Title

Single-nucleus transcriptional and chromatin accessibility analyses of maturing mouse Achilles tendon uncover the molecular landscape of tendon stem/progenitor cells

Authors

Hiroki Tsutsumi[1#], Tomoki Chiba[1#], Yuta Fujii[1,2], Takahide Matsushima[1], Tsuyoshi Kimura[3,4], Akinori Kanai[5], Akio Kishida[3], Yutaka Suzuki[5] and Hiroshi Asahara[1,2*]

Affiliations

1 Department of Systems BioMedicine, Graduate School of Medical and Dental Sciences, Institute of Science Tokyo, Tokyo, Japan

2 Department of Molecular and Cellular Biology, The Scripps Research Institute, La Jolla, USA

3 Laboratory for Biomaterials and Bioengineering, Institute of Integrated Research, Institute of Science Tokyo, Tokyo, Japan

4 Department of Biomedical Engineering, Faculty of Life Science, Toyo University, Saitama, Japan

5 Laboratory of Systems Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, the University of Tokyo, Chiba, Japan

#: These authors contributed equally to this work.

Contact info

*Correspondence: asahara@scripps.edu

Abstract

Tendons and ligaments are crucial connective tissues linking bones and muscles, yet achieving full functional recovery after injury remains challenging. We investigated the characteristics of tendon stem/progenitor cells (TSPCs) by focusing on the declining tendon repair capacity with growth. Using single-cell RNA sequencing on Achilles tendon cells from 2- and 6-week-old mice, we identified *Cd55* and *Cd248* as novel surface antigen markers for TSPCs. Combining single-cell RNA sequencing with single-nucleus RNA and ATAC sequencing analyses revealed that *Cd55* and *Cd248* positive fractions in tendon tissue represent TSPCs, as confirmed by their expression of established TSPC markers, with this population decreasing at 6 weeks. We also identified candidate upstream transcription

factors regulating these fractions. Functional analyses of isolated CD55/CD248 positive cells demonstrated high clonogenic potential and tendon differentiation capacity, forming functional tendon-like tissue *in vitro*. This study establishes CD55 and CD248 as novel TSPC surface antigens, potentially advancing tendon regenerative medicine and contributing to the development of new treatment strategies for tendon and ligament injuries.

Introduction

Tendons are connective tissues involved in the contractile movement of muscle to bone. Tendons are rich in extracellular matrix components, such as type 1 collagen and proteoglycans, which both have elastic and viscous properties to withstand overload (Andarawis-Puri et al., 2015; Asahara et al., 2017; Connizzo et al., 2013).

Tendons have a hierarchical structure. Collagen fibrils aggregate to form fibers, which then bundle into fascicles. Each fascicle is surrounded by a thin connective tissue layer called the endotenon. Fascicles are further enclosed by the epitenon, and in some cases, an additional outer layer called the paratenon. The number of fascicles and the presence of a paratenon can vary among different tendons and species (Jozsa et al., 1991; Walia et al., 2019).

Tendon and ligament injuries potentially account for the majority of musculoskeletal disorders and can lead to arthritis and spondylitis. (Gracey et al., 2020). The most popular treatment approaches are surgery and conservation (Steinmann et al., 2020). However, full functional recovery is often not achieved due to scarring and fibrosis in the injured area (Guilak et al., 2014; Thomopoulos S et al., 2015). Furthermore, the risk of postoperative rupture is high (Leong et al., 2020; Loiacono et al., 2019). In this context, tendon regeneration focused on tendon stem/progenitor cells (TSPCs) is attracting attention (Leong et al., 2016). Tendons have few cellular components, including tendon cells (tenocytes) and TSPCs (Tang et al., 2016; Bi et al., 2007). Tenocytes are responsible for tendon homeostasis, while TSPCs self-renew and differentiate into tenocytes. Thus, investigating TSPCs biology is important for understanding tendon regeneration and homeostasis.

Murine TSPCs were first reported in 2007 by Bi Y et al. (Bi et al., 2007). They demonstrated that TSPCs possess self-renewal capacity, colony-forming ability, and multi-differentiation potential *in vitro*. Furthermore, murine TSPCs were reported to simultaneously express stem cell markers such as *Cd44* and *Stem cells antigen-1* (*Ly6a*), as well as tendon-related genes like *Scleraxis* (*Scx)* and Collagen type I alpha 1 chain (*Col1a1*), and were identified as a subset of tendon cells within the tendon fascicle.

Subsequent research on murine TSPC localization has yielded diverse perspectives. Harvey et al. reported that *Tppp3*/*Pdgfra*-positive cells in the epitenon are induced upon tendon injury and contribute to repair (Harvey et al., 2019). Additionally, Yin et al. and

Tempfer et al. demonstrated the presence of TSPCs around blood vessels (Yin et al., 2016; Tempfer et al., 2009).

Markers used to characterize murine TSPCs include CD73, CD105, and CD90, which are known criteria for mesenchymal stem cells (MSCs) (Lui et al., 2015). Additionally, TSPCs have been reported to express *POU class 5 homeobox 1* (*Oct-4*), *Nanog*, *Nucleostemin*, stage-specific embryonic anitigen-4 *(SSEA-4)*, *c-Myc*, *SRY-box transcription factor (Sox*2), *Fucosyltransferase 4* (*Fut-4*), and other genes (Zhang et al., 2016). Expression of CD146 and CD44 has been confirmed as well (Ruzzini et al., 2014). The *Tppp3*/*Pdgfra*-positive TSPCs reported by Harvey et al. highly express *Cd34*, which is generally considered to have low expression in MSCs (Harvey et al., 2019; Tachibana, et al., 2022). *Cd34* is known to be highly expressed in mouse embryonic limb buds at E14.5 compared to E11.5 (Havis et al., 2014), suggesting that *Cd34* positive cells might reflect progenitor cells that constitute the limb bud, including tendons. However, these markers are not specific to TSPCs, thus, a definitive method to distinguish TSPCs from mature tendons *in vivo* has not been established yet (Li et al., 2019; Wang et al., 2008; Chen et al., 2012; Cho et al., 2018; Fang et al., 2022).

Regarding tendon regenerative capacity, Howell et al. reported an interesting observation. Tendons in juvenile mice can regenerate functional tissue after injury, but this ability is lost in mature mice, resulting in scar tissue formation (Howell et al., 2017). This finding suggests the possibility of abundant TSPCs in juvenile mouse tendons.

Given this background, we hypothesized that evaluating tendon tissue heterogeneity at the single-cell level using juvenile mouse tendons would enable a more detailed characterization of TSPCs. This approach is expected to advance the identification and characterization of TSPCs, which has been challenging with conventional methods.

Therefore, we performed scRNA-seq using cells collected from 2- and 6-week-old mouse Achilles tendons and investigated clusters that co-express known TSPCs markers, such as *Tppp3*, *Pdgfra*, and *Ly6a*. Then, we analyzed the expression of surface antigens in these clusters and identified *Cd55* and *Cd248* as novel candidate surface antigens for murine TSPCs.

Furthermore, snATAC-seq and snRNA-seq were performed simultaneously using cells collected from Achilles tendons to evaluate the validity of *Cd55* and *Cd248* as surface antigens for TSPCs and to identify the landscape of transcription factors (TFs) involved in tendon maturation. We sorted mouse Achilles tendon cells based on CD55 and CD248 and confirmed their phenotypes, demonstrating high clonogenicity and highly efficient induction into tendon cells. These results suggested that CD55 and CD248 are novel surface antigens of murine TSPCs and may be useful for understanding the process of tendon maturation and

for applications in cell therapy.

Material and Methods

Single cell isolation

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. 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. All animals were purchased from Sankyo Lab Service (Tokyo, Japan). Achilles tendons were digested to obtain a single-cell suspension. In total, 40 or 60 (2-week and 6-week, respectively) mouse Achilles tendons were processed together as a single sample. After washing with 1× phosphate-buffered saline (PBS) several times and cutting 1-mm-wide sections using scissors, tendons were digested for 1 h at 37°C with continuous shaking at 1,200 rpm and 37°C in a dissociation solution consisting of 30 mg/mL collagenase (Wako, Osaka, Japan). After digestion, the single-cell suspension was filtered for debris using a 40 μm cell strainer and washed twice with 1× PBS. The samples were centrifuged for 15 min at 1,500 rpm and 4°C. Dead cells were removed using a Dead Cell Removal Kit (Miltenyi Biotec, Bergisch Gladbach, Germany). Cell viability was 76% (2-week) and 73% (6-week), respectively. The experiment involving mice was approved by Animal Experimental Committee of the Institute of Science Tokyo.

scRNA-seq library construction and sequencing

Cells isolated from Achilles tendons of both 2-week-old and 6-week-old mice were processed separately. Cells were resuspended in PBS with 1% fetal bovine serum (FBS) at a concentration of 10,000 cells per μL. Then, 10,000 cells per sample were loaded on a Chromium Controller (10x Genomics, Pleasanton, CA, USA) for single cell capture. Libraries were prepared using Single Cell 3′ Library & Gel Bead Kit v3 (10x Genomics) following the manufacturer's instructions. A single-cell emulsion (Gel Bead-In-EMulsions, GEMs) was created by making barcoded cDNA unique to each individual emulsion. A recovery agent was added to break GEM and cDNA was then amplified. A library was produced via end repair, dA-tailing, adapter ligation, post-ligation cleanup with SPRIselect, and sample index PCR. The quality and concentration of the amplified cDNA were evaluated using the Bioanalyzer (Agilent 2100) on a High Sensitivity DNA chip (Agilent, #5065-4401; Santa Clara, CA, USA). Sequencing was performed using the HiSeq X system (Illumina, San Diego, CA, USA) to generate 28/90 bp paired-end reads.

Cell dissociation, nuclei isolation, and snRNA-seq and snATAC-seq library construction and sequencing

After obtaining cells from the mouse Achilles tendons of 2-week-old and 6-week-old mice (processed separately for each age group), nuclei were isolated following the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle (10XGenomics) and all of the buffers were made according to the manufacturer's instructions. Briefly, after centrifugation, the supernatant was discarded, and the cells were re-suspended in 1 mL of 1% FBS in PBS. Approximately 500,000 cells were transferred to a new tube for further lysis. To remove the supernatant, cells were centrifuged again for 5 min at 300 rcf and 4°C and resuspended in 100 mL of chilled lysis buffer. Cells were lysed for 9 min on ice and 1 mL of wash buffer was added to stop the reaction. Then, the suspension was filtered through a 40 μm cell strainer, cells were centrifuged and resuspended in 66.2 mL of chilled Diluted Nuclei Buffer aiming to target 7000 nuclei. The final concentration of nuclei was determined, followed by transposition, GEM generation, barcoding, and library construction according to the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression (10XGenomics). Libraries were sequenced with the parameters recommended by the manufacturer, using the NovaSeq 6000 (Illumina) to generate 28/90 bp paired-end reads for gene expression and 50/49 bp paired-end reads for ATAC sequencing.

scRNA-seq and snRNA-seq analyses

Sequencing reads were processed with the Cellranger_arc (10X Genomics, v2.0.0) using the mouse reference mm10. We performed separate analyses for scRNA-seq data (collected from both 2-week and 6-week samples) and snRNA-seq data (collected from both 2-week and 6-week samples) to enable comprehensive characterization of transcriptional landscapes. From the gene expression matrix, downstream analyses were carried out using R. Quality control, filtering, data clustering and visualization, and a differential expression analysis were carried out using Seurat (Butler et al., 2018). For each dataset, cells with unique feature counts over 2,500 or less than 200 and >5% mitochondrial counts were filtered. Then, heterotypic doublets (assuming 5% of barcodes represent doublets) were removed using DoubletFinder (McGinnis et al., 2019).

Unsupervised shared nearest neighbor (SNN) clustering was performed with varying resolution and the results were visualized using uniform manifold approximation and projection (UMAP) (Becht et al., 2019). Differentially expressed genes (DEGs) among each cell cluster were determined using the FindAllMarkers function in Seurat. The criteria of DEGs as a marker gene for the cluster was logFC > 0.25, adjusted p < 0.05, expression in

181 >25% of cells. We then annotate each cluster based on DEGs (supplementary table1-15).
182 To ensure consistency between our different sequencing approaches, we compared cell
183 clusters identified in the scRNA-seq analysis with those from the snRNA-seq analysis
184 (Supplemental Figure 5).
185
186 **Pseudotime analysis**
187 Monocle3 (Trapnell et al., 2014) was used to convert the snRNA-seq dataset into a cell
188 dataset object (CDS), preprocess data, correct for batch effects, embed with dimensional
189 reduction, and perform pseudotemporal ordering. Cicero (Pliner et al., 2018) was used to
190 generate pseudo-temporal trajectories for the snATAC-seq dataset.
191
192 **Cell-cell communication analysis**
193 Intercellular communication networks were quantitatively inferred and analyzed using
194 scRNA-seq data. The R package CellChat (Jin et al., 2021) was used to visualize the
195 interactions among different cell groups. Two hundred twenty-nine signaling pathway
196 families were grouped as a library to analyze cell–cell communication.
197
198 **GO analysis**
199 The R package ClusterProfiler (Yu et al., 2012) was used to perform a gene enrichment
200 analysis. The p-values were corrected by the Benjamini & Hochberg method.
201
202 **Motif enrichment analysis (ChiP seeker)**
203 Genomic regions containing snATAC-seq peaks were annotated using ChIPSeeker (Yu et
204 al., 2015) and the UCSC database on mouse (mm10).
205
206 **snATAC-seq analysis**
207 The Cell Ranger ATAC pipeline (1.2.0) (Satpathy et al., 2019) was used to preprocess the
208 data resulting from sequencing. First, Tn5 sites were mapped to the mouse reference
209 transcriptome mm10, and duplicate reads and background cells were removed. This
210 returned barcoded fragment files, which were loaded into Signac (Stuart et al., 2021) for
211 downstream analyses using the standard Signac/Seurat pipeline. Macs2 (Zhang et al.,
212 2008) was run on the fragment files to call peaks using the Signac 'CallPeaks' function.
213 Fragments were mapped to the Macs2 called peaks and assigned to cells using the Signac
214 'FeatureMatrix' function. Nucleosome signal strength and transcription start site (TSS)
215 enrichment for each cell were calculated using the Signac 'NucleosomeSignal' and
216 'TSSEnrichment' functions, respectively. Outliers in the QC metric categories were removed

as per Signac's standard processing guidelines. Latent semantic indexing (LSI), a form of dimensional reduction, was performed using the Signac 'RunTFIDF' and 'RunSVD' functions. The UMAP hyperparameters were varied to produce consistent object shapes (using R). Once hyperparameters were chosen, the Signac/Seurat's 'RunUMAP' function was run on the LSI dimensions chosen earlier for UMAP embedding. The Signac/Seurat 'FindNeighbors' function was run using the same LSI dimensions used for UMAP to compute the nearest neighbor graph. Signac/Seurat 'FindClusters' was then run at varying resolutions. A gene activity matrix was constructed by counting ATAC peaks within the gene body and 2 kb upstream of the transcriptional start site for protein-coding genes annotated in the Ensembl database. The gene activity matrix was log-normalized prior to label transfer with the aggregated snRNA-seq Seurat object using a canonical correlation analysis. The aggregated snATAC-seq object was filtered using a 97% confidence threshold for cell-type assignment following label transfer to remove heterotypic doublets. The filtered snATAC-seq object was reprocessed with TFIDF, SVD, and batch effect correction followed by clustering and annotation based on lineage- specific gene activity. Differential chromatin accessibility between cell types was assessed with the Signac FindMarkers function for peaks detected in at least 20% of cells using a likelihood ratio test and a log-fold-change threshold of 0.25. Bonferroni-adjusted p-values were used to determine significance at an FDR < 0.05.

Integration of snRNA-seq and snATAC-seq data

snRNA-seq and snATAC-seq data were integrated using the cluster-label transfer procedure as implemented in Signac and Seurat. Each snRNA-seq sample. as clustered individually and its cluster labels were projected onto the matching, individually clustered snATAC-seq sample or vice versa. Anchors were identified for condition-matched snRNA- and snATAC-seq samples using the FindTransferAnchors function and a canonical correlation analysis (CCA) was performed using the snRNA expression values and the snATAC-imputed gene expression values. The anchors were used to transfer cluster-label identifiers between the two data types using the TransferData function. Each cell in the query was assigned the cluster label with the highest prediction score, and label transfer was considered successful for query cells with prediction scores above 0.3.

SCENIC analysis

To identify TFs and characterize cell states, a cis-regulatory analysis was performed using the R package SCENIC (Lake et al., 2018), which infers gene regulatory networks based on co-expression and DNA motif analyses. The network activity was then analyzed for each

253 cell to identify recurrent cellular states. In short, TFs were identified using GENIE3 and
254 compiled into modules (regulons), which were subsequently subjected to cis-regulatory
255 motif analysis using RcisTarget with two gene-motif rankings: 10 kb around the TSS and
256 500 bp upstream. Regulon activity in every cell was then scored using AUCell.
257
258 **Immunohistochemistry**
259 Immunohistochemistry was performed as previously described (Tsutsumi et al., 2022).
260 Briefly, tissue samples were fixed in 4% paraformaldehyde overnight at 4°C and embedded
261 in paraffin. Sections of 10 μm in thickness were deparaffinized and activated with citric acid
262 buffer (10 mM sodium citrate, 1 mM EDTA; pH 6.0) at 80 °C for 60 min in a decloaking
263 chamber NxGen (BIOCARE Medical, CA, USA) or with 0.1% trypsin/PBS at 37 °C for
264 30 min. After they were blocked with Blocking One solution (Nacalai tesque, Kyoto, Japan)
265 for 60 min, they were incubated with rabbit anti-CD55 antibody (1:100; A13918, ABclonal,
266 Woburn, MA, USA) or rat anti-CD248 (1:100; MAB7535, R&D Systems, Minneapolis, MN,
267 USA) overnight at 4 °C. Following this step, they were incubated with Alexa Fluor plus 488
268 goat anti-rabbit antibody (1:1000; A32731, Thermo Fisher Scientific , Waltham, MA, USA)
269 or Alexa Fluor plus 488 donkey anti-rat antibody (1:1000; A21208, Thermo Fisher Scientific)
270 for 60 min. Hoechst 33342 (100ng/ml, Thermo Fisher Scientific) was added during this
271 incubation. The sections were then mounted with ProLong Glass Antifade Mountant
272 (P36980, Thermo Fisher Scientific).
273
274 **Fluorescence-activated cell sorting (FACS)**
275 Cells harvested from the Achilles tendon were incubated for 60 min on ice with APC-CD55
276 (#131812; BioLegend, San Diego, CA, USA) and a primary antibody against CD248 (#LS-
277 B2712, LSBio) using a FACS buffer (PBS containing 1 [v/v%] FBS). After washing, cells
278 were reacted with fluorescent-conjugated rabbit secondary antibody. Cell sorting was
279 performed using MoFlo XDP FACS (Beckman Coulter, Brea, CA, USA). DAPI was used as
280 the live/dead discriminator to the gate. CD55/CD248 dual positive and negative cell cluster
281 were sorted.
282
283 **Primary cell culture**
284 Primary TSPCs were isolated from the Achilles tendons of 2-week-old mice with
285 collagenase digestion described above. Single-cell suspensions were cultured in the
286 culture medium (MEMα + 20% FBS + 1% penicillin-streptomycin + 1 [v/v%] 100× non-
287 essential amino acid solution [NEAA; Gibco], 1 [v/v%] 100× GlutaMAX [Gibco]). At 80%–

90% confluence, cells were trypsinized, centrifuged, resuspended in culture medium as passage 1 cells, and incubated in 5% CO_2 at 37°C, with fresh medium every 2–3 days.

