Università degli Studi di Padova

Facoltà di Medicina e Chirurgia

Corso di Laurea Magistrale in Medical Biotechnologies

Dipartimento di Medicina Molecolare

Anno Accademico 2021/2022

Tesi di Laurea Magistrale

Transcriptomic profiling of mesenchymal cells in fetal tissues

Laureanda: Giorgia D'Ariano

Relatore: Prof. Graziano Martello

Co-relatore: Dott. Marco Pellegrini

Prof. Paolo Martini

Indice

Abstract

Mesenchymal cells play a crucial role in both tissue development and regeneration. During embryonic and fetal stages, besides directly giving rise the body’s connective tissues, the mesenchyme also regulates organ formation by secreting signalling factors controlling, among others, proliferation, differentiation, positioning and branching. Although the importance of mesenchymal cells during development is known, the level of mesenchymal cells heterogeneity within and between organs has not been fully characterised. scRNA-seq allows to identify cell populations in a heterogenous system (e.g., fetal tissues) and to assess distribution of gene expression. Therefore, it can be used to discovering rare cell types and their markers, to identify differential cell composition between different conditions/tissues and to understand cells' relationship during development. Here I analysed four scRNA-seq datasets encompassing four fetal tissues (kidney, intestine, skeletal muscle and lung). For each dataset, I performed clustering to identify the main cellular compartments and to isolate the stromal fraction. Then, integration of mesenchymal cells from each dataset allowed me to identify both tissues-specific and conserved populations, to characterize them and to identify putative new markers.

Introduction

1.1. Mesenchyme

Mesenchyme is a tissue comprised of loose cells embedded in a hydrated mesh of fibrous (glycol)proteins called extracellular matrix (ECM)[1]. It is a transitive tissue involved in morphogenesis and development during embryonic/fetal life, which will give rise to connective tissue in adult organs[2].

Most mesenchyme derives from mesoderm, while a small number of mesenchymal cells in the head region originate from the neural crest, a specialized structure of the ectoderm. As soon as gastrulation starts in early embryonic life, some cell populations from the mesoderm undergo epithelial to mesenchymal transition (EMT) [3]. EMT consists in the loss of adhesive properties of epithelial cells via downregulation of adherent proteins, and acquisition of invasive motility and mesenchymal characteristics (Fig. 1). This allows cells to invade the extracellular matrix and travel to specific targets in the embryo where they can differentiate and/or induce differentiation of other cells by producing signalling factors.

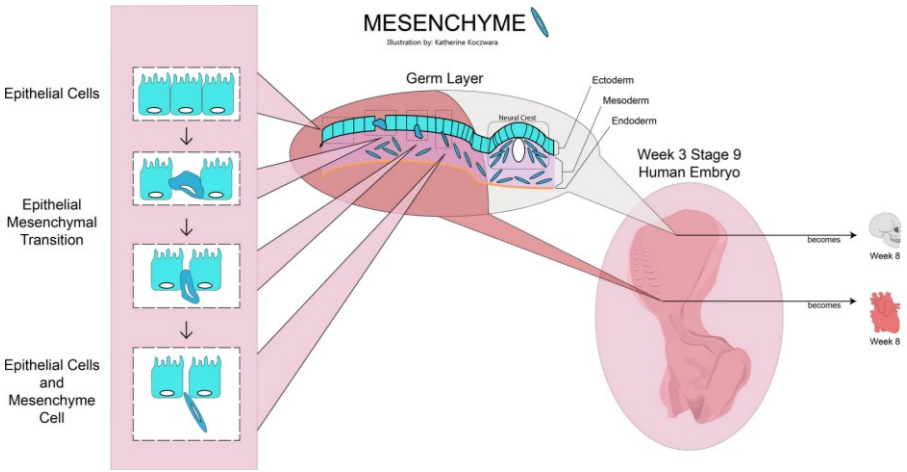


Figure 1[4] epithelial mesenchymal transition scheme in human embryo.

Some mesenchymal cells can also undergo the reverse process, the mesenchymal-epithelial transition (MET), that occurs when the loose cells of mesenchyme develop adhesive properties and arrange themselves into an organized sheet. This type of transition is also common during development and is involved, for example, in nephrons formation[5].

Due to its fluid nature, mesenchyme migrates easily, playing a crucial role in embryonic and fetal stages of animal life. Mesenchymal cells give origin to the body's connective tissues (bone, cartilage, stroma, muscle, lymphatic and circulatory system) and are involved in development of morphological structures by producing and remodelling the extracellular matrix (ECM) and by physical and functional interaction with the epithelium[3]. Therefore, in several developing tissues, despite giving rise to the connective compartment, mesenchyme has a trophic effect in organ morphogenesis by signalling the adjacent epithelium regulating its migration, proliferation and differentiation. Several steps of organ development rely on this epithelial-mesenchymal crosstalk.

1.2. Main mesenchymal cells populations

Despite mesenchymal cells populating the connective tissue in adult organs are widely investigated[6], the cell composition of the mesenchyme during development, and their function in organ maturation is still poorly understood.

1.2.1 Mesenchymal stromal cells

Mesenchymal stromal cells (MSCs) were firstly described as a population of bone-marrow derived cells, distinct from the Hematopoietic Stem Cell (HSC) population, displaying osteogenic properties in vivo and being clonogenic in vitro[7]. MSCs have been subsequently identified in several organs and have been assigned this name due to the ability to self-renew and differentiate *in vitro* into mesoderm cell types: osteoblasts, chondrocytes, adipocytes, and other cell types[8,9]. International Society for Cellular Therapy (ISCT) proposes minimal criteria defining *in vitro -expanded* MSCs[8]: 1) being plastic-adherent in culture; 2) exhibiting a set com- bination of surface antigens (CD73+, CD90+, CD105+, CD34-, CD45-, CD11b-, CD14-, CD19-, CD79a- and HLA-DR-); and 3) being able to differentiate in vitro into osteoblasts, chondrocytes and adipocytes. However, MSCs from different tissues have diverse phenotypic and functional behaviour, making tissues comparison difficult. Although the development of MSCs is poorly understood, MSCs-like cells have been identified in embryonic/fetal tissues like liver, skeletal muscle, blood, bone marrow, lung, kidney, aorta-gonad-mesonephros and yolk sac[6,10]. While the term “mesenchymal stem cells” refers to their *in vitro* properties, *in vivo* they represent a mesenchyme-derived reservoir of cells that are involved in tissue regeneration and repair via paracrine activity and EMC deposition/remodeling[11]. Although their function in fetal and embryonic tissues remains poorly investigated, it is suggested that they cover analogous functions helping organogenesis and tissues patterning[12]. Because MSCs are obtained by depletion of hematopoietic and endothelial cells from the tissues’ stromal fraction, they do not include a single cell type, rather a heterogeneous mix of cells. Also, the MSCs mixture is different across tissues, due to the different composition of the stromal compartment in each organ[6].

A relevant example of MSCs are Fibro Adipogenic Precursors (FAPs), also called adventitial fibroblasts, firstly described as muscle-resident stromal cells involved in skeletal muscle regeneration and subsequently identified in other tissues[11]. In muscle, they act as cellular sentinels and paracrine centres for tissue homeostasis and regeneration by secreting a mix of extracellular matrix components, diffusible cytokines, ligands, and immune-modulatory factors[13]. In addition to these functions, these cells respond to tissue injury by assuming a contractile phenotype (myofibroblasts). FAPs can also differentiate into specialized cells, including activated fibroblasts, adipocytes, and osteogenic cells after injury. It is known that fibroblasts of different tissues have different gene expression profiles and thus ECM composition varies across organs. However, the inter- and intra-organ heterogeneity of fibroblasts during fetal development has not been addressed yet.

Therefore, the relevance of MSCs in clinical practice is highlighted by both their differentiation properties and role as signalling cells promoting wound healing, haematopoiesis tissue regeneration and immunomodulation[6,14].

1.2.2. Mural cells

Pericytes and vascular smooth muscle cells (vSMCs) are contractile perivascular cells, such as cells localised around blood vessels, that are collectively called “mural cells”. They physically interact with vascular endothelial cells and communicate with them regulating vessel stability and blood flow. Because of their functions, mural cells are present in all vascularised tissues. Pericytes are discriminated from vSMCs mainly from a morphological point of view: while pericytes are embedded within the endothelial cells basal lamina and are found around smaller capillaries, vSMCs are localised around bigger vessels, like pre-capillary arteries and veins, and are placed outside the vascular basement membrane. From a molecular point of view, pericytes are described to be enriched in CSPG4 (or NG2), MCAM (or CD146), RGS5, PDGFRB and low levels of ACTA2, while smooth muscles cells specifically express high level of ACTA2, MYH11, CNN1 and SMTN. Although these markers are widely used, they are not absolute discriminants of pericytes and vSMCs. ACTA2, for example, is also found in myofibroblasts and pericytes’ marker MCAM is also reported to be expressed in endothelial cells and vSMCs[15]. This overlap in vSMCs and pericytes function and molecular signature, makes their discrimination difficult.

Pericytes encompass a series of functions spanning from angiogenesis, immunomodulation, clearance of toxic bio-products via their phagocytic activity, wound healing, and hematopoietic stem cells maintenance[16]. Remarkable molecules produced by pericytes that are involved in vessels maturation, are VEGF, ANGPT1, CSPG4 and TGF-β1[17]. All these functions make pericytes interesting for tissue engineering applications. For example, application of human adipose-derived pericytes on wounded skin of the rats had beneficial effects due to the increased angiogenesis, extensive collagen deposition, and epithelialization[18]. However, pericytes isolation is tricky due to the low specificity of known molecular markers, limiting their application in tissue engineering.

1.3. Aim of the study

For what has been discussed above, the study of mesenchyme and mesenchyme derived cells has a dual purpose: 1) having a comprehensive understanding of the different mesenchymal populations to understand their role during development and adult tissues regeneration 2) identify new cell populations with possible applications in the fields of regenerative medicine and tissue engineering.

Because mesenchymal cell heterogeneity has not been fully addressed, the aim of this thesis is to pave the way to understanding mesenchymal cells composition within and between tissues, analyse their dynamics across development and uncover differentially enriched biological processes. This information would help to define functional cell types and their regulation during development, gaining new insights into tissues morphogenesis and remodelling. The current work includes 4 tissues (kidney, lung, skeletal muscle and intestine) covering postconceptional weeks (PCW) 10-17 and represents the initial framework of our project. In the future we aim to include further tissues and samples (embryonic, peri-natal and adult) to observe how the composition of the mesenchymal compartment changes during development and in different organs.

To address this issue, I performed transcriptomic evaluation of kidney, lung, skeletal muscle and intestine data from available published datasets at single cell level. Indeed, single cell RNA sequencing (scRNAseq) analysis, allows to: 1) catalogue cell populations in a heterogenous system; 2) discover differential cell composition between different conditions/tissues and 3) assess distribution of gene expression. Starting from a matrix of RNA counts, tools like the R package “Seurat” allows to categorize cells into clusters based on the expression of highly variable genes. Each cluster contains cells with similar transcriptomic profile thus representing different cell types and/or expression phenotypes. Once cellular populations are identified, they can be characterised by evaluating the gene signature to obtain new insights into 1) cellular identity, 2) enriched pathways/cellular functions, 3) putative new markers useful to improve the current isolation protocols.

In this work, I analysed transcriptomic data of four different fetal organs, two of endodermal origin (lung and intestine), and two of mesodermal origin (skeletal muscle and kidney), focusing on the cell composition of their mesenchymal compartment, since cells of mesenchymal origin are known to regulate morphogenesis in all of them. As mentioned above, the fetal samples I analysed cover 10-17 PCW, so the first step was to understand how each tissue appears at this stage of development and which cell types are present.

1.3.1. Kidney

Nephrons, the basic functional units of the kidney, start to be induced as soon as the 7th week of gestation starts, and their characteristic histology is already established at the end of the PCW 8. Thus, fetal nephrons already shows differentiated cell types and specialised regions: glomerulus, proximal convoluted tubule, loop of Henle, macula densa and distant convoluted tubule. The glomerular apparatus shows presence of afferent and efferent arterioles, mesangial cells and podocytes. In the nephrogenic zone undifferentiated mesenchymal progenitors are still present during fetal life and start to decrease only at 30 weeks, when nephrogenesis is less intense. Therefore, at this stage of development we expect to see both differentiated and precursor epithelial cells, podocytes, differentiated and undifferentiated stromal cells and endothelium (Fig.2).

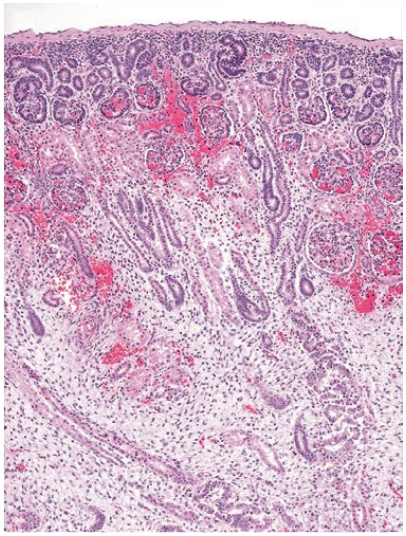


Figure 2[19]: at PCW 16, the lobe has already a recognizable structure: the peripheral portion consists of the cortex surrounding the medulla. The figure shows the cortex, also defined "nephrogenic region" at this stage. Indeed, glomeruli and their ducts are recognizable in the top part of the figure.

During the metanephron stage of mammalian kidney development (starting at around PCW 5 in human), the epithelial ureteric bud (UB) tips undergo repetitive branching until birth, giving rise to the collecting system (ureter, calyces and collecting ducts)[5,20,21]. Around the ureteric tips, a condensation of mesenchymal cells, called metanephric mesenchyme, are induced by the ureteric epithelia to start nephrogenesis. Part of metanephric mesenchyme forms the cap mesenchyme that will undergo MET to originate most of the cells of the nephron (parietal epithelial cells, podocytes, tubules, loops of Henle and connecting tubule)[5,21,22]. Interspersed in the cap mesenchyme, interstitial cells and interstitial precursors (called stromal mesenchyme) promote survival and differentiation of nephron progenitors and branching of the ureteric bud tips, as well as giving rise to fibroblasts, mural cells and endothelium[23]. A reciprocal crosstalk between epithelial cells and mesenchymal nephron progenitors is crucial for kidney morphogenesis[5,24]. On one hand, metanephric and cap mesenchyme release soluble ligands to signalling the UB tip to express specific transcription factors involved in outgrowth and morphogenesis. On the other hand, UB tip cells secrete signals to induce cap mesenchyme proliferation and differentiation toward nephron epithelia. Among these morphogens, the most relevant are GDNF-RET[25,26], WNT[27], FGF and NOTCH, which regulate self-renew, differentiation and positioning of nephron progenitors as well as UB branching. GDNF, for example, is a factor secreted from the metanephric mesenchyme and its receptor, RET, is located on the ureteric bud outgrowth. The GDNF-RET signalling guide ureteric budding via PI3K-AKT induced migration and proliferation. Indeed GDNF−/− and Ret−/− mice fail to form a ureteric bud and die perinatally with agenesis of both kidneys and ureters[25]. Many of the intercellular signalling networks controlling the morphogenesis of the urinary system are conserved in the development of other branched tubular organs such as the lung, mammary gland, and prostate. This suggest that similar functions are performed by mesenchymal cells of different organ.

1.3.2. Lung

Between PCW 8 and 16, the airways are in the so called "pseudo-glandular" stage of development (Fig. 3). This is the phase in which most of the organogenesis occurs: the bronchial tree forms, and primordial respiratory (acinar) structures appear for the first time. Indeed, airways branching is very intense at this stage: by PCW 14, 70% of the total airway generated at birth is formed and by the end of PCW 17, the formation of conducting airways and bronchioles is complete. Progressive differentiation of epithelial cells occurs, where 1) the more proximal airways become lined by ciliated and goblet cells 2) distally, cuboidal cells (immature type II cells) appear. Also, pulmonary arteries and veins are formed, even though capillaries – alveoli interface has not been established yet. Like the embryonic period, abundant loose interstitial mesenchyme appears condensed around bronchial buds. Cartilage formation around proximal airways and smooth muscle around airways and major vessels is also first appreciable during this stage.

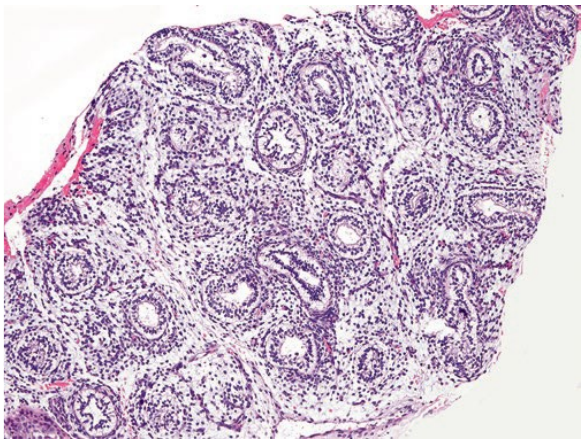


Figure 3[19]: between 10 -16 weeks of gestation, lung is composed of simple tubules embedded in a loose mesenchymal stroma.

The two primary epithelial buds that will give rise to the lungs, appear from the anterior foregut tube around PCW 4. Around the same period, the single foregut tube splits along the dorsal/ventral axis giving rise to the dorsal oesophagus and the ventral trachea that start elongation. The primary lung buds extend into the surrounding mesenchyme and begin the process of branching morphogenesis. Indeed, similarly to what just described in kidney, airways morphogenesis relies on inductive cues and reciprocal interaction between the pulmonary epithelium and the surrounding mesenchyme [28–30]. In particular, both dorsal/ventral patterning of the foregut tube and branching of the lung buds require signals released form the mesenchyme, including regulators of BMP, FGF, RA, TGFB and WNT pathways that control downstream effectors such as NMYC, ETV4/5, ERK1/2 and P38. For example, it is known that FGF10 and WNT2 are mesenchymal signal driving branching morphogenesis. Other processes controlled by the lung mesenchyme are alveolar maturation, vascular development and epithelial differentiation[31].

1.3.3. Intestine

The definitive histology of each gastro-intestinal (GI) tract is established within 14th week. The two main regions of the GI tube, the mucosa, submucosa and muscularis propria, acquire their unique features during fetal development.

Within PCW 14, the mucosa, a layer composed of the epithelium and the lamina propria, becomes specialised. Indeed, after a proliferative phase at week 6 (vacuolization) it starts to differentiate into the epithelium specific for each intestinal segment between PCW 8 and 10. At the base pf the epithelial crypts, neuroendocrine cells start to be present from PCW 11. Moreover, it is possible to appreciate enterocytes, goblet cells and epithelial crypts from PCW 12.

In the adult, the submucosa is composed of loose connective tissue containing nerves, ganglia, blood and lymphatic vessels, smooth muscle fibres and gut-associated lymphoid tissue. All these structures can be identified in intestinal histology already uring the fetal life. The muscularis propria appears for the first time in the oesophagus at PCW 6 and then spreads to the caudal region within PCW 8. The enteric nervous system, that includes specialised neural structure in both mucosa and muscularis propria, derives from neural

crest cells that start their migration toward the intestine at PCW 4. This process is completed within PCW 14. Thus, in fetal tissues of intestine it is possible to observe epithelial cells, mesenchyme, smooth muscle, neural cells, immune compartment and endothelial cells (Fig.3)

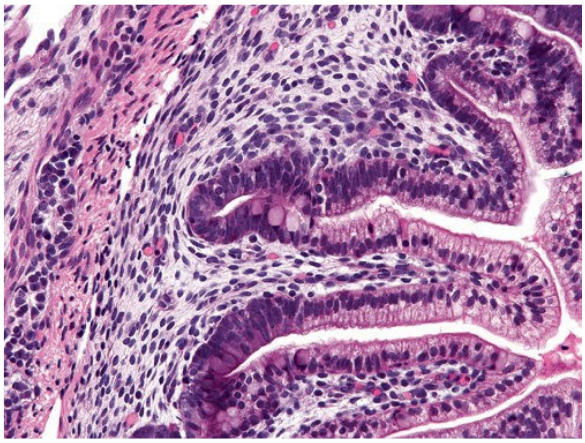


Figure 4[19]: Duodenum at 12 weeks postfertilization. The epithelium is already organised in crypts and can be observed both enterocytes and goblet cells. The muscularis propria is on the left, delimiting the mucosa.

In the adult intestine, PDGFRA+ stromal cells underly the epithelial layer and have been observed to act as signalling microenvironment for the epithelial stem cell compartment[32,33]. Indeed, the stem cells niche function is regulated by the stroma through signalling pathways including WNT, HH, NOTCH, BMP, and other growth factors. The PDGFRA+ population has a mesoderm origin, and it is already present in the mouse embryo where it induces epithelial sheet expansion and morphogenesis, orchestrated by stromal - epithelial crosstalk[33]. For example, the epithelium induces intestinal mesoderm expansion via Hh, Shh and Ihh secretion and in turn clusters of sub-epithelial PDGFRA-expressing cells promote villi formation (vilification), intestinal crypts establishment and intestinal stem cells maintenance. Also, during mice fetal development WNT and SHH signalling resulting from mesenchymal – epithelial interaction, induce intestinal regionalisation[34].

