

1 Single nucleus and spatial transcriptomic profiling of human healthy hamstring 2 tendon

3 Jolet Y. Mimpfen^{1*#}, Lorenzo Ramos-Mucci^{1#}, Claudia Paul^{1§}, Alina Kurjan^{1§}, Phillipa Hulley,
4 Chinemerem Ikwuanusi, Steve Gwilym, Mathew J. Baldwin, Adam P. Cribbs¹, Sarah J.B. Snelling¹

5 ¹*The Botnar Research Centre, Nuffield Department of Orthopaedics Rheumatology and*
6 *Musculoskeletal Science, University of Oxford, Oxford, OX3 7LD, UK*

7

8 # Jolet Y. Mimpfen and Lorenzo Ramos-Mucci contributed equally to this work

9 § Claudia Paul and Alina Kurjan contributed equally to this work

10

11 * Correspondence:

12 Dr Jolet Mimpfen

13 Botnar Research Centre

14 Old Road

15 Oxford

16 OX3 7LD

17 United Kingdom

18 jolet.mimpfen@ndorms.ox.ac.uk

19 ORCID: 0000-0003-4464-242X

20

21 Keywords: tendon, single nucleus transcriptomics, hamstring, fibroblasts, spatial transcriptomics,
22 skeletal muscle

23 **Abstract**

24

25 The molecular and cellular basis of health in human tendons remains poorly understood. Amongst
26 human tendons, the hamstrings are the least likely to be injured or degenerate, providing a prototypic
27 healthy tendon reference. The aim of this study was to define the transcriptome and location of all
28 cell types in healthy hamstring tendon. We profiled the transcriptomes of 10,533 nuclei from 4
29 healthy donors using single-nucleus RNA sequencing (snRNA-seq) and identified 12 distinct cell types.
30 We confirmed the presence of two fibroblast cell types, endothelial cells, mural cells, and immune
31 cells, and revealed the presence of cell types previously unreported for tendon sites, including
32 different skeletal muscle cell types, satellite cells, adipocytes, and nerve cells, which are undefined
33 nervous system cells. Location of these cell types within tendon was defined using spatial
34 transcriptomics and imaging, and transcriptional networks and cell-cell interactions were identified.
35 We demonstrate that fibroblasts have a high number of potential cell-cell interactions, are present
36 throughout the whole tendon tissue, and play an important role in the production and organisation of
37 extracellular matrix, thus confirming their role as key regulators of hamstring tendon tissue
38 homeostasis. Overall, our findings highlight the highly complex cellular networks underpinning tendon
39 function and underpins the importance of fibroblasts as key regulators of hamstring tendon tissue
40 homeostasis.

41 1 Introduction

42

43 Tendons are fibrous connective tissues that seamlessly connect muscles to bone. The fibrous
44 composition of the tendons allows them to withstand large mechanical loads, and store and deliver
45 substantial forces that facilitate joint movement and stability[1]. Despite their important role in
46 musculoskeletal function, there is still a lack of understanding of the underlying biology of tendon;
47 even the cellular composition and prominent markers remain poorly understood compared to
48 neighbouring musculoskeletal tissues. Consequently, treatment of tendon disease pathology such as
49 tendinopathy is limited to exercise programmes, pain management, and surgical intervention, but
50 largely fails to address disease pathology or allow focussed drug development[2].

51

52 The tendon extracellular matrix (ECM) is mainly made up of type I collagen, type III collagen and
53 proteoglycans, including decorin, cartilage oligomeric matrix protein, fibromodulin, and biglycan,
54 among others[3]. ECM components are produced and maintained by fibroblast-like cells, also called
55 tenocytes. These elongated fibroblasts are located inside collagen fibres and in the surrounding
56 endotenon, the loose connective tissue which surrounds collagen bundles in tendons. The endotenon
57 contains a more heterogeneous mix of vascular, nerve, and mesenchymal cells[4]. Furthermore, the
58 presence of both innate immune cells, such as macrophages, and adaptive immune cell types, such as
59 T cells, have been reported[5]. These studies demonstrate a diverse cellular environment, yet these
60 cell types and their interactions have not been comprehensively characterised. Insight into the
61 cellular composition of healthy tendon is vital to unravel pathways and relationships dysregulated in
62 disease, enabling the identification of new pharmaceutical targets and better design of biomaterials.
63 Currently, progress in the field of tendon biology is hindered by accessibility of healthy comparative
64 tissue and the technical challenges posed by ECM-rich tissues[6, 7].

65

66 Single-cell RNA sequencing (scRNA-seq) technologies are increasingly used to characterise major and
67 rare cell populations within tissues and infer important cellular markers and potential cellular
68 interactions. Initial studies by Kendal *et al.* (2020) and Akbar *et al.* (2021) presented CITE-seq and
69 scRNA-seq results of healthy and diseased tendons. Despite the limited number of cells captured in
70 these studies, ~440 cells and 3,040 cells, respectively [8, 9], identification of fibroblasts/stromal cells,
71 immune cells, and endothelial cells was possible. However, the low cell numbers captured limited a
72 detailed cellular profiling of healthy tendon. Therefore, a more comprehensive and spatially resolved
73 dataset is needed to infer the cellular states, functions, and interactions that underlie hamstring
74 tendon health. The aim of this study was to create a comprehensive single-cell healthy hamstring

- 75 reference dataset using single-nucleus RNA-sequencing (snRNA-seq), and resolve cellular locations
76 using spatial transcriptomics and immunofluorescence imaging.

77 2 Methods

78

79 2.1 Ethics

80 Ethical approval was granted for the Oxford Musculoskeletal Biobank (19/SC/0134) by the local
81 research ethics committee (Oxford Research Ethics Committee B) for all work on human hamstring
82 tendon, and informed written consent was obtained from all patients according to the Declaration of
83 Helsinki.

84

85 2.2 Tissue acquisition and processing

86 Healthy hamstring tissue (semitendinosus; proximal one-third which includes the midbody to the
87 myotendinous junction) was collected from patients undergoing ACL reconstruction for an ACL
88 rupture, in which healthy hamstring is used as a graft tissue. The age, gender, affected side, and body
89 mass index were collected (Table 1). Tendon was handled as previously described [11]. Tissue was
90 collected in cold DMEM-F12 media supplemented with 10% Foetal Bovine Serum and 1%
91 Penicillin/Streptomycin and transferred to the research facility for processing. Within two hours of
92 tissue collection, tendon was washed in PBS, fat and muscle were cut off, and the tendon was cut into
93 1 cm pieces and photographed to retain topographical reference. All the pieces were snap-frozen in
94 cryotubes using liquid nitrogen. Tissue was stored at -80°C until use.

95

96 **Table 1.** Patient characteristics, including sex, age, body-mass index (BMI), affected side, and
97 ethnicity.

	Sex	Age (yrs)	BMI	Side	Ethnicity	Single nucleus RNA-Seq	Spatial transcriptomics	Imaging
Hamstring 0	Male	37	24.38	Right	Unknown	Yes	No	No
Hamstring 1	Male	24	27.78	Right	White – British	Yes	No	No
Hamstring 2	Female	26	Unknown	Right	White – British	Yes	No	No
Hamstring 5	Female	18	29.73	Left	White – British	Yes	Yes	Yes

98

99 2.3 Single-nucleus RNA sequencing

100 2.3.1 Nuclei isolation

101 Nuclei were isolated using our previously published protocol[12]. In short, forceps, scalpels, and petri
102 dishes were pre-cooled on dry ice. Tissue was cut into thin flakes and stored in a pre-cooled 50 mL

103 Falcon tube. The tissue was then stored at -80°C until use. On the day of cell lysis, the tubes
104 containing the pre-cut tissue were thawed and 4mL of cold 1X CST buffer (salts and Tris buffer,
105 including NaCl, Tris-HCl pH 7.5, CaCl_2 , and MgCl_2 , with CHAPS hydrate (Sigma), BSA (Sigma), RNase
106 inhibitors (RNaseIn Plus (Promega) and SUPERase In (Invitrogen)), and protease inhibitor (cOmplete
107 tablet, Roche); full recipe can be found in the published protocol) was added. After 10 minutes of
108 incubation on a rotor at 4°C , the tissue/buffer mixture was poured through a $40\ \mu\text{m}$ strainer and the
109 tube used for tissue lysis was washed twice with 2mL PBS with 1% BSA. The nuclei solution was then
110 transferred to a 15mL Falcon tube, and the previous 50mL tube was washed once with 4 mL PBS with
111 1% BSA. The solution was centrifuged at 500g at 4°C for 5 minutes. After pouring off the supernatant,
112 the tubes were briefly spun down to get all the liquid to the bottom. The nuclei were then
113 resuspended in the remaining supernatant, and remaining volume was determined. Concentration of
114 nuclei was determined by staining the nuclei with DAPI and counting using a Neubauer Improved
115 haemocytometer (NanoEnTek).

116

117 **2.3.2 Library preparation and sequencing**

118 Nuclei suspensions were diluted (PBS with 1% BSA) to 200-1000 nuclei/ μl and loaded on the
119 Chromium Next GEM Chip G (10x Genomics) with a targeted nuclei recovery of 5,000-10,000 nuclei
120 per sample. Samples were then loaded to the Chromium Controller (10x Genomics) and libraries were
121 prepared using the Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (10x Genomics) following the
122 manufacturer's instructions and indexed with the single-index kit T Set A (10x Genomics). Quality
123 control of cDNA and final libraries was analysed using High Sensitivity ScreenTape assays on a 4150
124 TapeStation System (Agilent). Final libraries were pooled and sequenced on a NovaSeq 6000
125 (Illumina) by Genewiz (UK) at a minimum depth of $\sim 20,000$ read pairs per expected nuclei.

126

127 **2.3.3 Single nucleus RNA-seq data analysis**

128 Raw NGS data was processed using an in-house pipeline scflow quantnuclei
129 (<https://github.com/cribbslab/scflow>). Quality control of Fastq files was performed using fastqc. Fastq
130 files were mapped to the human genome hg38 using kallisto bustools kb count (release 99) with kmer
131 size = 31 (https://www.kallistobus.tools/kb_usage/kb_count/). Spliced and unspliced matrices were
132 merged to create the count matrix for downstream analysis.

133 The snRNA-seq analysis was performed using Seurat v4.0 [15]. Filtering thresholds for number of cells
134 (nCount, number of features(nFeature) and mitochondrial ratio (mitoRatio) were set manually for
135 each sample to remove poor-quality cells (Table 2). Doublets were detected and removed with
136 scDblFinder using the default settings[13]. Ambient RNA (score > 0.2) was removed using decontX

137 from the package celda [14] (Supplementary Figure 1). After quality control, samples were merged,
138 normalised (based on 2000 variable features) and scaled. Principal Component Analysis (PCA) was
139 performed and the first two PCs were visualised to assess variation in sample, batch, sex, and side
140 (Supplementary Figure 2). All four samples were integrated by calling Harmony from within the Seurat
141 workflow, using sample as the covariate [16]. Dimensionality reduction was performed on the
142 integrated data using UMAP and t-SNE, neighbours were identified, and clustering was performed (50
143 dimensions, resolution 0.15; Supplementary Figure 3). Clustering was visualised on UMAP to assess
144 variation in sample, batch, sex, or side (Supplementary Figure 4). Differential expression analysis was
145 performed using FindMarkers (min.pct = 0.25) and the top 5 positive DE genes were visualised using
146 pheatmap (<https://www.rdocumentation.org/packages/pheatmap/versions/1.0.12/topics/pheatmap>) .
147 Dotplots, volcano plots and violin plots were generated using ggplot2. Pathway analysis was
148 performed using gprofiler2[17] to identify over-represented pathways from Gene ontology biological
149 processes (GO:BP) and Reactome databases using the top 100 differentially expressed genes for each
150 clusters (log2 fold change > 0.5). R scripts are available at [https://github.com/Botnar-MSK-](https://github.com/Botnar-MSK-Atlas/hamstring_atlas)
151 [Atlas/hamstring_atlas](https://github.com/Botnar-MSK-Atlas/hamstring_atlas).

