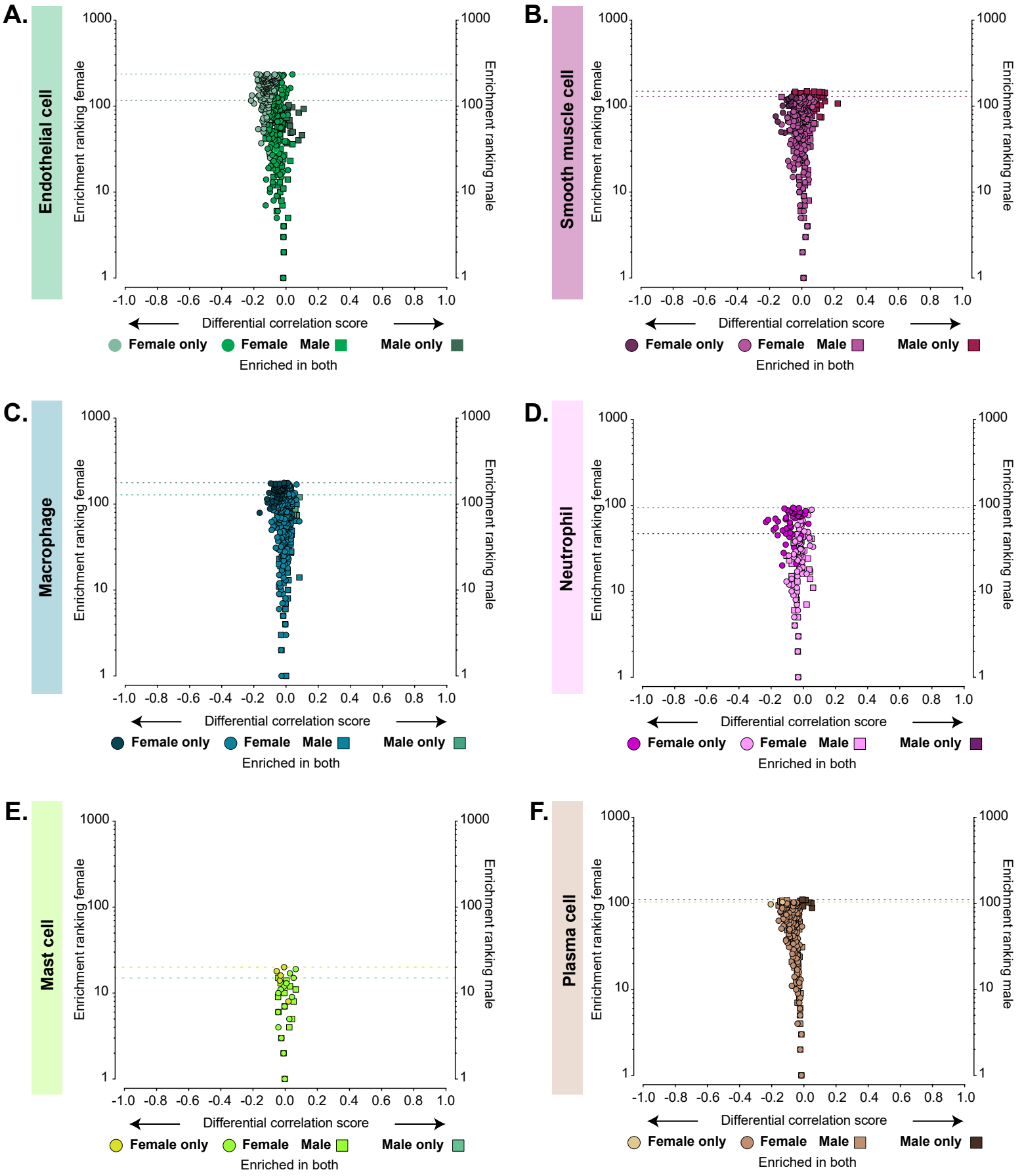

A.
Endothelial cell
B.
Smooth muscle cell
C.
Macrophage
D.
Neutrophil
E.
Mast cell
F.
Plasma cell
Enrichment ranking female
Enrichment ranking male
1000
100
10
1
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Differential correlation score
Female only
Female
Male
Male only
Enriched in both

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.

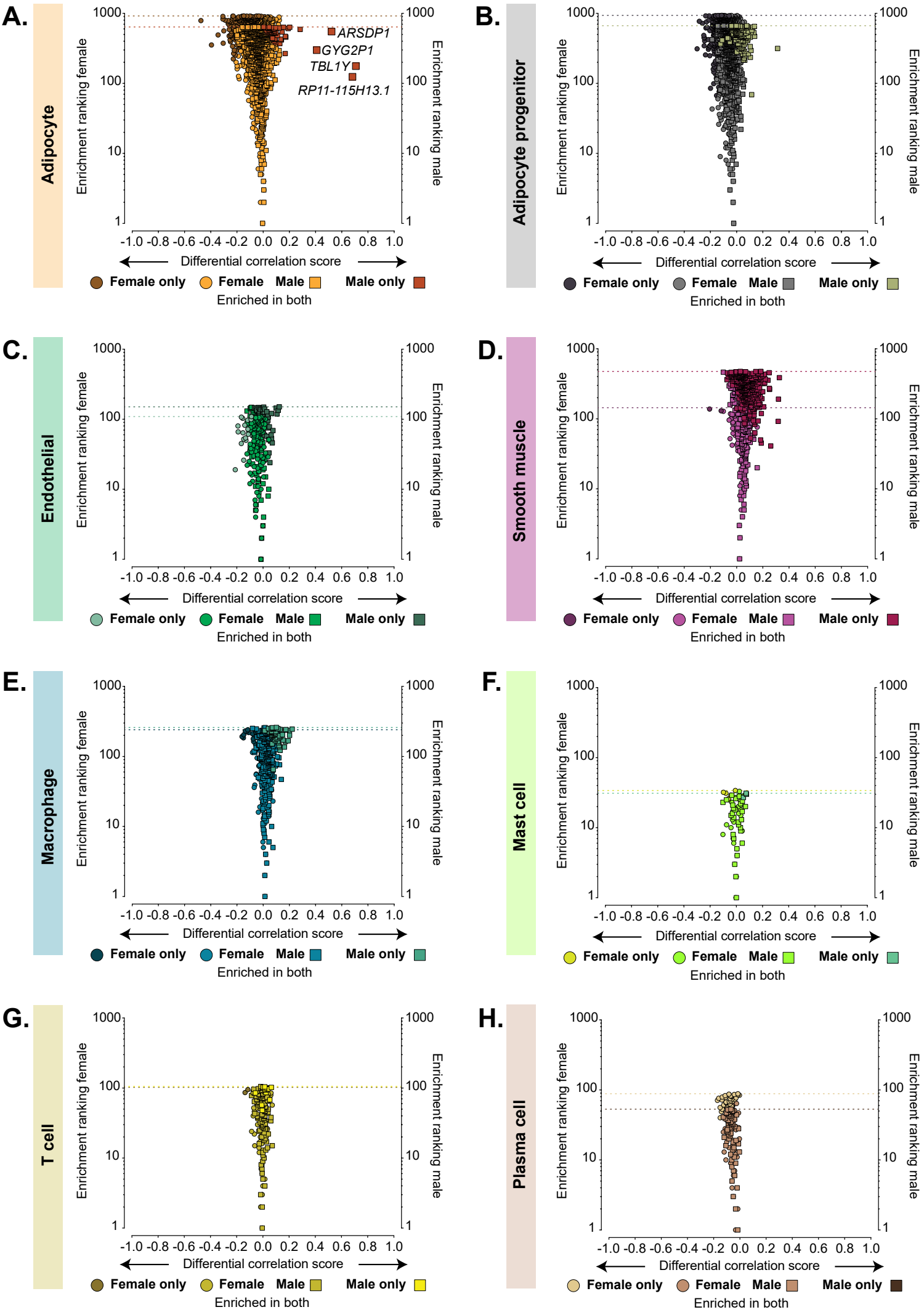

A.
Adipocyte
ARSDP1
GYG2P1
TBL1Y
RP11-115H13.1
B.
Adipocyte progenitor
C.
Endothelial
D.
Smooth muscle
E.
Macrophage
F.
Mast cell
G.
T cell
H.
Plasma cell
Enrichment ranking female
Enrichment ranking male
1000
100
10
1
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Differential correlation score
Female only
Female
Male
Male only
Enriched in both

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

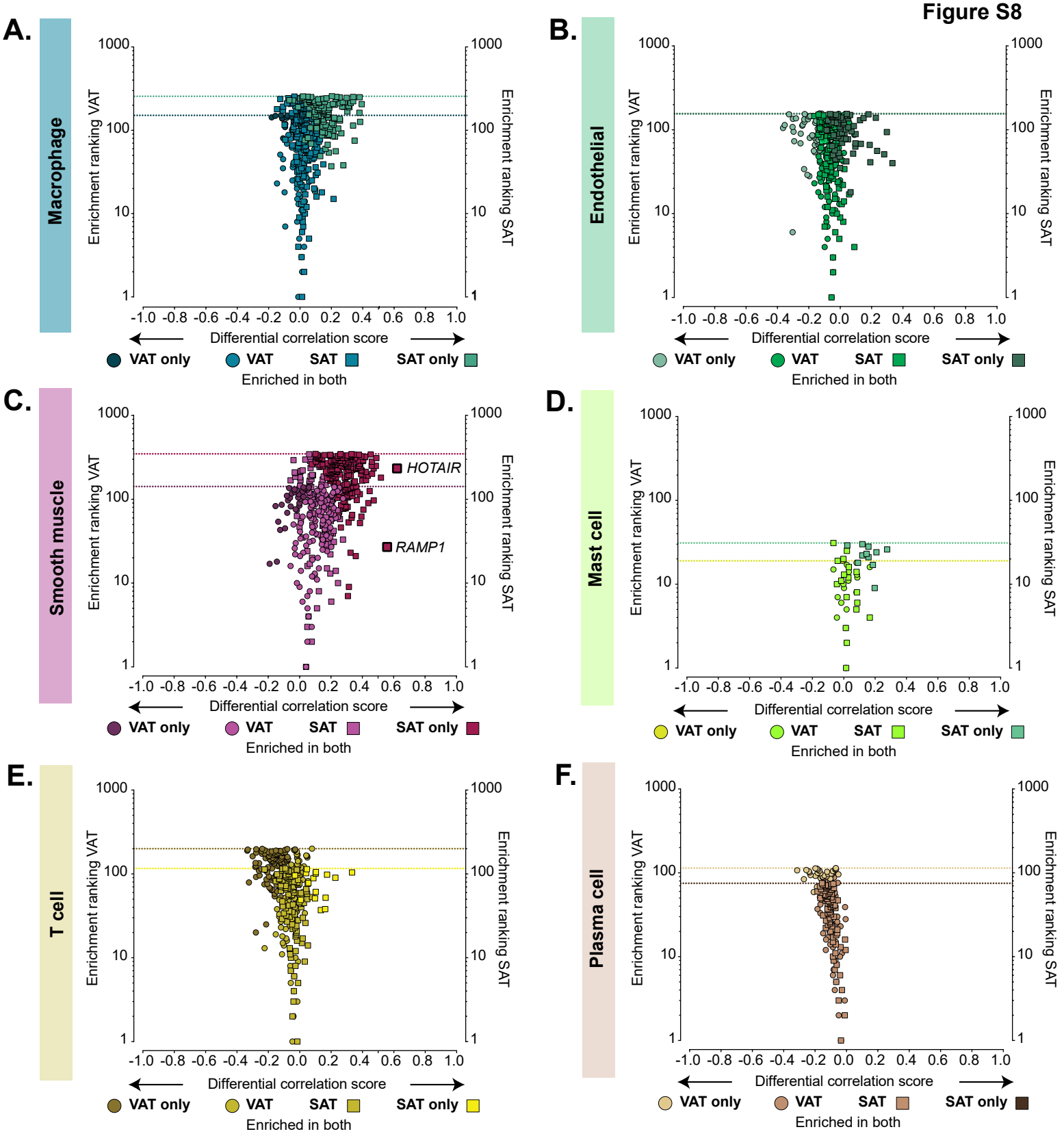

A.
Macrophage
B.
Endothelial
C.
Smooth muscle
HOTAIR
RAMP1
D.
Mast cell
E.
T cell
F.
Plasma cell
Enrichment ranking VAT
Enrichment ranking SAT
1000
100
10
1
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0
Differential correlation score
VAT only
VAT
SAT
SAT only
Enriched in both

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 II

A tissue centric atlas of cell type transcriptome enrichment signatures

P Dusart [1,2,3], S Öling[3], E Struck[3], M Norreen-Thorsen[3], M Zwahlen[2], K von Feilitzen[2], P Oksvold[2], M Bosic[4,5], MJ Iglesias[2], T Renné[6], J Odeberg[2,3,7,8], F Pontén[4], C Lindskog[4], M Uhlén[2], LM Butler [1,2,3,9*]

[1] Clinical Chemistry and Blood Coagulation Research, Department of Molecular Medicine and Surgery, Karolinska Institute, 171 76 Stockholm, Sweden

[2] Science for Life Laboratory, Department of Protein Science, Royal Institute of Technology (KTH), 171 21 Stockholm, Sweden

[3] Translational Vascular Research, Department of Clinical Medicine, The Arctic University of Norway, 9019 Tromsø, Norway

[4] Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, 752 37 Uppsala, Sweden

[5] Institute of Pathology, Faculty of Medicine, University of Belgrade, 11000 Belgrade, Serbia

[6] Institute for Clinical Chemistry and Laboratory Medicine, University Medical Centre Hamburg-Eppendorf, D-20246 Hamburg, Germany

[7] The University Hospital of North Norway (UNN), 9019 Tromsø, Norway

[8] Coagulation Unit, Department of Hematology, Karolinska University Hospital, 171 76 Stockholm, Sweden

[9] Clinical Chemistry, Karolinska University Laboratory, Karolinska University Hospital, 171 76 Stockholm, Sweden

Correspondence to:

Dr. L.M Butler, PhD

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: Cell profiling, bulk RNAseq, gene enrichment

SUMMARY

Genes with cell type specific expression typically encode for proteins that have cell type specific functions. Single cell RNAseq (scRNAseq) has facilitated the identification of such genes, but various challenges limit the analysis of certain cell types and lowly expressed genes. Here, we performed an integrative network analysis of over 6000 bulk RNAseq datasets from 15 human organs, to generate a tissue-by-tissue cell type enrichment prediction atlas for all protein coding genes. We profile all the major constituent cell types, including several that are fragile or difficult to process and thus absent from existing scRNAseq-based atlases. The stability and read depth of bulk RNAseq data, and the high number of biological replicates analysed, allowed us to identify lowly expressed cell type enriched genes that are difficult to classify using existing methods. We identify co-enriched gene panels shared by pancreatic alpha and beta cells, chart temporal changes in cell enrichment signatures during spermatogenesis, and reveal that cells in the hair root are a major source of skin enriched genes. In a cross-tissue analysis, we identify shared gene enrichment signatures between highly metabolic and motile cell types, and core identity profiles of cell types found across tissue types. Our study provides the only cell type gene enrichment atlas generated independently of scRNAseq, representing a new addition to our existing toolbox of resources for the understanding of gene expression across human tissues.

INTRODUCTION

Cell type can be categorised by function, origin, location, morphology and, more recently, global transcriptome. Transcriptional profiles depend on both intrinsic cell characteristics and transient states, but selective expression of genes typically required for cell type specialised functions currently underlie our definition of cell type. Large-scale projects, such as the Human Cell Atlas (www.humancellatlas.org) [1] and the Human Protein Atlas (www.proteinatlas.org/) [2,3] contain single-cell RNA sequencing (scRNA-seq) data from thousands of cells, which can be used to further understand human health and disease, through, for example, targeted biomarker discovery [4], or elucidation of disease associated gene expression [5,6].

However, scRNA-seq has limitations; cell processing can cause artefactual modification of gene expression, through induction of the stress response [7,8] or as a consequence of removal from the microenvironment [9]. Some cell types are sensitive to extraction protocols, e.g., kidney podocytes [8], whilst others require extensive, damaging proteolytic digestion to isolate e.g., adipocytes [10,11]; such cell types are absent from the major databases [3,12,13]. Single nuclei sequencing is an alternative tool for analysing such cell types [14], but resultant expression profiles are incomplete [15]. Compared to bulk RNA-seq, where all cell types in a tissue are sequenced without prior separation, scRNAseq produces less stable and more variable data, with a high number of zero values, particularly for lowly expressed genes [16-19], requiring computational imputation for interpretation [20,21], with methods remaining controversial [22]. Typically, tissues from a limited number of donors are analysed, resulting in underestimation of biological variance of gene expression and potential false discoveries when analysing differential expression between cell types or conditions [23-25]. Differentially expressed genes identified using scRNAseq typically have higher expression and smaller fold changes than those identified with bulk RNAseq [24].

We previously developed and validated an integrative correlation analysis method to identify cell type-enriched transcriptome profiles from unfractionated tissue RNAseq [26-28]. Our method circumvents some limitations of scRNAseq; hundreds of samples are analysed concurrently to reduce the influence of biological variation and batch effects, cell types that are technically challenging to process can be analysed, and lowly expressed cell enriched transcripts classified [28]. Here, we analysed over 6000 bulk RNAseq datasets from Genotype-Tissue Expression (GTEx) to generate a genome-wide, tissue-by-tissue cell type enrichment prediction atlas for all protein coding transcripts in 15 different human tissues. We provide gene enrichment signatures for all major constituent cell types, including those that are fragile or difficult to process, such as podocytes in the kidney and adipocytes in the breast, as well as

for minority cell types, such as those in the hair follicles of the skin. We identify co-enriched genes shared by related cell types, such as pancreatic alpha and beta cells, and chart temporal changes in gene enrichment during spermatogenesis. In a cross-tissue analysis, we identify common gene enrichment signatures, e.g., between respiratory ciliated cells and spermatids, endocrine cells in the pancreas, colon, thyroid, and stomach, and between cell types found in all or most tissues, such as endothelial and immune cell types. All data is available on the Human Protein Atlas (HPA) (www.proteinatlas.org/humanproteome/tissue+cell+type).

RESULTS

Cell type reference transcripts correlate across unfractionated tissue RNAseq data

Bulk RNAseq datasets for 15 human tissue types were retrieved from Genotype-Tissue Expression (GTEx) V8 (www.gtexportal.org) [29] (Figure 1A). To identify cell type-enriched transcript profiles, we performed an integrative correlation analysis on each dataset, using our previously published method [26-28].

As the tissue is unfractionated prior to sequencing, constituent cell types are present in different proportions in each sample (Figure 1 B.i [lung as an illustrative example]). Thus, each cell contributes mRNAs subsequently measured by RNAseq (Figure 1 B.ii), which can be: predominantly expressed in that cell type (cell type enriched), selectively expressed in two cell types (co-enriched), or expressed in several, or all, cell types within the tissue. For the main constituent cell types in each tissue (Figure 1 B.iii) marker ´reference transcripts´ [*Ref.T.*] were shortlisted (n=10-30), including: (**i**) those identified through in house tissue protein profiling [2] (**ii**) established markers identified in older ´none-omics´ studies, (**iii**) those identified by scRNAseq of mouse [13] or human [30] tissue, and (**iv**) markers from databases containing multiple studies e.g., Cell Marker [31], PanglaoDB [32] (Figure 1 B.iv). Spearman correlation coefficients were generated between all shortlisted candidate *Ref.T*. across each sample set, and three were selected to represent each cell type (for lung see Figure 1 B.v), based on the following criteria: (**i**) a high correlation between *Ref.T.* within each cell type panel (FDR <0.00001), consistent with cell type co-expression, (**ii**) a low correlation coefficient between *Ref.T.* in different cell type panels*,* consistent with high *specificity* of each panel (Figure 1 B.v) and (**iii**) a normal expression distribution of *Ref.T.* across samples. For all cell types, corresponding *Ref.T* and intra/inter *Ref.T* panel correlation coefficients in each tissue see Table S1, Tab 1, Table A-O.

Reference transcripts analysis can identify cell-type enriched gene signatures

For each tissue type analysed, the proportion of constituent cell types between samples vary, due to sampling and genetic factors [33,34], but ratios between constitutively expressed cell-specific genes remain relatively constant. Thus, high correlation of a given transcript with all *Ref.T.* in any one panel is consistent with selective expression in the corresponding cell type [28]. For all tissues, we generated correlation coefficients between each *Ref.T.* and all other sequenced transcripts (´test-transcripts´) and produced a list of provisional cell type-enriched transcripts, based on the following criteria: (**i**) the test-transcript had a mean correlation with a given *Ref.T.* panel ≥0.50 (FDR <0.0001), which was (**ii**) higher than the mean correlation with

any other Ref.T. panel. Resultant transcripts for each cell type were generally well separated from all others e.g., for lung: respiratory ciliated cells (RCC; Figure S1 A.i) and alveolar cell type 1 (AT1; Figure S1 Bi). However, in some cases, test-transcripts correlated well with more than one *Ref.T.* panel; panels typically representing closely related cell types, e.g., natural killer and T-cells (NK and TC; Figure S1 C.i), or those with functional commonalities, e.g., macrophages and alveolar type 2 (AT2) cells [35] (MC and AT2; Figure S1 D.i). To more carefully analyse the relationship between transcripts, the following was calculated for each to compare cell type lists: (**i**) the '*differential correlation score*', defined as the difference between the mean correlation of the test-transcript with the two sets of *Ref.T.*, e.g., respiratory ciliated cell (RCC) type panel [*ERICH3*, *DNAH12*, *SNTN*] and smooth muscle cell (SMC) panel [*TPM2*, *MYL9*, *TAGLN*] (Figure S1 A.ii) and (**ii**) the '*enrichment ranking*', based on the mean correlation value of the test-transcript with the *Ref.T.* panel (rank 1 = highest corr.). Transcripts that most highly correlated with the RCC *Ref.T.* panel separated well, from even the next closest cell type, SMC (Figure S1 A.ii), as did those most highly correlating with the alveolar cell type 1 (AT1) *Ref.T.* panel (Figure S1 B.ii). A panel of transcripts that most highly correlated with *Ref.T.* representing NK (Figure S1 C.ii, right side) or MC (Figure S1 D.ii, right side) had a low differential correlation score with *Ref.T.* for TC or AT2, respectively (Figure S1 C.ii and D.ii, left side), consistent with co-enrichment in both cell types, as we previously demonstrated [28]. scRNAseq data from human lung [36] was used to verify expression profiles of selected transcripts with predicted enrichment in one (Figure S1 A-D.iii and v) or both cell types (Figure S1 A-D.iv). For classification as single cell-type enriched, any transcript with a differential correlation score <0.15 vs. any *Ref.T.* panel representing a different cell type was excluded, on the basis of predicted co-enrichment (e.g., Figure S1 A-D.ii, grey shaded area). Application of these criteria across tissues generally resulted in intra-tissue cell-enriched gene panels that were well separated from each other (example for lung; Figure 1 C.i-xvi). For some cell types, these default thresholds were decreased when overlap with other *Ref.T*. panels was absent e.g., for erythroid cells in the liver (Figure S1 E.i and ii) or increased when overlap remained (details provided in Table S1, Tab 3). Gene ontology (GO) analysis [37], performed to identify over-represented classes and pathways among genes identified as cell type enriched produced resultant terms consistent with expected cell type functions, e.g. for lung respiratory ciliated cells, significant terms included *´cilium organisation´* (FDR 4.4 $x10^{-63}$) (Figure 1 C.i), and for plasma cells *´adaptive immune response´* (FDR 3.0 $x10^{-189}$) (Figure 1C.xiv). Tissue profiling for selected proteins encoded by predicted cell type enriched genes had expression consistent with our classifications (Figure 1 C.i-xvi).

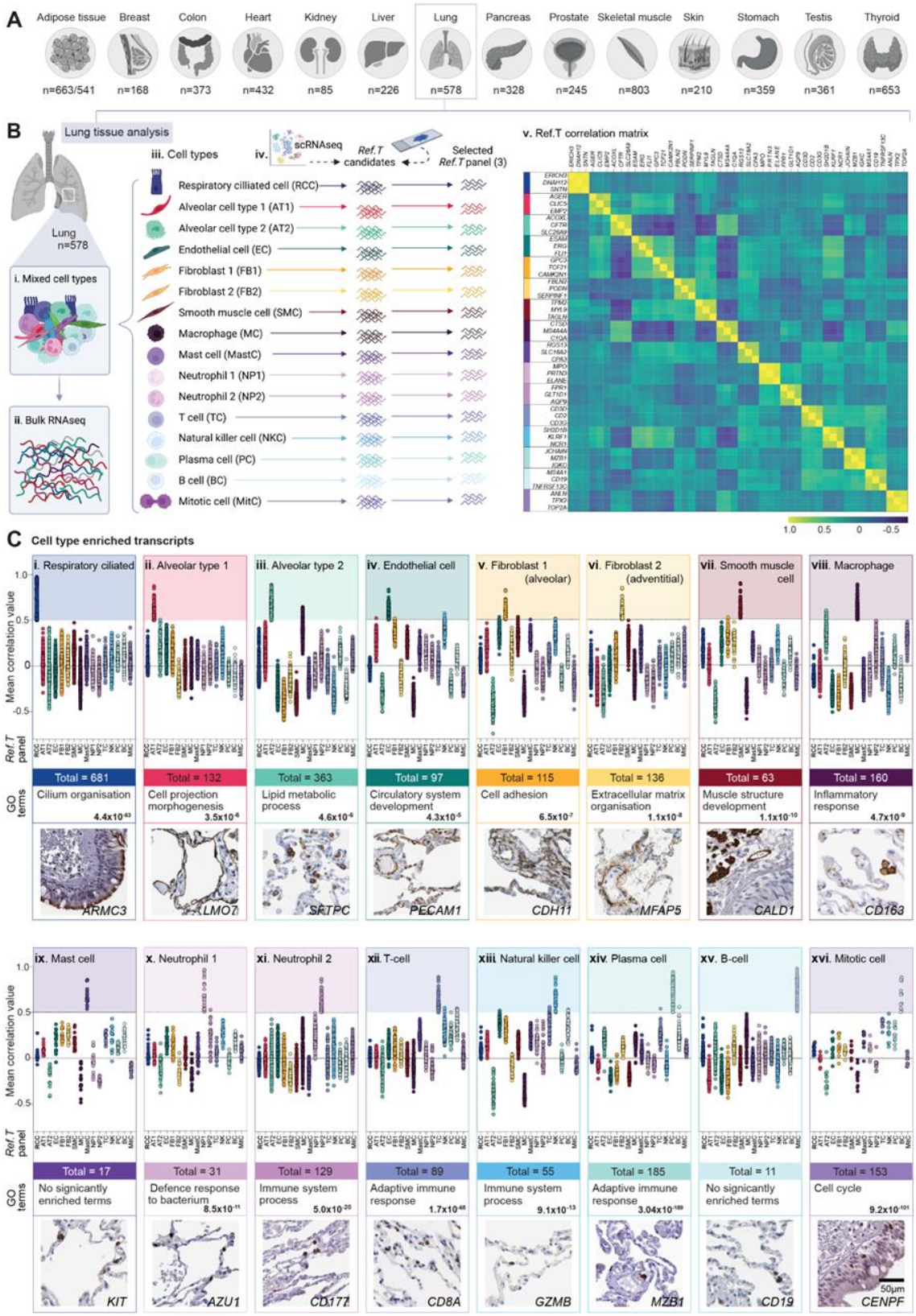


Figure 1. Integrative co-expression analysis of unfractionated human lung tissue RNAseq can resolve constituent cell type enriched genes. (A) Bulk RNAseq datasets were retrieved from GTEx V8 and analysed by tissue type (n=sample number). **(B)** Analysis concept, using lung as an illustrative example: **(i)** each sample (n=578) contained mixed cell types, contributing **(ii)** differing proportions of

mRNA to each sequenced dataset. To profile cell type-enriched transcriptomes **(iii)** constituent cell types for each tissue were identified and for each **(iv)** candidate reference transcripts (*Ref.T.*) for ´virtual tagging´ were shortlisted, primarily based on predicted cell specificity from existing literature and/or in house protein profiling. **(v)** Matrix of correlation coefficient values between selected *Ref.T.* across the sample set. (**C**) Mean correlation coefficients between genes above designated thresholds for classification as cell-type enriched in: **(i)** respiratory ciliated [RCC], **(ii)** alveolar type I [AT1], **(iii)** alveolar type II [AT2], **(iv)** endothelial [EC], **(v)** alveolar fibroblasts [FB1], **(vi)** adventitial fibroblasts [FB2], **(vii)** smooth muscle cell [SMC], **(viii)** macrophage [MC], **(ix)** mast cell [MastC], **(x-xi)** neutrophil [NP1 and NP2], **(xii)** T-cell [TC], **(xiii)** natural killer cell [NK], **(xiv)** plasma cell [PC], **(xv)** B-cell [BC], or **(xvi)** mitotic cell [MitC], and all *Ref.T.* panels. Total number, most significant gene ontology (GO) terms and illustrative protein profiling in human lung tissue are provided for each cell type. See also Table S1, Figure S1 and S2.

Weighted network correlation analysis supports cell type enrichment predictions

As our analysis method is based on manually selected *Ref.T.*, cell type classification is subject to an input bias. However, we previously showed that unbiased weighted network correlation analysis (WGCNA) [38], where correlation coefficients between all transcripts are calculated and subsequently clustered into related groups (based on expression similarity), supports *Ref.T.* based analysis cell type enrichment predictions [27,28]. Here, we performed WGNCA of lung and liver samples (Figure S2). Both *Ref.T* (Figure S2 A-B.i) and predicted cell-type enriched gene panels (Figure S2 A-B.ii-ix) clustered into the same, or closely related WGCNA groups when the differential correlation for exclusion was set at >0.15 (as described above) (Figure S2 A-B.v). When the differential correlation was increased in increments of 0.05 (Figure S2 A-B.vi-ix) the number of predicted cell type enriched genes outside the predominant WGCNA clusters decreased (see red dashed box), consistent with higher enrichment specificity. Gene enrichment could thus be categorised into *very high*, *high* or *moderate*, corresponding to a differential score vs. other profiled cell types within the tissue of >0.35, >0.25 or >0.15, respectively (see Table S1, Tab 3 for total number in each category for all cell types/tissues).

Specialised cell types have the highest number of enriched genes within tissues

The total number of genes with predicted cell type enrichment (*very high*, *high* or *moderate*) within each tissue ranged from 7041 (testis) to 829 (pancreas) (Figure 2 A) (Table S1, Tab 3). The number of cell types analysed in each tissue type ranged from 7-18; with the lowest number profiled in skeletal muscle and subcutaneous adipose tissue (n=7 and 8, respectively) and the highest in skin and lung (n=18 and 14, respectively) (Table S1, Tab 1).Tissue specialised cell types had the highest number of enriched genes, such as cardiomyocytes in the heart (number/total enriched in all cell types in that tissue: 916/1902 [48%]) (Figure 2 B.v), proximal tubular cells in the kidney (657/1778 [37%]) (Figure 2 B.vii), hepatocytes in the liver (1264/2393 [53%]) (Figure 2 B.xi), keratinocytes in the skin (945/2460 [38.4%]) (Figure 2 B.xiii),

gastric mucosal cells in the stomach (379/1361 [28%]) (Figure 2 B.xiv) and respiratory ciliated cells in the lung (681/2419 [28%]) (Figure 2 B.xv).

Of the 19,634 protein coding genes expressed in one or more tissues, 5644 (28.7%) were not predicted to be cell type enriched in any tissue (Figure 2 C.i). GO analysis identified the most significant over-represented pathways among these genes as *´metabolism of RNA´* (FDR 4.6 $x10^{-21}$), *´gene expression (transcription)´* (FDR 2.3 $x10^{-11}$) *´RNA polymerase II transcription´* (FDR 5.4 $x10^{-10}$) and *´rRNA processing ´* (FDR 5.8 $x10^{-10}$) (subgroups shown in Figure 2 D), consistent with housekeeping function. Indeed, 2893 of these 5644 genes (52.3%, $p<10^{-15}$) had been previously categorised as members of the housekeeping proteome [2].

5979 (30.4%) genes were classified as cell type enriched in only a single tissue (Figure 2 C.ii), the largest proportion of which were in testis (n=3141) (Table S1, Tab 4). GO term analysis of this gene group identified the most significant over-represented pathways as *´sexual reproduction´* (FDR 3.7 $x10^{-32}$) and *´spermatogenesis´* (FDR 2.9 $x10^{-30}$) (subgroups shown in Figure 2 E). Of the 8011 genes predicted to be cell type enriched in multiple tissues (Figure 2 C.iii), a small number (741, 9.2%) were enriched in seven or more; the majority of which were predicted to be immune cell-, endothelial cell- or stromal cell- enriched (Figure 2 F), i.e., in cell types profiled in all, or most, tissues. Enrichment scores for all genes in cell types by tissue type can be found in Table S2 (summary of cell type gene enrichment across tissue in Table S1, Tab 4).

Ref.T. analysis can predict source of tissue enriched genes

RNAseq data from unfractionated human tissues can be used to identify genes with higher expression in any given tissue, compared to others. For genes classified as *tissue enriched* in the Human Protein Atlas (HPA) [2], those we classified as cell type enriched were predominantly expressed in tissue specialised cell types, for example, heart enriched genes were predominantly cardiomyocyte enriched and liver-enriched genes predominantly hepatocyte enriched (Figure S3 A). A hypergeometric test was performed to determine similarity between predicted cell type enriched genes and the top 300 enriched genes in each tissue in the GTEx data [29] (as collated in the Harminozome database [39]); similar to the comparison with the HPA data, the highest statistical overlap between tissue enriched genes and cell enriched genes were predominantly with tissue specialised cell types (Figure S3 Bi-vi). This highlights the usefulness of our analysis of bulk RNAseq to disentangle cell type variance across the different tissues in the human body, independent of scRNAseq data.

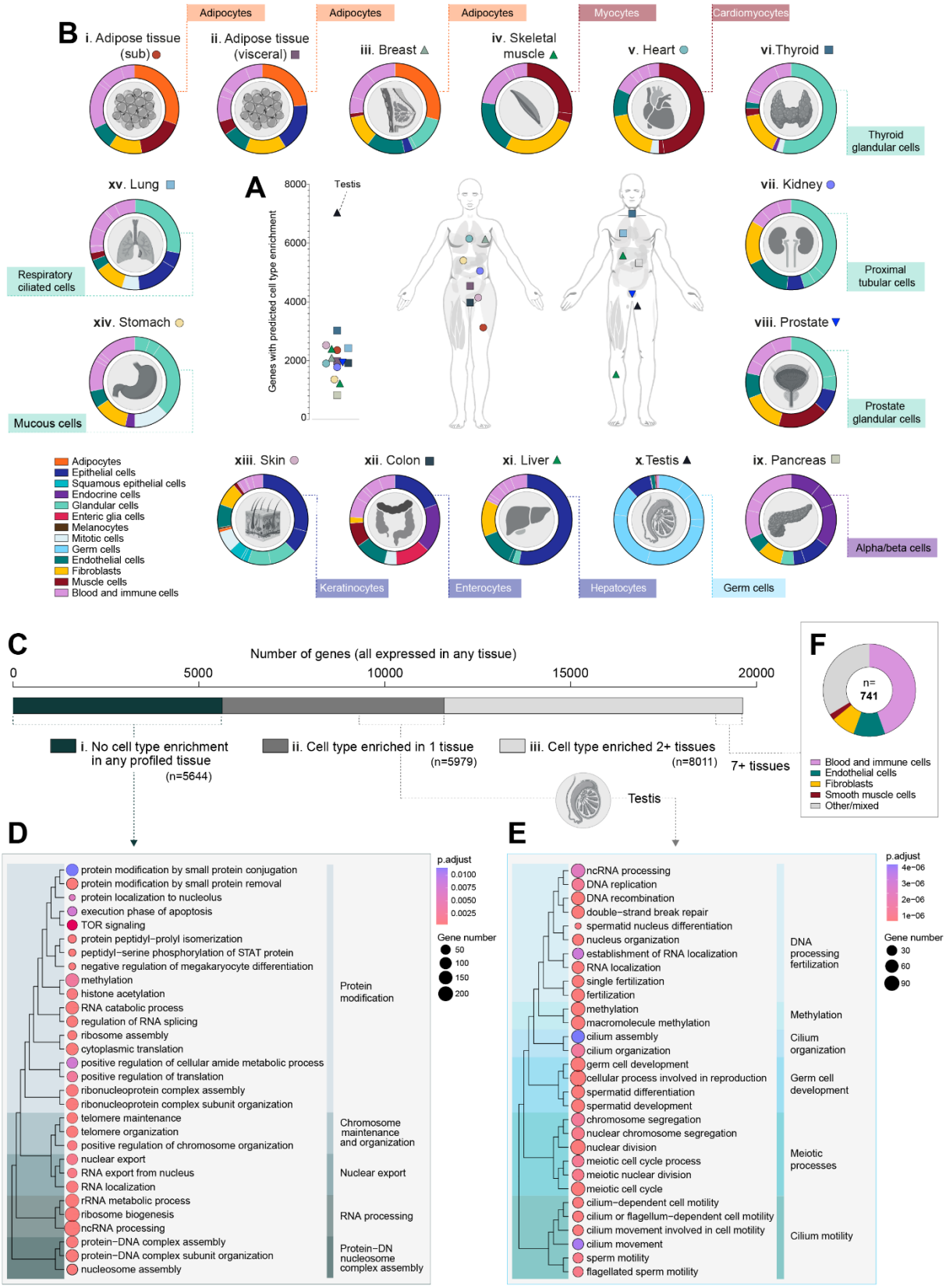


Figure 2. Overview of cell type enriched gene profiles across tissue types. Bulk RNAseq datasets were retrieved from GTEx V8 and cell type enriched transcriptome predictions made using integrative

correlation analysis. **(A)** Number of genes with predicted cell type enrichment in each analysed tissue type. **(B)** Circular plots showing broad classification of genes predicted to be cell type enriched in: **(i)** subcutaneous adipose tissue, **(ii)** visceral adipose tissue, **(iii)** breast, **(iv)** skeletal muscle, **(v)** heart, **(vi)** thyroid, **(vii)** kidney, **(viii)** prostate, **(ix)** pancreas, **(x)** testis, **(xi)** liver, **(xii)** colon, **(xiii)** skin, **(xiv)** stomach and **(xv)** lung, with majority cell types indicated in connected boxes. **(C)** Total number of expressed genes (in at least one tissue type) by respective status: **(i)** no cell type enrichment in any tissue, **(ii)** prediction as cell type enriched in one tissue, or **(iii)** predicted to be cell type enriched in two or more tissues. Gene ontology overrepresented terms for genes with: **(D)** no predicted cell type enrichment and **(E)** predicted enrichment only in testis. **(F)** Cell type enrichment predictions for genes classified as enriched in seven or more tissue types. See also Table S1 and S2 and Figure S3.

Pancreatic alpha and beta cells have both specific and shared gene enrichment profiles

Alpha and beta cells, the most abundant endocrine cell types in the pancreatic islet of Langerhans [40], are defined by their expression of the blood glucose elevating or lowering hormones, glucagon (*GCG*) and insulin (*INS*), respectively. As a general rule, transcripts predicted to be cell type enriched generally separated well from others, but analysis of pancreas samples (n=328) revealed that many transcripts that correlated most highly with the alpha cell *Ref.T.* panel also correlated well with the beta cell *Ref.T.* panel (Figure 3 A.i), and *vice versa* (Figure 3 A.ii). Analysis of individual transcripts revealed 131 genes highly and selectively correlated with the *Ref.T*. panels for *both* alpha and beta-cells (Figure 3B, [grey central panel; mean differential corr. between Ref.T panels <0.15]). GO and reactome analysis [41] of these 131 co-enriched genes revealed over-represented classes and pathways included *´regulation of secretion by cell´* (FDR 7.5 x10^{-11}), *´neuronal system´* (FDR 9.9 x10^{-7}) and *´synapse´* (FDR 1.5 x10^{-15}) (Table S3, Tab 1, Tables A-C). Synapse related proteins (n=44) included members of the synaptotagmin (*SYT4*, *5*, *7*, *13, 14*), and glutamate receptor (*GRIA2*,*3*) families (Table S3, Tab 2), many of which are reported to be important for pancreatic endocrine cell function e.g., *SYT4* [42] and *SYT13* [43,44], whilst the function of others in this context is not currently known e.g., *FRRS1L and NSG1.* Alpha and beta cell co-enriched genes included several encoding for transcription factors involved in islet cell specification, e.g., *NKX2-2*, [45], *NEUROD1* [46], *RFX6* [47], *INSM1* [48], *PAX6* [49] and *MYT1* [50], as well as those with no currently reported function in these cell types, e.g., *CELF3* and *MYT1L* one could speculate such genes likely have a role in neuroendocrine cell function. 91 genes had predicted alpha cell-enrichment, including *GCG*, *TTR* and *KCNH6* (Figure 3 B, left side); all of which are involved in glucose homeostasis [51-53], and other genes with, as yet, no described function in this cell type e.g., *SMIM24*, *CALY* and *C5orf38* (Figure 3 B, left side). 69 genes had predicted beta cell enrichment, including those encoding proteins with known beta cell-specific functions, e.g., *IAPP* and *MAFA* [54,55], as well as those with no reported function in this cell type, e.g., *HHATL, SNCB* and *SLC6A17* (Figure 3 B, right side).