1.3.4. Skeletal muscle

Skeletal muscle development is subdivided in 2 phases of myogenesis: embryonic (or primary) myogenesis and fetal (or secondary) myogenesis. Embryonic myogenesis is characterised by rise of myoblasts precursors from the paraxial mesoderm that start fusion at 7 weeks, forming primary myofibers. In the fetal period, secondary myotubes begin to form as by fusion of fetal myoblasts. Since fetal myoblasts are larger and contain more myofilaments that embryonic ones, these secondary myotubes display contractility. Therefore, in fetal muscle is possible to observe myoblasts and myocytes expressing contractile proteins (heavy and light myosin chains) but also undifferentiated mesenchyma and blood vessels.

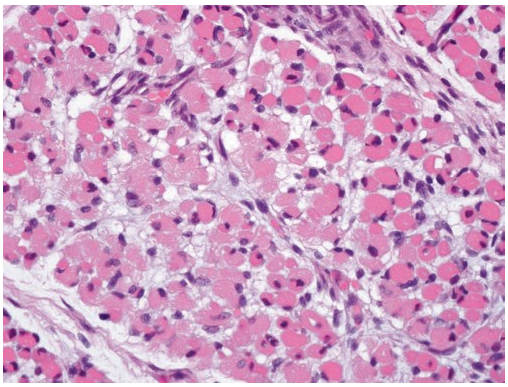


Figure 5 [19]: Cross-section of fetal skeletal muscle fibers at 17 weeks gestation.

The muscle connective tissue's main components are the extracellular matrix and its resident stromal cells, which dynamically reshape the muscle scaffold in embryonic development, homeostasis, and regeneration[11]. In this context, tissue-resident mesenchymal progenitors provide signalling cues that modulate other muscle-resident cells' function and remodel the ECM. In the adult, FAP cells secrete signals controlling muscle stem cells proliferation, fate and myogenesis after damage and in aging[13]. An analogous function of ECM remodelling and myogenesis regulation is carried out by mesodermal populations in the embryo and fetus. These cells have been observed to regulate spatiotemporal differentiation of myogenic progenitors, thus controlling limb patterning[35]. A mesenchymal – muscle cells crosstalk occurs during muscle morphogenesis. The muscle stromal compartment acts via secretion of signalling cues orchestrating myogenic cells repulsion, attraction, proliferation and differentiation. Also, they control the positioning of muscle progenitors by secreting and remodelling the ECM scaffold: these processes will define the size, shape and orientation of future myofibers.

Materials and methods

2.1. Data collection

Single-cell transcriptome data were extracted from the DESCARTES repository[36] (https://descartes.brotmanbaty.org/) where they are available for the public. I obtained 4 scRNA-seq datasets encompassing 4 fetal tissues: lung, kidney, intestine, skeletal muscle.

	Lung	**Kidney**	**Intestine**	**Skeletal muscle**
Number of cells in the dataset	180.223	178.603	59.470	47.537
Age range of the samples (weeks)	12 – 17	10 – 17	10 – 17	12 – 17
Number of reads	1.10×10^8	2.33×10^8	4.91×10^7	2.60×10^7

I removed all non-protein-coding genes by annotating them with the R-based package *org.Hs.eg.db* and retaining those reported as "protein coding". In this way I selected 19.582 coding genes.

2.2. Computational analysis

Clustering

I analysed the gene expression data downloaded from DESCARTES using the R-based package Seurat (version 4.1.2) [37], following the standard Seurat clustering workflow proposed by the Authors (https://satijalab.org/seurat/articles/pbmc3k_tutorial.html):

1) **Quality control**

   The quality control (QC) screening consists in the retaining of high-quality cells. QC metrics accepted by the community [38] are:

   - Number of genes in each cell: low-quality or empty droplets usually contain few genes.
   - Number of molecules detected in each cell: low-quality or empty droplets usually contain few molecules.
   - Percentage of reads mapping the mitochondrial genome: low-quality or dying cells are usually characterized by extensive mitochondrial contamination.

   After creation of the Seurat object, I filtered the dataset using the subset() function. I retained cells having:

   - Number of genes > 150
   - Number of counts >2
   - Percentage of mitochondrial genes < 5%

	Lung	Kidney	Intestine	Skeletal muscle
Number of cells in the dataset	180.223	178.603	59.470	47.537
Number of cells in the dataset post-filtering	**180.101**	**178.486**	**59.361**	**46.938**
Age range of the samples (weeks)	12 – 17	10 – 17	10 – 17	12 – 17

2) **Data normalization**
   I used the Seurat LogNormalize() function to normalize the gene expression measurements for each cell by the total expression, multiply it by a scale factor of 10.000 and log-transform the result.
3) **Identification of highly variable features**
   I used the Seurat FindVariableFeatures() function to calculate the 2.000 genes having high cell-to-cell variation in the dataset (i.e., those genes highly expressed in some cells and lowly expressed in others). The 2.000 variable features are used in the downstream analysis (PCA) since it has been observed [39] that focusing on them helps to highlight the most relevant biological differences among cells.
4) **Data scaling**
   I scaled the data using the ScaleData() function that:
   - Shifts the expression of each gene, so that the mean expression across cells is 0.
   - Scales the gene expression of each feature, so that the variance across cells is 1.

   In this way highly expressed genes are not given more weight in the downstream analysis. This step is required for the following dimensional reduction step.
5) **Perform linear dimensional reduction**
   I performed PCA (Principal Component Analysis) using the RunPCA() function on the scaled data and taking as input the variable features previously calculated.
6) **Cluster the cells**
   I performed clustering using the Seurat functions FindNeighbors() and FindCluster(). The first function takes as input the dimensionality of the dataset (I used the fist 10 PCs) to obtain a KNN (K-Nearest Neighbour) graph of the cells, with edges drawn between cells having similar transcriptome. The second function uses the Louvain algorithm to group cells together according to a chosen resolution. The resolution parameter determines the number of clusters (cell types) you get, so its value depends on dimension and heterogeneity of the dataset and must be chosen each time. To set this parameter, I used the *Clustree* tool [40] , that simulates how cells group at different resolution values. I usually set higher resolution - between 0.3 and 0.6 -

to cluster the entire tissues (a heterogeneous system) and a lower resolution – 0.2 – to cluster the mesenchymal cells (a more homogeneous group of cells).

DATASET	RESOLUTION
Lung	0.3
Kidney	0.4
Intestine	0.6
Skeletal muscle	0.3
Integration of mesenchymal cells from all the tissues	0.2

Clusters annotation and characterization

Clusters were firstly annotated manually and then cell types were confirmed by gene set enrichment analysis (GSEA). In the first case, expression of cell type specific markers from literature was calculated using the FindAllMarkers() and FoldChange() functions and evaluated via Dot Plots and Features Plots. In the second case, GSEA was performed by retrieving the cell type signature gene sets available on the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/) and calculating the enrichment scores and p-values of the gene sets via the R package *fgsea*. The gene sets with a p-value <0.5 were regarded as statistically significant.

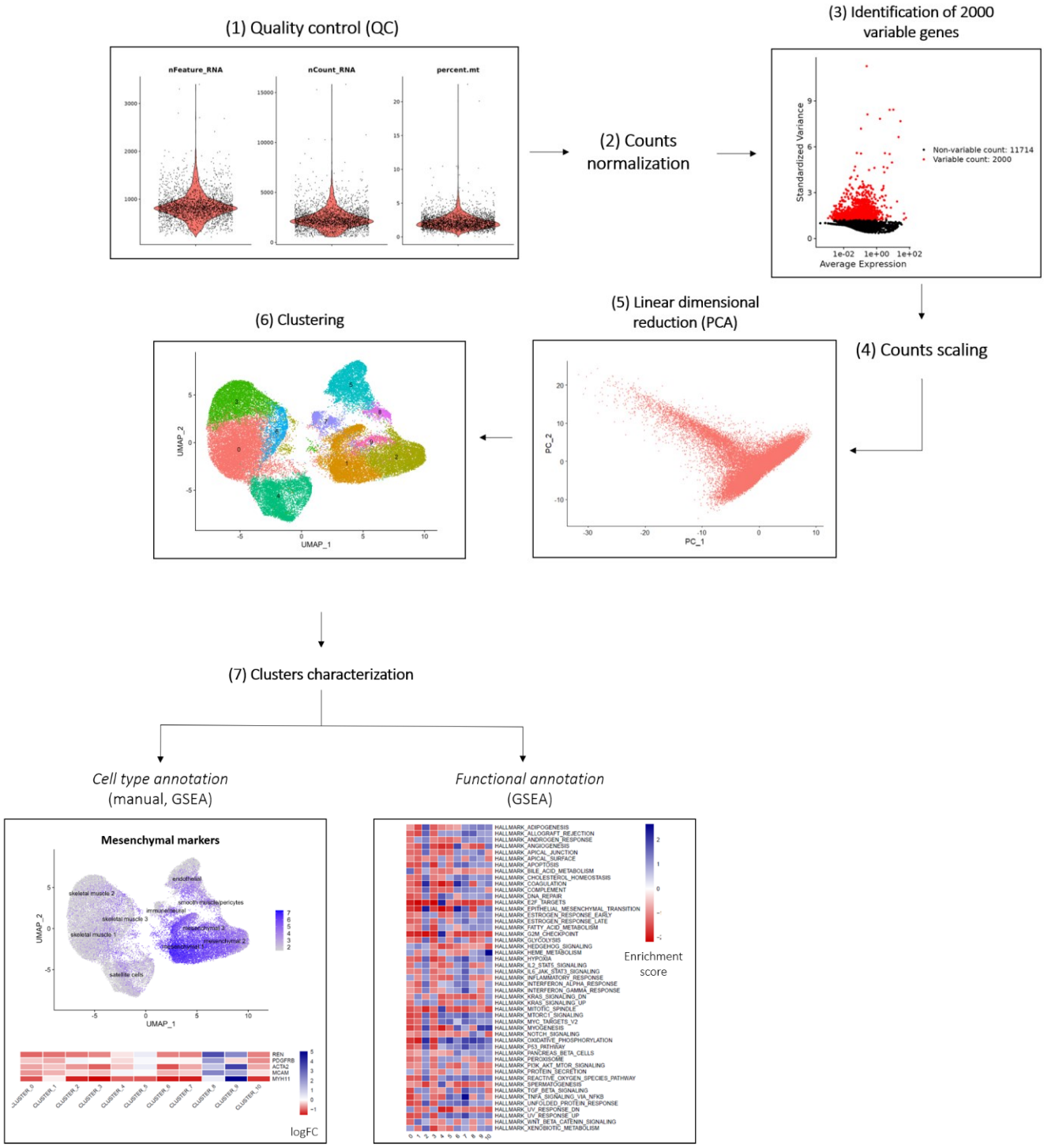
In the same way, GSEA was exploited to infer biological function and enriched pathways for each cluster by exploiting the hallmark gene sets and the Reactome gene sets available on the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/).

Genes used for the manual annotation are shown in the next table.

Cellular compartment	***Cell type***	***Markers (from literature)***
Mesenchymal	Pan - mesenchymal	COL1A1, COL1A2, COL3A1, VIM, CD90, CD105, CD34, CD44, CD73.
	Pericytes	ACTA2, PDGFRB, RGS5, MCAM, CSPG4.
	Smooth muscle	ACTA2, MYH11, MYOCD, SMNT, CNN1, TAGLN.
	Stromal mesenchyme (kidney)	FOXD1, LGALS1, MEIS1, NOTCH2, SFRP1, TBX3.
	Cap mesenchyme (kidney)	CITED1, EYA1, GDNF, HNF1A. ITGA8, MEIS1, SFRP1, SIX1, SIX2, TMEM100, WT1.
	Mesothelial	MSLN, UPK3B, WT1.
	Tenocytes (skeletal muscle)	TNMD.
	Pan – epithelial	EPCAM, CDH1.
	Enterocytes (intestine)	FABP1, FABP2, APOE, CDX1, CDX2.

Epithelial	Enteroendocrine (intestine)	CHGA, GHRL, NEUROD1, TPH1.
	AT1/AT2 (lung)	ETV5, MUC1, SFTPB, SFTPC, SFTPD.
	Nephron progenitors (kidney)	JAG1, LHX1, GATA3, ELF5, PAX8, PAX2, ITGA8, SIX2, EYA1.
	Podocytes (kidney)	LMX1B, MAFB, VEGFA, PODXL.
Immune	Pan – immune	CD45, HLA-DRA.
	Megakaryocytes	ITGA2B, ITGB3, CLEC1B, GP9, PPBP.
Endothelial	Vascular endothelial	CDH5, CLDN5, VWF, NOSTRIN, TGFBR2, PDGFB, KDR, KIT
	Lymphatic endothelial	LYVE1, PDPN, PROX1, FLT4 (VEGFR3).
Skeletal muscle	Myocyte/precursors	ACTA1, ACTC1, DES, MYH2, MYH7, MYOD1.
Satellite cells		PAX7, PAX3, MYF5, MYH8.
Neural	Pan – neural	PHOX2B, HAND2, TUBB2B.
	Neuroendocrine	ASCL1, CALCA, CHGA.
Erythroblasts		HBB, HBG1, HBG2, HBA2, HEMGN, AHSP.
Proliferation		MKI67, UBE2C, TOP2A

Workflow scheme



Mesenchymal cells subset

I used the subset() function to isolate clusters of interest.

FIRST ROUND OF SUBSET: MESENCHYMAL AND VASCULAR ENDOTHELIAL CELLS

For each tissue, I isolated:

- the mesenchymal clusters, by evaluating the expression of known pan-mesenchymal markers: COL1A1, COL1A2, COL3A1, VIM, CD90, CD105, CD34, CD44, CD73.
- the vascular endothelial clusters, by evaluating the expression of known markers: CDH5, CLDN5, VWF, NOSTRIN, TGFBR2, PDGFB, KDR, KIT.

I did not include clusters enriched in epithelial (EPCAM, CDH1), immune (CD45) and neural (PHOX2B, HAND2, TUBB2B) markers.

The standard Seurat workflow was applied to the mesenchymal/vascular endothelial mix as already described.

SECOND ROUND OF SUBSET: MESENCHYMAL CELLS

From the mesenchymal/endothelial subset I isolated the mesenchymal compartment, by evaluating the expression of pan - mesenchymal markers: COL1A1, COL1A2, COL3A1, VIM, CD90.

Integration

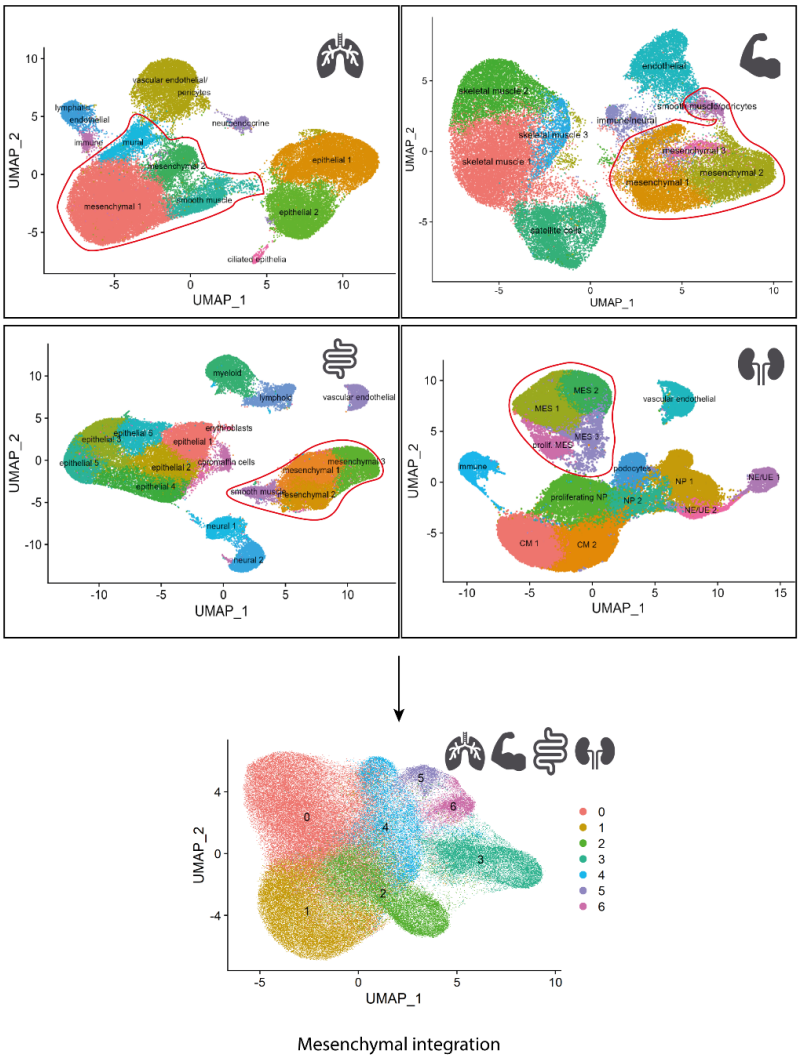
Mesenchymal cells from each tissue were merged and integrated to correct batch effect. I followed the standard Seurat workflow for datasets integration (https://satijalab.org/seurat/articles/integration_introduction.html):

1) **Split the Seurat object**
   The merged Seurat object was split in a list of 4 Seurat objects (corresponding to the 4 different tissues) using the SplitObject() function.
2) **Integration (batch effect correction)**
   Integration in Seurat is performed by mean of two functions:
   - FindIntegrationAnchors() identifies “anchors”, such as genes that can be used to integrate the dataset together.
   - IntegrateData() integrates the datasets.

Results

3.1. Whole tissue clustering

In the first phase, I performed clustering of four scRNA-sequencing datasets[36] obtained from fetal skeletal muscle, lung, intestine and kidney tissues ranging from 10 to 17 post-conception weeks (pcw). Each tissue has been clustered separately aiming to identify the main cellular compartments: epithelial, endothelial, immune, neural, mesenchymal and myogenic. Cell groups have been annotated manually searching for the expression of cell type specific markers from literature. The annotation was then validated via Gene Set Enrichment Analysis (GSEA), a method that correlates the cluster-specific gene expression with phenotypes available on the Molecular Signature Database. The “phenotypes” include cell identities, pathways, cellular functions, chromosome localization of the gene. GSEA feeds an enrichment score (ES) back: a positive ES indicates that cluster-specific genes are over-represented in the set of phenotype-specific genes, a negative ES indicates that cluster-specific genes are under-represented in the phenotype-specific ones. Once a general identity has been assigned, the mesenchymal group can be isolated to for further characterisation (Fig.6).



Mesenchymal integration

Figure 6 schematic representation of the analysis workflow. In a first step, clustering was performed on each tissue separately. Then, mesenchymal clusters (red circles) have been selected and merged, obtaining a mesenchymal dataset. Clustering was applied again to identify subclusters within the mesenchymal compartment.

3.1.1. Skeletal muscle clustering

Clustering of the skeletal muscle dataset at resolution 0.3 revealed the presence of 10 cellular subsets numbered from 0 to 9 (Fig.7A). A first manual screening of gene expression indicates the myogenic compartment includes clusters 0, 3, 4 and 6:

- Clusters 0, 3 and 6 express skeletal muscle cell markers (MYOD1, MYOG, MYH3, MYH7, ACTA1, DES), therefore representing fetal myocytes and myoblasts (Fig.7C).
- Cluster 4 is enriched in PAX7, marker of satellite cells (Fig.7C).

Pan – mesenchymal markers, including collagen genes and THY1 (CD90), are overexpressed in clusters 1, 2, 8 and 9 indicating that these clusters represent the mesenchymal compartment. Cluster 1 is expressing higher levels of COL1A1, COL1A2 and COL3A1. Also, clusters 8 is enriched in pericytes (ACTA2, PDGFRB, CSPG4, MCAM, RGS5) and smooth muscle (ACTA2, ACTG2, MYH11, MYOCD TAGLN) markers, indicating a mural cell phenotype. Apart from mural cells, no other stromal cell types could be clearly defined. Also, pericytes and smooth muscle cells cannot be resolved by simply increasing clustering resolution.

Despite all the cells of the dataset lack expression of PECAM1 (CD31), a well-known marker of vascular endothelial cells, cluster 5 has been annotated as “vascular endothelial cells” due to the enrichment in other canonical endothelial markers: CDH5, CLDN5, VWF, PDGRB, KDR, KIT, TEK, NOSTRIN, TGFBR2. Total absence of PECAM1 expression was also observed in the other three tissues, suggesting this is a bias related to sequencing procedure. A small percentage of cells (~10%) also expresses mural cell markers, and coherently the gene set enrichment analysis reveals an enrichment in smooth muscle features, although with lower enrichment score (Fig.7D).

Cluster 7 contains both immune and neural cells, as suggested by the co-expression of immune (PTPRC or CD45, HLA-DRA) and neural (S100B, SOX10, PLP1) genes. The cluster appears composed of two “lobes”: one expressing immune features and one with a neural phenotype (Fig.7E-F).

GSEA confirms these manually assigned identities (Fig.7D).