152

153 **Table 2.** Quality control of snRNA-Seq samples, including the number of nuclei before filtering, the
154 minimum and maximum number of Features (nFeature) and number of Counts (nCount), the
155 maximum mitochondrial ratio (mitoRatio), and the number of nuclei after filtering and doublet and
156 ambient RNA removal.

	Nuclei before filtering (n)	nFeature minimum	nFeature maximum	nCount minimum	nCount maximum	mitoRatio maximum	Final nuclei (n)
Hamstring 0	5,378	800	3,000	1,000	13,000	0.05	4,301
Hamstring 1	5,366	500	3,000	1,000	13,000	0.05	4,653
Hamstring 2	249	500	3,000	1,000	13,000	0.05	120
Hamstring 5	1,985	500	3,000	1,000	10,000	0.025	1,459
Total	12,978						10,533

157

158

159 2.3.4 SCENIC analysis

160 Analysis was carried out on the command line and in a Jupyter Notebook in an anaconda environment
161 with Python v3.7.12 and pyscenic v0.11.2, among other packages. Briefly, harmony-integrated Seurat
162 data was converted into the h5ad format using Sceasy[18]. The raw, unnormalized counts matrix as
163 well as the accompanying anndata observations and variables were made into a new object and saved
164 as a .loom file, which was used as input for the pySCENIC workflow steps[19, 20].

165 Gene regulatory network (GRN) inference was carried out using the GRNBoost2 algorithm by running
166 the “pyscenic grn” command using default settings. A predefined list of human transcription factors
167 used for this network inference step was retrieved from the Aertslab github repository
168 (<https://github.com/aertslab/SCENICprotocol/tree/master/example>, ‘allTFs_hg38.txt’, last updated 3
169 years ago). The resulting list of TF-gene interactions was then used to infer TF-gene co-expression
170 modules, identify enriched motifs and predict regulons by running the “pyscenic ctx” command with
171 default settings. Input genome rankings .feather and motif annotation .tbl files were retrieved from
172 the Aertslab cistarget resources webpage (v9 files; <https://resources.aertslab.org/cistarget/>). In total,
173 317 regulons were identified. Next, the activity of predicted regulons in individual cells was quantified
174 by running the “pyscenic aucell” command using default settings. For each regulon, cellular regulon
175 activity was then binarized as “on” or “off” using a Gaussian mixture model. Additionally, we assessed
176 regulon-cell type cluster specificity using the regulon specificity score (RSS).

177

178 **2.3.5 CellPhoneDB analysis**

179 Single-nuclei analysis of the interactions between different cells were analysed using cellphoneDB
180 v3.1.0 (<https://github.com/ventolab/CellphoneDB>)[21]. A table of the metadata and the expression
181 matrices were exported from the integrated and annotated hamstring Seurat object. For statistical
182 analysis, receptors and ligands expressed in more than 10% of nuclei in a specific cluster were
183 considered. Further analysis of CellPhoneDB predictions, including biological functions and pathways
184 associated with ligand and receptor genes, was completed with InterCellar (v2.0.0)[22].

185

186 **2.4 Spatial transcriptomics (Visium)**

187 Human hamstring tendon from one of the patients analysed for snRNA-seq was also taken for spatial
188 transcriptomics (see Table 1). Snap frozen tissue was embedded in cold OCT mounting medium (VWR)
189 on dry ice. Frozen sections were cut at 10µm thickness through the coronal plane of the tendon and
190 mounted on a 6.5 x 6.5mm capture area of a Visium spatially barcoded slide (Visium Spatial Gene
191 Expression, 10x Genomics). Following the manufacturer’s recommendations, sections were fixed,
192 stained with H&E, imaged, permeabilised and reverse transcribed. cDNA was then released, collected,
193 and prepared for sequencing. Libraries were sequenced on an Illumina Nextseq 500 at a depth of
194 48,217 mean reads per spot. Data and image alignment was processed using SpaceRanger (v.1.2.2,
195 10X Genomics). Feature-barcode matrices and H&E images were then used for downstream analysis.

196

197 Downstream analysis was conducted using Seurat [15], STUtility. Briefly with STUtility, samples
198 filtered barcodes were loaded and filtered (>20 features/spot and <25% mitochondrial reads)

199 remaining with 2694 spots after filtering. Data was normalised using SCTransform[23] and features of
200 interest were visualised using FeatureOverlay.

201

202 Visium data allowed for visualisation of cell types previously identified in single-nuclei transcriptomics
203 using Cell2location[24, 25]. Cell2location estimates the abundance of cell types in each Visium spot by
204 integrating the two datasets. For this analysis, the annotated single-nuclei dataset (Seurat object) was
205 converted to h5ad using Scesay and used as the reference dataset. First, the reference expression
206 signatures for all cell types (n=12) are estimated using a negative binomial regression which
207 accounted for donor effects. Second, the reference signatures were used to estimate absolute
208 abundance in the Visium spots with the following hyperparameters (cell abundance per spot = 3 and
209 detection alpha = 10). Cell abundance was estimated using histology and detection alpha was kept
210 low to allow for more lenient regularisation to account for large technical variability in RNA content.
211 The estimated cell abundance was plotted for each cell type or multiple cell types in the Visium slide.

212

213 **2.5 Immunofluorescence staining and imaging using Cell DIVE**

214 Snap-frozen tendon samples were embedded in OCT, before cutting 7 mm tissue sections. All
215 protocols were performed in accordance with the Cell DIVE Platform (GE Research, Niskayuna, NY,
216 USA). Slides were post-fixed using an ethanol-acetone (1:1) solution at 4°C for 1 minute. Slides were
217 blocked overnight at 4°C in PBS with 3% BSA and 10% donkey serum (Bio-Rad), stained with DAPI
218 (ThermoFisher) and mounted using mounting media (4% propyl gallate, 50% glycerol; Sigma-Aldrich).

219

220 Slides were imaged at 20X to acquire background autofluorescence, which was subtracted from the
221 following staining round. Slides were deconverslipped in PBS; an antibody mixture (unconjugated
222 primary antibodies for the first round of staining and conjugated antibodies for subsequent staining)
223 was prepared and incubated for 1 hour at RT or overnight at 4°C. Slides were washed thrice in PBS for
224 5 minutes with gentle agitation. An antibody mixture with appropriate secondary antibodies was
225 prepared and slides were incubated for 1 hour at RT, washed thrice with PBS, re-coverslipped, and
226 imaged. After imaging, a bleaching round was performed by deconverslipping and incubating the tissue
227 slides thrice for 15 minutes in 0.5M NaHCO₃ (pH 11.2) and 3% H₂O₂ with a 1-minute wash in between,
228 followed by 3 washes in PBS and a 2-minute DAPI recharge. Slides were then re-coverslipped and a
229 bleached image was acquired which was subtracted from the following staining round. Staining and
230 bleaching rounds were repeated until completion of 4 staining rounds (Table 3). Images were
231 analysed using QuPath (v0.3.2) and representative images are shown.

232

233 **Table 3.** Overview of primary and secondary antibodies used for immunofluorescence staining.

Primary antibodies					
Target	Isotype	Conjugation	Supplier	Prod. no.	Concentration
CD163	Mouse IgG1	647	Novus	NB110-40686	5 µg/ml
CD31	Mouse IgG	647	Novus	NBP2-33154	10 µg/ml
DES	Rabbit IgG	555	Abcam	ab203422	10 µg/ml
PDGFRa	Rabbit IgG	555	Abcam	ab248689	10 µg/ml
PRG4	Mouse IgG1k	Unconjugated	Merck	MABT401	10 µg/ml
TNNT1	Rabbit IgG	Unconjugated	Abcam	ab224583	1:100 (dilution)
TRDN	Rabbit IgG	Unconjugated	Atlas Antibodies	HPA038226	3 µg/ml
Target	Isotype	Conjugation	Supplier	Prod. no.	Concentration
Goat anti-mouse	IgG	555	Thermo	A28180	1:500 (2 µg/ml)
Donkey anti-rabbit	IgG	488 / 647	Thermo	A-21206 / A-31573	1:500 (2 µg/ml)

234
235

236 2.6 Trichrome staining

237 Snap-frozen tendon samples were embedded in OCT before cutting 6 µm tissue sections. All steps
238 were performed at room temperature. Slides were air-dried for 15 minutes, fixed in 10% formalin for
239 15 minutes and washed with ultrapure water. Nuclei were stained using Dako Mayer's haematoxylin
240 (Alignment Technologies) for 15 minutes, washed, blued with Dako Bluing Buffer (Alignment
241 Technologies) and washed again. The Trichrome Stain Kit – Masson (RHS-773-LG) was used. Slides
242 were stained with 1:1 Trichrome stain (A) and distilled water for 30 seconds, washed, and then
243 treated with fresh Phosphotungstic/Phosphomolybdic Acid solution for 8 minutes (1 Phosphotungstic
244 Acid: (B):1 Phosphomolybdic Acid (C): 2 distilled water). The slides were washed, then
245 counterstained in 1: 1 Light Green (D) and distilled water for 10 seconds, and washed again. Slides
246 were immersed in fresh 1% acetic acid solution for 8 minutes, washed, dehydrated, and mounted in
247 DPX. Slides were imaged using the Motic EasyScan 1 system.

248

249 2.7 DAB staining

250 Paraffin-embedded hamstring was sectioned at 6 µm and dewaxed sections subjected to antigen
251 retrieval using Vector Laboratories Tris-based (pH9.0) solution at 95°C for 20min. Sections were
252 rested as the buffer cooled for 20min, blocked for endogenous peroxidase with 0.3% H₂O₂ for 10
253 minutes, and permeabilised with 0.5% Triton-X for 10 minutes. Non-specific binding was blocked for
254 1hr at RT using 3% Horse serum and primary antibody applied overnight at 4°C in a humidified
255 chamber. Primary Pax 7 (ab218472) antibody was used at 1:1000 (0.2µg/ml) diluted in 3% BSA/1X PBS

256 and 3% horse serum. Detection was carried out using the VECTASTAIN® Elite® ABC-HRP Kit with
257 ImmPACT® DAB Substrate Kit and Vector Haematoxylin QS as counterstain. Slides were then
258 dehydrated and mounted in DPX. Slides were imaged using the Motic EasyScan 1 system.