Colony formation assay

For the colony formation assay, single-cell suspensions of TSPCs (1000 cells/well) were seeded and incubated in six-well plates for 12–14 days in the growth medium and fixed with 4% paraformaldehyde (PFA) (Sigma-Aldrich, St. Louis, MO, USA). Then, 0.1% crystal violet solution (Wako) was used to stain the cells. Colonies of >30–50 cells were defined as a single colony unit, and the number of clusters was counted using the ImageJ package ColonyArea (Guzmán et al., 2014)

RT-PCR

Total RNA was purified using TRIzol reagent (Invitrogen, Grand Island, NY, USA). Reverse transcription of mRNA was carried out using the PrimeScript RT Reagent Kit (Takara, Shiga, Japan). Quantitative PCR (qRT-PCR) was performed on cDNA with the Thunderbird SYBR mix (Toyobo, Osaka, Japan). B2m was used as a reference gene, and relative gene expression levels were calculated through the ΔCT method.

Tenocytes, cartilage, and osteocyte differentiation

For the differentiation experiment, TSPCs were cultured in 6-well plates (50,000 cells/well). Osteogenic, chondrogenic, and tenogenic differentiation were induced using a corresponding differentiation medium. The osteogenic differentiation medium contained a culture medium supplemented with 10 nM dexamethasone (Sigma-Aldrich), 5 mM β-glycerophosphate (APEXBIO), and 0.05 mM L-ascorbic acid 2-phosphate (Sigma-Aldrich). The chondrogenic differentiation medium contained a culture medium supplemented with 100 nM dexamethasone and 10 ng/mL BMP2 (Sigma-Aldrich). The tenogenic differentiation medium contained a culture medium supplemented with 10 ng/mL TGF-β1 (Peprotech, Rocky Hill, NJ, USA), 10 ng/mL GDF-5 (R&D Systems, Minneapolis, MN, USA), and 0.05 mM L-ascorbic acid 2-phosphate. After a 2 week induction period, cells were harvested for RT-PCR.

Bio-cultured tendon (Bio-tendon)

Generation of Bio-tendon has reported previously (Kataoka et al., 2020; Tsutsumi et al., 2022) Briefly, sorted cells were embedded in a 3D-culture cocktail, which is consistent with 2 mg/mL Cellmatrix (Type I-A, Nitta Gelatin Inc., Osaka, Japan), pro-survival cocktail (final concentrations: 100 nM Bcl-Xl BH4 4-23 (Merck Millipore, Burlington, MA, USA), 100 μM

carbobenzoxy-valyl-alanyl-aspartyl-[*O*-methyl]-fluoromethylketone (Z-VAD-FMK) (Promega, Madison, WI, USA)) in culture medium. The 3D chamber was coated with 1% gelatin and incubated at 37°C and 5% CO_2 for 30 min. After washing with 1× PBS 3 times, 1.0×10^6 cells were transferred to a 3D-culture cocktail mixture and incubated at 37°C and 5% CO_2 for 60 min for gelation. Following gelation, tenogenic differentiation medium was added to the chamber. Following 24 h of incubation, the chamber was placed within a cell stretching system (Shellpa Pro, Menicon Life Science, Aichi, Japan). Cyclic mechanical stretch was applied for 1 week, with gradual increase in stretch load: 1% on day 1, 2% on day 2, 3% on day 3, 4% on day 4, and 5% from days 5 to 7. The stretching cycle was programed at 0.25 Hz for 18 h/day, followed by resting for 6 h/day at 37°C and 5% CO_2. Daily medium changes were also required.

Decellularization by HHP and chemical treatment

The cultured bio-tendon was placed into a plastic pack filled with saline and sealed to prevent implosion and leakage during the procedure. The pack was then pressurized at 1,000 MPa at 30°C for 10 min using an HHP machine (Dr. CHEF; Kobe Steel, Hyogo, Japan). After pressurization, the bio-tendon was washed thrice with 30% ethanol (EtOH) by continuous shaking for 5 min at each step. Finally, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)/*N*-hydroxysuccinimide (NHS)-based cross-linking was performed by adding 70 mM EDC (Wako) and 70 mM NHS (Wako) with 30% EtOH for 24 h at 4°C. The processed bio-tendon was incubated in 1× PBS at 4°C until the experiment.

Electron microscopy

Bio-tendon tissues were dissected and fixed overnight in 2.5% glutaraldehyde in 0.1 M phosphate buffer (PB). For transmission electron microscopy (TEM), the specimens (n = 3) were rinsed with 0.1 M PB, post-fixed in 1% osmium buffered with 0.1 M PB for 2 h, and dehydrated in a graded ethanol series. The specimens were then embedded in Epon 812, sectioned into ultrathin sections (70 nm), mounted on copper grids, and double-stained with uranyl acetate and lead citrate. TEM observation was conducted JEM-1400Flash (JEOL, Tokyo, Japan). For scanning electron microscopy (SEM), the specimens (n = 2) were dried in a critical-point dryer (HCP-2; Hitachi, Tokyo, Japan) with liquid CO_2 and coated with platinum. The specimens were subsequently observed under SEM (JSM-7900F/JED-2300; JEOL). Fiber orientation within the microscopic sections was analyze using the OrientationJ image processing tool, a plugin for ImageJ.

Stretch test

The mechanical properties were measured using a creep meter (RE-3305S; Yamaden, Tokyo, Japan). After measuring the initial length (mm), diameter (mm), and thickness (mm) using a micrometer, the samples were fixed with two grips, which were pulled at a constant speed of 0.05 mm/s until failure, and the tensile strength (N) and failure strain (mm) were measured. The cross-sectional area (mm^2) was calculated using the initial diameter and thickness. The stiffness was manually determined from the slope in the linear region of the failure-stress curve.

The tensile strength (MPa), failure strain (%), stiffness, and Young's modulus were calculated using the following formulas:

Tensile strength (MPa) = tensile strength (N)/cross-sectional area (mm^2).

Failure strain (%) = failure strain (mm)/initial length (mm).

Stiffness = Stress (N)/Strain (mm)

Young's modulus = stiffness × initial length (mm)/cross-sectional area (mm^2)

Statistical analysis

All values are presented as means ± SEM. Statistically significant differences were assessed using unpaired two-tailed Student's *t*-tests and one-way analysis of variance (ANOVA) with Tukey's post hoc tests. Statistical significance was set at $p < 0.05$.

Results

Identification of Novel Surface Antigens for Tendon Stem/Progenitor Cells (TSPCs)

In mice, the healing capacity for tendon injuries is high up to 2 weeks of age, but decreases as the musculoskeletal system matures at 6 weeks (Howell et al., 2017). To investigate whether this phenomenon is due to changes in the cellular population of tendon tissue, including fluctuations in progenitor cells, we performed single-cell RNA sequencing on mouse Achilles tendons at these two time points.

Achilles tendons were harvested from mice at 2 and 6 weeks of age. Following collagenase digestion and dead cell removal, we employed massively parallel, droplet-enabled scRNA-seq analysis (10X Genomics Chromium). Data processing was conducted using the Cell Ranger pipeline (10X Genomics). We analyzed 10,314 cells from 2-week-old mice (median 3,167 genes/cell and 41,611 mean reads/cell) and 6,513 cells from 6-week-old mice (median 732 genes/cell and 60,386 mean reads/cell). After doublet removal using Doubletfinder, we merged the datasets and performed unbiased clustering using Seurat, identifying a total of 15 clusters (Fig. 1A). The top differentially expressed genes (DEGs) for

each cluster, based on log2 fold change and statistical significance, are summarized in Fig. 1B.

We identified two clusters (0 and 11) with high expression of *Mkx* and *Scx*, transcription factors involved in tendon development and growth. Cluster 0 showed higher expression of ECM-related genes such as *Col1a1* and *Fmod* compared to cluster 11, suggesting these were tenocytes (TC). Cluster 11, with high *Periostin* (*Postn)* expression (Jacobson, et al., 2020), was classified as myotendinous junction (MTJ). Other clusters were characterized as follows: cartilage (CA, high *SRY-box transcription factor 9* (*Sox9*) expression), lymphocytes (LC1, LC2, high *Lysozyme 2* (*Lyz2*) and *Complement component 1, q subcomponent, alpha polypeptide* (*C1qa*)), endothelial cells (EC, high *Platelet and endothelial cell adhesion molecule 1* (*Pecam1*)), red blood cells (RBC, high *Hemoglobin alpha, adult chain 1* (*Hba-a1*)), smooth muscle cells (SM, high *Myosin light chain 9* (*Myl9*)), proliferating cells (PC, high *Mki67* and *Stathmin 1* (*Stmn1*)), Schwann cells (SW, high *Myelin basic protein* (*Mbp*)), vascular endothelial cells (VEC, high *Multimerin 1* (*Mmrn1*)), and macrophages (MC, high *Protein tyrosine phosphatase receptor type C* (*Ptprc*)). In cluster3, *Col1a1* and *Fmod* were expressed; however, *Gm42418* and *AY036118* were highly expressed. These long non-coding RNAs are related to RN45s and therefore cluster3 represents ribosomal contamination (Isola et al., 2024). It is also speculated that the cartilage cluster (CA) reflects enthesis, with expression of *Scx* as well as *Sox9* (Zhan et al., 2023). DEG s of each cluster was summarized in supplementary table1-15.

To identify progenitor populations within these clusters, we analyzed expression patterns of previously reported markers *Tppp3* and *Pdgfra* (Harvey et al., 2019; Tachibana, et al., 2022), along with the known stem/progenitor cell marker *Ly6a* (Holmes et al., 2007; Sung et al., 2008; Hittinger et al., 2013; Sidney et al., 2014; Fang et al., 2022) (Supplemental Figure 1B). We identified subclusters within clusters 1 and 4 showing high expression of these genes, which we defined as SP1 and SP2. SP2 exhibited the highest expression of these genes, suggesting it had the strongest progenitor characteristics. It has been reported (Harvey et al., 2019) that the *Tppp3*-positive population is localized to peritenon, and SP clusters might reflect peritenon as well.

SP2 also showed strong expression of genes associated with early tendon development in mouse embryos, as identified by RNA-seq of mouse limb tendon cells at E14.5 (Havis et al., 2014) and the Eurexpress mouse embryo transcriptome atlas database (Diez-Roux et al., 2011) (Supplemental Figure 2A). Gene Ontology analysis revealed upregulation of tendon development-related pathways in SP2, including TGFβ production and collagen-containing extracellular matrix (Supplemental Figure 2B). Evaluation of cell numbers in

each cluster showed a decrease in SP2, TC, and MTJ at 6 weeks (Fig. 1C, Supplemental Figure 1A).

We then examined DEGs in SP2 compared to SP1 (Fig. 1D). Whole list of DEGs comparing SP2 and SP1 was summarized in supplementary table16. Among these, we identified *Cd34*, *Cd55*, and *Cd248* as genes encoding surface antigens. *Cd34* is known to be highly expressed in mouse embryonic limb bud E14.5 compared to E11.5 (Havis et al., 2014), suggesting that *Cd34* positive cluster might reflect a progenitor cell that consist of the limb bud, including tendons. However, *Cd55* and *Cd248* have not been discussed with this context. Expression of *Cd55* and *Cd248* was localized to areas corresponding to SP1 and SP2 (Fig. 1E). We evaluated the correlation coefficients between the expression of these genes and previously suggested TSPC markers (*Cd73*, *Cd90*, *Cd105*, *Cd44*, *Cd146*) as well as *Tppp3*, *Pdgfra*, and *Ly6a* at 2 weeks of age. *Cd34*, *Cd55*, and *Cd248* showed high correlation, suggesting these genes as new candidate surface antigens for TSPCs (Fig. 1F).

Comparing gene expression between 2 and 6 weeks, we observed a decrease in *Mkx* and *Scx* expression, consistent with previous reports (Grinstein et al., 2019). Furthermore, the expression of our candidate tendon progenitor cell markers *Cd34*, *Cd55*, and *Cd248* all decreased at 6 weeks. Inflammatory cells, including macrophages, increased at 6 weeks (Fig. 1G). Gene expression analysis revealed a shift from M2 macrophage-related genes (*Mannose receptor C-type 1* (*Mrc1*) and *Cathepsin C* (*Ctsc*)) at 2 weeks to M1 macrophage-related genes (*Chemokine (C-C motif) ligand 6* (*Ccl6*) and *Chemokine (C-C motif) ligand 9* (*Ccl9*)) at 6 weeks. Comparative analysis between 2-week and 6-week tendons was summarized as a heatmap in supplemental Figure 3.

Analysis of cell-cell communication changes in postnatal tendon and surrounding tissues (Lui et al., 2019) using CellChat (Jin et al., 2021) showed a significant decrease in overall interaction at 6 weeks compared to 2 weeks. These results demonstrate that the cellular profile of tendon tissue changes dramatically at 6 weeks, with decreased expression of stem progenitor markers and reduced interaction with surrounding tissues.

Single Nucleus RNA + ATAC Analysis of 2-Week-Old Mouse Achilles Tendon Cells

Gene regulation is mediated by the binding of transcription factors to cis-regulatory elements proximal to the gene. Consequently, epigenetic changes such as chromatin accessibility play a crucial role in gene expression (Miller et al., 2013). Moreover, transcription factors are typically expressed at low levels, potentially leading to false negatives in scRNA-seq due to detection limits. Chromatin accessibility changes often precede gene expression changes, potentially allowing us to predict transcriptional

466 changes (Ranzoni et al., 2020). Therefore, we performed simultaneous ATAC-seq and
467 RNA-seq at the single-cell level to identify gene expression changes and their associated
468 epigenetic alterations, aiming to elucidate the post-natal growth mechanisms of tendon
469 tissue and evaluate the validity of *Cd55* and *Cd248* as markers.
470 We extracted nuclei from cells isolated from 2-week-old mouse Achilles tendons and
471 conducted droplet-enabled multi-omics analysis (10X Genomics Chromium), performing
472 simultaneous snATAC-seq and snRNA-seq as a reference. We first analyzed the snATAC-
473 seq data. After mapping reads to the genome and calling peaks, we annotated the location
474 of peaks in terms of genomic features. The peaks were associated with promoters, introns,
475 exons, and intergenic regions.
476 We evaluated 6,571 cells (median high-quality fragments per cell: 17,408) and clustered
477 them using latent semantic indexing and Uniform Manifold Approximation and Projection
478 (UMAP) with the R package Signac (Stuart et al., 2021) (Figure 2A). Given the limited
479 knowledge of cell-specific chromatin accessibility, we assessed gene activity by computing
480 counts per cell within the gene body and promoter of protein-coding genes. Using this gene
481 activity data, we performed unsupervised clustering and identified 17 clusters (ground-truth
482 annotation). Clusters expressing tendon-related genes such as *Mkx* and *Scx* were
483 identified as clusters 1 and 6, defined as A1: TC1 and A6: TC2, respectively. Clusters 0, 14,
484 and 15 were identified as expressing stem progenitor markers including *Cd34*, *Ly6a*,
485 *Pdgfra*, *Cd55*, and *Cd248*, and were defined as A2-0: SP1, A2-14: SP2, and A2-15: SP3,
486 respectively. The characteristic gene activities and annotations for each cluster are
487 summarized in Fig. 2B.
488 Next, we analyzed the snRNA-seq data. We evaluated 10,314 cells (median 3,167
489 genes/cell and 41,611 mean reads/cell). After doublet removal, we performed unbiased
490 clustering using Seurat and identified 17 clusters (Supplemental Figure 4A). Clusters
491 expressing tendon-related genes *Mkx* and *Scx* were clusters 1 and 13, defined as R2-1:
492 TC_1 and R2-13: TC_2, respectively. Clusters 2, 3, and 7 were identified as expressing
493 stem/progenitor-related genes and defined as R2-2: SP_1, R2-3: SP_2, and R2-7: SP_3,
494 respectively. To validate the consistency of our approaches, we conducted a
495 comprehensive comparison between the cell clusters identified in our scRNA-seq analysis
496 (which included both 2-week and 6-week samples) and our snRNA-seq analysis. We
497 confirmed that all clusters identified in the scRNA-seq data could be similarly annotated in
498 the snRNA-seq data, demonstrating strong concordance between these complementary
499 approaches (Supplemental Figure 5). This validation supports the reliability of our cell type
500 identification across different single-cell methodologies. Interestingly, R2-5: The MTJ cluster
501 expressed not only *Tnc* but also *Sox9*, consistent with previous reports (Nagakura et al.,

2020). Using the Signac package's "FindTransferAnchors" function, we calculated predicted IDs for snATAC-seq clusters based on snRNA-seq annotations (predicted annotation) (Fig. 2C). Evaluation of the two annotation methods for snATAC-seq revealed that both methods allowed annotation of major cell types, and correlation between the two was maintained. Thus, the validity of predicted annotations based on snRNA-seq was confirmed (Figure 2D).

R2-7: SP3, considered the most undifferentiated stem/progenitor fraction in snRNA-seq, could be divided into A2-0: SP1 and A2-14: SP2 in snATAC-seq.

Comparing gene activity of tendon-related and stem/progenitor genes in each cluster of the ground truth annotation, A2-14: SP2 showed high expression of *Ly6a*, *Pdgfra*, and *Tppp3*, and inverse correlation with *Mkx*, *Scx*, and *Tnmd*. Furthermore, the newly identified surface antigens *Cd55* and *Cd248* also showed high expression in cluster 14. In contrast, *Cd34* did not show a high expression pattern in A2-14: SP2. These results suggest that A2-14: SP2 is the most immature cluster, with *Cd55* and *Cd248* showing characteristic expression (Fig. 2E). Subcluster analysis of R7, considered the most immature fraction in snRNA-seq, revealed two clusters based on *Cd55*/*Cd248* and *Cd34* expression, similar to the snATAC-seq analysis. The former (R7-1) showed high expression of *Ly6a*, *Tppp3*, and *Pdgfra* (Fig. 2F). Trajectory analysis using R2-7 and A2-14 as roots in snRNA-seq and snATAC-seq, respectively, showed increased expression of tendon-related genes *Mkx* and *Scx* along the trajectory, while expression of *Tppp3*, *Cd55*, and *Cd248* decreased (Figure 3A). Evaluation of peaks in each cluster based on gene expression levels in snRNA-seq revealed multiple peaks with heights correlated to gene expression for *Mkx* and *Scx*. Similar evaluation of *Cd55* and *Cd248* showed multiple peaks upstream of the TSS for *Cd248* (Fig. 3B). These results indicate correlation between snRNA-seq and snATAC-seq, and consistent with scRNA-seq results, suggest that A2-14 represents the most undifferentiated TSPCs based on *Cd55* and *Cd248* expression.