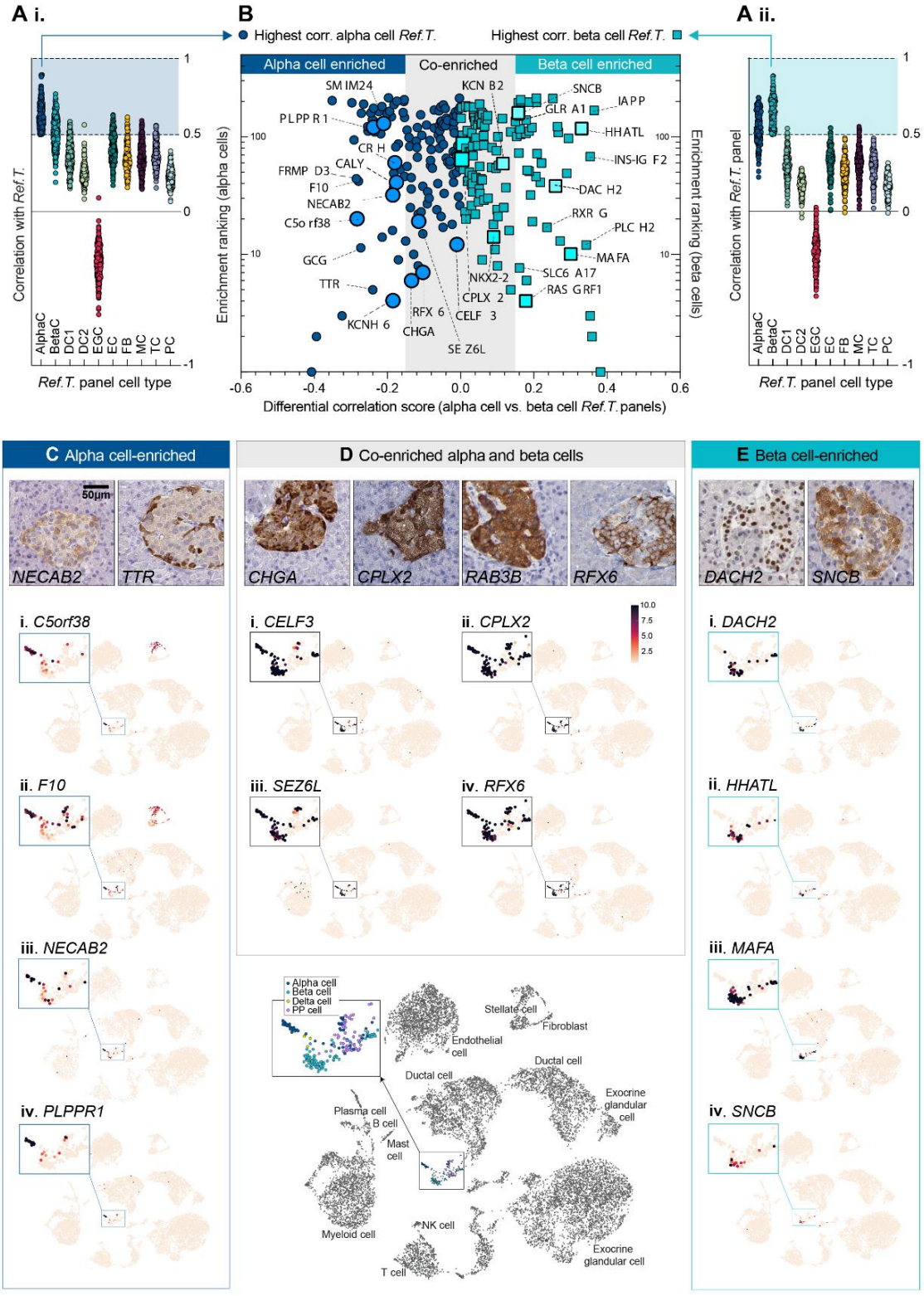


Figure 3. Pancreatic alpha and beta cells express respective cell type enriched genes and a panel of shared co-enriched genes. RNAseq datasets for human pancreas (n=328) were retrieved from GTEx V8 and correlation coefficients between selected cell type Ref T. and all others were generated. **(A)** Mean correlation values between protein coding genes that correlated most highly with **(i)** alpha or **(ii)** beta cell Ref.T. (above >0.50) and all Ref.T. panels. **(B)** For these transcripts, the 'differential

correlation score' (difference between mean correlation with alpha and beta cell Ref.T.) was plotted vs. 'enrichment ranking' (position in each respective list, highest correlation = rank 1). Shaded grey box highlights genes with differential correlation <0.15. Genes highlighted in bold correspond to those featured in the lower panels. Tissue protein profiling of selected genes predicted to be **(C)** alpha cell-enriched, **(D)** co-enriched in both alpha and beta cells, or **(E)** beta cell-enriched, in human pancreas samples. scRNAseq data from analysis of human pancreas was sourced from Tabula Sapiens (Tabula Sapiens et al., 2022), and used to generate UMAP plots, showing the expression profiles of example genes we predicted as being **(C)** alpha cell-enriched; **(i)** C5orf58, **(ii)** F10 **(iii)** NECAB2 and **(iv)** PLPPR1, **(D)** co-enriched in both alpha and beta cells; **(i)** CELF3, **(ii)** CPLX2, **(iii)** SEZ6L and **(iv)** RFX6, or **(E)** beta cell-enriched; **(i)** DACH2, **(ii)** HHATL, **(iii)** MAFA and **(iv)** SNCB. scRNAseq cell type annotations are displayed on lower central plot. AlphaC; alpha cell, BetaC; beta cell, DC1; ductal cell 1, DC2; ductal cell 2, EC; endothelial; FB1/2; fibroblast 1/2, SMC; smooth muscle cell, MC; macrophage, MastC; mast cell, NP1/2; neutrophil 1/2, TC; T-cell, PC; plasma cell. See also Table S3 and Figure S4.

Tissue profiling for selected genes showed protein expression consistent with our classifications (Figure C-D top panel). We sourced data from scRNAseq of human pancreas [36], to compare the expression profiles of selected predicted alpha- (Figure 3 C.i-iv), beta- (Figure 3 E.i-iv) or co- (Figure 3 D.i-iv) enriched genes; categorisation was largely consistent between datasets. A small number of genes we predicted to be alpha-, beta or co-enriched had a mean expression <0.1 TPM in the analysed bulk RNAseq dataset (gene n=11, 6 and 4, respectively, Figure S4 A). Despite this low expression, our predicted expression of these genes was consistent with the scRNAseq analysis; with most (21/22 [95%]) detected predominantly in the correspondingly annotated cell types (Figure S4 C-E). However, for several of these genes, detectable expression by scRNAseq was low, or only evident in a small number of cells within the cluster, e.g., *GLB1L3* (Figure S4 C.ii). The interpretation of such scRNAseq data is challenging; thus, our classifications, based on a completely independent data collection and analysis method, can be used to verify that low or irregular detection of gene expression by scRNAseq in an annotated cell type supports a real biological phenomenon, as opposed to noise or imputation artefact.

Temporal changes in gene enrichment signature underlie the process of spermatogenesis

Precise definitions, markers and terminology used for the respective cell types in the different stages of spermatogenesis vary between studies. Our analysis was based on four *Ref.T.* panels (S1-S4) that were selected to represent the temporal order of development: S1, germ cell expressed [*MAGEB2*, *KDM1B, PIWIL4*] (spermatogonia), S2, meiotic cell cycle expressed [*ANKRD31*, *RBM44, TOP2A*] (spermatocytes), S3, spermatid structure-related [*CEP55, KPNA5*, *PBK*] (round/early elongating spermatids) and S4, nuclear condensation/protamine repackaging factors [*PRM1*, *PRM2*, *TNP1*] (late/elongated spermatids) (Figure 4 A and Table S1, Tab 1, Table N). When the sample set was analysed by WGNCA, *Ref.T.* within each respective panel were all in a common module (Figure S5 A). Principle component analysis of

the corr. values of cell-enriched genes vs. all *Ref.T*. panels revealed the greatest proportion of variance in enrichment, and thus uniqueness vs. other cell types, was driven by cell types S1, S2, S3, S4 (Figure 4 B). Tissue profiling for proteins encoded by a panel of genes predicted to be enriched in cell types outside those in the spermatogenesis pathway revealed expression consistent with our classifications (Figure 4 C). 6179 genes were enriched in one or more of the germ cell types representing the different stages of spermatogenesis, vs. non-germ cell types (Figure S5 B.i and Table S4, Tab 1 [correlation with respective Ref.T panel >0.50, differential correlation vs. all non-germ cell types >0.15] columns H-K and Q). GO and reactome analysis of this gene list revealed that the most significantly over-represented terms included *´sexual reproduction´* (FDR 3.1 $x10^{-27}$), *´microtubule-based processes´* (FDR 2.2 $x10^{-26}$), *´male gamete generation´* (FDR 2.3 $x10^{-26}$) and *´cell cycle´* (FDR 4.6 $x10^{-19}$) (Table S4, Tab 2, Tables A and B) (Figure S5 B.ii [summary plot of GO terms, made with REVIGO [56]]). Genes that correlated with *Ref.T.* panels representing cells at different stages of spermatogenesis had two main profiles; they were enriched at a specific developmental stage, i.e., S1 (Figure 4 D.i), S2 (Figure 4 D.ii) S3 (Figure 4 D.iii) or S4 (Figure 4 D.iv) (for all see Figure S5 C .i and ii) or, they were co-enriched in adjacent cell types on the developmental trajectory: i.e., S1 and S2 (Figure 4 D.v), S2 and S3 (Figure 4 D.vi), S3 and S4 (Figure 4 D.vii) or S2, S3 and S4 (Figure 4 D.viii) (for all see Figure S5 D.i and ii). Each plot shows five illustrative genes for each enrichment profile type (Figure 5 Ei-vii), including genes encoding for proteins with a previously reported function at the corresponding stage of spermatogenesis e.g., for S1: *FGFR3* [57], and those with no known role in this context e.g., *C19orf84*(Figure 4 E.i). Protein profiling revealed spatial distribution for those encoded by genes classified as S1, S2, S3 or S4-enriched or co-enriched, with positive signals observed progressively closer to the centre of the seminiferous tubule with each subsequent developmental stage (Figure 4 E.i-vi). GO analysis revealed over-represented classes in genes predicted to be S1, S2- or S1 & S2 enriched included developmental, cell cycle and meiotic processes (Figure 4 F.i, ii and v), whilst organelle assembly, microtubule processes and cilium and flagellum organisation and motility associated genes were over represented in S3-, S4- and S3 & S4-enriched genes (Figure 4 F.iii, iv and vii) (Table S4, Tab 3). No transcripts were predicted to be co-enriched in non-adjacent cell stages along the developmental trajectory (e.g., S1 and S3, or S2 and S4), consistent with a coordinated temporal modification in gene enrichment signatures between subsequent stages. A single gene, *MEIOC*, was predicted to be enriched in 3 stages - S2, S3 and S4. *MEIOC* is required for germ cells to properly transition to a meiotic cell cycle program, together with binding partners *YTHDC2* and *RBM46* [58]; both of which we also predicted as enriched in cells in S2 and, to a lesser extent S3 (Table S4, Tab 1).

Data from scRNAseq of human adult testis [59] supported our predictions, showing *MEIOC* enrichment in cell clusters broadly corresponding to our classification of S2, S3 and S4 (Figure 4 G.i) (cell type annotation UMAP as in the original publication in Figure S5). In contrast, we predicted that the related transcript *MEIOB* had specific enrichment at stage S2 (Figure 4 D.ii and E.ii), which was also verified by scRNAseq (Figure 4 G.ii). scRNAseq for selected less well characterised genes that we predicted as enriched in either S1, S2, S3 or S4 cells (Figure S5 C.iii), or gene predicted to be co-enriched in two stages (Figure S5 D.iii) also showed agreement with our classifications. A number of genes that were predicted to be enriched in one or more of the germ cell stages were lowly expressed (n=240 with mean TPM<0.1), several of which did not appear in the scRNAseq dataset [59], presumably due to a lack of detection. Of the 100 most lowly expressed genes for which scRNAseq data was available, most (>80%) had expression profiles consistent with our predictions in the scRNAseq data (examples in Figure S6), but in many cases transcripts were detected at low levels in only a few cells in the corresponding cluster, e.g., *FZD10* (Figure S6 D), *LEP* (Figure S6 F) and *SIGLEC15* (Figure S6 J), making interpretation of this scRNAseq data in isolation challenging. Thus, we show that analysis of bulk RNAseq can identify differentially enriched genes associated with one or multiple stages of the developmental trajectory during spermatogenesis, including genes that are likely too lowly expressed for detection or classification as cell type enriched by scRNAseq.

RNAseq data from unfractionated tissue can be used to identify genes with enriched expression in testis vs. other tissues, as we previously described . The vast majority of genes with testis-enriched expression were predicted to be enriched in one or more germ cell type (845/871 [97%]), with a smaller number predicted to be enriched in sertoli (24/871 [2.8%]), Leydig (24/871 [0.1%]) or peritubular cells (24/871 [0.1%]) (Figure 4H). No testis enriched genes were classified as endothelial or macrophage-enriched in our analysis, reflecting their presence in other tissues.

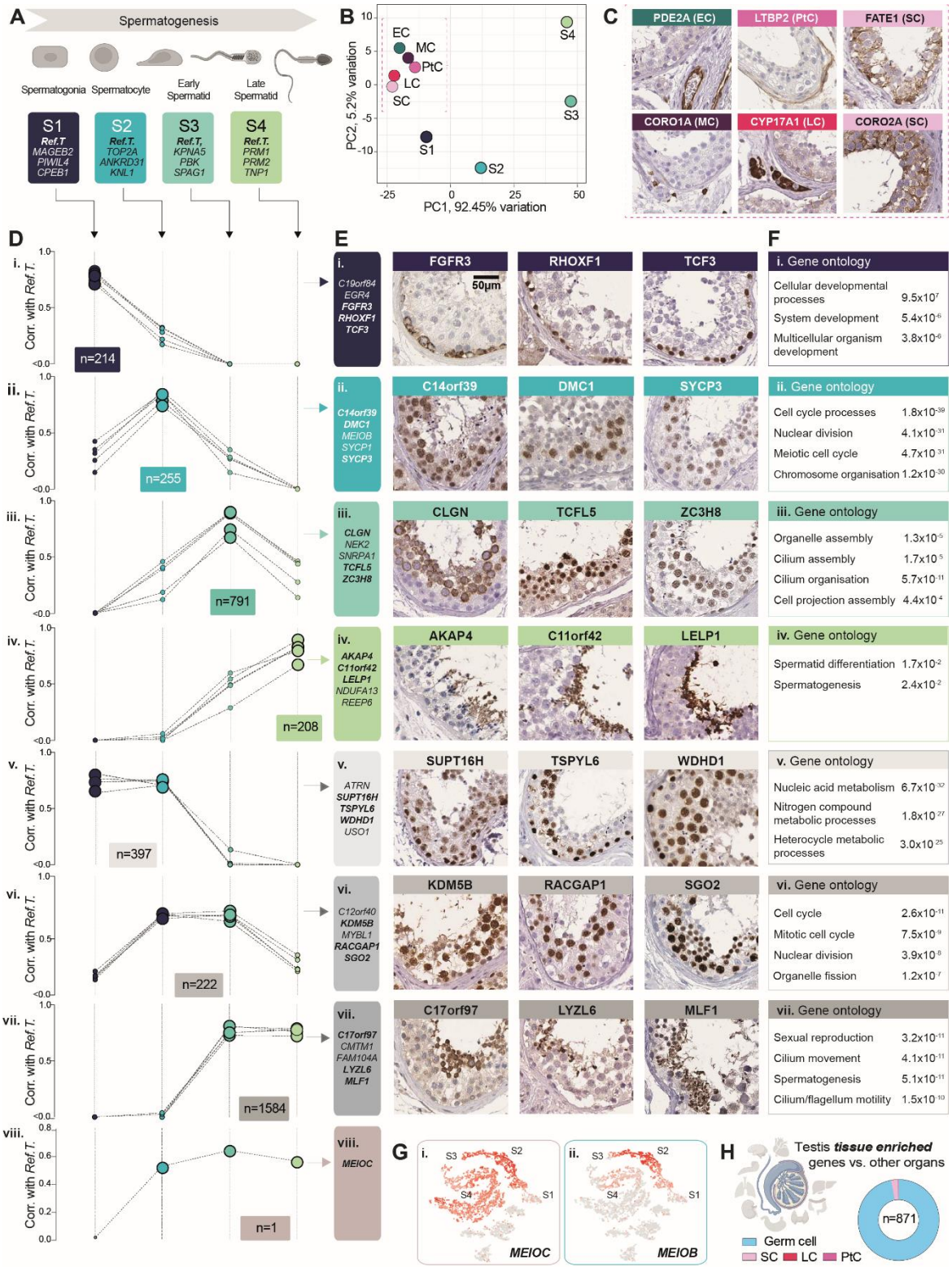


Figure 4. Analysis of pseudo temporal changes during spermatogenesis reveals stage-specific and common stage-shared gene enrichment signatures. (A) Cell types at the different stages of spermatogenesis were defined based on Ref.T. selected to broadly represent the developmental stage classifications spermatogonia [S1], spermatocytes [S2], early spermatids [S3] and late spermatids [S4]. **(B)** Principal component analysis of comparative correlation profiles of cell-enriched genes in S1, S2, S3, S4, sertoli cells (SC), Leydig cells (LC), peritubular cells (PtC), endothelial cells (EC) or macrophages (MC) vs. all Ref.T. panels. **(C)** Tissue profiling for proteins encoded by example genes

we predicted to be enriched in non-germ cell types. **(D)** Pseudo trajectories of gene enrichment signatures over time, showing enrichment values for each developmental stage, using **(E)** illustrative genes predicted to be: **(i)** S1, **(ii)** S2, **(iii)** S3, **(iv)** S4, **(v)** S1 and S2, **(vi)** S2 and S3, **(vii)** S2, S3 and S4, enriched, with corresponding tissue protein profiling. **(F)** Over-represented gene ontology terms and significance corrected FDR values for all genes classified as: **(i-iv)** highly enriched at a specific stage, or **(v-vii)** co-enriched at one or more stages of development. **(G)** UMAP plots showing gene expression profile in the Human Testis Atlas scRNAseq data (Guo et al., 2018) of: **(i)** the S2, S3 and S4 predicted enriched gene MEIOC and **(ii)** the S2 predicted enriched gene MEIOB. **(H)** Classification of testis tissue enriched genes that we predicted to be cell type enriched. See also Table S4, Figure S5 and S6.

Constituent cells of the skin hair root are the primary source of skin tissue enriched genes

The skin is one of the most complex tissue types in the human, with multiple layers and diverse constituent cell types. We profiled 18 different cell types in the skin, many of which are not represented in scRNAseq data in Tabula Sapiens [36] or the Human Protein Atlas (HPA) (www.proteinatlas.org/) [2,3], e.g. sebaceous gland cells, eccrine sweat gland cells, adipocytes, hair cortex and inner/outer root cells. Keratinocytes expressed the largest proportion of predicted cell type-enriched genes; 737 in the sub-granular layers (Figure 5 A.i) and 208 in the granular layer (Figure 5 A.ii). Analysis of the sub-granular keratinocyte layers at a higher cell type resolution was not possible, as a *Ref.T* panel with high, consistent, specificity for sub-population(s) of basal and suprabasal keratinocytes could not be identified. Similarly, when the dataset was analysed by WGNCA, most genes we predicted to be sub-granular keratinocyte enriched clustered in multiple groups on common clades (552/737 [75%]), the constituent groups of which contained a combination of genes considered basal e.g., *COL17A1* or suprabasal e.g., *DSG1* keratinocyte markers (Figure S7 A.i). In contrast, *Ref.T.* representing granular keratinocytes and the majority of genes predicted to be enriched in this cell type (181/208 [87%]), clustered in two groups on a single clade (Figure S7 A.ii). These results are consistent with keratinocyte development being associated with a shift in absolute gene expression levels, as opposed to a defined transition between highly distinct cell states that express many specific markers (prior to terminal differentiation in the granular layer).

For genes identified as cell type enriched, GO analysis revealed over-represented classes consistent with cell type annotation, e.g. for granular keratinocytes significant terms included *´epidermal cell differentiation´* (FDR 2.0 $x10^{-16}$) (Figure 5 A.ii) and for sebaceous gland cells *´lipid metabolic processes´* (FDR 2.3 $x10^{-32}$) (Figure 5 A.vi). Of the skin-specific cell types profiled, melanocytes had the fewest enriched genes (n=17) (Figure 5 A.v), including highly expressed genes with known cell type-specific functions e.g., *PMEL*, *DCT* (mean TPM in skin RNAseq 58.2 and 29.6, respectively). More lowly expressed melanocyte-enriched genes included *SLC24A5, CA14* and *SLC45A* (mean TPM in skin RNAseq 0.5, 1.9 and 5.7,

respectively). In skin scRNAseq data from Tabula Sapiens [36] (Figure S7 B.i) *SLC24A5* was predominantly expressed in a sub-population of cells in melanocyte annotated cluster (Figure S7 B.ii), but *CA14* and *SLC45A2* were not as clearly enriched in this cell type (Figure S7 B.iii and iv). However, our classifications of these genes as melanocyte-enriched are supported by other studies showing that *SLC45A2* has a role in deacidification of maturing melanosomes to support melanin synthesis [60] and that *CA14* is downregulated in vitiligo skin samples, compared to normal, along with other genes we classified as melanocyte enriched [61]. Furthermore, all three of these genes were clustered together with the melanocyte *Ref.T* when the dataset was analysed by WGNCA (Figure S7 A.iii). Thus, as we demonstrated for alpha and beta cells in the pancreas and germ cells in the testis, our analysis can identify cell-type enriched genes that are not always detectable as such by scRNAseq.

RNAseq data from unfractionated tissue was used to identify 188 genes as skin enriched vs. other tissues in the HPA tissue section [2], of which 151 were also identified as such in a similar analysis of tissues in GTEx [29], collated in the Harminozome database [39]. Of these, 96/151 [63%] were predicted to be cell type enriched in our analysis (Figure 5 B.i); most frequently in cells of the hair root (hair cortex or inner root sheath cell), granular keratinocytes or other keratinocytes. Other skin enriched genes were predicted to be enriched in melanocytes, sweat gland or sebaceous gland cells (Figure 5 B.i). Tissue profiling of proteins encoded by selected genes supported our classifications (Figure 5 B.ii). No skin enriched genes were predicted to be cell type enriched in endothelial cells, smooth muscle cells, fibroblasts, macrophages, or other immune cell types - consistent with their presence in multiple tissue beds, and thus lack of specificality to skin. Of those cells that were skin enriched, but not classified as cell type enriched in our analysis (Figure S7 C.i [rows lacking an enlarged circle]) most had co-enrichment profiles in multiple cell types in the hair root (Figure S7 C.ii). These genes included *PSORS1C2*, a member of the psoriasis susceptibility locus [62]. Enrichment of this gene in cell types of the hair follicle is supported by studies showing that ´near naked hairless´ mice, which have a spontaneous mutation preventing the development of a normal coat, have significantly reduced expression of *PSORS1C2* [63], together with others highlighted here e.g., *S100A3* and *KRTAP16-1* (Figure S7 C.ii) [63]. In depth skin tissue profiling showed expression of selected encoded proteins consistent with enrichment in the hair root (Figure S7 C.iii). Previously, keratinocytes, the majority cell type in the skin, have been annotated as the site of expression for the majority of skin enriched genes [3]. However, this is likely due to the lack of hair root cells in the scRNAseq data on which these classifications are based. Here, we show that a minority cell type represents the most common source of skin enriched genes.

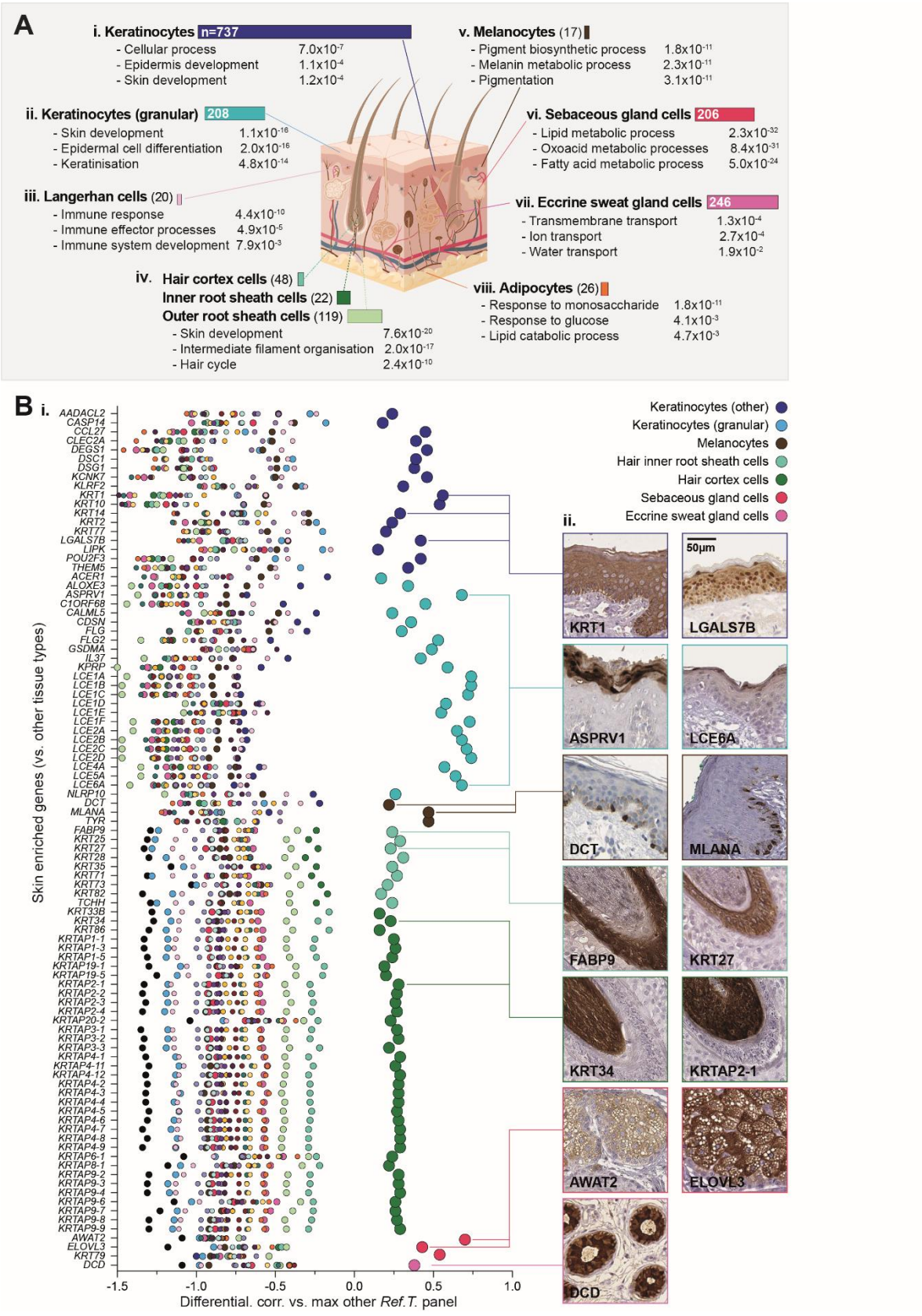


Figure 5. Constituent cells of the skin hair root are the primary source of skin tissue enriched genes. (A) Number of genes predicted to be cell type enriched, and corresponding over-represented gene ontology terms and significance corrected FDR values, for skin specialised cell types profiled: **(i)** keratinocytes, **(ii)** granular keratinocytes, **(iii)** Langerhans cells, **(iv)** hair cortex cells, inner and outer

root sheath cells, **(v)** melanocytes, **(vi)** sebaceous gland cells **(vii)** eccrine sweat gland cells and **(viii)** adipocytes. **(B) (i)** Genes enriched in skin vs. other organs, which were predicted to be cell type enriched in our analysis, were plotted to show the min differential values between the mean correlation coefficients with the Ref.T. panels for each cell type. Enlarged circles represent classification as predicted cell type enriched. **(ii)** Tissue profiling for proteins encoded by skin tissue enriched genes with predicted enrichment in the indicted cell type. See also Figure S7.

Cross-tissue analysis reveals similarities in cell type gene enrichment signatures

8011 genes were predicted to be cell type enriched in more than two tissue types. To explore the relationship between these cell type gene enrichment signatures, we performed a hypergeometric test to determine the degree of similarity between all cell types in all tissues. As cell type gene enrichment signatures are generated via a correlation-based analysis, independent of cell type absolute gene expression levels, such comparisons between tissue datasets can be made without correction for batch effects, the analysis platform used, or requirement for normalisation.

Organ-specific cell types can have common gene enrichment signature panels

Organ specific cell types (i.e. excluding those found in all or multiple organs, e.g., endothelial cells, fibroblasts and immune cell types) had gene enrichment signatures with: little or no similarity to other cell types e.g., hair inner root sheath cells and melanocytes (Figure 6 A.ii and iii), significant similarity to one other cell type, e.g., skeletal myocytes and cardiomyocytes (Figure 6 A.iv) or significant similarity with multiple other cell types, e.g., endocrine cells from several tissues; alpha and beta cells from the pancreas, enteroendocrine cells from the colon and stomach, and parafollicular cells from the thyroid (Figure 6 A.vi). We found a significant overlap between the enriched gene signatures of adipocytes (subcutaneous adipose, visceral adipose and breast), skin sebaceous gland cells, liver hepatocytes, and kidney proximal tubular cells (Figure 6 A.vii). GO analysis of the 41 genes predicted to be enriched in at least three of the aforementioned cell types (Figure 6 B.i [green box]) (Table S5, Tab 1, Table A) revealed significant terms all related to metabolic processes, including *´carboxylic acid metabolic processes´* (FDR 8.8 $x10^{-26}$) and *´organic acid metabolic processes´* (FDR 9.4 $x10^{-26}$) (Table S5, Tab 1, Table B) (Figure 6 B.ii). 22 of these 41 genes were also predicted to be enriched in cardiomyocytes, another highly metabolically active cell type with a significant overlap in gene enrichment signature with both adipocytes and hepatocytes (Figure 6 A and Table S5, Tab 1, Table A). Illustrative protein profiling showed selective expression of ACO1 and HADH in adipocytes in adipose tissue, sebaceous gland cells in skin, hepatocytes in liver and proximal tubular cells in kidney (Figure 6 B.iii). The enrichment of such genes in many highly metabolically active cells is consistent with a common shared function across tissue types. In contrast, cell type enriched genes classified as such in only one tissue are likely key

for highly specialised cell functions, e.g., complement and coagulation factor genes were predicted to be enriched only in hepatocytes (including *C4B*, *C8A*, *C9*, *CFHR1/2/4/5* and others) and specific solute transporters only in proximal tubular cells (e.g., *SLC13A1*, *SLC22A13*, *SLC22A6*, *SLC22A8*). Tissue profiling for proteins encoded by example genes predicted to be enriched in only one of these four cell types; adipocytes (*PRKAR2B*), sebaceous gland cells (*TMEM97*), hepatocytes (*OTC*) or proximal tubular cells (*TMEM174*) showed positive staining in only the respective cell types (Figure 6 B.iv). In contrast to ACO1 and HADH, which were expressed mean TMP >10 in the RNAseq datasets analysed (Figure 6 B.v), expression values of these example genes were highest in the corresponding tissue, with low or no expression in the others (Figure 6 B.vi).

Our analysis also revealed a significant overlap between the gene enrichment signatures of early and late spermatids in the testis and respiratory ciliated cells in the lung (Figure 6 A.v and C.i). GO term analysis of these 441 shared genes (Table S5, Tab 2, column A-B) revealed the most significant terms were related to cilia (Figure 6 C.ii), which are important for both clearance of fluid from the airways and movement of the sperm flagellum, including *´cilium organisation´* (FDR 3.6 x10^{-69}) and *´cilium movement´* (FDR 9.5 x10^{-64}) (Table S5, Tab 2, Table 1). The top 50 genes predicted to be most highly enriched in both early and late spermatids and RCC (Figure 6 C.iii) had variable absolute expression in the respective tissues. *LMNTD1* and *MROH9* had very low expression in the lung RNAseq (mean TMP 0.42 and 0.68, respectively) (Figure 6 C.iii) and scRNAseq data from human lung [36] revealed highly specific, but variable expression (or detection) of these genes in RCC (Figure S8 A.ii and iii). Predicted expression in S3 and S4 cells in testis was also supported by scRNAseq from the Human Testis Atlas [59] (Figure S8 B.ii and iii). Despite the highly specific enrichment profiles of *LMNTD1* and *MROH9*; neither were predicted to be enriched in any other cell type across all tissues analysed (Figure 6 C.iii), there are no existing studies of these genes in this context. Some other genes with highly predicted enrichment in early and late spermatids and RCC were also predicted to be enriched in several other cell types e.g., *PACRG* (Figure 6 C.iii), which is well studied in the context of motile cilia, particularly in sperm [64], but has also been proposed to have other roles, such as in primary cilia [65] and even inflammatory pathway signalling [66]; perhaps explaining its more widespread enrichment profiles in our analysis. Tissue profiling for proteins encoded by example genes enriched in both RCC and early and late spermatids (Figure 6 C.iv), or RCC only (Figure 6 C.v) showed expression consistent with our predictions. GO analysis of genes predicted to be highly enriched in spermatids, but not RCC, revealed the most significant terms were unrelated to cilia formation, including *´spermatogenesis´* (FDR 1. x10^{-25}), *´multicellular organism reproduction´* (FDR 1.5 x10^{-19}), *´spermatid development´* (FDR 6.7 x10^{-14}) and *´fertilisation´* (FDR 1.4 x10^{-10}) (Table S5, Tab 3, Column A-B and Table 1);

reflecting an enrichment for genes with highly specialised function within the testis only, e.g., *CALR3*, *LELP1* and *SMCP* (Figure 6 C.vi).

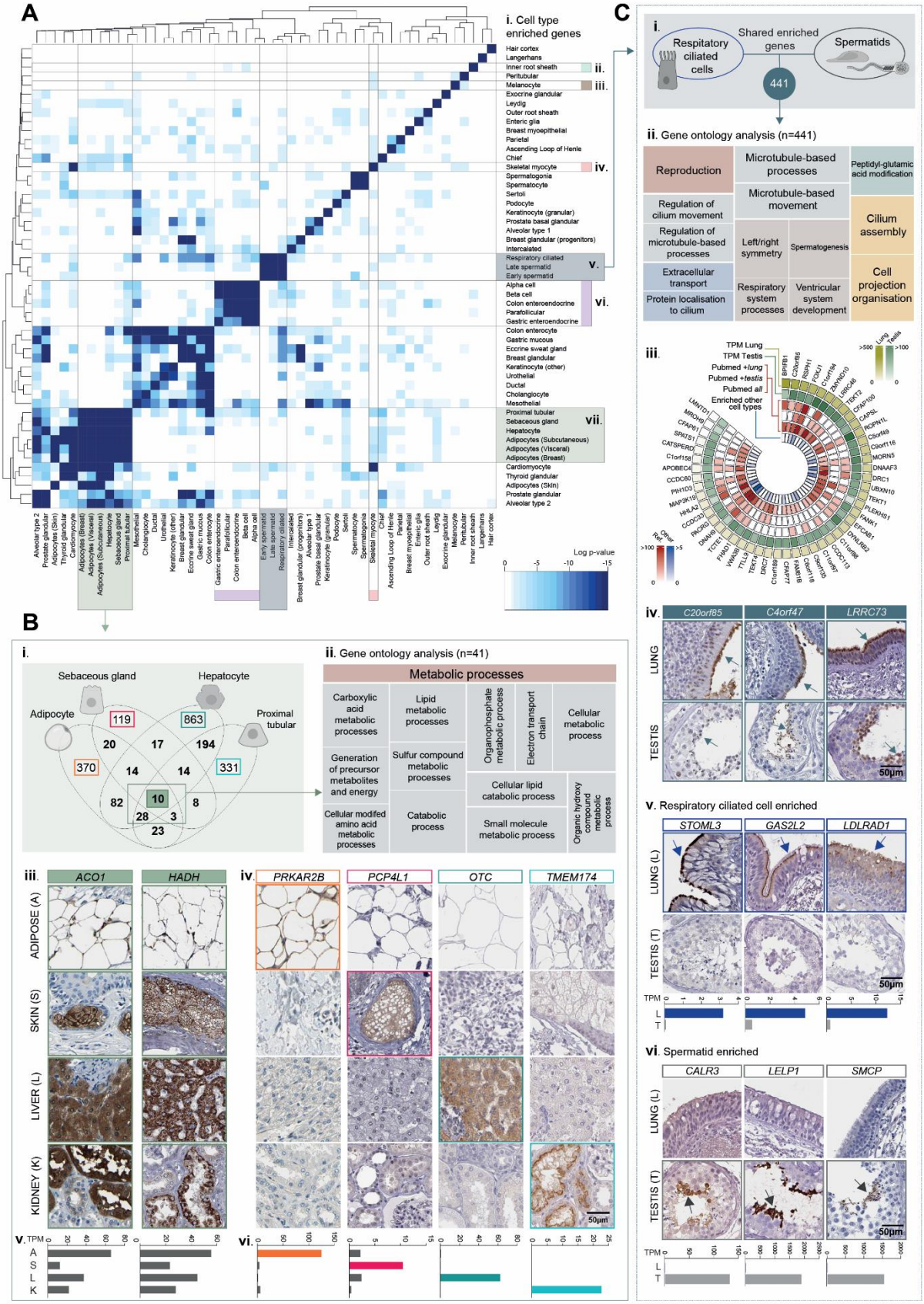


Figure 6. Organ specific cell types can have gene enrichment signature similarities. (A) Heatmap showing significance p-values for similarity scores between predicted cell type enriched genes, calculated using a hypergeometric test for: **(i)** all organ-specific cell type enriched gene signatures, **(ii)** skin inner root sheath hair cells, **(iii)** skin melanocytes, **(iv)** skeletal myocytes, **(v)** lung respiratory ciliated cells and testis spermatids, **(vi)** pancreatic alpha and beta cells, colon enteroendocrine cells, thyroid parafollicular and stomach gastric enteroendocrine cells and **(vii)** kidney proximal tubular cells, sebaceous gland cells, hepatocytes, and adipocytes in subcutaneous adipose tissue, visceral adipose tissue and breast. **(B) (i)** Number of individual or common enriched genes for adipocytes (in at least 2/3 tissues profiled) kidney proximal tubular cells, sebaceous gland cells and hepatocytes. **(ii)** Over-represented gene ontology terms among the 41 genes featuring in the gene enrichment signature of at least 3/4 cell types, displayed in tree map format, generated using REVIGO (areas proportional to Log10 significance values). Tissue profiling for proteins encoded by genes that were: **(iii)** part of the shared gene enrichment signature [ACO1, HADH] or **(iv)** classified as cell type enriched only in one cell type [PRKAR2B, TMEM97, OTC, TMEM174] and corresponding mean TMP expression in the corresponding tissue RNAseq datasets (v) and **(vi),** respectively. **(C) (i)** Number of genes that featured in gene enrichment signatures for respiratory ciliated cells, early and late spermatids and **(ii)** the over-represented gene ontology terms among these shared enriched genes. **(iii)** Circular plot showing up to the top 50 most enriched genes in respiratory ciliated cells and spermatids, displaying the mean TMP values in the lung and testis RNAseq datasets, the number of mentions in Pubmed of gene and corresponding tissue (´Pubmed + lung/testis´) or the gene alone (´Pubmed all´), and the number of other cell types in which the gene was also predicted to be enriched (´enriched other cell type´). Tissue profiling for proteins encoded by genes with predicted enrichment in: **(iv)** both respiratory ciliated cells and spermatids, **(v)** respiratory ciliated cells only and **(vi)** spermatids only and corresponding mean TMP values in the tissue RNAseq. See also Table S5 and Figure S8.

Core cell types have common gene enrichment signature panels across tissues

Eight cell types were profiled in all, or most, tissue types (termed “core cell types”); endothelial cells [n=15 tissues], smooth muscle cells [n=10], fibroblasts (including hepatic stellate cells (HSC) in the liver, and adipose progenitor cells (APC) in adipose tissue) [n=14], macrophages (including Kupffer cells in the liver) [n=15], neutrophils [n=8], mast cells [n=5], T-cells [n=13] and plasma cells [n=14] (Figure 7 A.i). Gene enrichment signatures of the same core cell type in different tissues had high similarity, with little, or no, crossover between different cell types (Figure 7A). Notable exceptions included hepatic stellate cells (HSC) and fibroblasts in liver and kidney, respectively, which had some commonality with smooth muscle cell gene enrichment signatures in other tissues (Figure 7 A.ii), in line with reports that liver HSC can have contractile properties [67] and potentially reflecting the presence of a kidney myofibroblast-like population, and lung neutrophils, which had some similarity to macrophages in several other tissues (Figure 7 A.iii). Enrichment signatures of core cell types had little or no cross over with those of organ specific cell types (Figure S8 C.i), except for lung macrophages, which had a significant similarity with the cell type group we previously identified as having shared gene enrichment signatures related to metabolic processes (Figure 6 B), including adipocytes, hepatocytes, proximal tubular cells (Figure S8 C.ii). One could speculate that this indicates macrophages in the lung have specific metabolic characteristics, in keeping with recent studies

indicating that their metabolic responses to infectious pathogens or other insults may be distinct from other macrophage subtypes [68].

