

**Title:** - *The transcription factor scleraxis differentially regulates gene expression in tenocytes isolated at different developmental stages*

**Paterson, Y.Z.<sup>1,2\*</sup>, Evans, N.<sup>2</sup>, Kan, S.<sup>2</sup>, Cribbs, A.<sup>3</sup>, Henson, F.M.D.<sup>1,2</sup> and Guest, D.J.<sup>2</sup>**

<sup>1</sup> Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

<sup>2</sup> Centre for Preventive Medicine, Animal Health Trust, Newmarket, UK.

<sup>3</sup> Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford UK

\*Corresponding author

**Author contact information: -**

Ms Yasmin Zoe Paterson

Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge, UK  
CB3 0ES

Tel: +44(0)1638 751000

Email: [yzp20@cam.ac.uk](mailto:yzp20@cam.ac.uk)

Ms Naomi Evans

Centre for Preventive Medicine, Animal Health Trust, Lanwades Park, Kentford, Newmarket, Suffolk,  
CB8 7UU, UK

Tel: +44(0)1638 751000

Email: [nevans5@rvc.ac.uk](mailto:nevans5@rvc.ac.uk)

Mr Shohei Kan

Centre for Preventive Medicine, Animal Health Trust, Lanwades Park, Kentford, Newmarket, Suffolk,  
CB8 7UU, UK

Tel: +44(0)1638 751000

Email: [skan3@sheffield.ac.uk](mailto:skan3@sheffield.ac.uk)

Dr Adam Cribbs

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of  
Oxford, Oxford UK

OX3 7LD

Tel: +44(0)1865 227 374

Email: [adam.cribbs@ndorms.ox.ac.uk](mailto:adam.cribbs@ndorms.ox.ac.uk)

Dr Frances M. D. Henson

Centre for Preventive Medicine, Animal Health Trust, Lanwades Park, Kentford, Newmarket, Suffolk,  
CB8 7UU, UK

Tel: +44(0)1638 751000

Email: [fmdh1@cam.ac.uk](mailto:fmdh1@cam.ac.uk)

Dr Deborah Jane Guest

Centre for Preventive Medicine, Animal Health Trust, Lanwades Park, Kentford, Newmarket, Suffolk,  
CB8 7UU, UK

Tel: +44(0)1638 751000

Email: [debbie.guest@aht.org.uk](mailto:debbie.guest@aht.org.uk)

## **Abstract**

The transcription factor scleraxis (SCX) is expressed throughout tendon development and plays a key role in directing tendon wound healing. However, little is known regarding its role in fetal or young postnatal tendons, stages in development that are known for their enhanced regenerative capabilities. Here we used RNA-sequencing to compare the transcriptome of adult and fetal tenocytes following SCX knockdown. SCX knockdown had a larger effect on gene expression in fetal tenocytes, effecting 477 genes in comparison to the 183 genes effected in adult tenocytes, indicating that scleraxis-dependent processes may differ in these two developmental stages. Gene ontology, network and pathway analysis revealed an overrepresentation of extracellular matrix (ECM) remodelling processes within both comparisons. These included several matrix metalloproteinases, proteoglycans and collagens, some of which were also investigated in SCX knockdown tenocytes from young postnatal foals. Using chromatin immunoprecipitation, we also identified novel genes that SCX differentially interacts with in adult and fetal tenocytes. These results indicate a role for SCX in modulating ECM synthesis and breakdown and provides a useful dataset for further study into SCX gene regulation.

## **Keywords:**

Scleraxis; Gene Regulation; Tendon Development; Transcriptomics; Knockdown; Chromatin Immunoprecipitation

## 1. Introduction

Tendon injuries and tendinopathies occur commonly, accounting for approximately 30% of musculoskeletal diseases in humans (Nourissat et al., 2015) and 46% of all limb injuries in racehorses (Williams et al., 2010). Tendons are composed of dense connective tissue that serves to transmit contractile forces from skeletal muscle to bone. Force transmitting tendons such as the human Achilles tendon (AT) and the equine superficial digital flexor tendon (SDFT), share remarkable similarities in terms of their composition and function (Patterson-Kane et al., 2012; Patterson-Kane and Rich, 2014) and are predominantly composed of highly organised collagen I fascicles with other collagens, proteoglycans, elastin and resident tendon cells (tenocytes) making up the remainder of the extracellular matrix (ECM). Following a tendon injury, the healing process is very slow and results in scarring, which is characterised by a disorganised fibrovascular matrix consisting of abnormal quantities of collagen III (Yang et al., 2013). As such tendon functionality is permanently compromised and chronic re-injury often occurs. At present, although stem cell and tissue engineering techniques offer significant potential for tendon regeneration, few effective treatments are currently available (Goldin and Malanga, 2013). The lack of effective treatments is, in part, due to our limited knowledge on the exact nature of resident tenocytes behaviour during tendon repair. Further understanding of this could be instrumental in the development of pro-regenerative tendon therapies. The use of the horse as a model for translational orthopaedic research has been well described as highlighted in guidance issued by authorities such as the US Food and Drugs Administration (FDA) and European Medicine Agency (EMA) (Hillmann et al., 2016; Patterson-Kane and Rich, 2014).

At present scleraxis (SCX), a basic helix-loop-helix transcription factor, is the most widely studied tendon marker, upon which much of the information regarding tendon development and tenocyte behaviour has been built (Schweitzer et al., 2001). Although not exclusively expressed in tendons, SCX has an indispensable role in the development of force transmitting tendons and is a key regulator in tenocyte differentiation of embryonic stem cells (ESCs) and mesenchymal stromal cells (MSCs) (Bavin et al., 2017; Li et al., 2015; Murchison et al., 2007). SCX expression is regulated through TGF- $\beta$  signalling (Murchison et al., 2007; Pryce et al., 2009) and via mechanotransduction (Eliasson et al., 2009; Maeda et al., 2011) and has been shown to be significantly upregulated following injury, playing

a central role in tendon healing in adult (Barsby and Guest, 2013; Sakabe et al., 2018; Scott et al., 2011) and neonatal animals (Howell et al., 2017). However, less is known about the role of SCX in fetal development and in young, postnatal tendons. Tendon tissue continues to grow and develop, with fibrillogenesis and remodelling continuing after birth (Connizzo et al., 2013; Theodossiou and Schiele, 2019). Fetal tissues including tendons have also been shown to exhibit regenerative capacities following injury (Ansorge et al., 2012; Beredjiklian et al., 2003; Cox et al., 2014; Favata et al., 2006; Porrello et al., 2011; Yates et al., 2012) and young animals generally undergo better tissue regeneration than mature animals (Ballas and Davidson, 2001; Chen et al., 2016; Iismaa et al., 2018). However, it has not yet been demonstrated if SCX has alternative roles in different stages of development and ageing.

How SCX functions to enable tenocyte differentiation and ECM remodelling during tendon healing remains unclear, and there is limited information on its downstream regulatory effects. Currently *COL1A1* and *TNMD* are the most well documented genes which have been shown to be directly regulated by SCX in adult tenocytes (Léjard et al., 2007; Shukunami et al., 2006). However, emerging evidence suggests that SCX has a key role in driving extracellular matrix (ECM) production during development and remodelling in other tissue types including the heart and periodontal ligaments (Czubryt, 2014; Roguljic et al., 2013). In cardiac fibroblasts, genes including *COL1A2*, *VIM*, *Snai1*, *Twist1*, *MMP2* and *FN1* have been identified as direct targets of SCX (Bagchi et al., 2016b, 2016a; Bagchi and Czubryt, 2012; Nagalingam et al., 2018) but the large-scale identification of SCX regulated genes is still lacking.

Our group has also shown that knocking down SCX expression in adult equine tenocytes has no effect on their ability to re-organise a three-dimensional (3D) matrix to generate artificial tendons, whereas SCX knock-down in fetal tenocytes completely prevents their ability to form 3D tendons (Bavin et al., 2017). SCX knockdown also caused considerably different effects on tendon related gene expression in fetal and adult tenocytes, further suggesting that SCX regulates gene expression differently during different developmental stages. In this study we expand on this work by taking a global RNA-sequencing approach to analyse the transcriptome of equine adult and fetal tenocytes following SCX knockdown. This identified previously unreported genes that are differentially expressed in adult and fetal tenocytes



as a result of SCX depletion, some of which were compared to tenocytes from young post-natal foals. We further used chromatin immunoprecipitation to identify genes that are directly under SCX control. Collectively, these results help to increase our understanding of SCX regulation during tendon development, which may help inform future therapeutic approaches for tendon repair.

## 2. Results

### *2.1 Scleraxis knockdown by lentiviral-delivered short hairpin RNA (shRNA) produces stable, consistent knockdown in adult and fetal tenocytes*

In order to identify downstream target genes of SCX a retroviral delivery system to express a shRNA was used to generate stable SCX knockdown lines (shSCX) or non-target controls (NT) in four biological replicates of fetal and adult tenocytes. The average percentage of SCX mRNA knockdown was 72.6% in adult tenocytes and 80% in fetal tenocytes, which was not significantly different (p-value = 0.336) (Fig.1A). Viral copy number integration was measured in each line of shSCX and NT expressing tenocytes and approximately one copy number event was detected per diploid cell across all cell lines (Fig.1B). Supplementary Fig.1A and B highlight the consistency in percentage knockdown and copy integrations achieved between biological replicates. A reduction in SCX protein expression was also confirmed using immunocytochemistry (Fig.1C). In some lines, SCX protein was reduced but still detectable, in other lines SCX protein was no longer detectable (Supplementary Fig.1C).

### *2.2 Scleraxis knockdown leads to differential changes in gene expression in adult and fetal tenocytes*

RNA-sequencing was performed to measure transcriptional changes in four lines of adult and four lines of fetal tenocytes following SCX knockdown. A total of 13,159 and 13,142 genes were detectable (out of 21,688) in the adult and fetal tenocytes respectively. Of these, 183 genes were differentially expressed using a  $\text{Log}_2\text{FC} \pm 1$   $p.\text{adj} < 0.05$  between adult shSCX and NT controls, with 120 genes being upregulated and 63 genes being downregulated (Fig.2A). In comparison, 477 genes were differentially expressed between fetal shSCX and NT controls, with 358 genes being upregulated and 119 genes being downregulated (Fig.2A). Of these differentially expressed genes only 117 were commonly

differentially regulated as a result of SCX knockdown in both adult and fetal tenocytes lines, with 87 genes being upregulated and 30 genes being downregulated. *MMP3* was found to be the most significantly upregulated gene common to both groups. *IGF2BP1* was the most significantly downregulated gene common to both adult and fetal groups following SCX knockdown, however *CLDN16* was the most significantly downregulated gene when looking at the effects of shSCX in adult tenocytes alone. The results of the differential gene expression analysis can be visualized in the volcano plots in Fig.2B.

To validate some of the potential direct or indirect targets of SCX, qPCR analysis was conducted using both the original shSCX and NT sequenced lines alongside an additional cohort of two adult and two fetal shSCX and NT tenocyte biological replicates (Fig.2C&D). Of the eight genes tested SCX, *MMP2* and *MMP3* were significantly differentially expressed as a result of SCX knockdown in both adult and fetal tenocytes, corroborating the RNA-sequencing results (Fig.2C&D). In the fetal tenocytes, *MMP9*, *COL14A1* and *COL1A2* had significant adjusted p-values following RNA-sequencing however upon analysis of further biological replicates these genes were no longer significant, whereas *VIM* reached significance with further biological replicates (Fig.2C&D). All of the RNA-sequencing data for the adult tenocytes was corroborated by the qPCR analysis. Overall, for all genes in both adult and fetal tenocytes, there was a 75% corroboration between the RNA-sequencing data and the qPCR analysis of the larger cohort.

It has previously been demonstrated that SCX either directly or indirectly regulates the expression of several collagens, proteoglycans, matrix metalloproteinases and other transcription factors in both tendon and heart tissue (Al-Hattab et al., 2018; Bagchi et al., 2016a, 2016c, 2016b; Bagchi and Czubryt, 2012; Czubryt, 2014; Havis et al., 2016; Léjard et al., 2007; Nagalingam et al., 2018; Nichols et al., 2018). Using the aforementioned literature available on SCX regulation, a panel of 16 genes previously demonstrated to be affected by SCX knockdown in either tendon or cardiac fibroblasts was assessed (Fig.2E), excluding those already investigated in the qPCR validation (Fig.2C&D). In fetal tenocytes 10 of the 16 genes were also found to significantly differ as a result of SCX knockdown ( $p_{adj} < 0.05$ ). In adult tenocytes only one gene, *MMP1*, was found to significantly differ following SCX knockdown.