Comparison of Single Nucleus RNA+ATAC Analysis between 2-week and 6-week Mouse Achilles Tendon Cells

To confirm the presence of the identified *Cd55*/*Cd248* fraction at 6 weeks, we performed snRNA-seq and snATAC-seq on 6-week-old mouse Achilles tendon cells. Using the 2-week snATAC-seq data as a reference, we annotated the 6-week snATAC-seq clusters. Clusters A2-14 and A2-0 corresponded to clusters A6-12 and A6-0, respectively (Supplemental Figure 6). We also annotated the 6-week snATAC-seq clusters using 2-week snRNA-seq with the same results (Supplemental Figure 7). Gene activity evaluation revealed that, similar to the 2-week data, the 6-week A6-12 and A6-0 clusters showed *Cd55*/*Cd248* high

538 and *Cd34* dim, and Cd55/*Cd248* dim and *Cd34* high patterns, respectively. Furthermore,
539 annotation of the 6-week snRNA-seq data using the 2-week snRNA-seq as a reference
540 revealed clusters with expression patterns similar to those observed in snATAC-seq. In both
541 cases, clusters with high *Cd55* and *Cd248* expression showed high gene activity (snATAC-
542 seq) and gene expression (snRNA-seq) of *Tppp3*, *Pdgfra*, and *Ly6a*. These results confirm
543 that the *Cd55* and *Cd248* high clusters identified in the 2-week snRNA-seq+snATAC-seq
544 analysis are similarly detected at 6 weeks.
545 We also compared the snATAC-seq data between 2 and 6 weeks. No significant
546 differences were observed in genomic annotations between the two time points
547 (Supplemental Figure 8A). After merging the datasets (Supplemental Figure 8B), we
548 compared gene activity of tendon/stem-related genes to evaluate changes in gene activity.
549 We found that the activity of *Tppp3*, *Pdgfra*, *Ly6a*, *Cd55*, *Cd248*, and *Cd34* all decreased at
550 6 weeks (Supplemental Fig 8C). This was consistent with the decreased gene expression
551 observed when comparing 2-week and 6-week data.
552
553 **Estimation of Transcription Factor Activity in 2-week Mouse Achilles Tendon Cells**
554 To estimate transcription factor activity in each cluster, we used the Single-Cell Regulatory
555 Network Inference and Clustering (SCENIC) package (Lake et al., 2018) to calculate gene
556 regulatory network activity from scRNA-seq gene expression data. SCENIC constructs
557 gene expression networks centered on transcription factors and infers transcription factor
558 activity in each cluster. We summarized the predicted transcription factor activities in
559 tendon and stem/progenitor-related clusters R2-7, R2-2, R2-3, R2-1, and R2-13 identified
560 by snRNA-seq (Figure 4).
561 Next, we evaluated the gene activity of these identified transcription factors in each
562 snATAC-seq cluster, along with motif activity calculated by chromVAR (Schep et al., 2017).
563 We also assessed the expression levels of transcription factors in each snRNA-seq cluster.
564 In the most immature fractions, represented by cluster A2-14 in snATAC-seq and cluster
565 R2-7 in snRNA-seq, *KLF transcription factor 3* (*Klf3*), *KLF transcription factor 4* (*Klf4*), and
566 *cAMP responsive element binding protein 5* (*Creb5*) showed consistent behavior in gene
567 activity, motif activity, and gene expression. Additionally, *Signal transducer and activator of*
568 *transcription 2* (*Stat2*) and cAMP responsive element binding protein 3 like 1 (*Creb3l1*)
569 showed high values in A2-0/A2-15 and A2-1, respectively. *Stat2* gene expression was not
570 observed, likely due to false positives resulting from low expression levels. The gene
571 activity and motif activity of each transcription factor estimated by snATAC-seq correlated
572 with the transcription factor activity calculated by SCENIC. Therefore, simultaneous

analysis of snRNA-seq and snATAC-seq could be used to more accurately evaluate the function of transcription factors in each cluster.

Furthermore, SCENIC can predict candidate transcription factors regulating each gene (Table 1). *Cd55* and *Cd248* were predicted to be regulated by *Klf3* and *Klf4*, while *Mkx* and *Scx* were predicted to be under the control of *Creb3l1*. These predictions were consistent with the data calculated from gene activity and motif activity in snATAC-seq.

In vitro Evaluation of CD55 and CD248

Immunostaining evaluation of CD55 and CD248, the identified candidate stem/progenitor markers, revealed expression in the peritenon (Figure 5C, 5A). This was consistent with previous reports of localized expression of *Tppp3*/*Pdgfra*-positive cells in the peritenon and our analysis showing co-expression of *Tppp3*/*Pdgfra* and *Cd55*/*Cd248*. The results support the possibility that SP clusters reflect peritenon. Given that *Cd55* and *Cd248* expression appeared to reflect *Tppp3*, *Pdgfra*, and *Ly6a* expression more sensitively than *Cd34*, we extracted tendon cells from 2-week-old mice and sorted them by FACS to determine the biological phenotype of Cd55 and Cd248 positive cells (Figure 5B).

Gene expression analysis of sorted cells showed that CD55/CD248 positive cells, compared to negative cells, had lower expression of *Mkx*, *Scx*, *Col1a1*, and *Creb3l1*, but higher expression of *Ly6a*, *Tppp3*, *Pdgfra*, *Creb5*, and *Klf3*, consistent with snRNA-seq analysis (Supplemental Figure 9). To assess stem/progenitor capacity, we performed colony formation assays. CD55/CD248 sorted cells showed significantly increased colony formation compared to CD55/CD248 negative cells (control) (Figure 5C).

We then evaluated the differentiation tendencies of these cells towards tendon, cartilage, and bone. Tendon differentiation resulted in clusters of spindle-shaped cells from CD55/CD248 positive cells, suggesting tenogenic differentiation (Figure 5D). Gene expression analysis revealed that while CD55/CD248 positive cells initially had lower *Mkx* and *Scx* expression than negative cells, this pattern reversed after differentiation (Figure 5E). Increased expression of other tendon-related genes such as *Col1a1* and *Bgn* was also observed. In contrast, no significant increases in *Sox9* and *Collagen type II alpha 1 chain* (*Col2a1*) (chondrogenic) or *RUNX family transcription factor 2* (*Runx2*), and *Alkaline phosphatase, biomineralization associated* (*Alpl*) (osteogenic) expression were observed compared to negative cells during cartilage and bone differentiation assays (Supplemental Figure 10). Cells negative for CD55/CD248 could be mixed cell populations, including hematopoietic lineages, cells from tendon mid substance, immune cells, and/or endothelial cells. However, given that CD55/CD248 negative cells have increased expression of genes characteristic of tendons such as *Mkx*, *Scx* and *Col1a1*, we could speculate that this

609 popularity might reflect terminally differentiated tenocytes and CD55/CD248 positive cells
610 possess tenogenic differentiation capacity.
611 To further evaluate the tenogenic potential of CD55/CD248 positive cells, we created
612 tendon-like tissue (bio-tendon) using our previously reported 3D stretch stimulation culture
613 system. SEM and TEM analysis to assess collagen fiber density and thickness showed that
614 bio-tendons derived from CD55/CD248 positive cells had an increased proportion of
615 collagen fibers parallel to the stretch direction and increased collagen fiber diameter. TEM
616 also revealed characteristic banding patterns and triple helix structures in these bio-
617 tendons, indicating mature collagen organization (Wieczorek et al., 2015) (Figure 5F-H).
618 Stretch tests to evaluate the mechanical capacity of the bio-tendons showed high tensile
619 strength (Figure 5I).
620 These results demonstrate that CD55/CD248 positive cells have a tendency to
621 differentiate into tendon cells.
622
623 **Discussion**
624 Tendons are known for their low cellular content, making complete functional recovery
625 after injury challenging and increasing the risk of re-rupture. As a result, cell therapy has
626 garnered attention as a novel treatment strategy, distinct from current conservative and
627 surgical approaches. However, this requires a thorough understanding of TSPCs. Research
628 on TSPCs has been limited due to the lack of identified characteristic surface antigens.
629 To our knowledge, no multi-omics analysis comparing juvenile and mature mouse tendon
630 cells has been conducted. In this study, we performed single-cell RNA-seq along with
631 snRNA-seq+snATAC-seq on nuclei isolated from 2-week and 6-week mouse Achilles
632 tendons. While comparisons using older mice (e.g., 12 weeks or more) were initially
633 considered, they were excluded due to extremely low cell yield and viability, making 2- and
634 6-week-old mice more suitable for this analysis. This approach allowed us to identify
635 candidate surface antigens for TSPCs and elucidate the dynamism of transcription factors
636 involved in post-developmental tendon growth. As a result, we discovered a novel
637 combination of *Cd55* and *Cd248* as surface antigens showing characteristic expression in
638 the TSPC fraction.
639 While scRNA-seq analysis also showed *Cd34* expression characteristic of the TSPC
640 fraction, consistent with previous reports, snRNA-seq + snATAC-seq analysis revealed that
641 the *Cd34* high-expression cluster could be further classified into two clusters based on
642 *Cd55* and *Cd248* expression patterns. The cluster showing high expression of both *Cd55*
643 and *Cd248* also exhibited high expression of *Tppp3*, *Pdgfra*, and *Ly6a*, suggesting that

644 isolating *Cd55* and *Cd248* positive fractions may more sensitively capture immature
645 populations compared to *Cd34*.
646 snRNA-seq+snATAC-seq analysis provided simultaneous information on gene expression
647 levels and open chromatin regions for each cluster. Notably, *Cd248* showed several peaks
648 upstream of the TSS in high-expression clusters, likely reflecting its expression regulation
649 mechanism. We also identified several peaks that increased proportionally with expression
650 for *Mkx* and *Scx*. Given the many unknowns in *Mkx* and *Scx* expression regulation
651 mechanisms (Guerquin et al., 2013; Otabe et al., 2015), we used the SCENIC package to
652 predict upstream transcription factors based on expressed genes in each snRNA-seq
653 cluster. Combined with gene activity and motif activity data from snATAC-seq, Creb3l1 was
654 identified as a transcription factor regulating *Mkx* and *Scx* expression. *Creb3l1* has been
655 reported to increase in expression throughout development (Liu et al., 2015).
656 *Klf3*, *Klf4*, and *Creb5* were identified as characteristic transcription factors in the TSPC
657 fraction. Recent reports have highlighted the role of the *Klf* family in limb development (Kult
658 et al., 2021). Given that tendons are components of the developing limb, these transcription
659 factors may have roles in tendon biology. *Creb5* has been reported to show increased
660 expression from E11 to E13 (Liu et al., 2015), suggesting a role in early developmental
661 stages. Furthermore, *Creb5* has been reported to regulate Proteoglycan 4 (*Prg4*)
662 expression (Zhang et al., 2021). Further investigation into the functions of these
663 transcription factors in TSPCs is necessary.
664 CD55 was identified by Hoffmann et al. in 1969 as a surface antigen functioning as a
665 complement inhibitory factor on erythrocytes. CD55 is known to be expressed from early
666 developmental stages and characterizes the initial differentiation stage of hematopoietic
667 stem cells (Guo et al., 2013). It is also expressed in MSCs and has been reported as an
668 early progenitor marker for mouse mammary epithelial cells (Pal et al., 2017). CD248 was
669 identified in 1992 as an antigen for the FB5 antibody reacting with vascular wall cells
670 (Rettig et al., 1992). CD248 is a transmembrane glycoprotein expressed in pericytes and
671 fibroblasts during developmental stages. Regarding their relationship, CD55 and CD248
672 have been reported to be expressed in stromal cells during the early stages of arthritis
673 (Choi et al., 2017), but their detailed mechanisms in tendons remain unclear.
674 Previously suggested TSPC surface antigen candidates such as *Cd73*, *Cd90*, and *Cd105*
675 showed poor correlation with the expression of genes like *Tppp3*. Moreover, the initial
676 report on TSPCs (Bi et al., 2007) described TSPCs as *Scx+Cd34-*. The high expression of
677 CD55 and CD248 in the peritenon, similar to *Tppp3* (Harvey et al., 2019; Staverosky et al.,
678 2019), suggests the possibility of cells with tenogenic differentiation potential exhibiting
679 *Scx+Cd34-* expression patterns within the tendon. Spatial information is important to

investigate further. Future studies using lineage tracing experiments with mice labeled for CD55 and CD248 or mouse model of selective ablation of CD55 and CD248 are necessary to analyze the developmental functions of CD55 and CD248 positive cells and their roles in injury healing in more detail.

Clinically, to our knowledge, few studies have sorted TSPCs based on surface antigens and examined their *in vitro* tendon tissue-generating ability. In this study, we found that cells sorted for CD55/CD248 showed higher clonogenicity compared to CD55-/CD248- cells and demonstrated superior tendon tissue-generating ability in an artificial tendon model. In the future, CD55/CD248 double-positive TSPC cells may prove useful in clinical applications such as *in vitro* artificial tendon creation (Tsutsumi et al., 2022) and cell therapy for tendon injuries (Huang et al., 2021).

Reference

Andarawis-Puri, N., Flatow, EL., Soslowsky, LJ. (2015). Tendon basic science: Development, repair, regeneration, and healing. J Orthop Res. 33(6), 780-4.

Asahara, H., Inui, M., Lotz, MK. (2017). Tendons and Ligaments: Connecting Developmental Biology to Musculoskeletal Disease Pathogenesis. J Bone Miner Res. 32(9), 1773-1782.

Becht, E., McInnes, L., Healy, J., Dutertre, CA., Kwok, IWH., Ng, LG., Ginhoux, F., Newell. EW. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 37, 38-44.

Bi Y, Ehirchiou D, Kilts TM, Inkson CA, Embree, MC., Sonoyama, W., Li, L., Leet, AI., Seo, BM., Zhang, L., Shi, S., Young, MF. (2007). Identification of tendon stem/progenitor cells and the role of the extracellular matrix in their niche. Nat Med. 13(10), 1219-27.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 36(5), 411-420.

Cao, Y., Liu, Y., Liu, W., Shan, Q., Buonocore, SD., Cui, L. (2002). Bridging tendon defects using autologous tenocyte engineered tendon in a hen model. Plast Reconstr Surg. 110(5), 1280-9.

Chen, B., Wang, B., Zhang, WJ., Zhou, G., Cao, Y., Liu, W. (2012). In vivo tendon engineering

with skeletal muscle derived cells in a mouse model. Biomaterials. 33(26), 6086-97.

Chen, L., Liu, JP., Tang, KL., Wang, Q., Wang, GD., Cai, XH., Liu, XM. (2014). Tendon derived stem cells promote platelet-rich plasma healing in collagenase-induced rat achilles tendinopathy. Cell Physiol Biochem. 34(6), 2153-68.

Cho, YT., Hsu, WY., Lin, LF., Lin, YN. (2018). Kinesio taping reduces elbow pain during resisted wrist extension in patients with chronic lateral epicondylitis: a randomized, double-blinded, cross-over study. BMC Musculoskelet Disord. 19(1), 193.

Connizzo, BK., Yannascoli, SM., Soslowsky, LJ. (2013). Structure-function relationships of postnatal tendon development: a parallel to healing. Matrix Biol. 32(2), 106-16.

Craft, A. M., Ahmed, N., Rockel, J. S., Baht, G. S., Alman, B. A., Kandel, R. A., Grigoriadis, A. E. and Keller, G. M. (2013). Specification of chondrocytes and cartilage tissues from embryonic stem cells. Development 140, 2597-2610.

Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., Lin-Marq, N., Koch, M., Bilio, M., Cantiello, I., Verde, R., De, Masi, C., Bianchi, SA., Cicchini, J., Perroud, E., Mehmeti, S., Dagand, E., Schrinner, S., Nürnberger, A., Schmidt, K., Metz, K., Zwingmann, C., Brieske, N., Springer, C., Hernandez, AM., Herzog, S., Grabbe, F., Sieverding, C., Fischer, B., Schrader, K., Brockmeyer, M., Dettmer, S., Helbig, C., Alunni, V., Battaini, MA., Mura, C., Henrichsen, CN., Garcia-Lopez, R., Echevarria, D., Puelles, E., Garcia-Calero, E., Kruse, S., Uhr, M., Kauck, C., Feng, G., Milyaev, N., Ong, CK., Kumar, L., Lam, M., Semple, CA., Gyenesei, A., Mundlos, S., Radelof, U., Lehrach, H., Sarmientos, P., Reymond, A., Davidson, DR., Dollé, P., Antonarakis, SE., Yaspo, ML., Martinez, S., Baldock, RA., Eichele, G., Ballabio, A. (2011). A High-Resolution Anatomical Atlas of the Transcriptome in the Mouse Embryo. PLoS Biol. 9(1), e1000582.

Gracey, E., Burssens, A., Cambré, I., Schett, G., Lories, R., McInnes, IB., Asahara, H., Elewaut, D. (2020). Tendon and ligament mechanical loading in the pathogenesis of inflammatory arthritis. Nat Rev Rheumatol. 16(4):193–207.

Fang, F., Xiao, Y., Zelzer, E., Leong, KW., Thomopoulos, S. (2022). A mineralizing pool of Gli1-expressing progenitors builds the tendon enthesis and demonstrates therapeutic potential. Cell Stem Cell. 29(12), 1669-1684.

Franken, NA., Rodermond, HM., Stap, J., Haveman, J., van Bree C. (2006). Clonogenic assay of cells in vitro. Nat Protoc. 2006. 1(5), 2315-9

Grinstein, M., Dingwall, HL., O'Connor, LD., Zou, K., Capellini, TD., Galloway, JL. (2019). A distinct transition from cell growth to physiological homeostasis in the tendon. Elife. 8, e48689.

Harvey, T., Flamenco, S., Fan, CM. (2019). A Tppp3 + Pdgfra + tendon stem cell population contributes to regeneration and reveals a shared role for PDGF signalling in regeneration and fibrosis. Nat Cell Biol. 21(12), 1490-1503.

Guilak, F., Butler, DL., Goldstein, SA., Baaijens, FP. (2014). Biomechanics and mechanobiology in functional tissue engineering. J Biomech. 47(9), 1933-40

Gumucio, JP., Sugg, KB., Mendias, CL. (2015). TGF-β superfamily signaling in muscle and tendon adaptation to resistance exercise. Exerc Sport Sci Rev. 43(2), 93-9.

Guzmán, C., Bagga, M., Kaur, A., Westermarck, J., Abankwa, D. (2014). ColonyArea: an ImageJ plugin to automatically quantify colony formation in clonogenic assays. PLoS One. 9(3), e92444.