Table 1 shows a summary of clusters annotation.

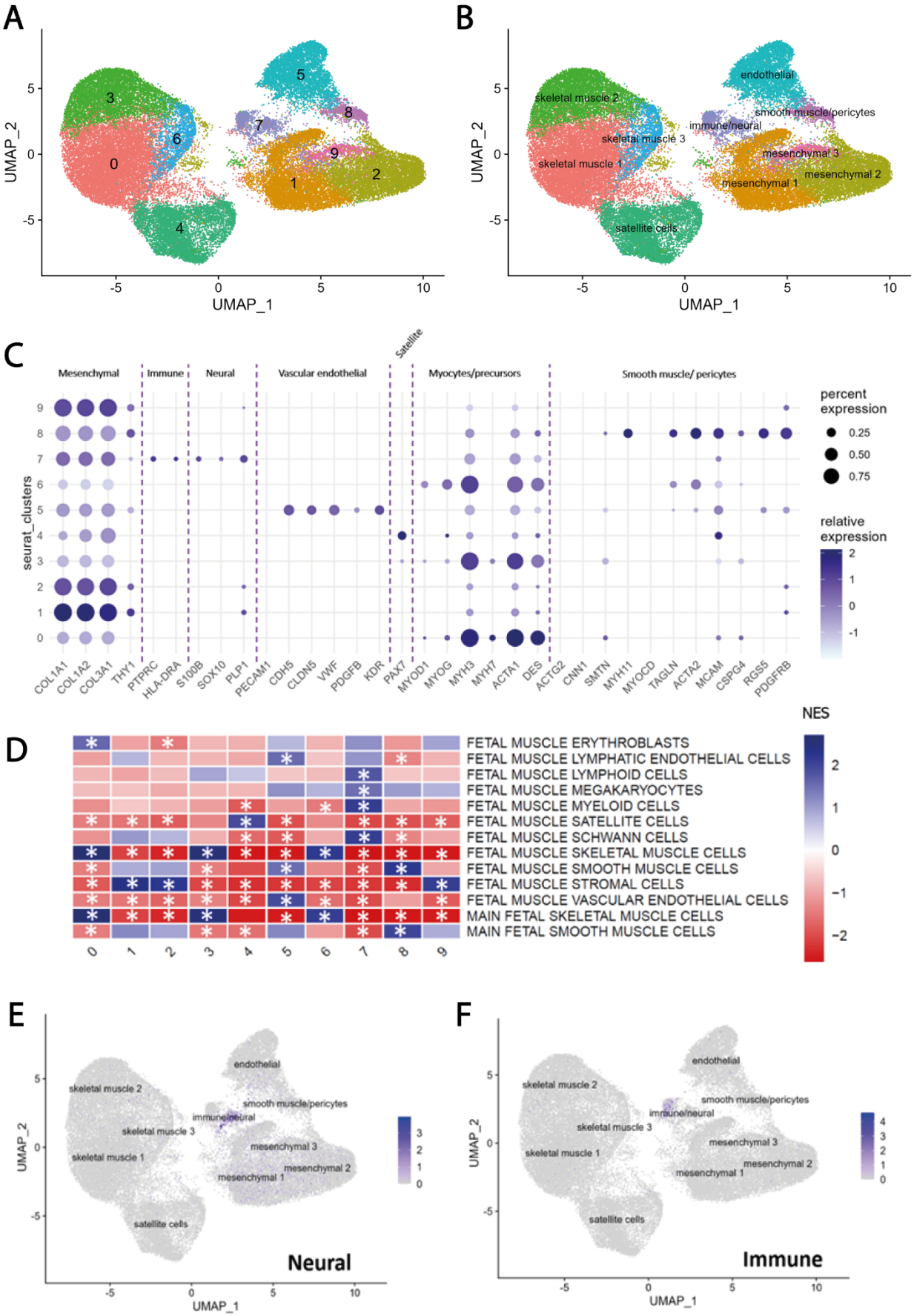

A
B
UMAP_2
UMAP_1
skeletal muscle 2
skeletal muscle 3
skeletal muscle 1
satellite cells
endothelial
smooth muscle/pericytes
immune/neural
mesenchymal 3
mesenchymal 2
mesenchymal 1
C
Mesenchymal
Immune
Neural
Vascular endothelial
Satellite
Myocytes/precursors
Smooth muscle/ pericytes
seurat_clusters
percent expression
relative expression
COL1A1 COL1A2 COL3A1 THY1 PTPRC HLA-DRA S100B SOX10 PLP1 PECAM1 CDH5 CLDN5 VWF PDGFB KDR PAX7 MYOD1 MYOG MYH3 MYH7 ACTA1 DES ACTG2 CNN1 SMTN MYH11 MYOCD TAGLN ACTA2 MCAM CSPG4 RGS5 PDGFRB
D
NES
FETAL MUSCLE ERYTHROBLASTS
FETAL MUSCLE LYMPHATIC ENDOTHELIAL CELLS
FETAL MUSCLE LYMPHOID CELLS
FETAL MUSCLE MEGAKARYOCYTES
FETAL MUSCLE MYELOID CELLS
FETAL MUSCLE SATELLITE CELLS
FETAL MUSCLE SCHWANN CELLS
FETAL MUSCLE SKELETAL MUSCLE CELLS
FETAL MUSCLE SMOOTH MUSCLE CELLS
FETAL MUSCLE STROMAL CELLS
FETAL MUSCLE VASCULAR ENDOTHELIAL CELLS
MAIN FETAL SKELETAL MUSCLE CELLS
MAIN FETAL SMOOTH MUSCLE CELLS
E
Neural
F
Immune

Figure 7 UMAP visualization of A) skeletal muscle dataset clustering and B) annotated clusters. C) Cell type GSEA of the 10 cell subsets. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. The asterisk indicates significant p-value < 0.05. D) Dot plot showing expression of cell type-specific genes in each cluster. UMAP showing the cumulative expression of E) immune and F) neural molecular markers.

Cluster	***Annotation***	***Manual markers***	***Top GSEA positively enriched phenotypes***
0	Skeletal muscle 1	MYOD1, MYOG, MYH3, MYH7, ACTA1, DES.	- Fetal muscle – skeletal muscle cells (p-val =1.00E-10) - Main fetal – skeletal muscle cells (p-val =1.00E-10) - Fetal muscle – erythroblasts (p-val =1.27E-04)
1	Mesenchymal 1	COL1A1, COL1A2, COL3A1, THY1 (CD90).	Fetal muscle – stromal cells (p-val =1.00E-10)
2	Mesenchymal 2	COL1A1, COL1A2, COL3A1, THY1 (CD90).	Fetal muscle – stromal cells (p-val =1.00E-10)
3	Skeletal muscle 2	MYOD1, MYOG, MYH3, MYH7, ACTA1, DES.	- Fetal muscle – skeletal muscle cells (p-val =1.00E-10) - Main fetal – skeletal muscle cells (p-val =1.00E-10)
4	Satellite cells	PAX7.	Fetal muscle – satellite cells (p-val =1.75E-08)
5	Vascular endothelial	CDH5, CLDN5, VWF, PDGFB, KDR.	- Fetal muscle – vascular endothelial (p-val =1.03E-10) - Fetal muscle – lymphatic endothelial (p-val =2.99E-03) - Fetal muscle – smooth muscle cells (p-val =7.10E-03)
6	Skeletal muscle 3	MYOD1, MYOG, MYH3, MYH7, ACTA1, DES.	- Fetal muscle – skeletal muscle cells (p-val =1.00E-10) - Main fetal – skeletal muscle cells (p-val =1.00E-10)
7	Immune and neural	Immune: PTPRC (CD45), HLA-DRA. Neural: S100B, SOX10, PLP1.	- Fetal muscle – myeloid cells (p-val =1.00E-10) - Fetal muscle – Schwann cells (p-val =1.00E-10) - Fetal muscle – lymphoid cells (p-val =1.67E-07)
8	Smooth muscle and pericytes	Smooth muscle: ACTA2, ACTG2,	- Fetal muscle – smooth muscle cells (p-val =1.00E-10)

		MYH11, MYOCD TAGLN, CNN1. Pericytes: ACTA2, PDGFRB, CSPG4, MCAM, RGS5.	- Main fetal – smooth muscle cells (p-val =1.00E-10)
9	Mesenchymal 3	COL1A1, COL1A2, COL3A1, THY1 (CD90).	Fetal muscle – stromal cells (p-val =1.00E-10)

Table 1 clusters annotation summary.

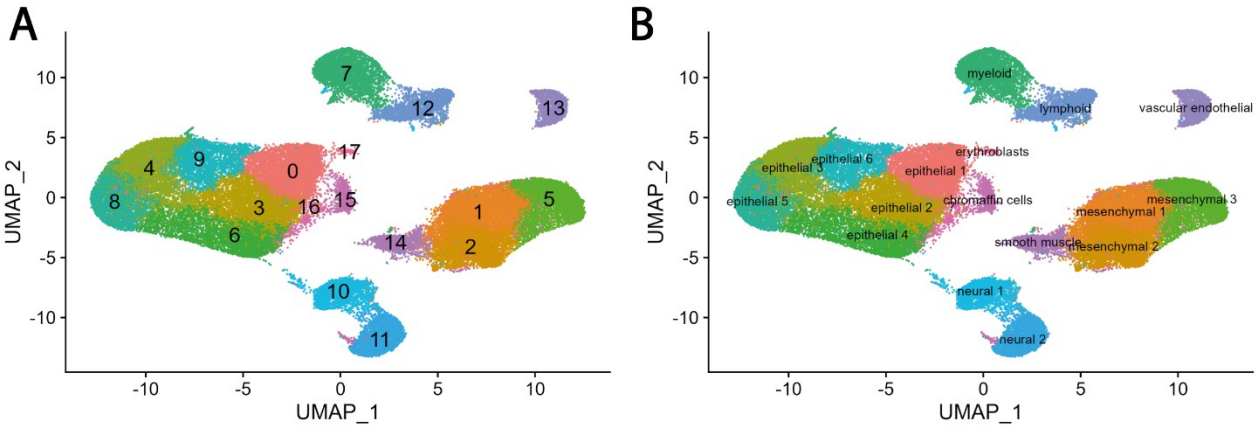
3.1.2. Intestine clustering

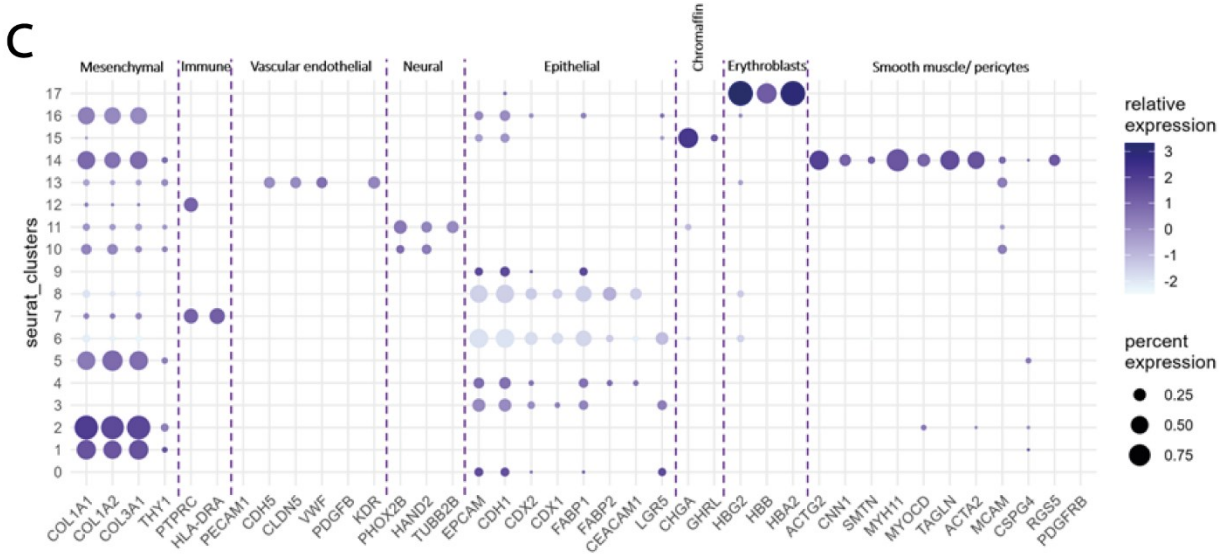
Clustering analysis of the intestine dataset revealed the presence of 18 clusters at resolution 0.6. Cellular groups 0, 3, 4, 6, 8, 9 and 15 show expression of pan – epithelial markers (EPCAM, CDH1) and variable levels of genes specifically expressed in the intestinal epithelia (CDX2, CDX1, FABP1, FABP2, CEACAM1), as shown in Fig.8C. Localised nearby the epithelial compartment, cluster 15 expresses chromaffin cells markers (CHGA, GHRL). LRG5, molecular discriminant of intestinal epithelial stem cells, is enriched in clusters 0, 3 and 6. Therefore, with the only exception of chromaffin cells, it is possible to isolate epithelial cells, but not to identify specific epithelial populations.

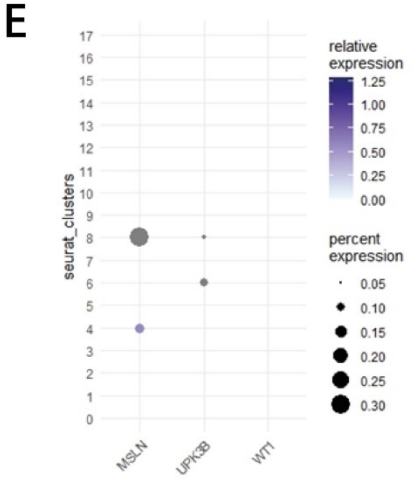
The stromal compartment includes clusters 1, 2, 5 and 14 all expressing pan - mesenchymal markers (COL1A1, COL2A1, COL3A1, THY1). Furthermore, cluster 14 shows a clear smooth muscle cells phenotype (enriched expression of ACTA2, ACTG, CNN1, SMTN, MYH11, MYOCD, TAGLN), while pericytes markers are expressed by less than 15% of cells. Cluster 16 shows co-expression of epithelial and mesenchymal markers. However, the list of DEGs (Differentially Expressed Genes) calculated with Seurat does not include COL1A1, COL2A1, COL3A1 and THY1 (CD90), while CDH1, like most of the genes, displays very low average log fold change (Fig.8F). Gene set enrichment analysis (Fig.8D) shows a statistically significant enrichment in the terms “epithelial cells”, “mesothelial cells”, “stromal cells” but with lower NES when compared with the other clusters. However, none of the cell groups expresses mesothelial cell markers (Fig.8E). In conclusion, it was not possible to annotate cluster 16 as clearly mesenchymal or epithelial and therefore it was not included in further analysis. Groups 7 and 12 constitute the immune compartment, as revealed by the expression of PTPRC (CD45) and HLA-DRA. Furthermore, GSEA indicates that myeloid and lymphoid phenotypes are enriched in cluster 7 and 12, respectively (Fig.8D). Endothelial cells are localized in cluster 13, that is expressing vascular markers (CDH5, CLDN5, VWF and KDR). Enteric nervous system cells are encompassed in groups 10 and 11 representing, respectively, enteric glial cells and neurons, as showed by the enrichment scores (Fig.8D). Finally, cluster 17 contains contaminating erythroblasts, as indicated by the expression of haemoglobin genes.

GSEA confirms these manually assigned identities (Fig.8D).

Table 2 shows a summary of clusters annotation.







F

cluster	gene	p-value	average log2FC
16	CLPS	1.63E-17	1.14
16	CEL	4.04E-09	1.14
16	CPA1	0.000129317	0.96
16	SPINK1	0.001919432	0.76
16	CFTR	0.006492002	0.67
16	PLEKHA7	0.008526993	0.64
16	ASIC2	9.30E-05	0.57
16	RHPN2	0.000167491	0.56
16	ZNRF3	0.000457384	0.55
16	SEMA3A	1.41E-18	0.51
16	ROBO2	9.14E-17	0.51
16	HPSE2	4.16E-38	0.50
16	GMDS	5.12E-06	0.48
16	DOCK1	3.11E-06	0.46
16	MAP4K3	0.001455009	0.45
16	DPP10	9.67E-06	0.45
16	DMBT1	2.16E-06	0.44
16	LIMK2	0.002381688	0.43
16	**CDH1**	0.001642247	0.43
16	PARD3	6.08E-10	0.42
16	HNF4A	2.38E-06	0.42
16	MID1	0.000464451	0.41
16	DIAPH3	0.000236304	0.41
16	DLG1	0.00200104	0.41
16	CLDN3	0.009728111	0.40
16	SHROOM3	0.009639235	0.40

Figure 8 UMAP visualization of A) intestine dataset clustering and B) annotated clusters. C) Cell type GSEA of the 18 cell subsets. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. The asterisk indicates significant p-value < 0.05. D) Dot plot showing expression of cell type-specific genes in each cluster. E) Dot plot showing expression of mesothelial cells markers. F) positive DEGs (Differentially Expressed Genes) calculated for cluster 16 and their statistics.

Cluster	***Annotation***	***Manual markers***	***Top GSEA positively enriched phenotypes***
0	Epithelial 1	EPCAM, CDH1, CDX2, CDX1, FABP1, FABP2, CEACAM1.	Fetal intestine – intestinal epithelial cells (p-val =8.53E-04)
1	Mesenchymal 1	COL1A1, COL1A2, COL3A1, THY1 (CD90).	Fetal intestine – stromal cells (p-val =2.72E-08)
2	Mesenchymal 2	COL1A1, COL1A2, COL3A1, THY1 (CD90).	- Fetal intestine – stromal cells (p-val =3.16E-04) - Fetal intestine – mesothelial cells (p-val =2.18E-08)
3	Epithelial 2	EPCAM, CDH1, CDX2, CDX1, FABP1, FABP2, CEACAM1.	Fetal intestine – intestinal epithelial cells (p-val =1.343E-08)
4	Epithelial 3	EPCAM, CDH1, CDX2, CDX1, FABP1, FABP2, CEACAM1.	Fetal intestine – intestinal epithelial cells (p-val =1.10E-10)
5	Mesenchymal 3	COL1A1, COL1A2, COL3A1, THY1 (CD90).	Fetal intestine – stromal cells (p-val =2.97E-05)
6	Epithelial 4	EPCAM, CDH1, CDX2, CDX1, FABP1, FABP2, CEACAM1.	Fetal intestine – intestinal epithelial cells (p-val =1.10E-10)
7	Myeloid	PTPRC (CD45), HLA-DRA.	Fetal intestine – myeloid cells (p-val =1.10E-10)
8	Epithelial 5	EPCAM, CDH1, CDX2, CDX1, FABP1, FABP2, CEACAM1.	Fetal intestine – intestinal epithelial cells (p-val =1.10E-10)
9	Epithelial 6	EPCAM, CDH1, CDX2, CDX1, FABP1, FABP2, CEACAM1.	Fetal intestine – intestinal epithelial cells (p-val =1.10E-10)
10	Neural 1	PHOX2B, HAND2, TUBB2B.	Fetal intestine – ENS glia (p-val =1.10E-10)
11	Neural 2	PHOX2B, HAND2, TUBB2B.	Fetal intestine – ENS neurons (p-val =1.10E-10)
12	Lymphoid	PTPRC (CD45), HLA-DRA.	Fetal intestine – lymphoid cells (p-val =1.10E-10)
13	Vascular endothelial	CDH5, CLDN5, VWF, PDGFB, KDR.	- Fetal intestine – vascular endothelial (p-val =3.02E-10)

			- Fetal intestine – lymphatic endothelial (p-val =2.62E-05)
14	Smooth muscle	ACTA2, ACTG2, MYH11, MYOCD TAGLN, CNN1.	Fetal intestine – smooth muscle cells (p-val =1.10E-10)
15	Chromaffin cells	CHGA, CHRL.	Fetal intestine – chromaffin cells (p-val =1.10E-10)
16	N/D		- Fetal intestine – intestinal epithelial cells (p-val =3.80E-06) - Fetal intestine – stromal cells (p-val =1.08E-02) - Fetal intestine – mesothelial cells (p-val =5.86E-03)
17	Erythroblasts	HBG2, HBB, HBA2.	Fetal intestine – erythroblasts (p-val =1.10E-10)

Table 2 clusters annotation summary.

3.1.3. Kidney clustering

Kidney dataset cells separate in 14 groups when clustered at resolution 0.4. The epithelial compartment includes clusters 0, 1, 2, 4, 6, 9, 11 and 13, which are enriched in genes of kidney epithelial precursors: ELF5, LHX1, JAG1, PAX8, PAX2, SI2, SIX1, EYA1, ITGA8 and GATA3 (Fig.9C). On the other hand, genes expressed in mature epithelia (EPCAM, CDH1) are mostly expressed in clusters 11 and 13. GSEA suggests the remaining epithelial clusters are epithelial precursors at different stages of differentiation (Fig.9D):

- Clusters 0 and 1 are enriched in “metanephric” and “cap mesenchyme” and thus represent nephron progenitors undergoing mesenchymal to epithelial transition.
- Cluster 4 is showing “cap mesenchyme” features and it is also expressing proliferation markers (Fig.9E), indicating it contains proliferating precursors.
- Cluster 6 is enriched in markers of “nephron progenitor”, the cells which derive from cap mesenchyme epithelialization during kidney development.
- Clusters 2, 10 and 13 are differentiating nephron progenitors and ureteric cells as showed by enrichment in gene sets of “nephron progenitors”, “proximal tubule cells”, “collecting duct cells”, “loop of Henle distal cells” and “ureteric bud cells”.
- Podocyte phenotype is associated with cluster 9 (LMX1B, MAFB, PODXL expression). The close association of podocytes with nephron progenitors is in agreement with their common origin (from cap mesenchyme).

