A single-cell transcriptional atlas identifies extensive heterogeneity in the cellular composition of tendons

Jacob B Swanson[1], Andrea J De Micheli[2], Nathaniel P Disser[1], Leandro M Martinez[1], Nicholas R Walker[1,3], Benjamin D Cosgrove[2], Christopher L Mendias[1,3,*]

[1]Hospital for Special Surgery, New York, NY, USA

[2]Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA

[3]Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY, USA

*Corresponding Author
Christopher Mendias, PhD
Hospital for Special Surgery
535 E 70th St
New York, NY 10021
USA
+1 212-606-1785
mendiasc@hss.edu

Keywords: tenocyte; tendon fibroblast; pericyte; single-cell RNA sequencing

Abstract

Tendon is a dense connective tissue that transmits forces between muscles and bones. Cellular heterogeneity is increasingly recognized as an important factor in the biological basis of tissue homeostasis and disease, but little is known about the diversity of cells that populate tendon. Our objective was to explore the heterogeneity of cells in mouse Achilles tendons using single-cell RNA sequencing. We identified 13 distinct cell types in tendons, including 4 previously undescribed populations of fibroblasts. Using pseudotime trajectory analysis, we provide additional support for the notion that pericytes are progenitor cells for the fibroblasts that compose adult tendons. These findings identify notable heterogeneity between and within tendon cell populations. This heterogeneity may have important implications for our understanding of how the tendon extracellular matrix is assembled and maintained, and for the design of therapies to treat tendinopathies.

Introduction

Tendon is a dense connective tissue that transmits forces between muscles and bones. The extracellular matrix (ECM) of tendon is composed of type I collagen, as well as other collagens, proteoglycans, and matrix proteins (1). Tendon fibroblasts, or tenocytes, are the main cell type in tendons and are thought to be responsible for ECM production, organization, and maintenance (1). During development and early postnatal stages, tendon is a relatively cellular tissue with high rates of cell proliferation, but by 3 weeks after birth the tendons of mice become hypocellular with low rates of cellular turnover (2, 3). Tendon allows for forces to be efficiently transmitted through the highly organized network of ECM proteins, but in the case of excess loading or ECM damage, the low cellularity and limited capacity of resident tendon fibroblasts to repair damaged matrix can lead to frank tendon ruptures and degenerative tendinopathies (1, 4).

Cellular heterogeneity is increasingly recognized as important for the biological function of tissues, and in the development of new therapies for diseases (5). Single-cell RNA sequencing (scRNAseq) is a technique that allows for the investigation of cellular heterogeneity within tissues (5). While single-cell qPCR has been used to analyze selected gene expression in cultured tendon fibroblasts (6), to our knowledge comprehensive transcriptional profiling of single-cells isolated directly from tendon tissue has not been performed. Therefore, our objective was to use scRNAseq to establish a transcriptional atlas of tendon to explore the cellular heterogeneity of this functionally important connective tissue.

Results and Discussion

We isolated Achilles tendons, which are the major load bearing tendons of the hindlimb, from 6-week old male C57BL/6J mice. This age was selected to be reflective of early adulthood, as the cells within tendon have low proliferation rates and are likely specified at this point, but the ECM is still actively being synthesized as the skeleton continues to slowly lengthen (2, 3, 7). Tendons were digested to generate a single-cell suspension, and scRNAseq was performed. From the 1197 cells that were isolated and passed quality control validation, 13 distinct populations of cells in tendons were identified (Figure 1A). The top groups of genes that are uniquely expressed in each cell type are shown in Figure 1B. An online, interactive atlas is available at https://www.hss.edu/genomics-research-center.asp. Flow cytometry using cell surface antigens was also performed as a quantitative validation step for scRNAseq (Figure 1E-F).