*2.3 Gene ontology and pathway analysis reveals genes that are differentially expressed following SCX knockdown are involved in different biological processes in adult and fetal tenocytes*

To determine the biological processes altered as a result of SCX knockdown gene ontology (GO) analysis was performed. The top ten significantly enriched terms are listed in Fig.3A. In adult tenocytes SCX knockdown resulted in over-representation of genes involved in the regulation of responses to endogenous stimulus, stress, signal transduction and organic substrates. These included several mRNAs from the growth factor protein class as well as g-protein coupled receptors, and membrane traffic proteins. In fetal tenocytes SCX knockdown resulted in over-representation of genes involved in cellular migration and motility as well as tissue development and structure morphogenesis. These included several mRNAs from the growth factor protein class as well as ECM structural proteins, metalloproteases and membrane bound signalling molecules. In addition to GO analysis differentially expressed genes were also overlaid into the GeneAnalytics Pathway Analysis tool (Fig.3B). Genes differentially expressed following SCX knockdown in adult tenocytes resulted in significant enrichment of pathways including metabolism and ECM remodelling. Genes differentially expressed following SCX knockdown in fetal tenocytes resulted in significant enrichment of pathways including the TGF-Beta and differentiation pathways.

The STRING plugin from Cytoscape was then used to predict interactions between the top differentially expressed genes resulting from SCX knockdown, using a less stringent cut off of  $\text{Log}_2\text{FC} \pm 0.6$   $p_{\text{adj}} < 0.05$ . This approach allowed us to encompass a larger number of differentially expressed genes and prevented bias in the network by imposing too high a fold change cut-off (Evans, 2015). This cut-off resulted in a total of 465 differentially expressed genes as a result of shSCX expression in the adult tenocytes (305 upregulated and 160 downregulated) and 1149 genes in the fetal tenocytes (803 upregulated and 346 downregulated).

In the adult network 423 genes were recognised by the STRING software which resulted in a network of 332 connected genes and 91 singletons. Within this network the first to third direct neighbouring interactors of SCX can be found in Fig.3C, with the cytokine CCL26 representing the first direct neighbouring interaction. The top five functional annotations which are common throughout the network

are depicted by the coloured bars surrounding each node and include responses to stimulus and cell migration (Fig.3D).

In the fetal network 1060 genes were recognised by the STRING software which resulted in a network of 1001 connected genes and 59 singletons. Within this network the first and second direct neighbouring interactors of SCX can be found in Fig.3E. In this case only the first and second neighbours were considered in order to reduce the overall size of the network and allow for easier visualisation. This resulted in SOX9, BGN, DCN, GDF6 and CCL26 (the only direct neighbour found in the adult network) representing the first direct neighbouring interactions. The top five functional annotations which are common throughout the fetal network do not overlap with the adult network and include extracellular matrix organisation and skeletal system development (Fig.3F).

#### *2.4 Foal tenocytes are differentially affected by scleraxis knockdown compared to both adult and fetal tenocytes*

Young postnatal foal tenocytes (aged 54 days – 84 day postpartum) were examined to determine if target genes of SCX are differentially regulated at this developmental stage. Foal tenocytes were capable of remodelling an artificial 3D collagen gel with no significant difference in contraction rate or cell survival observed compared to adult or fetal tenocytes (Fig.4A). Foal tenocytes also express a panel of tendon associated markers with no significant differences compared to adult or fetal cells (Fig.4B).

SCX expression was knocked down by an average of 79.3% in the foal tenocytes (Fig.4C). SCX knockdown in foal tenocytes did not affect their ability to contract a collagen gel or their cell survival within the collagen gels (Fig.4D). Expression of 11 genes were measured by qPCR in shSCX and NT foal tenocytes. Only *COMP* was shown to be significantly differentially expressed in the foal tenocytes (Fig.4E). *TNC* and *MMP3* were consistently upregulated to a high degree in all foal replicates (4.76 and 14.66 average fold change relative to NT control), however no significant changes were detected due to the large variation in fold change increase between the biological replicates. Interestingly, of the

genes examined *COMP* and *TNC* were not differentially expressed in either fetal or adult tenocytes (Fig.4F).

## 2.5 The direct target genes of SCX are different in adult and fetal tenocytes

Chromatin immunoprecipitation by qPCR (ChIP-qPCR) demonstrated significant enrichment of SCX binding to the *COL1A2* promoter in both adult and fetal tenocytes (Fig.5) despite this gene not being differentially expressed following SCX knockdown (Fig.2D). *IGF2BP1*, which was the top downregulated gene in both adult and fetal tenocytes following SCX knockdown was significantly enriched in both adult and fetal tenocytes, however at different E-box binding sites (*IGF2BP1* #1 in fetal tenocytes and *IGF2BP1* #2 in adult tenocytes). *PDGFB*, which was significantly downregulated only in fetal tenocytes following SCX knockdown was significantly enriched in fetal tenocytes. Adult tenocytes also showed substantial enrichment (7.66-fold) however this was not significant due to the high variability between biological replicates. *CLDN16*, which was significantly downregulated only in adult tenocytes following SCX knockdown showed significant enrichment in SCX binding in adult tenocytes but not fetal tenocytes. *ACTA1* (within E-box region #1), which was significantly downregulated in fetal shSCX expressing cells only following differential expression analysis, surprisingly showed a trend of enrichment (>2-fold) in adult tenocytes but not in fetal tenocytes. This enrichment was however not significant. Similarly, *FGF19*, a gene which was significantly downregulated in both adult and fetal shSCX expressing cells showed a trend of enrichment (>2-fold) in fetal tenocytes alone, yet this was not significant. *KLF15* and *NOV*, which were significantly downregulated in both fetal and adult tenocytes and *MMP3* (within E-box region #2) which was significantly upregulated in fetal and adult tenocytes, also showed a trend of enrichment (>2-fold) in both adult and fetal tenocytes however again this was not significant. *ACTA1* (within E-box region #2), *FGF9* and *MMP3* (within E-box region #1 and #3) clearly showed no enrichment of SCX binding. As a control *TNMD*, which has previously been identified as being directly regulated by SCX in adult tenocytes (Shukunami et al., 2018, 2006), showed significant enrichment in both adult and fetal tenocytes.

### 3. Discussion

Scleraxis is a transcription factor required for the development of force transmitting tendons and is heavily implicated in tendon injury repair. Much of what is currently known regarding SCX has been generated from studies of early embryogenesis or in adult tenocytes and few studies have examined its role in fetal or early postnatal tendon, stages in development which are known for their increased regenerative capabilities. Similarly, the downstream transcriptional regulatory network of SCX is poorly defined. Identification of SCX downstream targets at different developmental stages will yield valuable insight into how adult reparative and fetal regenerative tenocytes act to facilitate scleraxis-dependent processes which may provide new targets for therapeutics.

To allow identification of genes downstream of SCX, a short-hairpin RNA against SCX was used which produces a significant reduction in SCX expression with no effect on cellular proliferation or morphology (Bavin et al., 2017). Although SCX knockdown was confirmed both using qPCR and immunocytochemistry, it should be noted that quantitative western blotting would be required to give a more accurate reflection of the degree of SCX knockdown at the protein level. In our study we used multiple SCX knockdown cell lines from different horses for RNA-sequencing. Limitations in using a single biological replicate for RNA-seq experiments have been extensively documented (Baccarella et al., 2018; Lamarre et al., 2018; Son et al., 2018) and are highlighted in Nichols et al., 2018. Here, SCX knockdown in a single adult equine tenocyte line resulted in significant downregulation of genes involved in focal adhesions. However, these genes were no longer significant upon validation in a larger biological cohort (Nichols et al., 2018). In comparison, our study revealed 75% corroboration between the RNA-seq data and cohort qPCR, similar to that which we have observed previously (Paterson et al., 2020).

SCX is postulated to be a key mediator in ECM and collagen production in a number of tissue types (Bagchi et al., 2016b; Czubryt, 2014) and several collagens, proteoglycans and matrix metalloproteinases (MMPs) have been shown to be regulated by SCX (Al-Hattab et al., 2018; Bagchi et al., 2016a, 2016c, 2016b; Bagchi and Czubryt, 2012; Czubryt, 2014; Havis et al., 2016; Léjard et al., 2007; Nagalingam et al., 2018; Nichols et al., 2018). Our results similarly indicate a significant

involvement of SCX in regulating MMPs 1, 2, 3, 12 and 13, collagens 4A1 and 18A1 as well as ECM proteoglycans *VIM* and *LUM*, with pathway analysis also highlighting ECM remodelling and TGF $\beta$  pathway signalling as being significantly overrepresented.

In this study SCX knockdown had a much larger effect on fetal tenocyte gene expression than adult tenocytes. Only 25-35% of the differentially expressed genes in both groups were downregulated, with the majority being upregulated as a result of the knockdown. As SCX has only been reported as being a transcriptional activator (Cserjesi et al., 1995; Furumatsu et al., 2010) (Fig.6), this would suggest that the high proportion of upregulated genes are not direct targets of SCX and are instead regulated by some downstream intermediate. However, this traditional view of classifying transcription factors as “activators” or “repressors” is being questioned, and in many cases “activators” often have an indirect repression effect by blocking the binding of other transcriptional activators (Hobert, 2008; Lambert et al., 2018) (Fig.6). Similarly, this high percentage of upregulated genes may be due to SCX activation of transcriptional repressors, such that when SCX is knockdown this reduced activation of transcriptional repressors may result in de-repression of other genes leading to their upregulation. As SCX has been reported to work with other co-factors (Carlberg et al., 2000; Cserjesi et al., 1995; Furumatsu et al., 2010), the observed differences in downstream genes in fetal and adult stages, may indicate that there is a difference in the availability of such co-factors at the different stages of development.

The fetal tenocytes used in this study were from around 85-90% of the way through gestation. The large effect on gene expression that we observed following SCX knockdown in fetal tenocytes suggests that SCX has a critical role in tendons at this developmental stage. This is supported by our previous work which demonstrated that SCX knockdown in fetal tenocytes prevented them from contracting a 3D collagen gel, whereas adult SCX knockdown tenocytes contracted the collagen gels as normal (Bavin et al., 2017). Our data demonstrates that SCX is regulating fewer genes in adult tenocytes. Although SCX is implicated in adult tendon injury and repair, it is regulated through pro-fibrotic TGF $\beta$  signalling and via mechanical loading in response to injury (Eliasson et al., 2009; Maeda et al., 2011; Pryce et al., 2009). It would therefore be of interest to look at the transcriptome of SCX knockdown adult tenocytes following such stimulus. Using a candidate gene approach, we further demonstrated that SCX-dependent processes differ in young postnatal tenocytes compared to both adult and fetal tenocytes.

These tenocytes were isolated from foals of around 3 months of age which represents a period of rapid tendon growth and development (Firth, 2006). Follow-up studies looking at the transcriptome and direct target genes of SCX in foal tenocytes are required to make more detailed comparisons.

In order to identify potential interactions between the differentially expressed genes resulting from SCX knockdown, network analysis was performed in STRING, software that uses text-mining and computational predictions in order to determine protein-protein interactions (Szklarczyk et al., 2018). In the adult network response to stimuli and cellular migration were overrepresented, with only CCL26 identified as a direct interacting partner of SCX. CCL26 is a cytokine that displays chemotactic activity to eosinophils and basophils and why a direct connection to SCX is made is unclear. Other genes within this network include MMP1 and 2, and Snai1. MMP2 and Snai1 are regulated by SCX in cardiac fibroblasts (Al-Hattab et al., 2018; Nagalingam et al., 2018). Why no direct connection is made is surprising and highlights the limitations in this form of analysis when investigating less well studied proteins where reference sources are limited. In the fetal tenocyte network a clear overrepresentation of ECM and skeletal system development processes were found with direct connections to SOX9, BGN, DCN, GDF6 and CCL26 being detected. Knockout of BGN and DCN in mouse embryos results in tendon defects and SOX9 and GDF6 are heavily implicated in cartilage development (Asou et al., 2002; Delgado Caceres et al., 2018), suggesting that SCX regulation or perhaps coordinated expression of these genes is critical to ensure normal tendon development.

We next sought to identify some of the direct targets of SCX. TNMD, a marker of tendons in humans and rats, is a direct downstream target of SCX (Jelinsky et al., 2009; Shukunami et al., 2018) and in this study we demonstrated that SCX was bound to its promoter region in both adult and fetal tenocytes. *COL1A2* is a direct downstream target gene of SCX in cardiac fibroblasts (Bagchi et al., 2016b; Bagchi and Czubryt, 2012). Following SCX knockdown in both adult and fetal tenocytes, no difference in mRNA expression was identified, however significant enrichment of SCX binding to the *COL1A2* promoter was detected. This could indicate that SCX directs similar but discrete signalling programs in a tissue-dependant manner that are modified by tissue specific interacting partners (Czubryt, 2014). *IGF2BP1* expression was significantly decreased in adult and fetal tenocytes following SCX knockdown and here we demonstrate that SCX directly binds to the *IGF2BP1* gene promoter in both adult and fetal tenocytes.



Although IGF2BP1 has not currently been implicated in tendon development, evidence suggests it plays a role in periodontal ligament (PDL) tissue (Song et al., 2013). Further functional studies to determine the effects of altering IGF2BP1 expression in tenocytes would therefore be of interest.