In conclusion, kidney clustering analysis shows a good separation of the different stages of epithelial differentiation, however it is not possible to have positional information since proximal tubule, collecting duct and loop of Henle features are co-enriched in clusters 2, 11, 13.

Collagen genes and THY1 (CD90) are expressed in groups 3, 5, 10, 12 and 14, therefore annotated as mesenchymal clusters. Among these, cluster 14 shows very high expression of collagen genes in 100% of cells (Fig.9C), suggesting a matrix remodelling. Cluster 10 has high and significant enrichment score in “stromal cells” and “cap mesenchyme” gene sets, suggesting it could represent mesenchymal nephron progenitors. Cluster 12 is also enriched in proliferation genes (Fig.9E). Intriguingly, none of the stromal group shows a robust smooth muscle or pericyte phenotype, since their molecular markers are expressed by a low percentage of cells in all the stromal subsets (Fig.9C). Vascular endothelial markers are expressed in cluster 7, while cluster 8 is enriched in PTPRC (CD45) and HLA-DRA and has been annotated as “immune”. The latter, also contains a group of contaminating erythroblasts as shown by the significant enrichment for the “erythroblasts” gene set and the expression of known erythroblasts genes (HBG2, HBA2, HEMGN, AHSP, HBG1, HBB (Fig.9F). Few erythroblasts are also adjacent to cluster 4.

GSEA confirms these manually assigned identities (Fig.9D).

Table 3 shows a summary of clusters annotation.

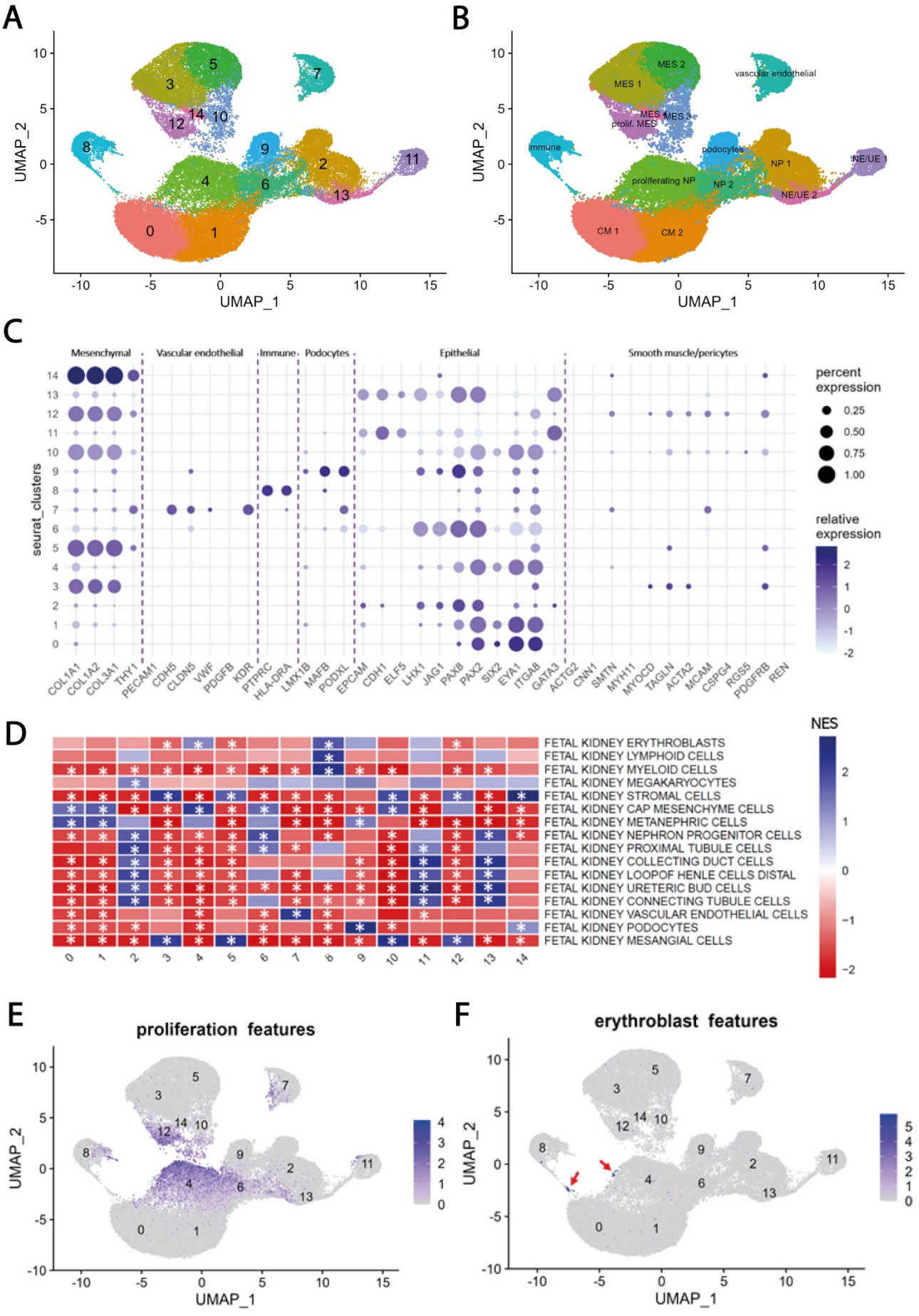

A
B
UMAP_2
UMAP_1
MES 1
MES 2
MES 3
prolif. MES
vascular endothelial
immune
podocytes
proliferating NP
NP 1
NP 2
NE/UE 1
NE/UE 2
CM 1
CM 2
C
Mesenchymal
Vascular endothelial
Immune
Podocytes
Epithelial
Smooth muscle/pericytes
seurat_clusters
percent expression
relative expression
COL1A1
COL1A2
COL3A1
THY1
PECAM1
CDH5
CLDN5
VWF
PDGFB
KDR
PTPRC
HLA-DRA
LMX1B
MAFB
PODXL
EPCAM
CDH1
ELF5
LHX1
JAG1
PAX8
PAX2
SIX2
EYA1
ITGA8
GATA3
ACTG2
CNN1
SMTN
MYH11
MYOCD
TAGLN
ACTA2
MCAM
CSPG4
RGS5
PDGFRB
REN
D
NES
FETAL KIDNEY ERYTHROBLASTS
FETAL KIDNEY LYMPHOID CELLS
FETAL KIDNEY MYELOID CELLS
FETAL KIDNEY MEGAKARYOCYTES
FETAL KIDNEY STROMAL CELLS
FETAL KIDNEY CAP MESENCHYME CELLS
FETAL KIDNEY METANEPHRIC CELLS
FETAL KIDNEY NEPHRON PROGENITOR CELLS
FETAL KIDNEY PROXIMAL TUBULE CELLS
FETAL KIDNEY COLLECTING DUCT CELLS
FETAL KIDNEY LOOPOF HENLE CELLS DISTAL
FETAL KIDNEY URETERIC BUD CELLS
FETAL KIDNEY CONNECTING TUBULE CELLS
FETAL KIDNEY VASCULAR ENDOTHELIAL CELLS
FETAL KIDNEY PODOCYTES
FETAL KIDNEY MESANGIAL CELLS
E
proliferation features
F
erythroblast features

Figure 9 UMAP visualization of A) kidney dataset clustering and B) annotated clusters (MES = mesenchyme; prolif. MES = proliferating mesenchyme; NP = nephron progenitors; CM = cap mesenchyme, NE/UE = nephron epithelia and ureteric epithelia; proliferating NP = proliferating nephron progenitors). C) Cell type GSEA of the 15 cell subsets. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. The asterisk indicates significant p-value < 0.05. D) Dot plot showing expression of cell type-specific genes in each cluster. E) cumulative expression of proliferation markers (MKI67, TOP2A, UBE2C). F) cumulative expression of erythroblasts markers (HBG2, HBA2, HEMGN, AHSP, HBG1, HBB).

Cluster	***Annotation***	***Manual markers***	***Top GSEA positively enriched phenotypes***
0	Cap mesenchyme 1	ELF5, LHX1, JAG1, PAX8, PAX2, SI2, SIX1, EYA1, ITGA8, GATA3.	- Fetal kidney – metanephric mesenchyme (p-val =3.05E-06) - Fetal kidney – cap mesenchyme (p-val =2.20E-03)
1	Cap mesenchyme 2	ELF5, LHX1, JAG1, PAX8, PAX2, SI2, SIX1, EYA1, ITGA8, GATA3.	- Fetal kidney – metanephric mesenchyme (p-val =2.68E-06) - Fetal kidney – cap mesenchyme (p-val =1.40E-04)
2	Nephron progenitors 1	ELF5, LHX1, JAG1, PAX8, PAX2, SI2, SIX1, EYA1, ITGA8, GATA3.	- Fetal kidney – proximal tubule cells (p-val =1.00E-10) - Fetal kidney – connecting tubule cells (p-val =1.00E-10) - Fetal kidney – loop of Henle (p-val =1.00E-10) - Fetal kindey – nephron progenitors (p-val =3.35E-04)
3	Mesenchymal 1	COL1A1, COL1A2, COL3A1, THY1 (CD90).	- Fetal kidney- stromal cells (p-val =1.00E-10) - Fetal kidney – mesangial cells (p-val =7.26E-10)
4	Proliferating nephron progenitors	Proliferation: MKI67, UBE2C, TOP2A. Nephron progenitors: ELF5, LHX1, JAG1, PAX8, PAX2, SI2, SIX1, EYA1, ITGA8, GATA3.	- Fetal kidney – erythroblasts (p-val =1.00E-10) - Fetal kidney – cap mesenchyme (p-val =1.00E-10)
5	Mesenchymal 2	COL1A1, COL1A2, COL3A1, THY1 (CD90).	- Fetal kidney- stromal cells (p-val =1.00E-10) - Fetal kidney – mesangial cells (p-val =1.00E-10)
6	Nephron progenitors 2	ELF5, LHX1, JAG1, PAX8, PAX2, SI2,	- Fetal kidney – erythroblasts (p-val =6.20E-08)

		SIX1, EYA1, ITGA8, GATA3.	- Fetal kidney – cap mesenchyme (p-val =6.42E-08)
7	Vascular endothelial	CDH5, CLDN5, VWF, PDGFB, KDR.	Fetal kidney – vascular endothelial (p-val =1.00E-10)
8	Immune	PTPRC (CD45), HLA-DRA.	- Fetal kidney – erythroblasts (p-val =1.00E-10) - Fetal kidney – myeloid (p-val=1.00E-10) - Fetal kidney – lymphoid (p-val=1.00E-10)
9	Podocytes	LMX1B, MAFB, PODXL	- Fetal kidney – podocytes (p-val = 1.00E-10) - Fetal kidney – metanephric mesenchyme (p-val=4.13E-02)
10	Mesenchymal 3	COL1A1, COL1A2, COL3A1, THY1 (CD90).	- Fetal kidney- stromal cells (p-val =1.00E-10) - Fetal kidney – mesangial cells (p-val =1.00E-10) - Fetal kidney – cap mesenchyme (p-val =5.69E-05)
11	Nephron epithelia and ureteric epithelia 1	EPCAM, CDH1, ELF5, LHX1, JAG1, PAX8, PAX2, SI2, SIX1, EYA1, ITGA8, GATA3.	- Fetal kidney – ureteric bud cells (p-val = 1.00E-10) - Fetal kidney – collecting duct cells(p-val = 1.00E-10) - Fetal kidney – connecting tubule cells (p-val = 4.81E-10) - Fetal kidney – loop of Henle (p-val=1.125-8)
12	Proliferating mesenchyme	Proliferation: MKI67, UBE2C, TOP2A. Mesenchyme: COL1A1, COL1A2, COL3A1, THY1 (CD90).	- Fetal kidney- stromal cells (p-val =1.00E-10) - Fetal kidney – mesangial cells (p-val =2.70E-05)
13	Nephron epithelia and ureteric epithelia 2	EPCAM, CDH1, ELF5, LHX1, JAG1, PAX8, PAX2, SI2, SIX1, EYA1, ITGA8, GATA3.	- Fetal kidney – ureteric bud cells (p-val =1.00E-10) - Fetal kidney – collecting duct cells (p-val =1.00E-10) - Fetal kidney – connecting tubule cells (p-val =1.00E-10) - Fetal kindey – nephron progenitors (p-val =6.04E-06) - Fetal kidney – loop of Henle (p-val =1.09E-08)

14	Mesenchymal 4	COL1A1, COL1A2, COL3A1, THY1 (CD90).	Fetal kidney- stromal cells (p-val =1.00E-10)

Table 3 clusters annotation summary.

3.1.4. Lung clustering

Lung dataset clusters in 11 cellular subsets. As shown in Fig.10C, the epithelial compartment includes clusters 1, 3, 8 and 10 which express general epithelial genes (EPCAM, CDH1) and groups 1, 3 and 10 express lung- specific epithelial markers (SFTPC, SFTPB, MUC1, ETV5). Cluster 10 is enriched in FOXJ1, gene identifying ciliated epithelial cells. In addition, cluster 8 has been annotated as “neuroendocrine” due to the expression of CALCA, CHGA, ASCL1.

The stromal subset contains clusters expressing collagen genes (COL1A1, COL1A2, COL3A1) and THY1 (CD90) and consists of groups 0, 4, 5, 6. Among these, cluster 5 has a smooth muscle phenotype (CNN1, MYH11, MYOCD, TAGLN, ACTA2), while a portion of cluster 6 shows a mixed phenotype between pericytes (ACTA2, MCAM, CSPG4, RGS5, PDGFRB) and smooth muscle (MYH11 and TAGLN). This suggests that cluster 5 could contain airway smooth muscle and cluster 6 consist of mural cells. As already observed in the skeletal muscle dataset, there is partial overlap of pericytes and vascular endothelial cells (CDH5, CLDN5, VWF, PDGFB, KDR) markers in cluster 2. Some vascular endothelial markers are also expressed by cluster 7, that is showing a clear lymphatic endothelial phenotype (positive for LYVE1, PDPN, PROX1). Furthermore, LYVE1 expression is shared with the immune cluster 9 (expressing CD45 and HLA-DRA), in agreement with the reported expression of this gene in macrophages[41,42].

GSEA confirms these manually assigned identities (Fig.10D).

Table 4 shows a summary of clusters annotation.

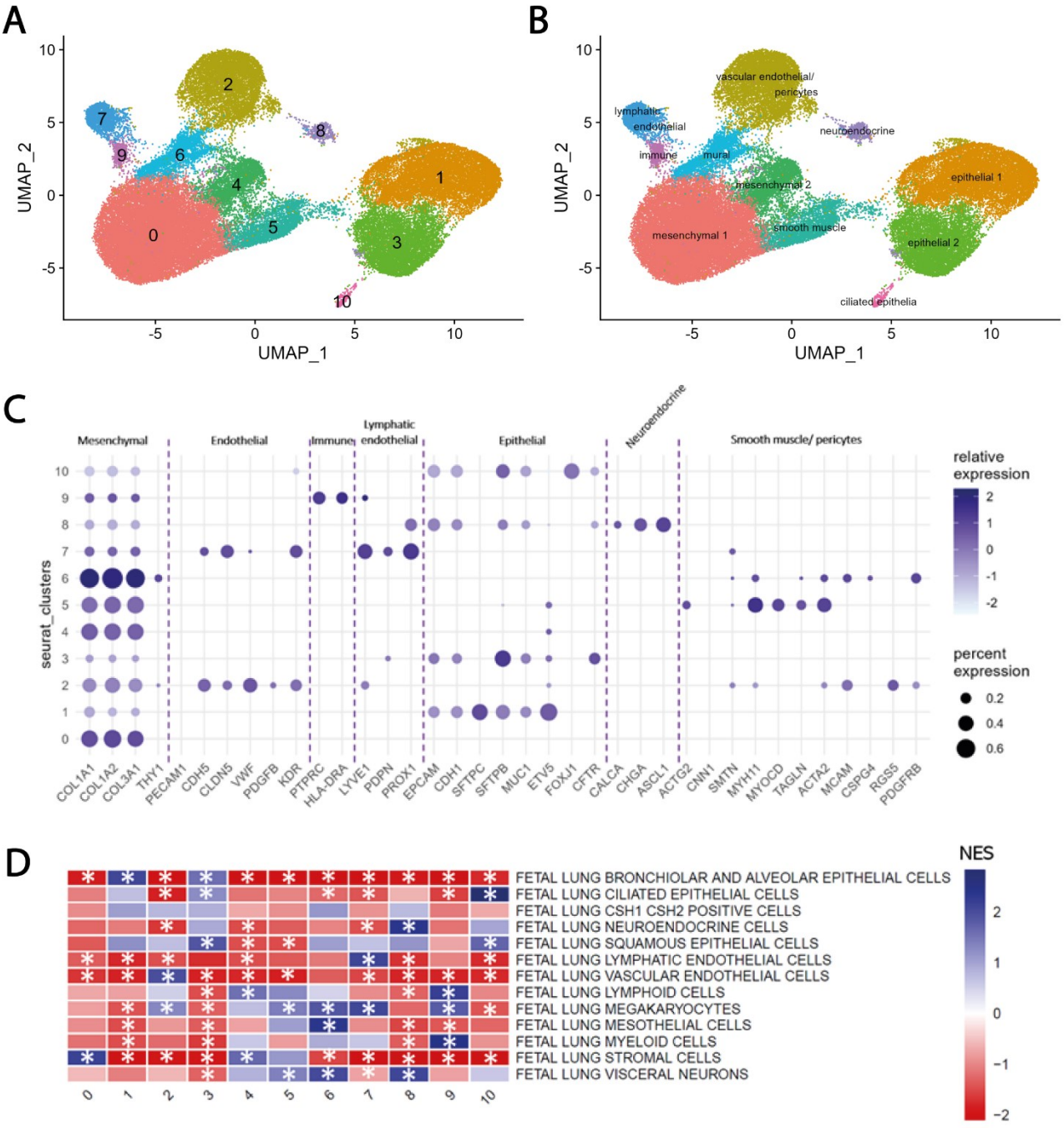


Figure 10 UMAP visualization of A) lung dataset clustering and B) annotated clusters. C) Cell type gene set enrichment analysis of the 11 cell subsets. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. The asterisk indicates significant p-value < 0.05. D) Dot plot showing expression of cell type-specific genes in each cluster.

Cluster	***Annotation***	***Manual markers***	***Top GSEA positively enriched phenotypes***
0	Mesenchymal 1	COL1A1, COL1A2, COL3A1, THY1 (CD90).	Fetal lung – stromal cells (p-val =7.09E-10)
1	Epithelial 1	EPCAM, CDH1, SFTPC, SFTPB, MUC1, ETV5.	Fetal lung – bronchiolar and alveolar epithelial cells (p-val =1.00E-10)
2	Vascular endothelial /pericytes	Endothelial: CDH5, CLDN5, VWF, PDGFB, KDR. Pericytes: ACTA2, MCAM, CSPG4, RGS5, PDGFRB	Fetal lung – vascular endothelial cels (p-val =1.79E-09)
3	Epithelial 2	EPCAM, CDH1, SFTPC, SFTPB, MUC1, ETV5.	Fetal lung – squamous epithelial cells (p-val =6.15E-09)
4	Mesenchymal 2	COL1A1, COL1A2, COL3A1, THY1 (CD90).	Fetal lung – stromal cells (p-val =3.04E-03)
5	Smooth muscle	CNN1, MYH11, MYOCD, TAGLN, ACTA2.	- Fetal lung megakaryocytes (p-val =5.47E-03) - Fetal lung – visceral neurons (p-val =1.25E-02)
6	Mural cells (smooth muscle and pericytes)	Smooth muscle: CNN1, MYH11, MYOCD, TAGLN, ACTA2. Pericytes: ACTA2, MCAM, CSPG4, RGS5, PDGFRB	- Fetal lung – mesothelial cells (p-val =1.00E-10) - Fetal lung – visceral neurons (p-val =5.83E-09)
7	Lymphatic endothelial	LYVE1, PDPN, PROX1.	- Fetal lung – lymphatic endothelial (p-val =1.73E-10) - Fetal lung - megakaryocytes (p-val =1.00E-10)
8	Neuroendocrine	CALCA, CHGA, ASCL1.	- Fetal lung – neuroendocrine cells (p-val =1.00E-10) - Fetal lung – visceral neurons (p-val =1.00E-10)
9	Immune	PTPRC (CD45), HLA-DRA	- Fetal lung – myeloid (p-val =1.00E-10) - Fetal lung – lymphoid (p-val =1.00E-10) - Fetal lung – megakaryocytes (p-val =1.80E-06)
10	Ciliated epithelial	FOXJ1.	- Fetal lung – ciliated epithelial cells (p-val =1.00E-10) - Fetal lung – squamous epithelial cells (p-val =7.20E-06)

Table 4 clusters annotation summary.