We identified three groups of fibroblasts that we refer to as tendon fibroblasts 1, 2, and 3 (Figure 1A-B). These cells express type I collagen (*Col1a1*) at a high level, and it is the basis of high *Col1a1* expression that we define these cells as fibroblasts (Figures 1A,B,E and 2A). For flow cytometry, fibroblasts 1 are present in the $CD31^{-}CD45^{-}CD34^{-}CD146^{-}$ population, and fibroblasts 2 and 3 are in the $CD31^{-}CD45^{-}CD34^{+}CD146^{-}$ population (Figures 1F-G). Each fibroblast type expresses somewhat distinct patterns of the ECM binding genes osteopontin (*Spp1*), dermatopontin (*Dpt*), and SMOC2 (*Smoc2*) (Figures 1A,B,E). We therefore sought to verify the expression patterns of *Spp* and *Dpt* using RNA *in situ* hybridization (ISH). Similar to scRNAseq, *Spp1* and *Dpt* are expressed in separate cells (fibroblasts 1 and 2, respectively), and occasionally in the same cell (fibroblasts 3, Figure 2B). Based on scRNAseq and ISH data, fibroblasts 1 are characterized by high levels of *Spp1*, fibroblasts 2 by low levels of *Spp1* and high levels of *Dpt*, and fibroblasts 3 by high levels of *Dpt* and *Smoc2* with occasional low *Spp1*

expression. Unlike RNA, at the protein level osteopontin, dermatopontin, and SMOC2 are mainly present in overlapping locations (Figure 2C). As tendon fibroblasts are arranged in linear clusters of several cells, these observations indicate that fibroblasts within clusters could be specialized to produce distinct proteins that constitute the ECM.

Scleraxis (*Scx*) is a transcription factor that is required for embryonic tendon development (8). Several studies have used a transgenic *Scx* GFP reporter mouse (ScxGFP), in which GFP is regulated by a 4kb upstream regulatory sequence of *Scx*, to visualize scleraxis expressing cells (8). Through 2 months of age, nearly all tendon fibroblasts of ScxGFP mice robustly express GFP, resulting in the widespread assumption that scleraxis is present in nearly every tendon fibroblast at this age (8). In addition to scleraxis, tenomodulin (*Tnmd*) is a transmembrane protein and mohawk (*Mkx*) is a transcription factor that are thought to also be consistent markers of the tenogenic lineage (8). We therefore expected to observe widespread *Scx*, *Tnmd*, and *Mkx* expression in fibroblasts. To our surprise, only a small portion of tendon fibroblasts 1 express *Scx*, and a minority of fibroblasts 1 and 2 express *Tnmd* (Figures 1A,B,E). *Mkx* was also only expressed in a subset of fibroblasts 1 and 2. We performed RNA ISH for *Scx* and *Tnmd* to confirm these findings. Consistent with scRNAseq, while all of the fibroblasts in tendon express *Col1a1*, only a subset express *Scx* and *Tnmd* (Figure 2A,D). Some cells expressed *Tnmd* but did not express *Scx* (Figure 2E). We also confirmed these findings with scleraxis lineage tracing mice, in which *CreERT2* is driven from the native *Scx* locus and a flox-stop-flox-tdTomato sequence is expressed from the ubiquitous *Rosa26* locus. This allows for *Scx* expressing cells to permanently express *tdTomato* upon treatment with tamoxifen, referred to as *Scx*:*R26*$^{\text{tdTomato}}$ mice (Figure 2G). Following 5 days of tamoxifen treatment, only a small portion of fibroblasts of *Scx*:*R26*$^{\text{tdTomato}}$ mice contained tdTomato (Figure 2F), providing further support

that *Scx* is only expressed in a subset of adult tendon fibroblasts. As *ScxGFP* contains only portions of the regulatory elements of native *Scx*, it is likely that ScxGFP mice overestimate *Scx* expression in adult tendons. While *Col1a1* is expressed in a small portion of endothelial and nerve cells, *Col1a1* may be more useful than *Scx* in lineage tracing or conditional deletion studies across all fibroblasts in tendons.