*PDGFB* was significantly decreased following SCX knockdown in fetal tenocytes only. PDGF proteins are produced by a variety of cell types and are heavily implicated in tendon wound healing (Kendall and Feghali-Bostwick, 2014; Yang et al., 2013). Significant enrichment of SCX binding to *PDGFB* was detected in fetal tenocytes suggesting that SCX directly activates *PDGFB* expression in fetal tenocytes. Enrichment was also detected in adult tenocytes; however, this was not significant due to the large variation in fold changes. *CLDN16* was the most significantly downregulated gene in adult tenocytes following SCX knockdown yet was unchanged in fetal tenocytes. Little is known in regard to *CLDN16*, a claudin protein that is an important component of tight junctions between cells. We confirmed that SCX directly interacts with the *CLDN16* gene promoter in adult tenocytes, having significant enrichment in SCX binding whereas no binding was observed in the fetal tenocytes. Functional studies to determine *CLDN16* role in tendon cells are therefore warranted. *MMP3* was the most significantly upregulated gene in both adult and fetal tenocytes following SCX knockdown. Although SCX has only been reported to be a transcriptional activator, SCX was bound to *MMP3* (*MMP3*#2 E-box region) in both adult and fetal tenocytes. This could suggest that either SCX has some previously unreported repressive effects, or that it is passively binding the promoter region which is blocking the binding of a different transcriptional activator for *MMP3* (Fig.6). Identifying histone modifications around this promoter site alongside SCX binding may help determine if SCX is indeed involved in *MMP3* regulation.

Taken together we therefore hypothesise that these novel downstream target genes allow SCX to regulate ECM production during tendon development and healing. In this study we only looked at a small number of differentially expressed genes using ChIP-qPCR and were thus unable to investigate all of the differentially expressed genes resulting from the RNA-seq experiment. Using this approach we designed primers within gene promoter regions, however evidence is emerging that some genes are regulated by introns (Rose, 2019). This emphasises the need for larger scale studies such as genome-wide chromatin immunoprecipitation-sequencing to produce a broader picture of SCX downstream regulation. Similarly, it is well reported that transcription factors generally do not work

alone and details on their interactions are sorely lacking (Lambert et al., 2018). The E2A splice variants E47 and E12 have been identified as common binding partners of SCX, therefore it would be beneficial to determine if the differentially expressed genes are co-activated by SCX and E47/E12 by comparing ChIP-seq profiles or using co-immunoprecipitation or electrophoretic mobility shift assays (EMSA) techniques.

In conclusion we have identified SCX as a key regulator of ECM gene expression in tendon cells and demonstrate that SCX regulation differs during adult and fetal development. This dataset provides a crucial resource for directing further studies into SCX regulation which, given the significant role SCX plays in tissue repair, may prove useful for the development of therapeutics for tendon injury and disease.

## **4. Materials and Methods**

### **4.1 Tenocyte Cell Culture**

Equine (*Equus caballus*) tendon tissue was harvested with the approval of the Animal Health Trust ethical review committee (AHT02\_2012) as described previously (Barsby et al., 2014). Tenocytes were isolated from the mid-metacarpal region, ensuring the paratenon/epitenon was removed and only the central tendon core was used, of healthy SDFTs of six adult Thoroughbred horses (aged 2-10 years) and three Thoroughbred foals (aged 54 days – 84 day postpartum), euthanised for reasons unrelated to this project. Fetal tenocytes were derived from the tendons of six spontaneously aborted Thoroughbred fetuses which were 271, 289, 316, 319, 320, and 321 days through gestation (full term = 322-387 days). All tenocytes lines used in this study were authenticated for key tendon markers using qPCR and their ability to form artificial tendons confirmed. Briefly, dissected tissue was digested in 1 mg/mL collagenase (Sigma, Poole, Dorset, UK) solution at 37°C overnight. Dulbecco's Modified Eagle's Medium (DMEM) (Gibco, Invitrogen, Carlsbad, CA, USA) containing 10% Fetal Bovine Serum (FBS) (Gibco), 2 mM L-glutamine (LQ) (Gibco) and 1% penicillin-streptomycin (P/S) (Gibco) was used for cellular expansion and culture, with conditions being maintained at 37°C in 5% CO<sub>2</sub>. Medium was

replaced every 2-3 days with cells being passaged using 0.25% trypsin/EDTA (Sigma) once they reached 80% confluency.

#### 4.2 Three-Dimensional Cell Culture

Three-dimensional cell culture was conducted as previously described (Barsby et al., 2014; McClellan et al., 2019). Briefly, pairs of 0.2 mm-diameter minuten pins (Interfocus fine science tools) were embedded at 15 mm distances into silicon-coated six-well plates (Sylgard 184 Silicone elastomer; Dow Corning). Tenocytes were suspended at a concentration of  $4 \times 10^5$  cells/ml in a mixture of eight parts chilled PureCol (Bovine collagen type I; Advanced Biomatrix, Carlsbad, CA, USA) to 2 parts chilled cell culture medium (pH adjusted with 1 M sodium hydroxide from 7.2 to 7.6). Cell counts were performed using a haemocytometer. The chilled cell suspension mix (200  $\mu$ l) was pipetted around each pair of pins before allowing to set for 60-90 minutes at 37°C. Once fully set, 3 ml of tenocyte medium was added per well with the medium changed every 3-4 days during the 14-day culture period. Images were obtained throughout the culture period and contraction analysis was conducted using ImageJ software (National Institutes of Health) with results being displayed as a percentage of the Day 0 value. For each time point from day 0 to 14, contraction analysis was carried out on 3-18 constructs per biological line. Cell survivals were also performed as previously described (Barsby et al., 2014). Briefly, 3D constructs were digested in 1 mg/ml type 1 collagenase produced by *C. histolyticum* (Sigma) for 1-2 hours at 37°C. Survivals were carried out on days 3, 7 and 14 of culture. Live cell counts using trypan blue exclusion staining (Sigma) were performed, with cell counts determined using a haemocytometer. Cell survival was calculated as a percentage of the number of cells originally seeded and were carried out on three to six constructs per time point for each biological line.

#### 4.3 Lentiviral Transfection

TRC2-pLKO.1-puro vector plasmids containing a shRNA specific for human SCX (95.24% identity to the equine SCX sequence) (clone NM\_001008271.1-95s21c1; Sigma) or a non-target, scrambled (NT) shRNA sequence (SHC202; Sigma) were utilised for stable cell line generation. One microgram of the pLKO.1 plasmid alongside 750 ng psPAX2 (#11260; Addgene, Cambridge, MA) and 250 ng pMD2.G (#12259; Addgene) were transfected into HEK293T packaging cells, plated at  $1 \times 10^5$  cells per well of a 6 well plate, using the FuGENE 6 transfection reagent as per the manufacturer's instructions (Promega,

Hampshire, UK). Packaging cell supernatant containing infectious lentiviral material was then collected 48-, 72- and 96-hours post transfection, filtered through a 0.45 µm filter (Millipore, Billerica, MA, USA) to remove cellular debris and frozen at -70°C. Target cells were seeded at 1x10<sup>5</sup> cells per well of a 6-well plate 24-hours before infection with 1x10<sup>7</sup> IU/ml of virus. Puromycin (Sigma) antibiotic selection was then carried out at a concentration of 4 µg/ml (determined optimal based on an antibiotic kill curve of non-infected cells). The qPCR Lentivirus Titration Kit (LV900) from abm (Applied Biological Materials Inc, Richmond, CA) was used to calculate the viral titer of the frozen preparations. Lentiviral infection was optimised using titers varying from 5x10<sup>5</sup> IU/ml to 1x10<sup>8</sup> IU/ml, with 1x10<sup>7</sup> IU/ml obtaining the best viral efficiency with limited cytotoxic effects and good copy number integration (data not shown).

#### 4.4 Lentiviral Copy Number Integration

Genomic DNA was extracted from 1x10<sup>6</sup> lentiviral infected cells using the QIAamp DNA mini kit (Qiagen, Manchester, UK) according to manufacturer's instruction. Transgene copy number was quantified using a combination of the protocols described in Joshi *et al.*, 2008 and Barczak *et al.*, 2014. Briefly qPCR was conducted using 100 ng of DNA per 25 µl reactions using SensiMix SYBR No-ROX (Bioline, London, UK). Primers were designed for the detection of the Woodchuck Hepatitis Virus Post-Transcriptional Regulatory Element (WPRE) sequence, a lentiviral specific fragment, and the 18s housekeeping gene (Table 1). All reactions were conducted using the cycling conditions described in Table 2. The mass of one copy of the SCX shRNA and NT shRNA plasmids were calculated using the formula below: -

$$m = \frac{M}{N_A}$$

Where m = mass of one copy, M = molecular weight of the plasmid (Dalton) calculated based on the base composition of the plasmid and the molecular weight of nucleotide (650 Daltons/base pair) and N<sub>A</sub> = Avogadro's number (6.02 x 10<sup>23</sup> copies/mole).

It was calculated that each copy of SCX shRNA and NT shRNA has a mass of 5.7 ag and 1 ng of plasmid therefore had approximately 1.23 x 10<sup>9</sup> copies of WPRE transgene. Plasmid concentrations were quantified using the Quant-iT™ PicoGreen™ dsDNA assay kit (Invitrogen, Renfrewshire, UK)

according to the manufacturer's instructions. A standard curve was constructed by carrying out serial dilutions of each plasmid, containing the relevant sequence (WPRE), from 1 pg to 10 ag allowing quantification of unknown samples. All dilutions were made in the presence of 10 ng/μl carrier DNA from wildtype equine tenocytes (not containing the WPRE transgene) in order to increase the stability of the dilutions. The number of transgene lentiviral integrating events in 100 ng of DNA was calculated from their respective C<sub>T</sub> (cycle threshold) value using a linear equation calculated from each plasmid standard curve. WPRE copy number per diploid cell was then calculated based on an average DNA yield of 28.33 μg obtained from 1x10<sup>6</sup> cells as shown in the equation below: -

$$\begin{aligned} \text{Average number of cells per 100 ng DNA} &= \frac{28.33 \mu\text{g}}{100 \text{ ng}} = 283.3 \text{ ng} \\ &= \frac{1 \times 10^6}{283.3} = 353 \text{ cells} \end{aligned}$$

Therefore, it was calculated that 100 ng of DNA contains the DNA from approximately 353 diploid cells.

#### 4.5 RNA Isolation

Four biological lines of adult and fetal non-target shRNA infected (NT) and SCX shRNA infected (shSCX) tenocytes (passages (P) 5-8) were used in the RNA-sequencing experiments. Quantitative PCR (qPCR) validation of the RNA-sequencing data was conducted using six biological lines of adult and fetal NT and shSCX tenocytes (P 5-8). Three biological non-transduced lines were used for comparing adult, fetal and foal tenocytes for key tendon markers using qPCR (P 3-8). Three biological lines of NT and shSCX foal tenocytes were used in foal knockdown qPCR analysis (P 5-9). RNA was harvested in 1 ml Tri-reagent (Sigma) per confluent 10 cm<sup>2</sup> plate of cells. RNA was extracted and purified using a RNeasy mini kit (Qiagen). Contaminating DNA was removed using the Ambion DNA-free kit (Life Technologies, Paisley, UK). A Nanodrop was used to measure RNA concentration (ThermoFisher, Loughborough, UK). RNA integrity Numbers (RIN<sup>®</sup>) were confirmed on an Agilent 2200 Tapestation by an external provider (Cambridge Genomics, Cambridge, UK) with values of 9.9-10 obtained.

#### 4.6 RNA Sequencing

mRNA library prep and sequencing were performed by an external provider (Edinburgh Genomics, Edinburgh, UK) using a TruSeq stranded mRNA kit (Illumina, Cambridge, UK), with sequencing

performed on the NovaSeq6000 (Illumina). Per sample, on average 31.75 million reads of approximately 100 base pair paired end were generated. Resulting FASTQ files were quality control checked using FASTQC and FASTQ Screen (Babraham Bioinformatics, Cambridge, UK). The pseudoaligner Salmon (Patro et al., 2017) in Quasi-mapping based mode was used to align the reads to the Ensemble version v96 EquCab 3.0 transcriptome with GC-bias correction (-gcBias) applied. Gene-level abundance data was imported into R (v.3.5.2) using Tximport (Soneson et al., 2015) and the R/Bioconductor DSeq2 (v.1.22.2) software used to conduct differential expression analysis as described in Love et al., 2014. Genes with Log2FC of  $\pm 1$  and adjusted p-value (p-adj) of  $<0.05$  were considered differentially expressed (DE).

#### *4.7 Gene, Pathway and Network Analysis*

Panther (<http://www.pantherdb.org/>) was used to conduct Gene Ontology (GO) analysis, with a false discovery rate (FDR) of  $<0.05$  being defined as significantly enriched. Gene Analytics from the LifeMap's GeneCards Suite (<http://geneanalytics.genecards.org>) was used to perform pathway analysis, with an entity score of  $>5$  being equivalent to a corrected p-value of  $<0.05$  and therefore defined as significantly enriched. Both GO and pathway analysis were conducted on genes with Log2FC of  $\pm 1$  and adjusted p-value (p-adj) of  $<0.05$ . Network analysis was conducted using the STRING Protein network analyser plug-in of Cytoscape (v.3.7.2) using the human reference set. Networks were generated using a less stringent cut off of  $\text{Log2FC} \pm 0.6$  p.adj $<0.05$  to encompass a larger number of differentially expressed genes.