259 3 Results

260

261 A single-nucleus transcriptomic atlas of healthy human hamstring tendon

262 To gain a comprehensive view of healthy human hamstring tendon, we utilised snRNA-seq and spatial
263 transcriptomics, in combination with immunohistochemistry and immunofluorescence. Using sn-
264 RNASeq, we profiled the transcriptomes of 10,533 nuclei from 4 healthy donors. Isolated nuclei from
265 each sample were subjected to droplet-based 3' end sequencing (10X Chromium). After batch
266 correction and integration using Harmony (Supplementary Figure 3), we performed unsupervised
267 graph clustering and annotation of our individual cell populations (Figure 1A), each displaying
268 canonical marker genes (Figure 1B). We identified a range of different cell types, including 2 fibroblast
269 subsets (*COL1A1*, *COL1A2*, *COL3A1*, *DCN*), 3 skeletal muscle cell clusters (*TRDN*, *DES*), satellite cells
270 (*PAX7*, *CALCR*, *GREM1*), 2 endothelial clusters (*PECAM1*, *PTPRB*, *VWF*), mural cells (*NOTCH3*, *PDGFRB*,
271 *MYO1B*), adipocytes (*GRAM*, *AQP7*, *ADIPOQ*, *PLIN1*), immune cells (*PTPRC*, *CD247*, *CD69*, *BLNK*,
272 *CD163*), and “nerve cells”, which are undefined nervous system cells, expressing markers including
273 *TENM2*, *NTRK3*, *NRP2*, *COL23A1*, *ROR1*, and *GRID2*. The two fibroblast subsets are subsequently
274 referred to as MKX+ and PDGFRA+ fibroblasts, due to their expression of these markers (see also
275 Figure 3). Fast- and slow-twitch skeletal muscle cells were identified, along with a third cluster which
276 likely represents cells transitioning between these two states (see also Figure 4). All cell populations
277 were represented across all healthy donors, apart from the slow-twitch skeletal muscle cells and
278 adipocytes, which were only present in 3/4 donors (Supplementary Figure 5).

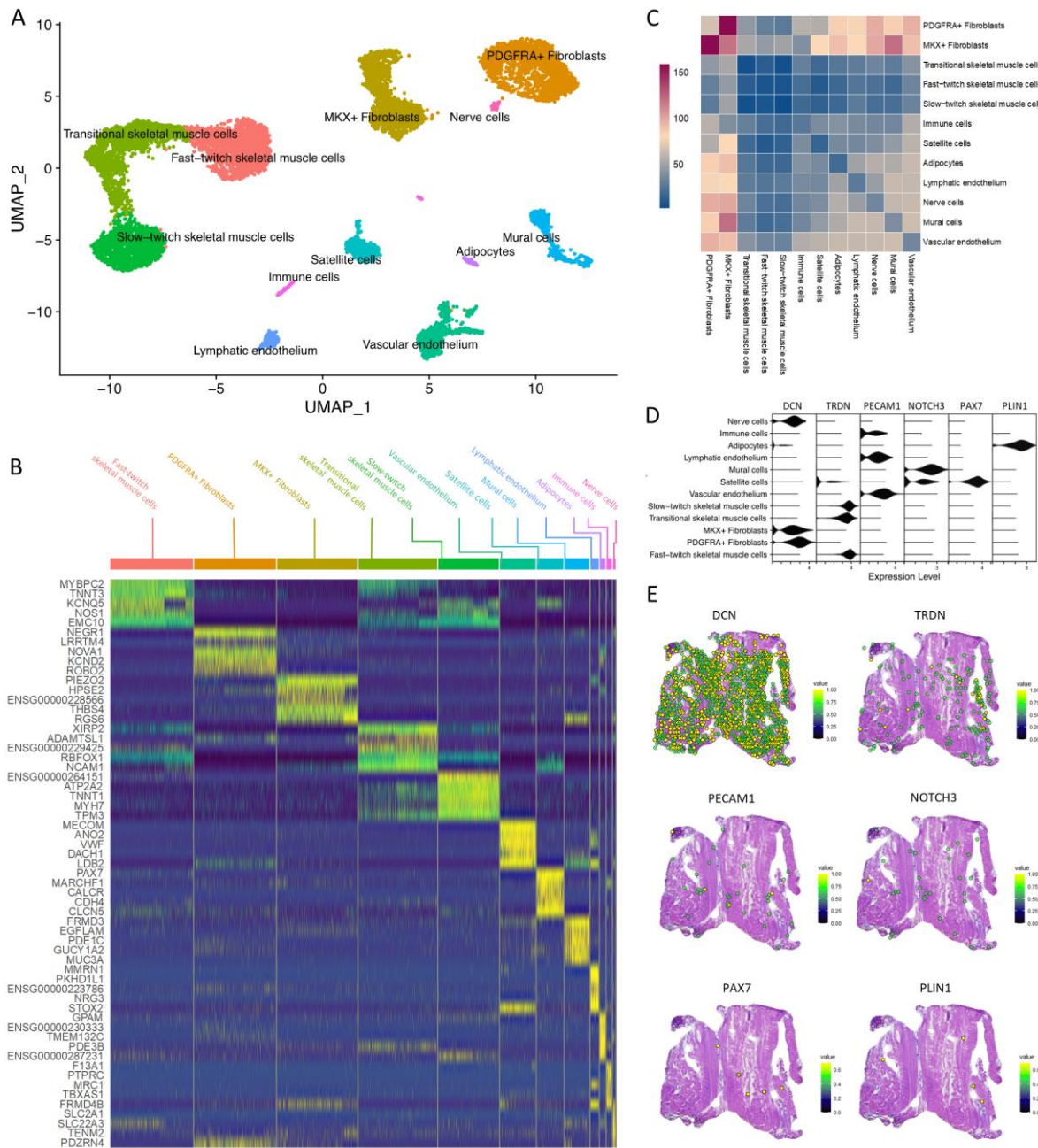
279

280 CellPhoneDB analysis was used to predict the potential ligand-receptor interactions within and
281 between the identified cell types. This analysis revealed that the two fibroblast subsets had the
282 highest number of ligand-receptor interactions (165), showing interactions within and between each
283 fibroblast cluster, and with other clusters such as adipocytes, satellite cells, and endothelial cell
284 subsets, suggesting that fibroblasts are key regulators of hamstring tendon tissue homeostasis (Figure
285 1C). All three skeletal muscle cell types and immune cells have very few predicted interactions with
286 the same or other clusters. All other cell types have some predicted interactions with each other, but
287 have the highest number of potential interactions with the two fibroblast cell types.

288

289 To acquire an understanding of the spatial distribution of cells across the hamstring tendon tissue we
290 performed spatial transcriptomics (10X Visium, n=1). Canonical gene expression markers that were
291 used to identify the major cell types in the snRNA-seq data (*DCN* for fibroblasts, *TRDN* for skeletal
292 muscle cells, *PECAM1* for endothelial cells, *NOTCH3* for mural cells, *PAX7* for satellite cells, and *PLIN1*

293 for adipocytes) (Figure 1D) were also identified in the spatial transcriptomics data (Figure 1E). Overall,
 294 our snRNA-seq datasets identify major cell types residing within healthy human hamstring tendon,
 295 including skeletal muscle cells, adipocytes, and nerve cells, which have not been identified with
 296 transcriptomic methods previously.
 297



298
 299 **Figure 1.** Transcriptomic analysis reveals 12 different cell types in healthy hamstring tendon. (A)
 300 Uniform manifold approximation and projection (UMAP) embedding of 10,533 cells from 4 individuals
 301 reveals 12 clusters. (B) Heatmap of the top 5 positive differentially expressed genes for each cluster. If

302 gene symbol names were not available, ensemble names were used. (C) CellphoneDB analysis of
303 predicted ligand-receptor interactions between different cell types in the snRNA-seq datasets. (D)
304 Expression of *DCN* (fibroblasts), *TRDN* (skeletal muscle cells), *PECAM1* (endothelial cells), *NOTCH3*
305 (mural cells), *PAX7* (satellite cells), and *PLIN1* (adipocytes) in snRNA-seq datasets. (E) Spatial
306 expression of *DCN*, *TRDN*, *PECAM1*, *NOTCH3*, *PAX7*, and *PLIN1* in the healthy hamstring.

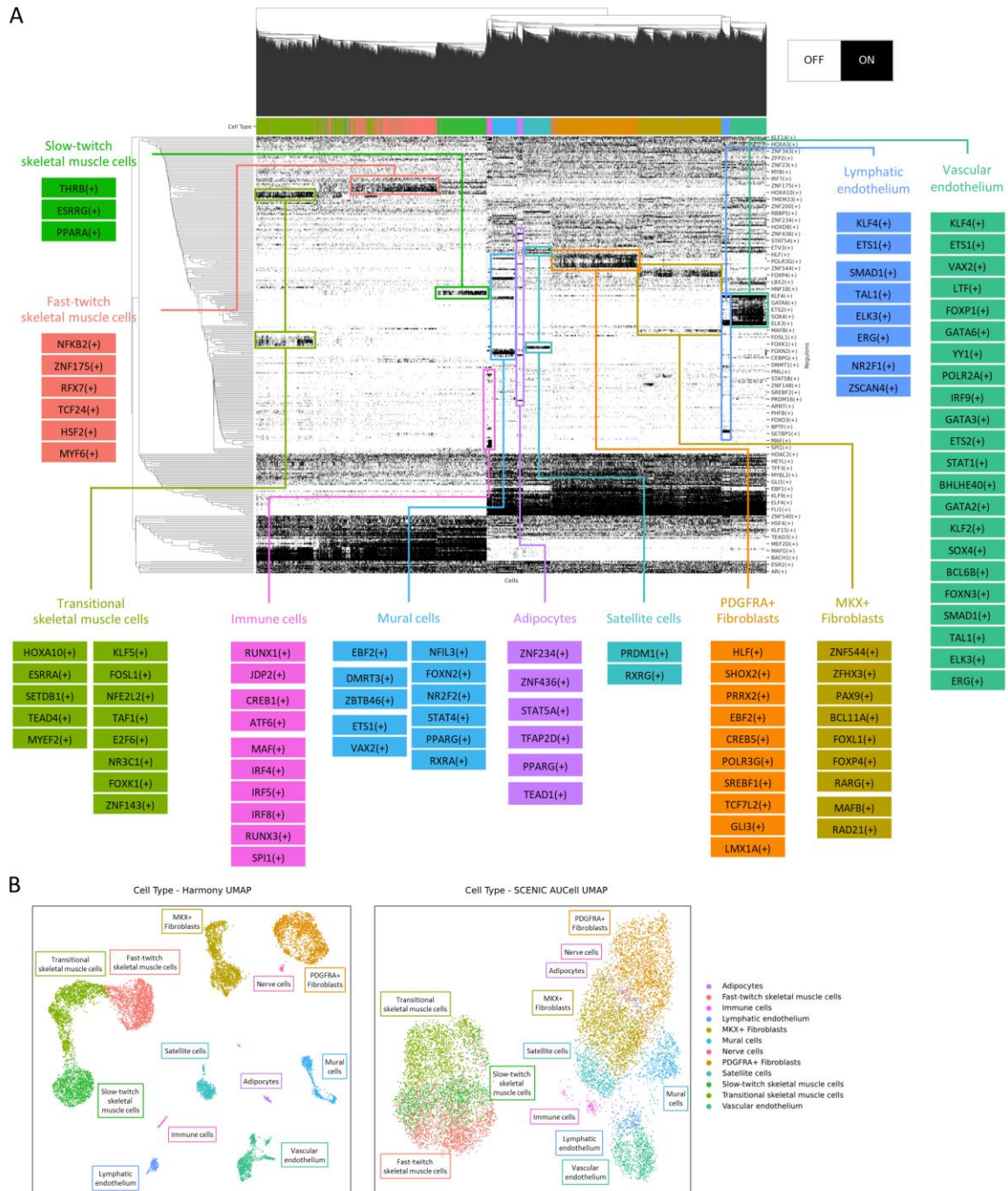
307

308

309 **Mechanistic insights into the regulatory networks driving healthy hamstring tendon cellularity**

310 We next sought to evaluate gene regulatory network reconstruction and cell-state identification using
311 SCENIC. SCENIC evaluates which regulons, or collections of transcription factors and co-factors, are
312 significantly enriched in each cell. We identified 317 regulons with significantly enriched motifs, and
313 the resulting binary regulon activity matrix (Figure 2A) clearly stratifies similar cell types identified by
314 clustering the integrated snRNA-seq data. The three skeletal muscle cell clusters, located on the left
315 side of the matrix, have the most distinct regulon activity when compared to all the other clusters.
316 Interestingly, satellite cells, multipotent cells present in skeletal muscle, did not cluster with the
317 skeletal muscle cells. Both of the fibroblast cell types had distinct sets of activated regulons. Cell
318 clustering based on regulon activity reflected the findings from the regulon activity matrix (Figure 2B).
319 All cell subsets, apart from the small nerve cells cluster, could easily be detected as a separate cluster.
320 The three skeletal muscle cell types, especially the slow-twitch and fast-twitch skeletal muscle cell
321 subsets, showed a lot of overlapping regulon activity but were different from non-skeletal muscle
322 cells. Satellite cells clustered close to the fibroblast subsets rather than close to the skeletal muscle
323 cell subsets. Although the two fibroblasts cell clusters were present near each other, they were clearly
324 distinct. Finally, the regulon specificity score (RSS) was calculated to assess regulon-cell type cluster
325 specificity, and the top 5 regulons per cell type were highlighted (Supplementary Figure 6).