A fourth group of fibroblasts was identified that displayed moderate *Col1a1* expression. These cells also expressed transcripts for type XXII collagen (*Col22a1*) which is known to be present at tissue junctions, as well as the synovium markers BGLAP2 (*Bglap2*) and matrilin 4 (*Matn4*). Bursae are thin sheets of connective tissue that secrete a protective synovial fluid to reduce friction, and are found at the enthesis of tendons. We suspect these cells are likely fibroblasts of bursal tissue, Sharpey's fibers, or unique fibroblasts at the enthesis, and we have termed these cells bursal fibroblasts.

Pericytes, which express the transmembrane protein CD146 (*Mcam*), are thought to be a progenitor cell population for adult tendon fibroblasts (9). To further explore this relationship, we used the reverse graph embedding machine learning software Monocle (10) to model cell fate decisions in pseudotime. In support of a role for pericytes as progenitor cells, Monocle analysis predicted a differentiation trajectory of pericytes into two branches consisting of bursal fibroblasts and tendon fibroblasts 2 and 3, and tendon fibroblasts 1 and 2 (Figure 1C-D).

Several other cell types were identified with scRNAseq (Figure 1A-B), and select populations were quantified with flow cytometry (Figure 1F-G). The chondrogenic transcription factor *Sox9* is expressed in a portion of cells in all three tendon fibroblast groups, which is consistent with $Scx^{+}/Sox9^{+}$ and $Scx^{-}/Sox9^{+}$ cells found as the tendon ECM transitions into fibrocartilage at the enthesis (8). Three clusters of endothelial cells were identified based on

CD31 (*Pecam1*) expression and absence of the hematopoietic lineage marker CD45 (*Ptprc*), and quantified in flow cytometry as $CD31^{+}CD45^{-}$ cells. These three clusters are likely cells with specialized blood or lymphatic vascular functions. Red blood cells were identified by hemoglobin expression (*Hba-a1, Hba-a2,* and *Hbb-bt*). CD45 demarcated immune cells, with most expressing the M2 macrophage marker CD206 (*Mrc1*), consistent with a tissue resident M2 macrophage population (9). These cells were detected in flow cytometry as $CD31^{-}CD45^{+}CD206^{-}$ cells for other hematopoietic cells, and $CD31^{-}CD45^{+}CD206^{+}$ cells for M2 macrophages.

We identified three groups of nerve cells, with nerve cells 1 and 2 having high expression of myelin protein zero (*Mpz*). Nerve cells 3 also expressed *Mpz*, the voltage gated sodium channel *Scn7a*, the sphingomyelin production and myelin maintenance gene *Abca8a*, and the neuronal cell fatty acid metabolism protein *Dbi* (11). Nerve cells 1 and 2 likely represent the Schwann cells which surround the afferent sensory axons that innervate the Golgi tendon organs and Pacinian corpuscle sensory nerve structures of tendons. Nerve cells 3 could be cells that support the structural and metabolic activities of Schwann cells.

There are several limitations to this study. We used a single age and sex of mice, and while we do not know how cell populations will change with aging, we expect our findings are relevant to both sexes. There were fewer total cells analyzed for scRNAseq than other tissues, but this is likely due to the hypocellular nature of tendons. Despite these limitations, our findings reveal tremendous heterogeneity in the cells that populate tendons. We identify previously undescribed populations of tendon fibroblasts based on relatively distinct expression of ECM proteins, and support the notion that pericytes are progenitor cells of adult fibroblasts. The findings also indicate that fibroblasts could specify to produce distinct components of the tendon ECM, which may have implications in the treatment of tendinopathies.