#### *4.8 cDNA Synthesis and quantitative PCR*

The SensiFAST cDNA Synthesis Kit (Bioline) was used to reverse transcribed 1  $\mu\text{g}$  of total RNA. To ensure no genomic DNA contamination, reactions lacking the reverse transcriptase were carried out in parallel. Equine gene specific primers were designed using NCBI Primer-Blast (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) and are found in Table 1. The programme mfold (<http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form>) was used to ensure each amplicon (50-150bp in size) had a melting temperature ( $T_m$ ) of between 58-62°C and were devoid of secondary structure at  $T_m$  60°C. qPCR was conducted on the Bio-Rad C1000 Touch Thermal Cycler (Bio-Rad),

using SensiMix SYBR No-ROX (Bioline). qPCR cycling conditions can be found in Table 2. All primer sets were optimised to work under these conditions to give efficiencies between 90-110% (data not shown). Reactions were quantified relative to the 18s rRNA housekeeping gene (Bavin et al., 2017), with RefFinder (Xie et al., 2012) being used to determine the most suitable, stable housekeeping gene (data not shown). The relative gene expression was calculated using:  $2^{-\Delta\Delta CT}$  (Livak and Schmittgen, 2001).

#### 4.9 Immunocytochemistry on 2D Coverslips

Four biological lines of adult and fetal non-target shRNA infected (NT) and SCX shRNA infected (shSCX) tenocytes were cultured on gelatin-coated (Sigma) coverslips. Coverslips were fixed for 20 minutes in 3% paraformaldehyde and subsequently permeabilized in 0.1% triton-X-100 (Sigma) for 1 hour. Blocking was carried out in 2.5% normal horse serum (Vector Laboratories, Peterborough, UK) for 20 minutes before incubating with anti-SCX primary antibody (#ab58655; abcam) at a 1:100 dilution in 2.5% normal horse serum. Primary antibody incubations were carried out overnight at 4°C. Subsequent detection using fluorescently labelled secondary antibody was carried out using anti-rabbit IgG Alexaflor 594 antibody (#A11012; ThermoFisher) at a 1:200 dilution in 2.5% normal horse serum. Secondary antibody incubations were carried out for 3 hours at room temperature. Images were captured using the Zeiss Axioplan 2 imaging suite.

#### 4.10 Chromatin Immunoprecipitation Assays

Three lines of non-transduced adult and fetal tenocytes (approximately  $25 \times 10^6$  cells) were washed in PBS and fixed in 11% formaldehyde for 10 minutes at room temperature with gentle shaking. Fixation was quenched by adding glycine at a final concentration of 125 mM. The iDeal ChIP-qPCR Kit (#C01010180; Diagenode, Seraing, Belgium) was used to prepare chromatin, conducted according to the manufacturer's instructions. Chromatin shearing was performed using the Misonix Sonicator XL2020 Ultrasonic Liquid Processor (Misonix, NT, USA) using the cup horn water bath probe. Shearing optimisation determined that a programme of 7 minutes and 40 seconds, with pulse parameters of 20 seconds on and 20 seconds off provided the most efficient shearing producing chromatin fragments between 100-600bp (data not shown). Shearing quality was determined via gel electrophoresis, by running RNase treated (RNase cocktail - Ambion) reverse cross-linked samples on a 1.5% agarose gel

in 1x TAE buffer as described in the iDeal ChIP-qPCR Kit Protocol (# C01010180; Diagenode). Sheared chromatin was immunoprecipitated at 4°C overnight with gentle rotation with either 4 µg of anti-SCX antibody (#PA5-23943; ThermoFisher) or 4 µg rabbit non-immune IgG (#C15410206; Diagenode) according to the iDeal ChIP-qPCR Kit instructions. Immunoprecipitated DNA was subjected to qPCR as previously described. Primer generation was conducted using the sequences of equine proximal promoter regions of 11 genes in order to identify putative E-box regions (consensus sequence CANNTG (N = any nucleotide)) in which SCX may bind (Table 1). As promoter regions in the horse are currently not well annotated, the sequence 2 kb upstream of the transcription start sites were used to define the “promoter” region. Some genes contained multiple E-boxes within their predicted promoter regions, therefore where necessary multiple primers were designed in order to capture these. Negative control regions included genes which had no expression in equine tenocytes, were not within a predicted promoter region and had no E-box binding sites within 250 bp either side of the primers (TEX33 and RNASE9) and intronic regions which contained no E-box binding sites within 1.5 kb of the primers (Intronic #1 and Intronic #2). Fold enrichment was calculated ( $\Delta C_t$ ) and represents the ratio of SCX-bound DNA to the average of the negative controls (TEX33, RNASE9, Intronic #1 and Intronic #2) normalised for input DNA.

#### 4.11 Statistical Analysis

Statistical analysis of qPCR and gel contraction data was performed using XLSTAT (version 22.1.3). Histograms were plotted to visualise the distribution of the datasets and all were confirmed to have a Gaussian distribution using the Shapiro Wilks normality test and subsequent visualisation of Q-Q plots. For comparisons of two groups the Student's *t*-test (unpaired, two-tailed) was used. For comparisons of more than two groups with equal variance ANOVA was used, followed by the Tukey's post hoc test. When unequal variance was observed (as determined by the k-sample comparison of variance test, Levene's Test) Welch's ANOVA was used, followed by the Games-Howell post hoc test. In all cases the significance threshold was set at  $p < 0.05$ . The number of independent biological replicates used is indicated in the figure legends.



## Acknowledgments

The authors are grateful to Miss Emma Goodfellow for collecting the tissue samples and Ms Ellen Schofield for advice with the RNA-sequencing analysis. Infographics were created with the Mind the Graph platform (<https://mindthegraph.com>).

## Funding

The authors are grateful to the PetPlan Charitable Trust for funding this work (S17-419-457). YP is funded by a Biotechnology and Biological Sciences Research Council (BBSRC) doctoral training partnership PhD studentship.

**Declarations of interest:** none

## Data Availability

Sequence data generated have been submitted to the National Centre for Biotechnology Information Gene Expression Omnibus (NCBI GEO, [www.ncbi.nlm.nih.gov/geo](http://www.ncbi.nlm.nih.gov/geo)) accession number GSE149570 accessible using token glytcukvhmzvjk. Differential expression analysis and normalized counts are included in Supplementary File 1.

## References

- Al-Hattab, D.S., Safi, H.A., Nagalingam, R.S., Bagchi, R.A., Stecy, M.T., Czubryt, M.P., 2018. Scleraxis regulates Twist1 and Snai1 expression in the epithelial-to-mesenchymal transition. *Am. J. Physiol. Circ. Physiol.* 315, H658–H668. <https://doi.org/10.1152/ajpheart.00092.2018>
- Ansorge, H.L., Hsu, J.E., Edelstein, L., Adams, S., Birk, D.E., Soslowsky, L.J., 2012. Recapitulation of the Achilles tendon mechanical properties during neonatal development: A Study of differential healing during two stages of development in a mouse model. *J. Orthop. Res.* 30, 448–456. <https://doi.org/10.1002/jor.21542>
- Asou, Y., Nifuji, A., Tsuji, K., Shinomiya, K., Olson, E.N., Koopman, P., Noda, M., 2002. Coordinated expression of scleraxis and Sox9 genes during embryonic development of tendons and cartilage. *J. Orthop. Res.* 20, 827–833. [https://doi.org/10.1016/S0736-0266\(01\)00169-3](https://doi.org/10.1016/S0736-0266(01)00169-3)
- Baccarella, A., Williams, C.R., Parrish, J.Z., Kim, C.C., 2018. Empirical assessment of the impact of

639 sample number and read depth on RNA-Seq analysis workflow performance. BMC  
 640 Bioinformatics 19, 423. <https://doi.org/10.1186/s12859-018-2445-2>  
 641 Bagchi, R.A., Czubryt, M.P., 2012. Synergistic roles of scleraxis and Smads in the regulation of  
 642 collagen 1 $\alpha$ 2 gene expression. Biochim. Biophys. Acta - Mol. Cell Res. 1823, 1936–1944.  
 643 <https://doi.org/10.1016/j.bbamcr.2012.07.002>  
 644 Bagchi, R.A., Lin, J., Wang, R., Czubryt, M.P., 2016a. Regulation of fibronectin gene expression in  
 645 cardiac fibroblasts by scleraxis. Cell Tissue Res. 366, 381–391. [https://doi.org/10.1007/s00441-](https://doi.org/10.1007/s00441-016-2439-1)  
 646 016-2439-1  
 647 Bagchi, R.A., Roche, P., Aroutiounova, N., Espira, L., Abrenica, B., Schweitzer, R., Czubryt, M.P.,  
 648 2016b. The transcription factor scleraxis is a critical regulator of cardiac fibroblast phenotype.  
 649 BMC Biol. 14, 21. <https://doi.org/10.1186/s12915-016-0243-8>  
 650 Bagchi, R.A., Wang, R., Jahan, F., Wigle, J.T., Czubryt, M.P., 2016c. Regulation of scleraxis  
 651 transcriptional activity by serine phosphorylation. J. Mol. Cell. Cardiol. 92, 140–148.  
 652 <https://doi.org/10.1016/j.yjmcc.2016.02.013>  
 653 Ballas, C.B., Davidson, J.M., 2001. Delayed wound healing in aged rats is associated with increased  
 654 collagen gel remodeling and contraction by skin fibroblasts, not with differences in apoptotic or  
 655 myofibroblast cell populations. Wound Repair Regen. 9, 223–237. [https://doi.org/10.1046/j.1524-](https://doi.org/10.1046/j.1524-475x.2001.00223.x)  
 656 475x.2001.00223.x  
 657 Barsby, T., Bavin, E.P., Guest, D.J., 2014. Three-dimensional culture and transforming growth factor  
 658 beta3 synergistically promote tenogenic differentiation of equine embryo-derived stem cells.  
 659 Tissue Eng. - Part A 20, 2604–2613. <https://doi.org/10.1089/ten.tea.2013.0457>  
 660 Barsby, T., Guest, D., 2013. Transforming Growth Factor Beta3 Promotes Tendon Differentiation of  
 661 Equine Embryo-Derived Stem Cells. Tissue Eng. Part A 19, 2156–2165.  
 662 <https://doi.org/10.1089/ten.tea.2012.0372>  
 663 Bavin, E.P., Atkinson, F., Barsby, T., Guest, D.J., 2017. Scleraxis Is Essential for Tendon  
 664 Differentiation by Equine Embryonic Stem Cells and in Equine Fetal Tenocytes. Stem Cells Dev.  
 665 26, 441–450. <https://doi.org/10.1089/scd.2016.0279>  
 666 Beredjiklian, P.K., Favata, M., Cartmell, J.S., Flanagan, C.L., Crombleholme, T.M., Soslowsky, L.J.,  
 667 2003. Regenerative Versus Reparative Healing in Tendon: A Study of Biomechanical and  
 668 Histological Properties in Fetal Sheep. Ann. Biomed. Eng. 31, 1143–1152.

669 <https://doi.org/10.1114/1.1616931>

670 Carlberg, A.L., Tuan, R.S., Hall, D.J., 2000. Regulation of Scleraxis Function by Interaction with the  
671 bHLH Protein E47. *Mol. Cell Biol. Res. Commun.* 3, 82–86.  
672 <https://doi.org/10.1006/mcbr.2000.0195>

673 Chen, J., Zhang, W., Liu, Z., Zhu, T., Shen, W., Ran, J., Tang, Q., Gong, X., Backman, L.J., Chen, X.,  
674 Chen, X., Wen, F., Ouyang, H., 2016. Characterization and comparison of post-natal rat Achilles  
675 tendon-derived stem cells at different development stages. *Sci. Rep.* 6, 22946.  
676 <https://doi.org/10.1038/srep22946>

677 Connizzo, B.K., Yannascoli, S.M., Soslowsky, L.J., 2013. Structure–function relationships of postnatal  
678 tendon development: A parallel to healing. *Matrix Biol.* 32, 106–116.  
679 <https://doi.org/10.1016/j.matbio.2013.01.007>

680 Cox, B.C., Chai, R., Lenoir, A., Liu, Z., Zhang, L., Nguyen, D.-H., Chalasani, K., Steigelman, K.A.,  
681 Fang, J., Cheng, A.G., Zuo, J., 2014. Spontaneous hair cell regeneration in the neonatal mouse  
682 cochlea in vivo. *Development* 141, 816–829. <https://doi.org/10.1242/dev.103036>

683 Cserjesi, P., Brown, D., Ligon, K.L., Lyons, G.E., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Olson,  
684 E.N., 1995. Scleraxis: A basic helix-loop-helix protein that prefigures skeletal formation during  
685 mouse embryogenesis. *Development* 121, 1099–1110.

686 Czubryt, M.P., 2014. A tale of 2 tissues: the overlapping role of scleraxis in tendons and the heart.  
687 *Can. J. Physiol. Pharmacol.* 92, 707–712. <https://doi.org/10.1139/cjpp-2013-0489>

688 Delgado Caceres, M., Pfeifer, C.G., Docheva, D., 2018. Understanding Tendons: Lessons from  
689 Transgenic Mouse Models. *Stem Cells Dev.* 27, 1161–1174.  
690 <https://doi.org/10.1089/scd.2018.0121>

691 Eliasson, P., Andersson, T., Aspenberg, P., 2009. Rat Achilles tendon healing: mechanical loading  
692 and gene expression. *J. Appl. Physiol.* 107, 399–407.  
693 <https://doi.org/10.1152/japplphysiol.91563.2008>

694 Evans, T.G., 2015. Considerations for the use of transcriptomics in identifying the “genes that matter”  
695 for environmental adaptation. *J. Exp. Biol.* 218, 1925–1935. <https://doi.org/10.1242/jeb.114306>

696 Favata, M., Beredjiklian, P.K., Zgonis, M.H., Beason, D.P., Crombleholme, T.M., Jawad, A.F.,  
697 Soslowsky, L.J., 2006. Regenerative properties of fetal sheep tendon are not adversely affected  
698 by transplantation into an adult environment. *J. Orthop. Res.* 24, 2124–2132.