326



327

328

329

330

331

332

333

334

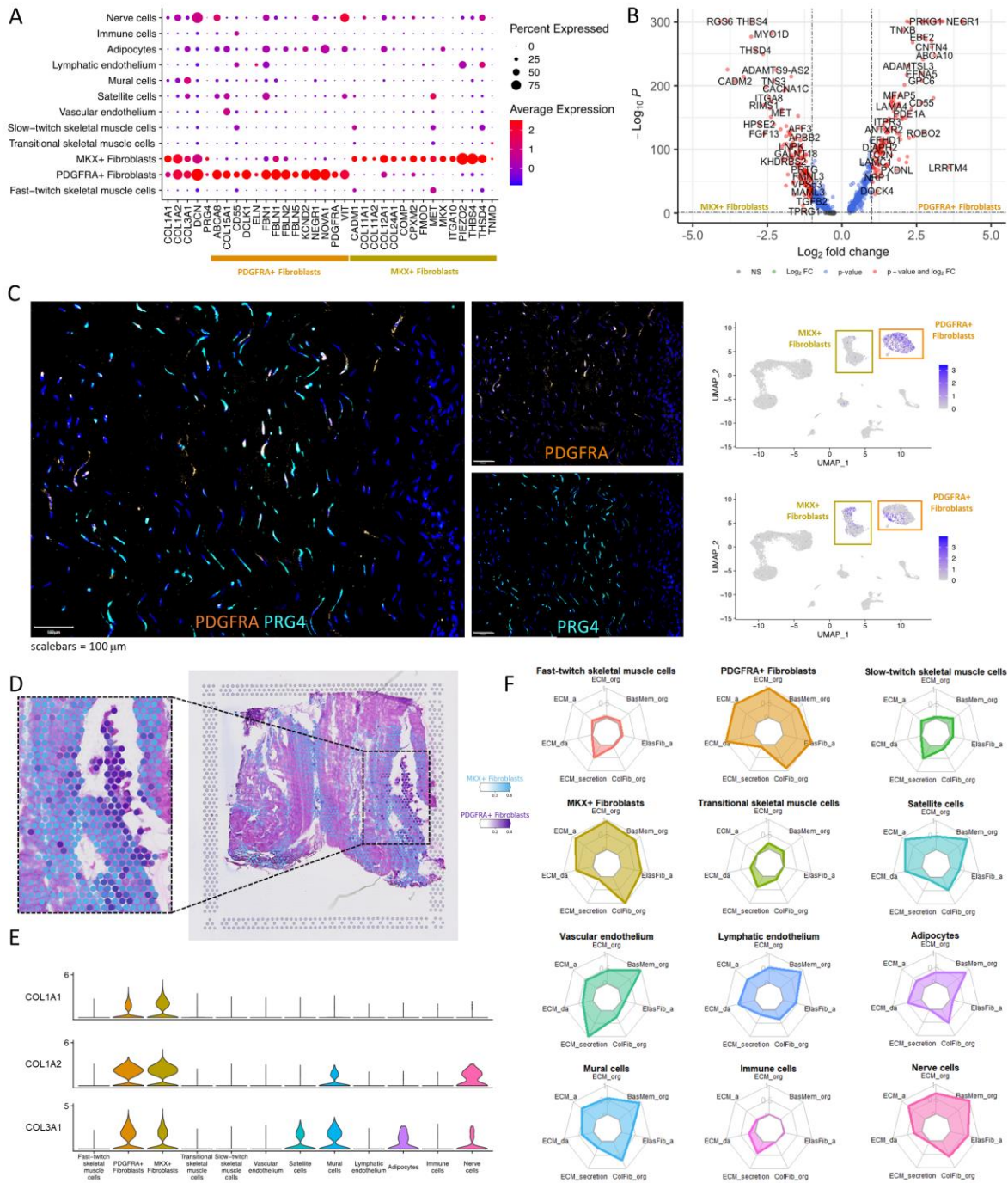
335

Figure 2. SCENIC analysis provides mechanistic insights into the regulatory networks driving healthy hamstring tendon function (A) Binary regulon activity matrix of the 317 regulons that were found to have significantly enriched motifs for the corresponding transcription factors. (B) Comparison of the UMAP using Harmony integration (left) or the SCENIC regulon activity matrix (right).

336 **Two distinct fibroblast cell types are detected in healthy hamstring tendon**

337 In order to better understand the similarities and differences between the MKX+ and PDGFRA+
338 fibroblasts, we compared their gene expression profiles (Figure 3A). Markers previously described as
339 specific for tendon fibroblasts, including mohawk (*MKX*), tenomodulin (*TNMD*), thrombospondin-4
340 (*THBS4*), and cartilage oligomeric matrix protein (*COMP*), were all exclusively present in MKX+
341 fibroblasts. Other highly expressed genes included piezo-type mechanosensitive ion channel
342 component 2 (*PIEZO2*), which is important for proprioception, and different collagens and ECM
343 proteins (*COL11A1*, *COL11A2*, *COL12A1*, *COL14A1*, *FMOD*). In contrast, PDGFRA+ fibroblasts
344 specifically express high levels of neuronal growth regulator 1 (*NEGR1*), RNA-binding protein Nova-1
345 (*NOVA1*), and vitrin, as well as fibrillin-1 (*FBN1*), fibulins (*FBLN1*, *FBLN2*, *FBLN5*), and elastin (*ELN*),
346 which are all key components of elastic fibres. Differences between these two fibroblasts subsets are
347 further illustrated by the differential expression analysis (Figure 3B). Pathway analysis using highly
348 expressed genes for each fibroblast subset revealed that both cell types had strong upregulation of
349 “ECM organisation”, “collagen formation”, “collagen biosynthesis and modifying enzymes”, and “ECM
350 proteoglycans”. However, increased expression of “elastin fibre formation” and “molecules
351 associated with elastic fibres” were only found in PDGFRA+ fibroblasts, while “signalling by receptor
352 tyrosine kinases”, “signalling by MET”, “MET promotes cell motility”, and “MET activates PTK2
353 signalling” were only significant in MKX+ fibroblasts (Supplementary Figure 7). Immunofluorescence
354 staining of healthy human hamstring tendon showed that these tendon fibroblasts, recognisable by
355 their typical elongated shape, are positive for fibroblast marker lubricin (*PRG4*), while only a subset of
356 fibroblasts showed expression of PDGFRA (Figure 3C). We performed cell2location analysis to
357 leverage the cell type annotations identified in the snRNA-seq data and decompose the cell type
358 spatial location of each cluster. We found that both fibroblast cell types, but in this tissue mainly
359 MKX+ fibroblasts, are predicted to be located throughout the whole hamstring, while the PDGFRA+
360 fibroblasts were predicted to be especially abundant in areas close to skeletal muscle (Figure 3D).
361 Although tendon fibroblasts are thought to be mainly responsible for the production and organisation
362 of ECM in tendon, not only fibroblasts, but other cell types such as satellite cells, mural cells,
363 adipocytes, and nerve-like cells were also found to be potential producers of type I and type III
364 collagens (Figure 3E). Pathway analysis confirmed that while both fibroblast subsets had a high
365 enrichment score for ECM-related GO:BP terms, other cells, including satellite cells, vascular and
366 lymphatic endothelium, mural cells, and nerve-like cell, are also involved in these processes (Figure
367 3F).

368



369

370

371

372

373

374

375

376

377

Figure 3. Healthy hamstring tendon contains two different fibroblast types. (A) Expression of fibroblast markers across all different cell types in the snRNA-Seq datasets. (B) Differential expression analysis of PDGFRA+ Fibroblasts and MKX+ Fibroblasts, grey = not significant, green = log₂ fold change (Log₂FC) at least ±1, but p > 0.05, blue = p < 0.05 but Log₂FC between 1 and -1, red = p < 0.05 and Log₂FC at least ±1. (C) Expression of fibroblast marker lubricin (PRG4) and PDGFRA+ fibroblasts marker platelet derived growth factor receptor alpha (PDGFRA) in healthy human tendon tissue compared to their RNA expression in the feature plots. (D) Predicted location of MKX+ fibroblasts (blue) and PDGFRA+ fibroblasts (purple) in the spatial transcriptomics data (Visium) using

378 cell2location analysis. (E) Expression *COL1A1*, *COL1A2*, and *COL3A1* in snRNA-seq datasets. (F) Radar
379 plots of the enrichment score (0-1) of gene ontology biological processes (GO:BP) related to
380 extracellular matrix (ECM). Clockwise starting at the top: ECM organisation, basement membrane
381 organisation, elastic fibre assembly, collagen fibre organisation, ECM secretion, ECM disassembly,
382 ECM assembly.