Endothelial cells had strong gene enrichment signature similarities across tissues (Figure 7 A), with the exception of liver sinusoidal endothelial cells (LSEC), where over half of the enriched genes (19/34 [56%]) were not enriched in endothelial cells in any other tissue, consistent with their unique structural and phenotypic features, and highly specialised function [69]. Despite this, overall, they did have greatest similarity with vascular endothelial cells vs. any other core cell type (Figure 7 A.iv). Tissue profiling for proteins encoded by *CLEC4G* (Figure 7 B.i) and *CD36* (Figure 7 B.ii) showed expression consistent with our predictions of LSEC enrichment only, or vascular endothelial and LSEC enrichment, respectively.

To define key components of the gene enrichment signature for each core cell type, we identified genes predicted to enriched in at least half of the tissues profiled (Table S6), e.g., in at least 8/15 tissues for EC and MC (Figure 7 C-Gi and ii and Table S6, Tab 1). To assess existing reports for each gene in each given cell type, we used PubMed to search for the number of studies citing both gene name and cell type together (Figure 7C-G.iii), or gene name alone (Figure 7 C-G.iv). Many were well characterised genes on which a plethora of studies have been performed, e.g., *CD3E* and *CD2* in T-cells (Figure 7 C.iii and iv), others were poorly studied in the cell type context. For example, *SHANK3,* predicted to be endothelial cell enriched in 10 tissues (Figure 7 D), has been researched predominantly in the context of neurons and autism [70]. *TSPAN7* (Figure 7D)*,* predicted to be endothelial cell enriched in 8 tissues, has only been identified in endothelial cells in the context of tumour associated vasculature and metastasis [71]. There is little information in the literature about the function of *LRRN4CL* (Figure 7 E), a gene we predicted to be fibroblast enriched in 7 tissues, except for its elevated expression in skin melanoma metastases and breast cancer samples [72,73]. In contrast, *MFAP4* (Figure 7 E) is a well-known gene in this cell context [74]. *TBXAS1* was identified as a macrophage core enriched gene, and its enzymatic product, thromboxane A2, is linked to vasoconstriction and platelet aggregation, with links to innate immunity [75], but little knowledge exists in the macrophage context. Tissue profiling for proteins encoded by predicted endothelial enriched genes *SHANK3* (Figure 7 D.v) and *TSPAN7* (Figure 7 D.vi), fibroblast enriched genes *LRRN4CL* (Figure 7 E.v) and *MFAP4* (Figure 7 E.vi) and the macrophage enriched gene *TBXAS1* (Figure 7 G.vi), revealed selective expression consistent with our predictions.

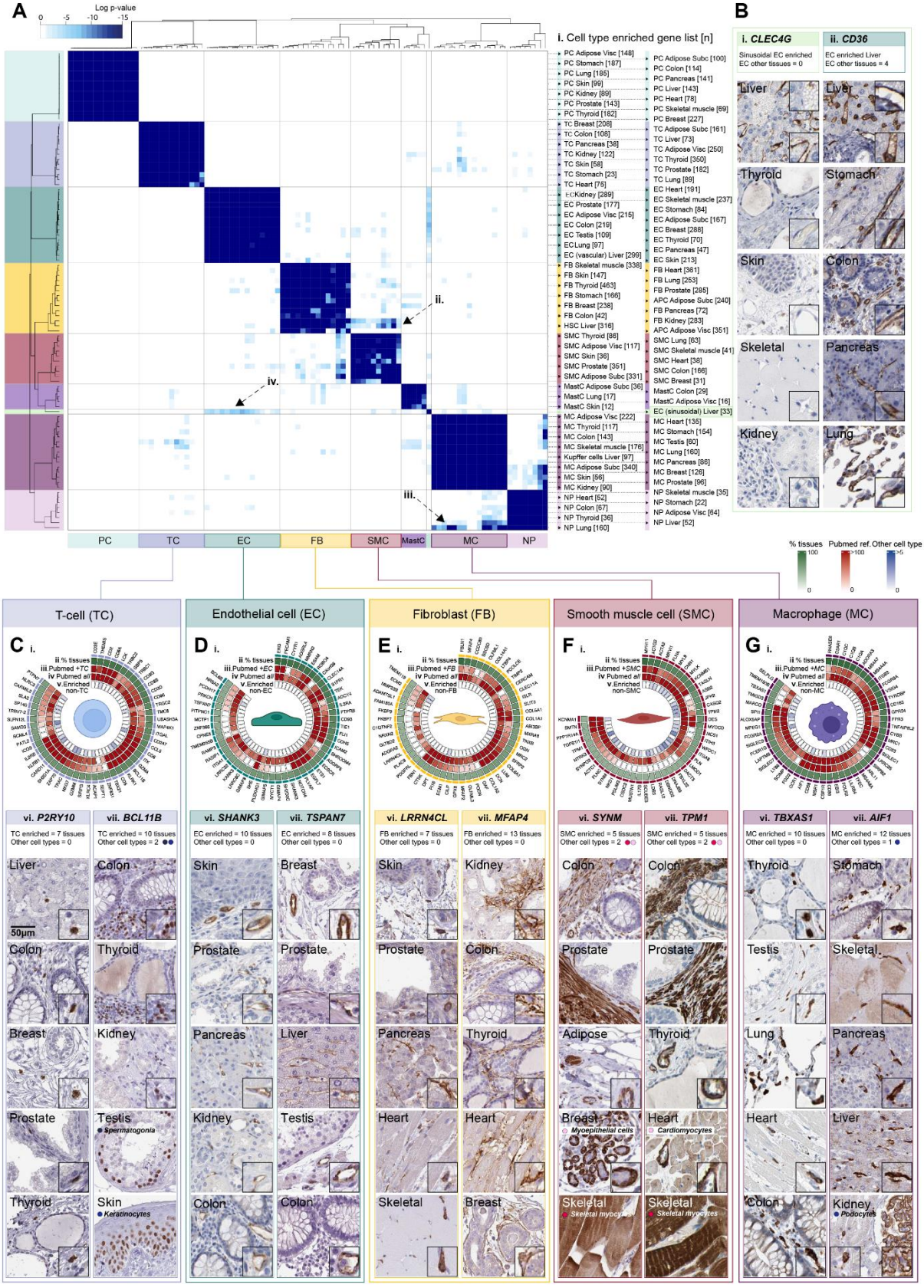


Figure 7. Core cell types share gene enrichment signatures across organs. (A) Heatmap showing significance p-values for similarity scores in cell type gene enrichment signatures, calculated using a hypergeometric test, between **(i)** plasma cells (PC), T-cells (TC), endothelial cells (EC), fibroblasts (FB),

smooth muscle cells (SMC), mast cells (MastC), macrophages (MC) and neutrophils (NP) in different tissues. **(B)** Tissue profiling for proteins encoded by **(i)** the sinusoidal EC enriched gene CLEC4G and **(ii)** the vascular and sinusoidal EC enriched gene CD36, in different tissue beds. Circular plots showing up to the top 50 genes most frequently predicted as enriched in **(C)** TC, **(D)** EC, **(E)** FB, **(F)** SMC and **(G)** MC in different organs, displaying **(i)** the percentage of tissues in which the gene was classified as enriched in the given cell type (´% tissues´), the number of mentions in Pubmed of **(ii)** gene and corresponding core cell type (´Pubmed + cell type´) together or **(iii)** the gene alone (´Pubmed all´), and **(iv)** the number of other cell types (including non-core cell types) in which the gene was also predicted as enriched (´enriched non-cell type´). **(v-vi)** Tissue profiling of proteins encoded by selected genes predicted to be core cell type enriched. See also Table S6 and Figure S8.

Whereas most core endothelial and fibroblast enriched genes were not predicted to be enriched in any other cell types in our analysis (Figure 7 D-E.v), several T-cell (Figure 7 C.v) smooth muscle cell (Figure 7 F.v), and macrophage (Figure G.v) enriched genes where predicted to be enriched in an additional cell type(s). *BCL11B*, a gene we predicted to be T-cell enriched gene in 10 tissues (Figure 7 C) with a known role in T-cell development [76], was also predicted to be enriched in skin keratinocytes, consistent with its role in dermal development in mice [77], and unexpectedly, in the absence of any existing reports, in spermatogonia in the testis; expression profiles we verified by protein profiling (Figure 7 C.vii). *AIF1*, predicted to be macrophage enriched in 12 tissues (Figure 7 G.i and viii), consistent with its known expression in this cell type [78], was also classified as kidney podocyte enriched; previously reported in only a single study [79], a prediction we again verified with tissue protein profiling (Figure 7 G.vi). Of all the core cell types, smooth muscle cell enriched genes were most likely to have predicted enrichment in another cell type, most frequently cardiomyocytes, skeletal myocytes or breast myoepithelial cells, e.g., *SYNM* (Figure 7 F.v) and *TPM1* (Figure 7 F.vi).

DISCUSSION

Here we present a tissue-centric, cell type gene enrichment atlas, generated from the analysis of hundreds of biological replicates. Although it is frequently stated that cell-type gene expression profiles cannot be extracted from bulk RNAseq, e.g., [80,81], here we have identified cell type enriched or co-enriched genes, and charted temporal transcriptome changes underlying cell type differentiation. We made comparisons between cell type enrichment signatures across tissues, without the requirement for normalisation or batch effect adjustments, a significant issue when handling scRNAseq datasets, for which currently no universal solution [82,83]. Our analysis included cell types that are difficult to extract from tissue, e.g., adipocytes, and those that are sensitive to processing, e.g., kidney podocytes; issues that can hinder analysis [8,10], but are circumvented here as cell removal from tissue was not required. We identify lowly expressed transcripts as cell type enriched, many of which can be detected only in a small minority of cells annotated as a given type by scRNAseq, possibly due to limited read depth and high number of drop-out events [18]. Transcript level alone is not sufficient to predict protein levels [84] and so potential function of proteins encoded by such genes may have been overlooked.

Our study is the only cell type gene enrichment atlas generated independently of scRNAseq. Comparison of scRNAseq datasets generated from analysis of the same tissue type can reveal surprisingly low agreement between studies [22,23], possibly due to the low number of samples typically analysed, and the associated lack of biological varience. For top cell type enriched genes in adipose tissue, agreement between data generated using our analysis method and several scRNAseq studies was equivalent or greater than between the scRNAseq studies themselves [28]. This could reflect the large sample set analysed and the associated biological variance represented. Our method also has scope for well-powered comparisons of cell type enrichment profiles between healthy and disease states, sexes [28], ages, developmental stages, or metabolic states, using existing RNAseq resources for which phenotypic data is available, such as GTEx [33] or TCGA (https://www.cancer.gov/tcga) .

Various deconvolution algorithms have been developed to determine proportions of constituent cell types in bulk RNAseq, e.g., CIBERSORT and others [85-87]. Such analyses typically depend on input expression matrices of cell type reference genes, generated from transcriptome analysis of isolated cells or cell types. The accuracy of input matrices can be affected by various factors, such as technical artifacts due to cell extraction and processing, the presence of contaminating cell types, and limited input data availability for some cell types. Cross checking input matrices against our dataset could optimise such analysis, by identifying the most likely highly enriched genes *in vivo*.

Limitations of the study

There are limitations to our study. In some tissues, we do not profile specific cell subtypes, e.g., basal and suprabasal keratinocytes in the skin, which were handled as one cell type in our analysis. In such cases, we failed to identify genes that fulfilled the criteria for use as input *Ref.T.*. In keeping with our observations, scRNAseq analysis of skin showed that genes considered to be basal keratinocyte markers e.g., *COL17A1* and *KRT5*, were indeed most highly expressed in this cell type, but were also co-enriched within the tissue in suprabasal keratinocytes [3]. Thus, such cell subtype definitions are likely primarily governed by variation in absolute mRNA expression levels, rather that the presence or absence of a large number of uniquely enriched genes.

As our analysis end point is a gene enrichment score, we do not provide information on absolute mRNA expression profiles on a cell type basis, such as that generated by scRNAseq analysis.

As the prediction of cell type enriched genes is dependent on known input *Ref.T.*, we cannot identify novel cell (sub)types for which *Ref.T.* have not yet been described.

We analysed samples from a total of 933 individuals from the GTEx portal [33], with diverse health status, whose ages skewed older (ages 20-29: 8.5%, 30-39: 8.1%, 40-49: 15.6%, 50-59: 31.9%, 60-69: 32.4%, 70-79: 3.4%). Thus, the input dataset represents a limited age demographic, and a health status that may not represent the general population.

Expression of certain genes are strongly modified by environmental (e.g., eating, exercise, inflammation etc.), or genetic factors [88]. Such genes may therefore lack correlation with the constitutively expressed *Ref.T.* selected to represent the cell type in which they are predominantly transcribed, and thus could be considered a type of false negative in our analysis. One such example is *SELE*, an endothelial cell specific gene that is highly upregulated during inflammation and expressed at very low levels, if at all, in resting state [89]. *SELE*, despite its highly endothelial cell restricted expression profile, is not classified as EC enriched in our analysis, due to the variable nature of its expression.

We used relatively high thresholds for classification of cell type enriched genes. It is likely that some cell type enriched genes may be false negatives in our analysis, as they fall just below the thresholds required for classification as such. For example, the gene *KANK3* is classified as endothelial cell enriched in 9 tissue types, in the remaining 6 the highest enrichment score is also in endothelial cells, vs. all other types profiled, although it did not reach classification

threshold. Thus, our classifications are intended only as a guide, and the reader should consider the data on a transcript-by-transcript basis.

All data generated in this study is available on the Human Protein Atlas in the ´Tissue Cell Type´ section (www.proteinatlas.org/humanproteome/tissue+cell+type), and can be used alongside data generated from scRNAseq in the ´Single Cell Type Section´ [3], and the antibody based tissue protein profiling in the ´Tissue Section´ [2].

METHODS AND RESOURCES

LEAD CONTACT AND MATERIALS AVAILABILITY

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

This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bulk RNAseq data analysed in this study was obtained from the Genotype-Tissue Expression (GTEx) Project (gtexportal.org) V8 [29] on 2021/04/26 (dbGaP Accession phs000424.v8.p2). Protein coding genes were categorised according to Biotype definitions in ENSEMBL release 102 [90] inclusive of those defined as “protein_coding”, “IG_C_gene”, “IG_D_gene”, “IG_J_gene”, “IG_V_gene”, “TR_C_gene”,“TR_D_gene” ,“TR_J_gene” and “TR_V_gene”. All other categorisations were classified as “non-protein coding” and were excluded from the analysis. Human tissue protein profiling was performed in house as part of the Human Protein Atlas project [2,91,92] (www.proteinatlas.org). Normal 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

Sample inclusion

All samples in each GTEx tissue type dataset were included in the analyses, with the exception of: (i) ***Skin-not Sun Exposed (suprapubic)***: *Ref.T.* selected to represent hair root cells were absent or very lowly expressed in a large number of samples, presumably due to the lack of such structures in the selected region of tissue analysed. Thus, only samples with mean TPM >0.1 for hair follicle expressed transcripts trichohyalin (*TCHH*), keratins 25 (*KRT25*) and 71 (*KRT71*) were included for in the analysis (n=177). (ii) ***Breast – Mammary Tissue***: The GTEx breast dataset contains samples from both male and female donors, we analysed those from only from females. In both cases, sample IDs included in the analysis can be found in Table S2, Tab ´Sample IDs´.

QUANTIFICATION AND STATISTICAL ANALYSIS

Reference transcript-based correlation analysis

This method was based on that we previously developed [26-28]. Pairwise Spearman correlation coefficients between reference transcripts (*Ref.T.*), selected as proxy markers for each cell type (see Table S1, Tab 1, Table A-O), and all other transcripts were calculated in R using the *corr.test* function from the *psych* package (v 1.8.4). False Discovery Rate (FDR) adjusted p-values (using Bonferroni correction) <0.0001 were considered significant. Genes were predicted to be cell type enriched if they fulfilled the criteria as described in the results section. In cases where a given cell type was represented by more than one *Ref.T* panel, or they could be considered related sub-cell types, the minimum differential score required vs. other *Ref.T.* panels was calculated excluding each the other (i.e., genes that correlated highly with both *Ref.T.* panels representing the same (sub)cell type were *not* excluded from classification as cell type enriched, but included in both – see Table S1, Tab 2).

Weighted correlation network (WGCNA) analysis

The R package WGCNA [38] was used to perform co-expression network analysis for gene clustering, on log2 expression TPM values. The analysis was performed according to recommended conditions in the WGCNA manual. Non-protein coding transcripts and transcripts with too many missing values were excluded using the goodSamplesGenes() function.

Gene Ontology

The Gene Ontology Consortium [41] and PANTHER classification resource [93,94] were used to identify over represented terms in gene lists from the GO ontology (release date 2022-07-01) or reactome (release date 2021-10-01) databases. Plots of GO terms were created using the Clusterprofiler package in R [95] or REVIGO [56], as specified.

Additional datasets and analysis

Single cell RNAseq data from Tabula Sapiens [36] was downloaded and UMAP plots created using the Seurat package in R [96]. Human testis scRNAseq data was sourced from the human testis atlas [59]. Tissue enriched genes were downloaded from the Human Protein Atlas (HPA) tissue atlas [2] or GTEx database [29], as collated in the Harminozome database [39].

Tissue Profiling: Human tissue sections

Immunohistochemistry (IHC) stained tissue sections were stained, as previously described [2,91]. Briefly, formalin fixed and paraffin embedded tissue samples were sectioned, de-paraffinised 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 minutes using Diaminobenzidine (Lab Vision) as the chromogen. Slides were counterstained in Mayers hematoxylin (Histolab) and scanned using Scanscope XT (Aperio). Primary antibodies used for IHC staining are listed in Table S7.

Other visualisation and analysis tools

Graphs and plots were made using Graphpad prism or the ggplot2 package in R [97], unless otherwise specified. Circular plots were constructed using the R package *circlize* [98] and pubmed data was extracted using the easyPubMed package in R (https://CRAN.R-project.org/package=easyPubMed). Some figure illustrations were created using BioRender.com.

DATA AVAILABILITY

Data for all protein coding genes and antibody-based protein profiling is provided on the Human Protein Atlas (Tissue Cell Type section) (www.proteinatlas.org/humanproteome/tissue+cell+type). This article also includes all individual tissue datasets generated (Table S2) and cell type enrichment categorisations (Table S1, Tab 4).

AUTHOR CONTRIBUTIONS

Conceptualisation: LMB. Methodology: PD, LMB. Formal analysis: PD, SO, ES, MNT, MJI, LMB. Investigation: PD, SO, ES, MNT, MJI, FP, CL, LMB. Resources: FP, CL, JO, MU, LMB. Writing – Original Draft: PD, LMB. Writing – Review & Editing: All, Visualisation: PD, LMB. Supervision: LMB, PD. Funding Acquisition: JO, FP, CL, MU, LMB.

ACKNOWLEDGEMENTS

This work was supported by funding granted to LMB from Hjärt Lungfonden (20170759, 20170537) and the Swedish Research Council (2019-01493). Main funding for the Human Protein Atlas was provided from the Knut and Alice Wallenberg Foundation (WCPR) and the Erling Persson Foundation (KCAP). We acknowledge the staff of the Human Protein Atlas program and the Science for Life Laboratory (SciLifeLab) for their valuable contributions. **Data usage:** We used data from the Genotype-Tissue Expression (GTEx) Project (gtexportal.org) [29], supported by the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL TABLE LEGENDS

Table S1. Reference transcript selection and analysis summary

[Tab 1]: Correlation coefficient values between selected *Ref.T.* in each tissue. [Tab 2]: Cell subtypes represented by different *Ref.T*. panels within a single tissue and corresponding annotations in the Tabula Sapiens and HPA databases. [Tab 3]: Analysis criteria and totals for very high, high and moderately enriched genes within each cell type. [Tab 4]: Cell type enrichment predictions for all protein coding genes.

Table S2. Sample IDs and tissue-by-tissue data

[Tab: Sample IDs]: Analysed sample IDs (GTEx). [Other tabs]: Details for each tissue type (see key).

Table S3. Gene ontology (GO) terms in alpha and beta cells of pancreas

[Tab 1]: GO biological process, reactome, and cellular component analysis for genes predicted to be co-enriched in alpha and beta cells of the pancreas. [Tab 2]: Synapse-linked genes with predicted co-enrichment in alpha and beta cells of the pancreas.

Table S4. Values and gene ontology (GO) analysis of germ cell enriched genes

[Tab 1]: Genes predicted to be enriched in germ cells of the testis (see key). [Tab 2]: GO biological process and reactome analysis of germ cell enriched genes. [Tab 3]: GO biological process analysis for germ cell subtype predicted enriched genes.

Table S5. GO analysis of genes enriched in multiple cell types

[Tab 1]: Table A: Genes predicted to be enriched in 3 or more of: adipocytes, sebaceous gland cells, hepatocytes, and proximal tubular cells. Table B: GO Biological Process analysis for genes in Table A. [Tab 2]: Table A: Genes predicted to be enriched in respiratory ciliated cells of the lung and S3 and/or S4 cells (early or late spermatids) of the testis. Table B: Enriched GO biological process anaysis for genes listed in Table A. [Tab 3]: Table A: Genes predicted to be enriched in S3 and/or S4 cells (early/late Spermatids) of the testis, but not in respiratory ciliated cells of the lung. Table B: GO biological process analysis for genes listed in Table A.

Table S6. Core cell type predicted enriched genes

Genes predicted to be enriched in the same cell type in at least half of the tissues where profiled.

Table S7. Primary antibodies

IDs for all primary antibodies used to stain all immunohistochemistry images used in this study.

REFERENCES

1. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The Human Cell Atlas. Elife *6*. 10.7554/eLife.27041.
2. 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. 10.1126/science.1260419.
3. 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*. 10.1126/sciadv.abh2169.
4. Iglesias, M.J., Kruse, L.D., Sanchez-Rivera, L., Enge, L., Dusart, P., Hong, M.G., Uhlen, M., Renne, T., Schwenk, J.M., Bergstrom, G., et al. (2021). Identification of Endothelial Proteins in Plasma Associated With Cardiovascular Risk Factors. Arterioscler Thromb Vasc Biol *41*, 2990-3004. 10.1161/ATVBAHA.121.316779.
5. Cano-Gamez, E., and Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. Front Genet *11*, 424. 10.3389/fgene.2020.00424.
6. Wang, R.D.-Y., L.; Jiang, Y. (2021). EPIC: inferring relevant cell types for complex traits by integrating genome-wide association studies and single-cell RNA sequencing. www.biorxiv.org. https://doi.org/10.1101/2021.06.09.447805.
7. van den Brink, S.C., Sage, F., Vertesy, A., Spanjaard, B., Peterson-Maduro, J., Baron, C.S., Robin, C., and van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat Methods *14*, 935-936. 10.1038/nmeth.4437.
8. Denisenko, E., Guo, B.B., Jones, M., Hou, R., de Kock, L., Lassmann, T., Poppe, D., Clement, O., Simmons, R.K., Lister, R., and Forrest, A.R.R. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol *21*, 130. 10.1186/s13059-020-02048-6.
9. Massoni-Badosa, R., Iacono, G., Moutinho, C., Kulis, M., Palau, N., Marchese, D., Rodriguez-Ubreva, J., Ballestar, E., Rodriguez-Esteban, G., Marsal, S., et al. (2020). Sampling time-dependent artifacts in single-cell genomics studies. Genome Biol *21*, 112. 10.1186/s13059-020-02032-0.
10. Rondini, E.A., and Granneman, J.G. (2020). Single cell approaches to address adipose tissue stromal cell heterogeneity. Biochem J *477*, 583-600. 10.1042/BCJ20190467.
11. Viswanadha, S., and Londos, C. (2006). Optimized conditions for measuring lipolysis in murine primary adipocytes. J Lipid Res *47*, 1859-1864. 10.1194/jlr.D600005-JLR200.
12. Quake, T.T.S.C.S.R. (2021). The Tabula Sapiens: a single cell transcriptomic atlas of multiple organs from individual human donors. www.biorxiv.org *https://doi.org/10.1101/2021.07.19.452956*.
13. Tabula Muris, C., Overall, c., Logistical, c., Organ, c., processing, Library, p., sequencing, Computational data, a., Cell type, a., Writing, g., et al. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367-372. 10.1038/s41586-018-0590-4.
14. Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., et al. (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat Methods *14*, 955-958. 10.1038/nmeth.4407.

15. 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. 10.1016/j.celrep.2020.108189.
16. Haque, A., Engel, J., Teichmann, S.A., and Lönnberg, T. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med *9*, 75. 10.1186/s13073-017-0467-4.
17. Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015). The technology and biology of single-cell RNA sequencing. Mol Cell *58*, 610-620. 10.1016/j.molcel.2015.04.005.
18. Hicks, S.C., Townes, F.W., Teng, M., and Irizarry, R.A. (2018). Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics *19*, 562-578. 10.1093/biostatistics/kxx053.
19. Zheng, Y., Zhong, Y., Hu, J., and Shang, X. (2021). SCC: an accurate imputation method for scRNA-seq dropouts based on a mixture model. BMC Bioinformatics *22*, 5. 10.1186/s12859-020-03878-8.
20. Chen, G., Ning, B., and Shi, T. (2019). Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front Genet *10*, 317. 10.3389/fgene.2019.00317.
21. Hou, W., Ji, Z., Ji, H., and Hicks, S.C. (2020). A systematic evaluation of single-cell RNA-sequencing imputation methods. Genome Biol *21*, 218. 10.1186/s13059-020-02132-x.
22. Jiang, R., Sun, T., Song, D., and Li, J.J. (2022). Statistics or biology: the zero-inflation controversy about scRNA-seq data. Genome Biol *23*, 31. 10.1186/s13059-022-02601-5.
23. Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat Commun *12*, 5692. 10.1038/s41467-021-25960-2.
24. Denninger, J.K., Walker, L.A., Chen, X., Turkoglu, A., Pan, A., Tapp, Z., Senthilvelan, S., Rindani, R., Kokiko-Cochran, O.N., Bundschuh, R., et al. (2022). Robust Transcriptional Profiling and Identification of Differentially Expressed Genes With Low Input RNA Sequencing of Adult Hippocampal Neural Stem and Progenitor Populations. Front Mol Neurosci *15*, 810722. 10.3389/fnmol.2022.810722.
25. Trostle, A.J., Wang, J., Li, L., Wan, Y., and Liu, Z. (2022). Most High Throughput Expression Data Sets Are Underpowered. biorxiv.org. 10.1101/2022.08.03.502688
26. Butler, L.M., Hallstrom, B.M., Fagerberg, L., Ponten, F., Uhlen, M., Renne, T., and Odeberg, J. (2016). Analysis of Body-wide Unfractionated Tissue Data to Identify a Core Human Endothelial Transcriptome. Cell Syst *3*, 287-301 e283. 10.1016/j.cels.2016.08.001.
27. Dusart, P., Hallstrom, B.M., Renne, T., Odeberg, J., Uhlen, 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 e1694. 10.1016/j.celrep.2019.09.088.
28. Norreen-Thorsen, M., Struck, E.C., Oling, S., Zwahlen, M., Von Feilitzen, K., Odeberg, J., Lindskog, C., Ponten, F., Uhlen, M., Dusart, P.J., and Butler, L.M. (2022). A human adipose tissue cell-type transcriptome atlas. Cell Rep *40*, 111046. 10.1016/j.celrep.2022.111046.

29. Consortium, G.T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648-660. 10.1126/science.1262110.
30. 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. 10.1038/s41586-020-2157-4.
31. 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. 10.1093/nar/gky900.
32. 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 (Oxford) *2019*. 10.1093/database/baz046.
33. Consortium, G.T. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318-1330. 10.1126/science.aaz1776.
34. Wang, X., Park, J., Susztak, K., Zhang, N.R., and Li, M. (2019). Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nat Commun *10*, 380. 10.1038/s41467-018-08023-x.
35. Debbabi, H., Ghosh, S., Kamath, A.B., Alt, J., Demello, D.E., Dunsmore, S., and Behar, S.M. (2005). Primary type II alveolar epithelial cells present microbial antigens to antigen-specific CD4+ T cells. Am J Physiol Lung Cell Mol Physiol *289*, L274-279. 10.1152/ajplung.00004.2005.
36. Tabula Sapiens, C., Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. 10.1126/science.abl4896.
37. Gene Ontology, C. (2021). The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res *49*, D325-D334. 10.1093/nar/gkaa1113.
38. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559. 10.1186/1471-2105-9-559.
39. 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 (Oxford) *2016*. 10.1093/database/baw100.
40. Moede, T., Leibiger, I.B., and Berggren, P.O. (2020). Alpha cell regulation of beta cell function. Diabetologia *63*, 2064-2075. 10.1007/s00125-020-05196-3.
41. 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. The Gene Ontology Consortium. Nature genetics *25*, 25-29. 10.1038/75556.
42. Huang, C., Walker, E.M., Dadi, P.K., Hu, R., Xu, Y., Zhang, W., Sanavia, T., Mun, J., Liu, J., Nair, G.G., et al. (2018). Synaptotagmin 4 Regulates Pancreatic beta Cell Maturation by Modulating the Ca(2+) Sensitivity of Insulin Secretion Vesicles. Dev Cell *45*, 347-361 e345. 10.1016/j.devcel.2018.03.013.
43. Tarquis-Medina, M., Scheibner, K., Gonzalez-Garcia, I., Bastidas-Ponce, A., Sterr, M., Jaki, J., Schirge, S., Garcia-Caceres, C., Lickert, H., and Bakhti, M. (2021). Synaptotagmin-13 Is a Neuroendocrine Marker in Brain, Intestine and Pancreas. Int J Mol Sci *22*. 10.3390/ijms222212526.
44. Bakhti, M., Bastidas-Ponce, A., Tritschler, S., Tarquis-Medina, M., Nedvedova, E., Scheibner, K., Jaki, J., Cota, P., Salinno, C., Boldt, K., et al. (2021). Synaptotagmin 13

orchestrates pancreatic endocrine cell egression and islet morphogenesis. www.biorxiv.org *https://doi.org/10.1101/2021.08.30.458251*.
45. Churchill, A.J., Gutierrez, G.D., Singer, R.A., Lorberbaum, D.S., Fischer, K.A., and Sussel, L. (2017). Genetic evidence that Nkx2.2 acts primarily downstream of Neurog3 in pancreatic endocrine lineage development. Elife *6*. 10.7554/eLife.20010.
46. Bohuslavova, R., Smolik, O., Malfatti, J., Berkova, Z., Novakova, Z., Saudek, F., and Pavlinkova, G. (2021). NEUROD1 Is Required for the Early alpha and beta Endocrine Differentiation in the Pancreas. Int J Mol Sci *22*. 10.3390/ijms22136713.
47. Soyer, J., Flasse, L., Raffelsberger, W., Beucher, A., Orvain, C., Peers, B., Ravassard, P., Vermot, J., Voz, M.L., Mellitzer, G., and Gradwohl, G. (2010). Rfx6 is an Ngn3-dependent winged helix transcription factor required for pancreatic islet cell development. Development *137*, 203-212. 10.1242/dev.041673.
48. Liang, X., Duan, H., Mao, Y., Koestner, U., Wei, Y., Deng, F., Zhuang, J., Li, H., Wang, C., Hernandez-Miranda, L.R., et al. (2021). The SNAG Domain of Insm1 Regulates Pancreatic Endocrine Cell Differentiation and Represses beta- to delta-Cell Transdifferentiation. Diabetes *70*, 1084-1097. 10.2337/db20-0883.
49. Hart, A.W., Mella, S., Mendrychowski, J., van Heyningen, V., and Kleinjan, D.A. (2013). The developmental regulator Pax6 is essential for maintenance of islet cell function in the adult mouse pancreas. PLoS One *8*, e54173. 10.1371/journal.pone.0054173.
50. Wang, S., Zhang, J., Zhao, A., Hipkens, S., Magnuson, M.A., and Gu, G. (2007). Loss of Myt1 function partially compromises endocrine islet cell differentiation and pancreatic physiological function in the mouse. Mech Dev *124*, 898-910. 10.1016/j.mod.2007.08.004.
51. Su, Y., Jono, H., Misumi, Y., Senokuchi, T., Guo, J., Ueda, M., Shinriki, S., Tasaki, M., Shono, M., Obayashi, K., et al. (2012). Novel function of transthyretin in pancreatic alpha cells. FEBS Lett *586*, 4215-4222. 10.1016/j.febslet.2012.10.025.
52. Noguchi, G.M., and Huising, M.O. (2019). Integrating the inputs that shape pancreatic islet hormone release. Nat Metab *1*, 1189-1201. 10.1038/s42255-019-0148-2.
53. Yang, J.K., Lu, J., Yuan, S.S., Asan, Cao, X., Qiu, H.Y., Shi, T.T., Yang, F.Y., Li, Q., Liu, C.P., et al. (2018). From Hyper- to Hypoinsulinemia and Diabetes: Effect of KCNH6 on Insulin Secretion. Cell Rep *25*, 3800-3810 e3806. 10.1016/j.celrep.2018.12.005.
54. Westermark, P., Andersson, A., and Westermark, G.T. (2011). Islet amyloid polypeptide, islet amyloid, and diabetes mellitus. Physiol Rev *91*, 795-826. 10.1152/physrev.00042.2009.
55. Nishimura, W., Takahashi, S., and Yasuda, K. (2015). MafA is critical for maintenance of the mature beta cell phenotype in mice. Diabetologia *58*, 566-574. 10.1007/s00125-014-3464-9.
56. Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800. 10.1371/journal.pone.0021800.
57. Ewen, K.A., Olesen, I.A., Winge, S.B., Nielsen, A.R., Nielsen, J.E., Graem, N., Juul, A., and Rajpert-De Meyts, E. (2013). Expression of FGFR3 during human testis development and in germ cell-derived tumours of young adults. Int J Dev Biol *57*, 141-151. 10.1387/ijdb.130022er.
58. Qian, B., Li, Y., Yan, R., Han, S., Bu, Z., Gong, J., Zheng, B., Yuan, Z., Ren, S., He, Q., et al. (2022). RNA binding protein RBM46 regulates mitotic-to-meiotic transition in spermatogenesis. Sci Adv *8*, eabq2945. 10.1126/sciadv.abq2945.

59. Guo, J., Grow, E.J., Mlcochova, H., Maher, G.J., Lindskog, C., Nie, X., Guo, Y., Takei, Y., Yun, J., Cai, L., et al. (2018). The adult human testis transcriptional cell atlas. Cell Res *28*, 1141-1157. 10.1038/s41422-018-0099-2.
60. Le, L., Escobar, I.E., Ho, T., Lefkovith, A.J., Latteri, E., Haltaufderhyde, K.D., Dennis, M.K., Plowright, L., Sviderskaya, E.V., Bennett, D.C., et al. (2020). SLC45A2 protein stability and regulation of melanosome pH determine melanocyte pigmentation. Mol Biol Cell *31*, 2687-2702. 10.1091/mbc.E20-03-0200.
61. Yu, R., Broady, R., Huang, Y., Wang, Y., Yu, J., Gao, M., Levings, M., Wei, S., Zhang, S., Xu, A., et al. (2012). Transcriptome analysis reveals markers of aberrantly activated innate immunity in vitiligo lesional and non-lesional skin. PLoS One *7*, e51040. 10.1371/journal.pone.0051040.
62. Abbas Zadeh, S., Mlitz, V., Lachner, J., Golabi, B., Mildner, M., Pammer, J., Tschachler, E., and Eckhart, L. (2017). Phylogenetic profiling and gene expression studies implicate a primary role of PSORS1C2 in terminal differentiation of keratinocytes. Exp Dermatol *26*, 352-358. 10.1111/exd.13272.
63. Liu, Y., Das, S., Olszewski, R.E., Carpenter, D.A., Culiat, C.T., Sundberg, J.P., Soteropoulos, P., Liu, X., Doktycz, M.J., Michaud, E.J., and Voy, B.H. (2007). The near-naked hairless (Hr(N)) mutation disrupts hair formation but is not due to a mutation in the Hairless coding region. J Invest Dermatol *127*, 1605-1614. 10.1038/sj.jid.5700755.
64. Li, W., Tang, W., Teves, M.E., Zhang, Z., Zhang, L., Li, H., Archer, K.J., Peterson, D.L., Williams, D.C., Jr., Strauss, J.F., 3rd, and Zhang, Z. (2015). A MEIG1/PACRG complex in the manchette is essential for building the sperm flagella. Development *142*, 921-930. 10.1242/dev.119834.
65. Dawe, H.R., Farr, H., Portman, N., Shaw, M.K., and Gull, K. (2005). The Parkin co-regulated gene product, PACRG, is an evolutionarily conserved axonemal protein that functions in outer-doublet microtubule morphogenesis. J Cell Sci *118*, 5421-5430. 10.1242/jcs.02659.
66. Meschede, J., Sadic, M., Furthmann, N., Miedema, T., Sehr, D.A., Muller-Rischart, A.K., Bader, V., Berlemann, L.A., Pilsl, A., Schlierf, A., et al. (2020). The parkin-coregulated gene product PACRG promotes TNF signaling by stabilizing LUBAC. Sci Signal *13*. 10.1126/scisignal.aav1256.
67. Soon, R.K., Jr., and Yee, H.F., Jr. (2008). Stellate cell contraction: role, regulation, and potential therapeutic target. Clin Liver Dis *12*, 791-803, viii. 10.1016/j.cld.2008.07.004.
68. Khaing, P., and Summer, R. (2020). Maxed Out on Glycolysis: Alveolar Macrophages Rely on Oxidative Phosphorylation for Cytokine Production. Am J Respir Cell Mol Biol *62*, 139-140. 10.1165/rcmb.2019-0329ED.
69. Shetty, S., Lalor, P.F., and Adams, D.H. (2018). Liver sinusoidal endothelial cells - gatekeepers of hepatic immunity. Nat Rev Gastroenterol Hepatol *15*, 555-567. 10.1038/s41575-018-0020-y.
70. Delling, J.P., and Boeckers, T.M. (2021). Comparison of SHANK3 deficiency in animal models: phenotypes, treatment strategies, and translational implications. J Neurodev Disord *13*, 55. 10.1186/s11689-021-09397-8.
71. Sawada, J., Hiraoka, N., Qi, R., Jiang, L., Fournier-Goss, A.E., Yoshida, M., Kawashima, H., and Komatsu, M. (2022). Molecular Signature of Tumor-Associated High Endothelial Venules That Can Predict Breast Cancer Survival. Cancer Immunol Res *10*, 468-481. 10.1158/2326-6066.CIR-21-0369.
72. van der Weyden, L., Harle, V., Turner, G., Offord, V., Iyer, V., Droop, A., Swiatkowska, A., Rabbie, R., Campbell, A.D., Sansom, O.J., et al. (2021). CRISPR

activation screen in mice identifies novel membrane proteins enhancing pulmonary metastatic colonisation. Commun Biol *4*, 395. 10.1038/s42003-021-01912-w.
73. Zhang, Y., Tong, G.H., Wei, X.X., Chen, H.Y., Liang, T., Tang, H.P., Wu, C.A., Wen, G.M., Yang, W.K., Liang, L., and Shen, H. (2021). Identification of Five Cytotoxicity-Related Genes Involved in the Progression of Triple-Negative Breast Cancer. Front Genet *12*, 723477. 10.3389/fgene.2021.723477.
74. Lin, Y.J., Chen, A.N., Yin, X.J., Li, C., and Lin, C.C. (2020). Human Microfibrillar-Associated Protein 4 (MFAP4) Gene Promoter: A TATA-Less Promoter That Is Regulated by Retinol and Coenzyme Q10 in Human Fibroblast Cells. Int J Mol Sci *21*. 10.3390/ijms21218392.
75. Sellers, M.M., and Stallone, J.N. (2008). Sympathy for the devil: the role of thromboxane in the regulation of vascular tone and blood pressure. Am J Physiol Heart Circ Physiol *294*, H1978-1986. 10.1152/ajpheart.01318.2007.
76. Hosokawa, H., Romero-Wolf, M., Yang, Q., Motomura, Y., Levanon, D., Groner, Y., Moro, K., Tanaka, T., and Rothenberg, E.V. (2020). Cell type-specific actions of Bcl11b in early T-lineage and group 2 innate lymphoid cells. J Exp Med *217*. 10.1084/jem.20190972.
77. Golonzhka, O., Liang, X., Messaddeq, N., Bornert, J.M., Campbell, A.L., Metzger, D., Chambon, P., Ganguli-Indra, G., Leid, M., and Indra, A.K. (2009). Dual role of COUP-TF-interacting protein 2 in epidermal homeostasis and permeability barrier formation. J Invest Dermatol *129*, 1459-1470. 10.1038/jid.2008.392.
78. Zhao, Y.Y., Yan, D.J., and Chen, Z.W. (2013). Role of AIF-1 in the regulation of inflammatory activation and diverse disease processes. Cell Immunol *284*, 75-83. 10.1016/j.cellimm.2013.07.008.
79. Tsubata, Y., Sakatsume, M., Ogawa, A., Alchi, B., Kaneko, Y., Kuroda, T., Kawachi, H., Narita, I., Yamamoto, T., and Gejyo, F. (2006). Expression of allograft inflammatory factor-1 in kidneys: A novel molecular component of podocyte. Kidney Int *70*, 1948-1954. 10.1038/sj.ki.5001941.
80. Li, W.V., and Li, J.J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat Commun *9*, 997. 10.1038/s41467-018-03405-7.
81. Mou, T., Deng, W., Gu, F., Pawitan, Y., and Vu, T.N. (2020). Reproducibility of Methods to Detect Differentially Expressed Genes from Single-Cell RNA Sequencing. Front Genet *10*, 1331. 10.3389/fgene.2019.01331.
82. Chu, S.K., Zhao, S., Shyr, Y., and Liu, Q. (2022). Comprehensive evaluation of noise reduction methods for single-cell RNA sequencing data. Brief Bioinform *23*. 10.1093/bib/bbab565.
83. Mandelboum, S., Manber, Z., Elroy-Stein, O., and Elkon, R. (2019). Recurrent functional misinterpretation of RNA-seq data caused by sample-specific gene length bias. PLoS Biol *17*, e3000481. 10.1371/journal.pbio.3000481.
84. Liu, Y., Beyer, A., and Aebersold, R. (2016). On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell *165*, 535-550. 10.1016/j.cell.2016.03.014.
85. 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. 10.1038/nmeth.3337.
86. 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. 10.1016/j.ajhg.2019.03.025.