699 <https://doi.org/10.1002/jor.20271>

700 Firth, E.C., 2006. The response of bone, articular cartilage and tendon to exercise in the horse. *J.*  
701 *Anat.* 208, 513–526. <https://doi.org/10.1111/j.1469-7580.2006.00547.x>

702 Furumatsu, T., Shukunami, C., Amemiya-Kudo, M., Shimano, H., Ozaki, T., 2010. Scleraxis and E47  
703 cooperatively regulate the Sox9-dependent transcription. *Int. J. Biochem. Cell Biol.* 42, 148–156.  
704 <https://doi.org/10.1016/j.biocel.2009.10.003>

705 Goldin, M., Malanga, G.A., 2013. Tendinopathy: A Review of the Pathophysiology and Evidence for  
706 Treatment. *Phys. Sportsmed.* 41, 36–49. <https://doi.org/10.3810/psm.2013.09.2019>

707 Havis, E., Bonnin, M.-A., Esteves de Lima, J., Charvet, B., Milet, C., Duprez, D., 2016. TGF $\beta$  and  
708 FGF promote tendon progenitor fate and act downstream of muscle contraction to regulate  
709 tendon differentiation during chick limb development. *Development* 143, 3839–3851.  
710 <https://doi.org/10.1242/dev.136242>

711 Hillmann, A., Ahrberg, A.B., Brehm, W., Heller, S., Josten, C., Paebst, F., Burk, J., 2016. Comparative  
712 Characterization of Human and Equine Mesenchymal Stromal Cells: A Basis for Translational  
713 Studies in the Equine Model. *Cell Transplant.* 25, 109–124.  
714 <https://doi.org/10.3727/096368915X687822>

715 Hobert, O., 2008. Gene Regulation by Transcription Factors and MicroRNAs. *Science* (80-. ). 319,  
716 1785–1786. <https://doi.org/10.1126/science.1151651>

717 Howell, K., Chien, C., Bell, R., Laudier, D., Tufa, S.F., Keene, D.R., Andarawis-Puri, N., Huang, A.H.,  
718 2017. Novel Model of Tendon Regeneration Reveals Distinct Cell Mechanisms Underlying  
719 Regenerative and Fibrotic Tendon Healing. *Sci. Rep.* 7, 45238.  
720 <https://doi.org/10.1038/srep45238>

721 Iismaa, S.E., Kaidonis, X., Nicks, A.M., Bogush, N., Kikuchi, K., Naqvi, N., Harvey, R.P., Husain, A.,  
722 Graham, R.M., 2018. Comparative regenerative mechanisms across different mammalian  
723 tissues. *npj Regen. Med.* 3, 6. <https://doi.org/10.1038/s41536-018-0044-5>

724 Jelinsky, S.A., Archambault, J., Li, L., Seeherman, H., 2009. Tendon-selective genes identified from  
725 rat and human musculoskeletal tissues. *J. Orthop. Res.* 28, 289–97.  
726 <https://doi.org/10.1002/jor.20999>

727 Kendall, R.T., Feghali-Bostwick, C.A., 2014. Fibroblasts in fibrosis: novel roles and mediators. *Front.*  
728 *Pharmacol.* 5, 123. <https://doi.org/10.3389/fphar.2014.00123>

729 Lamarre, S., Frasse, P., Zouine, M., Labourdette, D., Sainderichin, E., Hu, G., Le Berre-Anton, V.,  
 730 Bouzayen, M., Maza, E., 2018. Optimization of an RNA-seq differential gene expression  
 731 analysis depending on biological replicate number and library size. *Front. Plant Sci.* 9, 108.  
 732 <https://doi.org/10.3389/fpls.2018.00108>  
 733 Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes,  
 734 T.R., Weirauch, M.T., 2018. The Human Transcription Factors. *Cell* 172, 650–665.  
 735 <https://doi.org/10.1016/j.cell.2018.01.029>  
 736 Léjard, V., Brideau, G., Blais, F., Salingcarnboriboon, R., Wagner, G., Roehrl, M.H.A., Noda, M.,  
 737 Duprez, D., Houillier, P., Rossert, J., 2007. Scleraxis and NFATc Regulate the Expression of the  
 738 Pro- $\alpha$ 1(I) Collagen Gene in Tendon Fibroblasts. *J. Biol. Chem.* 282, 17665–17675.  
 739 <https://doi.org/10.1074/jbc.M610113200>  
 740 Li, Y., Ramcharan, M., Zhou, Z., Leong, D.J., Akinbiyi, T., Majeska, R.J., Sun, H.B., 2015. The Role of  
 741 Scleraxis in Fate Determination of Mesenchymal Stem Cells for Tenocyte Differentiation. *Sci.*  
 742 *Rep.* 5, 13149. <https://doi.org/10.1038/srep13149>  
 743 Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time  
 744 quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* 25, 402–408.  
 745 <https://doi.org/10.1006/meth.2001.1262>  
 746 Maeda, T., Sakabe, T., Sunaga, A., Sakai, K., Rivera, A.L., Keene, D.R., Sasaki, T., Stavnezer, E.,  
 747 Iannotti, J., Schweitzer, R., Ilic, D., Baskaran, H., Sakai, T., 2011. Conversion of mechanical  
 748 force into TGF- $\beta$ -mediated biochemical signals. *Curr. Biol.* 21, 933–941.  
 749 <https://doi.org/10.1016/j.cub.2011.04.007>  
 750 McClellan, A., Evans, R., Sze, C., Kan, S., Paterson, Y., Guest, D., 2019. A novel mechanism for the  
 751 protection of embryonic stem cell derived tenocytes from inflammatory cytokine interleukin 1  
 752 beta. *Sci. Rep.* 9, 2755. <https://doi.org/10.1038/s41598-019-39370-4>  
 753 Murchison, N.D., Price, B.A., Conner, D.A., Keene, D.R., Olson, E.N., Tabin, C.J., Schweitzer, R.,  
 754 2007. Regulation of tendon differentiation by scleraxis distinguishes force-transmitting tendons  
 755 from muscle-anchoring tendons. *Development* 134, 2697–2708.  
 756 <https://doi.org/10.1242/dev.001933>  
 757 Nagalingam, R.S., Safi, H.A., Al-Hattab, D.S., Bagchi, R.A., Landry, N.M., Dixon, I.M.C., Wigle, J.T.,  
 758 Czubryt, M.P., 2018. Regulation of cardiac fibroblast MMP2 gene expression by scleraxis. *J.*

759 Mol. Cell. Cardiol. 120, 64–73. <https://doi.org/10.1016/j.yjmcc.2018.05.004>

760 Nichols, A.E.C., Settlage, R.E., Werre, S.R., Dahlgren, L.A., 2018. Novel roles for scleraxis in  
 761 regulating adult tenocyte function. BMC Cell Biol. 19, 14. [https://doi.org/10.1186/s12860-018-](https://doi.org/10.1186/s12860-018-0166-z)  
 762 0166-z

763 Nourissat, G., Berenbaum, F., Duprez, D., 2015. Tendon injury: from biology to tendon repair. Nat.  
 764 Rev. Rheumatol. 11, 223–233. <https://doi.org/10.1038/nrrheum.2015.26>

765 Paterson, Y.Z., Cribbs, A., Espenel, M., Smith, E.J., Henson, F.M.D., Guest, D.J., 2020. Genome-  
 766 wide transcriptome analysis reveals equine embryonic stem cell-derived tenocytes resemble  
 767 fetal, not adult tenocytes. Stem Cell Res. Ther. 11, 184. [https://doi.org/10.1186/s13287-020-](https://doi.org/10.1186/s13287-020-01692-w)  
 768 01692-w

769 Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-  
 770 aware quantification of transcript expression. Nat. Methods 14, 417–419.  
 771 <https://doi.org/10.1038/nmeth.4197>

772 Patterson-Kane, J.C., Becker, D.L., Rich, T., 2012. The Pathogenesis of Tendon Microdamage in  
 773 Athletes: the Horse as a Natural Model for Basic Cellular Research. J. Comp. Pathol. 147, 227–  
 774 247. <https://doi.org/10.1016/j.jcpa.2012.05.010>

775 Patterson-Kane, J.C., Rich, T., 2014. Achilles tendon injuries in elite athletes: Lessons in  
 776 pathophysiology from their equine Counterparts. ILAR J. 55, 86–99.  
 777 <https://doi.org/10.1093/ilar/ilu004>

778 Porrello, E.R., Mahmoud, A.I., Simpson, E., Hill, J.A., Richardson, J.A., Olson, E.N., Sadek, H.A.,  
 779 2011. Transient Regenerative Potential of the Neonatal Mouse Heart. Science (80-. ). 331,  
 780 1078–1080. <https://doi.org/10.1126/science.1200708>

781 Pryce, B.A., Watson, S.S., Murchison, N.D., Staverosky, J.A., Dünker, N., Schweitzer, R., 2009.  
 782 Recruitment and maintenance of tendon progenitors by TGFB signaling are essential for tendon  
 783 formation. Development 136, 1351–1361. <https://doi.org/10.1242/dev.027342>

784 Roguljic, H., Matthews, B.G., Yang, W., Cvija, H., Mina, M., Kalajzic, I., 2013. In vivo Identification of  
 785 Periodontal Progenitor Cells. J. Dent. Res. 92, 709–715.  
 786 <https://doi.org/10.1177/0022034513493434>

787 Rose, A.B., 2019. Introns as Gene Regulators: A Brick on the Accelerator. Front. Genet. 9, 672.  
 788 <https://doi.org/10.3389/fgene.2018.00672>

789 Sakabe, T., Sakai, K., Maeda, T., Sunaga, A., Furuta, N., Schweitzer, R., Sasaki, T., Sakai, T., 2018.  
 790 Transcription factor scleraxis vitally contributes to progenitor lineage direction in wound healing  
 791 of adult tendon in mice. *J. Biol. Chem.* 293, 5766–5780.  
 792 <https://doi.org/10.1074/jbc.RA118.001987>  
 793 Schweitzer, R., Chyung, J.H., Murtaugh, L.C., Brent, A.E., Rosen, V., Olson, E.N., Lassar, A., Tabin,  
 794 C.J., 2001. Analysis of the tendon cell fate using Scleraxis, a specific marker for tendons and  
 795 ligaments. *Development* 128, 3855–3866.  
 796 Scott, A., Sampaio, A., Abraham, T., Duronio, C., Underhill, T.M., 2011. Scleraxis expression is  
 797 coordinately regulated in a murine model of patellar tendon injury. *J. Orthop. Res.* 29, 289–296.  
 798 <https://doi.org/10.1002/jor.21220>  
 799 Shukunami, C., Takimoto, A., Nishizaki, Y., Yoshimoto, Y., Tanaka, S., Miura, S., Watanabe, H.,  
 800 Sakuma, T., Yamamoto, T., Kondoh, G., Hiraki, Y., 2018. Scleraxis is a transcriptional activator  
 801 that regulates the expression of Tenomodulin, a marker of mature tenocytes and  
 802 ligamentocytes. *Sci. Rep.* 8, 3155. <https://doi.org/10.1038/s41598-018-21194-3>  
 803 Shukunami, C., Takimoto, A., Oro, M., Hiraki, Y., 2006. Scleraxis positively regulates the expression  
 804 of tenomodulin, a differentiation marker of tenocytes. *Dev. Biol.* 298, 234–247.  
 805 <https://doi.org/10.1016/j.ydbio.2006.06.036>  
 806 Son, K., Yu, S., Shin, W., Han, K., Kang, K., 2018. A Simple Guideline to Assess the Characteristics  
 807 of RNA-Seq Data. *Biomed Res. Int.* 2018, 1–9. <https://doi.org/10.1155/2018/2906292>  
 808 Soneson, C., Love, M.I., Robinson, M.D., 2015. Differential analyses for RNA-seq: transcript-level  
 809 estimates improve gene-level inferences. *F1000Research* 4, 1521.  
 810 <https://doi.org/10.12688/f1000research.7563.1>  
 811 Song, J.S., Hwang, D.H., Kim, S.-O., Jeon, M., Choi, B.-J., Jung, H.-S., Moon, S.J., Park, W., Choi,  
 812 H.-J., 2013. Comparative Gene Expression Analysis of the Human Periodontal Ligament in  
 813 Deciduous and Permanent Teeth. *PLoS One* 8, e61231.  
 814 <https://doi.org/10.1371/journal.pone.0061231>  
 815 Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M.,  
 816 Doncheva, N.T., Morris, J.H., Jensen, L.J., Von Mering, C., 2018. STRING v11: protein-protein  
 817 association networks with increased coverage, supporting functional discovery in genome-wide  
 818 experimental datasets. *Nucleic Acids Res.* 47, 607–613. <https://doi.org/10.1093/nar/gky1131>