Materials and Methods

Animals. This study was approved by the Hospital for Special Surgery/Weill Cornell Medical College/Memorial Sloan Kettering Cancer Center IACUC. Male 6-week old C57BL/6J mice (strain 000664) were obtained from the Jackson Laboratory (Bar Harbor, ME, USA). *Scx*^{CreERT2} mice in which an IRES-CreERT2 sequence was inserted between the stop codon and 3' UTR in exon 2 of scleraxis (12), were kindly provided by Dr. Ronen Schweitzer (Shriners Hospitals for Children, Portland, OR, USA). We also obtained *R26*^{tdTomato} reporter mice (13) in which the constitutively expressed *Rosa26* locus was modified to contain a stop codon cassette flanked by loxP sites upstream of the red fluorescent tdTomato gene (Jackson Labs strain 007909). *Scx*^{CreERT2} mice were crossed to *R26*^{tdTomato} mice, and backcrossed again until homozygosity for both alleles was achieved. Six week old male *Scx*$^{CreERT2/CreERT2}$: *R26*$^{tdTomato/tdTomato}$ mice received an intraperitoneal injection of 1mg of tamoxifen (Millipore Sigma, St. Louis, MO, USA) dissolved in 50µL of corn oil for a period of 5 days as described (14) to induce recombination at the *R26* locus in scleraxis-expressing cells, generating animals referred to as *Scx*:*R26*^{tdTomato} mice (Figure 2G).

Surgical procedure. Mice were euthanized by exposure to CO_2 followed by cervical dislocation. To remove Achilles tendons, a longitudinal incision through the skin was made down the midline of the posterior aspect of the lower limb, superficial to the gastrocnemius and Achilles tendon. The paratenon was reflected, and a sharp transverse incision was made just distal to the myotendinous junction and again just proximal to the enthesis, and the Achilles tendon was carefully removed. The procedure was performed bilaterally.

Single cell isolation. The Achilles tendons were digested to obtain a single cell suspension, as modified from a previous study (15). Two Achilles tendons from each animal

were processed together as a single sample, and the tendons of N=4 mice were used. Tendons were finely minced using a scalpel, and then digested for 1 h at 37ºC in a vigorously shaking solution consisting of 16 mg of Collagenase D (Roche, Pleasanton, CA, USA), 3.0 U of Dispase II (Roche), 640 ng of DNase I (Millipore Sigma, St. Louis, MO, USA), 20 µL of 4% bovine serum albumin (BSA, Millipore Sigma) in 2mL of low glucose DMEM (Corning, Corning, NY, USA). After digestion, the single cell suspension was filtered for debris using a 70 µm cell strainer and resuspended in 0.04% BSA in PBS.

Single cell RNA sequencing and analysis. Single cell RNA sequencing and analysis was performed, as modified from a previous study (15). Libraries were prepared using a Chromium Single Cell 3' Reagent Kit (version 3, 10X Genomics, Pleasanton, CA, USA) following the directions of the manufacturer. Cells were loaded into the kit and processed using a Chromium Controller (10X Genomics). Following library preparation and quantification, libraries were sequenced by the Weill Cornell Medical College Epigenomics Core using a HiSeq 2500 system (Illumina, San Diego, CA, USA) Libraries were sequenced to generate approximately 250 million reads per sample, which resulted in on average approximately 960,000 reads per cell. Sequencing data has been deposited to NIH GEO (ascension GSE138515).

Gene expression matrices were generated from the sequencing data using Cell Ranger (version 3.0.1, 10X Genomics) and the mm10 reference genome. Downstream analyses were carried out with R version 3.5.2 (2018-12-20) and the Seurat 2.3.4 R package (16). We combined the four gene expression matrices for more powerful statistical analyses after careful evaluation of batch effect and differences in population number. Genes expressed in less than 3 cells as well as cells <1000 unique molecular identifiers (UMIs) and <200 genes were removed from the gene expression matrix. We also filtered out cells with >20% UMIs mapping to mitochondrial genes.