87. Jew, B., Alvarez, M., Rahmani, E., Miao, Z., Ko, A., Garske, K.M., Sul, J.H., Pietiläinen, K.H., Pajukanta, P., and Halperin, E. (2020). Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. Nat Commun *11*, 1971. 10.1038/s41467-020-15816-6.
88. Gibson, G. (2008). The environmental contribution to gene expression profiles. Nat Rev Genet *9*, 575-581. 10.1038/nrg2383.
89. Vestweber, D., and Blanks, J.E. (1999). Mechanisms that regulate the function of the selectins and their ligands. Physiol Rev *79*, 181-213. 10.1152/physrev.1999.79.1.181.
90. Yates, A.D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R., et al. (2020). Ensembl 2020. Nucleic Acids Res *48*, D682-D688. 10.1093/nar/gkz966.
91. Ponten, F., Jirstrom, K., and Uhlen, M. (2008). The Human Protein Atlas - a tool for pathology. J Pathol *216*, 387-393. 10.1002/path.2440.
92. Uhlen, M., Zhang, C., Lee, S., Sjostedt, 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*. 10.1126/science.aan2507.
93. 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*, D336-342. 10.1093/nar/gkv1194.
94. Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. Nature protocols *8*, 1551-1566. 10.1038/nprot.2013.092.
95. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb) *2*, 100141. 10.1016/j.xinn.2021.100141.
96. 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 e3529. 10.1016/j.cell.2021.04.048.
97. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
98. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. Bioinformatics *30*, 2811-2812. 10.1093/bioinformatics/btu393.

Supplementary data: Paper II

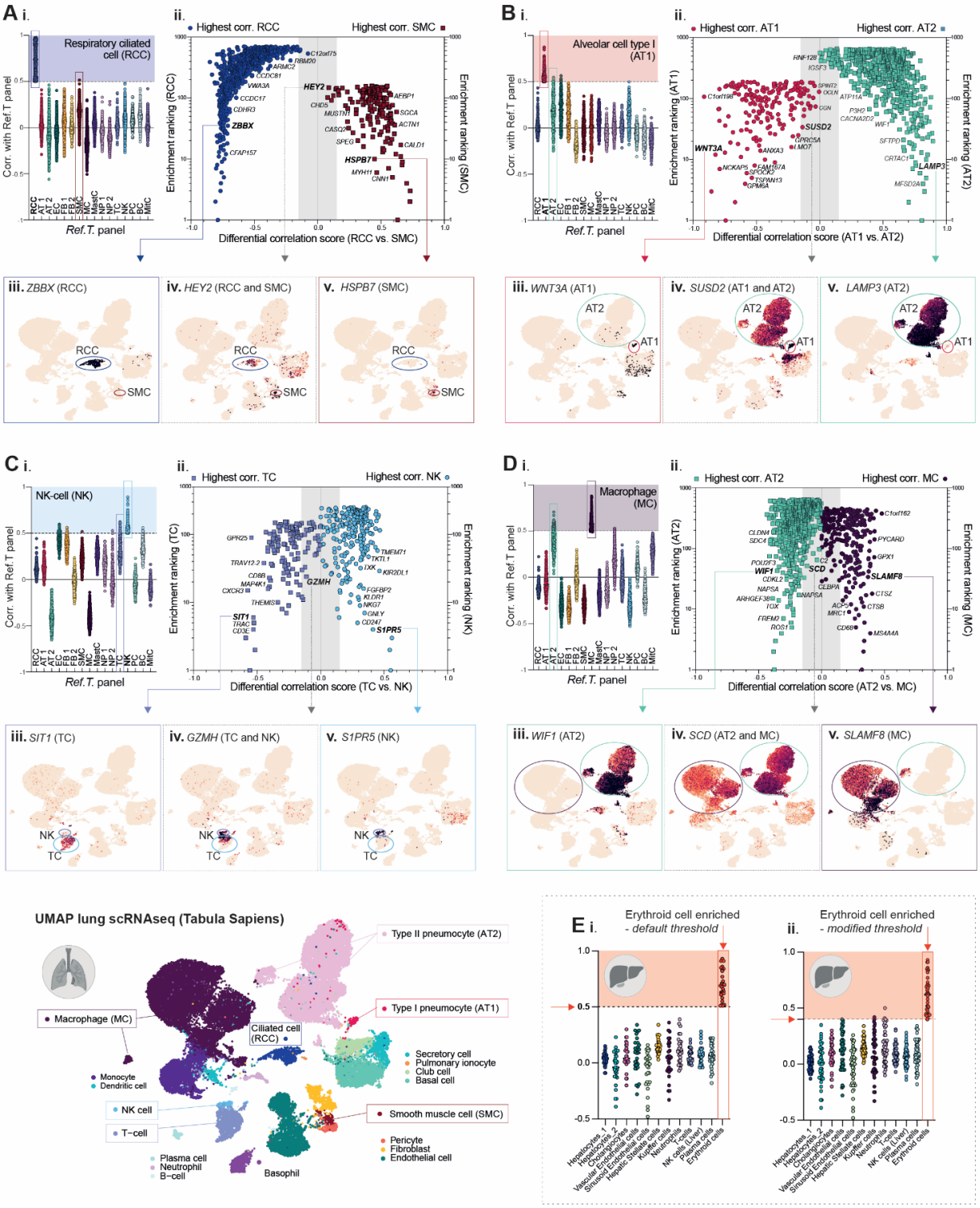

A i.
Respiratory ciliated cell (RCC)
Corr. with Ref.T panel
Ref.T. panel
ii.
Highest corr. RCC
Highest corr. SMC
Enrichment ranking (RCC)
Enrichment ranking (SMC)
Differential correlation score (RCC vs. SMC)
C12orf75
RBM20
ARMC2
CCDC81
VWA3A
CCDC17
CDHR3
ZBBX
CFAP157
HEY2
CHD5
MUSTN1
CASQ2
SPEG
HSPB7
MYH11
CNN1
AEBP1
SGCA
ACTN1
CALD1
iii. ZBBX (RCC)
iv. HEY2 (RCC and SMC)
v. HSPB7 (SMC)
RCC
SMC
B i.
Alveolar cell type I (AT1)
ii.
Highest corr. AT1
Highest corr. AT2
Differential correlation score (AT1 vs. AT2)
RNF128
IGSF3
C1orf198
SPINT2
OCLN
ATP11A
CGN
P3H2
CACNA2D2
WIF1
SUSD2
GPR35A
LMO7
WNT3A
ANXA3
NCKAP5
FAM107A
SPOCD1
TSPAN13
GPM6A
SFTPD
CRTAC1
LAMP3
MFSD2A
iii. WNT3A (AT1)
iv. SUSD2 (AT1 and AT2)
v. LAMP3 (AT2)
AT1
AT2
C i.
NK-cell (NK)
ii.
Highest corr. TC
Highest corr. NK
Differential correlation score (TC vs. NK)
GPR25
TRAV12-2
CD8B
MAP4K1
CXCR3
THEMIS
SIT1
TRAC
CD3E
TMEM71
TXK
KIR2DL1
GZMH
FCGBP2
KLRD1
NKG7
IKZF3
CD247
S1PR5
iii. SIT1 (TC)
iv. GZMH (TC and NK)
v. S1PR5 (NK)
NK
TC
D i.
Macrophage (MC)
ii.
Highest corr. AT2
Highest corr. MC
Differential correlation score (AT2 vs. MC)
C1orf162
CLDN4
SDC4
PYCARD
POU2F3
WIF1
CDKL1
SCD
GPX1
SLAMF8
NAPSA
ARHGEF38
TOX
CEBPA
CTSZ
CTSB
FREM2
ROS1
MRC1
CD68
MS4A4A
iii. WIF1 (AT2)
iv. SCD (AT2 and MC)
v. SLAMF8 (MC)
UMAP lung scRNAseq (Tabula Sapiens)
Type II pneumocyte (AT2)
Type I pneumocyte (AT1)
Macrophage (MC)
Ciliated cell (RCC)
Secretory cell
Pulmonary ionocyte
Club cell
Basal cell
Monocyte
Dendritic cell
NK cell
T−cell
Smooth muscle cell (SMC)
Pericyte
Fibroblast
Endothelial cell
Plasma cell
Neutrophil
B−cell
Basophil
E i.
Erythroid cell enriched
- default threshold
ii.
Erythroid cell enriched
- modified threshold
Hepatocytes_1
Hepatocytes_2
Cholangiocytes
Vascular Endothelial cells
Sinusoid Endothelial cells
Hepatic Stellate cells
Kupffer cells
Neutrophils
T cells (liver)
NK cells (liver)
Plasma cells
Erythroid cells

Figure S1. Integrative co-expression analysis of unfractionated human tissue RNAseq can resolve constituent cell type enriched genes. Related to Figure 1.RNAseq datasets for human lung (n=578) were retrieved from GTEx V8 and correlation coefficients between selected cell type Ref. T. and all other sequenced transcripts generated. Correlation values vs. all other cell type Ref.T. panels for transcripts reaching the designated threshold with Ref. T. for **(A) (i)** respiratory ciliated cells (RCC) **(B) (i)** alveolar type I cells (AT1), **(C) (i)** natural killer cells (NK) or **(D) (i)** macrophages (MC). The 'differential correlation score' and respective enrichment rankings for transcripts reaching the designated threshold with Ref. T. for **(A) (ii)** RCC or SMC, **(B) (ii)** AT1 or AT2, **(C) (ii)** NK or TC and **(D) (ii)** MC and AT2. scRNAseq data from analysis of human lung was sourced from Tabula Sapiens (Tabula Sapiens et al., 2022) and used to generate UMAP plots, showing the expression profiles of example genes we predicted as being enriched in **(A) (iii)** RCC only, **(iv)** RCC and SMC or **(v)** SMC only, **(B) (iii)** AT1 only, **(iv)** AT1 and AT2 or **(v)** AT2 only, **(C) (iii)** TC only, **(iv)** TC and NK or **(v)** NK only, or **(D) (iii)** AT2 only, **(iv)** AT2 and MC or **(v)** MC only. **(E).** RNAseq datasets for human liver (n=226) were retrieved from GTEx V8 and analysed as described for lung. Correlation values vs. all cell type Ref.T. panels for transcripts reaching the **(i)** designated or **(ii)** modified threshold for classification as erythroid cell enriched. EC; Endothelial cell, FB1/FB2; fibroblast, MC; macrophage, MastC; mast cell, NP1/NP2; neutrophil, TC; T-cell, NK; natural killer cell, PC; plasma cell, BC; B-cell.

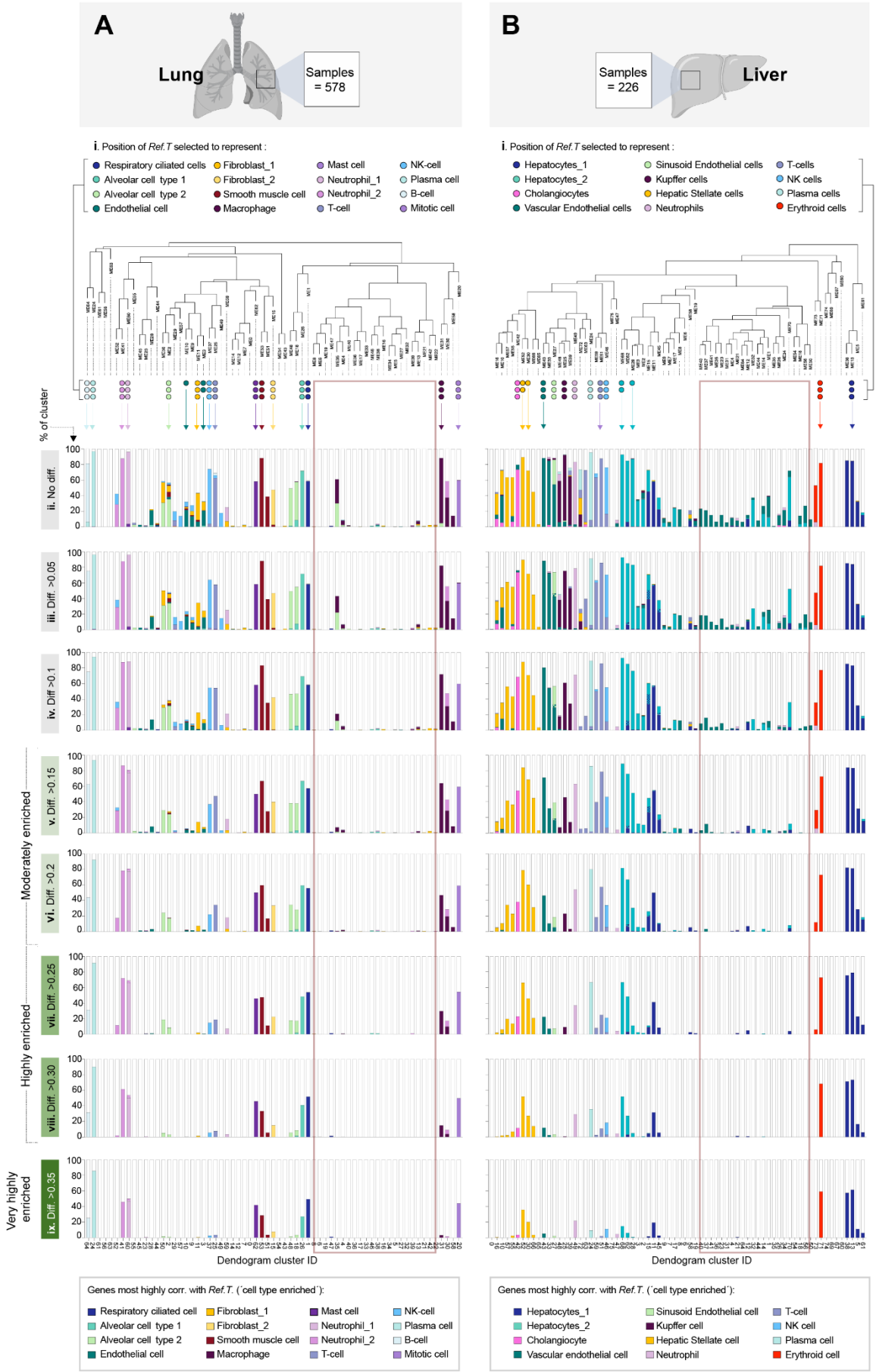

A
Lung
Samples = 578
B
Samples = 226
Liver
i. Position of Ref.T selected to represent :
Respiratory ciliated cells
Alveolar cell type 1
Alveolar cell type 2
Endothelial cell
Fibroblast_1
Fibroblast_2
Smooth muscle cell
Macrophage
Mast cell
Neutrophil_1
Neutrophil_2
T-cell
NK-cell
Plasma cell
B-cell
Mitotic cell
Hepatocytes_1
Hepatocytes_2
Cholangiocytes
Vascular Endothelial cells
Sinusoid Endothelial cells
Kupffer cells
Hepatic Stellate cells
Neutrophils
T-cells
NK cells
Plasma cells
Erythroid cells
% of cluster
ii. No diff.
iii. Diff. >0.05
iv. Diff >0.1
v. Diff. >0.15
vi. Diff. >0.2
vii. Diff. >0.25
viii. Diff. >0.30
ix. Diff. >0.35
Moderately enriched
Highly enriched
Very highly enriched
Dendogram cluster ID
Genes most highly corr. with Ref.T. (´cell type enriched´):
Respiratory ciliated cell
Alveolar cell type 1
Alveolar cell type 2
Endothelial cell
Fibroblast_1
Fibroblast_2
Smooth muscle cell
Macrophage
Mast cell
Neutrophil_1
Neutrophil_2
T-cell
NK-cell
Plasma cell
B-cell
Mitotic cell
Hepatocytes_1
Hepatocytes_2
Cholangiocyte
Vascular endothelial cell
Sinusoid Endothelial cell
Kupffer cell
Hepatic Stellate cell
Neutrophil
T-cell
NK cell
Plasma cell
Erythroid cell

Figure S2. Unsupervised weighted network correlation analysis (WGNCA) is consistent with *Ref.T.* analysis. Related to Figure 1. RNAseq data from human **(A)** lung (n=578 individuals) or **(B)** pancreas (n=328) was subject to weighted correlation network analysis (WGCNA). In the resultant dendrograms, the position of **(i)** *Ref.T*. selected to represent each cell type and **(ii)** the % of the cluster containing transcripts that had a correlation with any *Ref.T*. panel above the designated threshold, are indicated; colour representing the cell type classification (see bottom panel) (Table S1, Tab 5 for thresholds). Distribution of transcripts for each cell type classification when the highest correlation with any given *Ref.T*. panel was a minimum of **(ii)** 0, **(iii)** 0.05, **(iv)** 0.10, **(v)** 0.15 [moderately enriched], **(vi)** 0.20, **(vii)** 0.25 [highly enriched] or **(viii)** 0.30 or **(ix)** 0.35 [very highly enriched] greater that the next highest with a different *Ref.T.* panel (´differential correlation score´).

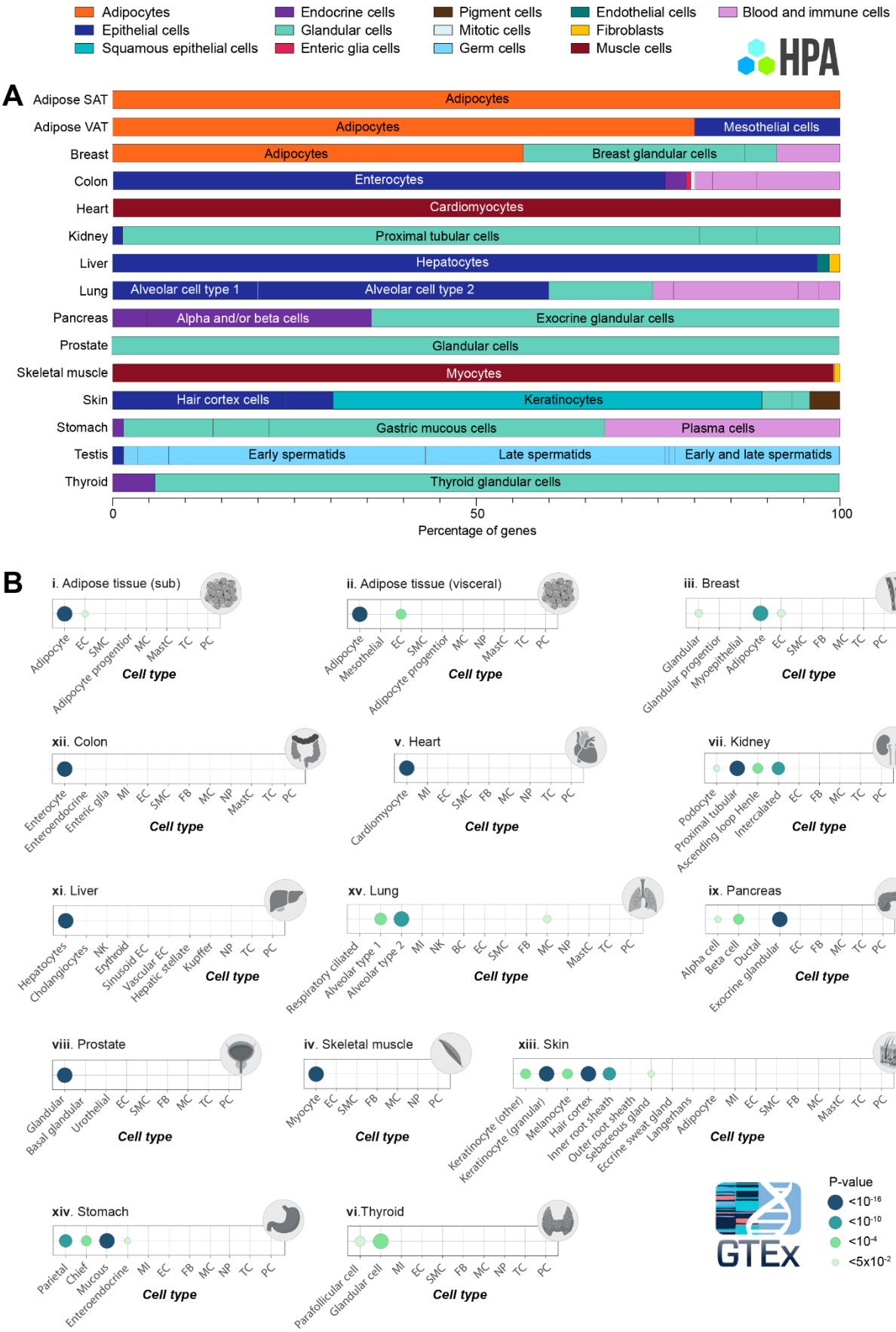

Adipocytes
Epithelial cells
Squamous epithelial cells
Endocrine cells
Glandular cells
Enteric glia cells
Pigment cells
Mitotic cells
Germ cells
Endothelial cells
Fibroblasts
Muscle cells
Blood and immune cells
HPA
A
Adipose SAT
Adipocytes
Adipose VAT
Adipocytes
Mesothelial cells
Breast
Adipocytes
Breast glandular cells
Colon
Enterocytes
Heart
Cardiomyocytes
Kidney
Proximal tubular cells
Liver
Hepatocytes
Lung
Alveolar cell type 1
Alveolar cell type 2
Pancreas
Alpha and/or beta cells
Exocrine glandular cells
Prostate
Glandular cells
Skeletal muscle
Myocytes
Skin
Hair cortex cells
Keratinocytes
Stomach
Gastric mucous cells
Plasma cells
Testis
Early spermatids
Late spermatids
Early and late spermatids
Thyroid
Thyroid glandular cells
0
50
100
Percentage of genes
B
i. Adipose tissue (sub)
ii. Adipose tissue (visceral)
iii. Breast
xii. Colon
v. Heart
vii. Kidney
xi. Liver
xv. Lung
ix. Pancreas
viii. Prostate
iv. Skeletal muscle
xiii. Skin
xiv. Stomach
vi.Thyroid
Cell type
GTEx
P-value
<10-16
<10-10
<10-4
<5x10-2

Figure S3. Integrative co-expression analysis of unfractionated human tissue RNAseq can resolve tissue enriched genes into single cell type expression source. Related to Figure 2. (A) Bar plot showing the fraction of predicted cell type enriched genes among the tissue, or tissue-group, enriched genes in Human Protein Atlas (HPA). Colour indicates cell type group. The cell type with the most shared enriched genes with tissues are labelled. **(B)** Bubble plots showing the significance (indicated by dot size and colour) of similarity between the top 300 tissue enriched genes in GTEx and the predicted cell type enrichment signatures Where overlap is not statistically significant (hypergeometric test, $P > 0.05$), the corresponding dot is removed. EC; endothelial cell, SMC; Smooth muscle cell, MC; macrophage, MastC; mast cell, TC; T-cell, PC; Plasma cell, NP; Neutrophil, MI; Mitotic cell, NK; Natural killer cell.

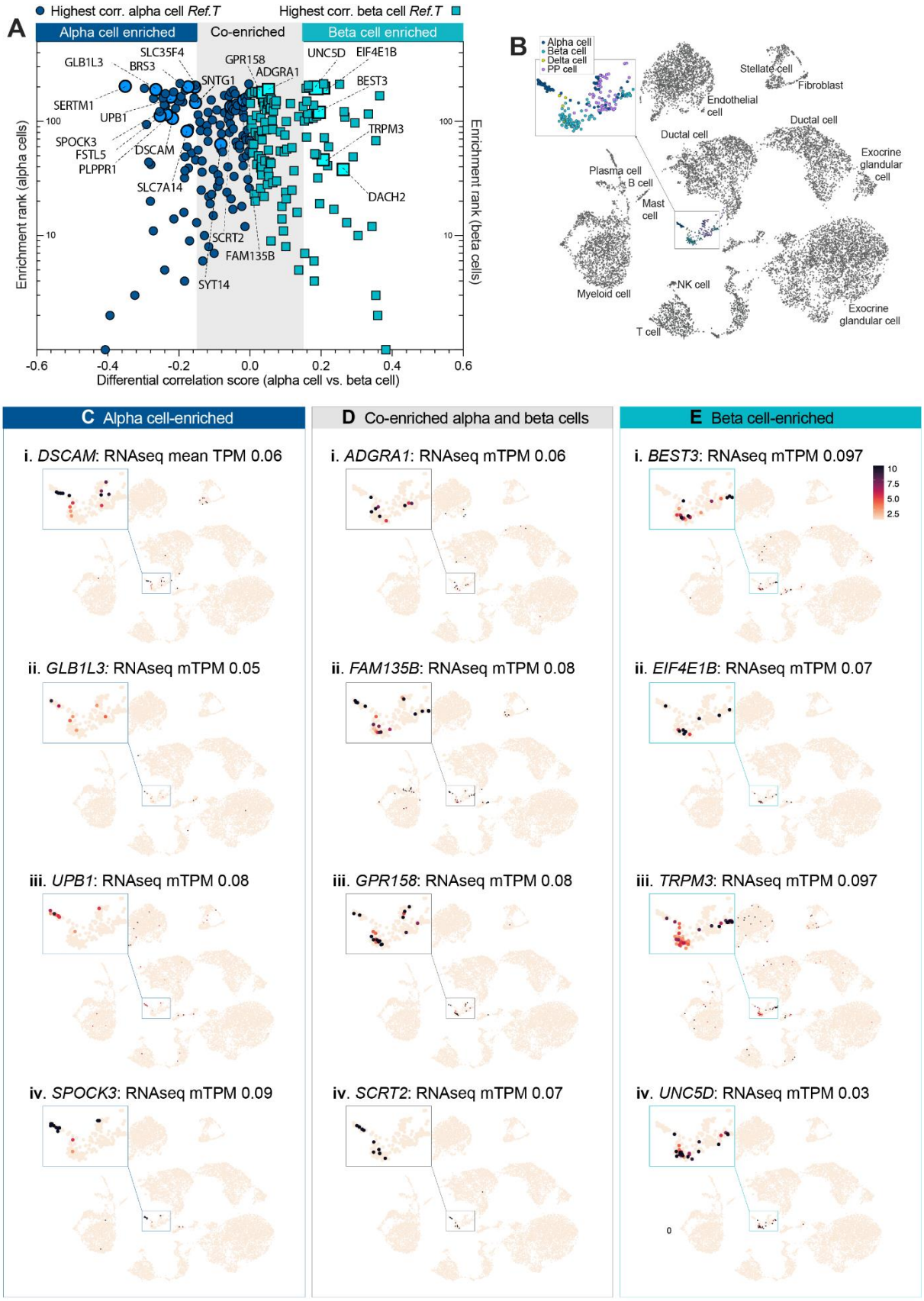

A
Highest corr. alpha cell Ref.T
Highest corr. beta cell Ref.T
Alpha cell enriched
Co-enriched
Beta cell enriched
GLB1L3
SLC35F4
BRS3
GPR158
ADGRA1
UNC5D
EIF4E1B
SNTG1
BEST3
SERTM1
UPB1
TRPM3
SPOCK3
FSTL5
PLPPR1
DSCAM
SLC7A14
DACH2
SCRT2
FAM135B
SYT14
Enrichment rank (alpha cells)
Enrichment rank (beta cells)
Differential correlation score (alpha cell vs. beta cell)
B
Alpha cell
Beta cell
Delta cell
PP cell
Stellate cell
Fibroblast
Endothelial cell
Ductal cell
Exocrine glandular cell
Plasma cell
B cell
Mast cell
Myeloid cell
NK cell
T cell
C Alpha cell-enriched
i. DSCAM: RNAseq mean TPM 0.06
ii. GLB1L3: RNAseq mTPM 0.05
iii. UPB1: RNAseq mTPM 0.08
iv. SPOCK3: RNAseq mTPM 0.09
D Co-enriched alpha and beta cells
i. ADGRA1: RNAseq mTPM 0.06
ii. FAM135B: RNAseq mTPM 0.08
iii. GPR158: RNAseq mTPM 0.08
iv. SCRT2: RNAseq mTPM 0.07
E Beta cell-enriched
i. BEST3: RNAseq mTPM 0.097
ii. EIF4E1B: RNAseq mTPM 0.07
iii. TRPM3: RNAseq mTPM 0.097
iv. UNC5D: RNAseq mTPM 0.03

Figure S4. Reference transcript-based identification of lowly expressed pancreatic alpha and beta cell type enriched and co-enriched genes. Related to Figure 3. RNAseq datasets for human pancreas (n=328) were analysed to generate correlation coefficient values between all protein coding genes and Ref T. **(A)** For genes that correlated most highly with alpha (dark blue) or beta cell (turquoise) Ref.T (above >0.50), the 'differential correlation score' (difference between mean corr. with alpha and beta cell Ref.T.) was plotted vs. 'enrichment ranking' (position in each respective list, highest corr. = rank 1). Shaded grey box highlights genes enriched in both cell types (co-enriched). Genes highlighted in bold correspond to those featured in the lower panels. scRNAseq data from analysis of human pancreas was sourced from Tabula Sapiens (Tabula Sapiens et al., 2022), and used to generate UMAP plots showing **(B)** scRNAseq cell type annotations, and the expression profiles of genes we predicted as being **(C)** alpha cell-enriched; **(i)** DSCAM, **(ii)** GLB1L3, **(iii)** UPB1 and **(iv)** SPOCK3, **(D)** co-enriched in both alpha and beta cells; **(i)** ADGRA1, **(ii)** FAM135B, **(iii)** GPR158 and **(iv)** SCRT2, or **(E)** beta cell-enriched; **(i)** BEST3, **(ii)** EIF4E1B, **(iii)** TRPM3 and **(iv)** UNC5D.

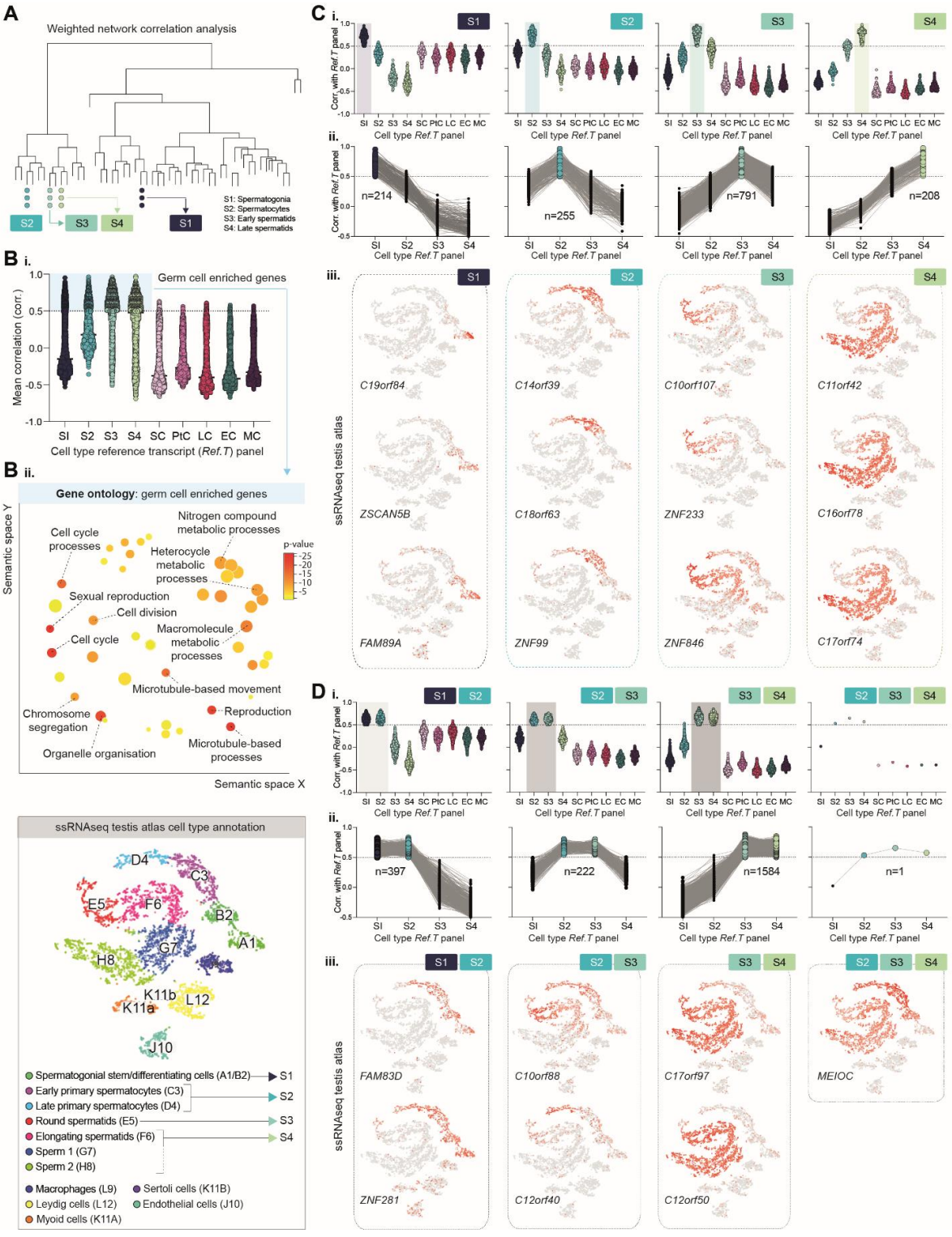

A
Weighted network correlation analysis
S1: Spermatogonia
S2: Spermatocytes
S3: Early spermatids
S4: Late spermatids
S2
S3
S4
S1
B i.
Germ cell enriched genes
Mean correlation (corr.)
Cell type reference transcript (Ref.T) panel
SI S2 S3 S4 SC PtC LC EC MC
B ii.
Gene ontology: germ cell enriched genes
Semantic space Y
Semantic space X
p-value
Nitrogen compound metabolic processes
Cell cycle processes
Heterocycle metabolic processes
Sexual reproduction
Cell division
Cell cycle
Macromolecule metabolic processes
Microtubule-based movement
Chromosome segregation
Reproduction
Organelle organisation
Microtubule-based processes
ssRNAseq testis atlas cell type annotation
D4
C3
E5
F6
B2
A1
G7
H8
K11b
L12
K11a
J10
Spermatogonial stem/differentiating cells (A1/B2)
Early primary spermatocytes (C3)
Late primary spermatocytes (D4)
Round spermatids (E5)
Elongating spermatids (F6)
Sperm 1 (G7)
Sperm 2 (H8)
Macrophages (L9)
Sertoli cells (K11B)
Leydig cells (L12)
Endothelial cells (J10)
Myoid cells (K11A)
C i.
Corr. with Ref.T panel
Cell type Ref.T panel
ii.
n=214
n=255
n=791
n=208
iii.
ssRNAseq testis atlas
C19orf84
C14orf39
C10orf107
C11orf42
ZSCAN5B
C18orf63
ZNF233
C16orf78
FAM89A
ZNF99
ZNF846
C17orf74
D i.
ii.
n=397
n=222
n=1584
n=1
iii.
FAM83D
C10orf88
C17orf97
MEIOC
ZNF281
C12orf40
C12orf50

Figure S5. Analysis of pseudo temporal changes during spermatogenesis reveals stage-specific and common stage-shared gene enrichment signatures. Related to Figure 4. (A) weighted network correlation analysis of human testis RNAseq data (n=361) annotated to show position of genes in Ref.T. panels (each indicated with single circle) selected to represent cell types at the different stages of spermatogenesis: S1 (spermatogonia), S2 (spermatocytes), S3 and S4 (early and late spermatids, respectively). **(B)** For genes with predicted cell-type enrichment in S1, S2, S3 or S4 **(i)** mean correlation coefficients with Ref.T. for S1, S2, S3, S4 and sertoli cells (SC), Leydig cells (LC), peritubular cells (PtC), endothelial cells (EC) or macrophages (MC) and **(ii)** over-represented gene ontology terms, summarised and visualised using REVIGO. For all genes predicted to be: **(C)** highly cell type enriched at one stage of spermatogenesis or **(D)** co-enriched at two or more stages of spermatogenesis (category indicated in top left of each plot): **(i)** mean correlation coefficients with Ref.T. for S1, S2, S3, S4, SC, LC, PtC, EC or MC, **(ii)** mean correlation coefficients with Ref.T. for S1, S2, S3, S4 with linkage lines connecting each individual gene **(iii)** expression profiles in Human Testis Atlas scRNAseq data (Guo et al., 2018) for selected lesser known genes appearing in each respective category. UMAP from the Human Testis Atlas shows original cell type annotations (bottom left), with arrows to indicating the broad equivalence classifications in our analysis.

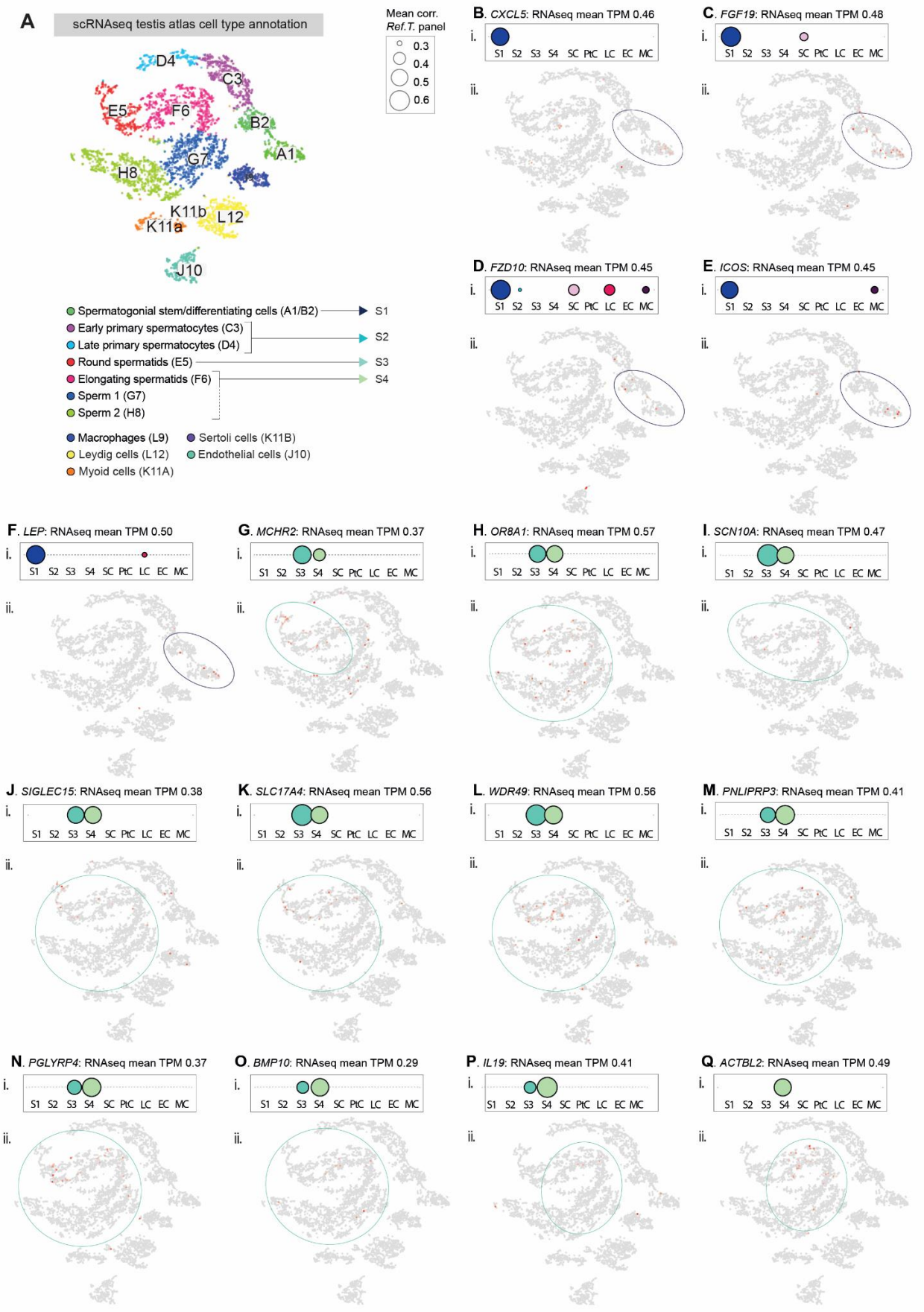

A
scRNAseq testis atlas cell type annotation
Mean corr. Ref.T. panel
0.3
0.4
0.5
0.6
D4
C3
E5
F6
B2
A1
G7
H8
K11b
L12
K11a
J10
Spermatogonial stem/differentiating cells (A1/B2)
Early primary spermatocytes (C3)
Late primary spermatocytes (D4)
Round spermatids (E5)
Elongating spermatids (F6)
Sperm 1 (G7)
Sperm 2 (H8)
S1
S2
S3
S4
Macrophages (L9)
Sertoli cells (K11B)
Leydig cells (L12)
Endothelial cells (J10)
Myoid cells (K11A)
B. CXCL5: RNAseq mean TPM 0.46
C. FGF19: RNAseq mean TPM 0.48
D. FZD10: RNAseq mean TPM 0.45
E. ICOS: RNAseq mean TPM 0.45
F. LEP: RNAseq mean TPM 0.50
G. MCHR2: RNAseq mean TPM 0.37
H. OR8A1: RNAseq mean TPM 0.57
I. SCN10A: RNAseq mean TPM 0.47
J. SIGLEC15: RNAseq mean TPM 0.38
K. SLC17A4: RNAseq mean TPM 0.56
L. WDR49: RNAseq mean TPM 0.56
M. PNLIPRP3: RNAseq mean TPM 0.41
N. PGLYRP4: RNAseq mean TPM 0.37
O. BMP10: RNAseq mean TPM 0.29
P. IL19: RNAseq mean TPM 0.41
Q. ACTBL2: RNAseq mean TPM 0.49
S1 S2 S3 S4 SC PtC LC EC MC

Figure S6. Reference transcript-based identification of lowly expressed germ cell enriched genes in the human testis. Related to Figure 4. (A) UMAP and cell type annotations as defined in the scRNAseq Human Testis Atlas (Guo et al., 2018), with arrows to indicate the broad equivalence classifications in our analysis. **(i)** Enrichment scores in all cell types profiled for genes predicted to be **(B-F)** S1 enriched, **(H-N)** S3 and S4 enriched or **(O-Q)** S4 enriched, with **(ii)** corresponding UMAP expression plots from the scRNAseq Human Testis Atlas (Guo et al., 2018).