- Theodossiou, S.K., Schiele, N.R., 2019. Models of tendon development and injury. *BMC Biomed. Eng.* 1, 32. <https://doi.org/10.1186/s42490-019-0029-5>
- Williams, R.B., Harkins, L.S., Hammond, C.J., Wood, J.L.N., 2010. Racehorse injuries, clinical problems and fatalities recorded on British racecourses from flat racing and National Hunt racing during 1996, 1997 and 1998. *Equine Vet. J.* 33, 478–486. <https://doi.org/10.2746/042516401776254808>
- Xie, F., Xiao, P., Chen, D., Xu, L., Zhang, B., 2012. miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* 80, 75–84. <https://doi.org/10.1007/s11103-012-9885-2>
- Yang, G., Rothrauff, B.B., Tuan, R.S., 2013. Tendon and ligament regeneration and repair: Clinical relevance and developmental paradigm. *Birth Defects Res. Part C Embryo Today Rev.* 99, 203–222. <https://doi.org/10.1002/bdrc.21041>
- Yates, C.C., Hebda, P., Wells, A., 2012. Skin Wound Healing and Scarring: Fetal Wounds and Regenerative Restitution. *Birth Defects Res. Part C - Embryo Today Rev.* <https://doi.org/10.1002/bdrc.21024>

## Figure Legends

**Fig.1.SCX expression is robustly knocked down in equine tenocytes.** (A) Adult and fetal tenocytes exhibit significant reduction in SCX mRNA expression following viral transduction with a specific shRNA against SCX (shSCX) when compared to cells transduced with a control non-target scrambled shRNA (NT). Asterisk (\*) represents  $p < 0.001$  using a two-tailed Student's *t*-test. No significant difference between percentage knockdown is detected between the adult and fetal lines ( $p = 0.336$ ). Error bars represent the SEM of 4 biological replicates. (B) Adult and fetal shSCX or NT copies per diploid cell determined by copy number assays. Water (H<sub>2</sub>O) and wild type, non-infected cells (WT) were used as negative controls. Using a two-tailed Student's *t*-test no significant difference in copy number was detected between fetal NT and shSCX lines ( $p = 0.462$ ), between adult NT and shSCX lines ( $p = 0.101$ ) or between the adult and fetal infected lines ( $p = 0.548$ ). Error bars represent the SEM of 4 biological replicates. (C) Immunocytochemistry confirms a reduction in SCX protein (red) in both adult and fetal



tenocytes following shSCX expression. Images representative of 4 biological replicates. Scale bar = 40  $\mu$ m. DAPI staining of the nuclei is shown in blue.

**Fig.2. RNA-sequencing reveals SCX knockdown in adult and fetal tenocytes has differing effects on global gene expression.** (A) VENN diagram showing overlap of DE genes in the adult (teal) and fetal (orange) tenocytes following SCX knockdown. Top 4 upregulated genes based on their Log2FC are shown in both groups next to the green arrow and the top 4 downregulated genes value are shown in both groups next to the red arrows. (B) Volcano plot displaying the differentially expressed genes between the shSCX and NT control in adult and fetal tenocytes. The x-axis corresponds to the Log2FC and the y-axis the mean expression value of  $\log_{10}(\text{p.adj})$ . The red dots represent those genes which are considered differentially expressed based a  $\text{Log2FC} \pm 1$   $\text{p.adj} < 0.05$  cut off. Positive values on the x-axis represent upregulated genes and negative values represent downregulated genes. (C) Validation of 8 DE genes detected from RNA-seq using qPCR on a larger cohort (6x biological replicates) of SCX knockdown lines. Expression shown relative to the NT control on a  $\log_{10}$  scale. \*  $p < 0.05$  using a two-tailed Student's *t*-test. Error bars represent the SEM of 6 biological replicates. (D) Comparison of significance in expression following SCX knockdown in RNA-seq and qPCR results from adult and fetal tenocytes. A significance threshold of  $< 0.05$  was used for both RNA-seq  $\text{p.adj}$  and q-PCR  $p$ -values. Red shaded boxes indicate no-significant difference. Yellow shaded boxed with a double asterisk (\*\*) represent a significant  $\text{p.adj}$  value, but do not meet the fold cut-off. Green shaded boxes indicate there is a significant difference based on both  $\text{p.adj}/p$ -value and Log2FC. (E) Bar chart of gene expression of collagens, proteoglycans, matrix metalloproteinases and other transcription factors previously demonstrated to be affected by SCX knockdown in tendon and cardiac fibroblasts in adult (teal) and fetal (orange) shSCX expressing tenocyte. Y-axis shows the gene expression in terms of the  $\log_{10}$  normalized counts. Significant differences based on both  $\text{Log2FC} \pm 1$   $\text{p.adj} < 0.05$  are depicted by a double asterisk (\*\*). Those genes which are significant based on  $\text{p.adj}$  but not Log2FC are depicted by a single asterisk (\*). Error bars represent the SEM of 4 biological replicates.

**Fig.3. Gene ontology, pathway and network analysis of DE genes following SCX knockdown in adult and fetal tenocytes.** (A) Summary of the top 10 significantly enriched gene ontology (GO) biological process terms for each pairwise comparison of DE genes in adult and fetal tenocytes following

SCX knockdown. Go terms have been arranged alphabetically to allow easier comparison between groups. (B) Summary of the top 10 Pathways based on entity score for each pairwise comparison of DE genes in adult and fetal tenocytes following SCX knockdown. (C-D) Interaction network showing first to third direct neighboring interactors for SCX within the DE genes in adult tenocytes following SCX knockdown. Pathways have been arranged alphabetically to allow easier comparison between groups. (E-F) Interaction network showing first and second direct neighboring interactors for SCX within the DE genes in fetal tenocytes following SCX knockdown. The top five functional annotations linking connecting nodes are colour coded and can be visualized around each node in the network. Networks as predicted by STRING and visualized in Cytoscape.

**Fig.4. Effect of knocking down SCX in young postnatal foal tenocytes.** (A) Adult, fetal and foal tenocytes are capable of contracting a collagen gel to the same degree. Contraction shown as the percentage of the day 0 value. Similarly, no significant difference in percentage cell survival was detected between the adult, fetal and foal tenocytes.  $p = >0.05$  using ANOVA. Error bars represent the SEM of 3 biological replicates per condition. (B) qPCR showed no significant differences in tendon gene expression between adult, fetal and foal tenocytes using Welch's ANOVA. Expression shown relative to the 18s rRNA housekeeping gene on a log10 scale. N.D. = expression not detected. Error bars represent the SEM of 3 biological replicates. (C) Foal tenocytes exhibit significant reduction in SCX mRNA expression following viral transduction with a specific shRNA against SCX (shSCX) when compared to cells transduced with a control non-target scrambled shRNA (NT). \*  $p < 0.001$  using a two-tailed Student's *t*-test. Expression shown relative to the 18s rRNA housekeeping gene. Error bars represent the SEM of 3 biological replicates. (D) Foal tenocytes expressing shSCX can contract a collagen matrix to the same degree as the NT control cells, ~16% of their starting size. Similarly, no significant difference in percentage cell survival was detected. Significance tested between the two conditions (NT vs. shSCX) using a two-tailed Student's *t*-test. Error bars represent the SEM of 3 biological replicates. (E) Gene expression in foal tenocytes following SCX knockdown produces a significant reduction in *COMP*. Error bars represent the SEM of 3 biological replicates. Relative expression to the NT control is plotted on a log10 scale. \*  $p < 0.001$  using a two-tailed Student's *t*-test. (F) Comparison of differentially expressed genes in adult, fetal and foal tenocytes. Table shows the *p*-values obtained following Student's *t*-testing of the fold change in gene expression of adult, fetal and

foal shSCX expressing tenocytes relative to the NT control. Red shaded boxes indicate no-significant difference and green shaded boxes indicate there is a significant difference  $p < 0.05$ . Data obtained from 3-6 biological replicates per condition.

**Fig.5. ChIP-qPCR demonstrated that SCX interacts with candidate gene promoter regions differently in adult and fetal tenocytes.** Chromatin immunoprecipitation was performed on adult and fetal tenocytes using anti-scleraxis antibody or IgG. Immunoprecipitates were subjected to qPCR using primers flanking E-box binding sites within the gene's promoter regions. Genomic DNA was used as a positive input control. Results of  $n = 3$  biological replicates are presented as a fold enrichment of the ratio of SCX-bound DNA to the average of the negative controls (TEX33, RNASE9, Intronic #1 and Intronic #2) normalised for input DNA, with error bars representing the SEM. Grey bars represent the fold change in enrichment of SCX-bound DNA for the average negative control genes. Blue bars represent the fold change in enrichment of SCX-bound DNA for each gene analysed. \*  $p < 0.05$  using a two-tailed Student's  $t$ -test.

**Fig.6. Mechanisms of SCX transcription factor (TF) binding.** SCX can form homodimers (Gene A) or heterodimer (Gene B) with other basic helix-loop-helix (bHLH) proteins in order to regulate gene expression, with the alternative splice variant of the E2A gene E12 and E47 being the most common (Bagchi et al., 2016a; Carlberg et al., 2000; Cserjesi et al., 1995; Shukunami et al., 2018). As dimers they bind to E-boxes in the promoter regions of target genes and upon binding can either directly recruit RNA polymerase or recruit accessory transcription factors in order to promote transcription (Gene C), one example of this being the interaction between SCX and SMAD (Bagchi et al., 2016b). SCX has also been shown to participate in cooperative binding where multiple proteins interact to regulate gene transcription (Gene D) (Furumatsu et al., 2010). Although SCX, as well as many other TFs, have been typically classified as either being an "activator" or a "repressor", this idea is being repeatedly questioned (Lambert et al., 2018). As SCX regulation is not very well studied there is a possibility that it too could have repressive actions. This could be through binding to promoter regions and blocking other TFs from binding, thus preventing transcription (Gene E), or alternatively through upregulating the expression of a transcriptional repressor.

**Supplementary Fig.1. Demonstration of consistency of SCX knockdown in each biological line of adult and fetal tenocytes.** (A) Four biological replicates of adult and fetal tenocytes exhibit consistent and comparable reduction in SCX mRNA expression following viral transduction with a specific shRNA against SCX (shSCX) when compared to cells transduced with a control non-target scrambled shRNA (NT). Arrows highlight the percentage decrease between the NT control and shSCX expressing cells. Using a two-tailed Student's *t*-test no significant difference between percentage knockdown is detected between the adult and fetal lines ( $p = 0.336$ ). Error bars represent the SD of two qPCR replicates. (B) Adult and fetal shSCX or NT copies per diploid cell determined by copy number assay's in each biological line. Water (H<sub>2</sub>O) and wild type (WT) non infected DNA were used as negative controls. Using a two-tailed Student's *t*-test no significant difference in copy number was detected between fetal NT and shSCX line ( $p = 0.462$ ), between adult NT and shSCX lines ( $p = 0.101$ ) or between the adult and fetal infected lines ( $p = 0.548$ ). **C) Immunocytochemistry confirms a reduction in SCX protein (red) in each of the four-biological replicates of adult and fetal tenocytes following shSCX expression used in the RNA-sequencing experiment. Scale bar = 40  $\mu$ m. DAPI staining of the nuclei is shown in blue.**