Havis, E., Bonnin, MA., Olivera-Martinez, I., Nazaret, N., Ruggiu, M., Weibel, J., Durand, C., Guerquin, MJ., Bonod-Bidaud, C., Ruggiero, F., Schweitzer, R., Duprez, D. (2014). Transcriptomic analysis of mouse limb tendon cells during development. Development. 141(19), 3683-96.

Hittinger, M., Czyz, ZT., Huesemann, Y., Maneck, M., Botteron, C., Kaeufl, S., Klein, CA., Polzer, B. (2013). Molecular profiling of single Sca-1+/CD34+,- cells--the putative murine lung stem cells. PLoS One. 8(12), e83917.

Hoffmann EM. (1969). Inhibition of complement by a substance isolated from human erythrocytes. II. Studies on the site and mechanism of action. Immunochemistry. 6(3), 405-19.

Holmes, C., Stanford, WL. (2007). Concise review: stem cell antigen-1: expression, function, and enigma. Stem Cells. 25(6), 1339-47.

Howell, K., Chien, C., Bell, R., Laudier, D., Tufa, SF., Keene, DR., Andarawis-Puri, N., Huang, AH. (2017). Novel Model of Tendon Regeneration Reveals Distinct Cell Mechanisms Underlying Regenerative and Fibrotic Tendon Healing. Sci Rep. 7, 45238.

Isola, JVV., Ocañas, SR., Hubbart, CR., Ko, S., Mondal, SA., Hense, JD., Carter, HNC., Schneider, A., Kovats, S., Alberola-Ila, J., Freeman, WM., Stout, MB. (2024). A single-cell atlas of the aging mouse ovary. Nat Aging. 4(1), 145-162.

Jacobson, KR., Lipp, S., Acuna, A., Leng, Y., Bu, Y., Calve, S. (2020). Comparative Analysis of the Extracellular Matrix Proteome across the Myotendinous Junction. 19(10), 3955-3967.

Jozsa, L., Kannus, P., Balint, JB., Reffy, A. (1991). Three-dimensional ultrastructure of human tendons. Acta Anat (Basel). 142(4), 306-12.

Kataoka, K., Kurimoto, R., Tsutsumi, H., Chiba, T., Kato, T., Shishido,. K., Kato, M., Ito, Y., Cho, Y., Hoshi, O., Mimata, A., Sakamaki, Y., Nakamichi, R., Lotz, MK., Naruse, K., Asahara, H. (2020). In vitro Neo-Genesis of Tendon/Ligament-Like Tissue by Combination of Mohawk and a Three-Dimensional Cyclic Mechanical Stretch Culture System. Front Cell Dev Biol. 2020;8:307.

Kendal, AR., Layton, T., Al-Mossawi, H., Appleton, L., Dakin, S., Brown, R., Loizou, C., Rogers, M., Sharp, R., Carr, A. (2020). Multi-omic single cell analysis resolves novel stromal cell populations in healthy and diseased human tendon. Sci Rep. 10(1): 13939.

Leong, DJ., Sun, HB. (2016). Mesenchymal stem cells in tendon repair and regeneration: basic understanding and translational challenges. Ann. N. Y. Acad. Sci. 1383, 88–96.

Leong, NL., Kator, JL., Clemens, TL., James, A., Enamoto-Iwamoto, M., Jiang, J. (2020). Tendon and Ligament Healing and Current Approaches to Tendon and Ligament Regeneration. J Orthop Res. 38(1), 7-12.

Li, Y., Wu, T., Liu, S. (2021). Identification and Distinction of Tenocytes and Tendon-Derived Stem Cells. Front Cell Dev Biol. 9, 629515.

Liu, H., Xu, J., Liu, CF., Lan, Y., Wylie, C., Jiang, R. (2015). Whole transcriptome expression

profiling of mouse limb tendon development by using RNA-seq. J Orthop Res. 33(6), 840–848.

Loiacono, C., Palermi, S., Massa, B., Belviso, I., Romano, V., Gregorio, AD., Sirico, F., Sacco, AM. (2019). Tendinopathy: Pathophysiology, Therapeutic Options, and Role of Nutraceutics. A Narrative Literature Review. Medicina (Kaunas). 55(8), 447.

Lui, PP. (2015). Markers for the identification of tendon-derived stem cells in vitro and tendon stem cells in situ - update and future development. Stem Cell Res Ther. 6(1), 106.

Lui, PPY., Wong, CM. (2019). Biology of Tendon Stem Cells and Tendon in Aging. Front Genet. 10, 1338.

McGinnis, CS., Murrow, LM., Gartner, ZJ. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Syst. 8(4), 329-337.e4.

Mendias, CL., Gumucio, JP., Bakhurin, KI., Lynch, EB., Brooks, SV. (2012). Physiological loading of tendons induces scleraxis expression in epitenon fibroblasts. J Orthop Res. 30(4), 606-12.

Morikawa, S., Mabuchi, Y., Kubota, Y., Nagai, Y., Niibe, K., Hiratsu, E., Suzuki, S., Miyauchi-Hara, C., Nagoshi, N., Sunabori, T., Shimmura, S., Miyawaki, A., Nakagawa, T., Suda, T., Okano, H., and Matsuzaki, Y. (2009). Prospective identification, isolation, and systemic transplantation of multipotent mesenchymal stem cells in murine bone marrow. J. Exp. Med. 206, 2483-2496.

Nagakura, R., Yamamoto, M., Jeong J, Hinata N, Katori Y, Chang WJ, Abe S. (2020). Switching of Sox9 expression during musculoskeletal system development. Sci Rep. 10(1), 8425.

Nusspaumer, G., Jaiswal, S., Barbero, A., Reinhardt, R., Ishay Ronen, D., Haumer, A., Lufkin, T., Martin, I. and Zeller, R. (2017). Ontogenic identification and analysis of mesenchymal stromal cell populations during mouse limb and long bone development. Stem Cell Rep. 9, 1124-1138.

Pliner, HA., Packer, JS., McFaline-Figueroa, JL., Cusanovich, DA., Daza, RM., Aghamirzaie,

D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., Adey, AC., Steemers, FJ., Shendure, J., Trapnell, C. (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol. Cell. 71, 858–871. e8.

Ruzzini, L., Abbruzzese, F., Rainer, A., Longo, UG., Trombetta, M., Maffulli, N., Denaro, V. (2014). Characterization of age-related changes of tendon stem cells from adult human tendons. Knee Surg Sports Traumatol Arthrosc. 22(11), 2856-66.

Satpathy, AT., Granja, JM., Yost, KE., Qi, Y., Meschi, F., McDermott, GP., Olsen, BN., Mumbach, MR., Pierce, SE., Corces, MR., Shah, P., Bell, JC., Jhutty, D., Nemec, CM., Wang, J., Wang, L., Yin, Y., Giresi, PG., Chang, ALS., Zheng, GXY., Greenleaf, WJ., Chang, HY. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol. 37(8), 925-936.

Sidney, LE., Branch, MJ., Dunphy, SE., Dua, HS., Hopkinson, A. (2014). Concise review: evidence for CD34 as a common marker for diverse progenitors. Stem Cells. 32(6), 1380-9.

Somerville, JM., Aspden, RM., Armour, KE., Armour, KJ., Reid, DM. (2004). Growth of C57BL/6 mice and the material and mechanical properties of cortical bone from the tibia. Calcif Tissue Int. 74(5), 469-75.

Staverosky, JA., Pryce, BA., Watson, SS., and Schweitzer, R. (2019). Tubulin polymerization-promoting protein family member 3, Tppp3, is a specific marker of the differentiating tendon sheath and synovial joints. Dev Dyn. 238(3), 685-92.

Steinmann, S., Pfeifer, CG., Brochhausen, C., Docheva, D. (2020). Spectrum of Tendon Pathologies: Triggers, Trails and End-State. Int J Mol Sci. 21(3), 844.

Still, C, 2nd., Chang, WT., Sherman, SL., Sochacki, KR., Dragoo, JL., Qi, LS. (2021). Single-cell transcriptomic profiling reveals distinct mechanical responses between normal and diseased tendon progenitor cells. Cell Rep Med. 2(7), 100343.

Stuart, T., Hanna, J., Gutruf, P. (2021). Single-cell chromatin state analysis with Signac. Nat Methods. 18(11), 1333-1341.

Sung, JH., Yang, HM., Park, JB., Choi, GS., Joh, JW., Kwon, CH., Chun, JM., Lee, SK., Kim,

SJ. (2008). Isolation and characterization of mouse mesenchymal stem cells. Transplant Proc. 40(8), 2649-54.

Tachibana, N., Chijimatsu, R., Okada, H., Oichi, T., Taniguchi, Y., Maenohara, Y., Miyahara, J., Ishikura, H., Iwanaga, Y., Arino, Y., Nagata, K., Nakamoto, H., Kato, S., Doi, T., Matsubayashi, Y., Oshima, Y., Terashima, A., Omata, Y., Yano, F., Maeda, S., Ikegawa, S., Seki, M., Suzuki, Y., Tanaka, S., Saito, T. (2022). RSPO2 defines a distinct undifferentiated progenitor in the tendon/ligament and suppresses ectopic ossification. Sci Adv. 8(33), eabn2138.

Takakura, N., Yoshida, H., Ogura, Y., Kataoka, H., Nishikawa, S., and Nishikawa, S. (1997). PDGFR alpha expression during mouse embryogenesis: immunolocalization analyzed by whole-mount immunohistostaining using the monoclonal anti-mouse PDGFR alpha antibody APA5. J. Histochem. Cytochem. 45, 883-893.

Tang, JB., Wu, YF., Cao, Y., Chen, CH., Zhou, YL., Avanessian, B., Shimada, M., Wang, XT., Liu, PY. (2016). Basic FGF or VEGF gene therapy corrects insufficiency in the intrinsic healing capacity of tendons. Sci Rep. 6, 20643.

Tempfer, H., Wagner, A., Gehwolf, R., Lehner, C., Tauber, M., Resch, H., Bauer, HC. (2009). Perivascular cells of the supraspinatus tendon express both tendon- and stem cell-related markers. Histochem Cell Biol. 131(6), 733-41.

Thomopoulos, S., Parks, WC., Rifkin, DB., Derwin KA. (2015). Mechanisms of tendon injury and repair. J Orthop Res. 33(6), 832-9.

Trapnell, C., Cacchiarelli. D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, NJ., Livak, KJ., Mikkelsen, TS., Rinn, JL. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 32(4), 381-386.

Tsutsumi, H., Kurimoto, R., Nakamichi, R., Chiba, T., Matsushima, T., Fujii, Y., Sanada, R., Kato, T., Shishido, K., Sakamaki, Y., Kimura, T., Kishida, A., Asahara, H. (2022). Generation of a tendon-like tissue from human iPS cells. J Tissue Eng. 13, 20417314221074018.

Walia, B., Huang, AH. (2019). Tendon stem progenitor cells: Understanding the biology to inform therapeutic strategies for tendon repair. J Orthop Res . 37(6), 1270-1280.

Wang, B., Liu, W., Zhang, Y., Jiang, Y., Zhang, WJ., Zhou, G., Cui, L., Cao, Y. (2008). Engineering of extensor tendon complex by an ex vivo approach. Biomaterials. 29(20): 2954-61.

Yin, Z., Hu, JJ., Yang, L., Zheng, ZF., An, CR., Wu, BB., Zhang, C., Shen, WL., Liu, HH., Chen, JL., Heng, BC., Guo, GJ., Chen, X., Ouyang, HW. (2016). Single-cell analysis reveals a nestin + tendon stem/progenitor cell population with strong tenogenic potentiality. Sci Adv. 2(11), e1600874.

Yu, G., Wang, LG., Han, Y., He, QY. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 16(5), 284-7.

Yu, G., Wang, LG., He, QY. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 31(14), 2382-3.

Zhang, C., Zhu, J., Zhou, Y., Thampatty, BP., Wang, JH. (2019). Tendon Stem/Progenitor Cells and Their Interactions with Extracellular Matrix and Mechanical Loading. Stem Cells Int .2019, 3674647.

Zhang, T., Wan, L., Xiao, H., Wang, L., Hu, J., Lu, H.(2023) Single-cell RNA sequencing reveals cellular and molecular heterogeneity in fibrocartilaginous enthesis formation. Elife. 12: e85873.

Acknowledgements

We thank all the members of the Department of Systems BioMedicine at Institute of Science Tokyo for their support. We also thank the Research Core at Institute of Science Tokyo for supporting cell sorting.

Author contributions

Conceptualization, H.T., T.C., and H.A.; Methodology, H.T., T.C., and H.A.; Formal analysis, H.T.; Investigation, H.T., F. Y., T. M., T. K., A. K., K. A., Y. S.; Writing – Original Draft, H.T.; Writing – Review & Editing, T.C., and H.A.; Funding Acquisition, H.A.; Supervision, H.A.

Data availability

FASTQ data of RNA-Seq and ATAC-seq are deposited in DDBJ under accession number

PRJDB18857

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by JSPS KAKENHI (Grant Numbers JP15H02560, JP20H05696, 16H06279 (PAGS)), AMED (Grant Numbers JP21gm0810008, JP23ym0126805, JP24gm0010009), and NIH (Grant Number R01AR080127) to H.A.

Figure legends

Figure 1. scRNA-seq of tendon cells from 2-week-old and 6-week-old mice and the identification of surface markers of TSPC.

(A) Integrated uniform manifold approximation and projection (UMAP) scRNA-seq clustering of cells harvested from 2-week-old and 6-week-old mouse Achilles tendons.

(B) Dot plot of average gene expression levels of the indicated genes in each scRNA-seq cluster. The size of the dot reflects the percentage of cells in the cluster that express each gene. TC, tenocyte; SP1, tendon stem/progenitor cell_1; CA, cartilage; RB, ribosomal RNA; SP2, tendon stem/progenitor cell_2; LC1, lymphocyte_1; EC, endothelial cell; RBC, red blood cell; SM, smooth muscle cell; PC, proliferating cell; LC2, lymohocytes_2; MTJ, myotendinous junction cell; SW, Schwann cell; VEC, vascular endothelial cell; MC, macrophage.

(C) Proportions of cells in clusters identified from scRNA-seq. Clusters are colored according to cluster type.

(D) Volcano plot of gene expression in the SP2 cluster and the identification of candidate TSPC marker genes (red under line).

(E) Feature plot of *Cd55* and *Cd248* expression.

(F) Correlation of gene expression of TSPC candidate genes in 2-week data.

(G) Violin plots presenting the gene expression changes for a selection of differentially expressed genes.

Figure 2. snATAC-seq of tendon cells from a 2-week-old mouse and the validation of Cd55 and Cd248 as candidate markers of TSPC.

(A) UMAP snATAC-seq clustering of cells derived from the Achilles tendon of a 2-week-old mouse. Annotation was based on each gene activity (ground-truth annotation). SP1, tendon stem/progenitor cell_1; TC1, tenocyte_1; MTJ, myotendinous junction cell; SM, smooth

1004 muscle cell; MC, macrophage; CA, cartilage; TC2, tenocyte_2; SK, skeletal muscle cell; SW,
1005 Schwann cell; LC, lymphocyte; NC, neutrophil; EC, endothelial cell; RBC, red blood cell; UC,
1006 unspecified cell; SP2, tendon stem/progenitor cell_2; SP3, tendon stem/progenitor cell_3;
1007 VEC, vascular endothelial cell.
1008 (B) Dot plot of average gene activity of the indicated genes in each snATAC-seq cluster. The
1009 size of the dot reflects the percentage of cells in the cluster that express each gene.
1010 (C) UMAP visualization and predicted annotation of 2-week snATAC-seq after integration and
1011 label transfer of 2-week snRNA-seq data.
1012 (D) Identification of matching cell clusters between the 2-week snRNA- and 2-week snATAC-
1013 seq data from visualized as heatmap. The heatmap shows the proportions of cells from each
1014 snATAC-seq cluster across all sample conditions assigned to each snRNA-seq cluster as
1015 part of the label-transfer process.
1016 (E) Violin plot of tenocytes and TSPC-related gene expression in each cluster.
1017
1018 **Figure 3. Trajectory analysis and peak visualization of snRNA-seq and snATAC-seq**
1019 **data for the tendon and tendon stem/progenitor cell-related cluster.**
1020 (A) UMAP representation of snRNA-seq differentiation trajectory of tenocytes and TSPC
1021 lineage and pseudotime-dependent gene expression changes of *Tppp3*, *Cd55*, *Cd248*, *Mkx,*
1022 and *Scx,* as inferred using Monocle3.
1023 (B) UMAP representation of snATAC-seq differentiation trajectory of tenocytes and the TSPC
1024 lineage and pseudotime-dependent gene expression changes, as inferred using Cicero.
1025 (C) Coverage plots of *Mkx*, *Scx*, *Cd55,* and *Cd248*. Selected peaks that differ across each
1026 cluster are highlighted.
1027
1028 **Figure 4. Transcription factor landscapes of 2-week mouse Achilles tendons.**
1029 SCENIC analysis of transcription factor activity based on 2-week snRNA-seq data for
1030 tenocytes and the TSPC lineage (Left). Validation was performed based on the gene activity
1031 and motif activity of 2-week snATAC-seq and gene expression of 2-week snRNA-seq (right).
1032
1033 **Figure 5. *In vitro* analysis of CD55+/CD248+ TSPCs.**
1034 (A) Immunohistochemical image of 10-week mouse Achilles tendons. Scale bars show
1035 100μm. CD55 and CD248, green; Hoechst 33342, blue.
1036 (B) Schema of the *in vitro* assessment of the capacity of CD55+/CD248+ TSPCs as the
1037 differentiation toward tenocytes.
1038 (C) Colony-forming efficiency of CD55+/CD248+ and CD55-/CD248- (negative) TSPCs.
1039 Colonies were stained with crystal violet (n = 6). CD55+/CD248+ TSPC exhibited higher

clonogenic capacity. Data are presented as means ± SEM. $^{**}p < 0.01$.

(D) Morphological changes of CD55+/CD248+ and negative TSPCs after tenogenic induction.

(E) Quantitative PCR of tendon-related gene expression in CD55+/CD248+ and negative TSPCs after tenogenic induction (n = 3). Data are presented as means ± SD. $^{**}p < 0.01$, $^{*}p < 0.05$.

(F) SEM and TEM imaging of artificial tendons derived from CD55+/CD248+ and negative TSPCs. Data are presented as means ± SD. $^{**}p < 0.01$, $^{*}p < 0.05$.

(G) Proportions of fiber alignment for each artificial tendon (n = 4). Data are presented as means ± SEM. $^{***}p < 0.005$, $^{**}p < 0.01$, $^{*}p < 0.05$.

(H) Diameter of collagen fiber in each artificial tendon based on TEM imaging (n = 4). Data are presented as means ± SD. $^{***}p < 0.005$.

(I) Tensile strength (MPa) of each artificial tendon (n = 5). Data are presented as means ± SEM. $^{***}p < 0.005$.