383

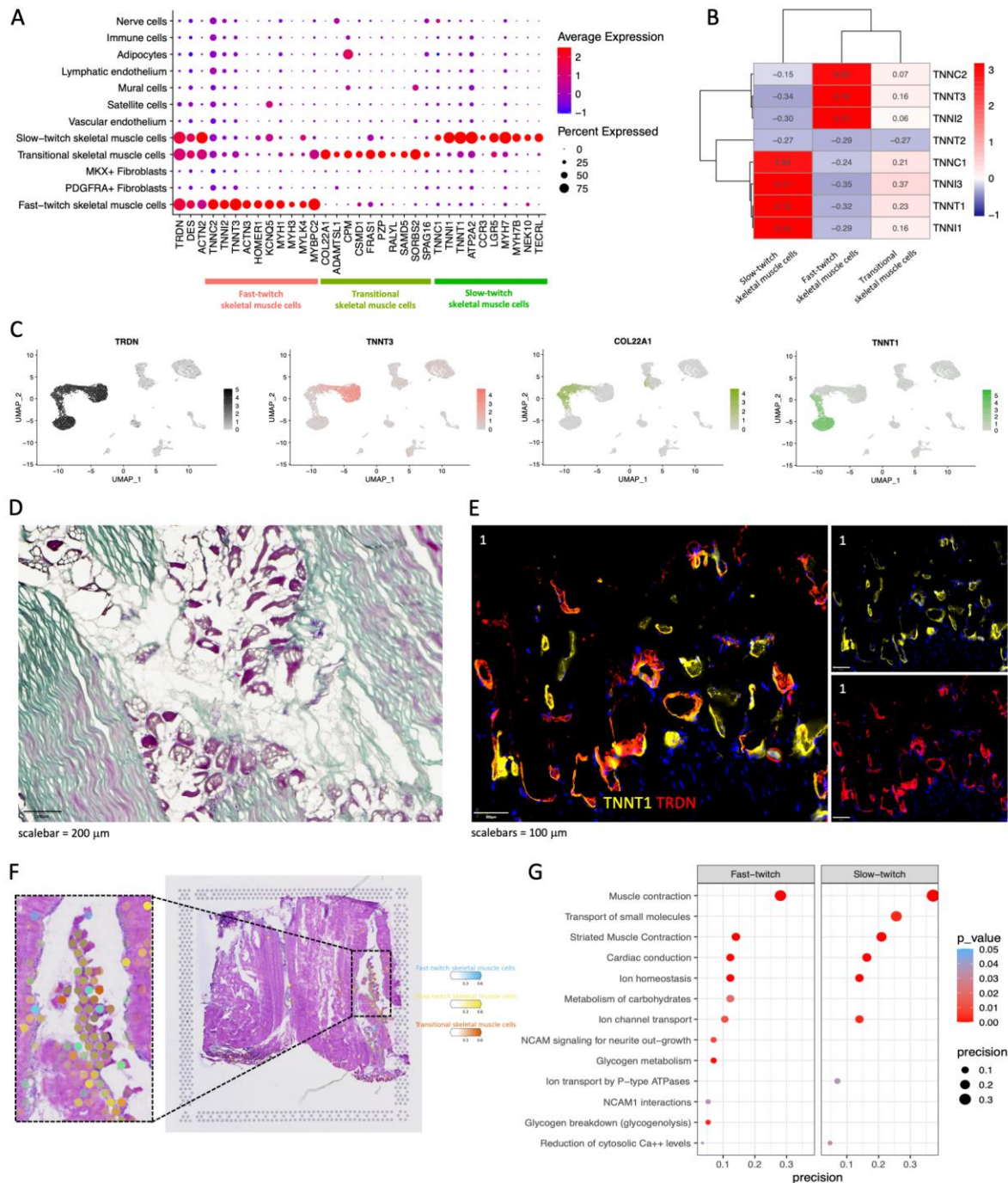
384

385 **Three types of skeletal muscle-like cells are present in human hamstring tendon**

386 Human tendons are highly transitional tissues that transition from bone attachment (enthesis)
387 through tendon midbody to the muscle attachment (myotendinous junction, MTJ). The tissue used in
388 this study includes the region of hamstring tendon from the midbody to the MTJ. Although
389 macroscopic muscle at the MTJ was removed from tendon samples during tissue processing for
390 sequencing, tendon regions are a continuum and it is unsurprising that skeletal muscle was present in
391 this area. Three cell types were identified with high expression of the skeletal muscle markers triadin
392 (*TRDN*) and desmin (*DES*)(Figure 4A-C). Two of the three clusters expressed high levels of troponins:
393 while one cluster expressed high levels troponin isoforms associated with fast-twitch skeletal muscle
394 (*TNNT3*, *TNNC2*, and *TNNI2*), the other cluster expressed high levels of isoforms associated with slow-
395 twitch skeletal muscle (*TNNT1*, *TNNI1*, and *TNNC1*). The third cell type was not associated with high
396 expression of troponin isoforms, but instead expressed high levels of *COL22A1*, a transcript for an
397 ECM protein which is mainly located in tissue junctions. Therefore, these cells were designated
398 “transitional skeletal muscle cells”. Trichrome staining clearly shows the presence of skeletal muscle
399 (red) infiltrating the collagen dense matrix of tendon (Figure 4D). Furthermore, immunofluorescence
400 confirms the presence of both triadin (TRDN) and troponin T1 (TNNT1) in healthy hamstring tissue
401 (Figure 4E). Visually, all three skeletal muscle cell types are predicted to be located simultaneously
402 within one region (as highlighted). In addition, each skeletal muscle cell type is also located
403 throughout some other parts of the tissue section (Figure 4F). Pathway analysis revealed that both
404 fast- and slow-twitch skeletal muscle cells were associated with “muscle contraction”, “striated
405 muscle contraction”, and “ion homeostasis”, while no Reactome pathways were significantly enriched
406 in the transitional skeletal muscle cells (Figure 4G). As all three of these cell types only have a limited
407 number of potential ligand-receptor interactions (Figure 1C), it is more likely that their interactions
408 are mechanical interactions rather than molecular in hamstring tendon MTJ.

409

Transcriptomic profiling of healthy hamstring tendon



410

411 **Figure 4.** Three types of skeletal muscle cells can be identified in healthy human hamstring tendon

412 using snRNA-seq. (A) Highly expressed genes across the three types of skeletal muscle cells. (B)

413 Heatmap of the expression of all troponins across the three skeletal muscle cell types, identifying fast-

414 and slow-twitch skeletal muscle cells. (C) Feature plots of the expression of *TRDN*, *TNNT3*, *COL22A1*,

415 and *TNNT1* in the snRNA-seq datasets. (D) Trichrome staining of human healthy hamstring tendon

416 tissue, including muscle (red) and collagen (green). (E) Immunofluorescence staining of troponin T1

417 (TNNT1, yellow) and triadin (TRDN, red) in human healthy hamstring tendon tissue. (F) Cell2location

418 prediction of the location of the fast-twitch skeletal muscle cells (blue), slow-twitch skeletal muscle

419 cells (yellow), and transitional skeletal muscle cells (orange) using the spatial transcriptomics (10X
420 Visium) data. (G) Enriched pathways in the Reactome database for each of the three skeletal muscle
421 cell types (no significant pathways for transitional skeletal muscle cells).

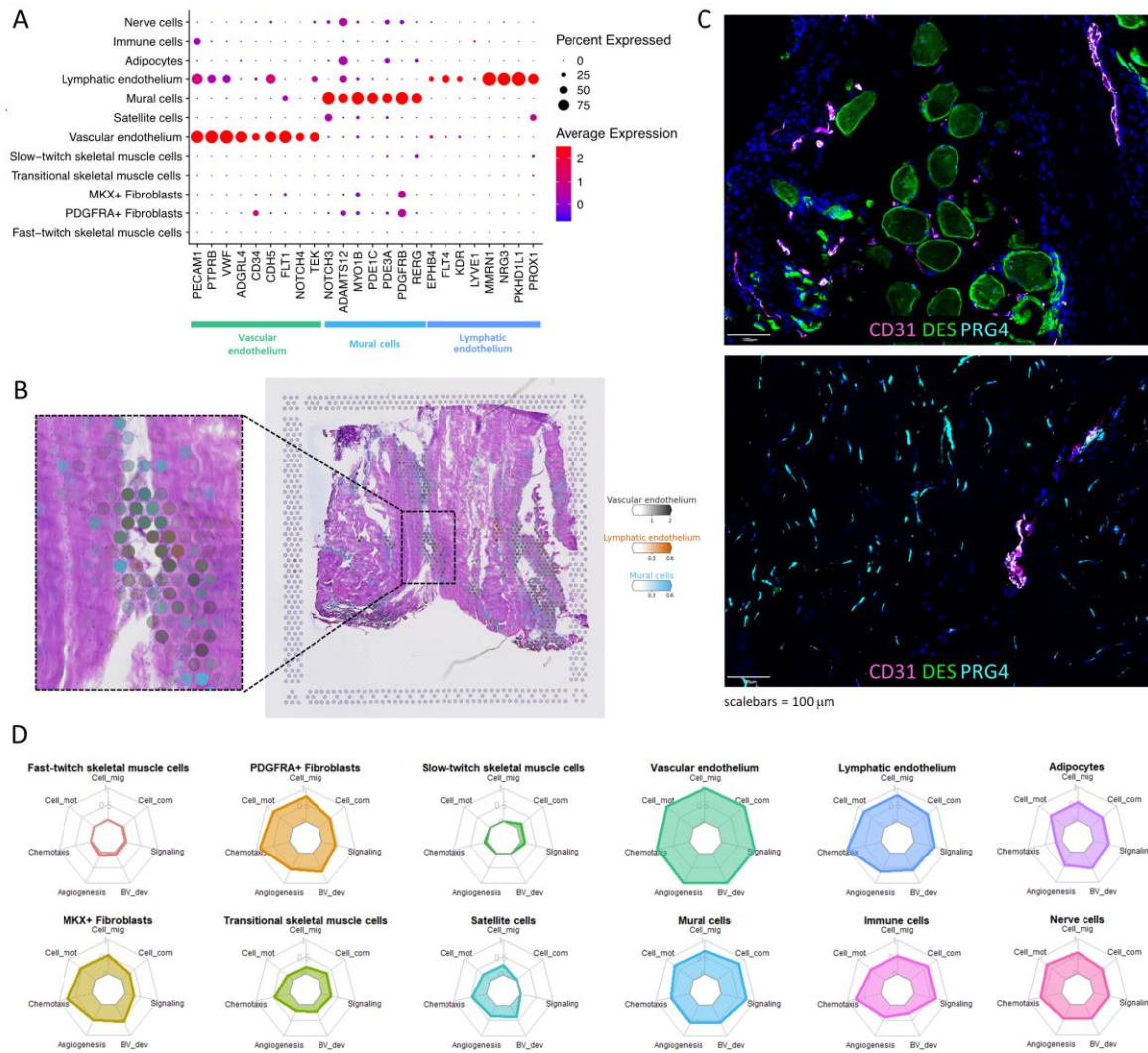
422

423

424 **The vascular compartment of hamstring tendon is composed of three distinct populations**

425 The vascular compartment includes 3 distinct populations: vascular endothelium (*FLT1*, *CD34*,
426 *NOTCH4*), lymphatic endothelial (*MMRN1*, *PROX1*, *KDR*, *FLT4*), and mural cells (*NOTCH3*, *PDGFRB*,
427 *MYO1B*) (Figure 5A). Cell2location analysis demonstrates a large overlap in the predicted locations of
428 vascular endothelial cells (black), mural cells (blue), and lymphatic endothelium (orange), suggesting
429 possible interactions between these cell types (Figure 5B, Figure 1C). Immunofluorescence confirms
430 the presence of vessels across healthy hamstring tendon tissue. However, vessels seem to be much
431 more abundant in areas close to muscle fibres than in the main body of the tendon (Figure 5C). GO:BP
432 pathways associated with vascular processes, including cell migration, cell communication, signalling,
433 blood vessel development, angiogenesis, chemotaxis, and cell motility, were all strongly enriched in
434 vascular endothelium. Strong enrichment of most of these pathways was also found in all other cell
435 types, except for skeletal muscle cells and satellite cells, which only showed minimal enrichment
436 (Figure 5D). Enriched pathways from the Reactome database were “signal transduction” and
437 “signalling by Rho GTPases” for vascular endothelium, “signal transduction” and “signalling by GPCR”
438 in mural cells, and “RHO GTPase cycle” in lymphatic endothelium (Supplementary Figure 8). The
439 predicted ligand-receptor interactions show a high number of predicted interactions between the
440 three vascular cell types and the two fibroblast cell types, especially between mural cells and MKX+
441 fibroblast (Figure 1D). Among the top predicted ligand-receptor interactions from mural cells to MKX+
442 fibroblast are pairs involved in the GO:BP pathways signal transduction (COPA & EGFR), angiogenesis
443 (FGF1 & TGFBR3) and BMP signalling (BMP5 & multiple receptor complexes). Among the top
444 predicted ligand-receptor interactions from MKX+ fibroblasts to mural cells are immune response /
445 cell adhesion (THBS1 & CD36), FGF2 & CD44 which are involved in ECM organisation and regulation of
446 ERK pathways, and BMP signalling (BMP5 and BMP6 & multiple receptor complexes) (Supplementary
447 Table 1).