**Table 1.** Primer sequences

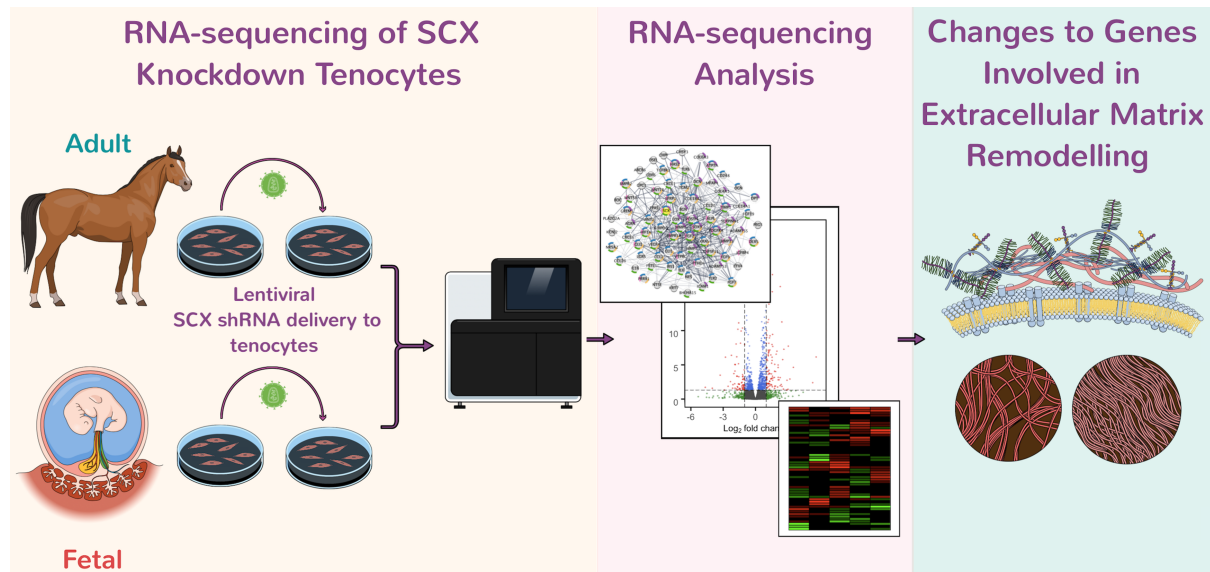
Gene	Forward Primer	Reverse Primer
<i>qPCR Primers</i>		
18s rRNA	CCCAGTGAGAATGCCCTCTA	TGGCTGAGCAAGGTGTTATG
COL1A1	TGCGAAGACACCAAGAAGCTG	GACTCCTGTGGTTTGGTCGT
COL1A2	GTGCCTAGCAACATGCCAAT	GTCCTCTATCTCCGGTTGGG
COL14A1	CTTCATGTTCTGCCTACGGG	CGTCTTGGACAGGGGTGAAT
COMP	AGAACATCATCTGGGCCAAC	CGCTGGATCTCGTAGTCCTC
MMP2	CAGGAGGAGAAGGCTGTGTT	AGGGTGCTGGCTGAGTAGAC
MMP3	TGGACCTGGAAAAGTTTGG	GACCAAGTTCATGAGCAGCA
MMP9	GAGATCGGGAATCATCTCCA	CCAAGAGTCGCCAGTACCTC
SCX	CCCAAACAGATCTGCACCTT	ATCCGCCTCTAACTCCGAAT
SOX9	GCTCTGGAGACTTCTGAACGA	GTAATCCGGGTGGTCCTTCT
THBS4	GGGAAATGGGGTTACCTGTT	CGGGTAGCAGGGATGATATT
TNC	AACCCGTCCAAAGAGACCTT	GCGTGGGATGGAAGTATCAT
TNMD	GTCCCTCAAGTGAAGGTGGA	CCTCGACGGCAGTAAATACAA
VIM	GAGAGCACCTGCAATCC	AAGGTCAAGACGTGCCAAAGA
WPRE	AGCTGACAGGTGGTGGCAAT	TGCTGACGCAACCCCCACTGGT

### ChIP-qPCR Primers

Intro #1	TTTGGGGGAGGTTTTGCTTC	AACCGAAGTCCTGATACTGC
Intro #2	AGAGGAGATTCAACATTCTGC	CCTTTGCAGTATTCAGTGTACC
RNASE9	CCAGCCCTTGAAAGATCCATT	GCTTCCAACCCTATCTTTGCT
TEX33	CTCACAACGATGGAGAAAGGG	CTCCTCTCAGCTCCACAAC
ACTA #1	CATTCCCCTTCAAAAGGTCCA	GGGAATTGCTTACATTTTGGC
ACTA #2	TCCTCTGACCCCCATTCTG	CCTGCTGAAGCGGTTCTATTT
CLDN16	CGTTCAGCATGAGTGACAGA	AAAGCATGGCAAAGTTGGAA
COL1A2 #1	AGAACTGACCAGACACCCTT	CAAAGGCATTGAGTTGGGAT
COL1A2 #2	CTAGTGGGGAGTAGAGAGTGG	TCTCTGGTAACACCCCAAAC
COL1A2 #3	CCAACAGTAGGCGTCCTC	GGTTCTGTCTGTGGAGGGTT
FGF9	CAATACAGCTTGCGCTTGTG	CATGTTGCATTGCGCTAGGA
FGF19	AGCGGTTGGAGGAATAATGAG	AGGGGACTTTGGCTCAACAC
IGF2BP1 #1	GTGAACGGCATGAAATCGTCT	TTCTTGTGATTTTGTGTGCAG
IGF2BP1 #2	GCTTCTCTTTGTCTCTCTCGG	GAGTTTGCCCACCCTACCTC
KLF15	CCTCCACGCATGCTTACTTC	GATTCTGGGAAAGCTCACGA
MMP3 #1	CAAACGAGCTCCAACCTACC	ACAACCTACGCAGAGTCAAGA
MMP3 #2	GGTAAGTAGGTGAGGTTGGTG	AGACCCTAAACATTCCCAGA
MMP3 #3	CCTGAACTAACTGCCACCTT	AGGGTCAGATAATGGCTGGT
NOV	CCCAAACCTTCAGGCATCCTTC	CAAGCAGGGGCTGCGAATGTA
PDGFB	TTTAGCCGGCGAGTGAAGAC	CAAGAGGAAAAAGAACACGGC
TNMD	GGTGGGGATGAAAAGTTGCC	ACAGGTTCTTTGCTCTCCTCT

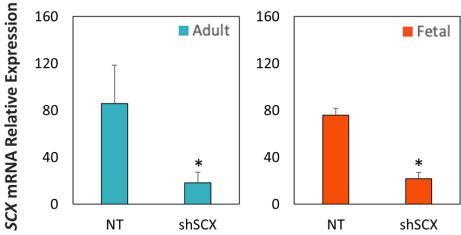
**Table 2:** Cycling conditions for qPCR

Stage	Process	Temperature (°C)	Time (Seconds)
1	Initial denaturation	95	600
2	Denaturation	95	15
3	Annealing	55	15
4	Elongating	72	15
<i>Stages 2-4 repeated 44 times.</i>			
5	Ramp from 60-90°C in 0.5°C increments.		

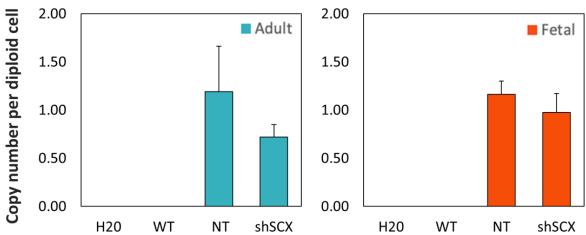


Graphical Abstract

**Fig.1 A**



**B**



**C**

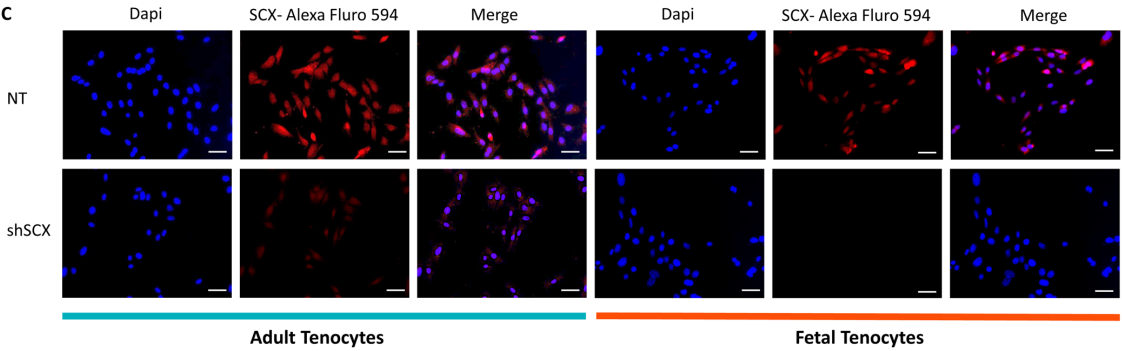
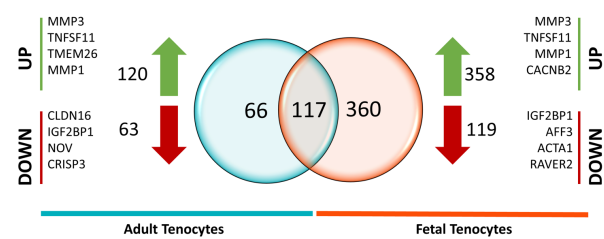
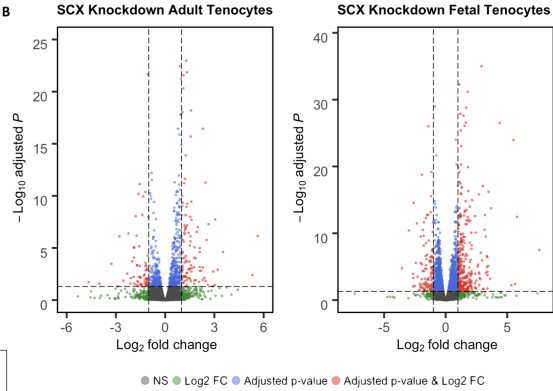


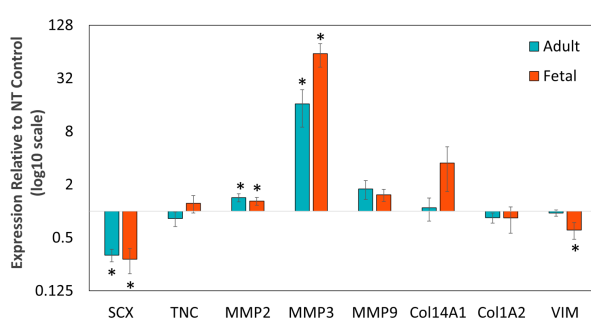
Fig.2 A



B



C



D

Gene	Adult Knockdown		Fetal Knockdown	
	RNA-seq (p-adj.)	Cohort qPCR (p-value)	RNA-seq (p-adj.)	Cohort qPCR (p-value)
SCX	1.26E-09	1.86E-08	4.89E-07	1.32E-05
TNC	0.9131	0.5411	0.2062	0.4242
MMP2	**0.0013	0.0307	**0.0087	0.0498
MMP3	3.05E-08	0.0471	9.49E-15	0.0089
MMP9	0.6110	0.1990	**0.0038	0.0514
COL14A1	0.8404	0.6729	4.02E-27	0.2023
COL1A2	0.6519	0.1529	**0.0001	0.5768
VIM	0.9208	0.7900	0.5380	0.0145

E

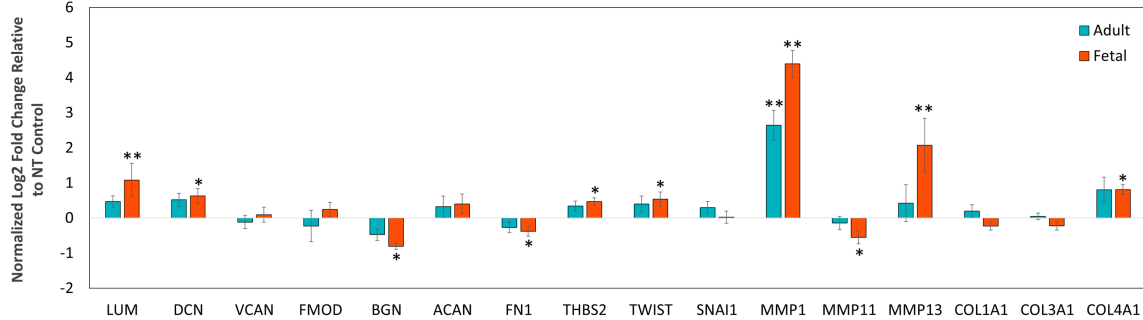




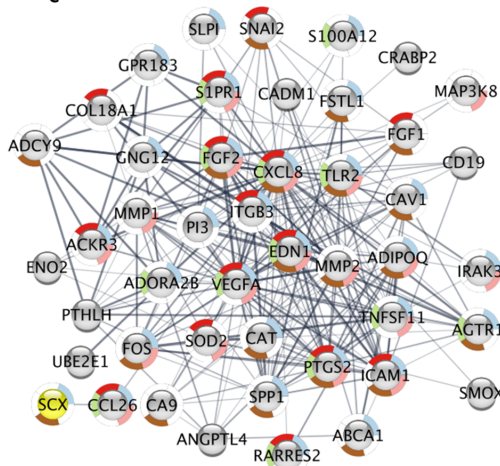
Fig.3 A

PANTHER GO-Slim Overrepresented Biological Process	REFLIST COUNT (14707)	INPUT COUNT	INPUT (expected)	INPUT (Fold Enrichment)	INPUT (Raw P- value)	INPUT (FDR)
<b>Adult Knockdown</b>						
Cellular response to endogenous stimulus (GO:0071495)	638	21	5.32	3.95	1.35E-07	6.22E-04
Cellular response to organic substance (GO:0071310)	1503	34	12.53	2.71	1.16E-07	8.05E-04
Cellular response to organic substance (GO:0071310)	1178	27	9.82	2.75	2.23E-06	6.18E-03
Regulation of response to stimulus (GO:0048583)	2435	42	20.3	2.07	4.74E-06	9.41E-03
Regulation of signal transduction (GO:0009966)	1845	35	15.38	2.28	5.57E-06	8.59E-03
Regulation of signaling (GO:0023051)	2095	38	17.46	2.18	5.82E-06	8.09E-03
Regulation of system process (GO:0044057)	275	12	2.29	5.24	5.07E-06	8.79E-03
Response to endogenous stimulus (GO:0009719)	697	23	5.81	3.96	3.09E-08	4.29E-04
Response to organic substance (GO:0010033)	1488	32	12.4	2.58	8.70E-07	3.02E-03
Response to stress (GO:0006950)	1755	34	14.63	2.32	3.89E-06	8.99E-03
<b>Fetal Knockdown</b>						
Anatomical structure morphogenesis (GO:0009653)	1382	67	30.48	2.2	3.15E-09	4.86E-06
Cellular response to chemical stimulus (GO:0070887)	1503	67	33.15	2.02	1.03E-07	7.52E-05
Positive regulation of response to stimulus (GO:0048584)	1334	62	29.42	2.11	5.88E-08	4.80E-05
Positive regulation of signal transduction (GO:0009967)	990	52	21.84	2.38	1.94E-08	1.92E-05
Regulation of cell migration (GO:0030334)	558	39	12.31	3.17	9.37E-10	2.17E-06
Regulation of cell motility (GO:2000145)	595	39	13.12	2.97	5.06E-09	6.39E-06
Regulation of cellular component movement (GO:0051270)	642	41	14.16	2.9	4.06E-09	5.64E-06
Regulation of locomotion (GO:0040012)	647	39	14.27	2.73	4.28E-08	3.71E-05
Tissue development (GO:0009888)	1072	59	23.64	2.5	3.31E-10	9.20E-07
Tube development (GO:0035295)	586	37	12.92	2.86	3.17E-08	2.94E-05