Table1: Estimated TF activity in each gene

Supplemental Figure 1. 2w- and 6w-scRNA seq results.

(A) Integrated uniform manifold approximation and projection (UMAP) scRNA-seq clustering of cells harvested from 2-week-old and 6-week-old mouse Achilles tendons. Clusters are colored according to cluster type.

(B) Feature plot of *Tppp3*, *Pdgfra* and *Ly6a* expression.

Supplemental Figure 2. Comparison of 2w- and 6w-scRNA seq results.

(A) Violin plot of gene expression enriched in the limb bud (Development (2014) 141, 3683-3696) in each cluster.

(B) Gene Ontology (GO) terms associated with genes with upregulated expression in the SP2 cluster.

(C) Number of inferred interactions and interaction strength of genes expressed within 2w- and 6w-mouse scRNA-seq data.

Supplemental Figure 3. Comparative analysis of 2w- and 6w-scRNA seq results.

Heatmap showing differential gene expression between red (2w) and green (6w).

Supplemental Figure 4. Analysis of 2-week snRNA-seq data.

(A) UMAP snRNA-seq clustering of cells harvested from 2-week mouse Achilles tendons.

(B) Dot plot of average gene expression levels of the indicated genes in 2-week snRNA-seq clusters. RB, ribosomal RNA; TC1, tenocyte_1; SP1, stem/progenitor cell_1; SP2,

stem/progenitor cell_2; SM, smooth muscle cell; MTJ, myotendinous junction cell; CA, cartilage; SP3, stem/progenitor cell_3; LC, lymphocyte; MC, macrophage; RBC, red blood cell; SW, Schwann cell; EC, endothelial cell; TC2, tenocyte_2; PC, proliferating cell; NC, neutrophil; UC, unspecified cell.

(C) Violin plot of tenocytes and TSPC-related gene expression in each cluster.

(D) Feature plot of TSPC-related gene expression.

Supplemental Figure 5. Analysis of 6-week snATAC-seq.

(A) UMAP snATAC-seq clustering of cells harvested from 6-week mouse Achilles tendons. Annotation was based on gene activity (ground-truth annotation, left) and the predicted annotation inferred from 2-week snATAC-seq (right).

(B) Identification of matching cell clusters between the 2-week and 6-week snATAC-seq data, visualized as a heatmap.

(C) Violin plot of tenocytes and TSPC-related gene activity in each cluster.

Supplemental Figure 6. Comparison of snRNA-seq and scRNA-seq.

(A) Dot plot of representative DEGs in snRNA-seq (2w) and snRNA-seq (2w+6w).

(B) Relevance of annotation in each data.

Supplemental Figure 7. Analysis of 6-week snRNA-seq.

(A) UMAP snRNA-seq clustering of cells harvested from 6-week mouse Achilles tendons. Annotation was based on each gene expression (ground-truth annotation, left) and predicted annotation inferred from 2-week snRNA-seq (right).

(B) Identification of matching cell clusters between the 2-week and 6-week snRNA-seq data visualized as a heatmap.

(C) Violin plot of tenocytes and TSPC-related gene expression in each cluster.

Supplemental Figure 8. Comparison of 2-week and 6-week snATAC-seq.

(A) Circle plot of annotated differentially accessible regions for each data.

(B) Integrated UMAP snATAC-seq clustering of cells harvested from 2-week and 6-week mouse Achilles tendons.

(C) Feature plot of TSPC-related gene activity in each dataset.

Supplemental Figure 9. Gene expression changes in CD55+/CD248+ and negative TSPCs.

Quantitative PCR of gene expression in CD55+/CD248+ and negative TSPCs (n = 4). Data

1112 are presented as means ± SEM. *$p < 0.05$, ***$p < 0.005$.

1113

1114 **Supplemental Figure 10. Chondrogenic and osteogenic induction of CD55+/CD248+**
1115 **and negative TSPCs.**

1116 Quantitative PCR of tendon-related genes in CD55+/CD248+ and negative TSPCs after
1117 chondrogenic and osteogenic induction (n = 4). Data are presented as means ± SEM. **$p <$
1118 0.01, *$p < 0.05$.

Figure 1

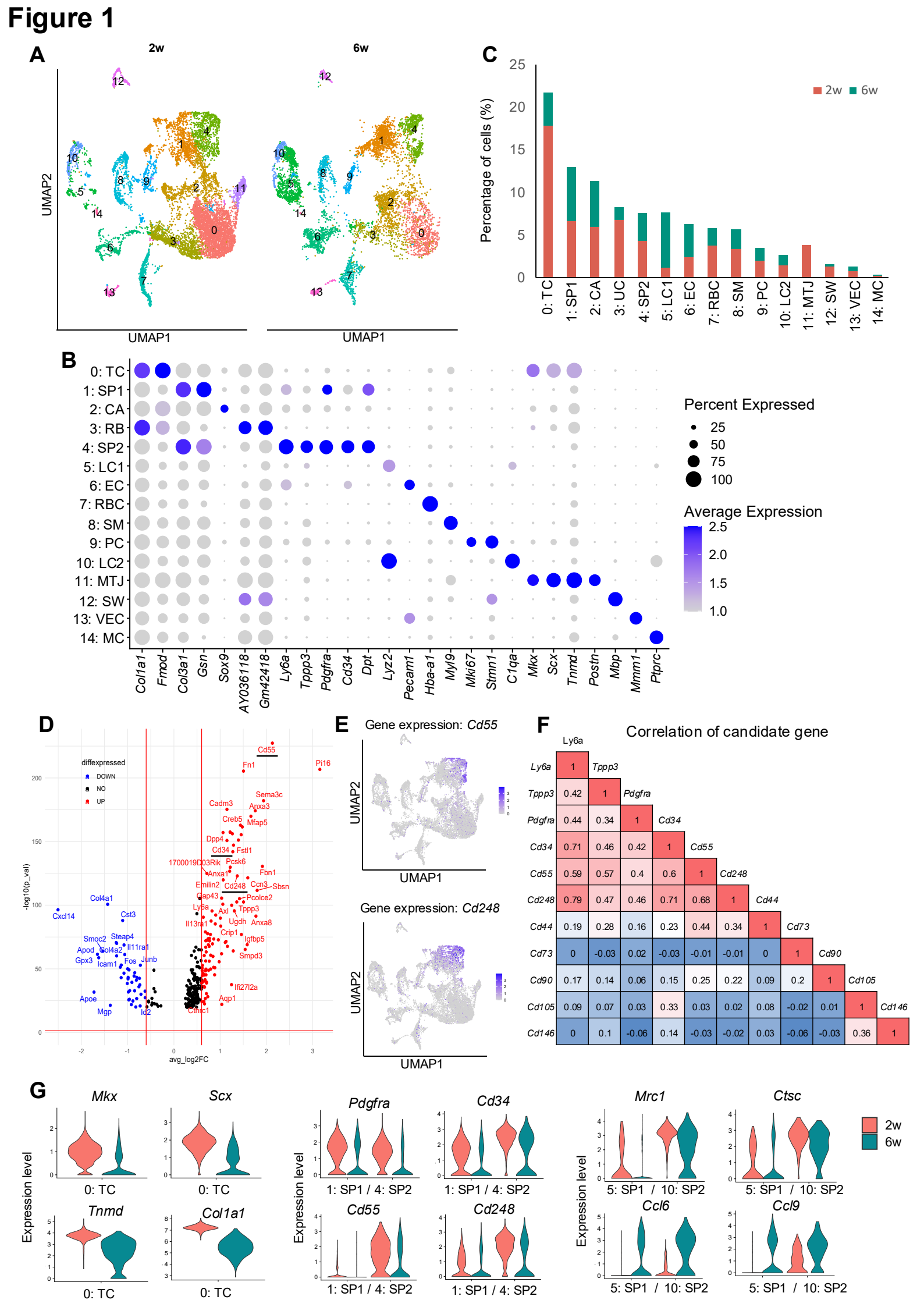

A
2w
6w
UMAP2
UMAP1
C
Percentage of cells (%)
2w
6w
0: TC
1: SP1
2: CA
3: UC
4: SP2
5: LC1
6: EC
7: RBC
8: SM
9: PC
10: LC2
11: MTJ
12: SW
13: VEC
14: MC
B
Percent Expressed
25
50
75
100
Average Expression
2.5
2.0
1.5
1.0
Col1a1
Fmod
Col3a1
Gsn
Sox9
AY036118
Gm42418
Ly6a
Tppp3
Pdgfra
Cd34
Dpt
Lyz2
Pecam1
Hba-a1
Myl9
Mki67
Stmn1
C1qa
Mkx
Scx
Tnmd
Postn
Mbp
Mmm1
Ptprc
D
diffexpressed
DOWN
NO
UP
-log10(p_val)
avg_log2FC
Cd55
Fn1
Pi16
Sema3c
Cadm3
Anxa3
Creb5
Mfap5
Dpp4
Cd34
Fstl1
1700019D03Rik
Pcsk6
Anxa1
Fbn1
Emilin2
Cd248
Ccn3
Sbsn
Gap43
Pcolce2
Ly6a
Axl
Tppp3
Il13ra1
Ugdh
Anxa8
Crip1
Igfbp5
Smpd3
Ifi27l2a
Aqp1
Cthrc1
Col4a1
Cxcl14
Cst3
Smoc2
Steap4
Apod
Col4a2
Il11ra1
Gpx3
Icam1
Fos
Junb
Apoe
Mgp
Id2
E
Gene expression: Cd55
Gene expression: Cd248
UMAP2
UMAP1
F
Correlation of candidate gene
Ly6a
Tppp3
Pdgfra
Cd34
Cd55
Cd248
Cd44
Cd73
Cd90
Cd105
Cd146
G
Mkx
Scx
Pdgfra
Cd34
Mrc1
Ctsc
Tnmd
Col1a1
Cd55
Cd248
Ccl6
Ccl9
Expression level
0: TC
1: SP1 / 4: SP2
5: SP1 / 10: SP2
2w
6w

Figure 2

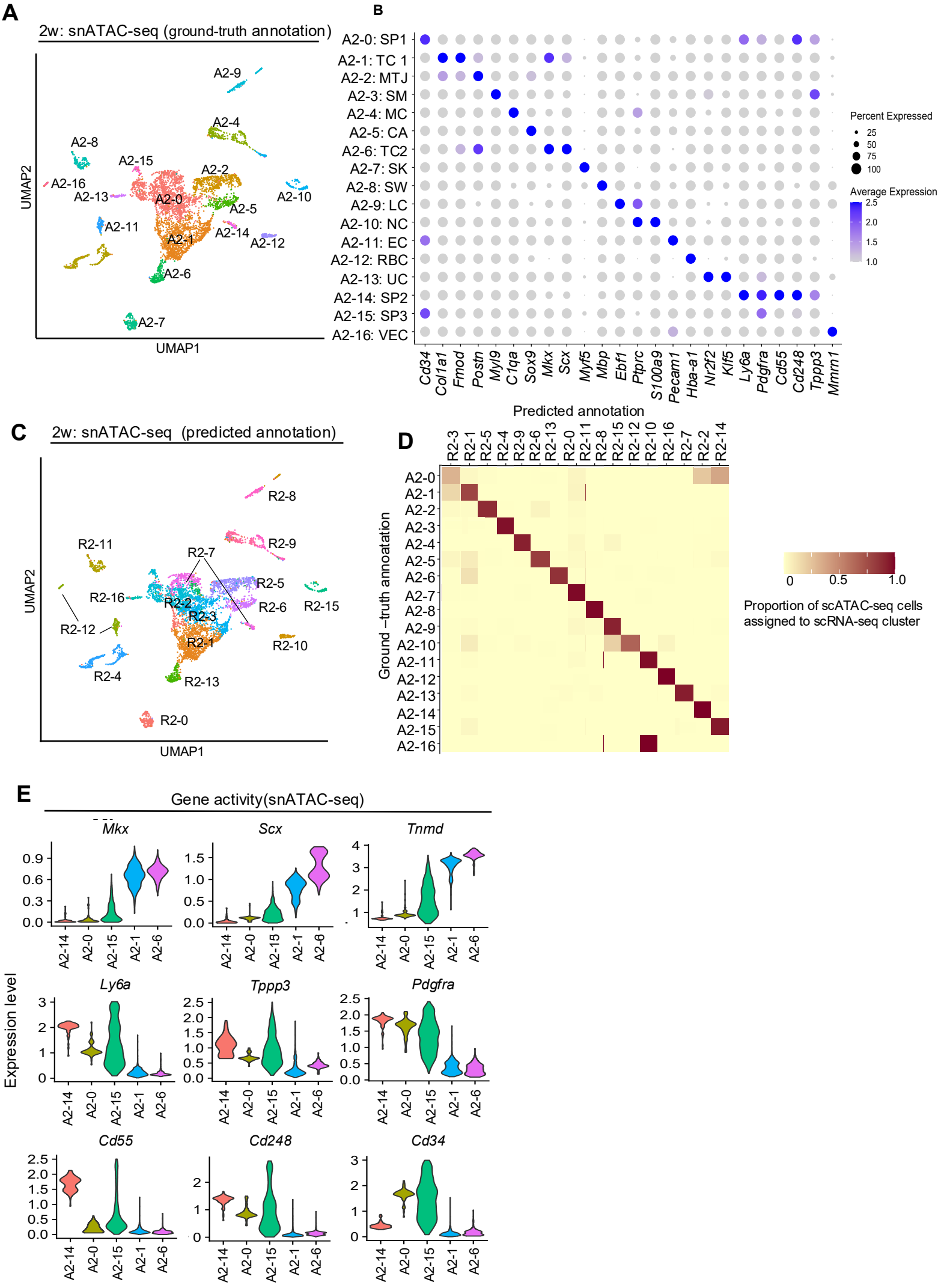

A
2w: snATAC-seq (ground-truth annotation)
UMAP2
UMAP1
A2-0
A2-1
A2-2
A2-4
A2-5
A2-6
A2-7
A2-8
A2-9
A2-10
A2-11
A2-12
A2-13
A2-14
A2-15
A2-16
B
A2-0: SP1
A2-1: TC 1
A2-2: MTJ
A2-3: SM
A2-4: MC
A2-5: CA
A2-6: TC2
A2-7: SK
A2-8: SW
A2-9: LC
A2-10: NC
A2-11: EC
A2-12: RBC
A2-13: UC
A2-14: SP2
A2-15: SP3
A2-16: VEC
Cd34
Col1a1
Fmod
Postn
Myl9
C1qa
Sox9
Mkx
Scx
Myf5
Mbp
Ebf1
Ptprc
S100a9
Pecam1
Hba-a1
Nr2f2
Klf5
Ly6a
Pdgfra
Cd55
Cd248
Tppp3
Mmrn1
Percent Expressed
25
50
75
100
Average Expression
2.5
2.0
1.5
1.0
C
2w: snATAC-seq (predicted annotation)
UMAP2
UMAP1
R2-0
R2-1
R2-2
R2-3
R2-4
R2-5
R2-6
R2-7
R2-8
R2-9
R2-10
R2-11
R2-12
R2-13
R2-15
R2-16
D
Predicted annotation
Ground –truth annotation
0
0.5
1.0
Proportion of scATAC-seq cells assigned to scRNA-seq cluster
E
Gene activity(snATAC-seq)
Expression level
Mkx
Scx
Tnmd
Ly6a
Tppp3
Pdgfra
Cd55
Cd248
Cd34
A2-14
A2-0
A2-15
A2-1
A2-6

Figure 3

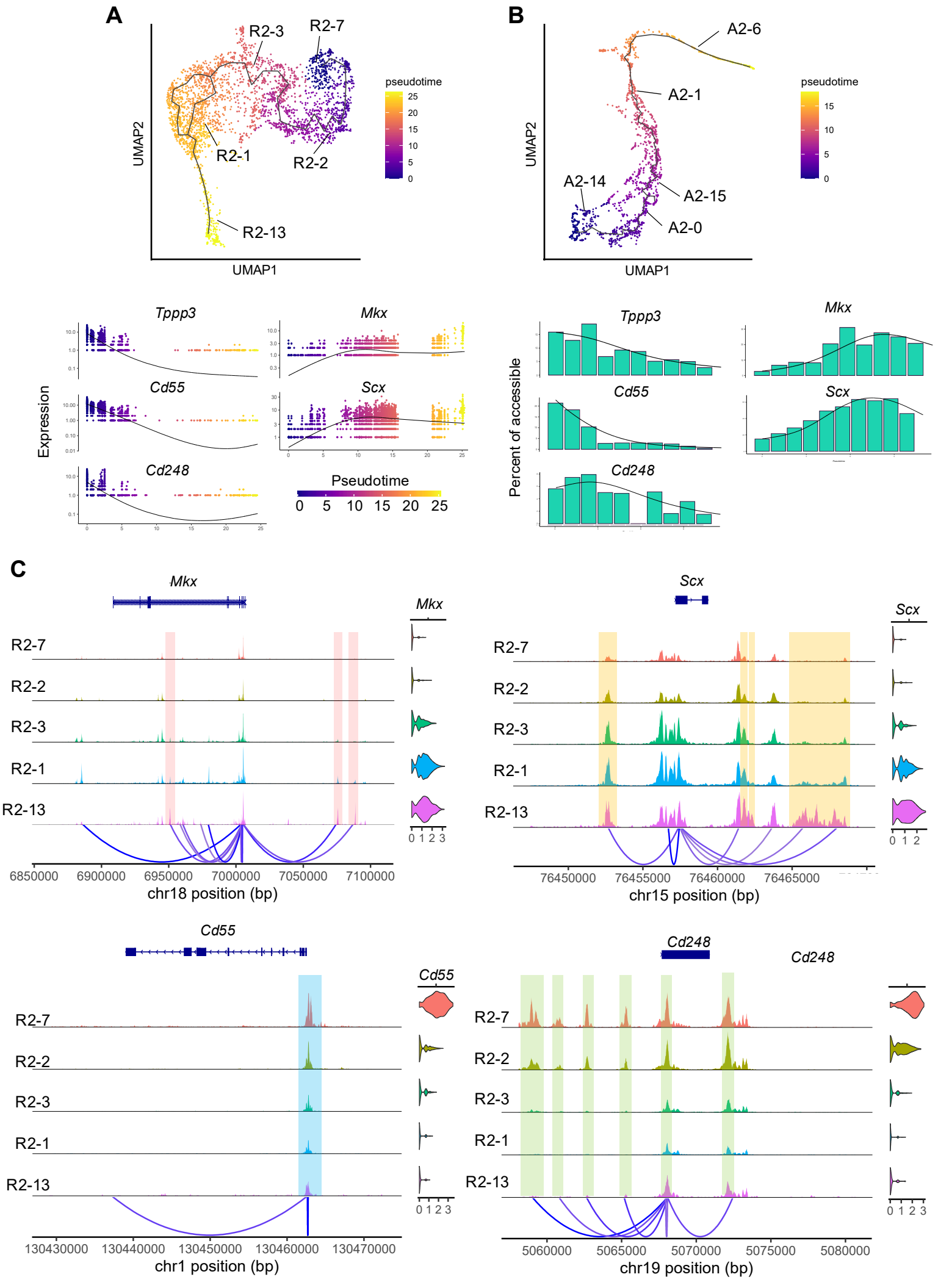

A
R2-3
R2-7
R2-1
R2-2
R2-13
UMAP2
UMAP1
pseudotime
25
20
15
10
5
0
B
A2-6
A2-1
A2-14
A2-15
A2-0
UMAP2
UMAP1
pseudotime
15
10
5
0
Tppp3
Mkx
Cd55
Scx
Cd248
Expression
Pseudotime
0 5 10 15 20 25
Tppp3
Mkx
Cd55
Scx
Cd248
Percent of accessible
C
Mkx
R2-7
R2-2
R2-3
R2-1
R2-13
6850000 6900000 6950000 7000000 7050000 7100000
chr18 position (bp)
0 1 2 3
Scx
R2-7
R2-2
R2-3
R2-1
R2-13
76450000 76455000 76460000 76465000
chr15 position (bp)
0 1 2
Cd55
R2-7
R2-2
R2-3
R2-1
R2-13
130430000 130440000 130450000 130460000 130470000
chr1 position (bp)
0 1 2 3
Cd248
R2-7
R2-2
R2-3
R2-1
R2-13
5060000 5065000 5070000 5075000 5080000
chr19 position (bp)
0 1 2 3