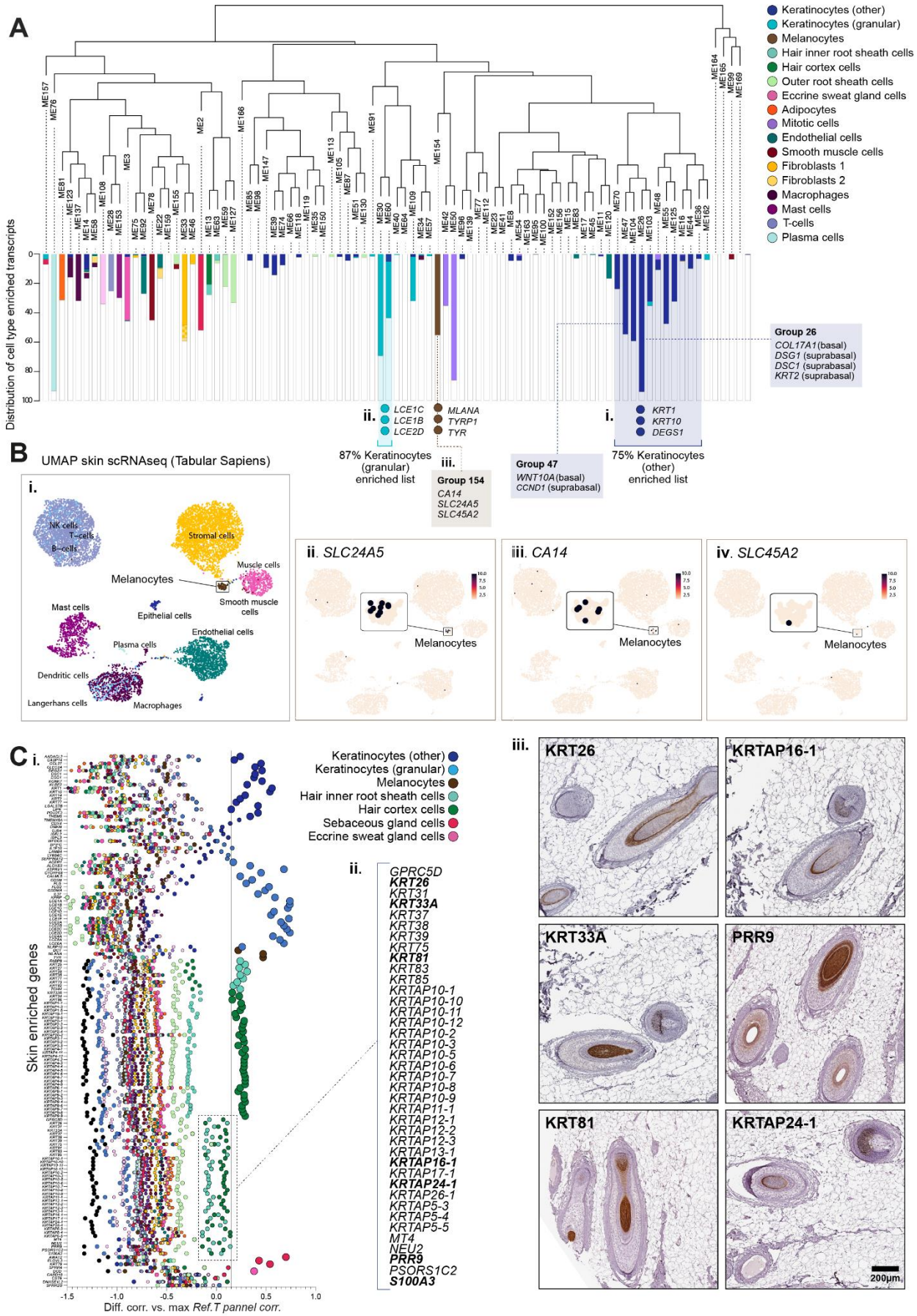

A
Keratinocytes (other)
Keratinocytes (granular)
Melanocytes
Hair inner root sheath cells
Hair cortex cells
Outer root sheath cells
Eccrine sweat gland cells
Adipocytes
Mitotic cells
Endothelial cells
Smooth muscle cells
Fibroblasts 1
Fibroblasts 2
Macrophages
Mast cells
T-cells
Plasma cells
Distribution of cell type enriched transcripts
Group 26
COL17A1 (basal)
DSG1 (suprabasal)
DSC1 (suprabasal)
KRT2 (suprabasal)
ii. LCE1C LCE1B LCE2D
87% Keratinocytes (granular) enriched list
iii. MLANA TYRP1 TYR
Group 154
CA14
SLC24A5
SLC45A2
Group 47
WNT10A (basal)
CCND1 (suprabasal)
i. KRT1 KRT10 DEGS1
75% Keratinocytes (other) enriched list
B
UMAP skin scRNAseq (Tabular Sapiens)
i.
NK cells
T-cells
B-cells
Stromal cells
Muscle cells
Melanocytes
Smooth muscle cells
Mast cells
Epithelial cells
Endothelial cells
Plasma cells
Dendritic cells
Langerhans cells
Macrophages
ii. SLC24A5
iii. CA14
iv. SLC45A2
Melanocytes
C
i.
Keratinocytes (other)
Keratinocytes (granular)
Melanocytes
Hair inner root sheath cells
Hair cortex cells
Sebaceous gland cells
Eccrine sweat gland cells
Skin enriched genes
Diff. corr. vs. max Ref.T pannel corr.
ii.
GPRC5D
KRT26
KRT31
KRT33A
KRT37
KRT38
KRT39
KRT75
KRT81
KRT83
KRT85
KRTAP10-1
KRTAP10-10
KRTAP10-11
KRTAP10-12
KRTAP10-2
KRTAP10-3
KRTAP10-5
KRTAP10-6
KRTAP10-7
KRTAP10-8
KRTAP10-9
KRTAP11-1
KRTAP12-1
KRTAP12-2
KRTAP12-3
KRTAP13-1
KRTAP16-1
KRTAP17-1
KRTAP24-1
KRTAP26-1
KRTAP5-3
KRTAP5-4
KRTAP5-5
MT4
NEU2
PRR9
PSORS1C2
S100A3
iii.
KRT26
KRTAP16-1
KRT33A
PRR9
KRT81
KRTAP24-1
200µm

Figure S7. Constituent cells of the skin hair root are the primary source of skin tissue enriched genes. Related to Figure 5. (A) Weighted network correlation analysis (WGNCA) of human skin samples (n=210) with coloured coded bars showing distribution of genes predicted to be cell type enriched. Position of Ref.T. and example cell-type enriched genes are highlighted for: **(i)** supra-basal keratinocytes, **(ii)** granular keratinocytes and **(iii)** melanocytes. **(B)** scRNAseq data and cell type definitions were sourced for human skin from Tabula Sapiens (Tabula Sapiens et al., 2022) and used to generate UMPA plots showing: **(i)** cell type annotations or expression profiles for genes we predicted to be melanocyte enriched **(ii)** SLC24A5, **(iii)** CA14 and (iv) SLC45A2. **(C)** Skin enriched genes (vs. other tissue types) were identified and **(i)** corresponding cell type enrichment profiles in skin plotted, a panel of which **(ii)** did not reach the threshold for classification as enriched in a single cell type but had highest enrichment scores in one or more hair cell types. **(iii)** Expression of proteins encoded by selected examples were profiled in human skin tissue containing hair roots.

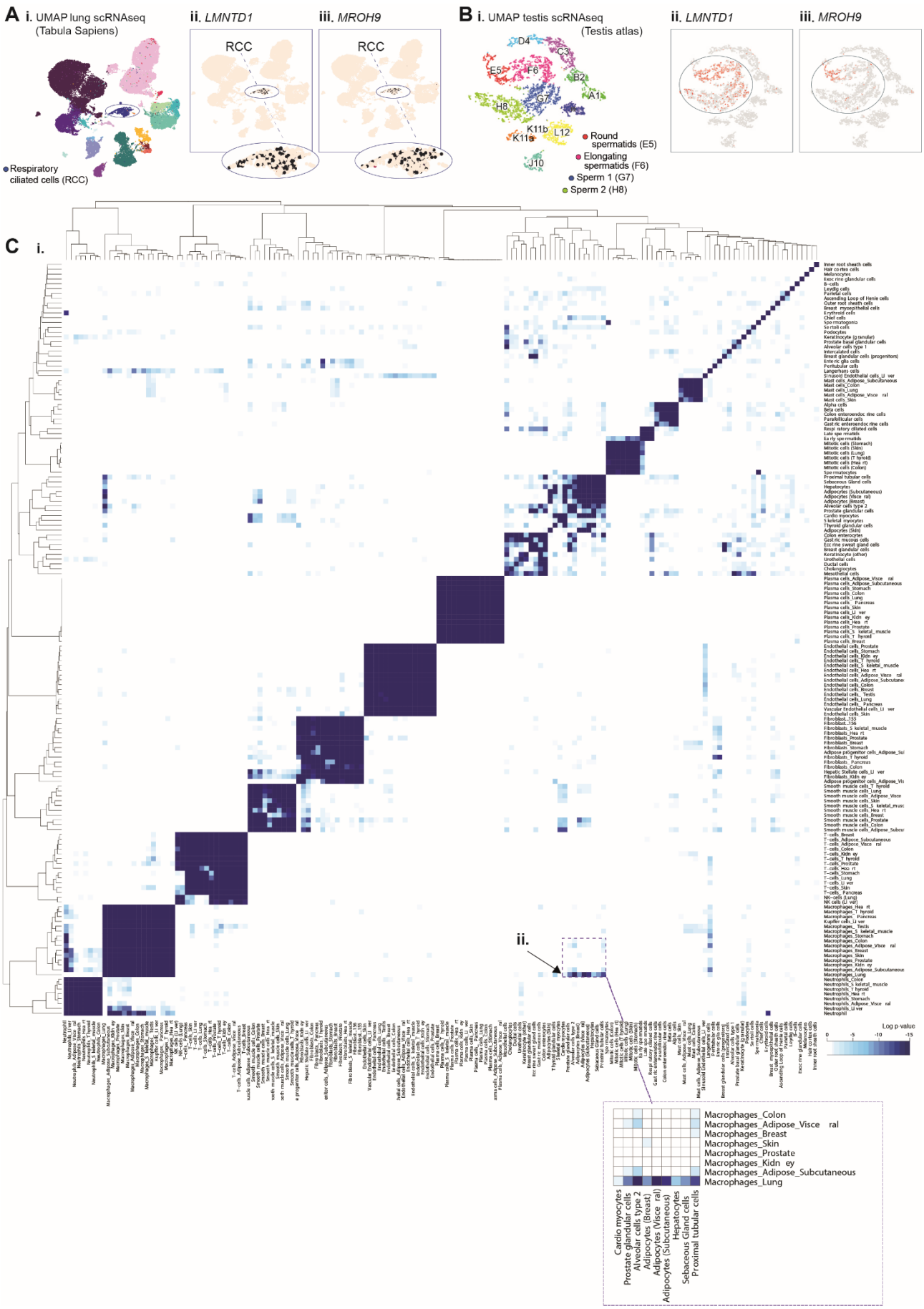

A i. UMAP lung scRNAseq
(Tabula Sapiens)
ii. LMNTD1
iii. MROH9
RCC
RCC
Respiratory
ciliated cells (RCC)
B i. UMAP testis scRNAseq
(Testis atlas)
ii. LMNTD1
iii. MROH9
D4
C3
E5
F6
B2
A1
G7
H8
K11b
L12
K11a
J10
Round
spermatids (E5)
Elongating
spermatids (F6)
Sperm 1 (G7)
Sperm 2 (H8)
C i.
ii.
Log p value
0
-5
-10
-15
Macrophages_Colon
Macrophages_Adipose_Visce ral
Macrophages_Breast
Macrophages_Skin
Macrophages_Prostate
Macrophages_Kidn ey
Macrophages_Adipose_Subcutaneous
Macrophages_Lung
Cardio myocytes
Prostate glandular cells
Alveolar cells type 2
Adipocytes (Breast)
Adipocytes (Visce ral)
Adipocytes (Subcutaneous)
Hepatocytes
Sebaceous Gland cells
Proximal tubular cells

Figure S8. Cell type enriched signature comparisons. Related to Figure 6 and 7. scRNAseq data was sourced for human **(A)** lung from Tabula Sapiens (Tabula Sapiens et al., 2022) or **(B)** testis from the Human Testis Atlas (Guo et al., 2018), and used to generate UMAP plots to show **(i)** cell type annotation as according to the original studies, or expression profiles of **(ii)** LMNTD1 or **(iii)** MROH9. **(C)** Heatmap showing significance p-values for similarity scores, calculated using a hypergeometric test, between: **(i)** all predicted cell type enriched genes, and **(ii)** lung macrophages vs. other non-macrophage cell types.

Paper III

Adhesion G protein-coupled receptor L4 modifies the endothelial cell response to tumour necrosis factor

M Norreen-Thorsen[1], L Gottlob[2], JB Kral-Pointner, S Öling[1], E Struck[1], J Odeberg[1,2,3,4], C Naudin[1,2], P Dusart[2,6], LM Butler [1,2,5*]

[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 Haematology, Karolinska University Hospital, 171 76 Stockholm, Sweden

[5] Clinical Chemistry and Blood Coagulation Research, Department of Molecular Medicine and Surgery, Karolinska Institute, 171 76 Stockholm, Sweden, *and* Clinical Chemistry, Karolinska University Laboratory, Karolinska University Hospital, 171 76 Stockholm, Sweden

[*] **Lead contact/corresponding author**:
Dr. L.M Butler, PhD
Email: Lynn.butler@ki.se or lynn.m.butler@uit.no

Key words: ADGRL4, protein profiling, HUVEC, endothelial cell, TNF-activation, tissue factor, interferon, inflammatory response

SUMMARY

Endothelial cells (EC), which form a monolayer lining the lumen of all blood vessels, have a key role in the regulation of haemostasis, blood pressure, inflammatory responses, the transportation of fluids and molecules across tissues, and angiogenesis. The endothelial-restricted expression of the highly conserved adhesion G protein-coupled receptor L4 (ADGRL4) indicates it likely has an important role in EC specific function. Previous reports indicate a role for ADGRL4 in EC angiogenesis and proliferation, but studies have failed to show any activation or coupling to canonical GPCR signalling pathways. Here, using a global proteomics profiling approach, we show that ADGRL4 depletion modifies the EC response to the inflammatory cytokine tumour necrosis factor (TNF), specifically augmenting TNF-induced tissue factor expression. Time course analysis indicated that initial transcription of tissue factor mRNA was not affected by ADGRL4 depletion, but subsequent resolution to baseline was inhibited. Both thrombin formation and fibrin deposition from plasma and whole blood, respectively, was correspondingly potentiated following ADGRL4-depletion, demonstrating an associated increase in pro-coagulant activity. Concurrently, expression of a panel of TNF-induced interferon stimulated genes were selectively inhibited by ADGRL4-depletion. Our results indicate that ADGRL4 has a currently unappreciated and potentially complex role in the EC response during inflammation.

INTRODUCTION

The endothelium, a cellular monolayer lining the lumen of all blood vessels, has a key role in the regulation of haemostasis, blood pressure, inflammatory responses, the transportation of fluids and molecules across tissues, and angiogenesis [1-3]. We previously identified a panel of human endothelial cell (EC) enriched genes in multiple tissue types [4-9], and these genes are likely important for EC specific phenotype and function. Proteins coded by such genes include cadherin-5 (CDH5), claudin-5 (CLDN5) and endothelial cell-selective adhesion molecule (ESAM), which play established roles in EC integrity, polarity and shape, vessel permeability and signalling [10-13]**,** von Willebrand factor (VWF) which is important for haemostasis [14], and the vascular endothelial growth factor receptor 1 (FLT1) and 2 (KDR), which are central to angiogenesis [15]. However, we also identified many endothelial enriched genes that encoded for currently poorly characterised proteins [4-8], including the adhesion G protein-coupled receptor L4 (ADGRL4). ADGRL4, formerly known as ELTD1, belongs to the latrophilin group that, along with eight other subfamilies, make up the adhesion G protein-coupled receptor (aGPCR) family. The topological compartmentation of ADGRL4 separates into a canonical 7-transmembrane domain, shared by all GPCRs, a short intracellular C-terminal domain and an extracellular domain consisting of an epidermal growth factor (EGF)-like domain, a calcium binding EGF-like domain and the aGPCR specific GPCR autoproteolysis-inducing domain. The latter contains a GPCR proteolytic site where cleavage is predicted to create a N-terminal and C-terminal fragment that associates non-covalently [16]. ADGRL4 is well conserved amongst vertebrates [17], implying an important functional role. No binding partner for ADGRL4 has been identified, rendering investigation of protein activation and signal-transduction challenging [16], with previous studies failing to show any activation or coupling to any canonical GPCR signalling pathway, even though similar methods have proven successful for other aGPCRs [18]. Despite this, studies have identified a potential role for ADGRL4 in EC angiogenesis and proliferation. ADGRL4 likely has a role in the NOTCH signalling pathway; silencing in cultured EC causes upregulation of *DLL4* and downregulation of *JAG1,* while the opposite is true for ADGRL4 overexpression [18, 19]. ADGRL4 itself is upregulated by VEGF and downregulated by DLL4 [20]. Overexpression of ADGRL4 in EC leads to induction of sprouting angiogenesis and reduced proliferation [18], while silencing reduced sprouting and tip-formation [20]. Another study found that EC overexpression of ADGRL4 induced endothelial-mesenchymal transition [21], with a loss of cell-cell contacts, increased network formation, stress-fibre production and endothelial sprouting. A role for ADGRL4 has been implicated in several cancer types, with upregulation in tumour cells or the surrounding vessels, with contrasting associations to prognosis and survival [20, 22-27].

Here, we investigated the effect of EC ADGRL4-depletion on the global protein expression profile. We observed a modest increase in the expression of EC proteins associated with cellular morphology and structure and marked suppression of cell-cycle associated proteins. When EC were treated with the inflammatory cytokine tumour necrosis factor (TNF), the associated increase in tissue factor expression was strongly potentiated by ADGRL4-depletion, and cellular coagulability correspondingly elevated, as measured by real-time thrombin formation and fibrin deposition. Concurrently, expression of a panel of TNF-induced interferon stimulated genes were selectively inhibited by ADGRL4-depletion. Our results indicate that ADGRL4 has a currently unappreciated and potentially complex role in the EC response during inflammation.

RESULTS

ADGRL4 is an endothelial enriched gene across human tissue types

Using a bioinformatics based integrative correlation analysis of bulk RNA sequencing samples, we predicted *ADGRL4* to be an EC enriched gene across tissue types [4-7, 9]. To verify our predictions, we sourced single cell transcriptomics (scRNAseq) data from the analysis of 24 human organs in Tabula Sapiens [28], where cells from all organs were grouped into the major cell type compartments (Figure 1A). *ADGRL4* transcripts were consistently detected in the EC compartment (Figure 1Aiv), but not in the immune- (Figure 1Ai), epithelial- (Figure 1Aii)- or stromal- (Figure 1Aiii) cell compartments. Thus, scRNAseq data supports our predictions that *ADGRL4* has an EC enriched expression profile, indicating a potential role for *ADGRL4* in cell type specific function.

Mass spectrometry-based analysis of *ADGRL4*-depleted unstimulated EC

Primary EC were isolated from human umbilical cords and transfected with siRNA (Assay ID: 6325) targeting *ADGRL4* (siRNA-*ADGRL4*-EC), or a scrambled control (siRNA-control-EC), before collection for subsequent analysis 72h later (n=3 biological replicates, annotated as sample sets [SS] 1,2,3). We used shotgun mass spectrometry to measure global EC protein profiles and to calculate the fold change (FC) in detected proteins between siRNA-*ADGRL4*-EC and siRNA-control-EC.

ADGRL4 silencing modifies the EC expression of structural and mitosis-associated proteins

A total of 92 proteins were upregulated in unstimulated (´resting´) siRNA-*ADGRL4*-EC (threshold FC ≥ 1.3 in all sample sets) vs. siRNA-control-EC (Figure 1Bi; Figure S2Ai-ii; Table S1, Tab 3). Among the most consistently up regulated proteins were transgelin (TAGLN) (FC [siRNA-*ADGRL4*-EC *vs.* siRNA-control-EC, SS 1,2,3]: 3.99, 3.54, and 4.06), protein phosphatase 1 regulatory subunit 14A (PPP1R14A) (FC: 4.82, 2.67, and 3.40) and tetraspanin-2 (TSPAN2) (FC: 2.61, 2.24 and 3.01). We performed gene ontology (GO) enrichment analysis [29-31] (biological processes) to identify overrepresented terms among these upregulated proteins, which included '*cell adhesion*' ($p=1.5x10^{-10}$), '*wound healing*' ($p=1.3x10^{-7}$), '*actin filament bundle assembly*' ($p=8.0x10^{-7}$), '*blood vessel morphogenesis*' ($p = 4.2x10^{-6}$) and '*regulation of cell motility*' ($p = 9.1x10^{-6}$) (Figure 1Ci) (for full list, see Table S1, Tab 4).

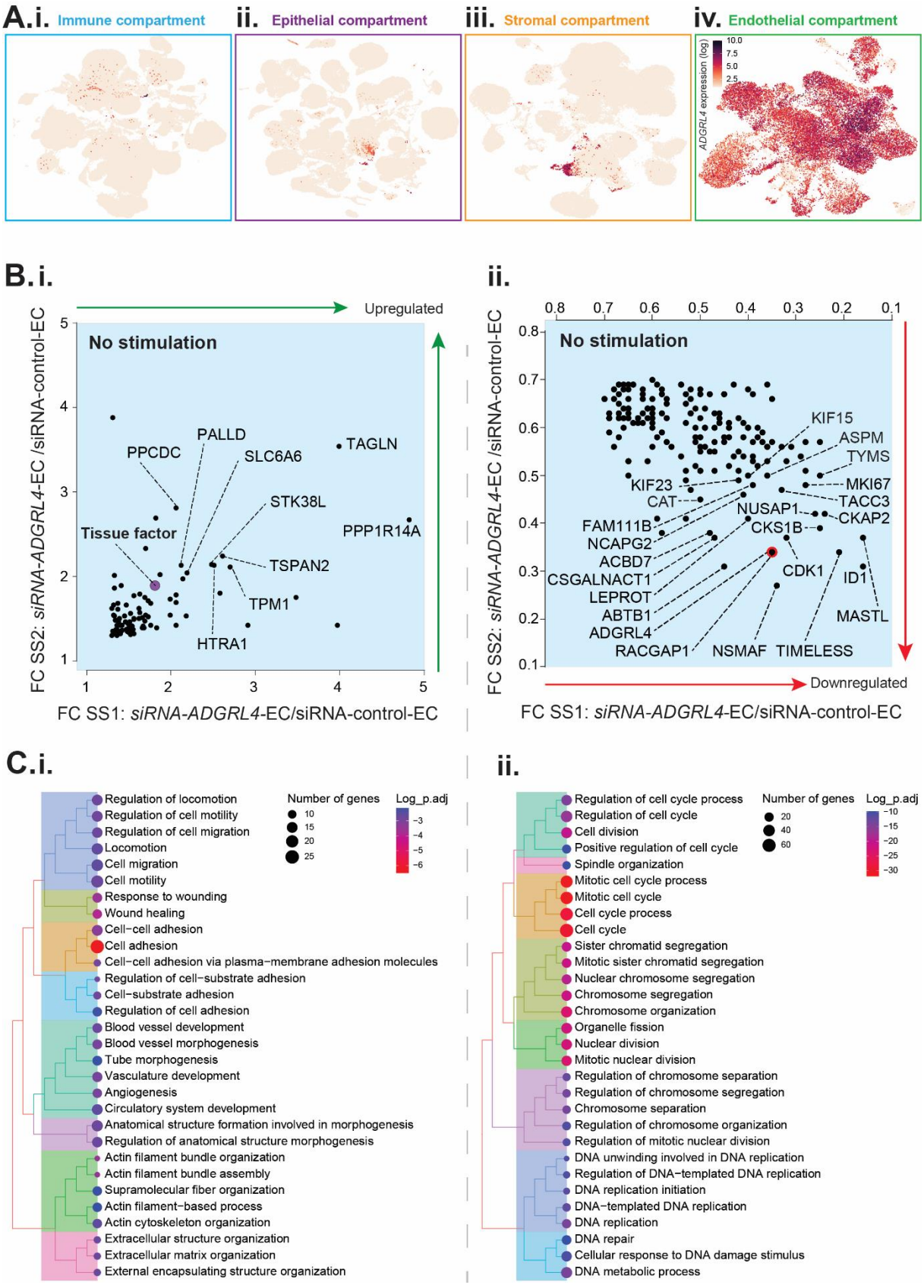


Figure 1. Endothelial *ADGRL4* depletion causes changes in structural and cell cycle related proteins. (A) scRNAseq data from analysis of cell types from 24 human tissues was sourced from Tabula Sapiens (Tabula Sapiens., 2021 [28]) and used to generate UMAP plots to show *ADGRL4*

expression in **(i)** immune-, **(ii)** epithelial-, **(iii)** stromal, and **(iv)** endothelial cell compartments (See Figure S1 for cluster annotations). **(B-C)** Human umbilical vein endothelial cells (HUVEC, n=3 independent sample sets [SS 1,2,3]) were transfected with siRNA targeting *ADGRL4* (siRNA-*ADGRL4*-EC) or a scrambled-control sequence (siRNA-control-EC) and analysed after 72 hours, using mass spectrometry. **(B)** Plots showing **(i)** upregulated or **(ii)** downregulated proteins in siRNA-*ADGRL4*-EC (displayed on the plot as a fold change (FC) vs. siRNA-control-EC, for 2 biological replicates - sample set [SS] 1 and 2). Proteins with ratio >1.3 (upregulated) or <0.7 (downregulated) across all sample sets are displayed. **(C)** Gene ontology (biological processes) overrepresented terms for the: **(i)** upregulated or **(ii)** downregulated proteins.

A total of 177 proteins were downregulated in unstimulated (´resting´) siRNA-*ADGRL4*-EC (threshold FC ≤ 0.7 in all sample sets) vs. siRNA-control-EC (Figure 1 Bii; Figure S2 Aiii-iv; Table S1, Tab 5). Among the most consistently down regulated proteins were timeless homolog (TIMELESS) (FC [siRNA-*ADGRL4*-EC vs. siRNA-control-EC, SS 1,2,3]: 0.21, 0.34 and 0.32), DNA-binding protein inhibitor ID-1 (ID1) (FC: 0.16, 0.31 and 0.40), serine/threonine-protein kinase greatwall (MASTL) (FC; 0.16, 0.37 and 0.35), cyclin-dependent kinases regulatory subunit 1 (CKS1B) (FC: 0.25, 0.39 and 0.35), cytoskeleton-associated protein 2 (CKAP2) (FC: 0.24, 0.42 and 0.34), nucleolar and spindle-associated protein 1 (NUSAP1) (FC: 0.26, 0.42 and 0.37), cyclin-dependent kinase 1 (CDK1) (FC: 0.32, 0.37 and 0.35) and Rac GTPase-activating protein 1 (RACGAP1) (FC: 0.35, 0.34 and 0.37). We performed GO enrichment analysis (biological processes) to identify over represented terms among the down regulated proteins, which included ‘mitotic cell cycle’ (p=6.2x10-36), ‘chromosome organization’ (p=3.9x10-27), ‘nuclear division’ (p=1.1x10-25), ‘chromosome segregation’ (p=1.1x10-24) and ‘DNA replication’ (p = 5.4x10-17) (Figure 1Cii; full list in Table S1, Tab 6). These results indicate a potential role for ADGRL4 in EC cell-cycle regulation.

Mass spectrometry-based analysis of *ADGRL4*-depleted inflammatory cytokine-stimulated EC

EC play a key role in inflammation and haemostasis and a variety of cytokines can induce EC activation in response to tissue damage or infection, leading to changes in the transcriptome and proteome profile [32-37]. To investigate if ADGRL4 has a potential role in these processes, we analysed global EC protein profiles in siRNA-*ADGRL4*-EC and siRNA-control-EC, following stimulation with the inflammatory cytokine tumour necrosis factor (TNF) for 24 hours (n=2 biological replicates, annotated as sample sets [SS] 4 and 5). As above, we used shotgun mass spectrometry to measure global EC protein profiles and to calculate the fold change (FC) in detected proteins between TNF-stimulated siRNA-control-EC and siRNA-*ADGRL4*-EC.

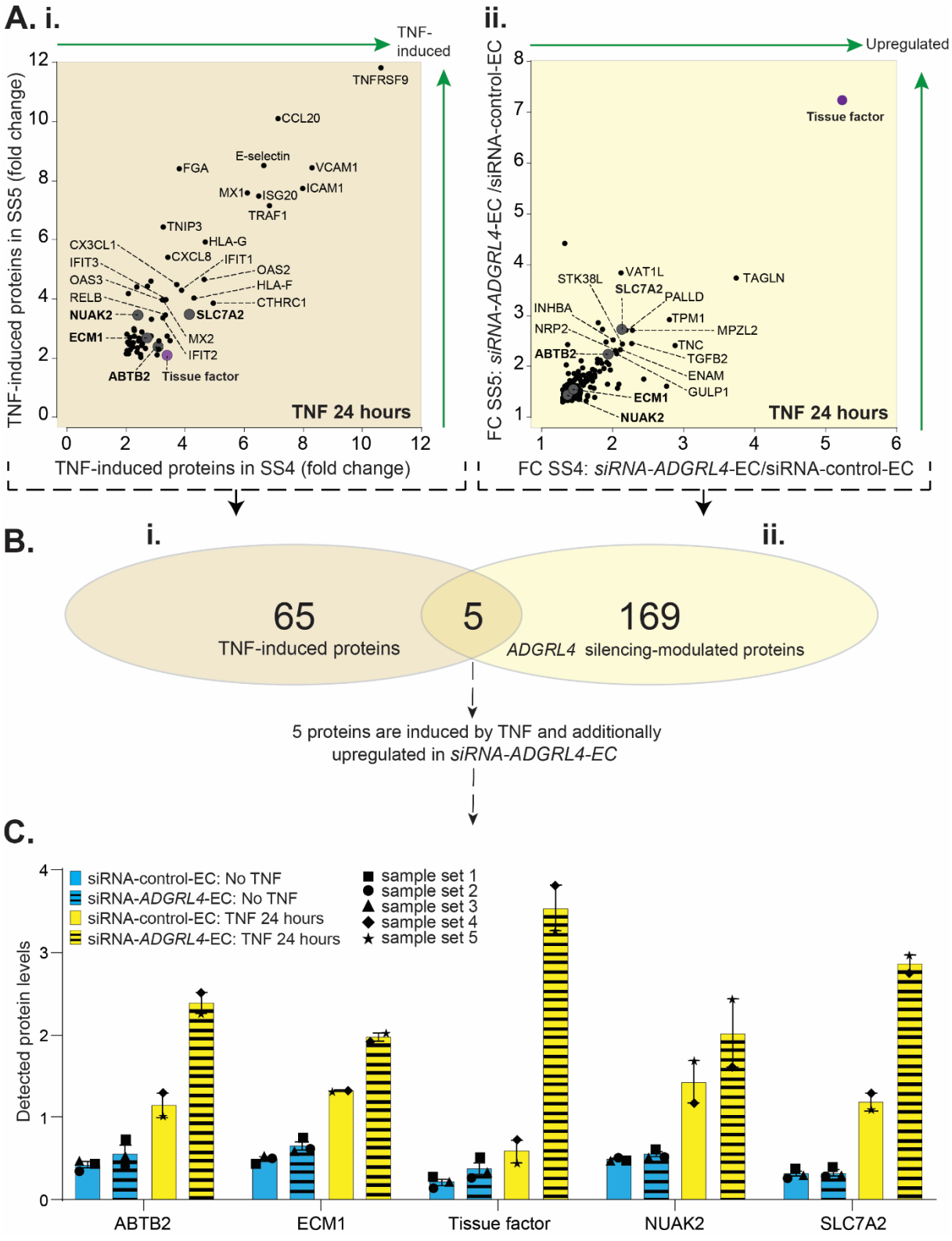


Figure 2. Endothelial *ADGRL4* depletion potentiates TNF-induced tissue factor expression. siRNA-*ADGRL4*-EC or siRNA-control-EC were treated with or without TNF for 24 hours before analysis by mass spectrometry. (**A**) Plots showing proteins: **(i)** induced by TNF (siRNA-control-EC + TNF vs. siRNA-control-EC unstimulated = >2 fold change) or **(ii)** upregulated by TNF *and* potentiated in siRNA-*ADGRL4*-EC (fold change siRNA-*ADGRL4*-EC + TNF vs. siRNA-control-EC + TNF = >1.3 fold). (**B**) Venn-diagram showing overlap between (**i**) the number of proteins induced by TNF and (**ii**) the number

of proteins potentiated by *ADGRL4* siRNA. The five proteins that were induced by TNF and potentiated in siRNA-*ADGRL4*-EC are labelled in bold on plots in (**A**), and data points for expression in individual samples shown in (**C**). Error bars represent standard error of the mean.

TNF-induced tissue factor expression is potentiated by ADGRL4 silencing

TNF treatment of siRNA-control-EC induced the expression of 70 proteins, which included vascular cell adhesion protein 1 (VCAM1) and E-selectin (SELE) (Figure 2Ai and Table S1, Tab 7), both leukocyte adhesion receptor proteins known to be induced by TNF [38, 39]. GO enrichment analysis was performed to identify overrepresented terms among TNF-induced proteins (FC ≥ 2, TNF siRNA-control-EC vs. unstimulated siRNA-control-EC), which included '*immune response*' ($p=1.2x10^{-29}$), '*response to cytokine*' ($p=5.8x10^{-28}$), '*defense response*' ($p=1.3x10^{-22}$) and '*cytokine-mediated signaling pathway*' ($p=4.2x10^{-16}$) (Table S1, Tab 8). We identified 174 proteins expressed at higher levels in TNF-stimulated siRNA-*ADGRL4*-EC, compared to TNF-stimulated siRNA-control-EC (FC ≥1.3 in both sample sets [SS] 4 and 5) (Figure 2Aii and Table S1, Tab 9), the majority of which were not TNF-responsive. Indeed, GO enrichment analysis returned overrepresented terms including '*cell migration*' ($p=2.1x10^{-17}$), '*cell motility*' ($p=1.3x10^{-16}$), '*cell adhesion*' ($p=6.1x10^{-16}$) and *'wound healing*' ($p=1.6x10^{-14}$) (Table S1, Tab 10), similar to that reported for proteins upregulated in unstimulated siRNA-*ADGRL4*-EC vs. siRNA-control-EC (Figure 1Ci). Only five TNF-induced proteins were upregulated in siRNA-*ADGRL4*-EC vs. siRNA-control-EC (Figure 2B); with the most markedly being tissue factor, the initiator of the extrinsic pathway of coagulation (FC [TNF-treated siRNA-*ADGRL4*-EC *vs.* TNF-treated siRNA-control-EC, SS 4 and 5]: 5.23 and 7.24). The others were more modestly potentiated - NUAK family SNF1-like kinase 2 (NUAK2) (FC: 1.37 and 1.44), cationic amino acid transporter 2 (SLC7A2) (FC: 2.13 and 2.73), ankyrin repeat and BTB/POZ domain-containing protein 2 (ABTB2) (FC: 1.94 and 2.24) and extracellular matrix protein 1 (ECM1) (FC:1.45 and 1.55) (Figure 2Aii and 2C; Table S1, Tab 9). Thus, of all TNF-induced proteins, EC tissue factor expression was most significantly and selectively potentiated by *ADGRL4* depletion.

ADGRL4 silencing inhibits the resolution of tissue factor after TNF-induction

To further explore the effect of ADGRL4 silencing on EC tissue factor expression, we measured mRNA and surface protein expression, using an alternative sequence *ADGRL4* siRNA (Assay ID: s34480, referred to as siRNA-*ADGRL4*^{1}-EC), in addition to the original (here referred to as siRNA^{2}-*ADGRL4*-EC), to rule out potential off target effects. Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*^{1}-EC, or siRNA^{2}-ADGRL4-EC were treated with or without TNF for 4, 24 or 48 hours. Knock-down of *ADGRL4* in siRNA-*ADGRL4*^{1}-EC and

siRNA[2]-*ADGRL4*-EC was confirmed on both the mRNA (Figure S2B) and protein level (Figure S2C).

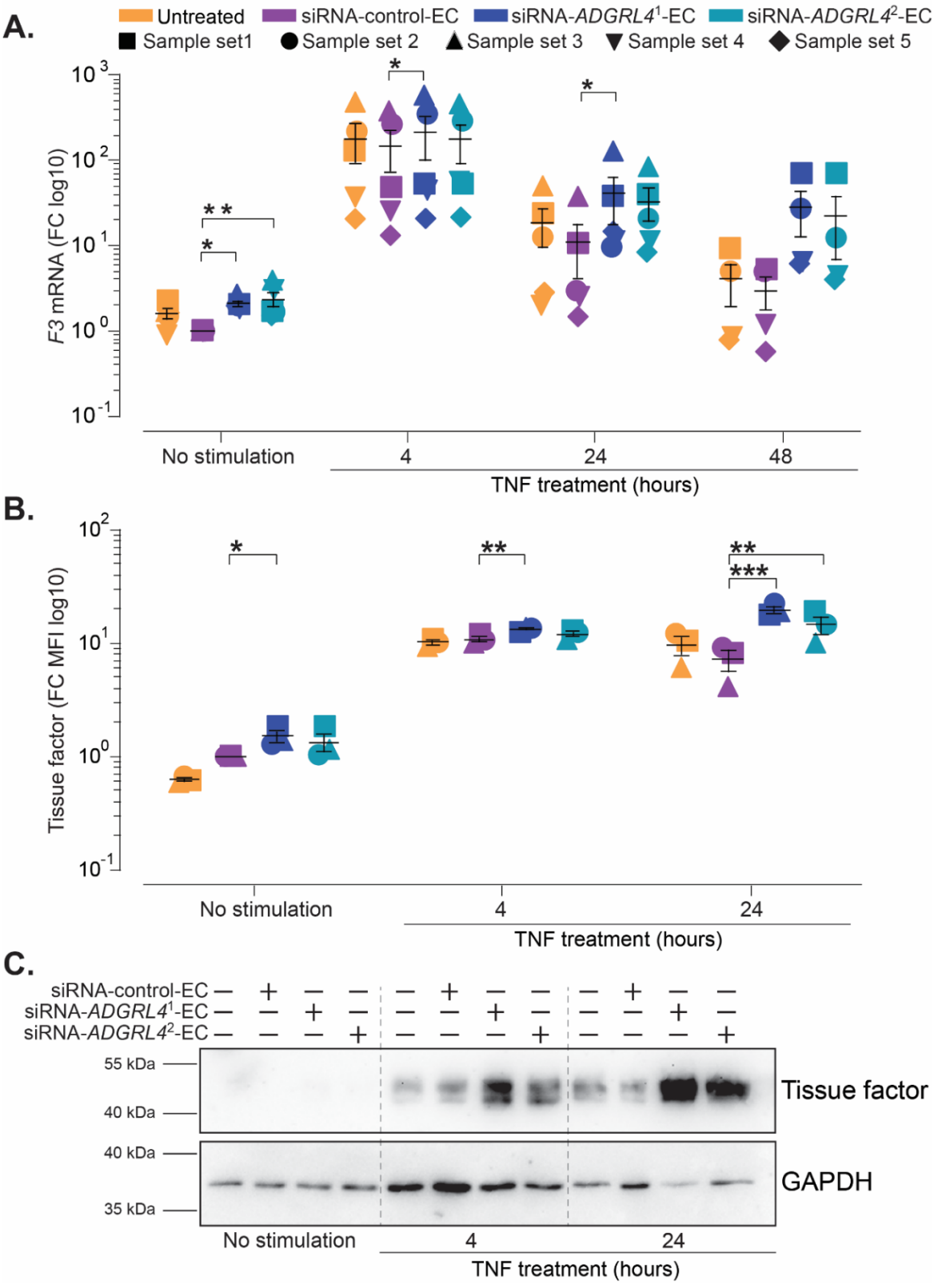


Figure 3. Endothelial *ADGRL4* depletion leads to reduced levels of TNF-induced tissue factor during the later phase of the response. Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*[1]-

EC, or siRNA[2]-*ADGRL4*-EC were treated with or without TNF for 4, 24, and 48 hours, before samples before collection and measurement of **(A)** relative *F3* mRNA (fold change (FC) vs. unstimulated siRNA-control-EC)) or tissue factor protein **(B)** on the cell surface by flow cytometry, displayed as mean fluorescence intensity (MFI) relative to untreated siRNA-control-EC or **(C)** in the entire cell lysate by western blot analysis. Mean values with SEM are indicated on graphs *p<0.05, **p<0.01, ***p<0.001 by one-way ANOVA for each time-point.