A total of 1197 cells remained after applying these criteria. We performed principal component analysis (PCA) and used the first 15 PCs for population clustering (unsupervised shared nearest neighbor, SNN, resolution=0.4) and data visualization (UMAP). Finally, differential expression analysis was achieved using the "FindAllMarkers" function in Seurat using a likelihood test that assumes a negative binomial distribution (min $\log_2$ fold-change > 0.25, min fraction > 25%). We refer to normalized gene expression values as the number of log-normalized counts per gene relative to the total number of counts per cell. Supplemental Table S1 contains the normalized gene expression data for each cell presented in the atlas in Figure 1A.

Single cell trajectory analysis. We used the Monocle v. 2.8.0 R package (10) to infer a hierarchical organization between subpopulations of pericytes, tendon fibroblasts, and bursal fibroblasts, to organize these cells in pseudotime. We took these subpopulations from the Seurat dataset from which we reperformed SNN clustering and differential expression analysis as described above. We then selected the top 300 differentially expressed genes based on fold-change expression with a minimum adjusted p-value of 0.01 for Monocle to order the cells using the DDRTree method and reverse graph embedding. We then used branch expression analysis to identify branch-dependent differentially expressed genes.

Flow cytometry. Tendons of N=3 mice were digested to obtain a single cell suspension as described above. Cells were treated with FC Block (BD, San Diego, CA, USA) and labeled with antibodies against CD31 (FAB3628N, R&D Systems, Minneapolis, MN, USA), CD34 (551387, BD), CD45 (103115, BioLegend, San Diego, CA, USA), CD146 (134707, BioLegend), and CD206 (141711, BioLegend), as well as DAPI (Millipore Sigma). Cytometry was performed using a FACSCanto (BD) and FlowJo software (version 10, BD). Forward Scatter (FSC) and Side Scatter (SSC) plots were used to identify cell populations. FSC-A (area) and FSC-H

(height) were used to exclude doublets. Dead cells were excluded by DAPI signal. Cell populations are expressed as a percentage of total viable cells.

RNA in situ hybridization. RNA in situ hybridization (ISH) was performed using a RNAscope HiPlex kit (ACD Bio-Techne, Newark, CA, USA) and detection probes against *Col1a1*, *Spp1*, *Dpt*, *Scx*, and *Tnmd* following directions of the manufacturer. Mouse Achilles tendons (N=4) for RNA ISH were snap frozen in Tissue-Tek OCT Compound (Sakura Finetek, Torrance, CA, USA) and stored at -80°C until use. Longitudinal 10 μm sections of tendons were prepared using a cryostat, and tissue sections were digested with protease, hybridized with target probes, amplified, and labeled with fluorophores. Tissue was counterstained with DAPI to visualize nuclei, and slides were imaged in a LSM 880 confocal microscope (Zeiss, Thornwood, NY, USA).

Histology. Histology was performed as described (9, 17). Longitudinal sections of tendons were fixed with 4% paraformaldehyde, permeabilized with 0.1% Triton X-100. For antibody immunofluorescence (N=4), sections were blocked with a Mouse on Mouse Blocking Kit (Vector Labs, Burlingame, CA, USA) and incubated with primary antibodies against osteopontin (NB100-1883, Novus Biologicals, Centennial, CO, USA), dermatopontin (10537, Proteintech, Rosemont, IL, USA), and SMOC2 (sc-376104, Santa Cruz Biotechnology, Santa Cruz, CA, USA). Secondary antibodies conjugated to AlexaFluor 488, 555, and 647 (Thermo Fisher) were used to detect primary antibodies. For *Scx*:*R26*^{tdTomato} fluorescent imaging (N=4), sections were blocked with 5% goat serum. Nuclei were identified in both sets of experiments with DAPI. Slides were visualized as described above.