Figure 4

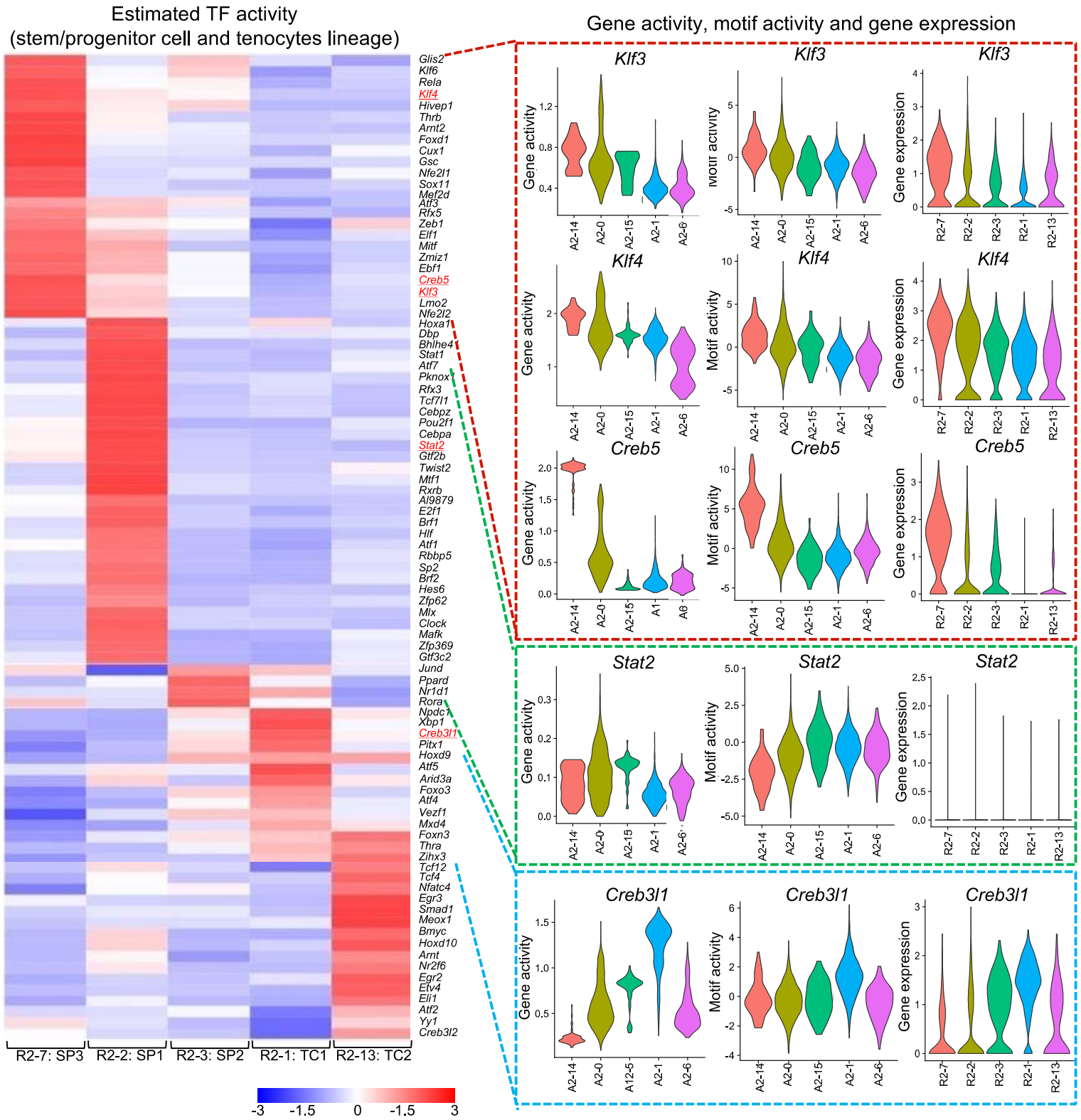

Estimated TF activity
(stem/progenitor cell and tenocytes lineage)
Glis2
Klf6
Rela
Klf4
Hivep1
Thrb
Arnt2
Foxd1
Cux1
Gsc
Nfe2l1
Sox11
Mef2d
Atf3
Rfx5
Zeb1
Elf1
Mitf
Zmiz1
Ebf1
Creb5
Klf3
Lmo2
Nfe2l2
Hoxa1
Dbp
Bhlhe4
Stat1
Atf7
Pknox1
Rfx3
Tcf7l1
Cebpz
Pou2f1
Cebpa
Stat2
Gtf2b
Twist2
Mtf1
Rxrb
Al9879
E2f1
Brf1
Hlf
Atf1
Rbbp5
Sp2
Brf2
Hes6
Zfp62
Mlx
Clock
Mafk
Zfp369
Gtf3c2
Jund
Ppard
Nr1d1
Rora
Npdc1
Xbp1
Creb3l1
Pitx1
Hoxd9
Atf5
Arid3a
Foxo3
Atf4
Vezf1
Mxd4
Foxn3
Thra
Zihx3
Tcf12
Tcf4
Nfatc4
Egr3
Smad1
Meox1
Bmyc
Hoxd10
Arnt
Nr2f6
Egr2
Etv4
Eli1
Atf2
Yy1
Creb3l2
R2-7: SP3
R2-2: SP1
R2-3: SP2
R2-1: TC1
R2-13: TC2
-3
-1.5
0
-1.5
3
Gene activity, motif activity and gene expression
Klf3
Klf4
Creb5
Stat2
Creb3l1
Gene activity
Motif activity
Gene expression
A2-14
A2-0
A2-15
A2-1
A2-6
A1
A6
A12-5
R2-7
R2-2
R2-3
R2-1
R2-13

Figure 5

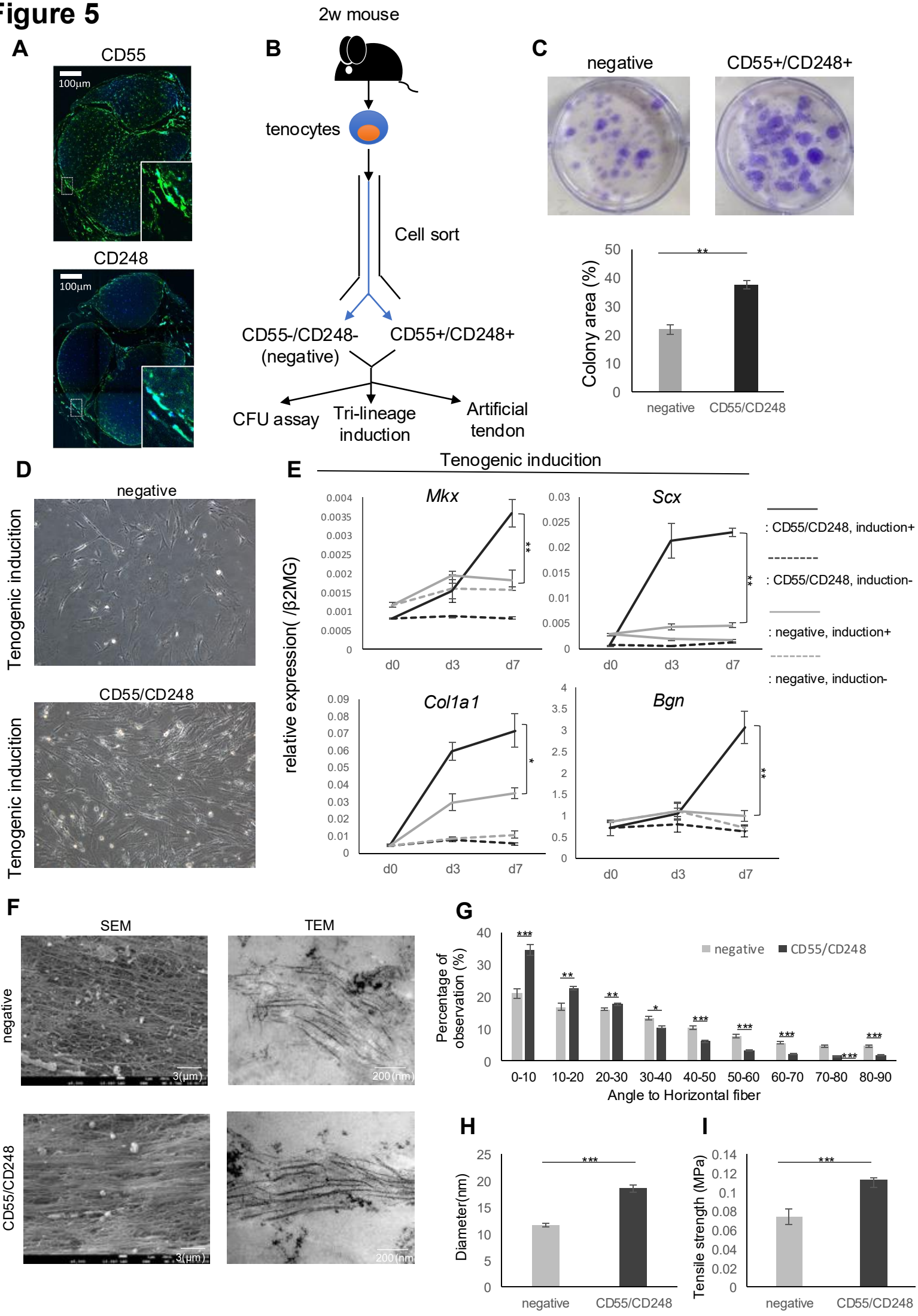

A
CD55
100μm
CD248
100μm
B
2w mouse
tenocytes
Cell sort
CD55-/CD248-
(negative)
CD55+/CD248+
CFU assay
Tri-lineage
induction
Artificial
tendon
C
negative
CD55+/CD248+
Colony area (%)
50
40
30
20
10
0
**
negative
CD55/CD248
D
negative
Tenogenic inducition
CD55/CD248
Tenogenic inducition
E
Tenogenic inducition
relative expression(/β2MG)
Mkx
Scx
Col1a1
Bgn
d0
d3
d7
: CD55/CD248, induction+
: CD55/CD248, induction-
: negative, induction+
: negative, induction-
F
SEM
TEM
negative
CD55/CD248
3(μm)
200(nm)
G
Percentage of observation (%)
negative
CD55/CD248
0-10
10-20
20-30
30-40
40-50
50-60
60-70
70-80
80-90
Angle to Horizontal fiber
H
Diameter(nm)
negative
CD55/CD248
I
Tensile strength (MPa)
negative
CD55/CD248

Table1

Estimated TF activity : Cd55

TF	spearman correlation
Klf4	0.264195541
Klf3	0.262413411
Zeb1	0.219593135
Pbx1	0.204085218
Hic1	0.202481054
Klf2	0.196232716
Nfe2l2	0.196076187

Estimated TF activity : Cd248

TF	spearman correlation
Hic1	0.375565336
Klf3	0.277768834
Zmiz1	0.272991765
Klf4	0.260526766
Zeb1	0.250583629
Sp3	0.2502976
Kdm5b	0.193229132

Estimated TF activity : Mkx

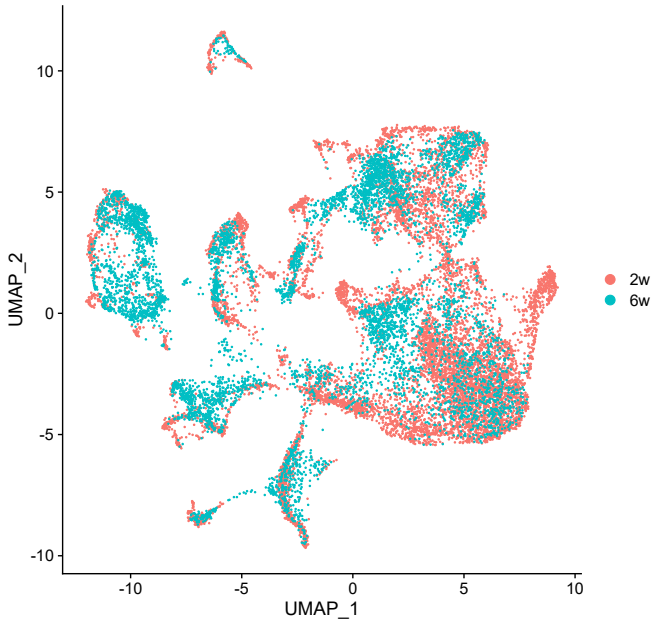
TF	spearman correlation
Creb3l1	0.320638653
Zfhx3	0.169341424
Mxi1	0.149048856
Bhlhe40	0.144301452
Ets2	0.137445497
Max	0.116134957
Elk3	0.098325126

Estimated TF activity : Scx

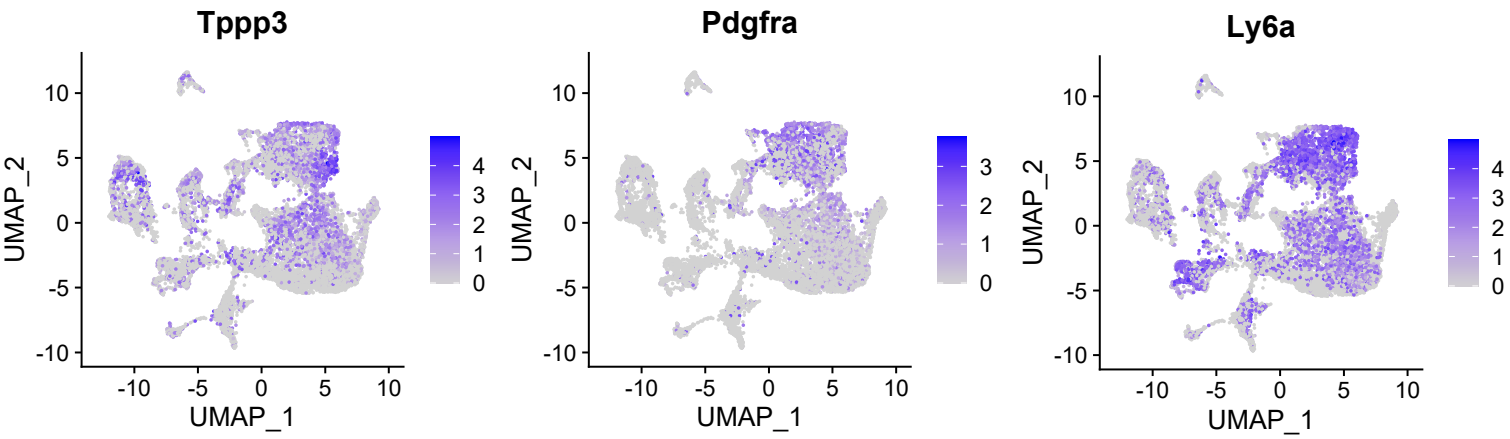
TF	spearman correlation
Creb3l1	0.316522613
Npdc1	0.163452737
Ets2	0.145463784
Mix1	0.121248221
Bhlhe41	0.115930755
Elk3	0.10803805
Bmyc	0.104968457