3.2. Mesenchymal and endothelial cells sub-clustering

In the skeletal muscle and lung datasets, pericytes and vascular endothelial groups tend to cluster close to each other with partial overlap. Moreover, expression of pericyte-specific markers can be observed in a small subset of the endothelial compartment. Aiming to resolve them from endothelial cells and to enrich my analysis in mesenchymal cell types, I isolated the two sub-groups and I applied a second round of clustering (sub-clustering). This process was repeated separately for each tissue.

The following table shows which clusters were included in the mesenchymal/endothelial sub clustering.

Dataset	***Endothelial cluster(s)***	***Mesenchymal clusters***
Skeletal muscle	5	1, 2, 8, 9
Kidney	7	3, 5, 10, 12, 14
Intestine	13	1, 2, 5, 14
Lung	2, 7	0, 4, 5, 6

3.2.1. Skeletal muscle

Sub-clustering of skeletal muscle endothelium and mesenchyme doesn't allow resolution of pericytes and vascular cells. Indeed, clusters (5 and 7) co-expressing markers of the two cell types can still be observed (Fig.11B). Also, GSEA indicates the presence of unexpected myogenic cells, particularly in clusters 3, 7 and 8 (Fig.11C). Therefore, they were excluded in the subsequent analysis as well as the mixed pericytes/endothelial population.

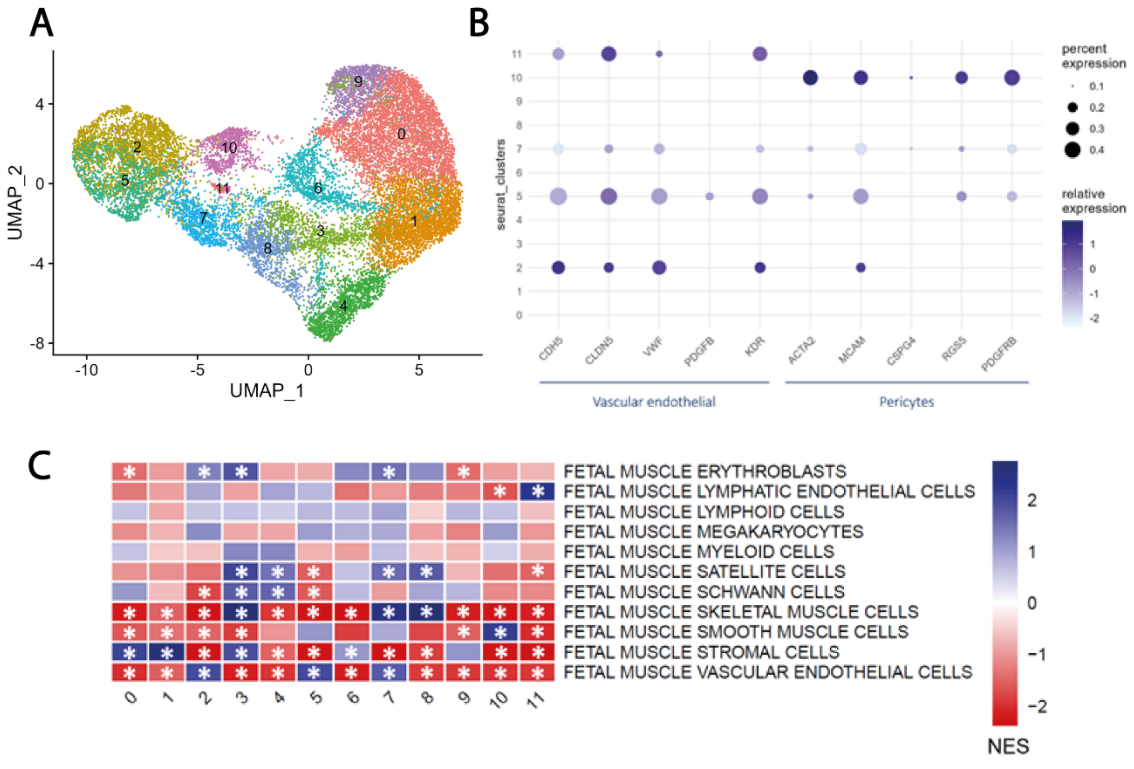


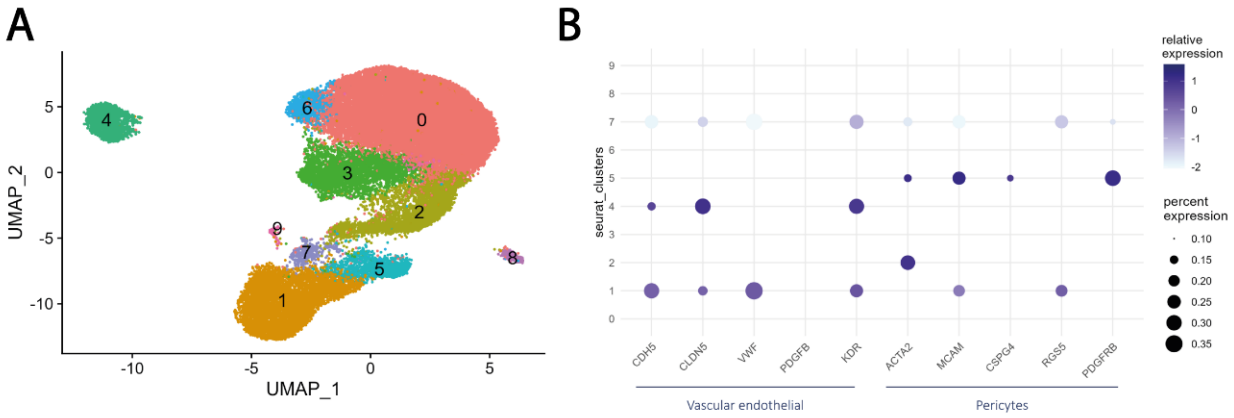
Figure 11 A) UMAP visualization of mesenchymal and endothelial cells subset clustering in skeletal muscle. B) Dot plot showing expression of pericytes and endothelial-specific genes in each cluster

C) Cell type gene set enrichment analysis of the 12 cell subsets. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. The asterisk indicates significant p-value < 0.05. D) Dot plot showing expression of cell type-specific genes in each cluster.

3.2.2. Lung

Similarly, mesenchymal-endothelial sub clustering of the lung dataset doesn't allow separation of pericytes and vascular cells (Fig.12B). Here, clusters of pure vascular endothelial cells and pericytes can be observed, respectively clusters 4 and 5. However, co-expression of the two set of genes is still present in the groups 1 and 7, with higher levels in cluster 1. Furthermore, cluster 1 contain RGS5+/VWF+, RGS5+/KDR+ and RGS5+/CLDN5+ cells (Fig.12C), where RGS5 is largely used marker of pericytes and VWF,KDR and CLDN5 stain endothelial cells.

This sub clustering allowed for the isolation of non-mesenchymal cell types (Fig.12D) including mesothelial cells (cluster 8) and megakaryocytes (cluster 9). Tracing back these cells in the previous step of whole tissue clustering, cluster 9 cells were localised in the endothelial compartment, while the mesothelial group derives from the stromal compartment. The fact that these cell types do not appear in the previous analysis is probably due to the low number of cells they include (304 in cluster 8 and 181 in cluster 9), which couldn't be resolved in the highly heterogeneous dataset. Megakaryocytes, endothelial and mesothelial cells were not included in the subsequent analysis.



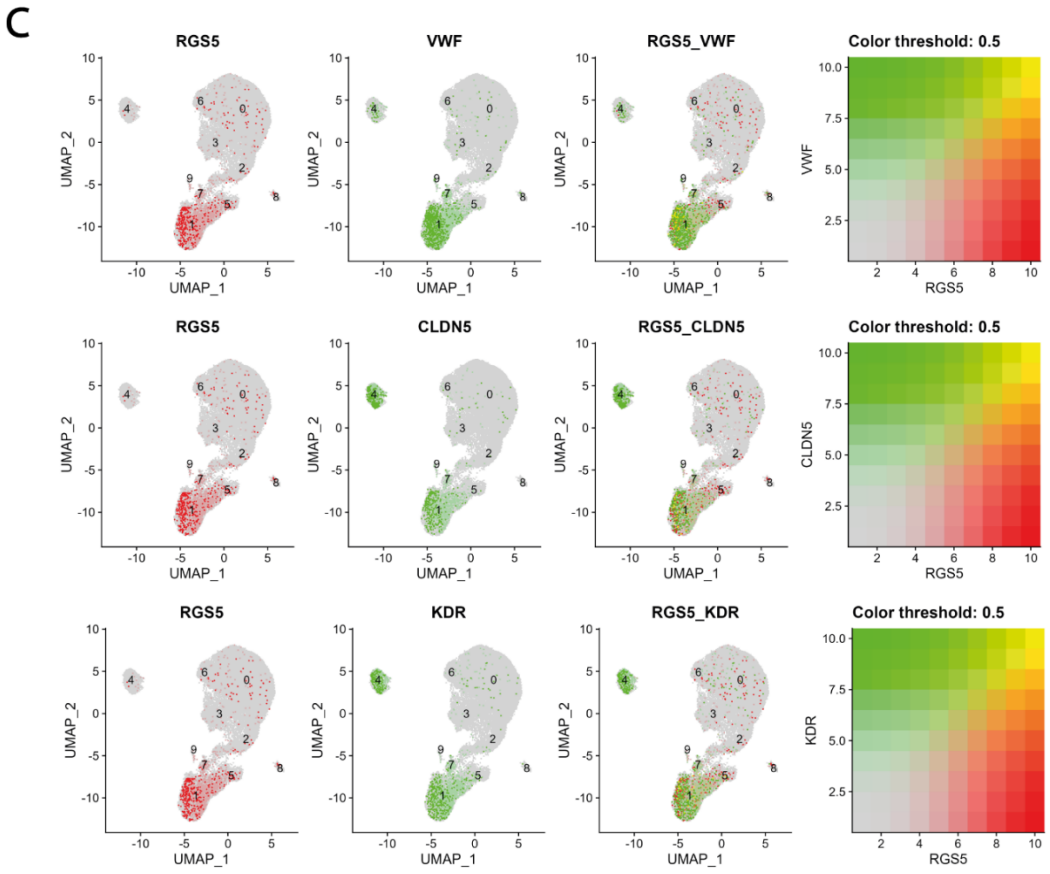


Figure 12 A) UMAP visualization of mesenchymal and endothelial cells subset clustering in lung. B) Dot plot showing expression of pericytes and endothelial-specific genes in each cluster C) UMAP showing the simultaneous expression of RGS5 and endothelial markers. Red indicates RGS5 expression, green indicates VWF, KDR and CLDN5 expression. Cells co-positive for the two markers are visualised in yellow. D) Dot plot showing expression of megakaryocytes and mesothelial cells markers.

3.2.3. Intestine and kidney

I performed mesenchymal-endothelial sub-clustering in the intestine and kidney datasets (Fig.13A-B), where mesenchymal and endothelial clusters separated very well. As expected, there is no overlap between the mesenchymal and vascular clusters. Mesothelial cells were identified in the intestine dataset: these cells were excluded in the subsequent analysis.

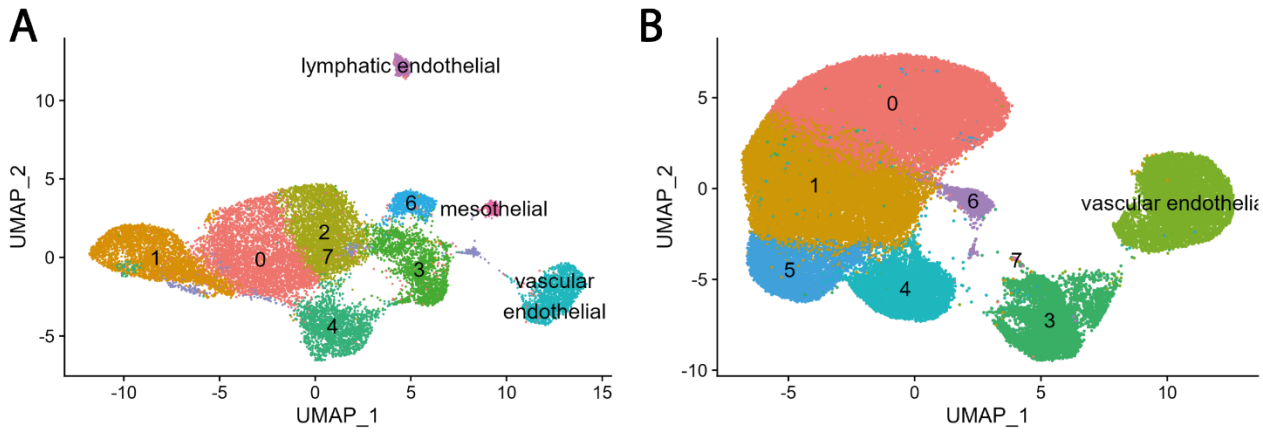


Figure 13 UMAP visualization of mesenchymal and endothelial cells subset clustering A) intestine and B) kidney.

In conclusion, the mesenchymal-endothelial sub clustering did not resolved areas of co-expression of pericyte and endothelial markers observed in lung and muscle. In particular, cells co-expressing pericytes and endothelial markers were identified in these two organs (Fig.12C). Nevertheless, this step allowed the identification and removal of non-mesenchymal contaminants (mesothelial cells in intestine and lung; myoblasts in skeletal muscle). As a consequence, the following analysis of the mesenchymal compartment was performed only on *bona fide* mesenchymal cells. More in detail, the selected mesenchymal clusters used in the subsequent analysis are showed in the following table:

Dataset	***Mesenchymal clusters***
Skeletal muscle	1, 2, 8, 9
Kidney	3, 5, 10, 12, 14
Intestine	1, 2, 5, 14
Lung	0, 4, 5, 6

3.3. Mesenchymal cells integration

3.3.1. Tissues integration

Once non-mesenchymal contaminants (myoblasts and mesothelial cells) were removed and bona fide mesenchymal groups were isolated for each tissue, the lasts were pooled together. Batch effect correction has been applied to remove technical differences among the datasets (Fig.14A-B). Then, clustering workflow identified 7 mesenchymal sub clusters (Fig.14C). Nonetheless, tow clusters are overrepresented by specific organs (Fig.14D):

- Cluster 0 is lung – enriched (82% of lung cells).
- Cluster 4 is kidney – enriched (92% of kidney cells).

Because of the big gap in the number of cells among organs, data were also normalised to the total cells per dataset (Table 5), confirming this observation. Together with tissues contribution, I also investigated distribution of fetal ages, encompassing pcw from 10 to 17. This could be useful to identify age – specific cluster, likely identifying age – specific cell

types. Looking at normalised data (Table 6) it looks like cluster 3 decreases over time, while the other cluster don't have a week-related composition.

Cluster characterization was carried on by analysing the expression of cell-types specific markers from literature and GSEA. Putative markers for each cluster were evaluated by selection of DEGs conserved throughout the four organ datasets.

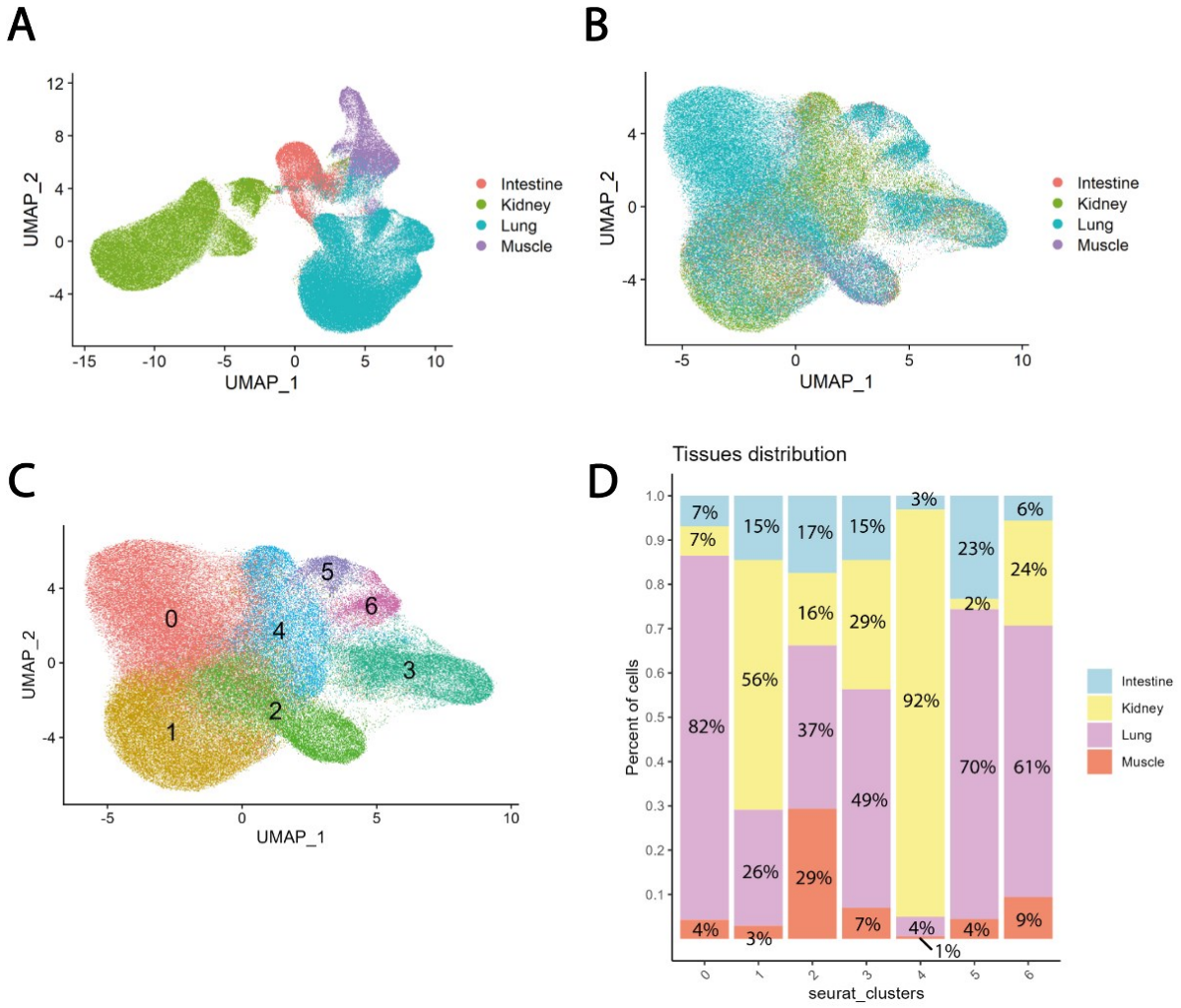


Figure 14 A) UMAP showing clustering without batch effect correction (merging). In this case tissues do not co – cluster. B) UMAP showing clustering without batch effect correction (integration) reveals a more homogeneous organ distribution. C) UMAP visualisation of the 7 mesenchymal sub populations. C) bar plot showing tissues contribution to each cluster. Y axis indicates percentage of tissue-specific cells in each cluster.

A

number of cells

	cluster 0	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	*total cells per dataset*
LUNG	41396	8698	8730	8652	737	2915	3261	74389
KIDNEY	3363	18718	3880	5121	15386	98	1262	47828
INTESTINE	3479	4816	4117	2543	524	969	299	16747
MUSCLE	2149	965	6953	1214	93	186	499	12059
total cells per cluster	50387	33197	23680	17530	16740	4168	5321	**151023** total cells

B

data normalised to the number of cells per dataset

	cluster 0	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6
LUNG	55.65%	11.69%	11.74%	11.63%	0.99%	3.92%	4.38%
KIDNEY	7.03%	39.14%	8.11%	10.71%	32.17%	0.20%	2.64%
INTESTINE	20.77%	28.76%	24.58%	15.18%	3.13%	5.79%	1.79%
MUSCLE	17.82%	8.00%	57.66%	10.07%	0.77%	1.54%	4.14%

Table 5 A) absolute and B) normalised frequencies of tissues-specific cells in each cluster. Data are normalised to the total number of cells per dataset.