Relative qPCR showed tissue factor mRNA expression in non-transfected and siRNA-control-EC was strongly induced by 4 hours post-TNF stimulation, with levels dropping by 24h, and further again by 48 hours (Figure 3A). Relative qPCR showed significant upregulation of tissue factor mRNA expression in both siRNA[1]-*ADGRL4*-EC and siRNA[2]-*ADGRL4*-EC relative to siRNA-control-EC at baseline only (siRNA[1]-*ADGRL4*-EC p=0.0017, siRNA[2]-*ADGRL4*-EC p=0.0458), although a clear downregulatory trend also was observed for 24 and 48 hours of TNF stimulation. Little or no difference in upregulation was observed at 4 hours post TNF-treatment (Figure 3A).

TNF-induced EC surface tissue factor expression, measured by flow cytometry, was significantly elevated following *ADGRL4* silencing at 24 hours (siRNA-*ADGRL4*[1]-EC p=0.001, siRNA[2]-*ADGRL4*-EC p=0.009) (Figure 3B; Figure S3A). Under baseline conditions, and at 4 hours post TNF-stimulation, similar trends were observed, but differences were not consistently statistically significant in both siRNA-*ADGRL4*[1]-EC and siRNA[2]-*ADGRL4*-EC. Tissue factor protein measurement in whole cell lysates (Figure 3C, Figure S3B) showed results consistent with cell surface levels. Taken together, these results suggest that initial transcription of tissue factor mRNA was not affected by ADGRL4 depletion, but subsequent resolution to baseline levels was inhibited.

Silencing of endothelial *ADGRL4* enhances tissue factor induced thrombin generation

To test for functional effects of elevated TNF-induced tissue factor expression we observed on siRNA-*ADGRL4*[1]-EC and siRNA[2]-*ADGRL4*-EC, we used an in-house developed assay to measure real time thrombin generation in plasma (Figure 4, and Figure S4 Ai-iv for assay description). Recombinant tissue factor was used as an internal reference and positive control, which strongly induced thrombin generation (Figure 4 Ai-iii).

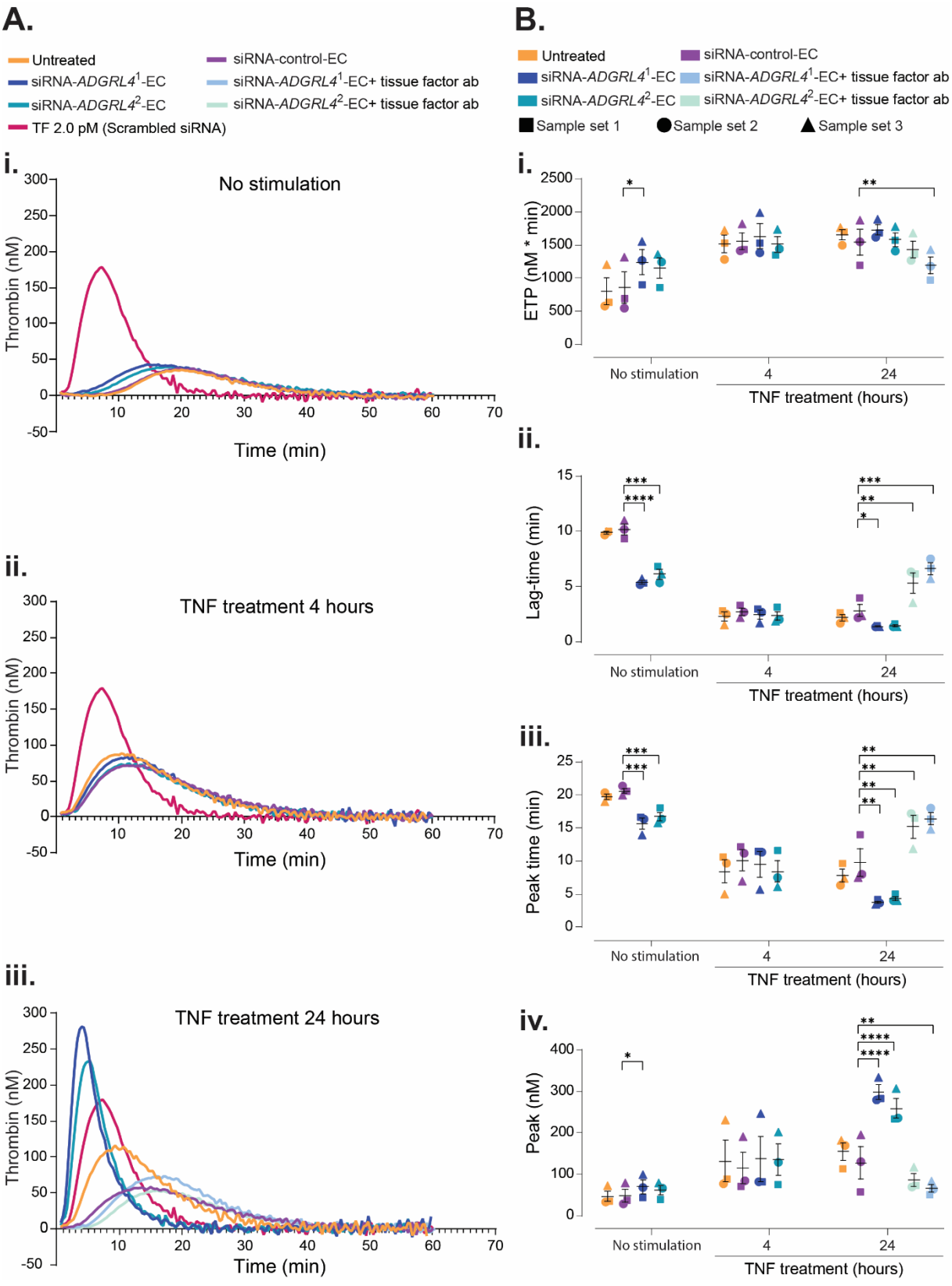

A.
Untreated
siRNA-ADGRL4¹-EC
siRNA-ADGRL4²-EC
TF 2.0 pM (Scrambled siRNA)
siRNA-control-EC
siRNA-ADGRL4¹-EC+ tissue factor ab
siRNA-ADGRL4²-EC+ tissue factor ab
i.
No stimulation
Thrombin (nM)
Time (min)
ii.
TNF treatment 4 hours
Thrombin (nM)
Time (min)
iii.
TNF treatment 24 hours
Thrombin (nM)
Time (min)
B.
Untreated
siRNA-ADGRL4¹-EC
siRNA-ADGRL4²-EC
siRNA-control-EC
siRNA-ADGRL4¹-EC+ tissue factor ab
siRNA-ADGRL4²-EC+ tissue factor ab
Sample set 1
Sample set 2
Sample set 3
i.
ETP (nM * min)
No stimulation
4
24
TNF treatment (hours)
ii.
Lag-time (min)
No stimulation
4
24
TNF treatment (hours)
iii.
Peak time (min)
No stimulation
4
24
TNF treatment (hours)
iv.
Peak (nM)
No stimulation
4
24
TNF treatment (hours)

Figure 4. Endothelial *ADGRL4* depletion enhances tissue factor induced thrombin generation. Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*^{1}-EC, or siRNA^{2}-*ADGRL4*-EC were treated with or without TNF for 4 or 24 hours. Tissue factor blocking antibody or recombinant tissue factor were added to some wells prior to assay. **(A)** Thrombin generation in human plasma was measured after **(i)** 0, **(ii)** 4 or **(iii)** 24 hours of TNF stimulation (plots representative of 3 experiments). **(B)** Individual data points for each experiment showing **(i)** endogenous thrombin potential (ETP), **(ii)** lag-time, **(iii)** peak-time and **(iv)** peak for all conditions. Mean values with SEM are indicated on graphs *p<0.05, **p<0.01, ***p<0.001, ****p<00001 by one-way ANOVA for each time-point.

The resultant thrombin generation curve was used for comparison of thrombin generation efficiency in the absence (Figure 4Ai), or presence, of prior TNF-stimulation, for 4 or 24 hours (Figure 4Aii-iii) (Figure S4Bi-iii for other replicates). Under unstimulated conditions, thrombin generation was modestly enhanced on siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC, compared to siRNA-control-EC, measured by lag-time (mean ± SEM; siRNA-control-EC [10.17±0.48 mins] vs. siRNA^{1}-*ADGRL4*-EC [5.39±0.15 mins, p<0.0001], or vs. siRNA^{2}-*ADGRL4*-EC [6.11±0.43 mins, p=0.0001]) (Figure 4Bii) and peak-time (mean ± SEM; siRNA-control-EC [20.61±0.39 mins] vs. siRNA^{1}-*ADGRL4*-EC [15.67±0.84 mins, p<0.0001], or vs. siRNA^{2}-*ADGRL4*-EC [16.72±0.66 mins, p=0.0005]) (Figure 4Biii). Thrombin generation under all conditions was enhanced following 4 hours of TNF-stimulation (Figure 4Aii), with no significant difference in endogenous thrombin potential (ETP), lag-time, peak-time or peak hight (Figure 4B) between siRNA-control-EC and siRNA^{1}-*ADGRL4*-EC or siRNA^{2}-*ADGRL4*-EC. Following 24 hours of TNF-stimulation, thrombin generation for both siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC was markedly enhanced, compared to siRNA-control-EC (Figure 4Aiii) as indicated for peak-time (mean ± SEM; siRNA-control-EC [9.83±2.09 mins] vs. siRNA^{1}-*ADGRL4*-EC [3.77±0.19 mins, p=0.0019], or vs. siRNA^{2}-*ADGRL4*-EC [4.33±0.33 mins, p=0.0036]) (Figure 4Bii) and peak hight (mean ± SEM; siRNA-control-EC [127.00±39.82 nM] vs. siRNA^{1}-*ADGRL4*-EC [298.70±17.46 nM, p<0.0001), or vs. siRNA^{2}-*ADGRL4*-EC [258.90±23.81 nM, p<0.0001]) (Figure 4Biv). When siRNA^{1}-*ADGRL4*-EC or siRNA^{2}-*ADGRL4*-EC were pre-treated with tissue factor function blocking antibodies, these effects were largely abolished (Figure 4Aiii and 4Bi-iv and Figure S3B), consistent with tissue factor being the driver behind the increased thrombin generation on siRNA-*ADGRL4*^{1}-EC and siRNA^{2}-*ADGRL4*-EC.

Silencing of endothelial *ADGRL4* enhances tissue factor induced fibrin deposition

As the observed increase in tissue factor corresponded with an increase in thrombin generation, we went on to investigate the impact on downstream fibrin deposition, which occurs as a result of activated thrombin cleaving fibrinogen into fibrin, whilst concurrently activating coagulation factor XIII, which cross-links fibrin fibres to create a stable mesh [40].

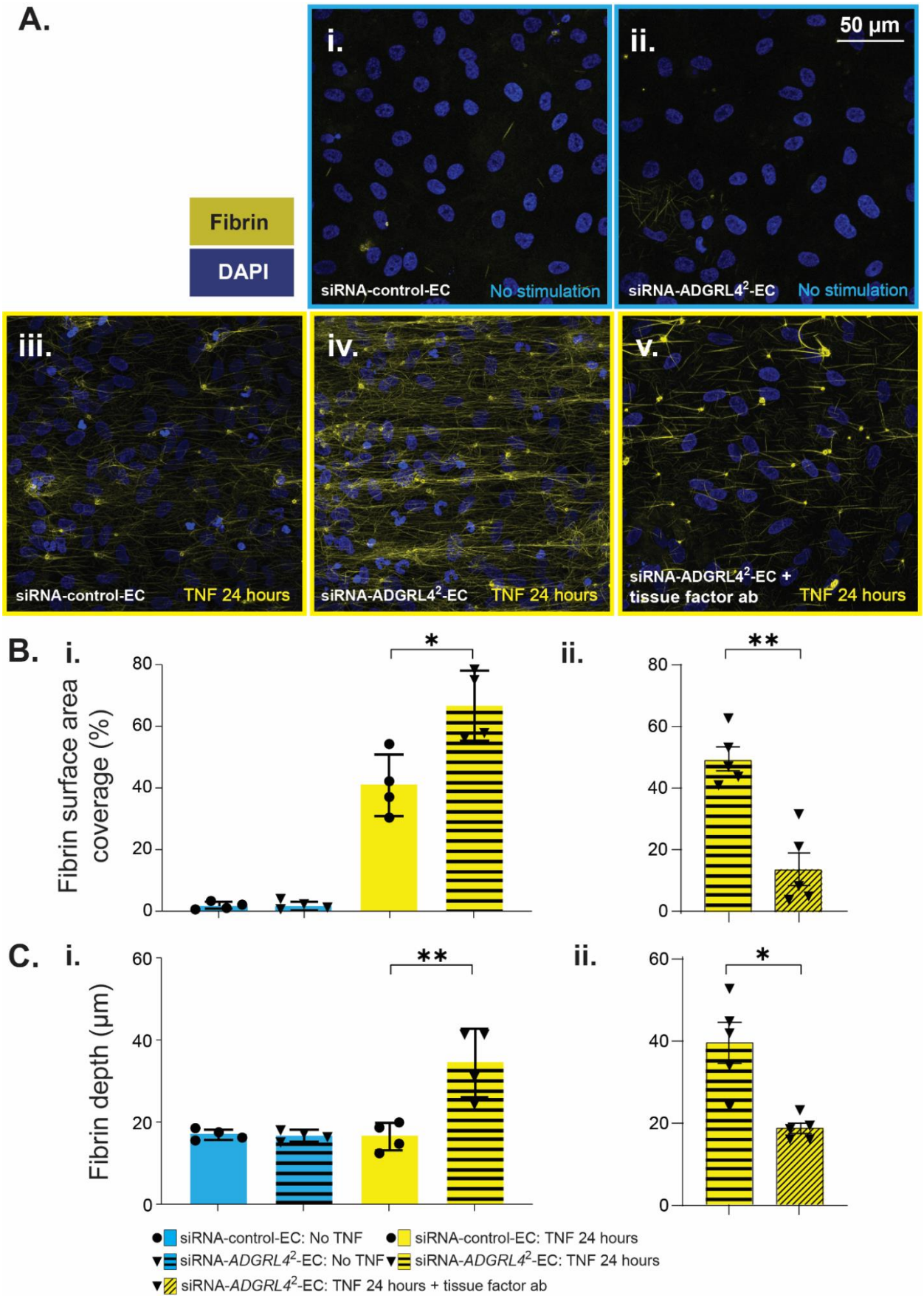


Figure 5: Endothelial *ADGRL4* depletion enhances tissue factor-induced fibrin deposition. siRNA^2-*ADGRL4*-EC or siRNA-control-EC were grown to confluence in Ibidi chamber slides and treated

with or without TNF for 24 hours. Re-calcified citrated human whole blood was flowed over the EC surface at a shear rate of 1 dyn/cm^2 for 10 minutes. **(A)** Representative immunofluorescence images of fibrin deposition on unstimulated: **(i)** siRNA-control-EC and **(ii)** siRNA^2-*ADGRL4*-EC, and TNF stimulated: **(iii)** siRNA-control-EC, **(iv)** siRNA^2-*ADGRL4*-EC and **(v)** siRNA^2-*ADGRL4*-EC pre-treated with tissue factor blocking antibody. Fibrin **(B)** surface area coverage and **(C)** depth was measured **(i)** with and without TNF stimulation and **(ii)** with and without tissue factor function blocking antibody treatment, prior to assay. Mean values with SEM are indicated on graphs *p<0.05, **p<0.01, by paired two-tailed t-test.

We used an in-house developed fibrin deposition assay. Briefly, siRNA-control-EC or siRNA^2-*ADGRL4*-EC were cultured in narrow channelled chamber slides before treatment with or without TNF for 24 hours. The channel was connected, via silicon tubing, to a syringe pump containing anticoagulated whole human blood, which was passed over the cell surface at a consistent flow rate. Recalcification buffer was injected and mixed into the blood just prior to entry into the channel (to remove the anticoagulant induced inhibition of the coagulation cascade). Under non-TNF stimulated conditions, fibrin deposition was low on both siRNA-control-EC (Figure 5Ai) and siRNA^2-*ADGRL4*-EC (Figure 5Ai), with no difference in surface area coverage (Figure 5Bi) or depth (Figure 5Ci). 24 hr post TNF stimulation, fibrin deposition under both conditions was increased (Figure 5Aiii and iv), but fibrin surface area (mean ± SEM; siRNA-control-EC [40.89±5.05 %] vs. siRNA^2-*ADGRL4*-EC [66.75±5.70 %, p=0.019]) and depth (mean ± SEM; siRNA-control-EC [16.53±1.74 μm] vs. siRNA^2-*ADGRL4*-EC [34.54±4.25 μm, p=0.0094]) was significantly enhanced on siRNA^2-*ADGRL4*-EC (Figure 5 Bi and Ci). Pre-treatment of siRNA^2-*ADGRL4*-EC with tissue factor blocking antibody (TF-ab) caused a significant reduction in fibrin surface area (mean ± SEM; siRNA^2-*ADGRL4*-EC [49.52±3.86 %] vs. siRNA^2-*ADGRL4*-EC +TF-ab [13.87±5.41 %, p=0.0015]) (Figure 5Bii) and depth (mean±SEM; siRNA^2-*ADGRL4*-EC [39.39±4.87 μm] vs. siRNA^2-*ADGRL4*-EC +TF-ab [18.56±1.29 μm, p=0.0164]) (Figure 5Cii).

ADGRL4 depletion inhibits TNF-induced interferon pathway protein expression

When EC were stimulated with TNF, a total of 145 proteins were detected at lower levels in EC-siRNA-*ADGRL4,* relative to EC-siRNA-control (threshold FC ≤ 0.7 in all biological replicates) by mass spectrometry (Figure 6Ai; Table S1, Tab 11). By cross-comparison with the previously identified TNF-induced proteins (Figure 6Aii), 129 were non-TNF responsive proteins and, as for unstimulated siRNA-ADGRL4-EC vs. siRNA-control-EC, these were associated predominantly with cell cycle (Figure 6B) (Table S1, Tab 12). 16 were TNF-induced proteins that were inhibited in siRNA-ADGRL4-EC vs. siRNA-control-EC (Figure 6B; Table S1, Tab 11), which included interferon-induced GTP-binding protein Mx2 (MX2) (FC [siRNA-*ADGRL4*-EC *vs.* siRNA-control-EC, SS 4 and 5]: 0.44 and 0.31), interferon-induced protein with tetratricopeptide repeats 2 (IFIT2) (FC: 0.42 and 0.34), interferon-induced protein with

tetratricopeptide repeats 3 (IFIT3) (FC: 0.46 and 0.34), 2'-5'-oligoadenylate synthase 2 (OAS2) (FC 0.46 and 0.35) and interferon-induced protein with tetratricopeptide repeats 1 (IFIT1) (FC: 0.48 and 0.34) (Figure 6B and 6C; Table S1, Tab 11). Corresponding overrepresented GO terms included '*response to virus*' ($p = 4.96x10^{-10}$), '*response to other organism*' ($p = 3.19x10^{-9}$), '*negative regulation of viral genome replication*' ($p = 1.32x10^{-7}$) and '*response to interferon-alpha*' ($p = 5.62x10^{-7}$) (Figure 6D; Table S1, Tab 13). An automated PubMed search (performed using easypubmed package in R) for these 16 proteins revealed that, for 7 the majority of published papers were linked to interferon (search: 'gene name' + 'interferon'), and all had some existing reports in this context (Figure S5A).

To further verify the effect of ADGRL4 silencing on interferon response gene (ISG) expression, we measured mRNA levels of MX1, IFIT1, Interferon-stimulated gene 20 kDa protein (ISG20) and interferon regulatory factor 1 (IRF1), a transcription factor that can induce the aforementioned candidates [41], in non-transfected EC, siRNA^{1}-*ADGRL4*-EC, siRNA^{2}-*ADGRL4*-EC or siRNA-control-EC with or without pre-treatment with TNF for 2, 4, 24 or 48 hours. Selected stimulation time points for investigation were based on previously reported peak times for IRF1 (2hr) and ISGs (24hr) [42].

ISG20 mRNA was inhibited in both siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC, relative to siRNA-control-EC under unstimulated conditions (siRNA^{1}-*ADGRL4*-EC p=0.0014, siRNA^{2}-*ADGRL4*-EC p=0.0006) and at 24 (siRNA^{1}-*ADGRL4*-EC p=0.0004, siRNA^{2}-*ADGRL4*-EC p=0.018) and 48 hours (siRNA^{1}-*ADGRL4*-EC p=0.0068, siRNA^{2}-*ADGRL4*-EC p=0.018) post TNF-treatment (Figure 7Ai) Significant inhibition at 48 hours were found for TNF-induced *MX1* (siRNA^{1}-*ADGRL4*-EC p=0.022, siRNA^{2}-*ADGRL4*-EC p=0.018) and *IFIT1* (siRNA^{1}-*ADGRL4*-EC p=0.028, siRNA^{2}-*ADGRL4*-EC p=0.03), although 24 hours indicated the same trend (Figure 8Aii-iii).

There was no significant difference in *IRF1* mRNA expression for siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC relative to siRNA-control-EC at any time point (Figure 8Aiv). Consistent with the mass spectrometry and gene expression data, western blotting showed reduced TNF-induced MX1 and IFIT1 protein in both siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC, vs. siRNA-control-EC at later time points (Figure 7B), but no such observation was made for IRF1 protein (Figure 7C). These data indicate that ADGRL4 has a currently unappreciated role in EC TNF-induced interferon response gene regulation.

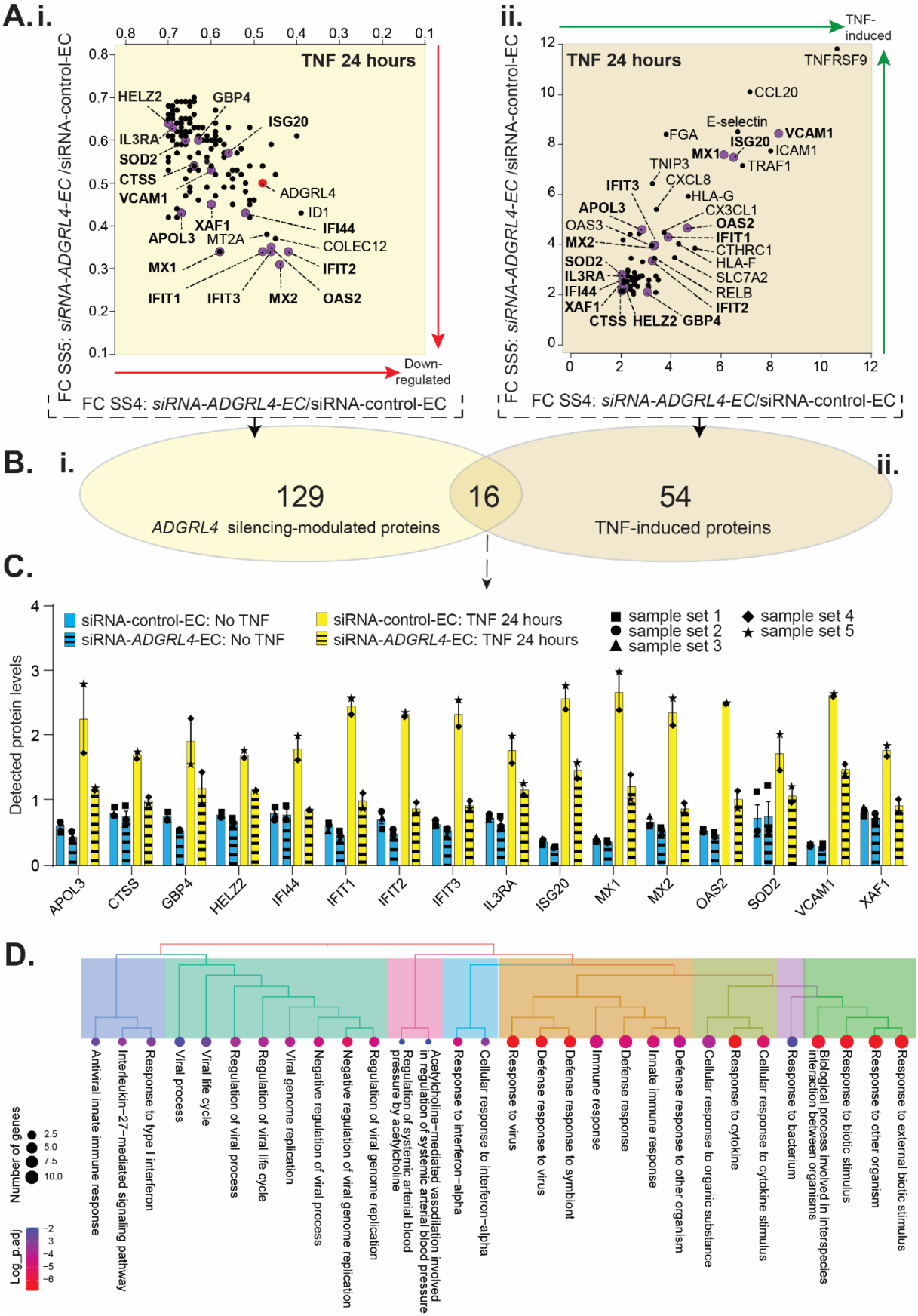


Figure 6. Endothelial *ADGRL4* depletion inhibits TNF-induced expression of interferon-related proteins. siRNA-*ADGRL4*-EC or siRNA-control-EC were treated with or without TNF for 24 hours before

analysis by mass spectrometry. **(A)** Plots showing proteins: **(i)** induced by TNF *but* downregulated in siRNA-*ADGRL4*-EC (fold change siRNA-*ADGRL4*-EC + TNF vs. siRNA-control-EC + TNF =<0.7 fold) or **(ii)** induced by TNF (siRNA-control-EC + TNF vs. siRNA-control-EC unstimulated = >2 fold change) The most outstanding proteins on each plot are labeled, the 16 proteins that were induced by TNF and downregulated in siRNA-ADGRL4-EC are labeled in bold and ADGRL4 is shown with a red dot for convenience. **(B)** Venn-diagram showing overlap between (**i**) the number of proteins downregulated by TNF and (**ii**) the number of proteins induced by *ADGRL4* siRNA. **(C)** Data for individual samples, error bars represent standard error of the mean. **(D)** Gene ontology (biological processes) overrepresented terms for the 16 downregulated proteins. Mean values with SEM for various conditions are indicated on graph.

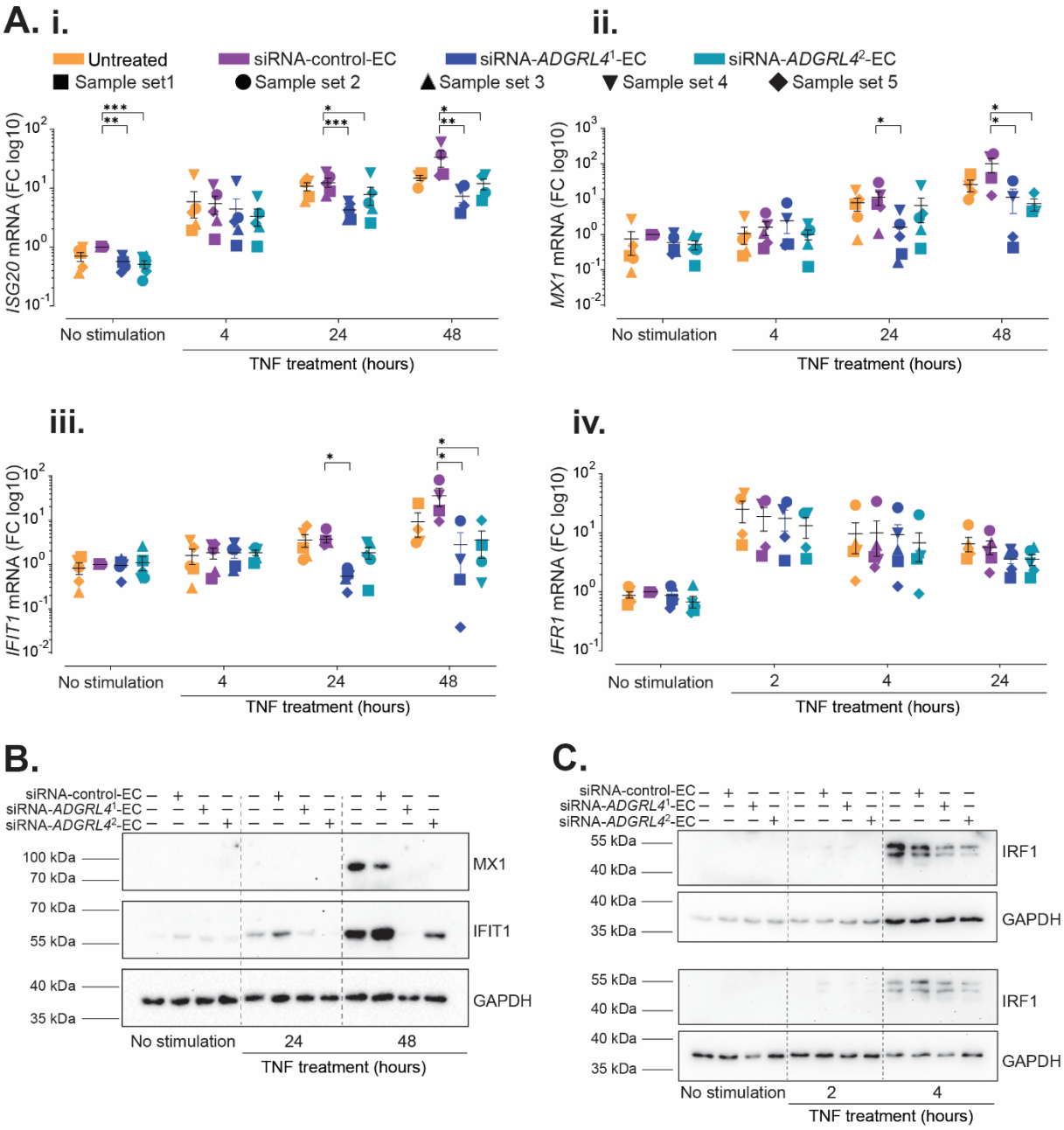


Figure 7. Endothelial *ADGRL4* depletion inhibits TNF-induced expression of interferon-related genes. Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*[1]-EC, or siRNA[2]-*ADGRL4*-EC were

treated with or without TNF. Samples were collected at 2, 4, 24 and 48 hours before measurement of **(A)** relative mRNA ((fold change [FC] vs. unstimulated siRNA-control-EC) for **(i)** *ISG20*, **(ii)** *MX1*, **(iii)** *IFIT1* and **(iv)** *IRF1*, or protein content of **(B)** MX1 and IFIT1 or **(C)** IRF1 in entire cell lysates by western blot analysis. Mean values with SEM are indicated on graphs *p<0.05, **p<0.01, ***p<0.001 by one-way ANOVA for each time-point.

ADGRL4 depletion inhibits TNF-induced leukocyte adhesion receptor expression

Following cytokine exposure EC enter an activated state, marked by induction of adhesion molecules important for leukocyte recruitment, such as VCAM1, ICAM1, and ESEL [32, 38, 39, 43]. As described above, mass spectrometry analysis showed a downregulation of VCAM1 protein expression after *ADGRL4* silencing (Figure 6Ai). Here, we measured the effect of *ADGRL4* silencing on mRNA levels and EC surface expression of VCAM1, together with E-selectin and ICAM1, which are similarly regulated by TNF-induced nuclear factor NF-kappa-B (NFKB) activation [44]. Non-transfected EC, siRNA^{1}-*ADGRL4*-EC, siRNA^{2}-*ADGRL4*-EC or siRNA-control-EC were treated with or without TNF for 4, 24 or 48 hours. TNF-induced *VCAM1* mRNA tended to be lower in siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC vs. siRNA-control-EC, although the observed differences were not statistically significant (Figure 8Ai).However, flow cytometry analysis (measured in unit of FC mean fluorescent intensity relative to unstimulated siRNA-control-EC) confirmed a pronounced and significant downregulation of VCAM1 following 24 hours of TNF-stimulation for both siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC (siRNA^{1}-*ADGRL4*-EC p=0.026, siRNA^{2}-*ADGRL4*-EC p=0.007) (Figure 8 Bi). Western blot analysis of whole cell lysates also showed consistent results (Figure 8C, Figure S5B)

There was no statistically significant difference in TNF-induced *SELE* mRNA expression across conditions (Figure 8Aii), but protein levels were downregulated in both siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC at baseline (siRNA^{1}-*ADGRL4*-EC p=0.001, siRNA^{2}-*ADGRL4*-EC p=0.0026), following TNF-stimulation for 4 hours (siRNA^{1}-*ADGRL4*-EC p=0.007, siRNA^{2}-*ADGRL4*-EC p=0.036) and 24 hours (siRNA^{1}-*ADGRL4*-EC p=0.006, siRNA^{2}-*ADGRL4*-EC p=0.015) (Figure 7Bii).

TNF-induced *ICAM1* mRNA (indicated in unit of FC relative to unstimulated siRNA-control-EC) was significantly downregulated in siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC following 24 hours of TNF-stimulation (siRNA^{1}-*ADGRL4*-EC p=0.014, siRNA^{2}-*ADGRL4*-EC p=0.004) and 48 hours (siRNA^{1}-*ADGRL4*-EC p=0.03, siRNA^{2}-*ADGRL4*-EC, p=0.008) (Figure 8Aiii). ICAM1 protein expression was not consistently affected by ADGRL4 depletion (Figure 8Biii). The observation that these adhesion molecules to a variable degree are affected by ADGRL4 silencing, and that their downregulatory effects are present earlier in the TNF-response,

relative to tissue factor which is also NFKB transcribed [45, 46], can indicate a regulatory pattern that might extend beyond changes in NFKB activation.

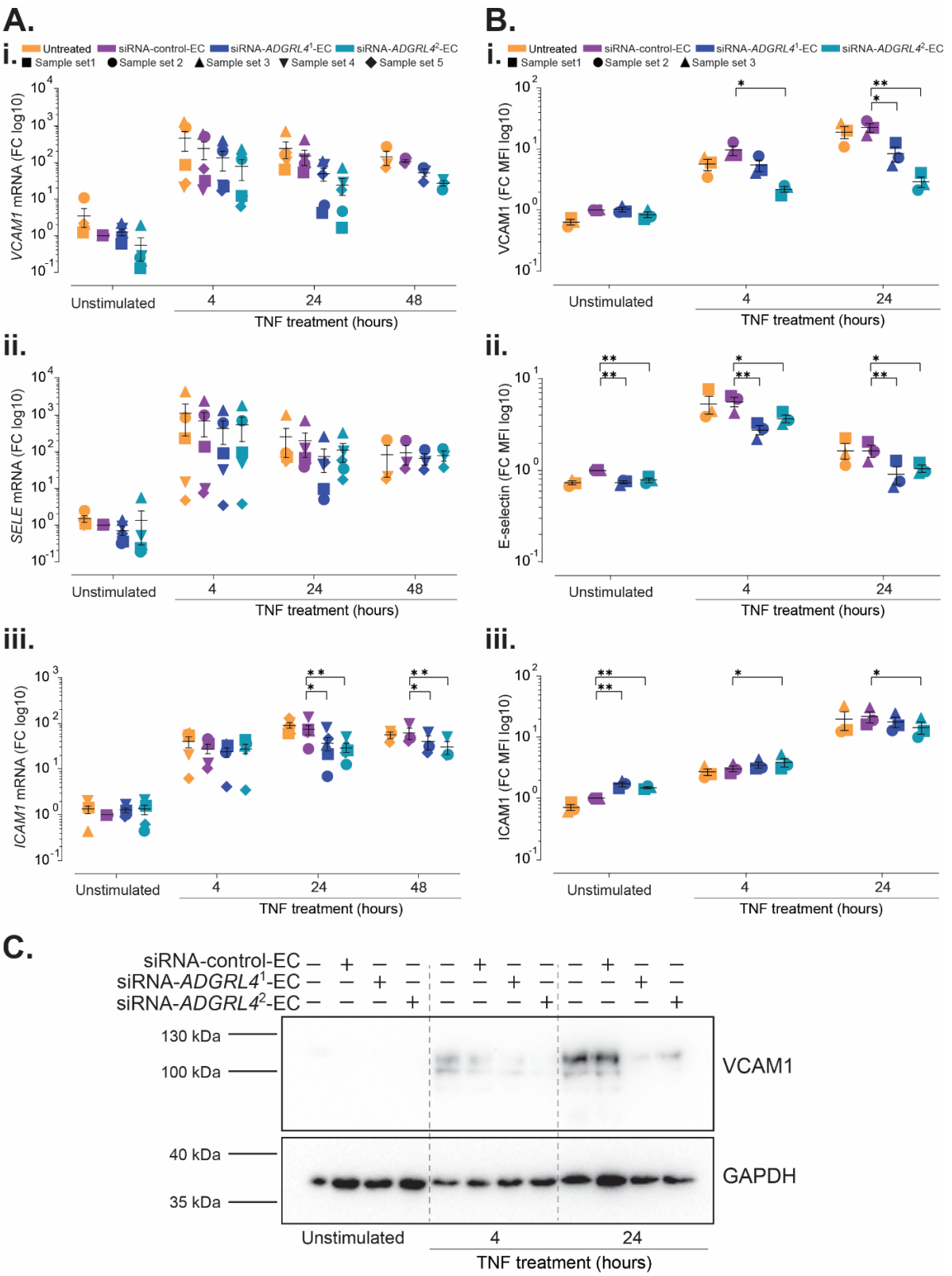


Figure 8. Endothelial *ADGRL4* depletion inhibits TNF-induced leukocyte adhesion receptor expression. Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*[1]-EC, or siRNA[2]-*ADGRL4*-EC

were treated with or without TNF. Samples were collected at 2, 4, 24 and 48 hours before measurement of **(A)** relative mRNA (fold change (FC) vs. unstimulated siRNA-control-EC) for **(i)** *VCAM1*, **(ii)** *SELE* and **(iii)** *ICAM1* or **(B)** on the cell surface by flow cytometry, displayed as mean fluorescence intensity (MFI) relative to untreated siRNA-control-EC for **(i)** VCAM1, **(ii)** E-selectin) or **(iii)** ICAM1 or **(C)** protein content of VCAM1 in entire cell lysates by western blot analysis. Mean values with SEM are indicated on graphs *p<0.05, **p<0.01 by one-way ANOVA for each time-point.

Investigation of ADGRL4 modulation of the TNF pathway

As ADGRL4 silencing seemed to cause a complex regulatory pattern with bidirectional effects, which is observed to different degrees along the TNF induction trajectory, it seemed less likely that these events solely resulted from modification in the early TNF signalling pathway. Indeed, the mass spectrometry data showed that ADGRL4 depletion did not affect the expression of tumour necrosis factor receptor superfamily member 1A and 1B (TNFRSF1A and TNFRSF1B), various receptor accessory proteins, or nuclear factor NF-kappa-B (NFKB) proteins with associated regulators (Figure 8A). 28 mitogen activated protein kinases (MAPKs), which are potential regulators of the TNF signalling pathway, were detected by mass spectrometry and of these 26 were not affected by ADGRL4 depletion (Figure S6 B). The remaining two were downregulated - dual specificity mitogen-activated protein kinase 1 (MAP2K1) (FC [siRNA-*ADGRL4*-EC *vs.* siRNA-control-EC, SS 1,2,3]: 0.49, 0.56 and 0.56) and dual specificity mitogen-activated protein kinase 6 (MAP2K6) (FC: 0.65, 0.50 and 0.51) (Figure S6B). 13 other kinases were upregulated or downregulated by ADGRL4 depletion (Figure S6C), including EIF2AK2. MAP2K1, MAP2K6 and EIF2AK2 have been previously reported as being involved in ERK-signalling (MAPK1 and MAPK3) [47] or the p38 MAPK pathway (MAPK11, MAPK12, MAPK13 and MAPK14) [48-50], but Western blot analysis showed no clear difference in phosphorylation status of p38 or ERK1/2 between siRNA^{1}-*ADGRL4*-EC and siRNA^{2}-*ADGRL4*-EC, relative to siRNA-control-EC, at 15- or 30-minutes post TNF-stimulation (Figure S6D). Although these results did not indicate changes, downregulation of the MAPKs could potentially have a broad spectre of cellular effects, which remains to be elucidated on.