Supplemental figure 1

A

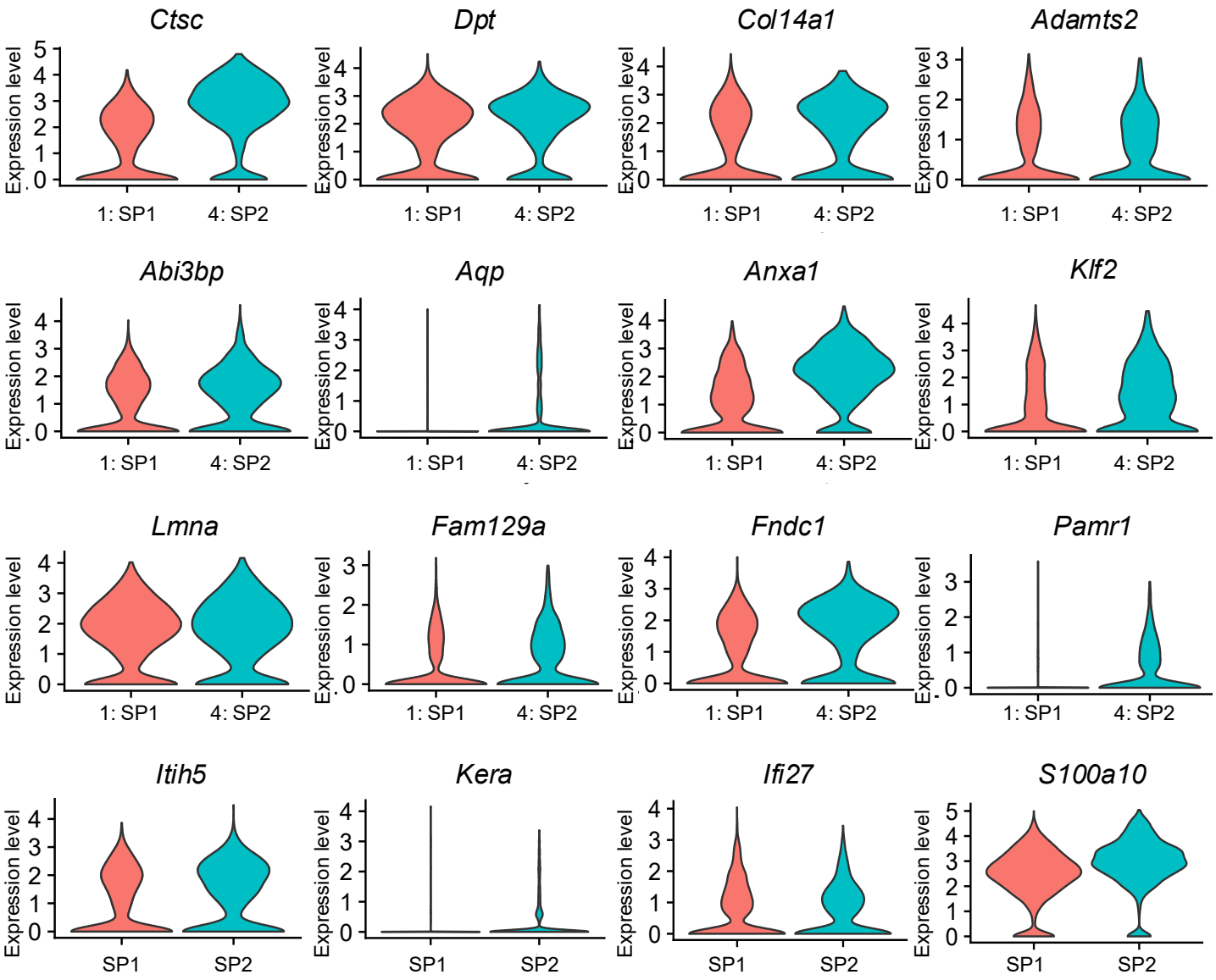


B

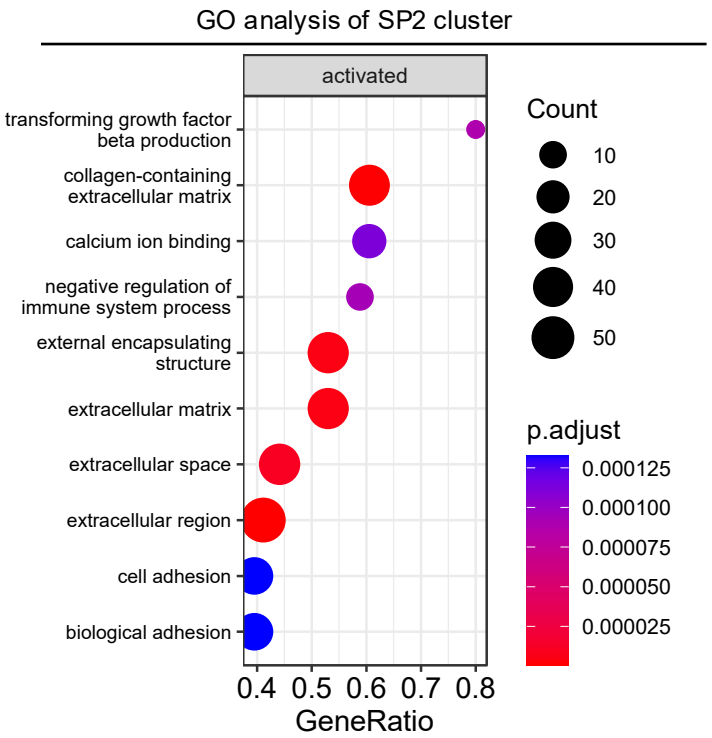


Supplemental Figure 2

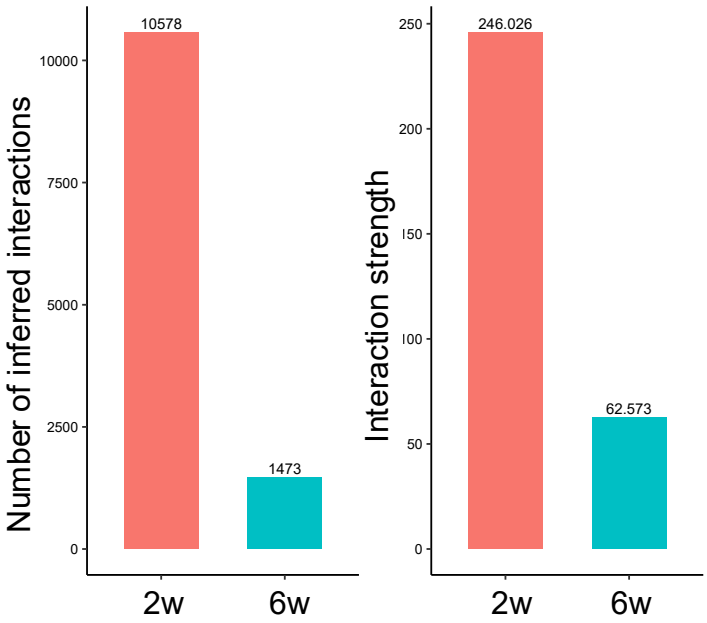
A



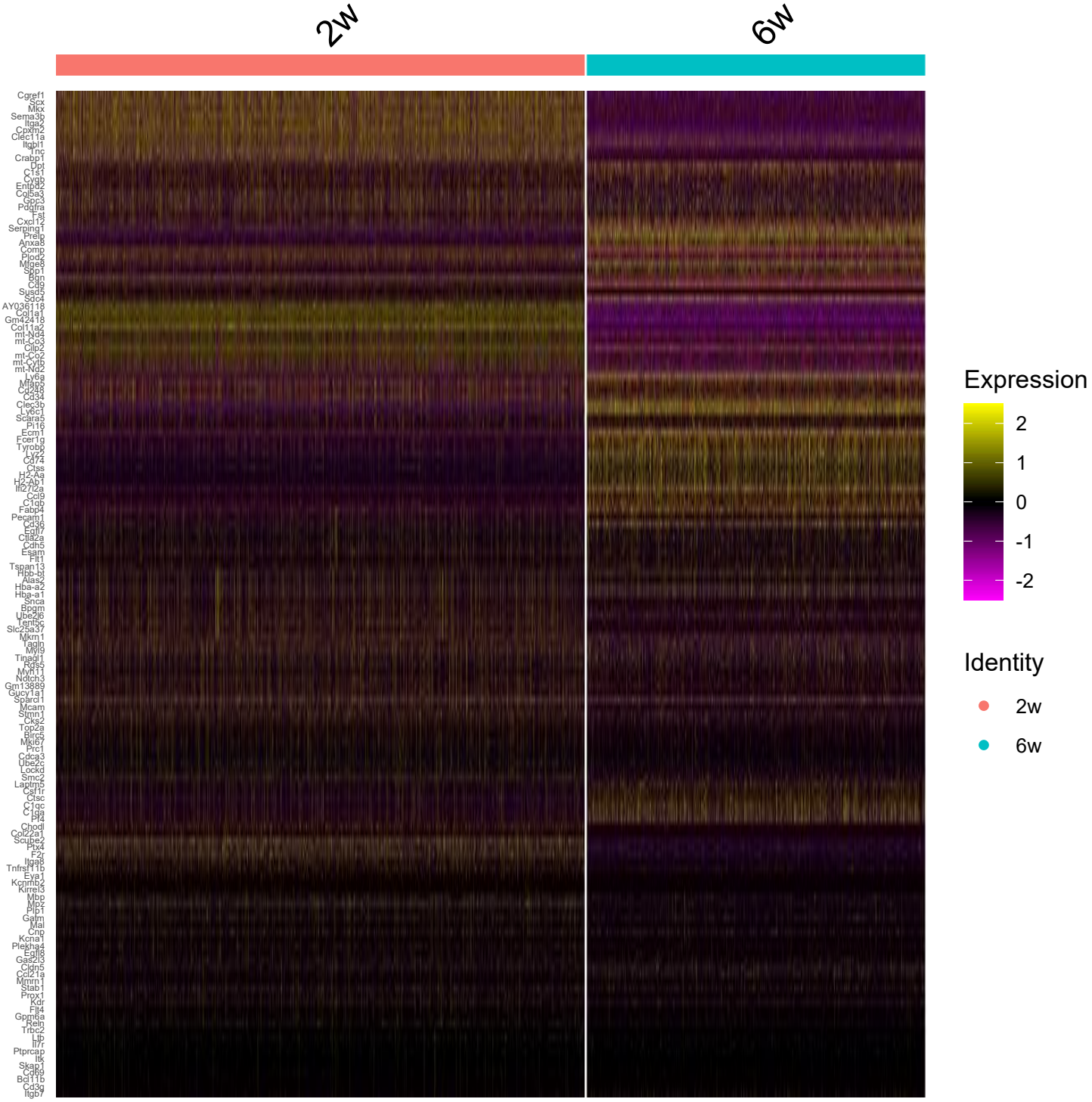
B



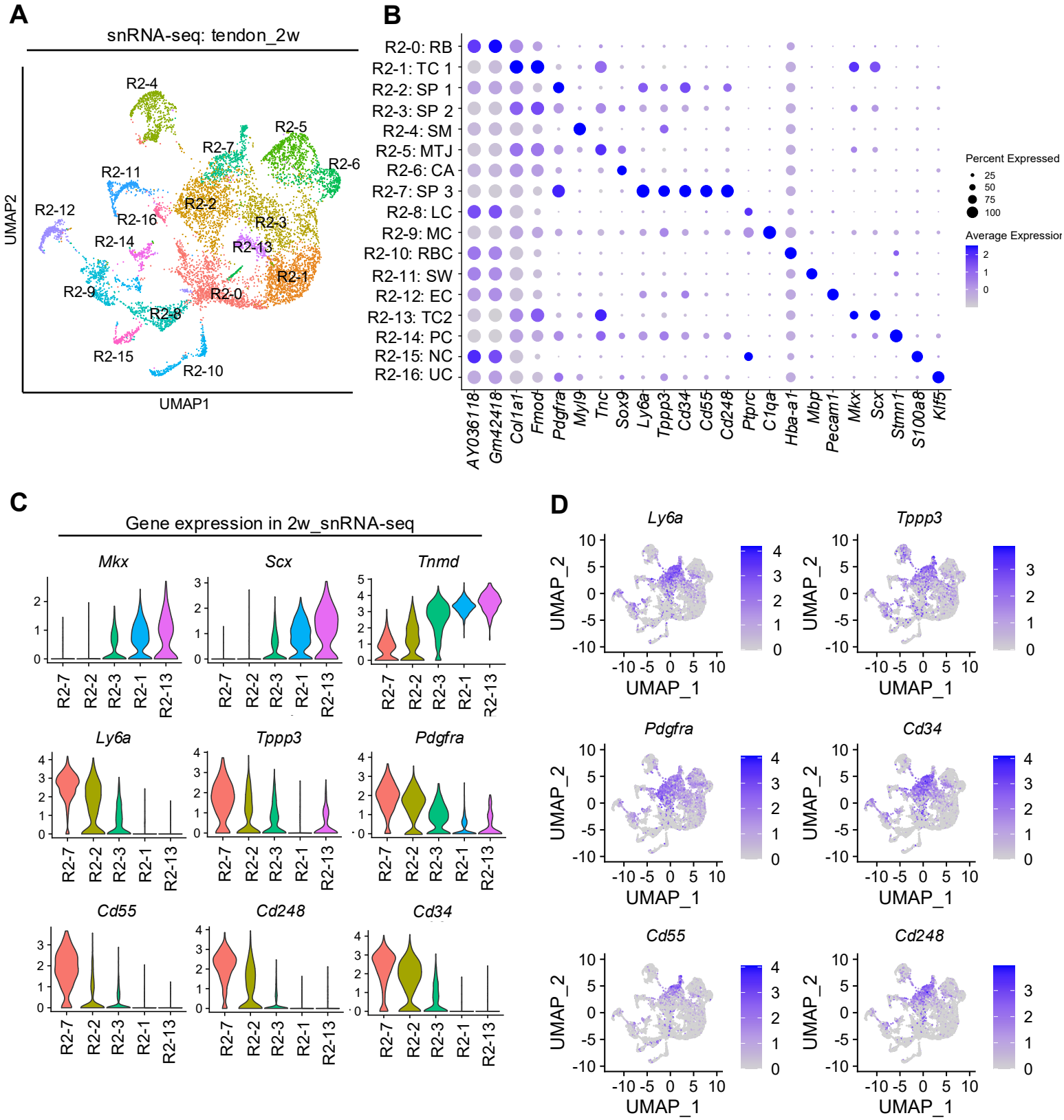
C



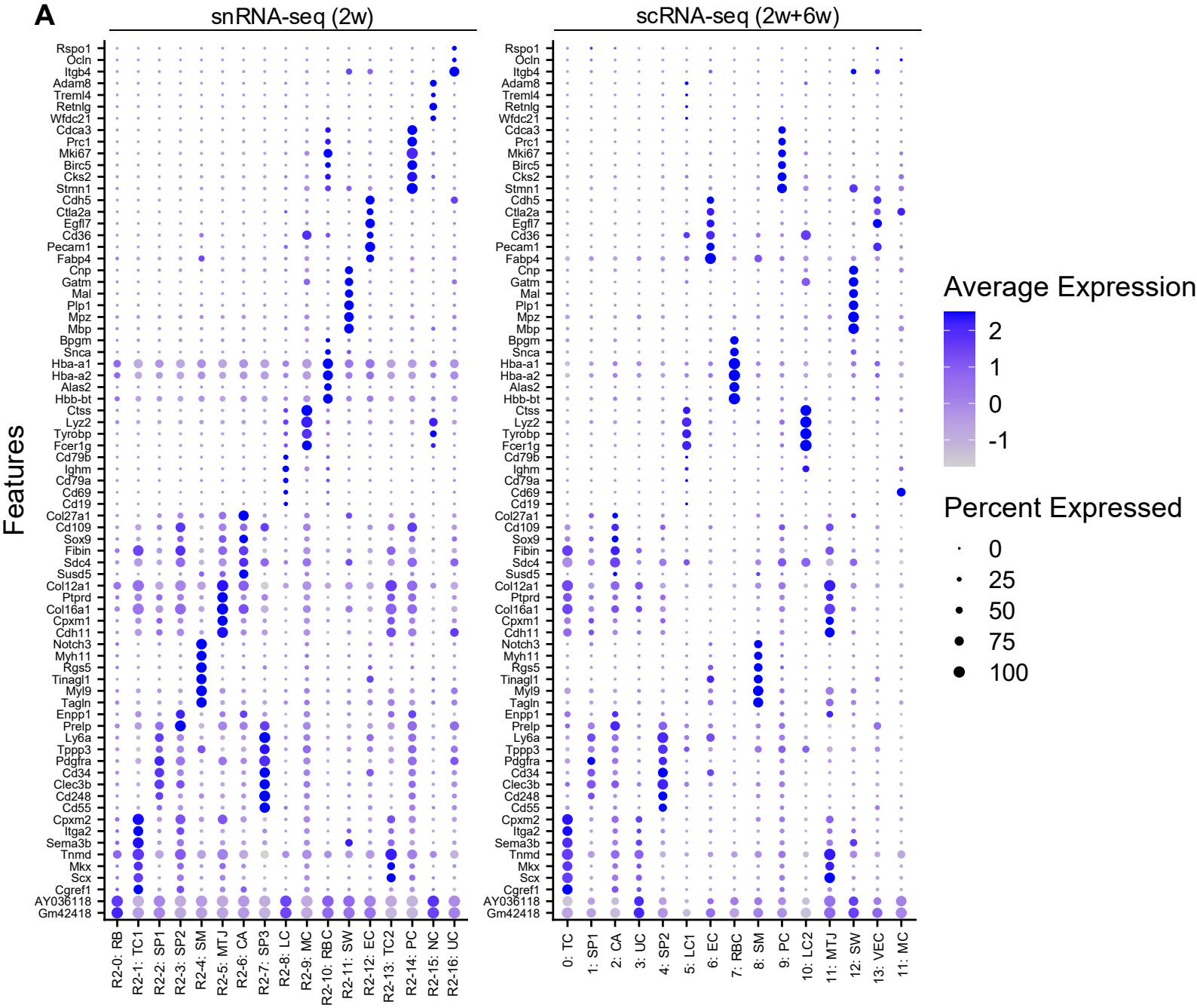
Supplemental Figure 3

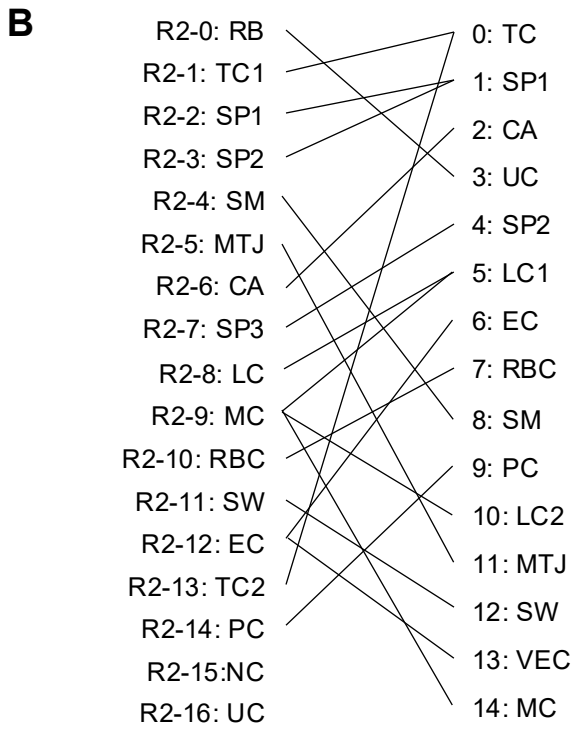


Supplemental Figure 4