A

number of cells

	INTESTINE	KIDNEY	LUNG	MUSCLE	cluster 0	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	*total cells per age*
10 weeks	x	x			2373	2683	1552	2113	1305	861	441	11328
12 weeks	x	x	x	x	14613	5817	5713	5413	1760	759	869	34944
13 weeks	x	x	x	x	4022	1977	1714	1550	503	288	395	10449
14 weeks	x	x		x	211	2537	1197	1069	3004	22	255	8295
15 weeks		x	x		13979	4438	4594	2426	1549	873	1580	29439
16 weeks	x	x	x	x	10975	14418	5467	4023	8345	696	1238	45162
17 weeks	x	x	x	x	4214	1327	3443	936	274	669	543	11406
					50387	33197	23680	17530	16740	4168	5321	**151023** total cells

B

data normalised to the number of cells per age

	INTESTINE	KIDNEY	LUNG	MUSCLE	cluster 0	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6
10 weeks	x	x			20.95%	23.68%	13.70%	18.65%	11.52%	7.60%	3.89%
12 weeks	x	x	x	x	41.82%	16.65%	16.35%	15.49%	5.04%	2.17%	2.49%
13 weeks	x	x	x	x	38.49%	18.92%	16.40%	14.83%	4.81%	2.76%	3.78%
14 weeks	x	x		x	2.54%	30.58%	14.43%	12.89%	36.21%	0.27%	3.07%
15 weeks		x	x		47.48%	15.08%	15.61%	8.24%	5.26%	2.97%	5.37%
16 weeks	x	x	x	x	24.30%	31.93%	12.11%	8.91%	18.48%	1.54%	2.74%
17 weeks	x	x	x	x	36.95%	11.63%	30.19%	8.21%	2.40%	5.87%	4.76%

Table 6 A) absolute and B) normalised frequencies of fetal pcw distribution in each cluster. Data are normalised to the total number of cells per age.

3.3.2. Cluster 5 characterization

Cluster 5 presents a “smooth muscle” phenotype, as shown by the expression of specific markers extensively described in literature (Fig.15A) and the enrichment in pathways like “contraction” and “oxidative phosphorylation” (Fig.15D). The most representative tissues in cluster 5 are lung and intestine, while muscle and kidney are poorly represented (2% and 4% of cells, respectively). This is probably due to the different cellular composition of the tissues. Indeed, two main types of smooth muscle cells exist:

1. Vascular smooth muscle cells have a perivascular localization and, together with pericytes, form the “mural cells”. They are wrapped around blood vessels where their contraction can regulate blood flow.
2. Visceral smooth muscle cells.

While vascular smooth muscle cells can be found in every tissues[16,43], visceral smooth muscles are present in tissues characterised by involuntary contraction: lung contains airways smooth muscle and intestine has 2 smooth muscle layers guiding for peristalsis. On the other hand, in kidney and skeletal muscle the main contribution to the smooth muscle compartment comes from mural cells. Currently, there are no molecular markers distinguishing vascular and visceral smooth muscle cells. Cluster 5 up-regulated differentially expressed genes (DEGs) are well established smooth muscle markers (Fig.10B). Thus, no novel markers could be identified. The expression of HHIP, ACTG2, DES, MYOCD, NRXN3, GREM2 is not conserved across tissues. Indeed, they are more abundantly expressed in the intestine and lung datasets (Fig.15C), indicating that they are either tissue-specific genes or possibly molecular discriminant between vascular and visceral smooth muscle cells.

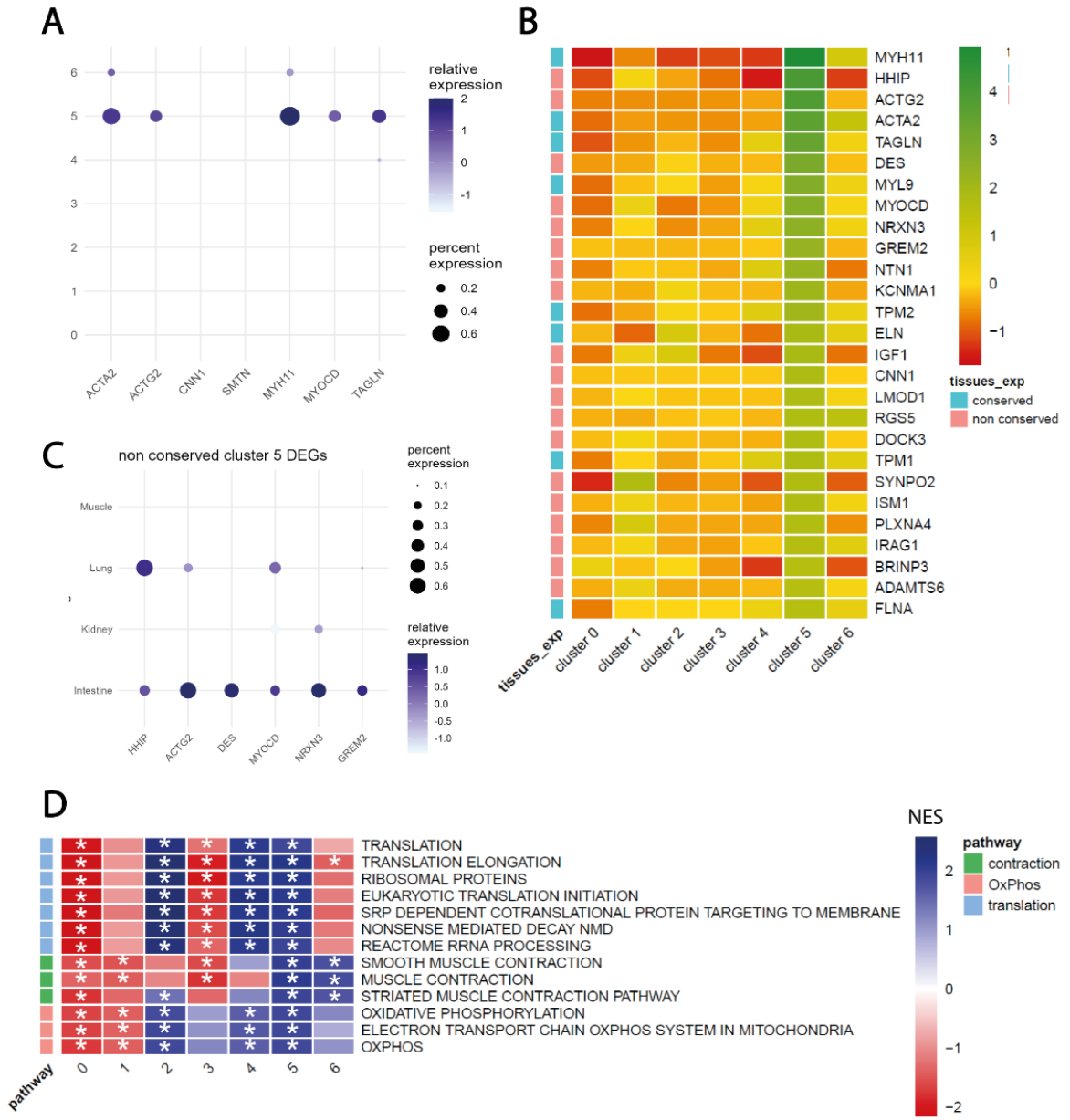


Figure 15 A) Dot plot showing expression of knoen smooth muscle cells markers from literature. B) Heatmap showing top positive DEGs in cluster 5 (LogFC >= 1.5, p-val < 0.05), with green representing positive LogFC and red representing negative LogFC. “conserved” indicates genes calculated as DEGs in all four tissues, “non conserved” indicates genes calculated as DEGs in three or less tissues. D) Dot plot showing tissue expression of non-conserved cluster 5 DEGs. D) Cellular pathways GSEA. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. All the pathways reported for cluster 6 have NES>=1.9. The asterisk indicates p-value < 0.05.

3.3.3. Cluster 6 characterization

Cluster 6 includes pericytes, as shown by the expression of specific markers from literature (Fig.21A) which are also present in the list up-regulated DEGs (Fig.21B). All analysed tissues contribute to the pericytes cluster, suggesting that they are a shared cellular population as expected. The cells in this cluster are also enriched in renin (REN), a kidney-specific marker expressed in the adult and during development by a common precursor of pericytes and vascular smooth muscle cells (Fig.21C)[44]. GSEA (Fig.21E) shows an enrichment in:

1) "Contractile features". Pericytes, indeed, are contractile cells[16].
2) "NOTCH pathway". Notch genes, especially NOTCH3, have been shown to be crucial for angiogenesis, mural cells recruitment and survival[45,46]. Coherently, NOTCH3 is one of the enriched genes in cluster 6 and it is also conserved across tissues.
3) "Inositol phosphate metabolism". The role of inositol phosphate pathway in pericytes in not known. However, alteration of this signalling pathway has been observed in diabetic retinopathy, a disease that correlates with capillaries walls thickening and pericytes loss [16,47].
4) "VEGFR pathway". Pericytes are known to promote angiogenesis by secreting VEGFA, that is sensed by VEGFR2+ endothelial cells promoting their survival[48,49].
5) "Pon channels". Among the enriched genes of cluster 6 there are diverse ion and solute channels like KCNQ5, TRPC6, SLC35F1, CNNM2 and KCNIP4. Ion channels are known to have a role in exchanging solutes between pericytes and the blood flow[16,50]. Similar role has been observd in mesangial cells [51], a specialised form of pericytes in the kidney. In my analysis, these channels are not equally distributed across tissues (Fig.21D), possibly indicating a tissue-specific expression of this class of proteins: they are more strongly expressed in the lung dataset. Moreover, while KCNIP4 and KCNQ5 are respectively kidney- and lung-specific, SLC35F1 is the only one conserved across tissues (Fig.21B). My analysis has come to light SLC35F1 as highly specific of pericytes in the mesenchymal compartment, despite it is a poorly characterised solute channel in literature.

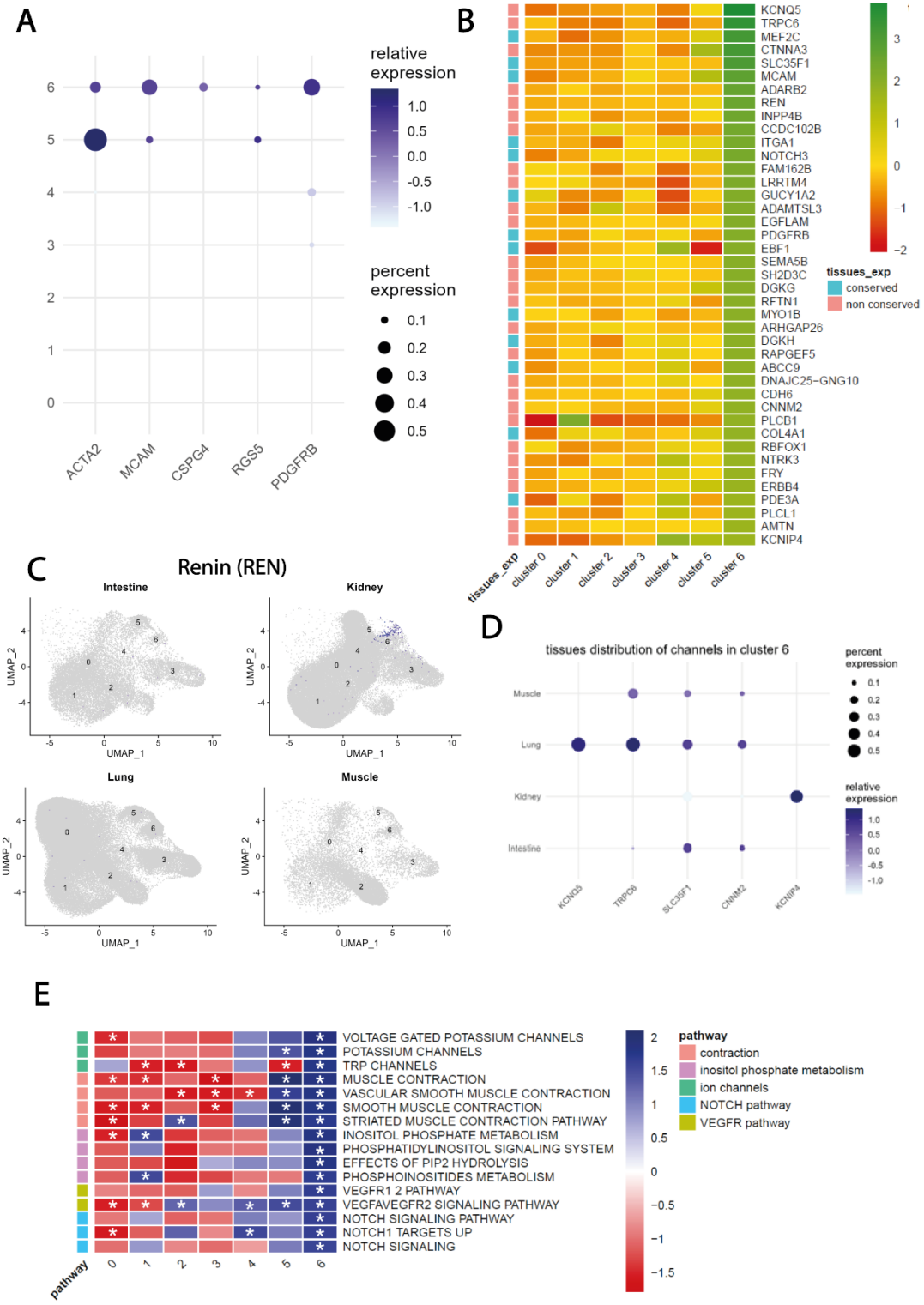


Figure 16 A) Dot plot showing expression of known pericytes markers from literature. B) UMAP showing expression of REN (renin) in the four tissues. C) Cellular pathways GSEA. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. All the pathways reported for cluster 6 have NES>=1.9. Asterisks indicate p-value < 0.05. D) Dot plots showing expression ion channels enriched in pericytes for each tissue. E) Heatmap showing top positive DEGs in cluster 6 (LogFC >= 1.5, p-val < 0.05), with green representing positive LogFC and red

representing negative LogFC. “conserved” indicates genes calculated as DEGs in all four tissues, “non conserved” indicates genes calculated as DEGs in three or less tissues.

3.3.4. Cluster 4 characterization

Cluster 4 is a kidney- enriched cluster, as 92% of cells derive from the kidney dataset. Looking at the UMAP visualization (Fig.22A), cluster 4 is adjacent to clusters 5 and 6, respectively annotated as “smooth muscle cells” and “pericytes”. Also, by decreasing clustering resolution, clusters 4-5-6 form a single group, suggesting some similarities in their transcriptome. Regarding clusters 5 and 6 this is not surprising since pericytes and vascular smooth muscle cells, also called mural cells, are similar cell types: they are perivascular cells regulating vascular stability and function, have a contractile phenotype and share some molecular markers including ACTA2 [16,52–54]. However, cluster 4 is not enriched in differentiated pericytes or vascular smooth muscle markers, with the only exception of the transcription factor EBF1, that is expressed at high levels by 50-60% of cluster 4 cells (Fig.22B-C). EBF1 has not only been observed to be co-expressed with mature pericytes markers in human placenta, brain and lung, but it is also involved in pericytes commitment[55,56]. Indeed, EBF1- silenced brain pericytes have reduced expression of pericytes’ markers and angiogenic factors, suggesting a role of this gene in pericytes differentiation. Despite EBF1, cluster 4 is characterised by the following transcriptomic signature CNTN5, SLC8A1, MEIS2, LRRC4C, MECOM, EMCN, DPP6, DHRS3, KIAA1217, PCDH9, HMGA2, IGF2, NTRK2, LDB2, IL1RAPL1 (Fig.22C). Although these gene have not been characterised in the developing kidney, mRNA expression of SLC8A1, MEIS2, PCDH9, IGF2, LDB2 and NTRK2 has been observed in the cap mesenchyme, the mesenchymal progenitor of nephron cells[57–59]. Thus, cluster 4 could contains precursors of kidney cell-types.

From a functional point of view, cluster 4 is enriched in protein translation pathways (Fig.12D)

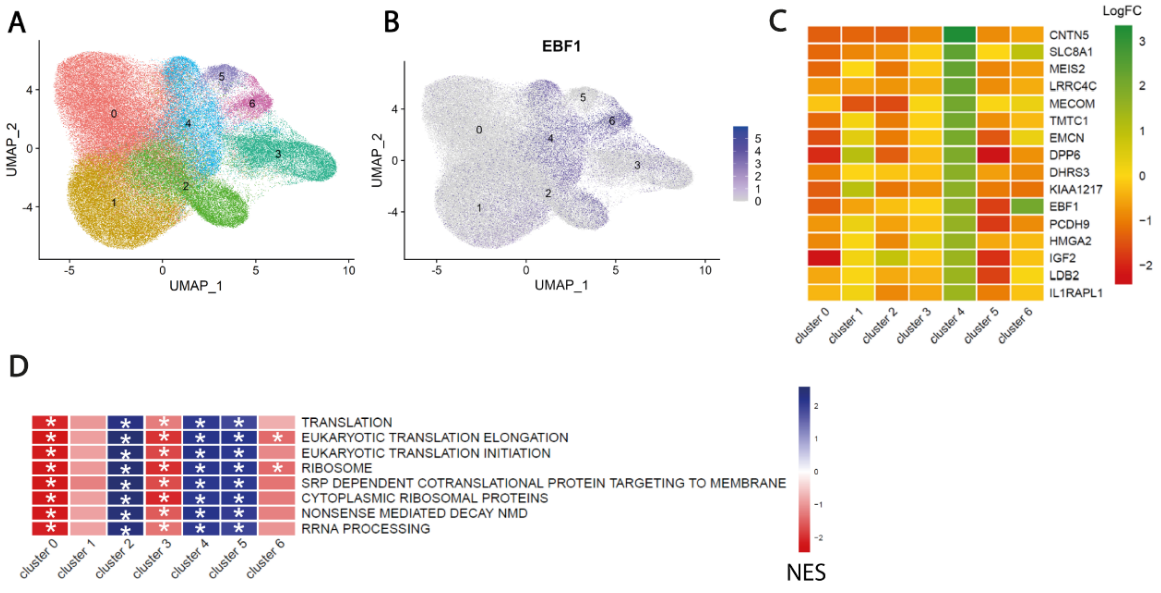


Figure 17 A) UMAP visualization of mesenchymal cells integration and clustering. B) UMAP showing EBF1 expression. C) Heatmap showing top positive DEGs in cluster 4 (LogFC >= 1.5, p-val < 0.05), with green representing positive LogFC and red representing negative LogFC. D) Cellular pathways GSEA. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. All the pathways reported for cluster 6 have NES>=1.9. Asterisks indicate p-value < 0.05.

3.3.5. Cluster 2 characterization

As show in Table 5, around half skeletal muscle cells localised in cluster 2, indicating that this population is abundant in the muscle sample. This group of cells strongly express collagen type I and III and functional characterization shows an enrichment in ECM (extracellular matrix) secretion and remodelling, as well as ribosomal activation and EMT (Epithelial to Mesenchymal Transition) (Fig.18A). Also, a "stem cell upregulated" gene set results significantly enriched in cluster 2. Groups 5 and 6 are characterised by protein translation activity as well. Cluster 2 DEGs are shown in Fig.14B. Some of these genes, like LUM, POST and DCN, are known markers of adult Fibro-Adipogenic Progenitors (FAPs) and of FAPs-like cells observed in developing human and mouse limbs[6,11,60,61]. Several FAPs genes, however, display baseline expression also in the rest of the dataset (Fig.18D-E). Therefore, to assess if cluster 2 is significantly associated with a FAPs phenotype, I performed GSEA exploiting the adult and embryonic/fetal set of genes shown in Fig.13D-E. This analysis demonstrates a positive and significant enrichment (Fig.18C). Moreover, the expression of FAP marker genes is more abundant in the muscle portion of cluster 2 (Fig.18F), in agreement to prevalent muscular origin of FAPs in the adult[6,13].

Hence, cluster 2 could represent fetal FAP-like cells, characterised by processes of ECM deposition and remodelling, active protein translation and epithelial to mesenchymal transition. The facts that all datasets contribute to cluster 2 and many DEGs are conserved across datasets (Fig.18B), suggest that this population is present in all analysed organs. Although fetal/embryonic FAP-like cells have been observed [6,11,60,61], their *in vivo* functions during development is still not clear. In mice, depletion of these cells causes muscle patterning defects, and since FAP-like cells do not express myogenic genes, they have been suggested to regulate myogenesis acting via secreted factors and ECM remodelling, similarly to their adult counterpart[6,11]. Moreover, DLK1 results to be a DEG of cluster 2 (Fig.18B). DLK1 is involved in muscle pattering during development[62] and in the adult DLK1+ stromal cells have been observed to appear upon muscle damage[63], suggesting that this gene may regulate the myogenic program, a known function of FAPs.