448



449

450 **Figure 5.** Vascular compartment in healthy hamstring tendon. (A) Expression of endothelial and

451 cell markers across all different cell types in the snRNA-Seq datasets. (B) Predicted location of vascular

452 endothelium (black), lymphatic endothelium (orange), and mural cells (blue) fibroblasts in the spatial

453 transcriptomics data (Visium) using cell2location analysis. (C) Immunofluorescence. (D) Radar plots of

454 the enrichment score (0-1) of gene ontology biological processes (GO:BP) related to vascular function.

455 Clockwise starting at the top: cell migration, cell communication, signalling, blood vessel

456 development, angiogenesis, chemotaxis, and cell motility.

457

458

459 Macrophage, T cell, and dendritic cells are present in healthy hamstring tendon

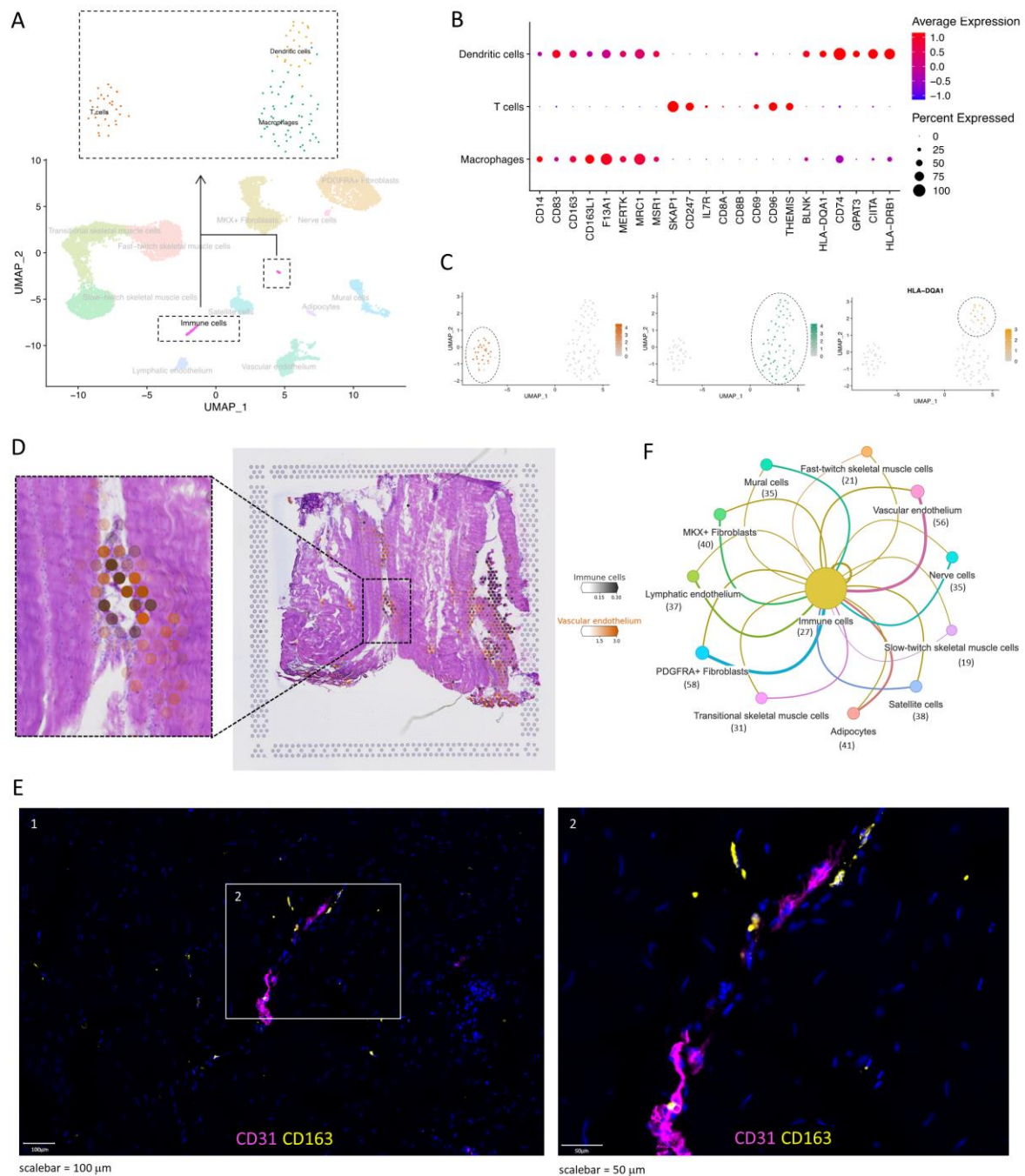
460 Re-clustering of the identified immune cells (108 nuclei, ~1% of all nuclei) revealed two distinct cell

461 clusters representing three different cell types. The three immune cell subsets were identified as

462 macrophages (*CD14*, *CD163*, *MERTK*, *MRC1*, *MSR1*), T cells (*CD247*, *IL7R*, *CD69*, *THEMIS*), and

463 dendritic cells (*HLA-DQA1*, *CD74*, *GPAT3*, *CIITA*, *HLA-DRB1*) (Figure 6A-C). Although further delineation

464 of the T cell clusters was not possible due to the small number of cells and the low expression of CD4
465 and CD8 classical T cell markers, the expression of *CD69* could indicate that at least a subset of these
466 cells are tissue-resident memory T cells. The strong expression of *MERTK* in the macrophage
467 population also suggest that this population is mostly comprised of tissue-resident macrophages.
468 Cell2location predicts the presence of immune cells in similar areas as vascular endothelium, but also
469 in areas where skeletal muscle cells are predicted to be located (Figure 6D). Presence of immune cells
470 in hamstring tissue was confirmed by immunofluorescence staining, which showed that immune cells
471 could mainly be found around vessels and colocalization of CD163 macrophages with CD3 T-cells
472 (Figure 6E). A relatively low number of potential ligand-receptor interactions were identified for
473 immune cells, indicating a low activation state of the immune cells in our datasets, consistent with the
474 samples coming from healthy hamstring (Figure 6F). Reactome pathways enriched in immune cells
475 included “Signal Transduction”, “Innate Immune System”, and “Adaptive Immune System”
476 (Supplementary Figure 9).
477



478

479

480

481

482

483

484

485

486

Figure 6. Immune cells in healthy hamstring tendon. (A) The immune cell cluster was re-clustered, revealing two distinct clusters representing T cells and macrophages. (B) Expression of immune cell markers in the two cell types within the immune cell cluster. (C) Featureplots of *SKAP1* and *MRC1* in the immune cell cluster. (D) Cell2location prediction of the location of immune cells and vascular endothelium using the spatial transcriptomics (10X Visium) data. (E) Immunofluorescence confirms the presence of CD163+ macrophages in healthy hamstring tendon tissue. (F) Predicted number of interactions of immune cells with the other cell types within the snRNA-seq datasets using CellPhoneDB and InterCellar.

487

488

489 **Adipocytes, satellite cells, and nerve cells are found close to fibroblasts in healthy hamstring tendon**

490 Finally, three cell types that were not previously reported in human tendon single cell transcriptomics

491 studies, namely adipocytes, satellite cells, and nerve cells, which are undefined nervous system cells,

492 were identified in our datasets (Figure 7A). This small cluster of nerve cells, or undefined nervous

493 system cells, represented only ~0.5% of all cells and expressed markers related to both nerve cells

494 and glial cells. Adipocytes and satellite cells were identified by expression of known markers (Figure

495 7A). Cell2location analysis (Figure 7B) showed that adipocytes were predicted to be located in similar

496 areas as PDGFRA+ fibroblasts and skeletal muscle cells (as per Figure 3D and Figure 4F, respectively),

497 and satellite cells and nerve cells were predicted to be located in similar areas as fibroblasts and cells

498 from the vascular compartment (as per Figure 3D and 5B, respectively). Immunohistochemistry

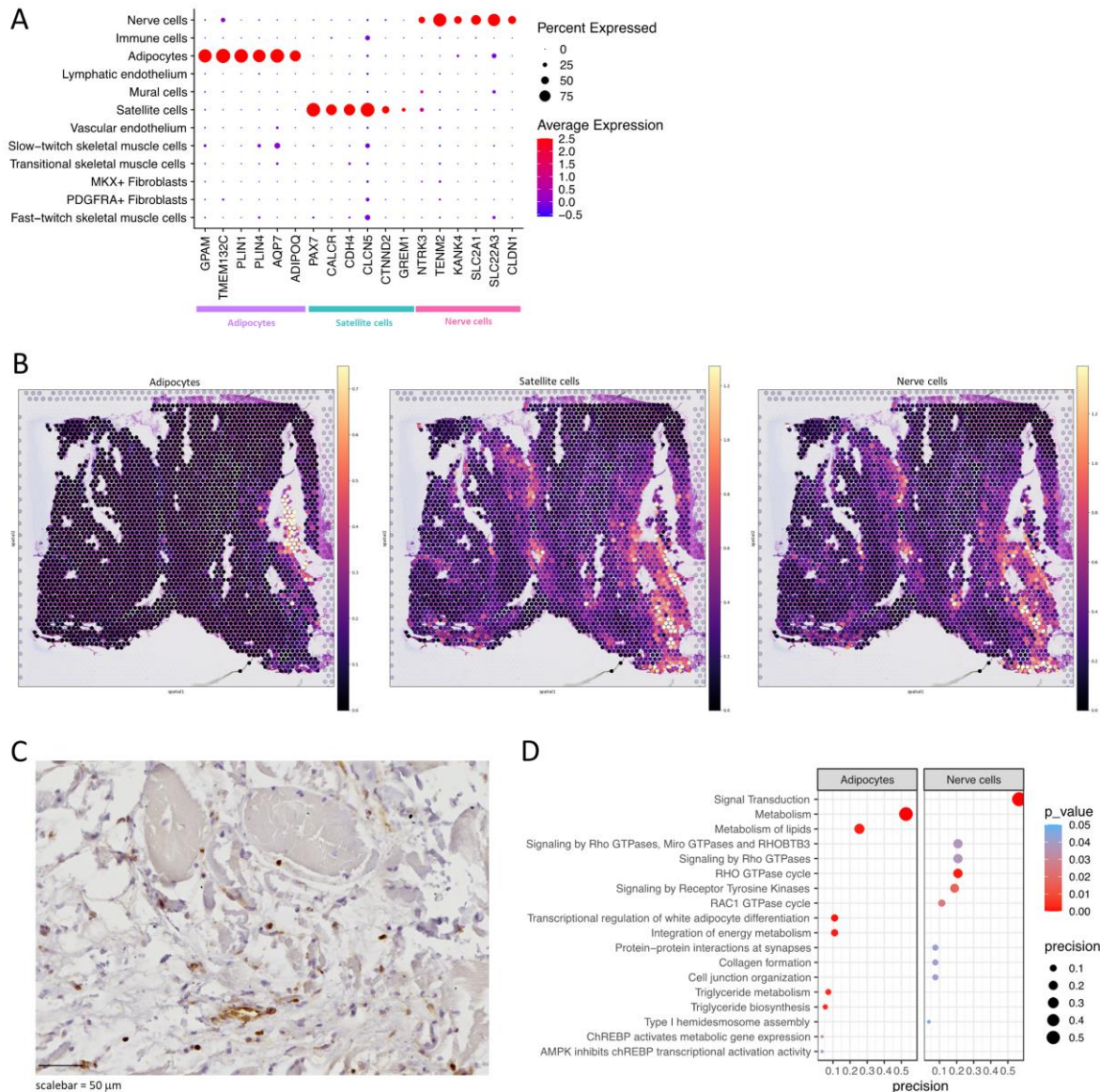
499 staining of PAX7 revealed that satellite cells are present in close proximity to muscle fibres within

500 tendon tissue (Figure 7C). No enriched Reactome pathways were found for satellite cells. However,

501 Reactome pathways “Metabolism” and “Metabolism of lipids” were significantly enriched in

502 adipocytes and “Signal Transduction” in nerve cells, among others (Figure 7D).