Acknowledgements

This work was supported by NIH grants R01-AR063649 (CLM) and R01-AG058630 (BDC), the Tow Foundation for the David Z Rosensweig Genomics Center (CLM), and by a Glenn Medical Research Foundation and American Federation for Aging Research Grant for Junior Faculty (BDC). We would like to thank Jonathan Daley, Richard Lee, and Marc Strum from the Hospital for Special Surgery for assistance in preparing the online single-cell atlas.

References

1. J. P. Gumucio, K. B. Sugg, C. L. Mendias, TGF-β Superfamily Signaling in Muscle and Tendon Adaptation to Resistance Exercise. *Exerc Sport Sci Rev* **43**, 93–99 (2015).
2. M. Grinstein, *et al.*, A distinct transition from cell growth to physiological homeostasis in the tendon. *Elife* **8**, 2716 (2019).
3. C. L. Mendias, J. P. Gumucio, K. I. Bakhurin, E. B. Lynch, S. V. Brooks, Physiological loading of tendons induces scleraxis expression in epitenon fibroblasts. *J Orthop Res* **30**, 606–612 (2012).
4. T. Stauber, U. Blache, J. G. Snedeker, Tendon tissue microdamage and the limits of intrinsic repair. *Matrix Biol* (2019) https:/doi.org/10.1016/j.matbio.2019.07.008.
5. C. Paolillo, E. Londin, P. Fortina, Single-Cell Genomics. *Clin. Chem.* **65**, 972–985 (2019).
6. Z. Yin, *et al.*, Single-cell analysis reveals a nestin+tendon stem/progenitor cell population with strong tenogenic potentiality. *Sci Adv* **2**, e1600874–e1600874 (2016).
7. J. M. Somerville, R. M. Aspden, K. E. Armour, K. J. Armour, D. M. Reid, Growth of C57BL/6 mice and the material and mechanical properties of cortical bone from the tibia. *Calcif. Tissue Int.* **74**, 469–475 (2004).
8. A. H. Huang, H. H. Lu, R. Schweitzer, Molecular regulation of tendon cell fate during development. *J Orthop Res* **33**, 800–812 (2015).
9. A. J. Schwartz, *et al.*, p38 MAPK Signaling in Postnatal Tendon Growth and Remodeling. *PLoS ONE* **10**, e0120044 (2015).
10. X. Qiu, *et al.*, Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

11. K. Bouyakdan, *et al.*, A novel role for central ACBP/DBI as a regulator of long-chain fatty acid metabolism in astrocytes. **133**, 253–265 (2015).

12. K. Howell, *et al.*, Novel Model of Tendon Regeneration Reveals Distinct Cell Mechanisms Underlying Regenerative and Fibrotic Tendon Healing. *Sci Rep* **7**, 45238 (2017).

13. L. Madisen, *et al.*, A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).

14. N. P. Disser, *et al.*, Insulin-like growth factor 1 signaling in tenocytes is required for adult tendon growth. *The FASEB Journal* **33**, 12680–12695 (2019).

15. A. J. De Micheli, *et al.*, Single-cell analysis of the muscle stem cell hierarchy identifies heterotypic communication signals involved in skeletal muscle regeneration. bioRxiv:671032 (17 June 2019).

16. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

17. K. B. Sugg, *et al.*, Postnatal tendon growth and remodeling require platelet-derived growth factor receptor signaling. *AJP - Cell Physiology* **314**, C389–C403 (2018).