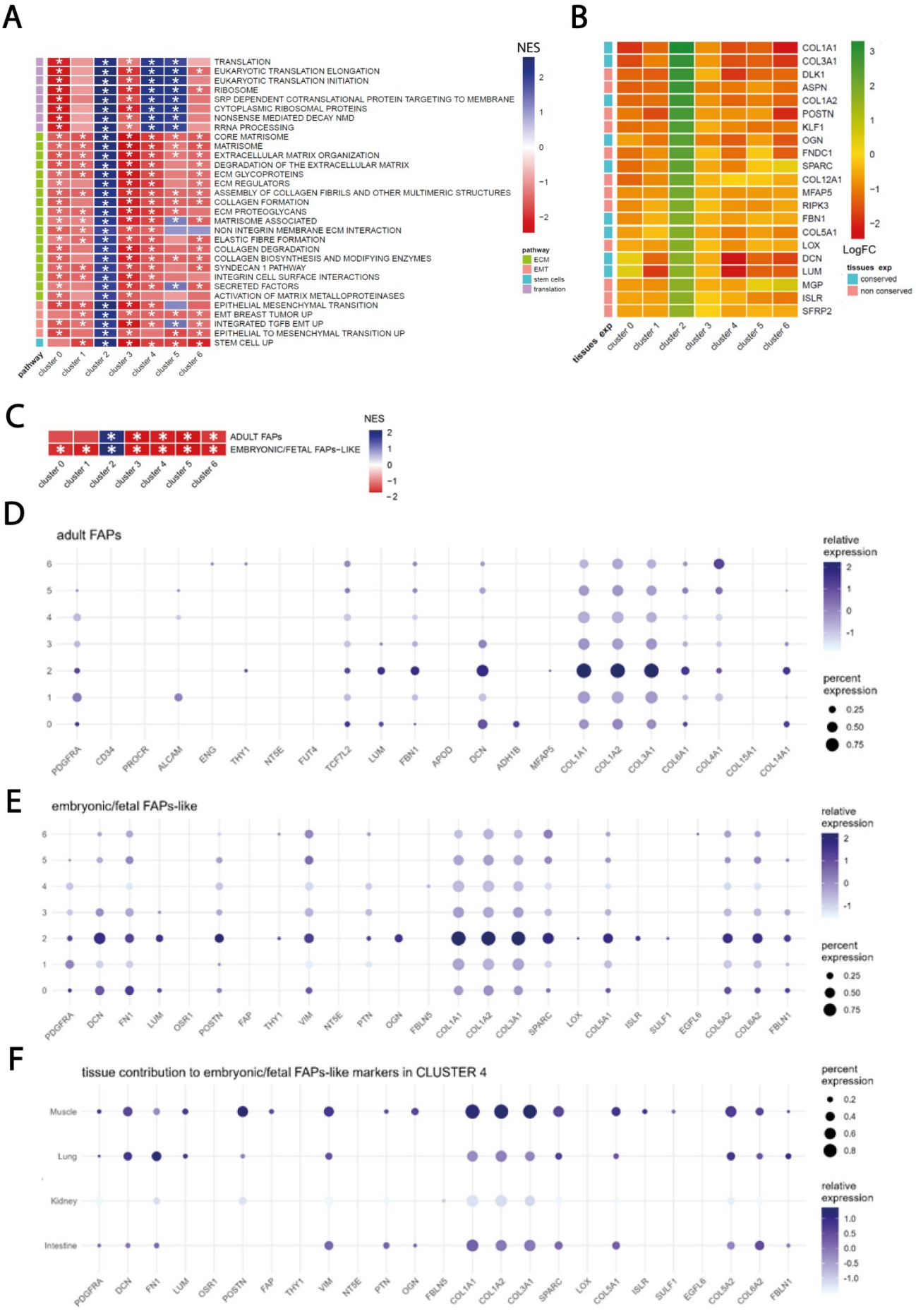

A
TRANSLATION
EUKARYOTIC TRANSLATION ELONGATION
EUKARYOTIC TRANSLATION INITIATION
RIBOSOME
SRP DEPENDENT COTRANSLATIONAL PROTEIN TARGETING TO MEMBRANE
CYTOPLASMIC RIBOSOMAL PROTEINS
NONSENSE MEDIATED DECAY NMD
RRNA PROCESSING
CORE MATRISOME
MATRISOME
EXTRACELLULAR MATRIX ORGANIZATION
DEGRADATION OF THE EXTRACELLULAR MATRIX
ECM GLYCOPROTEINS
ECM REGULATORS
ASSEMBLY OF COLLAGEN FIBRILS AND OTHER MULTIMERIC STRUCTURES
COLLAGEN FORMATION
ECM PROTEOGLYCANS
MATRISOME ASSOCIATED
NON INTEGRIN MEMBRANE ECM INTERACTION
ELASTIC FIBRE FORMATION
COLLAGEN DEGRADATION
COLLAGEN BIOSYNTHESIS AND MODIFYING ENZYMES
SYNDECAN 1 PATHWAY
INTEGRIN CELL SURFACE INTERACTIONS
SECRETED FACTORS
ACTIVATION OF MATRIX METALLOPROTEINASES
EPITHELIAL MESENCHYMAL TRANSITION
EMT BREAST TUMOR UP
INTEGRATED TGFB EMT UP
EPITHELIAL TO MESENCHYMAL TRANSITION UP
STEM CELL UP
pathway
cluster 0
cluster 1
cluster 2
cluster 3
cluster 4
cluster 5
cluster 6
NES
pathway
ECM
EMT
stem cells
translation
B
COL1A1
COL3A1
DLK1
ASPN
COL1A2
POSTN
KLF1
OGN
FNDC1
SPARC
COL12A1
MFAP5
RIPK3
FBN1
COL5A1
LOX
DCN
LUM
MGP
ISLR
SFRP2
tissues_exp
LogFC
tissues_exp
conserved
non conserved
C
ADULT FAPs
EMBRYONIC/FETAL FAPs-LIKE
NES
D
adult FAPs
PDGFRA
CD34
PROCR
ALCAM
ENG
THY1
NT5E
FUT4
TCF7L2
LUM
FBN1
APOD
DCN
ADH1B
MFAP5
COL1A1
COL1A2
COL3A1
COL6A1
COL4A1
COL15A1
COL14A1
relative expression
percent expression
E
embryonic/fetal FAPs-like
PDGFRA
DCN
FN1
LUM
OSR1
POSTN
FAP
THY1
VIM
NT5E
PTN
OGN
FBLN5
COL1A1
COL1A2
COL3A1
SPARC
LOX
COL5A1
ISLR
SULF1
EGFL6
COL5A2
COL6A2
FBLN1
relative expression
percent expression
F
tissue contribution to embryonic/fetal FAPs-like markers in CLUSTER 4
Muscle
Lung
Kidney
Intestine
percent expression
relative expression

Figure 18 A) Cellular pathways GSEA. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. All the pathways reported for cluster 6 have NES>=1.9. Asterisks indicate p-value < 0.05. B) Heatmap showing top positive DEGs in cluster 2 (LogFC >= 1.5, p-val < 0.05), with green representing positive LogFC and red representing negative LogFC. "conserved" indicates genes calculated as DEGs in all four tissues, "non conserved" indicates genes calculated as DEGs in three or less tissues. C) GSEA of the set of genes in Fig.13 D-E. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. All pathways shown for cluster 2 have have NES>=1.9 and p-value < 0.05. D-E) Dot plots showing expression of adult FAPs (D) and fetal/embryonic FAP-like cells (E) in each cluster. F) Dot plots showing expression of fetal/embryonic FAP-like cells for each tissue (specifically in cluster 2).

3.3.6. Cluster 0 characterization

Cluster 0 mostly consists of lung cells, that constitute 82% of the cluster (Fig.19). It has few differentially expressed genes (DEGs): NALF1, SLIT2, MEOX2, LRRC14B, RSPO2, PRKCB, TMEM108, USP17L27, ELMO1, BMP5, ADH1B and FAT3 (Fig.10). However, neither cellular identity nor enriched cellular functions couldn't be appreciated, since GSEA returns low and non-statistically significant enrichment scores. This can be due to the presence of cells with a low sequencing quality. Indeed, the ratio between number total RNA counts and number of cells is the lowest in cluster 0 (Table 7), possibly explaining the badly definable identity. However, cluster 2 shows a comparable ratio and no issues in its characterisation.

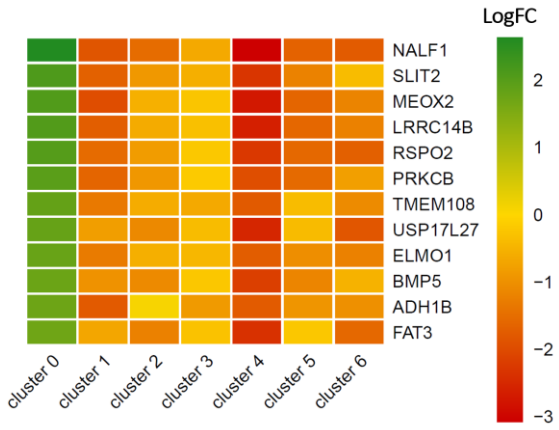


Figure 19 Heatmap showing top positive DEGs in cluster 0 (LogFC >= 1.5, p-val < 0.05), with green representing positive LogFC and red representing negative LogFC.

Cluster	Number of cells	Total RNA counts	counts/cells
0	50.387	2.05E+07	4.06E+02
1	33.197	2.49E+07	7.50E+02
2	23.680	1.10E+07	4.64E+02
3	17.530	1.28E+07	7.27E+02
4	16.740	1.31E+07	7.81E+02
5	4.168	2.24E+06	5.37E+02
6	5.321	3.13E+06	5.89E+02

Table 7 number of cells and RNA counts in each identified cluster.

3.3.7. Cluster 1 characterization

Cluster 1 has contribution from all the datasets, however only 3% of them derive from the skeletal muscle dataset, while half of the cluster consists of kidney cells. Furthermore, top positive DEGs of cluster 1 are shown in Fig.20A and none of them is conserved across dataset, since they are mainly expressed by the kidney subset (Fig.20D). The lung and intestine portions express myofibroblasts and smooth muscle specific genes including ACTA2, MYOCD, ACTG2, HHIP, WNT5A, TAGLN, MYH11[56,64] (Fig.20B). However, nor smooth muscle or myofibroblasts phenotypes are enriched in cluster 1, probably due to the low number of cells expressing these markers (Fig.20C). GSEA doesn't allow further characterisation of this cluster, due to the low and non-statistically significant enrichment scores.

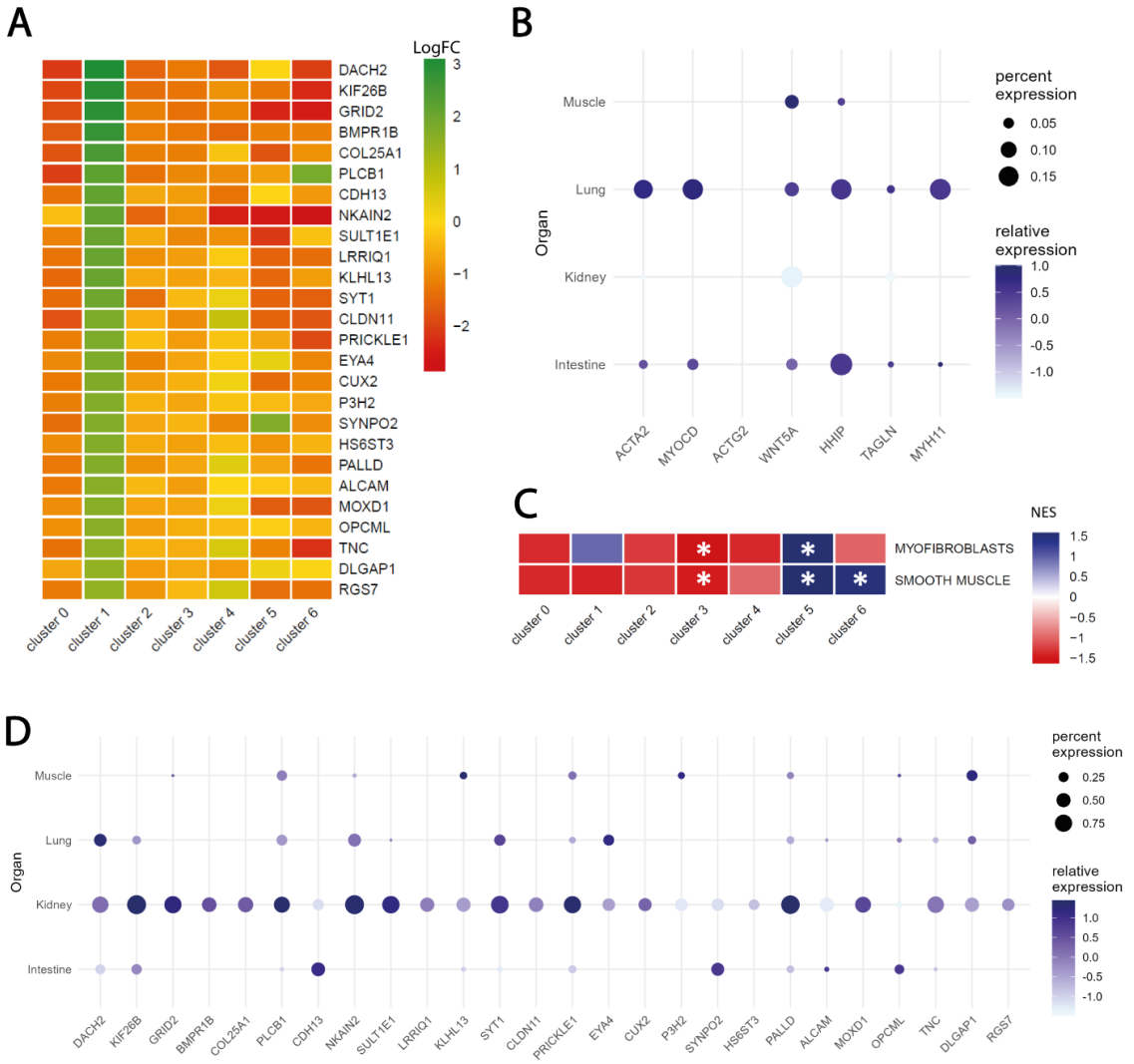


Figure 20 A) Heatmap showing top positive DEGs in cluster 5 (LogFC >= 1.5, p-val < 0.05), with green representing positive LogFC and red representing negative LogFC. B) Dot plot showing the tissue-expression of myofibroblasts and smooth muscle genes in cluster 1. C) GSEA for myofibroblasts (ACTA2, MYH11, TAGLN, TGFBI, HHIP, ENPP2, WNT5A, PDGFRA) and smooth muscle (ACTA2, MYH11, TAGLN, CNN1, ANGPT1) phenotypes. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. Asterisks indicates p-value>0.05.

3.3.8. Cluster 3 characterization

Cluster 3 contains cells from all the four tissues that are significantly enriched in mitosis and DNA repair pathways (Fig.21A), indicating that the tissues contain a portion of proliferating cells. Moreover, cluster 3 is more abundant in earlier pcw, suggesting it could contain proliferating precursors (Fig.21B), but it would be necessary to integrate further fetal ages to assess this hypothesis. Despite being enriched in proliferation features, it is not clear whether the cluster has a homogeneous cellular composition, or it contains different cell types with a common proliferation signature. To address this issue, a sub-clustering of cluster 3 has been performed.

It resulted that cell populations composing Cluster 3 phenocopy those in the non-proliferating compartment (Fig.22A), suggesting cluster 3 includes cell types of other clusters in active proliferation. Cluster 3.1, for examples, expresses mural cells markers (Fig.22B). However, smooth muscle cells markers are less abundant that pericytes' ones, suggesting that the proliferating counterpart of mural cells mainly consists of pericytes. Cluster 3.2 , on the other hand, resembles the FAP-like phenotype cells of the non-proliferating cluster 2 (Fig.22B). Cluster 0 couldn't be annotated; however, it shares markers with both non proliferating group 0 and 1. Also, cluster 3.3 contains contaminating erythroblasts (Fig.22B)

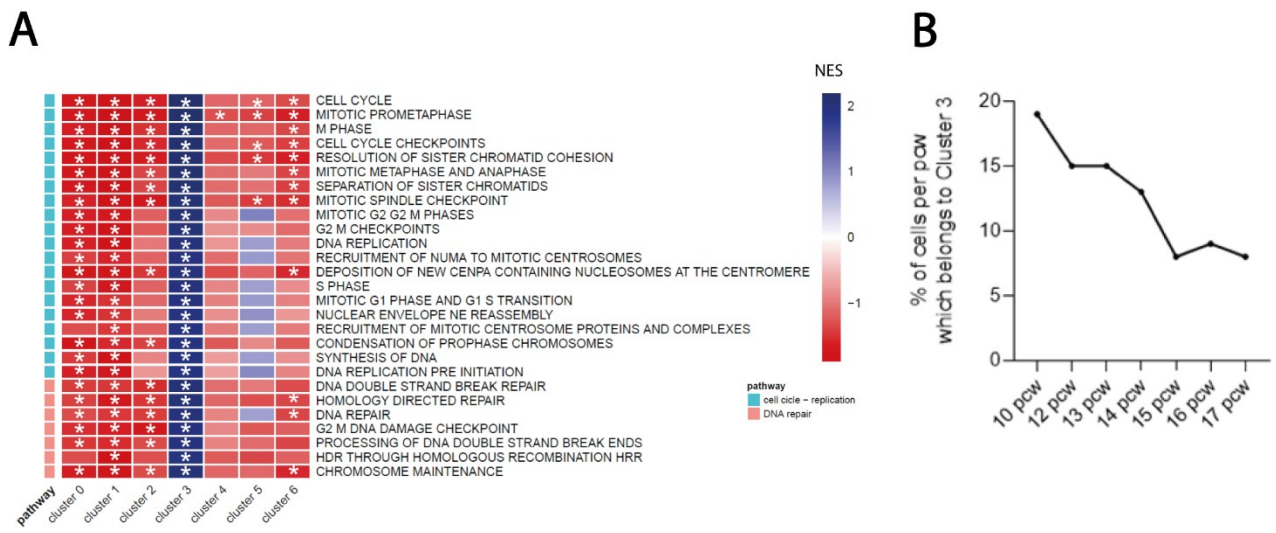


Figure 21 A) Cellular pathways gene set enrichment analysis. The heatmap plot shows the normalized enrichment scores (NES) for each gene sets among with blue representing positive enrichment and red representing negative enrichment. All the pathways reported for cluster 3 have NES>=1.9 and p-value < 0.05. B) Cluster 3 distribution during development.

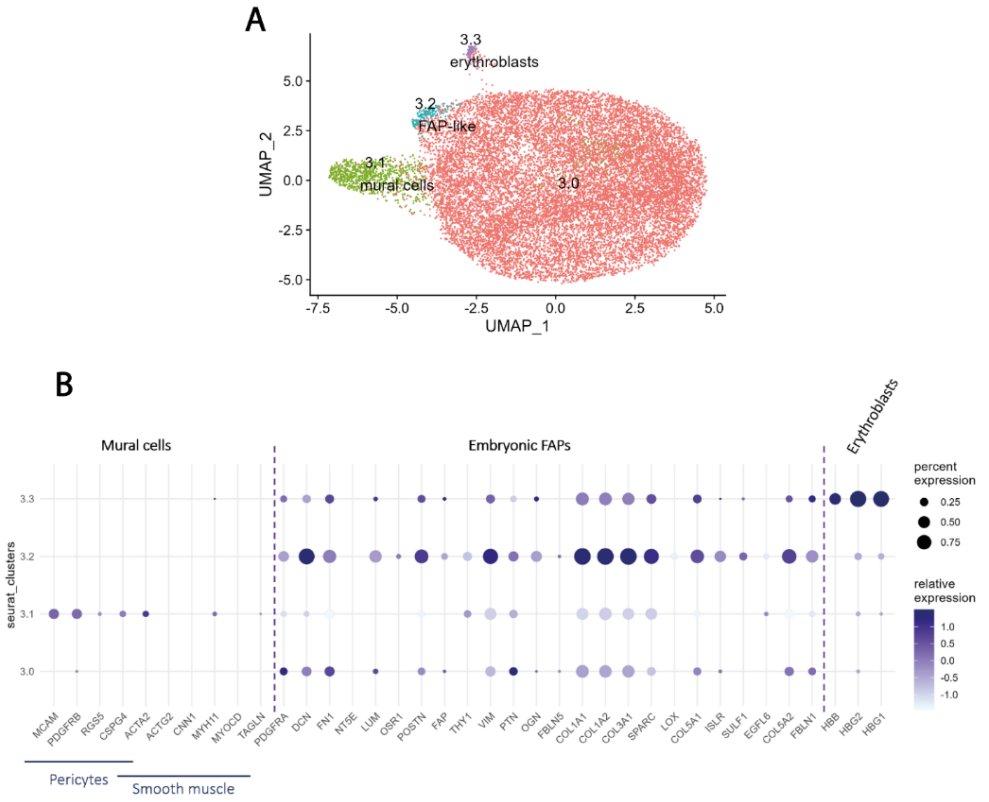


Figure 22 A) UMAP visualization of proliferating cells sub clustering. B) Dot plot showing the expression of pericytes, smooth muscle cells and embryonic FAPs from literature.

In conclusions, clusters were characterised and annotated as follow:

Cluster	***Identity***	***GSEA positively enriched pathways***
0	Lung - enriched	None
1	?	None
2	Fibroblasts	- ECM secretion and remodelling - Epithelial to mesenchymal transition - Translation
3	Proliferating	- Cell cycle progression - DNA repair
4	Kidney – enriched	- Translation
5	Smooth muscle	- Contraction - Oxidative phosphorylation - Translation
6	Pericytes	- Contraction - Ion channels - VEGFR pathway - NOTCH pathway

3.3.9. Selection of putative markers of distinct mesenchymal identities

Aiming to identify putative markers for each mesenchymal sub-population, I selected for each cluster the most up – regulated DEGs. When possible, I chose DEGs conserved across tissue to identify putative universal markers. In the case of cluster 5 no putative markers were evaluated since the most specific DEGs for this population are already known as smooth muscle markers. Selected genes are shown in the following table, asterisks indicate those conserved across tissues.