DISCUSSION:

Here, we investigate the role of ADGRL4 in EC and identify a previously unappreciated potential role in the EC response to inflammatory stimulation. ADGRL4 has been reported to have a role in EC angiogenesis and proliferation [18-20], but studies have failed to show any activation or coupling of ADGRL4 to any canonical GPCR signalling pathway [18]. One of the main functions of EC is the regulation of inflammatory processes [3, 51], such as the recruitment of leukocytes to sites of inflammation [52]. If ADGRL4, as an EC enriched gene, has a key role in aspects of the EC response to cytokine stimulation, it is possible that prior failure to identify its associated signalling pathways could be linked to the absence of such conditions in previous studies.

Under unstimulated conditions, we identified a panel of proteins that were upregulated when ADGRL4 in EC was silenced, which were prominently linked to biological processes centred around cell migration, adhesion, structure organization and angiogenesis. These findings are consistent with previous studies that showed a role for ADGRL4 in angiogenesis and vascular development [18, 20, 53], processes that depend on cellular structural rearrangements. Other aGPCR family members can affect cell size, shape, polarity, adhesion and migration, through cytoskeletal modifications, such as actin reorganization [16]. Indeed, among the most upregulated proteins in our dataset were the actin-crosslinking protein TAGLN, which has been linked to angiogenesis and EC elongation [54], and the cytoskeleton protein PALLD, which is involved in actin polymerization and organization [55-57], affecting cell motility and phagocytosis. ADGRL4 silencing in unstimulated EC also downregulated a group of proteins involved in proliferation, cell growth and cell cycle progression. These included NUSAP1, which associates with microtubules during mitosis and is important for proper chromosome segregation [58-60], CDK1, which is amongst the main drivers of the cell cycle [61-63], CKS1B, which is important for CDK function [64-67] and MASTL, which is important in DNA damage response and mitotic entry [68-71]. ADGRL4 overexpression has been reported to reduce proliferation [18], assessed by imaging and cell counting - results that appear potentially contradictory to our findings. However, we did not measure the effects of ADGRL4 on cell proliferation and the changes in cell cycle associated proteins could reflect various feedback effects or these may have negative regulatory functions on cell cycle progression. Nonetheless, our results also indicate that ADGRL4 is involved in proliferation in EC.

When treated with the inflammatory cytokine TNF, a variety of EC proteins are induced [32, 33, 38, 39, 43, 72]. When ADGRL4 was silenced, TNF-induced tissue factor protein was significantly and relatively specifically elevated. EC tissue factor is upregulated by many factors, including TNF, IL1, LPS, and shear stress ([46, 73-79]), and induction is fast, with

mRNA and protein peaking within 1-6 hours post stimulation [79-81]. At earlier time points post TNF stimulation, tissue factor mRNA and protein induction were comparable between control cells and those where ADGRL4 was silenced, indicating that the absence of ADGRL4 maintained or stabilized tissue factor mRNA and/or protein, prolonging the resolution phase after TNF induction, as opposed to increasing initial transcription or translation. The half-life of tissue factor is short, reported to be between 1-2 hours [82, 83], which suggest that tissue factor levels will decrease as soon as transcription is shut down. An ADGRL4 involvement in this turnover, in which ADGRL4 depletion would increase mRNA stability, could potentially give rise to our observed results. Tissue factor regulatory mechanisms have been reported, like for Vascular *miR-181b* that regulates the tissue factor induction, but this would not be consistent with our results. Another study identified Protein mono-ADP-ribosyltransferase PARP14 (PARP14), which in concert with the mRNA-destabilizing protein tristetraprolin (TTP), caused destabilization of tissue factor mRNA, and where PARP14 deficiency stabilized mRNA for a long period of time. It is therefore possible that ADGRL4 is involved in a similar destabilizing mechanism.

ADGRL4 depletion was found to selectively downregulate 16 TNF-induced interferon associated proteins, including MX1, IFIT1 and ISG20. The regulation of such proteins by inflammatory cytokines in EC under normal conditions is generally not well understood, but our observations of their induction in control EC at 24 hour post TNF-stimulation is consistent with a recent study which reported a late stage interferon response in EC, following TNF stimulation [42]. Expression of such genes is considered primarily driven by the production and subsequent signaling of interferon [84], via canonical (JAK-STAT) or non-canonical pathways [85], with IRF1 as a central driver, but we failed to detect any *de novo* production of interferon-beta or alpha in our system (data not shown). As IRF1 transcription is induced by NFKB, and IRF1 is a known ISG transcription factor [41], the increase in interferon associated gene expression following TNF stimulation could be driven by IRF1 directly. The potential link between ADGRL4 and the regulation of TNF-induced interferon related genes is currently unclear, but it is thus possible that IRF1 could be involved. Indeed, the expression of VCAM1, the transcription of which is also induced by IRF1 ([86, 87], was downregulated by ADGRL4 silencing. No change in IRF1 expression was induced when ADGRL4 was silenced, but IRF1 activity could be affected by post-translational modifications [88-92], which we did not measure.

In our data, we observed the downregulation of MAP2K1 and MAP2K6 after ADGRL4 silencing, irrespectively of TNF-influence. MAP2K1 belongs to the prototypical RAS-RAF-MEK-ERK signalling pathway and is activated by various stimuli and associated with

proliferation and growth [47, 93, 94], suggesting a link to the observed downregulation of cell cycle proteins, but no marked changes in phosphorylation of downstream targets MAPK1 or MAPK3 were indicated. MAP2K6 is activated in response to environmental stress and pro-inflammatory cytokines and activates p38-MAPKs (MAPK11-14), which subsequentially can affect cellular responses by targeting specific transcription factors [48, 95, 96]. Activated p38-MAPKs were detected in our samples with no indication of phosphorylation changes resulting from ADGRL4 silencing. A third downregulated kinase, EIF2AK2, plays a vital role in viral infection by inhibiting general translation [97, 98]. EIF2AK2 has been found to regulate signalling responses through regulation of IRF3, MAP2K6, p38-MAPKs and NFKB activity, with subsequent impact on gene transcription [49, 50, 99-103]. Considering all the possible targets and interactions for mentioned kinases, it is plausible that at least one could be linked to ADGRL4 through some for now undefined mechanism. Regardless, it appears ADGRL4 activation and subsequent signalling is complex, and the potential links between the different modulatory effects reported here warrant further exploration.

Four our study, we have utilized human umbilical cord endothelial cells (HUVEC), which are used at low passage, to avoid significant culture-related alteration of EC phenotype that can happen over time [104-106]. Even though HUVEC are a widely used model with several advantages [107], they are of foetal origin, and some studies report they have gene expression pattern that differ from other adult human EC [108]. Also, HUVEC in our study have not been cultured under flow, or in combination with other cell types, which can affect endothelial characteristics [109-111]. The study of ADGRL4 function in EC from adult tissue and in co-culture with other cells types could provide further insights. The sample size in this study is relatively low, and so we focussed only the most clearly modified proteins; it is possible that others could be of functional importance.

Taken together, we have uncovered some interesting new aspects of ADGRL4 function in EC, and by contributing with an extended list of proteins modulated by ADGRL4 silencing, we hope to contribute to future research and possibly narrow down the many pathways and interactions that could be interesting to investigate.

METHODS

Endothelial cell culture and treatments

For the mass spectrometry experiments, human umbilical vein endothelial cells (HUVEC) were isolated from human umbilical cords, collected from Karolinska University Hospital, Stockholm, Sweden, as previously described [112]. Ethical approval was granted by *Regionala etikprövningsnämnden i Stockholm* (2015/1294-31/2). For other experiments, EndoGRO HUVEC (Millipore) were used. HUVEC were cultured Medium 199 Earle's Salts (Gibco) supplemented with 20 % Fetal Bovine Serum (Sigma), 1x Penicillin-Streptomycin (Sigma), 1 µg/ml Hydrocortisone (Sigma), 1.25 µg/ml Amphotericin B (Gibco), 1 ng/ml murine Epidermal Growth Factor (Sigma).

siRNA transfection: Medium was exchanged with Opti-MEM Reduced Serum Medium (Gibco) before Lipofectamine RNAiMAX transfection reagent (Invitrogen) was mixed with either scrambled control siRNA (Silencer™ Select Negative Control No. 1 siRNA) (Invitrogen), *ADGRL4* siRNA 1 (Silencer® Select Pre-designed siRNA, Assay ID: s34480) (Invitrogen) or *ADGRL4* siRNA 2 (Silencer® Pre-designed siRNA, Assay ID: 6325) (Invitrogen) and HUVEC transfected according to manufacturers instruction. Cells were cultured for an additional 48 hours post-transfection in standard medium., prior to further treatment/assay. **Cytokine stimulation**: In some experiments, HUVEC were stimulated with 10 ng/ml human Tumour Necrosis Factor (TNF; Sigma) for time periods as stated in the results. **Tissue factor function blocking**: In some experiments, HUVEC were pre-incubated with mouse monoclonal tissue factor antibody (final concentration = 12.5µg/ml) (BD Pharmingen, Cat# 550252, RRID: AB_393557) for 1 hour prior to assay.

qPCR

Samples for relative qPCR were prepared using TaqMan Fast Advanced Cells-to-CT Kit (Invitrogen) according to manufactures instructions. TaqMan primer/probes targeting *F3* (Hs01076029_m1), *ICAM1* (Hs00164932_m1), *VCAM1* (Hs01003372_m1), *SELE* (Hs00174057_m1), *MX1* (Hs00895608_m1), *IFIT1* (Hs03027069_s1), *ISG20* (Hs00158122_m1), *IRF1* (Hs00971965_m1) and *ADGRL4* (Hs00223377_m1) were used along with Eukaryotic 18S rRNA Endogenous Control (all Applied Biosystems). The LightCycler 96 Instrument (Roche) with software version1.1, was used and all results are given as $\Delta\Delta$Ct log10, with values are calculated relative to *18S* and normalized to the non-stimulated scrambled control to give fold change measurements.

Western blotting

HUVEC lysates were mixed with Laemmli Sample Buffer containing 10 % DTT and heat denatured before separation by SDS-PAGE, and semi-dry transfer onto PVDF membranes. The following primary antibodies were used: tissue factor (dilution 1:1000) (Cell Signaling Technology, Cat# 55147, RRID:AB_2799478), VCAM1 (dilution 1:750) (Atlas Antibodies, Cat# HPA069867, RRID:AB_2686215), ICAM1 (1:1000) (Cell Signaling Technology, Cat# 67836, RRID:AB_2799738), IRF1 (1:750) (Atlas Antibodies, Cat# HPA063131, RRID:AB_2684945), MX1 (1:1000) (Cell Signaling Technology Cat# 37849, RRID:AB_2799122), IFIT1 (1:1000) (Cell Signaling Technology Cat# 14769, RRID:AB_2783869), ADGRL4 (1:1000) (Invitrogen, Cat# MA5-24705, RRID:AB_2664834), Phospho-p38 MAPK (Thr180/Tyr182) (1:1000) (Cell Signaling Technology, Cat#4511, RRID:AB_2139682) and Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204) (1:1000) (Cell Signaling Technology, Cat#9101, RRID:AB_331646), GAPDH (dilution 3 µg/mL) (Invitrogen, Cat# AM4300, RRID: AB_2536381) followed by secondary antibody staining with horseradish peroxidase (HRP)-conjugated goat anti-mouse (diluted 1:5000) (Bio-Rad Cat# 1706516, RRID: AB_11125547) or goat anti-rabbit (diluted 1:4000) (Bio-Rad Cat# 1706515, RRID: AB_11125142), and chemiluminescent detection was achieved using Clarity Western ECL Substrate (BioRAD) and iBrightCL1000 (Thermo Scientific). Stripping of blots was carried out after detection of protein of interest, followed by re-probing and detection of GAPDH which acted as loading control. Stripping and re-probing also applied in instances where several proteins of interest were investigated using the same blot, like for MX1 (first) and IFIT1 (second), and Phospho-MAPK11-14 (first) and Phospho-MAPK1+MAPK3 (second). PageRuler Prestained Protein Ladder (10 to 180 kD) (Thermo Scientific) was used for protein weight estimation purposes.

Flow cytometry

HUVEC viability was measured using LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit for 633 or 635 nm excitation (Invitrogen), followed by incubation with fluorophore conjugated antibodies targeting tissue factor (BioLegend, Cat# 365204, RRID: AB_2564566) or VCAM1 (BioLegend, Cat# 305808, RRID:AB_2214227), ICAM1 (BioLegend, Cat# 353110, RRID: AB_10900234) and E-selectin (BioLegend, Cat# 336016, RRID:AB_2800891) combined. CytoFLEX flow cytometer (Beckman Coulter) and data was analysed using CytExpert 2.0. software. All values are based on the selection of live cells and singlets only, where mean fluorescence intensity of cell population for each sample was calculated, followed by conversion to fold change relative to non-stimulated scrambled control.

Real-time thrombin generation assay

Cells were washed in PBS and blocked for 30 minutes using PBS with 1% BSA before thrombin generation was initiated in 120 µl reaction volume containing human citrated plasma, MP-reagent [final concentration =4µM phospholipids] (Thrombinoscope BV), and FluCa-kit buffer/substrate [final concentration =16.6 mM Ca^{2+} and 2.5 mM fluorogenic substrate (Z-Gly-Gly-Arg-AMC)] (Thrombinoscope BV), in addition to corn trypsin inhibitor (Final concentration = 2 µM) (Haematologic Technologies) for purposes of blocking intrinsic pathway. Some samples received either thrombin calibrator (Thrombinoscope BV), or Dade Innovin reagent containing tissue factor (final concentration 2 pM) (Dade Innovin), the latter as a positive control for thrombin generation potential. Thrombin generation was measured by a Thrombinoscope and quantified using the Thrombinoscope software package (Version 5.0.0.742) that reported means of duplicates ± SD. Results are based on three separate experiments with individual donors. See Figure S4 for more details on experimental set-up.

Mass spectrometry

Sample preparation: Cell pellets were lysed in 4% SDS, 25 mM HEPES, 1 mM DTT and heated to 95°C for 5 minutes followed by sonication for 1 minute and centrifugation at 14000 x g for 15 minutes. The supernatant was mixed with 1 mM DTT, 8 M urea, 25 mM HEPES, pH 7.6, and transferred to a 10-kDa cut-off centrifugation filtering unit (Pall, Nanosep), and centrifuged at 14000 x g for 15 minutes. Proteins were alkylated by 50 mM iodoacetamide (IAA) in 8 M urea, 25 mM HEPES for 10 minutes. The proteins were then centrifuged at 14000 x g for 15 minutes followed by 2 more additions and centrifugations with 8 M urea, 25 mM HEPES. Trypsin (Promega) in 250 mM urea, 50 mM HEPES was added to the proteins at a ratio of 1:50 trypsin:protein and incubated overnight at 37°C with gentle shaking. The filter units were centrifuged at 14000 x g for 15 min followed by another centrifugation with ultra-pure water (Milli-Q, Millipore) and the flow-through was collected. Peptides were labelled with TMT10-plex reagent according to the manufacturer's protocol (Thermo Scientific) and cleaned by a strata-X-C-cartridge (Phenomenex).

IPG-IEF of peptides: The TMT labelled peptides, 500 µg per TMT-10plex, were separated by immobilized pH gradient - isoelectric focusing (IPG-IEF) on a 3-10 strip. Peptides were extracted from the strips into 72 fractions by a prototype liquid handling robot, supplied by GE Healthcare Bio-Sciences AB. A plastic device with 72 wells was put onto each strip and 50 µl of ultra-pure water (Milli-Q, Millipore) was added to each well. After 30 minutes incubation, the liquid was transferred to a 96 well plate and the extraction was repeated 2 more times. The

extracted peptides were dried in a SpeedVac vacuum concentrator and dissolved in 3% acetronitrile (ACN), 0.1 % formic acid.

Q Exactive analysis: Before analysis on the Q Exactive Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Scientific), peptides were separated using an Ultimate 3000 RSLCnano system. Samples were trapped on an Acclaim PepMap nanotrap column (C18, 3 µm, 100Å, 75 µm x 20 mm), and separated on an Acclaim PepMap RSLC column (C18, 2 µm, 100Å, 75 µm x 50 cm), (Thermo Scientific). Peptides were separated using a gradient of A (5% DMSO, 0.1% FA) and B (90% ACN, 5% DMSO, 0.1% FA), ranging from 6 % to 37 % B in 30-90 min (depending on IPG-IEF fraction complexity) with a flow of 0.25 µl/min. The Q Exactive was operated in a data dependent manner, selecting top 10 precursors for fragmentation by HCD. The survey scan was performed at 70000 resolutions from 400-1600 m/z, with a max injection time of 100 ms and target of 1 x 106 ions. For generation of HCD fragmentation spectra, a max ion injection time of 140 ms and AGC of 1 x 105 were used before fragmentation at 30 % normalized collision energy, 35000 resolution. Precursors were isolated with a width of 2 m/z and put on the exclusion list for 70 s. Single and unassigned charge states were rejected from precursor selection.

Peptide and protein identification: All Orbitrap data was searched by SequestHT under the software platform Proteome Discoverer 1.4 (Thermo) against the Uniprot human protein database (canonical and isoforms, downloaded 20150406) and filtered to a 1% false discovery rate (FDR). A precursor mass tolerance of 10 ppm, and product mass tolerances of 0.02 Da for HCD-FTMS were used. Further settings used were: Trypsin with 2 missed cleavage; iodoacetamide on cysteine and TMT on lysine and N-terminal as fixed modifications; and oxidation of methionine as variable modification. Quantification of TMT-10plex reporter ions was done by Proteome Discoverer on HCD-FTMS tandem mass spectra using an integration window tolerance of 10 ppm. Only peptides unique to a protein group were used for quantitation.

Identification of differently regulated proteins: To identify differently regulated proteins within the dataset, the fold change between each ADGRL4- and scrambled siRNA sample set were calculated, and criteria for upregulation and downregulation was defined as fold change ≥ 1.3, and ≤ 0.7, respectively, and applied to each sample set representing non-stimulatory or stimulatory conditions. TNF induced proteins were defined based on changes in TNF stimulated relative to non-stimulated scrambled samples only, and criteria was given as fold change ≥ 2 for both TNF stimulated samples relative to average of the three non-stimulated samples.

Go ontology enrichment analysis

Go ontology enrichment analysis was performed using the ClusterProfiler package, version 4.6.0 [113] and annotation data from org.Hs.eg.db with the human genome annotation package, version 3.16.0, with mappings based on data provided by the Gene Ontology Consortium (2022-07-01) [29-31]. Default settings were used, including p-value correction by the Benjamin Hochberg adjustment method, with maxGSSize set to 2500.

PubMed search

For PubMed counts, the search was carried out in R using the easypubmed package, with results retrieved on 2023-02-13, and search was done using gene name alone or combined with interferon.

Fibrin deposition assay

HUVEC were grown in µ-Slide VI 0.4 channels (Ibidi) and connect via silicon tubing to a syringe pump containing freshly drawn citrated human blood and a re-calcification solution containing 0.1 % glucose, 75 mM CaCl2, 37.5 mM MgCl2 in Tyrone HEPES Buffer (126 mM NaCl, 2.7 mM KCl, 0.42 mM $NaH_2PO_4H_2O$, 5 mM HEPES, pH 7.45). Solutions were infused at a 10:1 ratio respectively for 10 min at a continuous flow corresponding to a shear stress of 1 dyn/cm². The channels were washed (shear stress of 0.5 dyn/cm²) for 8 min with a solution containing 0.1 % glucose, 0.1 % BSA, 2 mM $MgCl_2$ in Tyrone HEPES Buffer, then infused with 4 % paraformaldehyde in PBS and left to fixate for 10 min. The channels were blocked with 1% BSA in PBS. Deposited fibrin was stained using an in-house produced anti-fibrin mouse monoclonal antibody at 4°C overnight in 1% BSA in PBS. A secondary Alexa Fluor 488 goat anti-mouse antibody in 1% BSA in PBS was used for detection and cells treated with Vectashield mounting medium with DAPI (Vector Laboratories) before imaging using a Leica TCS SPE confocal microscope, equipped with a 63x oil immersion objective (Leica Microsystems GmbH), and images were analyzed using the image processing package Fiji/ImageJ.

UMAP plots for ADGRL4 expression

Data from scRNAseq analysis of cell types from 24 human tissues subdivided into the-immune-, epithelial-, stromal-, and endothelial compartment was downloaded from https://figshare.com/projects/Tabula_Sapiens/100973 [28], and cell type clustering and categorization was performed as originally described. The R Seurat package [114] and the FindAllMarkers function was used to generate illustrative UMAP plots to show ADGRL4

expression pattern. The Tabula Sapiens cell atlas website https://tabula-sapiens-portal.ds.czbiohub.org/ can be accessed for more detailed information and visualization of cell type expression patterns.

Use of protein names and abbreviation in text and figures

For simplicity and to avoid confusion, protein abbreviations are given as non-italic HGNC-approved gene symbols [115] in most text-situations and also used in figures and tables, but for contextualization, the full protein name recommended by UniProt [116] is also given the first time a protein is mentioned in the main text.

SUPPLEMENTAL TABLE LEGENDS

Table S1. Mass spectrometry protein profiling data. Related to Figure 1, 2, 6, S2. Human umbilical vein endothelial cells (HUVEC, n=3 independent biological replicates; divided into sample sets [SS 1,2,3,4,5]) were transfected with siRNA targeting *ADGRL4* (siRNA-*ADGRL4*-EC) or a scrambled-control sequence (siRNA-control-EC) and 48 hours post transfection were either non-stimulated or stimulated with TNF for 24 hours (SS1-3 no TNF, SS4-5 with TNF) before collection and analysis by mass spectrometry. **[Tab 1]** Key notes for mass spectrometry data. **[Tab 2]** Final list of all proteins identified in the analysis. **[Tab 3]** List of proteins identified as upregulated after ADGRL4 silenced, with no TNF treatment (fold change siRNA-*ADGRL4*-EC vs. siRNA-control-EC = >1.3 fold). **[Tab 4]** List with gene ontology overrepresented terms (biological processes) for proteins in Tab 3. **[Tab 5]** List of proteins identified as downregulated after ADGRL4 silencing, with no TNF treatment (fold change siRNA-*ADGRL4*-EC vs. siRNA-control-EC = <0.7 fold). **[Tab 6]** List with gene ontology overrepresented terms (biological processes) for proteins in Tab 5. **[Tab 7]** List of proteins identified as TNF-induced in control samples (siRNA-control-EC + TNF vs. siRNA-control-EC unstimulated = >2 fold change). **[Tab 8]** List with gene ontology overrepresented terms (biological processes) for proteins in Tab 7. **[Tab 9]** List of proteins identified as upregulated after ADGRL4 silencing, with TNF treatment (fold change siRNA-*ADGRL4*-EC + TNF vs. siRNA-control-EC + TNF = >1.3 fold). **[Tab 10]** List with gene ontology overrepresented terms (biological processes) for proteins in Tab 9. **[Tab 11]** List of proteins identified as downregulated after ADGRL4 silencing, with TNF treatment (fold change siRNA-*ADGRL4*-EC + TNF vs. siRNA-control-EC + TNF = <0.7 fold). **[Tab 12]** List with gene ontology overrepresented terms (biological processes) for proteins in Tab 11. **[Tab 13]** List with gene ontology overrepresented terms (biological processes) for 16 proteins in Tab 11 identified as TNF-induced but downregulated by ADGRL4 silencing.

REFERENCES

1. Krüger-Genge, A., et al., *Vascular Endothelial Cell Biology: An Update.* Int J Mol Sci, 2019. **20**(18).
2. Wang, M., et al., *Thrombotic Regulation From the Endothelial Cell Perspectives.* Arterioscler Thromb Vasc Biol, 2018. **38**(6): p. e90-e95.
3. Pober, J.S. and W.C. Sessa, *Evolving functions of endothelial cells in inflammation.* Nature Reviews Immunology, 2007. **7**(10): p. 803-815.
4. Norreen-Thorsen, M., et al., *A human adipose tissue cell-type transcriptome atlas.* Cell Rep, 2022. **40**(2): p. 111046.
5. Butler, Lynn M., et al., *Analysis of Body-wide Unfractionated Tissue Data to Identify a Core Human Endothelial Transcriptome.* Cell Systems, 2016. **3**(3): p. 287-301.e3.
6. 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.
7. Dusart, P., et al., *A tissue centric atlas of cell type transcriptome enrichment signatures.* bioRxiv, 2023: p. 2023.01.10.520698.
8. Karlsson, M., et al., *A single–cell type transcriptomics map of human tissues.* Science Advances. **7**(31): p. eabh2169.
9. Öling, S., et al., *A human stomach cell type transcriptome atlas.* bioRxiv, 2023: p. 2023.01.10.520700.
10. Dejana, E., *Endothelial cell–cell junctions: happy together.* Nature Reviews Molecular Cell Biology, 2004. **5**(4): p. 261-270.
11. Gavard, J., *Endothelial permeability and VE-cadherin: a wacky comradeship.* Cell Adh Migr, 2014. **8**(2): p. 158-64.
12. Scalise, A.A., et al., *The blood-brain and gut-vascular barriers: from the perspective of claudins.* Tissue Barriers, 2021. **9**(3): p. 1926190.
13. 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.
14. Hassan, M.I., A. Saxena, and F. Ahmad, *Structure and function of von Willebrand factor.* Blood Coagul Fibrinolysis, 2012. **23**(1): p. 11-22.
15. 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.
16. Hamann, J., et al., *International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G protein-coupled receptors.* Pharmacol Rev, 2015. **67**(2): p. 338-67.
17. Favara, D.M., A.H. Banham, and A.L. Harris, *ADGRL4/ELTD1 is a highly conserved angiogenesis-associated orphan adhesion GPCR that emerged with the first vertebrates and comprises 3 evolutionary variants.* BMC Evol Biol, 2019. **19**(1): p. 143.
18. Favara, D.M., et al., *Elevated expression of the adhesion GPCR ADGRL4/ELTD1 promotes endothelial sprouting angiogenesis without activating canonical GPCR signalling.* Scientific Reports, 2021. **11**(1): p. 8870.
19. Favara, D.M., et al., *ADGRL4/ELTD1 Silencing in Endothelial Cells Induces ACLY and SLC25A1 and Alters the Cellular Metabolic Profile.* Metabolites, 2019. **9**(12).
20. Masiero, M., et al., *A core human primary tumor angiogenesis signature identifies the endothelial orphan receptor ELTD1 as a key regulator of angiogenesis.* Cancer Cell, 2013. **24**(2): p. 229-41.

21. Sheldon, H., et al., *ELTD1 Activation Induces an Endothelial-EMT Transition to a Myofibroblast Phenotype.* Int J Mol Sci, 2021. **22**(20).
22. Sheldon, H., et al., *ADGRL4/ELTD1 Expression in Breast Cancer Cells Induces Vascular Normalization and Immune Suppression.* Mol Cancer Res, 2021. **19**(11): p. 1957-1969.
23. Huang, H., et al., *ELTD1 deletion reduces vascular abnormality and improves T-cell recruitment after PD-1 blockade in glioma.* Neuro Oncol, 2022. **24**(3): p. 398-411.
24. Li, J., et al., *ELTD1 facilitates glioma proliferation, migration and invasion by activating JAK/STAT3/HIF-1α signaling axis.* Sci Rep, 2019. **9**(1): p. 13904.
25. Towner, R.A., et al., *ELTD1, a potential new biomarker for gliomas.* Neurosurgery, 2013. **72**(1): p. 77-90; discussion 91.
26. Guihurt Santiago, J., et al., *Adhesion G protein-coupled receptor, ELTD1, is a potential therapeutic target for retinoblastoma migration and invasion.* BMC Cancer, 2021. **21**(1): p. 53.
27. Sun, J., et al., *ELTD1 promotes invasion and metastasis by activating MMP2 in colorectal cancer.* Int J Biol Sci, 2021. **17**(12): p. 3048-3058.
28. Jones, R.C., et al., *The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans.* Science, 2022. **376**(6594): p. eabl4896.
29. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
30. Mi, H., et al., *PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools.* Nucleic Acids Res, 2019. **47**(D1): p. D419-d426.
31. *The Gene Ontology resource: enriching a GOld mine.* Nucleic Acids Res, 2021. **49**(D1): p. D325-d334.
32. Vestweber, D., *How leukocytes cross the vascular endothelium.* Nature Reviews Immunology, 2015. **15**(11): p. 692-704.
33. Pober, J.S., *Endothelial activation: intracellular signaling pathways.* Arthritis Res, 2002. **4 Suppl 3**(Suppl 3): p. S109-16.
34. Cotran, R.S. and J.S. Pober, *Cytokine-endothelial interactions in inflammation, immunity, and vascular injury.* J Am Soc Nephrol, 1990. **1**(3): p. 225-35.
35. Redl, H., et al., *Response of the endothelium to trauma and sepsis. Adherence, cytokine effects and procoagulatory response.* Arzneimittelforschung, 1994. **44**(3a): p. 443-6.
36. Theofilis, P., et al., *Inflammatory Mechanisms Contributing to Endothelial Dysfunction.* Biomedicines, 2021. **9**(7).
37. Hot, A., V. Lenief, and P. Miossec, *Combination of IL-17 and TNFα induces a pro-inflammatory, pro-coagulant and pro-thrombotic phenotype in human endothelial cells.* Ann Rheum Dis, 2012. **71**(5): p. 768-76.
38. Haraldsen, G., et al., *Cytokine-regulated expression of E-selectin, intercellular adhesion molecule-1 (ICAM-1), and vascular cell adhesion molecule-1 (VCAM-1) in human microvascular endothelial cells.* J Immunol, 1996. **156**(7): p. 2558-65.
39. Fries, J.W., et al., *Expression of VCAM-1 and E-selectin in an in vivo model of endothelial activation.* Am J Pathol, 1993. **143**(3): p. 725-37.
40. Crawley, J.T., et al., *The central role of thrombin in hemostasis.* J Thromb Haemost, 2007. **5 Suppl 1**: p. 95-101.
41. Feng, H., et al., *Interferon regulatory factor 1 (IRF1) and anti-pathogen innate immune responses.* PLoS Pathog, 2021. **17**(1): p. e1009220.

42. Valenzuela, N.M., *Late phase endothelial cell inflammation is characterized by interferon response genes and driven by JAK/STAT, not NFkappaB.* Vascul Pharmacol, 2022. **146**: p. 107090.
43. Petzelbauer, P., et al., *Heterogeneity of dermal microvascular endothelial cell antigen expression and cytokine responsiveness in situ and in cell culture.* J Immunol, 1993. **151**(9): p. 5062-72.
44. Denk, A., et al., *Activation of NF-κB via the IκB Kinase Complex Is Both Essential and Sufficient for Proinflammatory Gene Expression in Primary Endothelial Cells *.* Journal of Biological Chemistry, 2001. **276**(30): p. 28451-28458.
45. Parry, G.C. and N. Mackman, *Transcriptional regulation of tissue factor expression in human endothelial cells.* Arterioscler Thromb Vasc Biol, 1995. **15**(5): p. 612-21.
46. Bode, M. and N. Mackman, *Regulation of tissue factor gene expression in monocytes and endothelial cells: Thromboxane A2 as a new player.* Vascul Pharmacol, 2014. **62**(2): p. 57-62.
47. Smits, P.J., et al., *Endothelial MAP2K1 mutations in arteriovenous malformation activate the RAS/MAPK pathway.* Biochem Biophys Res Commun, 2020. **529**(2): p. 450-454.
48. Goebeler, M., et al., *The MKK6/p38 stress kinase cascade is critical for tumor necrosis factor-alpha-induced expression of monocyte-chemoattractant protein-1 in endothelial cells.* Blood, 1999. **93**(3): p. 857-65.
49. Goh, K.C., M.J. deVeer, and B.R. Williams, *The protein kinase PKR is required for p38 MAPK activation and the innate immune response to bacterial endotoxin.* Embo j, 2000. **19**(16): p. 4292-7.
50. Silva, A.M., et al., *Protein kinase R (PKR) interacts with and activates mitogen-activated protein kinase kinase 6 (MKK6) in response to double-stranded RNA stimulation.* J Biol Chem, 2004. **279**(36): p. 37670-6.
51. Pate, M., et al., *Endothelial cell biology: role in the inflammatory response.* Adv Clin Chem, 2010. **52**: p. 109-30.
52. Filippi, M.D., *Mechanism of Diapedesis: Importance of the Transcellular Route.* Adv Immunol, 2016. **129**: p. 25-53.
53. Lu, S., et al., *Developmental vascular remodeling defects and postnatal kidney failure in mice lacking Gpr116 (Adgrf5) and Eltd1 (Adgrl4).* PLoS One, 2017. **12**(8): p. e0183166.
54. Tsuji-Tamura, K., et al., *The canonical smooth muscle cell marker TAGLN is present in endothelial cells and is involved in angiogenesis.* J Cell Sci, 2021. **134**(15).
55. Artelt, N., et al., *The Role of Palladin in Podocytes.* J Am Soc Nephrol, 2018. **29**(6): p. 1662-1678.
56. Sun, H.M., et al., *PALLD Regulates Phagocytosis by Enabling Timely Actin Polymerization and Depolymerization.* J Immunol, 2017. **199**(5): p. 1817-1826.
57. Goicoechea, S.M., D. Arneman, and C.A. Otey, *The role of palladin in actin organization and cell motility.* Eur J Cell Biol, 2008. **87**(8-9): p. 517-25.
58. Chou, H.Y., et al., *Phosphorylation of NuSAP by Cdk1 regulates its interaction with microtubules in mitosis.* Cell Cycle, 2011. **10**(23): p. 4083-9.
59. Li, C., et al., *NuSAP modulates the dynamics of kinetochore microtubules by attenuating MCAK depolymerisation activity.* Sci Rep, 2016. **6**: p. 18773.
60. Mills, C.A., et al., *Nucleolar and spindle-associated protein 1 (NUSAP1) interacts with a SUMO E3 ligase complex during chromosome segregation.* J Biol Chem, 2017. **292**(42): p. 17178-17189.

61. Malumbres, M., *Cyclin-dependent kinases.* Genome Biol, 2014. **15**(6): p. 122.
62. Santamaría, D., et al., *Cdk1 is sufficient to drive the mammalian cell cycle.* Nature, 2007. **448**(7155): p. 811-5.
63. Suski, J.M., et al., *CDC7-independent G1/S transition revealed by targeted protein degradation.* Nature, 2022. **605**(7909): p. 357-365.
64. van Zon, W., et al., *The APC/C recruits cyclin B1-Cdk1-Cks in prometaphase before D box recognition to control mitotic exit.* J Cell Biol, 2010. **190**(4): p. 587-602.
65. Salamina, M., et al., *Discriminative SKP2 Interactions with CDK-Cyclin Complexes Support a Cyclin A-Specific Role in p27KIP1 Degradation.* J Mol Biol, 2021. **433**(5): p. 166795.
66. Liberal, V., et al., *Cyclin-dependent kinase subunit (Cks) 1 or Cks2 overexpression overrides the DNA damage response barrier triggered by activated oncoproteins.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2754-9.
67. McGrath, D.A., et al., *Cks confers specificity to phosphorylation-dependent CDK signaling pathways.* Nat Struct Mol Biol, 2013. **20**(12): p. 1407-14.
68. Bisteau, X., et al., *The Greatwall kinase safeguards the genome integrity by affecting the kinome activity in mitosis.* Oncogene, 2020. **39**(44): p. 6816-6840.
69. Wong, P.Y., et al., *MASTL(Greatwall) regulates DNA damage responses by coordinating mitotic entry after checkpoint recovery and APC/C activation.* Sci Rep, 2016. **6**: p. 22230.
70. Adhikari, D., et al., *Mastl is required for timely activation of APC/C in meiosis I and Cdk1 reactivation in meiosis II.* J Cell Biol, 2014. **206**(7): p. 843-53.
71. Burgess, A., et al., *Loss of human Greatwall results in G2 arrest and multiple mitotic defects due to deregulation of the cyclin B-Cdc2/PP2A balance.* Proc Natl Acad Sci U S A, 2010. **107**(28): p. 12564-9.
72. Eisenreich, A., et al., *Cdc2-like kinases and DNA topoisomerase I regulate alternative splicing of tissue factor in human endothelial cells.* Circ Res, 2009. **104**(5): p. 589-99.
73. Folco, E.J., et al., *Neutrophil Extracellular Traps Induce Endothelial Cell Activation and Tissue Factor Production Through Interleukin-1α and Cathepsin G.* Arterioscler Thromb Vasc Biol, 2018. **38**(8): p. 1901-1912.
74. Okamoto, T., et al., *Endothelial connexin 32 regulates tissue factor expression induced by inflammatory stimulation and direct cell-cell interaction with activated cells.* Atherosclerosis, 2014. **236**(2): p. 430-7.
75. Kothari, H., U.R. Pendurthi, and L.V. Rao, *Analysis of tissue factor expression in various cell model systems: cryptic vs. active.* J Thromb Haemost, 2013. **11**(7): p. 1353-63.
76. Moriguchi, T. and B.E. Sumpio, *PECAM-1 phosphorylation and tissue factor expression in HUVECs exposed to uniform and disturbed pulsatile flow and chemical stimuli.* J Vasc Surg, 2015. **61**(2): p. 481-8.
77. Abe, R., et al., *Varying effects of hemodynamic forces on tissue factor RNA expression in human endothelial cells.* J Surg Res, 2011. **170**(1): p. 150-6.
78. Krikun, G., et al., *Lipopolysaccharide appears to activate human endometrial endothelial cells through TLR-4-dependent and TLR-4-independent mechanisms.* Am J Reprod Immunol, 2012. **68**(3): p. 233-7.
79. Szotowski, B., et al., *Procoagulant soluble tissue factor is released from endothelial cells in response to inflammatory cytokines.* Circ Res, 2005. **96**(12): p. 1233-9.

80. Friedl, J., et al., *Induction of permeability across endothelial cell monolayers by tumor necrosis factor (TNF) occurs via a tissue factor-dependent mechanism: relationship between the procoagulant and permeability effects of TNF.* Blood, 2002. **100**(4): p. 1334-9.
81. Ishii, H., et al., *Retinoic acid counteracts both the downregulation of thrombomodulin and the induction of tissue factor in cultured human endothelial cells exposed to tumor necrosis factor.* Blood, 1992. **80**(10): p. 2556-62.
82. Ahern, S.M., T. Miyata, and J.E. Sadler, *Regulation of human tissue factor expression by mRNA turnover.* J Biol Chem, 1993. **268**(3): p. 2154-9.
83. Brand, K., et al., *Tissue factor mRNA in THP-1 monocytic cells is regulated at both transcriptional and posttranscriptional levels in response to lipopolysaccharide.* Mol Cell Biol, 1991. **11**(9): p. 4732-8.
84. Venkatesh, D., et al., *Endothelial TNF receptor 2 induces IRF1 transcription factor-dependent interferon-β autocrine signaling to promote monocyte recruitment.* Immunity, 2013. **38**(5): p. 1025-37.
85. Mazewski, C., et al., *Type I Interferon (IFN)-Regulated Activation of Canonical and Non-Canonical Signaling Pathways.* Front Immunol, 2020. **11**: p. 606456.
86. Lechleitner, S., et al., *Interferon enhances tumor necrosis factor-induced vascular cell adhesion molecule 1 (CD106) expression in human endothelial cells by an interferon-related factor 1-dependent pathway.* J Exp Med, 1998. **187**(12): p. 2023-30.
87. Li, Y., et al., *Ataxin-10 Inhibits TNF-α-Induced Endothelial Inflammation via Suppressing Interferon Regulatory Factor-1.* Mediators Inflamm, 2021. **2021**: p. 7042148.
88. Kautz, B., et al., *SHP1 protein-tyrosine phosphatase inhibits gp91PHOX and p67PHOX expression by inhibiting interaction of PU.1, IRF1, interferon consensus sequence-binding protein, and CREB-binding protein with homologous Cis elements in the CYBB and NCF2 genes.* J Biol Chem, 2001. **276**(41): p. 37868-78.
89. Park, J., et al., *Elevated level of SUMOylated IRF-1 in tumor cells interferes with IRF-1-mediated apoptosis.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 17028-33.
90. Garvin, A.J., et al., *GSK3β-SCFFBXW7α mediated phosphorylation and ubiquitination of IRF1 are required for its transcription-dependent turnover.* Nucleic Acids Res, 2019. **47**(9): p. 4476-4494.
91. Sgarbanti, M., et al., *IκB kinase ε targets interferon regulatory factor 1 in activated T lymphocytes.* Mol Cell Biol, 2014. **34**(6): p. 1054-65.
92. Lin, R. and J. Hiscott, *A role for casein kinase II phosphorylation in the regulation of IRF-1 transcriptional activity.* Mol Cell Biochem, 1999. **191**(1-2): p. 169-80.
93. Guo, Y.J., et al., *ERK/MAPK signalling pathway and tumorigenesis.* Exp Ther Med, 2020. **19**(3): p. 1997-2007.
94. Lu, N. and C.J. Malemud, *Extracellular Signal-Regulated Kinase: A Regulator of Cell Growth, Inflammation, Chondrocyte and Bone Cell Receptor-Mediated Gene Expression.* Int J Mol Sci, 2019. **20**(15).
95. Goedert, M., et al., *Activation of the novel stress-activated protein kinase SAPK4 by cytokines and cellular stresses is mediated by SKK3 (MKK6); comparison of its substrate specificity with that of other SAP kinases.* Embo j, 1997. **16**(12): p. 3563-71.
96. Raingeaud, J., et al., *MKK3- and MKK6-regulated gene expression is mediated by the p38 mitogen-activated protein kinase signal transduction pathway.* Mol Cell Biol, 1996. **16**(3): p. 1247-55.