503



504

505 **Figure 7.** Adipocytes, satellite cells, and nerve cells in healthy human hamstring tendon. (A)

506 Expression of highly expressed markers in adipocytes, satellite cells, and nerve cells. (B) Cell2location

507 analysis predicted location of adipocytes, satellite cells, and nerve cells. (C) Reactome results of

508 enriched pathways for adipocytes and nerve cells (no significant pathways for satellite cells). (D)

509 Immunohistochemistry staining for PAX7 reveals that satellite cells are present in close proximity to

510 muscle fibres in tendon tissue.

511 **Discussion**

512

513 In this study, we systematically characterised healthy human hamstring tendon using snRNA-seq,
514 spatial transcriptomics, and imaging, resulting in a comprehensive atlas that can act as a reference
515 map for future studies. SnRNA-seq was utilised in this study as it can deliver considerably enhanced
516 extraction of nuclei from dense collagen-rich tissue and the removal of a stress-response to
517 prolonged enzymatic digestion. Unsupervised clustering was used to identify 12 cell types, which
518 were annotated according to known markers from the literature, differentially expressed genes, and
519 pathway analysis. We confirmed the presence of two distinct fibroblast subsets along with endothelial
520 cells, mural cells, and immune cells, and revealed the presence of less well-characterised cells,
521 including different skeletal muscle cell types, satellite cells, adipocytes, and nerve cells. The low
522 frequency of immune cells in our datasets (~1%) corroborates that this is indeed healthy tissue. Our
523 datasets suggest that fibroblasts are key regulators of hamstring tendon tissue homeostasis due to
524 their important role in the production and organisation of ECM as well as due the high number of
525 potential ligand-receptor interactions, both within each fibroblast cell type as well as with all other
526 cell types.

527

528 Our datasets show both similarities and differences with previous human single cell transcriptomics
529 studies[8, 9]. These studies identified several different fibroblasts, mural cells, endothelial cells, and
530 immune cells, including monocytes/macrophages, T cells, and dendritic cells. Most markers that were
531 used to differentiate between different tenocyte populations in the paper by Kendal *et al.* were
532 present in either both or neither of our fibroblast populations; markers used by Akbar *et al.* to identify
533 normal tenocyte1 and normal tenocyte2 mostly aligned with our PDGFRA+ fibroblasts and MKX+
534 fibroblasts, respectively (Supplementary Figure 10). However, these studies did not detect the
535 presence of adipocytes and nerve cells in their tissues and have low representation of skeletal muscle
536 cells or satellite cells. Differences between this study and previous studies could have arisen due to
537 the fact that a greater number of healthy nuclei was analysed in our study or due to the differences in
538 transcriptomics methods (single cell vs single nucleus approaches). While single cell transcriptomics
539 generally capture more RNA per cell, including stable transcripts in the cytoplasm that might not be
540 present in the nucleus, single nucleus transcriptomics enables the capture of a wider variety of cells,
541 including cells that are hard to capture with single cell techniques due to the buoyancy or larger size
542 of cells[26, 27].

543

544 Due to the origin of the tissue, it is unsurprising that skeletal muscle cells were found in hamstring
545 tendon samples. The multinucleated nature of skeletal muscle combined with the low cellularity of
546 tendon could explain the relatively high percentage of skeletal muscle cells (~45%) in our datasets.
547 The limited number of predicted ligand-receptor interactions of skeletal muscle cell types with all
548 other cell types suggests that skeletal muscle fibres might be present in healthy hamstring MTJ mostly
549 for mechanical attachment. It is interesting to note that the PDGFRA+ fibroblast subset enriched in
550 elastin biology sits spatially closer to the muscle and further study would be warranted to establish
551 functional specialisation linked to mechanical attachment. The single cell transcriptomics study by
552 Akbar *et al.* also identified a small muscle subset in their tendon datasets. However, the three
553 markers defining the muscle cell subset were all present in the satellite cells but not skeletal muscle
554 cells in our dataset, suggesting that these might have been satellite cells[9]. The satellite cells in our
555 datasets did not cluster together with skeletal muscle cells and their regulon activation profile was
556 more similar to non-skeletal muscle cells such as mural cells, endothelial cells, and fibroblasts, than
557 skeletal muscle cells. Interestingly, a previous study has shown that there is a loss of satellite cells in
558 skeletal muscle after tendon injury[28], illustrating that these cells can respond to changes in tendon.
559 However, more work is needed to better understand the role and localisation of satellite cells as well
560 as the potential interactions of these presumably skeletal muscle fibre-based cells with tendon-based
561 cells.

562

563 Analysis of regulatory networks using SCENIC corroborated our snRNA-seq cell type definitions and
564 also gave further insight into the potential functions of each identified cell type, especially the two
565 fibroblast subsets. Regulons upregulated in the MKX+ fibroblasts include PAX9, which is involved in
566 tendon development [29]. However, most other regulons have previously been associated with other
567 tissues: for example MAFB, which regulates macrophage differentiation[30], and ZFH3, which is
568 associated with neurons[31]. The link with neurons is intriguing as these MKX+ fibroblasts expressed
569 high levels of *PIEZO2*, piezo-type mechanosensitive ion channel component 2, which is known to be
570 critical for proprioception and has previously been identified in the Golgi tendon organs in mice[32].
571 Further studies will be required to validate the function of these regulons in tendon fibroblasts. In the
572 PDGFRA+ fibroblasts, the activated regulons include SHOX2, PRRX2, EBF2, CREB5, and TCF7L2, which
573 are known to be involved in musculoskeletal and (early) limb development, as well as tendon healing;
574 many of these factors are also involved in or can affect wnt/b-catenin signalling [33-39]. Interestingly,
575 PDGFRA+ cells have previously been identified as markers for progenitor cells in human skeletal
576 muscle[40, 41], and *Tpp3+ Pdgfra+* tendon cells have previously been identified as tendon-derived
577 stem cells (TDSC) in mice[42]. While only a small percentage of the PDGFRA+ fibroblasts in our

578 datasets expressed *TPP3*, this PDGFRA+ fibroblast population warrants further investigation in larger
579 datasets to explore the potential progenitor-like function of (a subpopulation) of these cells.

580

581 Limitations of this study include the small number of included samples in transcriptomic methods.
582 Although different cell types were identified that have not been reported on earlier, the addition of
583 further samples would enable annotation of cell types with more granularity, as well as comparisons
584 of features such as donor sex, age, and side of tendon. The possible functions of and interaction
585 between different cell types were predicted based on mRNA expression only. Two difficulties with this
586 are that snRNA-seq might not cover stable transcripts that are only present in the cytoplasm, and that
587 the presence or lack of mRNA expression does not always translate to the presence or lack of protein
588 expression. Therefore, additional studies are needed to validate cell-cell interactions and discover
589 their functional effects. This might especially give more insight into the differences and similarities of
590 the two different fibroblast cell types identified in this study. While spatial transcriptomics data was
591 able to add some information about the presence and distribution of different cell types across the
592 tissue, analysis is limited due to use of one sample in one tissue slice orientation. The lack of single-
593 cell resolution in spatial transcriptomics and difficulty in accurate normalisation due to high variability
594 in RNA expression levels between muscle and tendon areas were also limiting. The low cellularity of
595 tendon and low transcriptional activity of the cells results in low detection of RNA in many parts of the
596 tissue. The large differences in expression between high and low cellularity regions of the tissue could
597 have biased the predicted locations of cells. Therefore, other spatial analysis techniques might be
598 better suited to understand the spatial distribution of cells in tendon. Cell subsets were identified
599 using RNA markers, and where possible these were validated by imaging methods. However, in some
600 cases we were unable to obtain suitable antibodies (such as for mohawk).

601

602 In conclusion, this study enhances our understanding of cellular composition of healthy human
603 hamstring tendon by using both transcriptional and spatial analyses. We show that healthy human
604 hamstring tendon is comprised of wide range of cell types, including both previously demonstrated
605 cell types such as fibroblasts, endothelial cells, mural cells, and immune cells, as well as less well-
606 characterised cells, including different skeletal muscle cell types, satellite cells, adipocytes, and nerve
607 cells. We identified two distinct types of fibroblasts, which are suggested to be the major regulators
608 of tendon tissue homeostasis due their role in the production and organisation of ECM, the high
609 number of predicted interactions with all other cell types, and the wide distribution across the tissue.
610 Our datasets form the foundation of a comprehensive cellular atlas of healthy tendons that can act as

- 611 a reference map for future studies, and will help dissect mechanisms of disease pathogenesis and
612 identify new therapeutic targets.

613 **Funding information**

614 This research was funded by the Chan Zuckerberg Initiative (CZIF2019-002426) and supported by the
615 National Institute for Health Research (NIHR) and the NIHR Oxford Biomedical Research Centre. APC is
616 supported by a Medical Research Council Career Development Fellowship (MR/V010182/1). PH has
617 funding from the Pagets Association (PA21010).

618

619 **Disclosure statement**

620 APC is listed as an inventor on several patents filed by Oxford University Innovations concerning
621 single-cell sequencing technologies.

622

623 **Data availability statement**

624 All the code used for the data analysis of this paper is available here: [https://github.com/Botnar-MSK-](https://github.com/Botnar-MSK-Atlas/hamstring_atlas)
625 [Atlas/hamstring_atlas](https://github.com/Botnar-MSK-Atlas/hamstring_atlas). All the data in this manuscript will be openly available on Lattice on final
626 publication and is hosted on Lattice. For access prior to final publication please contact the authors.

627

628 **Acknowledgements**

629 We would like to thank our research assistant Louise Appleton, our research nurses Debra Beazley,
630 Bridget Watkins, Kim Wheway, and Lois Vesty-Edwards, and the knee surgeon team at the Nuffield
631 Orthopaedic Centre, for their invaluable help collecting human tissue for this study. We would like to
632 thank the CellDive team at the Kennedy Institute of Rheumatology, including Mark Coles, Dylan
633 Windell, and Ananya Bhalla, for their help with the immunofluorescence staining using the CellDive
634 system. Our thanks also goes to Dr Carla Cohen for helpful discussions and for reading of the
635 manuscript. Finally, we would like to thank all other members of the CZI Tendon Seed Network for
636 insightful discussions.