Cluster	**Putative markers**
0	SLIT2*
1	DACH2, KIF26B, GRID2, BMPR1B, COL25A1, DGKH
2	SFRP2, FNDC1, KLF1, RIPK3
4	CNTN5, SLC8A1, MEIS2, LRRC4C, MECOM
6	GUCY1A2*, ITGA1*, PDE3A*, SLC35F1*, MYO1B*, DGKH*

The expression of these putative markers has been evaluated at the whole tissue level, to assess whether they are expressed by other cell types (epithelial, immune etc.). An overview of the whole tissue validation is in Fig.23, and indicates that:

- Cluster 6. As expected, the new set of markers better identifies pericytes in those tissues where this population is well resolved (lung and skeletal muscle). On the other hand, markers expression is heterogeneous in the other tissues (kidney and intestine). Moreover, some of the molecular
- Cluster 0. The only conserved marker of cluster 0, SLIT2, results to be mesenchyme-specific only in the intestine and skeletal muscle dataset, while its expression is heterogeneous in lung and kidney. In the last one, SLIT2 is more expressed in nephron progenitors, coherently with its function in kidney epithelium development[21,65].
- Cluster 1. Putative cluster 1 markers are expressed in mesenchymal clusters of kidney and lung, while are poorly expressed in muscle. In intestine, these genes are strongly expressed in neural clusters.
- Cluster 2. Putative cluster 2 markers uniquely identify mesenchymal cells in all four organs.
- Cluster 4. The putative markers of this cluster never specifically identify mesenchymal cells as they are expressed, at variable levels, also in other cell types in all four analysed organs.

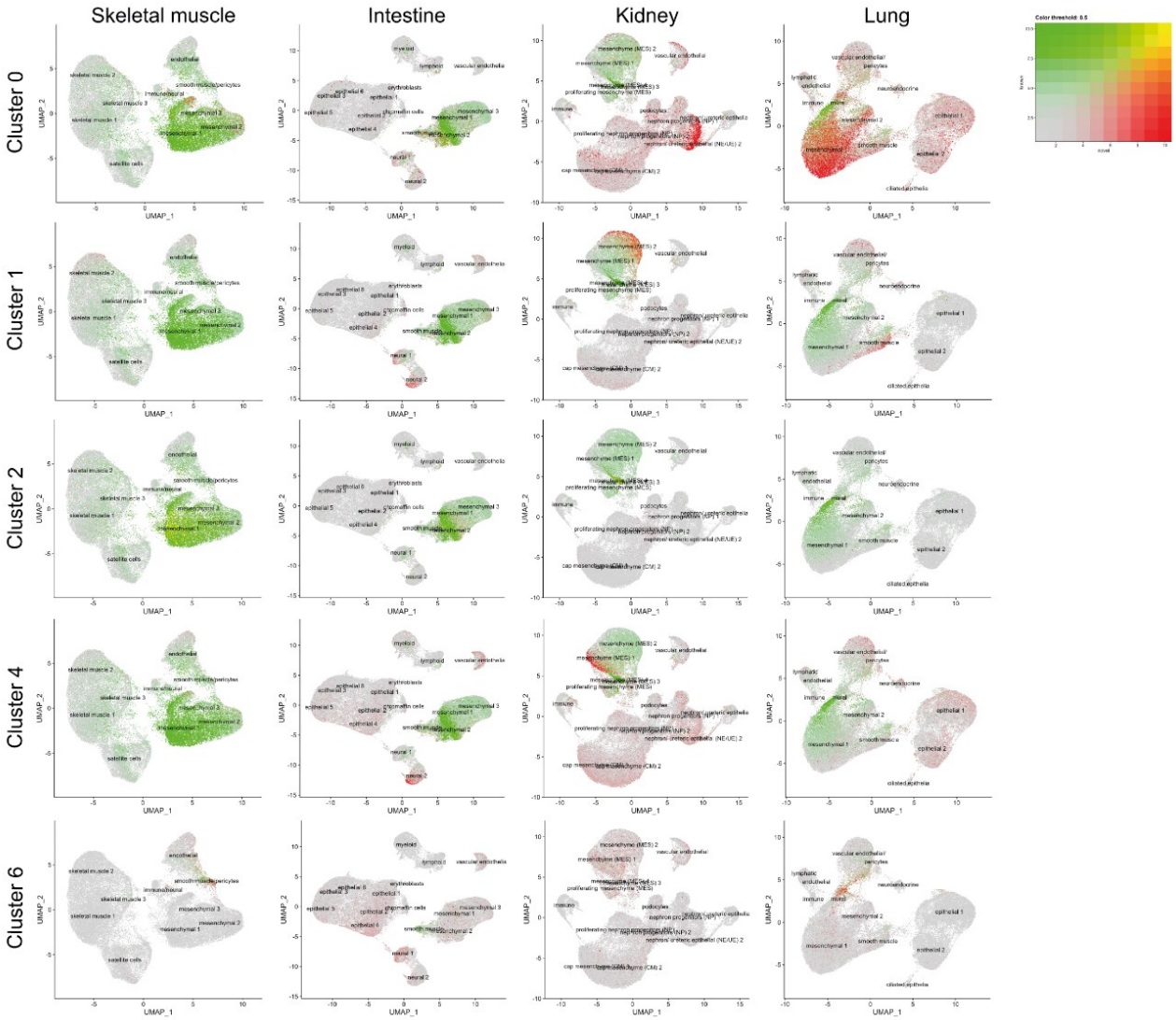


Figure 23 UMAP showing cumulative co-expression of literature markers (in green) and putative markers (in red). Literature genes are COL1A1, COL1A2, COL3A1 and THY1 for clusters 0, 1, 2 and 4, and PDGFRB, ACTA2, CSPG4, RGS5 and MCAM for cluster 6.

Discussion

Analysis of scRNA-seq human tissues is a useful approach to study cell composition and gene expression profile during development. The simultaneous study of different tissues can give insights into tissue-specific and conserved functional identities thus facilitating the investigation of the different cell populations during organogenesis. Moreover, scRNA-seq is a powerful tool to identify new molecular markers that can be used to study cell subpopulations throughout organogenesis.

My thesis aimed at applying a workflow based on scRNA-seq analysis to give insights into the fetal mesenchymal compartment. By comparing different tissues' mesenchymal transcriptome, I paved the way for a comprehensive analysis of the inter- and intra-organ heterogeneity of mesenchymal cells, identifying cell types shared among tissues and others specific of a certain organ.

The first step of clustering (whole tissue) enables the categorization of the main cellular compartments, including a subset of cells expressing collagen genes and other markers like CD90, thus representing bona fide mesenchymal cells. Interestingly, some cells co-expressing pericytes and vascular endothelial markers were identified; this raises some possibilities. An explanation could be the low specificity of adults pericytes markers during development, resulting in their expression in different cell types. On the other hand, some authors suggest that cross-contamination between pericytes and vascular endothelial cells can occur during RNA sequencing due to the physical association of the two cell types in vivo[50,66]. In alternative, a common origin of endothelial cells and pericytes, causing a temporary co-expression of their markers during development, has been suggested. Indeed, a common mesodermal-endothelial precursor called mesenchymoangioblast has been observed in vitro[67,68], which differentiate toward pericytes and endothelial cells; however, its presence in vivo has never been confirmed.

The subsequent integration and sub clustering of mesenchymal cells allows for deeper characterization of this population. My analysis highlighted seven mesenchymal sub-populations. Two of these, clusters 0 and 4 are lung- and kidney-enriched, respectively, suggesting organ – dependent specialization of mesenchymal cells already at this stage of development. On the other hand, pericytes (cluster 6), fibroblasts/FAP-like (cluster 2) and smooth muscle cells (cluster 5) are populations shared among organs. Some clusters (0, 1 and 4) cannot be characterised from a functional point of view. Possible ways to resolve this issue are 1) to perform an additional sub-clustering to assess whether these clusters are heterogeneous and contain further sub-populations. 2) to increase the number of principal components (PCs) in the computational workflow to better capture transcriptomic differences in the mesenchymal dataset.

It seems that merging of mesenchymal cells from diverse tissues better captures some populations. For example, in the intestine dataset, no pericytes could be observed nor in the whole tissue clustering, neither in the mesenchymal/endothelial sub-clustering, whereas combining mesenchymal cells from all tissues led to identification of a pericytes cluster having contribution of all the datasets, also the intestinal one, this highlighting a pericyte

population which was not identifiable in the starting whole-thissue analysis. Also, a proliferating counterpart of mural cells and fibroblasts can be appreciated. For each cluster some putative markers have been selected and an *in* silico first step of calidation at whole-tissue levels was performed. It resulted that markers specificity is highly variable across tissues, so no universal markers for the identified populations could be derived.

In conclusion, this analysis gives insight into mesenchyme composition within and between tissues, revealing the presence of both tissue – specific and shared populations.

To implement this work, some further validations can be performed:

1) Cluster 2 expresses markers of Fibro Adipogenic Progenitors (FAPs), a cell population involved in adult muscle regeneration via secretion of trophic factors and ECM remodelling[6]. Few works suggest the presence of FAP-like cells with similar phenotype in mouse and human embryonic/fetal environment [12,60]. A possible validation of a FAP identity in cluster 2 could be obtained by using RNA sequencing of human skeletal muscle FAPs and perform a co-clustering with the mesenchymal dataset. In this way it should be possible to evaluated if cluster 2 shows transcriptomic similarity with adult FAPs.
2) since this analysis focuses on mRNA levels, it is yet necessary to validate if the expression of the identified putative markers correlates with the presence of the proteins in the tissue. To prove this and to confirm putative new molecular markers, in situ tissues immunostaining would be required.
3) Inclusion of other tissues and stages of development (perinatal and adult) would be useful to further evaluate mesenchymal cell dynamics, functions and turnover throughout the development to adult life.
4) Further information could be gained by performing lineage progression analysis by exploiting computational tools capable of inferring developmental trajectories across clusters.

Bibliography

1. Yang, L. *et al.* The contributions of mesoderm-derived cells in liver development. *Semin. Cell Dev. Biol.* **92**, 63–76 (2019).
2. *An Atlas of Comparative Vertebrate Histology*. *An Atlas of Comparative Vertebrate Histology* (2018). doi:10.1016/c2012-0-06909-8.
3. Hay, E. D. The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. *Dev. Dyn.* **233**, 706–720 (2005).
4. Minot, C. S. & Hertwig, R. Mesenchyme [1]. 3–6.
5. Nishijima David L; Wisner, David H; Holmes, James F, D. K. S. Development of the Mammalian Kidney. *Physiol. Behav.* **176**, 139–148 (2016).
6. Soliman, H. *et al.* Multipotent stromal cells: One name, multiple identities. *Cell Stem Cell* **28**, 1690–1707 (2021).
7. Friedenstein, A. J., Piatetzky-Shapiro, I. I. & Petrakova, K. V. Osteogenesis in transplants of bone marrow cells. *J. Embryol. Exp. Morphol.* **16**, 381–390 (1966).
8. Dominici, M. *et al.* Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **8**, 315–317 (2006).
9. Sheng, G. The developmental basis of mesenchymal stem/stromal cells (MSCs). *BMC Dev. Biol.* **15**, 26–31 (2015).
10. Wang, X. Y. *et al.* Identification of mesenchymal stem cells in aorta-gonad-mesonephros and yolk sac of human embryos. *Blood* **111**, 2436–2443 (2008).
11. Contreras, O., Rossi, F. M. V & Theret, M. Origins , potency , and heterogeneity of skeletal muscle fibro-adipogenic progenitors — time for new definitions. 1–25 (2021).
12. Stumm, J. *et al.* Odd skipped-related 1 (Osr1) identifies muscle-interstitial fibro-adipogenic progenitors (FAPs) activated by acute injury. *Stem Cell Res.* **32**, 8–16 (2018).
13. Joe, A. W. B. *et al.* Muscle injury activates resident fibro/adipogenic progenitors that facilitate myogenesis. *Nat. Cell Biol.* **12**, 153–163 (2010).
14. Krampera, M. & Le Blanc, K. Mesenchymal stromal cells: Putative microenvironmental modulators become cell therapy. *Cell Stem Cell* **28**, 1708–1725 (2021).
15. Joshkon, A. *et al.* Role of CD146 (MCAM) in physiological and pathological angiogenesis—contribution of new antibodies for therapy. *Biomedicines* **8**, 1–19 (2020).
16. Birbrair, A. *Pericyte Biology - Novel Concepts*. *Anticancer research* vol. 1109 (Springer International Publishing, 2018).
17. Darland, D. C. *et al.* Pericyte production of cell-associated VEGF is differentiation-dependent and is associated with endothelial survival. *Dev. Biol.* **264**, 275–288 (2003).
18. Zamora, D. O. *et al.* Enhanced wound vascularization using a dsASCs seeded FPEG scaffold. *Angiogenesis* **16**, 745–757 (2013).
19. Ernst, L. M. & Ruchelli, E. D. *Color Atlas of Human Fetal and Neonatal Histology Second Edition*. (2019).
20. Chi, X. *et al.* Ret-Dependent Cell Rearrangements in the Wolffian Duct Epithelium Initiate Ureteric Bud Morphogenesis. *Dev. Cell* **17**, 199–209 (2009).

21. Faa, G. *et al.* Morphogenesis and molecular mechanisms involved in human kidney development. *J. Cell. Physiol.* **227**, 1257–1268 (2012).

22. Mugford, J. W., Sipilä, P., McMahon, J. A. & McMahon, A. P. Osr1 expression demarcates a multi-potent population of intermediate mesoderm that undergoes progressive restriction to an Osr1-dependent nephron progenitor compartment within the mammalian kidney. *Dev. Biol.* **324**, 88–98 (2008).

23. Das, A. *et al.* Stromal-epithelial crosstalk regulates kidney progenitor cell differentiation. *Nat. Cell Biol.* **15**, 1035–1044 (2013).

24. Reidy, K. J. & Rosenblum, N. D. Cell and Molecular Biology of Kidney Development. *Semin. Nephrol.* **29**, 321–337 (2009).

25. Cookson, M. D. & Stirk, P. M. R. Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. **367**, 380–383 (2019).

26. Moore, M. W. *et al.* Renal and neuronal abnormalities in mice lacking GDNF. **382**, 789–792 (1996).

27. Stark, K., Vainio, S., Vassileva, G. & McMahon, A. P. Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by Wnt-4. *Nature* **372**, 679–683 (1994).

28. Morrisey, E. E. & Hogan, B. L. M. Preparing for the First Breath: Genetic and Cellular Mechanisms in Lung Development. *Dev. Cell* **18**, 8–23 (2010).

29. Joshi, S. & Kotecha, S. Lung growth and development. *Early Hum. Dev.* **83**, 789–794 (2007).

30. Boucherat, O. *et al.* Lung development requires an active ERK/MAPK pathway in the lung mesenchyme. *Dev. Dyn.* **246**, 72–82 (2017).

31. McCulley, D., Wienhold, M. & Sun, X. The pulmonary mesenchyme directs lung development. *Curr. Opin. Genet. Dev.* **32**, 98–105 (2015).

32. Karlsson, L., Lindahl, P., Heath, J. K. & Betsholtz, C. Karlsson et al. - 2000 - Abnormal gastrointestinal development in PDGF-A and PDGFR- α deficient mice implicates a novel mesenchymal str. **3466**, 3457–3466 (2000).

33. Greicius, G. & Virshup, D. M. Stromal control of intestinal development and the stem cell niche. *Differentiation* **108**, 8–16 (2019).

34. Maimets, M. *et al.* Mesenchymal-epithelial crosstalk shapes intestinal regionalisation via Wnt and Shh signalling. *Nat. Commun.* **13**, 1–12 (2022).

35. Kardon, G., Harfe, B. D. & Tabin, C. J. A Tcf4-positive mesodermal population provides a prepattern for vertebrate limb muscle patterning. *Dev. Cell* **5**, 937–944 (2003).

36. Cao, J. *et al.* A human cell atlas of fetal gene expression - Supplementary Materials. *Science* **370**, (2020).

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

38. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17**, 1–15 (2016).

39. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1098 (2013).

40. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, 1–9 (2018).

41. Klotz, L. *et al.* Cardiac lymphatics are heterogeneous in origin and respond to injury. *Nature*

522, 62–67 (2015).

42. Lim, H. Y. *et al.* Hyaluronan Receptor LYVE-1-Expressing Macrophages Maintain Arterial Tone through Hyaluronan-Mediated Regulation of Smooth Muscle Cell Collagen. *Immunity* **49**, 326-341.e7 (2018).

43. Muhl, L. *et al.* Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. *Nat. Commun.* **11**, 1–18 (2020).

44. Pentz, E. S. *et al.* Renin Cells Are Precursors for Multiple Short Article Cell Types that Switch to the Renin Phenotype When Homeostasis Is Threatened University of Virginia University of North Carolina at Chapel Hill. **6**, 719–728 (2004).

45. Liu, H. *et al.* investment. **107**, 860–870 (2011).

46. Ando, K. *et al.* Peri-arterial specification of vascular mural cells from naïve mesenchyme requires notch signaling. *Dev.* **146**, (2019).

47. Microvessel, R., Changes, M. & Diabetic, C. O. F. Current Hypotheses for the Biochemical Basis of Diabetic Retinopathy. **15**, 1892–1901 (1992).

48. Zarb, Y., Franzoso, F. D. & Keller, A. *Pericyte Biology in Disease*. *Advances in Experimental Medicine and Biology* vol. 1147 (2019).

49. Hamik, A., Wang, B. & Jain, M. K. Transcriptional regulators of angiogenesis. *Arterioscler. Thromb. Vasc. Biol.* **26**, 1936–1947 (2006).

50. Hariharan, A., Weir, N., Robertson, C., He, L. & Betsholtz, C. The Ion Channel and GPCR Toolkit of Brain Capillary Pericytes. **14**, 1–45 (2020).

51. Pluznick, J. L. & Sansom, S. C. Ion Channels in Mesangial Cells : Function , Malfunction , or Fiction. 102–111 (2022).

52. Ramirez, M. *Pericyte biology in different organs*. (2019).

53. Betsholtz, C. Review Pericytes : Developmental , Physiological , and Pathological Perspectives , Problems , and Promises. (2011) doi:10.1016/j.devcel.2011.07.001.

54. Bischoff, J., Editor, G. & Interactions, E. P. Endothelial / Pericyte Interactions. 512–523 (2005) doi:10.1161/01.RES.0000182903.16652.d7.

55. Pagani, F., Tratta, E., Dell'Era, P., Cominelli, M. & Poliani, P. L. EBF1 is expressed in pericytes and contributes to pericyte cell commitment. *Histochem. Cell Biol.* **156**, 333–347 (2021).

56. Liu, X. *et al.* Categorization of lung mesenchymal cells in development and fibrosis. *iScience* **24**, (2021).

57. Stewart, B. J. *et al.* Spatiotemporal immune zonation of the human kidney. *Science (80-.).* **365**, 1461–1466 (2019).

58. Menon, R. *et al.* Single-cell analysis of progenitor cell dynamics and lineage specification in the human fetal kidney. *Dev.* **145**, (2018).

59. Dekel, B. *et al.* Multiple imprinted and stemness genes provide a link between normal and tumor progenitor cells of the developing human kidney. *Cancer Res.* **66**, 6040–6049 (2006).

60. Xi, H. *et al.* A Human Skeletal Muscle Atlas Identifies the Trajectories of Stem and Progenitor Cells across Development and from Human Pluripotent Stem Cells. *Cell Stem Cell* **27**, 158-176.e10 (2020).

61. Vallecillo-García, P. *et al.* Odd skipped-related 1 identifies a population of embryonic fibro-adipogenic progenitors regulating myogenesis during limb development. *Nat. Commun.* **8**, (2017).

62. Andersen, D. C. *et al.* Dual role of delta-like 1 homolog (DLK1) in skeletal muscle development and adult muscle regeneration. **3753**, 3743–3753 (2013).

63. Waddell, J. N. *et al.* Dlk1 Is Necessary for Proper Skeletal Muscle Development and Regeneration. **5**, (2010).

64. Boström, H. *et al.* PDGF-A signaling is a critical event in lung alveolar myofibroblast development and alveogenesis. *Cell* **85**, 863–873 (1996).

65. Grieshammer, U. *et al.* SLIT2-Mediated ROBO2 signaling restricts kidney induction to a single site. *Dev. Cell* **6**, 709–717 (2004).

66. He, L. *et al.* Pericyte-specific vascular expression of SARS-CoV-2 receptor ACE2 – implications for microvascular inflammation and hypercoagulopathy in COVID-19. *bioRxiv* 2020.05.11.088500 (2020).

67. Vodyanik, M. A. *et al.* A mesoderm-derived precursor for mesenchymal stem and endothelial cells. *Cell Stem Cell* **7**, 718–729 (2010).

68. Slukvin, I. I. & Kumar, A. The mesenchymoangioblast, mesodermal precursor for mesenchymal and endothelial cells. *Cell. Mol. Life Sci.* **75**, 3507–3520 (2018).