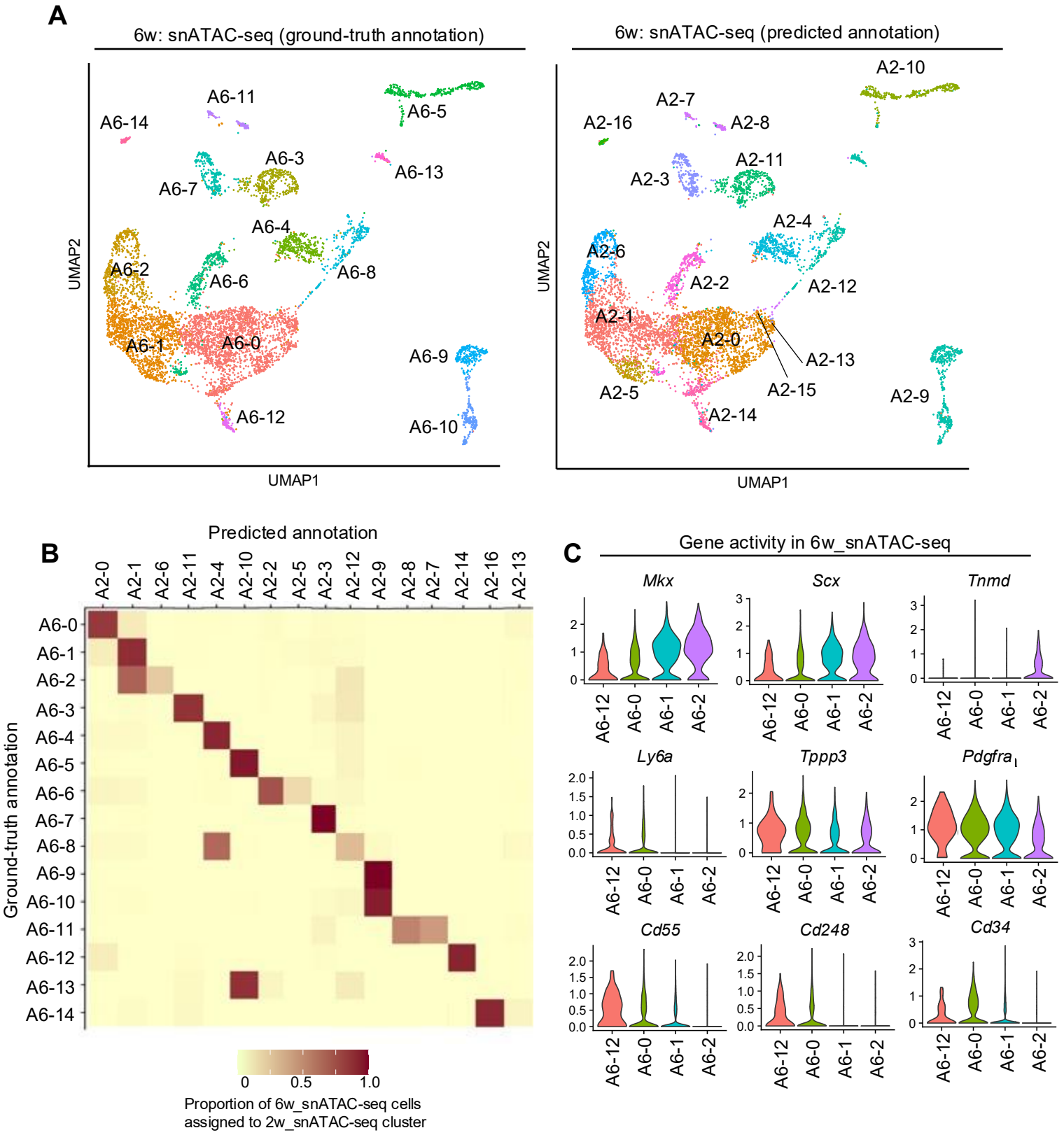


Supplemental Figure 5

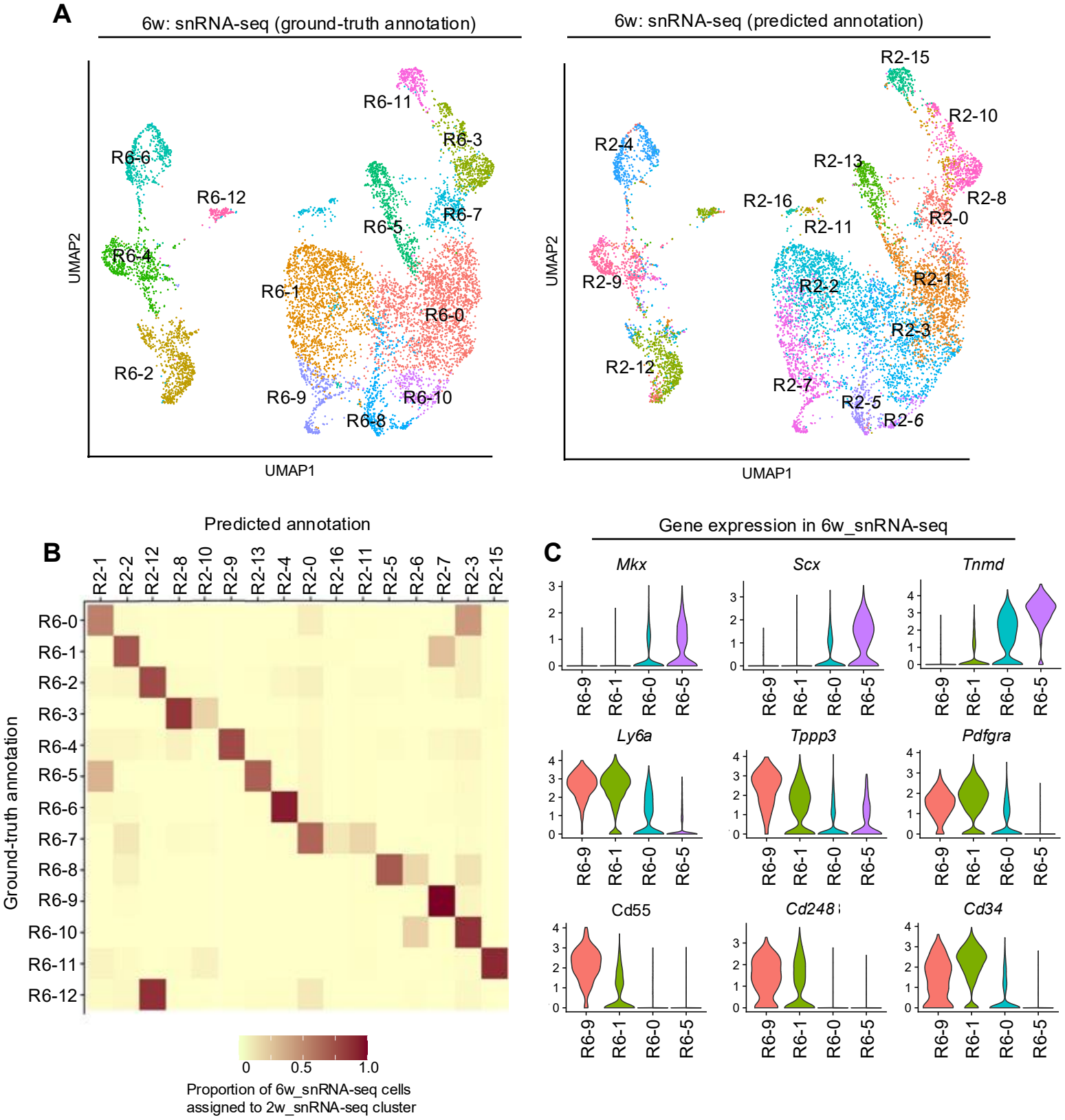




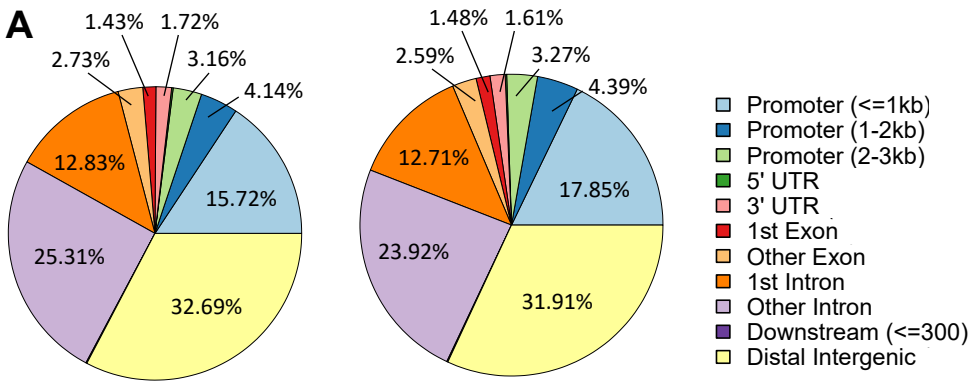
Supplemental Figure 6

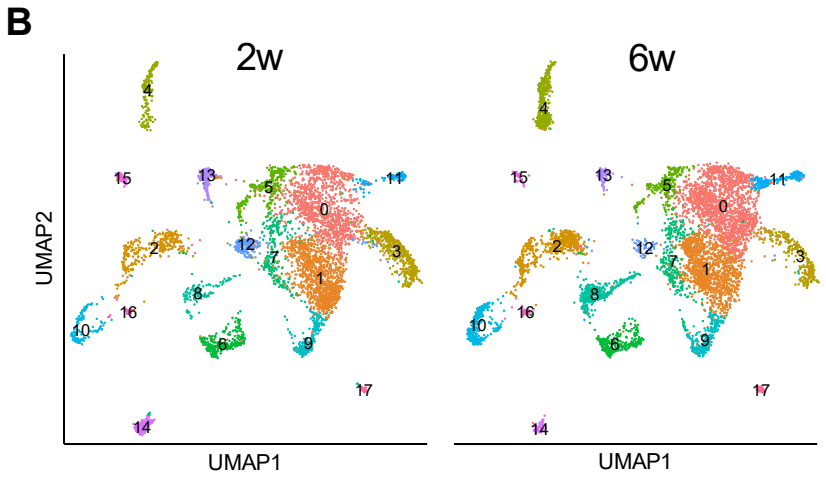


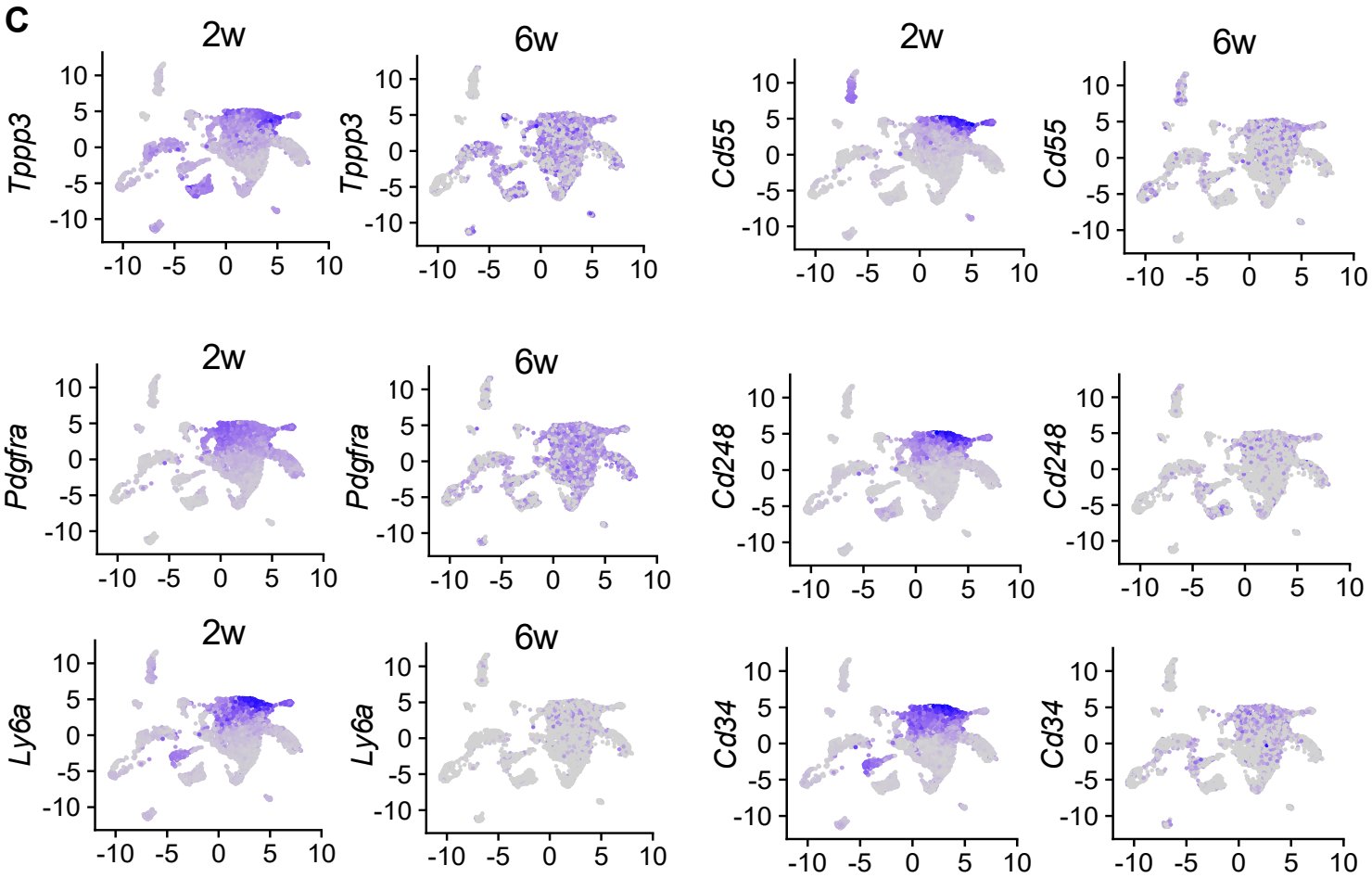
Supplemental Figure 7



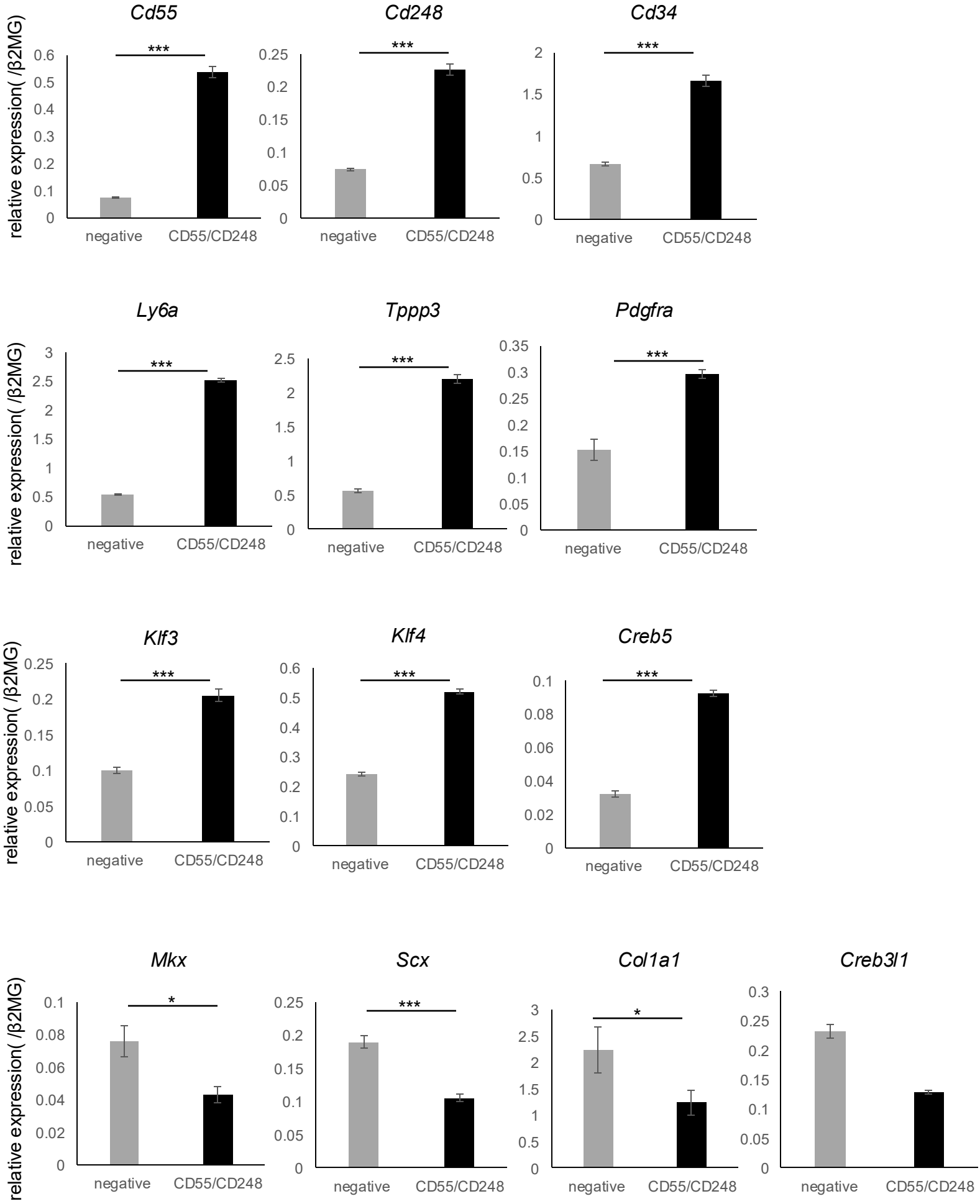
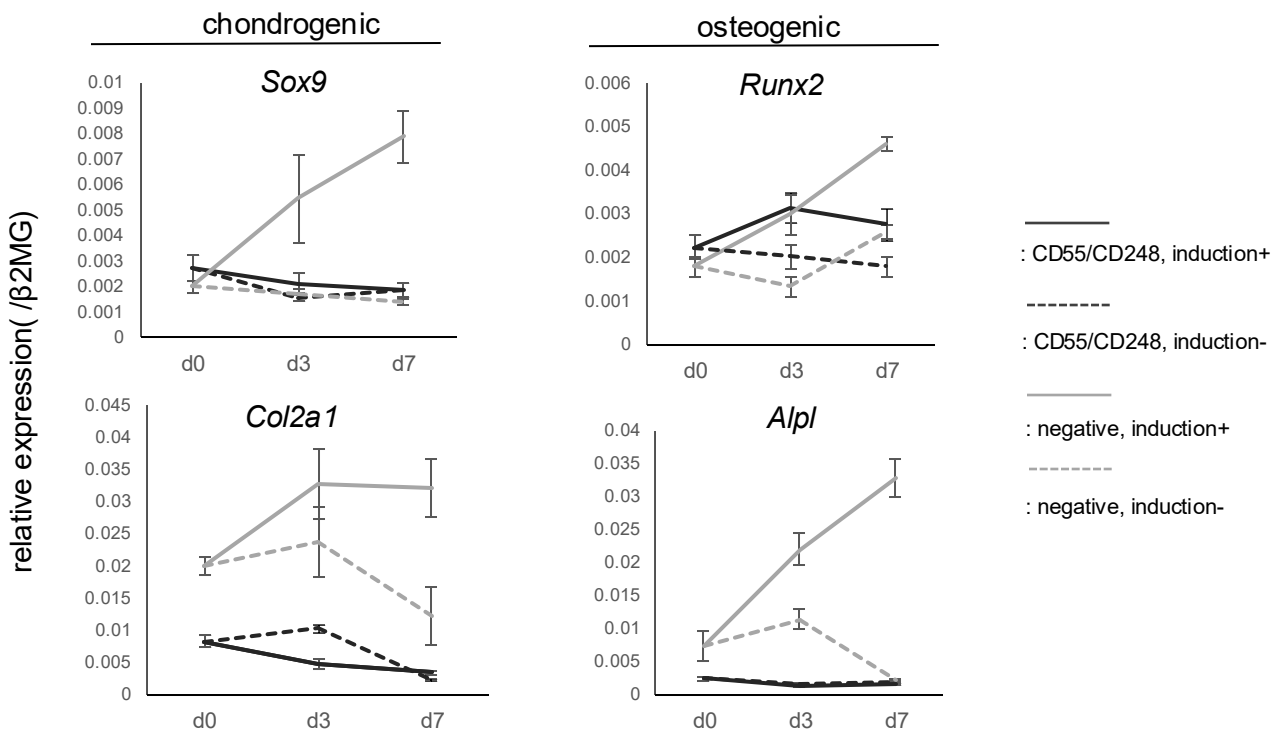
Supplemental Figure 8







Supplemental figure 9



Supplemental Figure 10