637 **References**

- 638 1. Bordoni B, Varacallo M (2022) Anatomy, Tendons. StatPearls Publishing
- 639 2. Millar NL, Silbernagel KG, Thorborg K, Kirwan PD, Galatz LM, Abrams GD, Murrell GAC,
640 McInnes IB, Rodeo SA (2021) Tendinopathy. Nat Rev Dis Prim.
641 <https://doi.org/10.1038/s41572-020-00234-1>
- 642 3. KJÆR M (2004) Role of Extracellular Matrix in Adaptation of Tendon and Skeletal Muscle to
643 Mechanical Loading. Physiol Rev 84:649–698
- 644 4. Dymant NA, Galloway JL (2015) Regenerative Biology of Tendon: Mechanisms for Renewal and
645 Repair. Curr Mol Biol Reports 1:124–131
- 646 5. Garcia-Melchor E, Cafaro G, Macdonald L, Crowe LAN, Sood S, Mclean M, Fazzi UG, Mcinnes
647 IB, Akbar M, Millar NL (2021) Novel self-amplificatory loop between T cells and tenocytes as a
648 driver of chronicity in tendon disease. Ann Rheum Dis 0:1–11
- 649 6. Ramos-Mucci L, Sarmiento P, Little D, Snelling S (2022) Research perspectives—Pipelines to
650 human tendon transcriptomics. J Orthop Res. <https://doi.org/10.1002/jor.25315>
- 651 7. Baldwin MJ, Cribbs AP, Guilak F, Snelling SJB (2021) Mapping the musculoskeletal system one
652 cell at a time. Nat Rev Rheumatol 17:247–248
- 653 8. Kendal AR, Layton T, Al-Mossawi H, Appleton L, Dakin S, Brown R, Loizou C, Rogers M, Sharp R,
654 Carr A (2020) Multi-omic single cell analysis resolves novel stromal cell populations in healthy
655 and diseased human tendon. Sci Rep 10:13939
- 656 9. Akbar M, MacDonald L, Crowe LAN, Carlberg K, Kurowska-Stolarska M, Ståhl PL, Snelling SJB,
657 McInnes IB, Millar NL (2021) Single cell and spatial transcriptomics in human tendon disease
658 indicate dysregulated immune homeostasis. Ann Rheum Dis 80:1494–1497
- 659 10. de Micheli AJ, Swanson JB, Disser NP, Martinez LM, Walker NR, Oliver DJ, Cosgrove BD,
660 Mendias CL (2020) Single-cell transcriptomic analysis identifies extensive heterogeneity in the
661 cellular composition of mouse Achilles tendons. Am J Physiol - Cell Physiol 319:C885–C894
- 662 11. Mimpfen JY, Paul C, Ramos-Mucci L, Tendon Seed Network, Cribbs A, Baldwin M, Snelling S
663 (2022) SOP for snap-freezing tissues.
664 <https://doi.org/https://dx.doi.org/10.17504/protocols.io.b2tqqemw>
- 665 12. Mimpfen JY, Paul C, Tendon Seed Network, Cribbs A, Snelling S (2021) Nuclei isolation from
666 snap-frozen tendon tissue for single nucleus RNA Sequencing. protocols.io.
667 <https://doi.org/https://dx.doi.org/10.17504/protocols.io.bc6xizfn>
- 668 13. Germain P-L, Lun A, Garcia Meixide C, Macnair W, Robinson MD (2022) Doublet identification
669 in single-cell sequencing data using scDbfFinder. F1000Research 10:979
- 670 14. Campbell J, Yang S, Wang Z, Corbett S, Koga Y (2022) celda: CELLular Latent Dirichlet Allocation.

- 671 R package version 1.14.0. In: Bioconductor.
- 672 15. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C,
673 Zager M, Hoffman P, Stoeckius M, Papalexi E, Mimitou EP, Jain J, Srivastava A, Stuart T,
674 Fleming LM, Yeung B, Rogers AJ, McElrath JM, Blish CA, Gottardo R, Smibert P, Satija R (2021)
675 Integrated analysis of multimodal single-cell data. *Cell* 184:3573-3587.e29
- 676 16. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh P,
677 Raychaudhuri S (2019) Fast, sensitive and accurate integration of single-cell data with
678 Harmony. *Nat Methods* 16:1289–1296
- 679 17. Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H (2020) gprofiler2 -- an R package for gene
680 list functional enrichment analysis and namespace conversion toolset g:Profiler.
681 *F1000Research* 9:709
- 682 18. Cakir B, Prete M, Huang N, van Dongen S, Pir P, Kiselev VY (2020) Comparison of visualization
683 tools for single-cell RNAseq data. *NAR Genomics Bioinforma.*
684 <https://doi.org/10.1093/nargab/lqaa052>
- 685 19. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F,
686 Marine J-C, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S (2017) SCENIC: single-
687 cell regulatory network inference and clustering. *Nat Methods* 14:1083–1086
- 688 20. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, Seurinck R, Saelens
689 W, Cannoodt R, Rouchon Q, Verbeiren T, De Maeyer D, Reumers J, Saeys Y, Aerts S (2020) A
690 scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc*
691 15:2247–2276
- 692 21. Garcia-Alonso L, Handfield L-F, Roberts K, Nikolakopoulou K, Fernando RC, Gardner L,
693 Woodhams B, Arutyunyan A, Polanski K, Hoo R, Sancho-Serra C, Li T, Kwakwa K, Tuck E, Lorenzi
694 V, Massalha H, Prete M, Kleshchevnikov V, Tarkowska A, Porter T, Mazzeo CI, van Dongen S,
695 Dabrowska M, Vaskivskiy V, Mahbubani KT, Park J, Jimenez-Linan M, Campos L, Kiselev VY,
696 Lindskog C, Ayuk P, Prigmore E, Stratton MR, Saeb-Parsy K, Moffett A, Moore L, Bayraktar OA,
697 Teichmann SA, Turco MY, Vento-Tormo R (2021) Mapping the temporal and spatial dynamics
698 of the human endometrium in vivo and in vitro. *Nat Genet* 53:1698–1711
- 699 22. Interlandi M, Kerl K, Dugas M (2022) InterCellar enables interactive analysis and exploration of
700 cell–cell communication in single-cell transcriptomic data. *Commun Biol* 5:21
- 701 23. Hafemeister C, Satija R (2019) Normalization and variance stabilization of single-cell RNA-seq
702 data using regularized negative binomial regression. *Genome Biol* 20:296
- 703 24. Bergenstråhle J, Larsson L, Lundeberg J (2020) Seamless integration of image and molecular
704 analysis for spatial transcriptomics workflows. *BMC Genomics* 21:482

- 705 25. Kleshchevnikov V, Shmatko A, Dann E, Aivazidis A, King HW, Li T, Elmentaite R, Lomakin A,
706 Kedlian V, Gayoso A, Jain MS, Park JS, Ramona L, Tuck E, Arutyunyan A, Vento-Tormo R,
707 Gerstung M, James L, Stegle O, Bayraktar OA (2022) Cell2location maps fine-grained cell types
708 in spatial transcriptomics. *Nat Biotechnol* 40:661–671
- 709 26. Norreen-Thorsen M, Christopher Struck E, Uhlé M, James Dusart P, Marie Butler
710 Correspondence L (2022) A human adipose tissue cell-type transcriptome atlas.
711 <https://doi.org/10.1016/j.celrep.2022.111046>
- 712 27. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, Hughes TK,
713 Wadsworth MH, Burks T, Nguyen LT, Kwon JYH, Barak B, Ge W, Kedaigle AJ, Carroll S, Li S,
714 Hacohen N, Rozenblatt-Rosen O, Shalek AK, Villani A-C, Regev A, Levin JZ (2020) Systematic
715 comparison of single-cell and single-nucleus RNA-sequencing methods. *Nat Biotechnol*
716 38:737–746
- 717 28. Gil-Melgosa L, Grasa J, Urbiola A, Llombart R, Susaeta Ruiz M, Montiel V, Ederra C, Calvo B, Ariz
718 M, Ripalda-Cemborain P, Prosper F, Ortiz-de-Solórzano C, Pons-Villanueva J, Pérez Ruiz A
719 (2021) Muscular and Tendon Degeneration after Achilles Rupture: New Insights into Future
720 Repair Strategies. *Biomedicines* 10:19
- 721 29. Peters H, Neubüser A, Kratochwil K, Balling R (1998) Pax9-deficient mice lack pharyngeal
722 pouch derivatives and teeth and exhibit craniofacial and limb abnormalities. *Genes Dev*
723 12:2735–2747
- 724 30. Hamada M, Tsunakawa Y, Jeon H, Yadav MK, Takahashi S (2020) Role of MafB in macrophages.
725 *Exp Anim* 69:1–10
- 726 31. Kim TS, Kawaguchi M, Suzuki M, Jung CG, Asai K, Shibamoto Y, Lavin MF, Khanna KK, Miura Y
727 (2010) The ZFH3 (ATBF1) transcription factor induces PDGFRB, which activates ATM in the
728 cytoplasm to protect cerebellar neurons from oxidative stress. *DMM Dis Model Mech* 3:752–
729 762
- 730 32. Woo S-H, Lukacs V, de Nooij JC, Zaytseva D, Criddle CR, Francisco A, Jessell TM, Wilkinson KA,
731 Patapoutian A (2015) Piezo2 is the principal mechanotransduction channel for proprioception.
732 *Nat Neurosci* 18:1756–1762
- 733 33. Vickerman L, Neufeld S, Cobb J (2011) Shox2 function couples neural, muscular and skeletal
734 development in the proximal forelimb. *Dev Biol* 350:323–336
- 735 34. Fowler DA, Larsson HCE (2020) The tissues and regulatory pattern of limb chondrogenesis. *Dev*
736 *Biol* 463:124–134
- 737 35. Bai WW, Tang ZY, Shan TC, Jing XJ, Li P, Qin WD, Song P, Wang B, Xu J, Liu Z, Yu HY, Ma ZM,
738 Wang SX, Liu C, Guo T (2020) Up-regulation of paired-related homeobox 2 promotes cardiac

- 739 fibrosis in mice following myocardial infarction by targeting of Wnt5a. *J Cell Mol Med*
740 24:2319–2329
- 741 36. Liu H, Xu J, Liu CF, Lan Y, Wylie C, Jiang R (2015) Whole transcriptome expression profiling of
742 mouse limb tendon development by using RNA-seq. *J Orthop Res* 33:840–848
- 743 37. Best KT, Korcari A, Mora KE, Nichols AE, Muscat SN, Knapp E, Buckley MR, Loisel AE (2021)
744 Scleraxis-lineage cell depletion improves tendon healing and disrupts adult tendon
745 homeostasis. *Elife* 10:1–64
- 746 38. Zhang CH, Gao Y, Hung HH, Zhuo Z, Grodzinsky AJ, Lassar AB (2022) Creb5 coordinates synovial
747 joint formation with the genesis of articular cartilage. *Nat Commun* 13:7295
- 748 39. Contreras O, Soliman H, Theret M, Rossi FMV, Brandan E (2020) TGF- β -driven downregulation
749 of the transcription factor TCF7L2 affects Wnt/ β -catenin signaling in PDGFR α + fibroblasts. *J*
750 *Cell Sci*. <https://doi.org/10.1242/jcs.242297>
- 751 40. Uezumi A, Fukada S, Yamamoto N, Ikemoto-Uezumi M, Nakatani M, Morita M, Yamaguchi A,
752 Yamada H, Nishino I, Hamada Y, Tsuchida K (2014) Identification and characterization of
753 PDGFR α + mesenchymal progenitors in human skeletal muscle. *Cell Death Dis* 5:e1186–e1186
- 754 41. Farahani RM, Xaymardan M (2015) Platelet-Derived Growth Factor Receptor Alpha as a
755 Marker of Mesenchymal Stem Cells in Development and Stem Cell Biology.
756 <https://doi.org/10.1155/2015/362753>
- 757 42. Harvey T, Flamenco S, Fan C-M (2019) A Tppp3+Pdgfra+ tendon stem cell population
758 contributes to regeneration and reveals a shared role for PDGF signalling in regeneration and
759 fibrosis. *Nat Cell Biol* 21:1490–1503
- 760