97. Kang, J.I., et al., *PKR protein kinase is activated by hepatitis C virus and inhibits viral replication through translational control.* Virus Res, 2009. **142**(1-2): p. 51-6.
98. Samuel, C.E., *The eIF-2 alpha protein kinases, regulators of translation in eukaryotes from yeasts to humans.* J Biol Chem, 1993. **268**(11): p. 7603-6.
99. Taghavi, N. and C.E. Samuel, *Protein kinase PKR catalytic activity is required for the PKR-dependent activation of mitogen-activated protein kinases and amplification of interferon beta induction following virus infection.* Virology, 2012. **427**(2): p. 208-16.
100. Pfaller, C.K., et al., *Protein kinase PKR and RNA adenosine deaminase ADAR1: new roles for old players as modulators of the interferon response.* Curr Opin Immunol, 2011. **23**(5): p. 573-82.
101. McAllister, C.S., N. Taghavi, and C.E. Samuel, *Protein kinase PKR amplification of interferon β induction occurs through initiation factor eIF-2α-mediated translational control.* J Biol Chem, 2012. **287**(43): p. 36384-92.
102. Takada, Y., et al., *Genetic deletion of PKR abrogates TNF-induced activation of IkappaBalpha kinase, JNK, Akt and cell proliferation but potentiates p44/p42 MAPK and p38 MAPK activation.* Oncogene, 2007. **26**(8): p. 1201-12.
103. Bonnet, M.C., et al., *PKR stimulates NF-kappaB irrespective of its kinase function by interacting with the IkappaB kinase complex.* Mol Cell Biol, 2000. **20**(13): p. 4532-42.
104. Chang, M.W., et al., *Comparison of early passage, senescent and hTERT immortalized endothelial cells.* Exp Cell Res, 2005. **309**(1): p. 121-36.
105. Ohori, M., et al., *Gene regulatory network analysis defines transcriptome landscape with alternative splicing of human umbilical vein endothelial cells during replicative senescence.* BMC Genomics, 2021. **22**(1): p. 869.
106. Liao, H., et al., *Effects of long-term serial cell passaging on cell spreading, migration, and cell-surface ultrastructures of cultured vascular endothelial cells.* Cytotechnology, 2014. **66**(2): p. 229-38.
107. Medina-Leyte, D.J., et al., *Use of Human Umbilical Vein Endothelial Cells (HUVEC) as a Model to Study Cardiovascular Disease: A Review.* Applied Sciences, 2020. **10**(3): p. 938.
108. Viemann, D., et al., *TNF induces distinct gene expression programs in microvascular and macrovascular human endothelial cells.* J Leukoc Biol, 2006. **80**(1): p. 174-85.
109. Heydarkhan-Hagvall, S., et al., *Co-culture of endothelial cells and smooth muscle cells affects gene expression of angiogenic factors.* J Cell Biochem, 2003. **89**(6): p. 1250-9.
110. Nakajima, H. and N. Mochizuki, *Flow pattern-dependent endothelial cell responses through transcriptional regulation.* Cell Cycle, 2017. **16**(20): p. 1893-1901.
111. Helle, E., et al., *HiPS-Endothelial Cells Acquire Cardiac Endothelial Phenotype in Co-culture With hiPS-Cardiomyocytes.* Front Cell Dev Biol, 2021. **9**: p. 715093.
112. Wu, T., et al., *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* Innovation (Camb), 2021. **2**(3): p. 100141.
113. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587 e29.
114. Seal, R.L., et al., *Genenames.org: the HGNC resources in 2023.* Nucleic Acids Res, 2023. **51**(D1): p. D1003-d1009.
115. *UniProt: the Universal Protein Knowledgebase in 2023.* Nucleic Acids Res, 2023. **51**(D1): p. D523-d531.

Supplementary data: Paper III

Figure S1

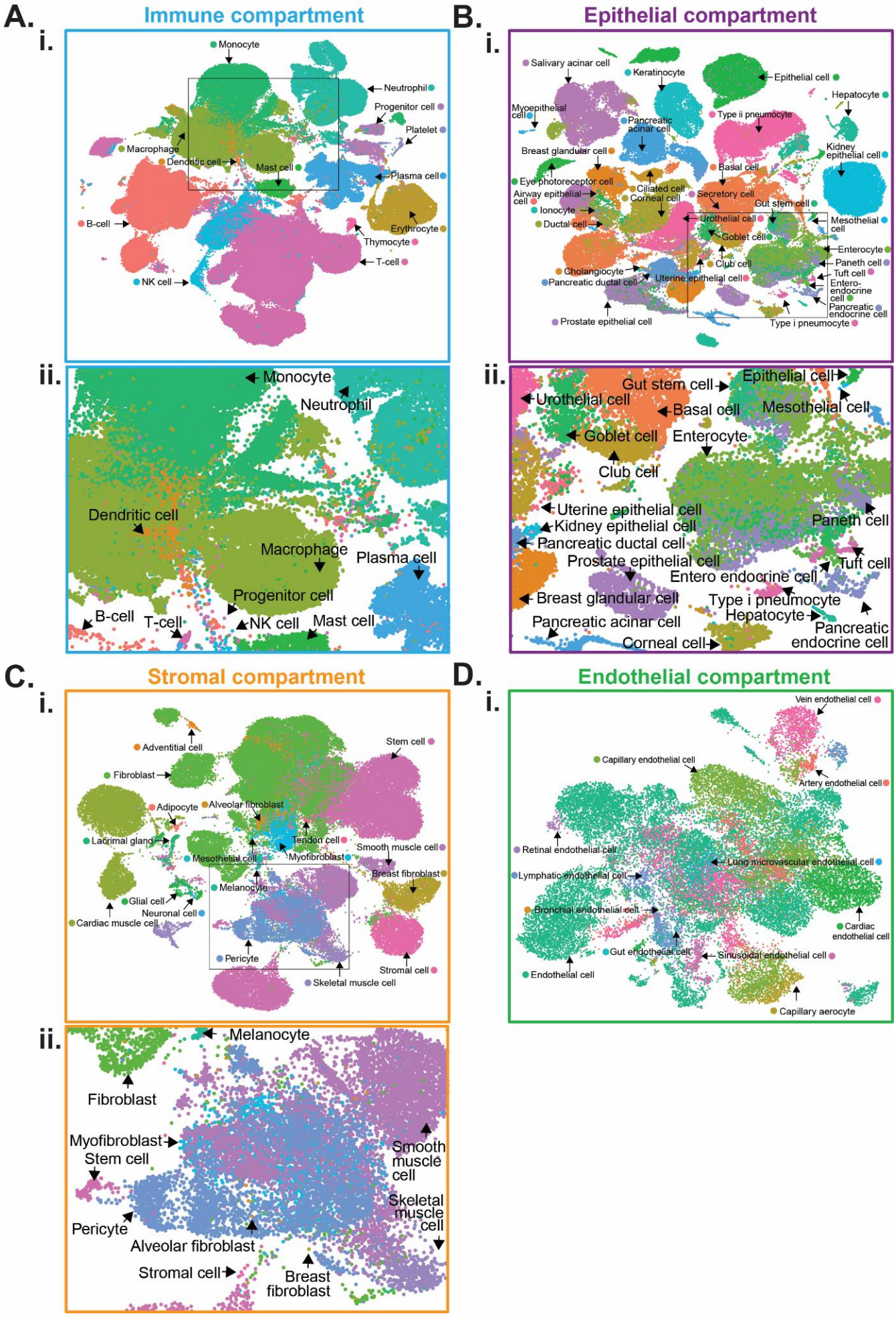

A.
Immune compartment
i.
Monocyte
Neutrophil
Progenitor cell
Platelet
Macrophage
Dendritic cell
Mast cell
Plasma cell
B-cell
Erythrocyte
Thymocyte
T-cell
NK cell
ii.
Monocyte
Neutrophil
Dendritic cell
Macrophage
Plasma cell
Progenitor cell
B-cell
T-cell
NK cell
Mast cell
B.
Epithelial compartment
i.
Salivary acinar cell
Keratinocyte
Epithelial cell
Hepatocyte
Myoepithelial cell
Pancreatic acinar cell
Type ii pneumocyte
Breast glandular cell
Kidney epithelial cell
Basal cell
Eye photoreceptor cell
Ciliated cell
Secretory cell
Airway epithelial cell
Corneal cell
Gut stem cell
Ionocyte
Urothelial cell
Mesothelial cell
Ductal cell
Goblet cell
Enterocyte
Club cell
Paneth cell
Cholangiocyte
Tuft cell
Pancreatic ductal cell
Uterine epithelial cell
Entero-endocrine cell
Pancreatic endocrine cell
Prostate epithelial cell
Type i pneumocyte
ii.
Gut stem cell
Epithelial cell
Urothelial cell
Basal cell
Mesothelial cell
Goblet cell
Enterocyte
Club cell
Uterine epithelial cell
Kidney epithelial cell
Paneth cell
Pancreatic ductal cell
Prostate epithelial cell
Tuft cell
Entero endocrine cell
Breast glandular cell
Type i pneumocyte
Hepatocyte
Pancreatic acinar cell
Pancreatic endocrine cell
Corneal cell
C.
Stromal compartment
i.
Adventitial cell
Stem cell
Fibroblast
Adipocyte
Alveolar fibroblast
Lacrimal gland
Tendon cell
Smooth muscle cell
Mesothelial cell
Myofibroblast
Breast fibroblast
Melanocyte
Glial cell
Neuronal cell
Cardiac muscle cell
Pericyte
Stromal cell
Skeletal muscle cell
ii.
Melanocyte
Fibroblast
Myofibroblast
Stem cell
Smooth muscle cell
Skeletal muscle cell
Pericyte
Alveolar fibroblast
Stromal cell
Breast fibroblast
D.
Endothelial compartment
i.
Vein endothelial cell
Capillary endothelial cell
Artery endothelial cell
Retinal endothelial cell
Lung microvascular endothelial cell
Lymphatic endothelial cell
Bronchial endothelial cell
Cardiac endothelial cell
Gut endothelial cell
Sinusoidal endothelial cell
Endothelial cell
Capillary aerocyte

Figure S1. UMAP plot overview of cell types. Related to Figure 1. **(A)** scRNAseq data from analysis of cell types from 24 human tissues subdivided into **(A)** immune-, **(B)** epithelial-, **(C)** stromal-, and **(D)** endothelial compartments was sourced from Tabula Sapiens (Tabula Sapiens., 2021), and used to generate UMAP plots. Plots display the cell type cluster annotations in each compartment: **(i)** at the same scale as in Figure 1A or **(ii)** with specific focus on of cell clusters that showed some ADGRL4 expression, outside the endothelial compartment.

.

Figure S2

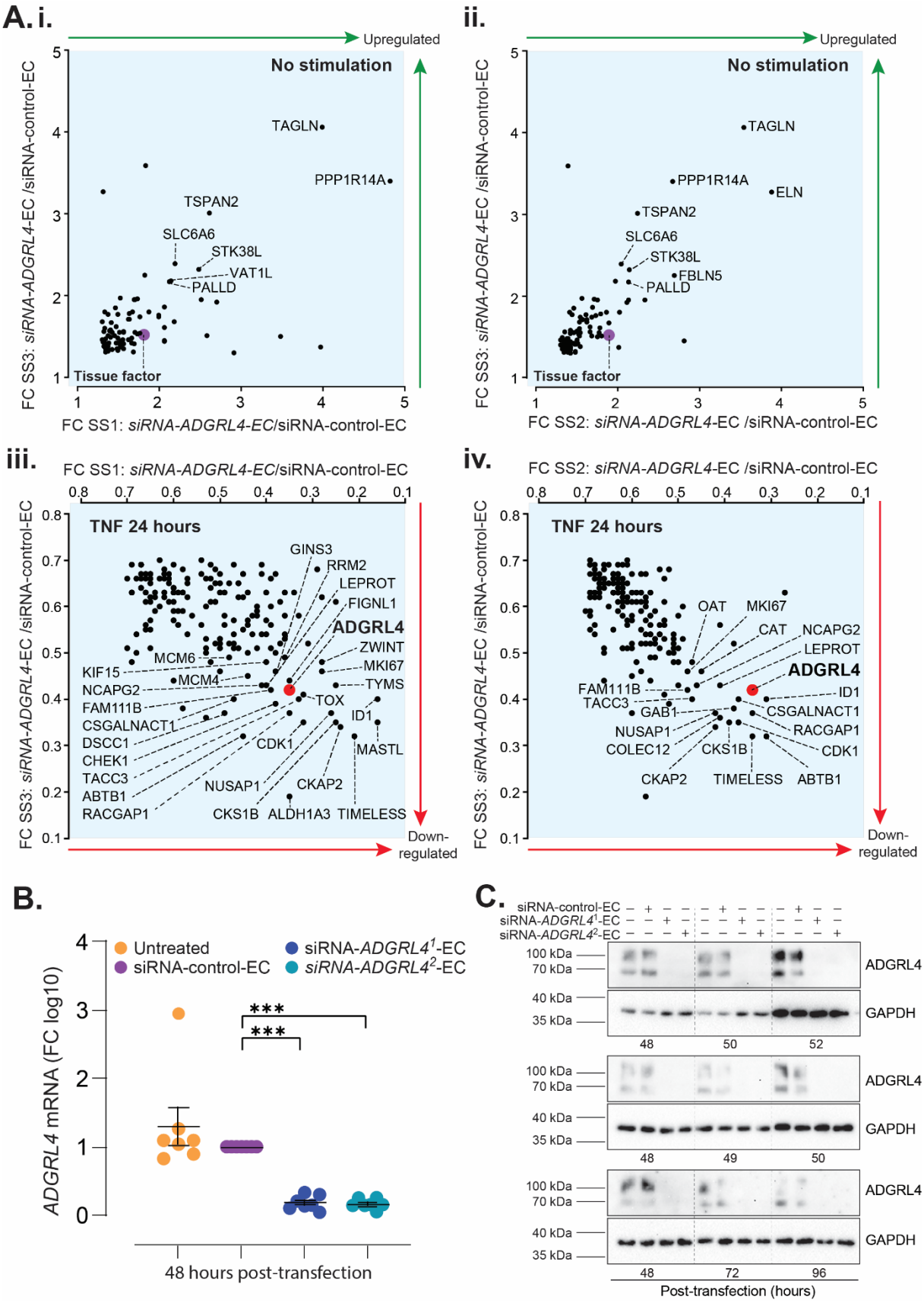


Figure S2. Protein profiling by mass spectometry: other biological replicates and evaluation of ADGRL4 silencing. Related to Figure 1 and 3 and Table S1. Human umbilical vein endothelial cells (HUVEC, n=3 independent sample sets [SS 1,2,3]) were transfected with siRNA targeting *ADGRL4*

Figure S2

(siRNA-*ADGRL4*-EC) or a scrambled-control sequence (siRNA-control-EC) and analysed after 72 hours, using mass spectrometry. **(A)** Plots showing **(i)** upregulated proteins SS3 vs SS1 **(ii)** upregulated proteins SS2 vs SS3 or **(iii)** downregulated proteins SS3 vs SS1 **(iv)** downregulated proteins SS2 vs SS3 in siRNA-*ADGRL4*-EC (plots display fold change (FC) vs. siRNA-control-EC). Only proteins with ratio >1.3 (upregulated) or <0.7 (downregulated) across all sample sets are displayed (Related to Figure 2B). **(B-C)** Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*[1]-EC, or siRNA[2]-*ADGRL4*-EC (n=3 biological replicates) were collected post-transfection at indicated time points before measurement of **(B)** relative *ADGRL4* mRNA (fold change (FC) vs. siRNA-control-EC) or **(C)** protein content of ADGRL4 in entire cell lysates by western blot analysis (n=3 biological replicates). Mean values with SEM are indicated on graph ***p<0.001 by one-way ANOVA.

Figure S3

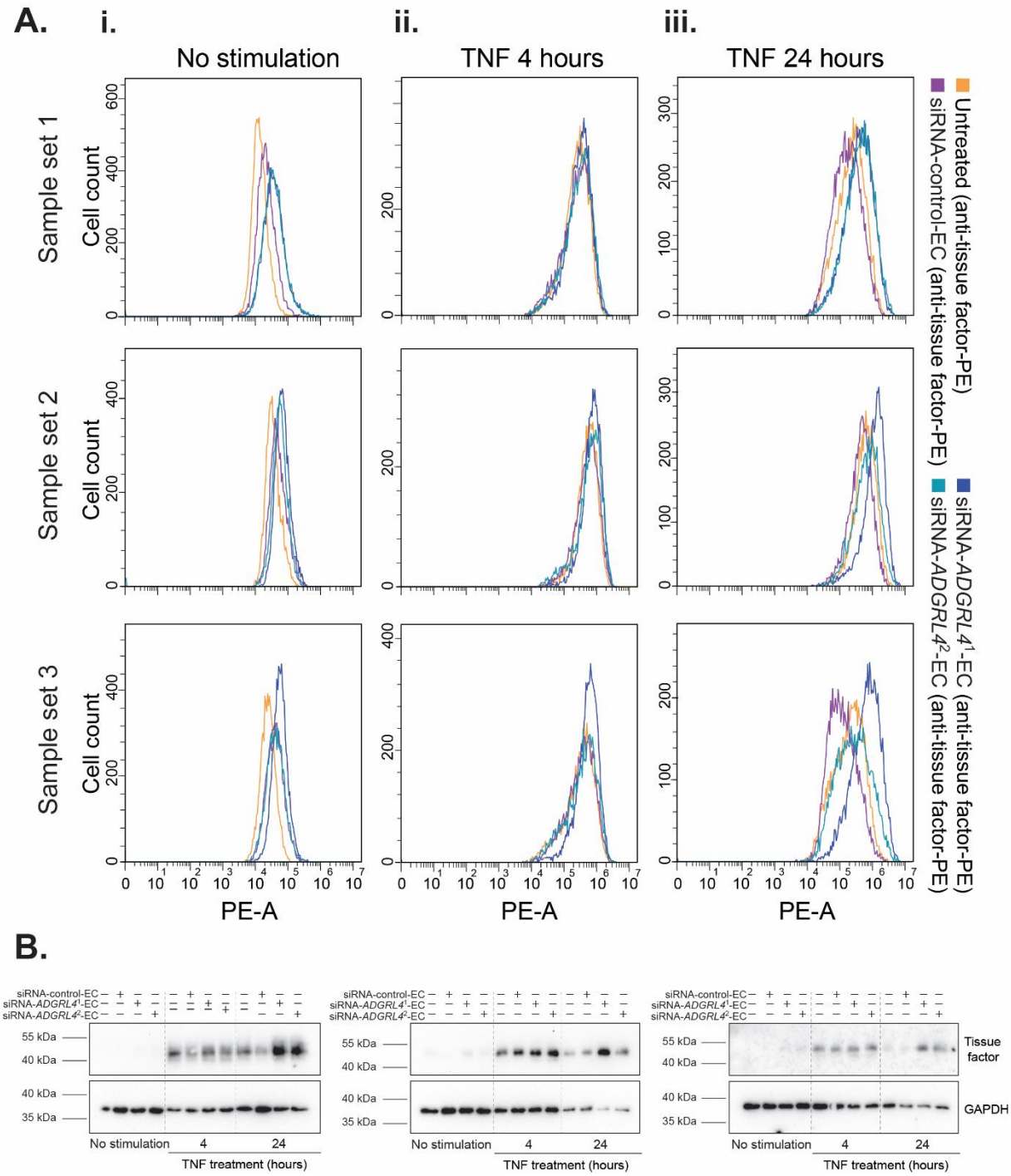
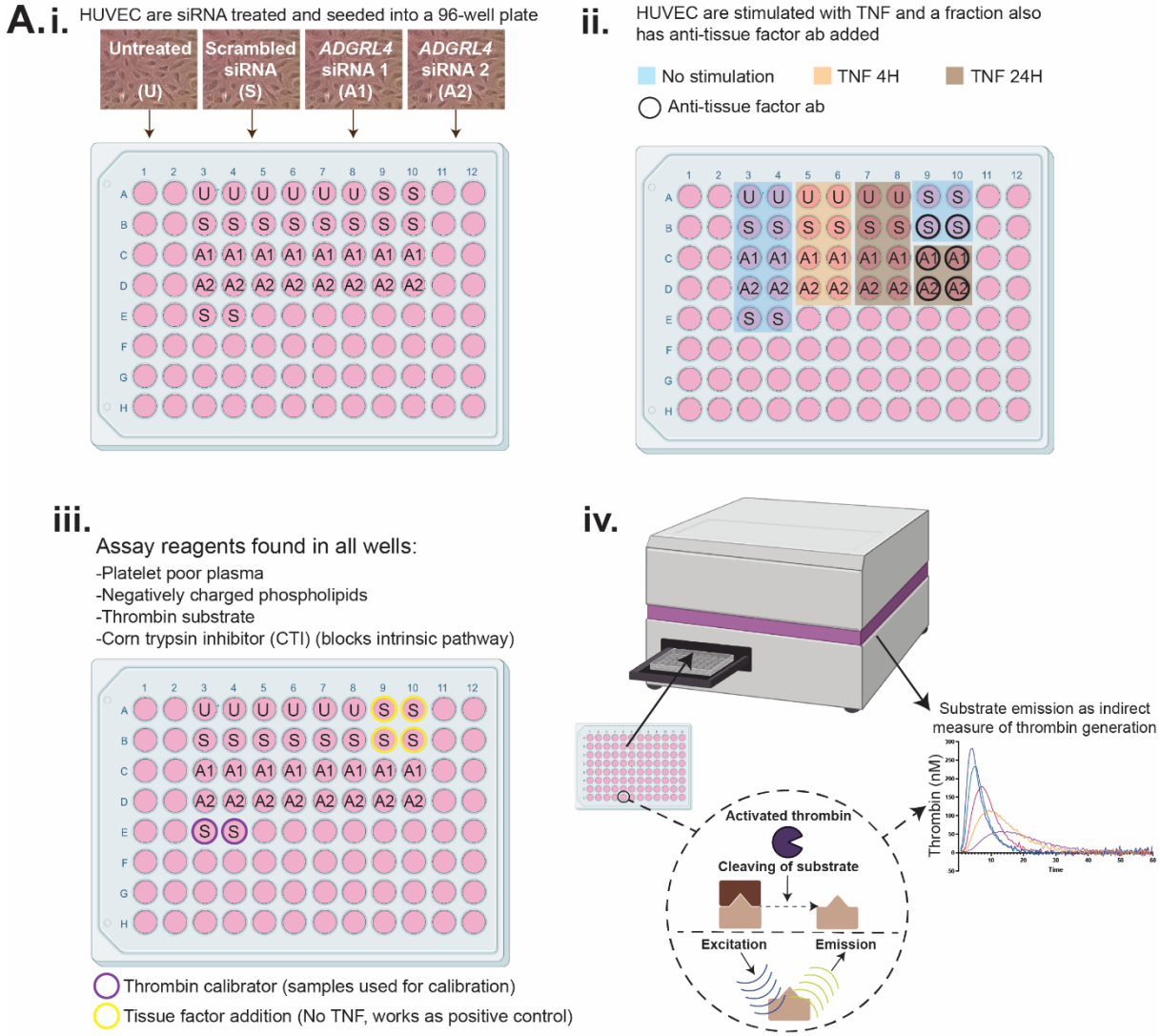


Figure S3. Endothelial *ADGRL4* depletion leads to reduced levels of TNF-induced tissue factor during the later phase of the response. Related to Figure 3. Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*^{1}-EC, or siRNA^{2}-*ADGRL4*-EC were treated with or without TNF for 4 and 24 before samples before collection and measurement of **(A)** tissue factor protein (PE-conjugated antibody) where signal intensity distribution for cells is plotted on x-axis (log-scale) and cell numbers on y-axis for various time points (Relates to Figure 3B), or **(B)** tissue factor protein in the entire cell lysate by western blot analysis (n=3 biological replicates) (Relates to Figure 3C).

Figure S4


A.i.
HUVEC are siRNA treated and seeded into a 96-well plate
Untreated (U)
Scrambled siRNA (S)
ADGRL4 siRNA 1 (A1)
ADGRL4 siRNA 2 (A2)
ii.
HUVEC are stimulated with TNF and a fraction also has anti-tissue factor ab added
No stimulation
TNF 4H
TNF 24H
Anti-tissue factor ab
iii.
Assay reagents found in all wells:
-Platelet poor plasma
-Negatively charged phospholipids
-Thrombin substrate
-Corn trypsin inhibitor (CTI) (blocks intrinsic pathway)
Thrombin calibrator (samples used for calibration)
Tissue factor addition (No TNF, works as positive control)
iv.
Substrate emission as indirect measure of thrombin generation
Activated thrombin
Cleaving of substrate
Excitation
Emission
Thrombin (nM)
Time

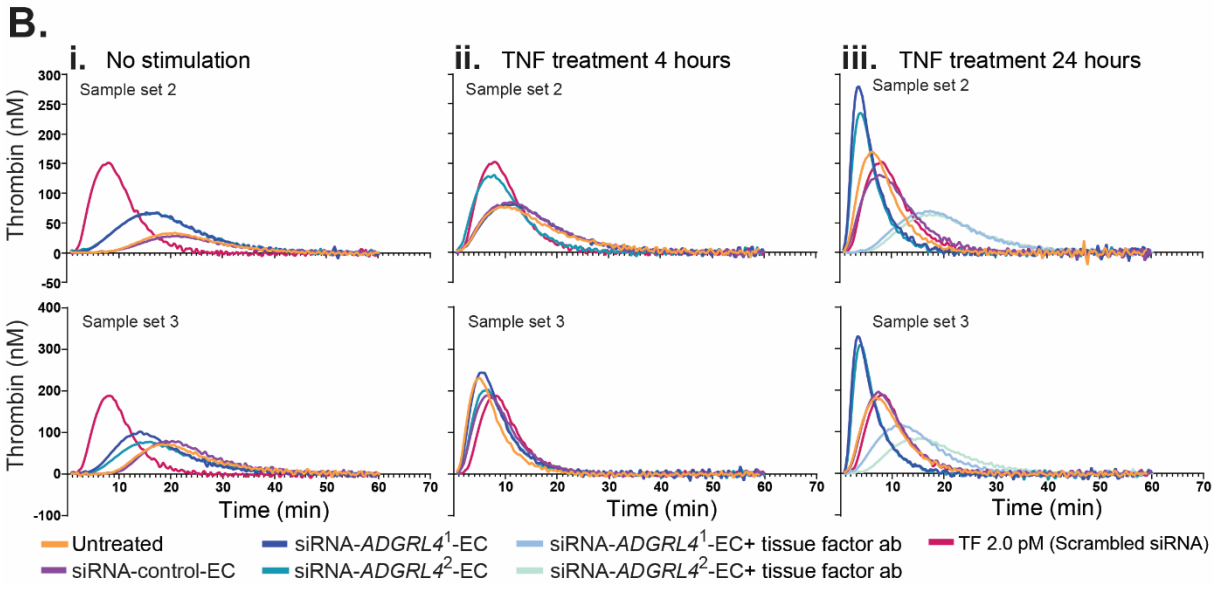

B.
i. No stimulation
ii. TNF treatment 4 hours
iii. TNF treatment 24 hours
Sample set 2
Sample set 3
Thrombin (nM)
Time (min)
Untreated
siRNA-control-EC
siRNA-ADGRL4¹-EC
siRNA-ADGRL4²-EC
siRNA-ADGRL4¹-EC+ tissue factor ab
siRNA-ADGRL4²-EC+ tissue factor ab
TF 2.0 pM (Scrambled siRNA)

Figure S4. Thrombin generation assay. Related to Figure 4. **(A) (i)** Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*[1]-EC, or siRNA[2]-*ADGRL4*-EC were seeded into 96-wells in duplicates and treated (ii) without (blue) or with TNF for 4 hours (orange) or 24 hours (brown). Selected wells were treated with tissue factor blocking antibody (black circles). **(iii)** Platelet poor plasma, negatively charged phospholipids and corn trypsin inhibitor (to block the intrinsic pathway) were added to all wells,. Two wells received thrombin calibrator reagent (purple circles) with a known concentration of thrombin, as to be used for calculating thrombin generated in all the other wells. Four wells with unstimulated siRNA-control-EC, acting as positive controls, received additional tissue factor (two with and two without tissue factor blocking antibody) to show the thrombin generation potential (yellow circles). **(iv)** Thrombin substrate was added and the fluorescent signal generated from the cleavage of thrombin substrate was measured can used to calculate thrombin generated in each sample. **(B)** Thrombin generation curves for biological replicates not presented in Figure 4 on: **(i)** unstimulated cells or cells receiving TNF-treatment for **(ii)** 4 hours and **(iii)** 24 hours.

Figure S5

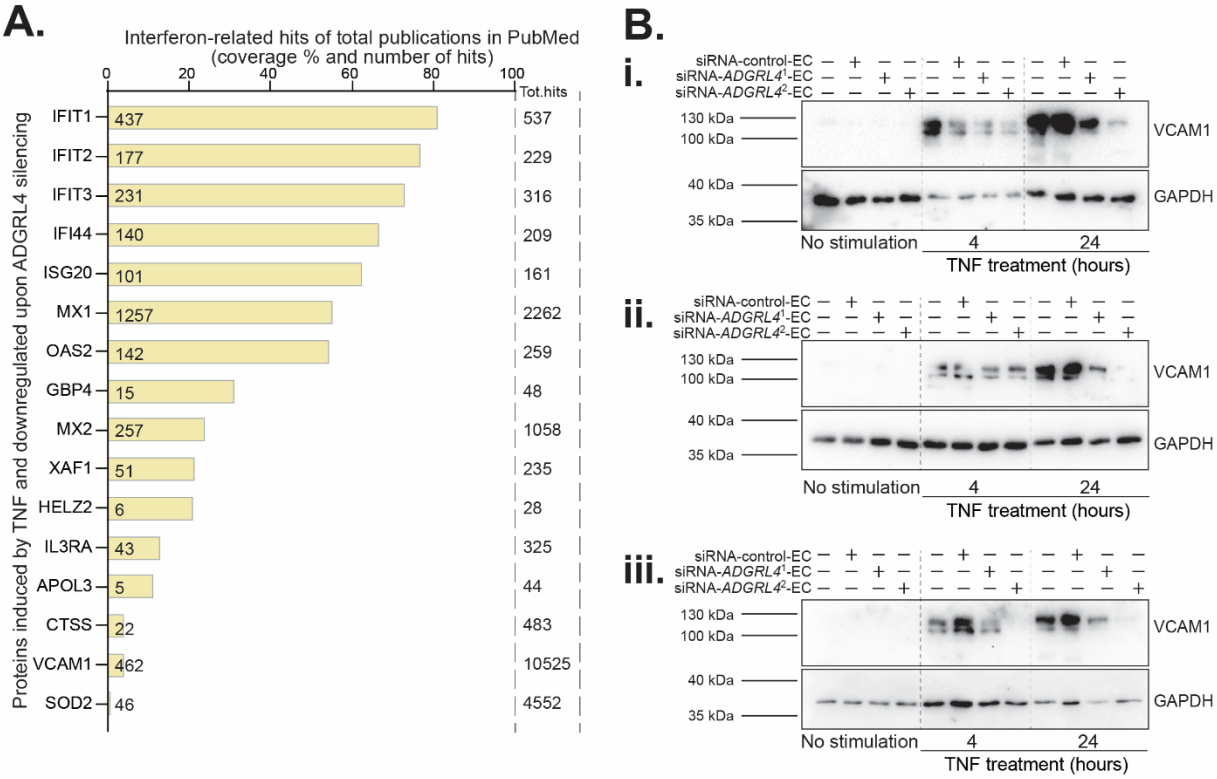
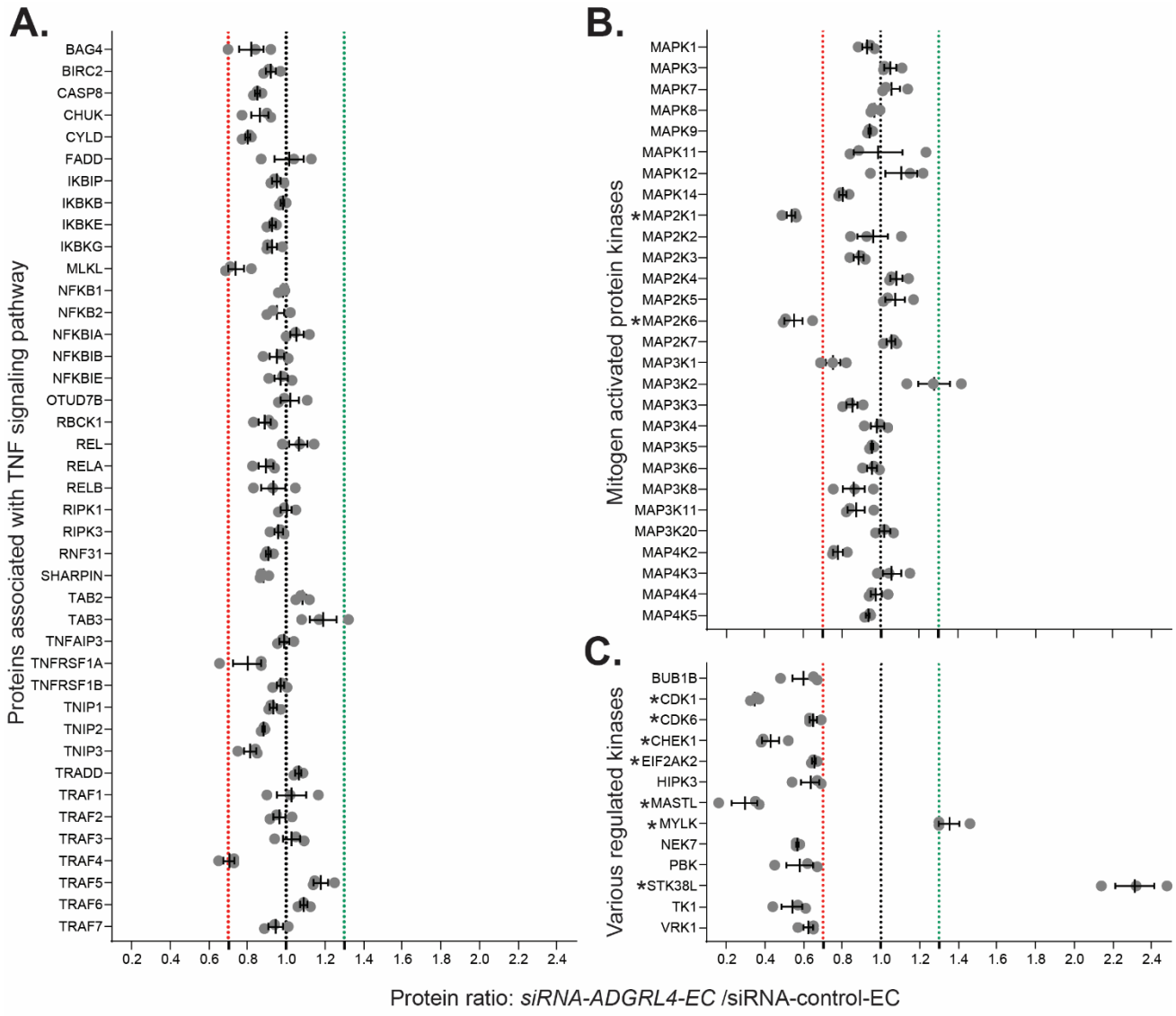


Figure S5: Endothelial *ADGRL4* depletion modifies TNF-stimulated interferon pathway proteins. Related to Figure 6 and 8. **(A)** An automated PubMed search was carried out to identify number of published hits for each of the individual 16 proteins identified as TNF-induced and downregulated in response to ADGRL4 silencing. The graph shows proteins ranked according to percentage of interferon-related hits (search: 'protein name' and 'interferon') out of the total hits- ('protein name') for an individual protein (Relates to Figure 6B-C). **(B)** Non-transfected EC, siRNA-control-EC, siRNA-*ADGRL4*^{1}-EC, or siRNA^{2}-*ADGRL4*-EC were treated with or without TNF. Samples were collected at 4 and 24 hours before measurement of protein content of VCAM1 in entire cell lysates by western blot analysis (n=3 biological replicates) (Related to Figure 8C).

Figure S6


A.
Proteins associated with TNF signaling pathway
BAG4
BIRC2
CASP8
CHUK
CYLD
FADD
IKBIP
IKBKB
IKBKE
IKBKG
MLKL
NFKB1
NFKB2
NFKBIA
NFKBIB
NFKBIE
OTUD7B
RBCK1
REL
RELA
RELB
RIPK1
RIPK3
RNF31
SHARPIN
TAB2
TAB3
TNFAIP3
TNFRSF1A
TNFRSF1B
TNIP1
TNIP2
TNIP3
TRADD
TRAF1
TRAF2
TRAF3
TRAF4
TRAF5
TRAF6
TRAF7
0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
B.
Mitogen activated protein kinases
MAPK1
MAPK3
MAPK7
MAPK8
MAPK9
MAPK11
MAPK12
MAPK14
*MAP2K1
MAP2K2
MAP2K3
MAP2K4
MAP2K5
*MAP2K6
MAP2K7
MAP3K1
MAP3K2
MAP3K3
MAP3K4
MAP3K5
MAP3K6
MAP3K8
MAP3K11
MAP3K20
MAP4K2
MAP4K3
MAP4K4
MAP4K5
C.
Various regulated kinases
BUB1B
*CDK1
*CDK6
*CHEK1
*EIF2AK2
HIPK3
*MASTL
*MYLK
NEK7
PBK
*STK38L
TK1
VRK1
0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 1.8 2.0 2.2 2.4
Protein ratio: siRNA-ADGRL4-EC /siRNA-control-EC

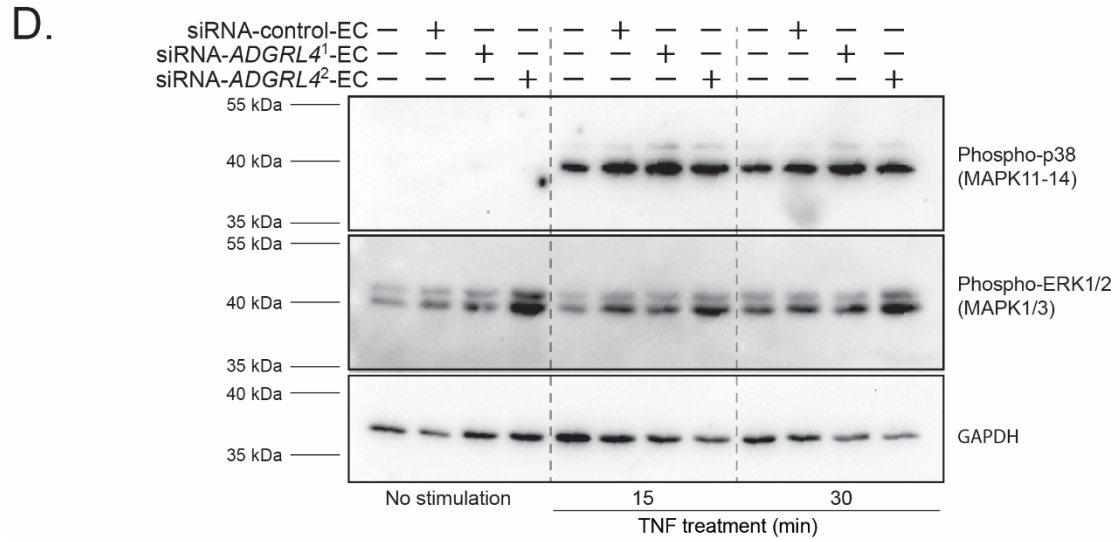

D.
siRNA-control-EC
siRNA-ADGRL4¹-EC
siRNA-ADGRL4²-EC
55 kDa
40 kDa
35 kDa
55 kDa
40 kDa
35 kDa
40 kDa
35 kDa
Phospho-p38 (MAPK11-14)
Phospho-ERK1/2 (MAPK1/3)
GAPDH
No stimulation
15
30
TNF treatment (min)

Figure S6: Effect of endothelial *ADGRL4* depletion on proteins related to the TNF-signaling pathway/kinases. Related to Figure 1,2,6. Human umbilical vein endothelial cells (HUVEC, n=3 independent sample sets [SS 1,2,3]) were transfected with siRNA targeting *ADGRL4* (siRNA-*ADGRL4*-EC) or a scrambled-control sequence (siRNA-control-EC) and analysed after 72 hours, using mass spectrometry. Each graph shows the fold change (FC) of siRNA-*ADGRL4*-EC relative to siRNA-control-EC. **(A)** Proteins involved in the TNF signaling pathway **(B)** All detected mitogen-activated protein kinases. **(C)** All kinases classified as regulated by ADGRL4 silencing. The dotted lines in green and red represent the ratio threshold for classification as up- (FC >1.3) or down- (FC <0.7) regulated, respectively, with each grey dot representing a biological replicate. **(D)** Protein levels of phosphorylated mitogen-activated protein kinase (MAPK) 11-14 (known as p38) and MAPK1 and MAPK3 (known as ERK1/2) assessed by western blot after ADGRL4 silencing and TNF treatment. **(B-C)** *Indicate same pattern of regulation upon TNF stimulation as for the ground state.


UiT The Arctic University of Norway