B

GeneAnalytics Pathway Analysis	Entity Score	Number of Genes in Superpath	Number of Genes Matched
<b>Adult Knockdown</b>			
Carbon Metabolism	9.64	143	6
Cell Adhesion_ECM Remodelling	9.2	61	4
GnRH Secretion	8.95	64	4
Metabolism	10.64	2621	37
Nitrogen Metabolism	11.29	17	3
PAK Pathway	9.28	683	14
PPAR Signaling Pathway	11.13	76	5
Statin Pathway	9.03	63	4
Transcription Role of VDR in Regulation of Genes Involved in Osteoporosis	10.81	45	4
Type II Diabetes Mellitus	9.37	59	4
<b>Fetal Knockdown</b>			
Activation of CAMP-Dependant PKA	20.03	630	34
Akt Signalling	21.83	682	37
CREB Pathway	24.1	529	33
Embryonic and Induced Pluripotent Stem Cell Differentiation Pathways and Lineage-specific Markers	17.03	133	13
ERK Signalling	35.5	1179	63
MAPK Signalling Pathway	17.04	321	21
Oxytocin Signalling Pathway	17.49	216	17
PAK Pathway	23.27	683	38
RhoGDI Pathway	18.65	181	16
TGF-Beta Pathway	20.34	653	35

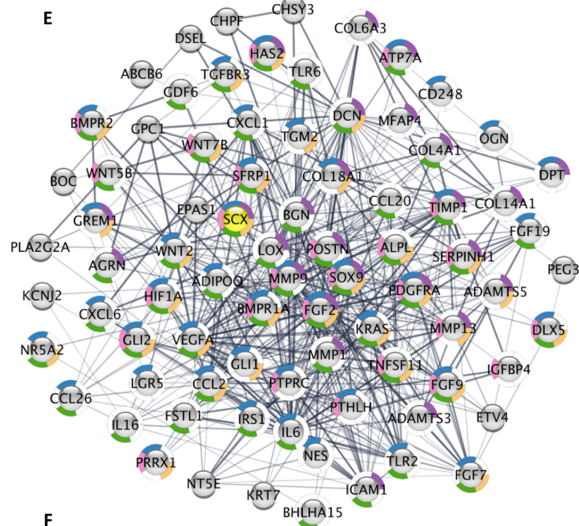
C



D

Colour	GO Term	Description	FDR Value
Blue	GO:0009605	Response to external stimulus	2.46E-15
Red	GO:0030335	Positive regulation of cell migration	2.54E-11
Green	GO:0032103	Positive regulation of response to external stimulus	3.37E-11
Brown	GO:0009719	Response to endogenous stimulus	6.55E-10
Pink	GO:0034097	Response to cytokines	7.29E-10

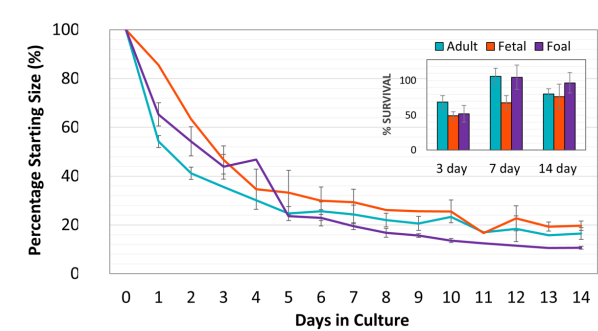
E



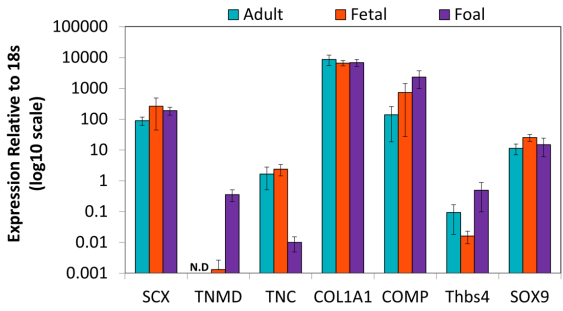
F

Colour	GO Term	Description	FDR Value
Purple	GO:0030198	Extracellular matrix organisation	2.27E-23
Dark Blue	GO:0042127	Regulation of cell population proliferation	2.48E-21
Light Blue	GO:0001501	Skeletal system development	8.74E-20
Yellow	GO:0010033	Response to organic substance	5.99E-19
Orange	GO:0034097	Animal organ morphogenesis	7.29E-10

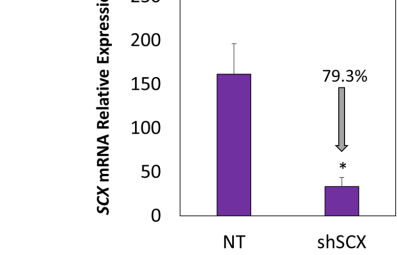
Fig.4 A



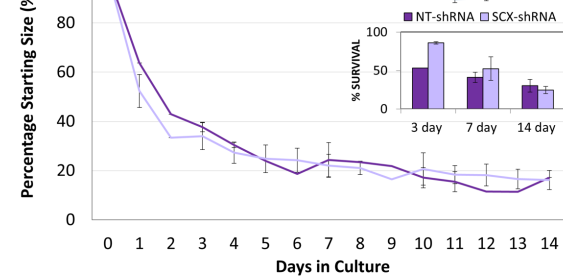
B



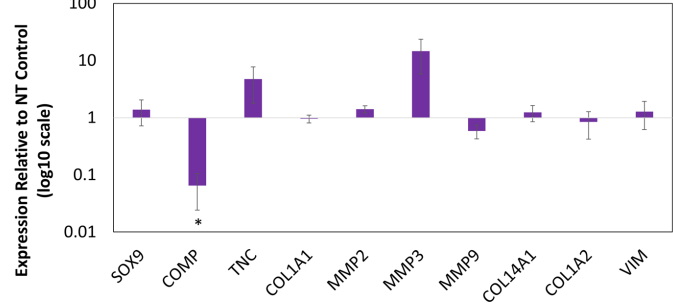
C



D



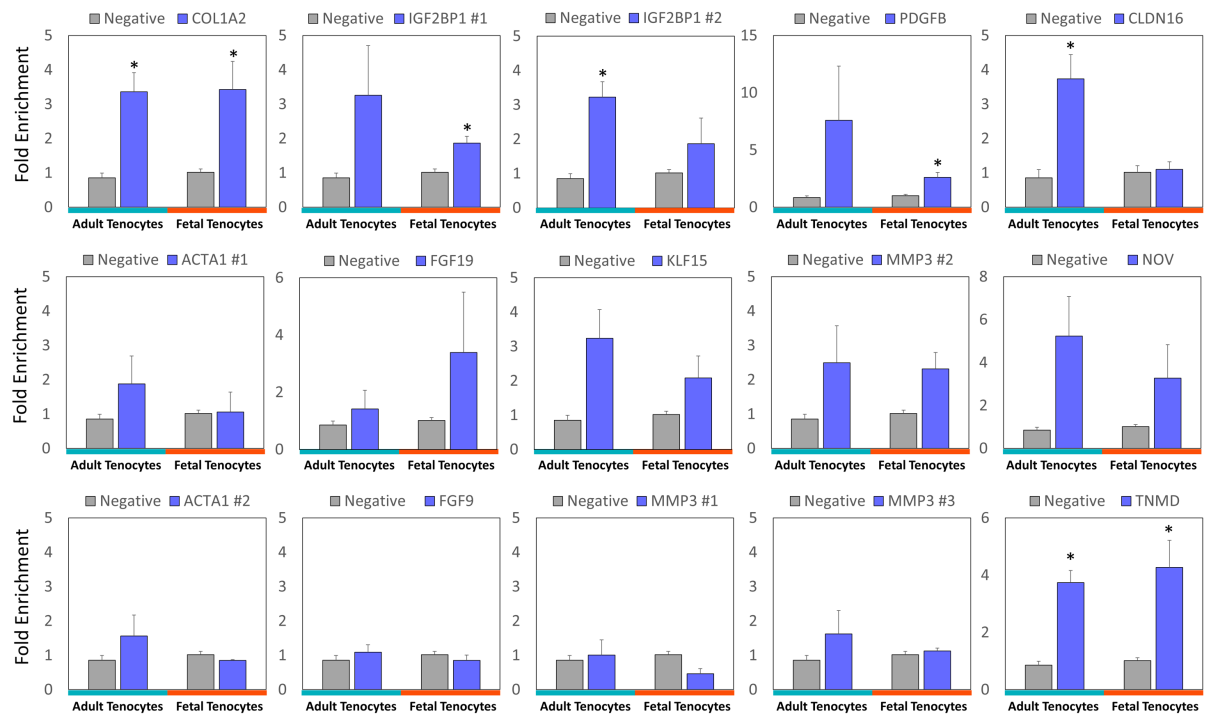
E



F

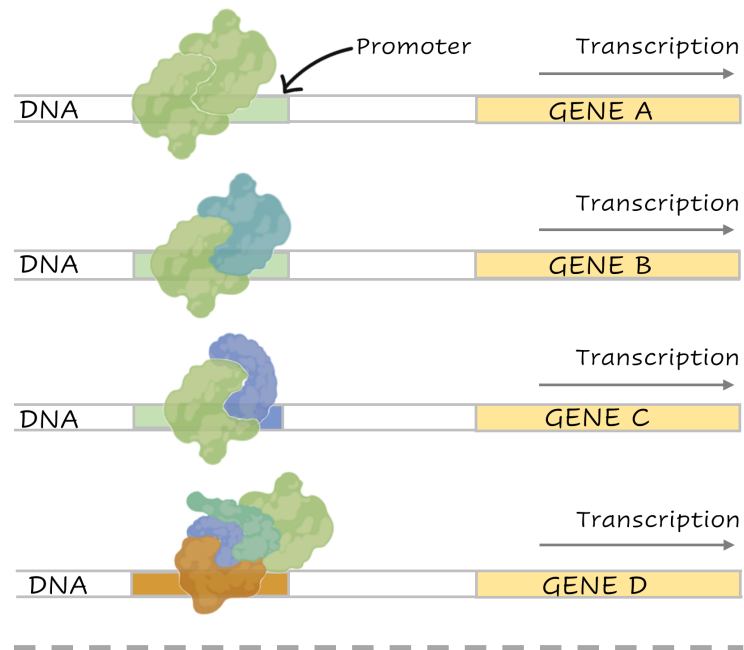
	Adult Knockdown	Fetal Knockdown	Foal Knockdown
Gene	qPCR (p-value)	qPCR (p-value)	qPCR (p-value)
SCX	1.86E-08	1.32E-05	0.0005
SOX9	0.4984	0.9022	0.5930
COMP	0.4032	0.8222	2.11E-05
TNC	0.5411	0.4242	0.2725
COL1A1	0.1109	0.0592	0.8093
MMP2	0.0307	0.0498	0.1099
MMP3	0.0471	0.0089	0.1894
MMP9	0.1990	0.0514	0.0588
COL14A1	0.6729	0.2023	0.5681
COL1A2	0.1529	0.5768	0.7355
VIM	0.7900	0.0145	0.6983

Fig.5



**Fig.6**

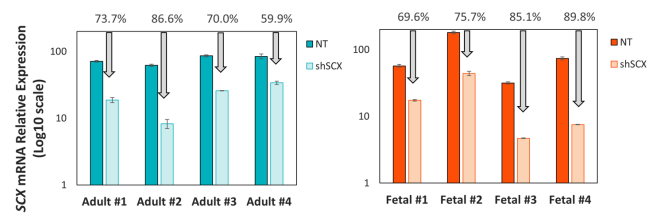
**Known Binding Interaction Capabilities**



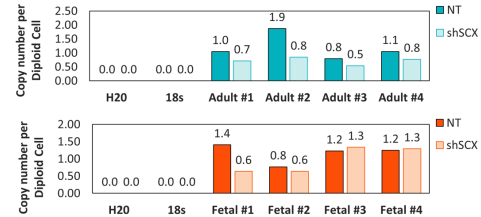
**Theoretical Binding Interaction**



Supplementary Fig.1 A



B



C