Figures

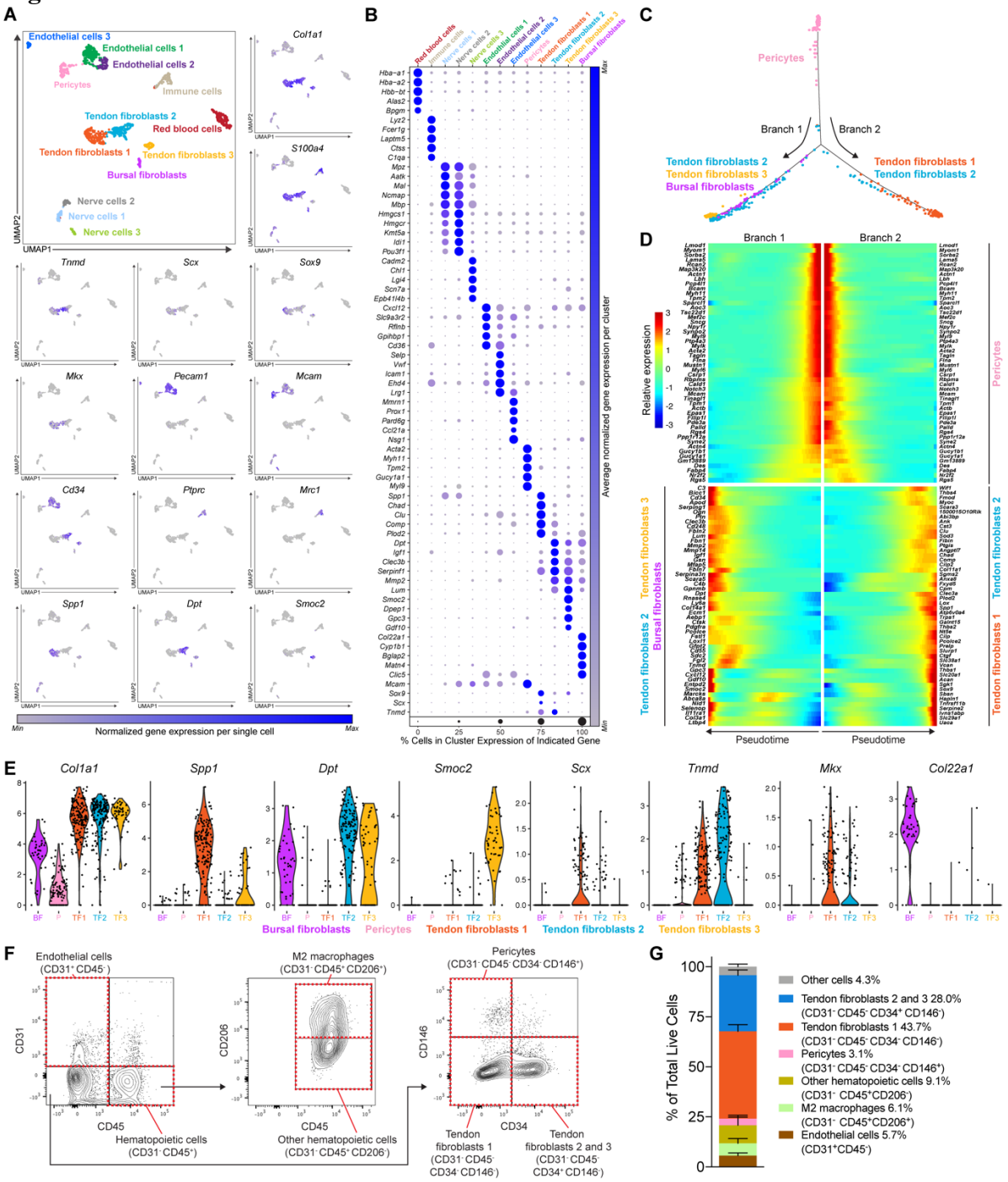

Figure 1. Single-cell transcriptional map. (A) Single-cell transcriptome atlas of Achilles tendons and selected feature plots indicating individual normalized expression for each gene. (B) Normalized expression of gene sets that are uniquely enriched in each cell type. (C) Single-cell pseudotime trajectory of pericytes into different fibroblast populations. (D) Heatmap demonstrating the top 50 genes in that are expressed across pseudotime progression from pericytes to tendon fibroblasts 2-3 and bursal fibroblasts, and to tendon fibroblasts 1 and 2. (E) Violin plots of select genes from fibroblasts and pericytes. Flow cytometry (F) contour plots and (G) cell population quantification.

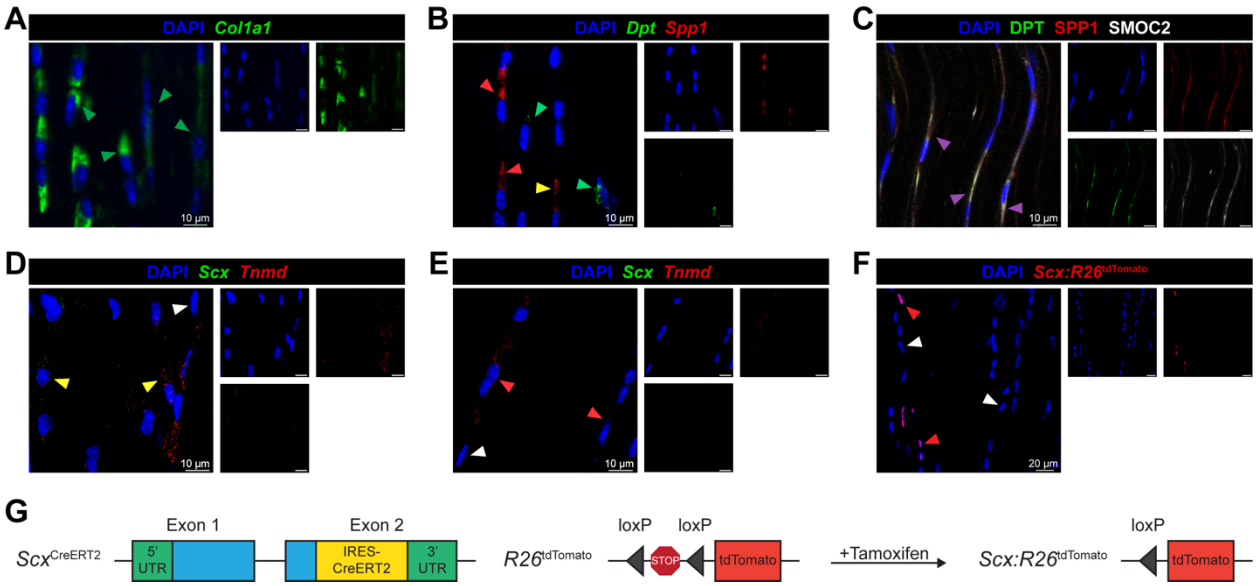

Figure 2. Histology. (A,B,D,E) RNA *in situ* hybridization and (C,F) protein immunofluorescence of Achilles tendons. (A) *Col1a1* RNA, green; green arrows exemplify *Col1a1*$^{+}$ fibroblasts. (B) *Dpt* RNA, green; *Spp1* RNA, red; green arrows exemplify *Spp1*$^{-}$*Dpt*$^{+}$ fibroblasts; red arrows exemplify *Spp1*$^{+}$*Dpt*$^{-}$ fibroblasts; yellow arrows exemplify occasional *Spp1*$^{+}$*Dpt*$^{+}$ fibroblasts. (C) Dermatopontin (DPT, green), osteopontin (SPP1, red), and SMOC2 (white) protein; magenta arrows exemplify overlap between DPT, SPP1, and SMOC2 protein. (D-E) *Scx* RNA (green); *Tnmd* RNA (red); red arrows exemplify *Scx*$^{-}$*Tnmd*$^{+}$ fibroblasts; yellow arrows exemplify *Scx*$^{+}$*Tnmd*$^{+}$ fibroblasts; white arrows exemplify *Scx*$^{-}$*Tnmd*$^{-}$ fibroblasts. (F) tdTomato expression in scleraxis expressing cells (red arrows) compared to non-scleraxis expressing cells (white arrows). Nuclei visualized with DAPI. (G) Overview of genetics of scleraxis lineage tracing mice.