Article

A Functional Single-Nucleotide Polymorphism Upstream of the Collagen Type III Gene Is Associated with Catastrophic Fracture Risk in Thoroughbred Horses

Esther Palomino Lago [1], Arabella Baird [2], Sarah C. Blott [3], Rhona E. McPhail [2], Amy C. Ross [1], Sian A. Durward-Akhurst [4] and Deborah J. Guest [1,*]

[1] Department of Clinical Sciences and Services, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield AL9 7TA, UK; epalominolago@rvc.ac.uk (E.P.L.); aross21@rvc.ac.uk (A.C.R.)
[2] Animal Health Trust, Lanwades Park, Kentford, Newmarket CB8 7UU, UK
[3] School of Veterinary Medicine and Science, University of Nottingham, Nottingham LE12 5RD, UK; sarah.blott@nottingham.ac.uk
[4] Department of Veterinary Clinical Sciences, University of Minnesota, Saint Paul, MN 55108, USA; durwa004@umn.edu
[*] Correspondence: djguest@rvc.ac.uk

Simple Summary: Bone fractures are a major welfare concern in Thoroughbred racehorses and often lead to euthanasia. In this study, we established a cell model to investigate the genetic factors involved in fracture risk. We found that collagen type III is expressed at lower levels in cells from horses with a high genetic risk of fracture and identified a novel DNA variant that is significantly associated with fracture and can affect the expression level of collagen type III.

Abstract: Fractures caused by bone overloading are a leading cause of euthanasia in Thoroughbred racehorses. The risk of fatal fracture has been shown to be influenced by both environmental and genetic factors but, to date, no specific genetic mechanisms underpinning fractures have been identified. In this study, we utilised a genome-wide polygenic risk score to establish an in vitro cell system to study bone gene regulation in horses at high and low genetic risk of fracture. Candidate gene expression analysis revealed differential expression of *COL3A1* and *STAT1* genes in osteoblasts derived from high- and low-risk horses. Whole-genome sequencing of two fracture cases and two control horses revealed a single-nucleotide polymorphism (SNP) upstream of *COL3A1* that was confirmed in a larger cohort to be significantly associated with fractures. Bioinformatics tools predicted that this SNP may impact the binding of the transcription factor SOX11. Gene modulation demonstrated SOX11 is upstream of *COL3A1*, and the region binds to nuclear proteins. Furthermore, luciferase assays demonstrated that the region containing the SNP has promoter activity. However, the specific effect of the SNP depends on the broader genetic background of the cells and suggests other factors may also be involved in regulating *COL3A1* expression. In conclusion, we have identified a novel SNP that is significantly associated with fracture risk and provide new insights into the regulation of the *COL3A1* gene.

Keywords: fracture risk; horse; genetic variation; bone; osteoblast; in vitro model

1. Introduction

Bone fractures caused by overloading rather than direct trauma can occur in Thoroughbred racehorses and, due to the complexities involved in treating fractures in a large animal that must remain weight-bearing, many of these can be catastrophic and result in euthanasia. Although the incidence of catastrophic fracture is low (0.38 per 1000 stars in the UK [1]), bone fractures are the leading cause of deaths occurring on the racecourse [2], with approximately 60 horses/year suffering from a fatal fracture during a race in the UK [1].

Fractures therefore have a significant welfare and economic impact on the racing industry. Many environmental risk factors have been identified for catastrophic fracture [3–7], but there is also a genetic contribution to risk. The heritability of distal limb fracture has been found to range from 0.21–0.37 in UK Thoroughbreds (flat and National Hunt) [8], and multiple genomic locations have been identified that may contribute to fracture risk, including a 5 Mb region on equine chromosome 18 (ECA18) in UK flat and national hunt Thoroughbreds [9,10]. A candidate SNP study using SNPs within this region also confirmed an association with fractures in Japanese flat racing Thoroughbreds [11]. Similarly, both genetic [12–15] and environmental [16,17] risk factors have been identified for stress fractures in humans. However, fracture risk is a complex trait resulting from the effects of multiple genes, and no specific functional mechanisms underpinning fracture have been defined in either species. To enable better prevention of catastrophic fractures, we need to develop an improved understanding of the biological processes affected by genetic factors.

The 5 Mb region on ECA 18 that is associated with catastrophic fracture risk in horses [9] contains 27 protein-coding genes with a known function. Of these, eleven genes have some association with fracture or bone formation, metabolism and remodelling: *ZNF804A* [9,18], *ITGAV* [19], *CALCRL* [20], GULP1 [21], *COL3A1* [22], *COL5A2* [23], *SLC40A1* [24], *MSTN* [25,26], *C2orf88* [27], *STAT1* [28–30], *GLS* [31]. Further investigation of these genes in equine fractures is therefore warranted.

However, studying the genetic mechanisms and biological processes predisposing horses to fracture is difficult to perform in vivo due to the large variation in exposure to environmental factors of fracture risk (e.g., training and racing histories) and the fact that bone tissue can only be isolated postmortem. Small animal models are very valuable for defining gene functions [32] but are unlikely to be useful in studying specific DNA variants, as causal variants are often poorly conserved between species [33]. To overcome these limitations, cell models have been widely used to study human diseases [34]. Often, these are simple inherited diseases, but for complex genetic diseases, cells from affected patients have been compared with cells from unaffected donors [35–37]. However, more recently, these models have been refined by the selection of cells based on their polygenic risk score [38,39]. In contrast to humans, the use of cell models to study inherited diseases in horses has not been widely reported.

The aim of this study was to develop a cell model to study the genetic basis of fracture risk in horses. We first established genome-wide polygenic risk scores (Genomic Best Linear Unbiased Prediction, GBLUP) [40,41] for fractures. The risk scores were then utilized to create an in vitro disease model using bone cells from horses at genetically high and low risk of fracture to identify candidate genes that were differentially expressed. Whole-genome sequencing was performed to identify potential regulatory DNA variants, and further cell models were then used to investigate the mechanisms of gene regulation.

2. Materials and Methods

An overview of the experimental approach is shown in Figure 1.

2.1. Skin Fibroblast Cell Isolation and Culture

Equine skin fibroblasts were isolated postmortem from 44 Thoroughbred horses (of unknown fracture status) and 3 Welsh Mountain ponies to establish the cell models. Thoroughbred samples were used exclusively in experiments that determined the effects of genetic variation for fracture. Welsh Mountain pony samples were used only in experiments that were not specifically related to fracture risk. All animals had been euthanised for reasons unrelated to this study and with the consent of the Animal Health Trust Ethical Review Committee (AHT_02_2012) and Royal Veterinary College Clinical Research Ethical Review Board (URN 2021 2035-2). Briefly, small dermal tissue pieces were collected into standard culture media (Dulbecco's modified Eagle's medium (DMEM) containing 10% foetal bovine serum, 2 mM L-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin (all ThermoFisher, Loughborough, UK) plus 2.5 µg/mL amphotericin B and stored at 4 °C

for up to 24 h. Tissue samples were then dissected into small pieces and digested in 1 mg/mL collagenase from *C. histolyticum* (Sigma-Aldrich, Dorset, UK) in standard culture media plus 2.5 μg/mL amphotericin B overnight at 37 °C, 5% CO_2. Dissociated cells were washed and cultured in standard culture media (without amphotericin B) and passaged using 0.25% trypsin-EDTA (Sigma-Aldrich) every 3–5 days.
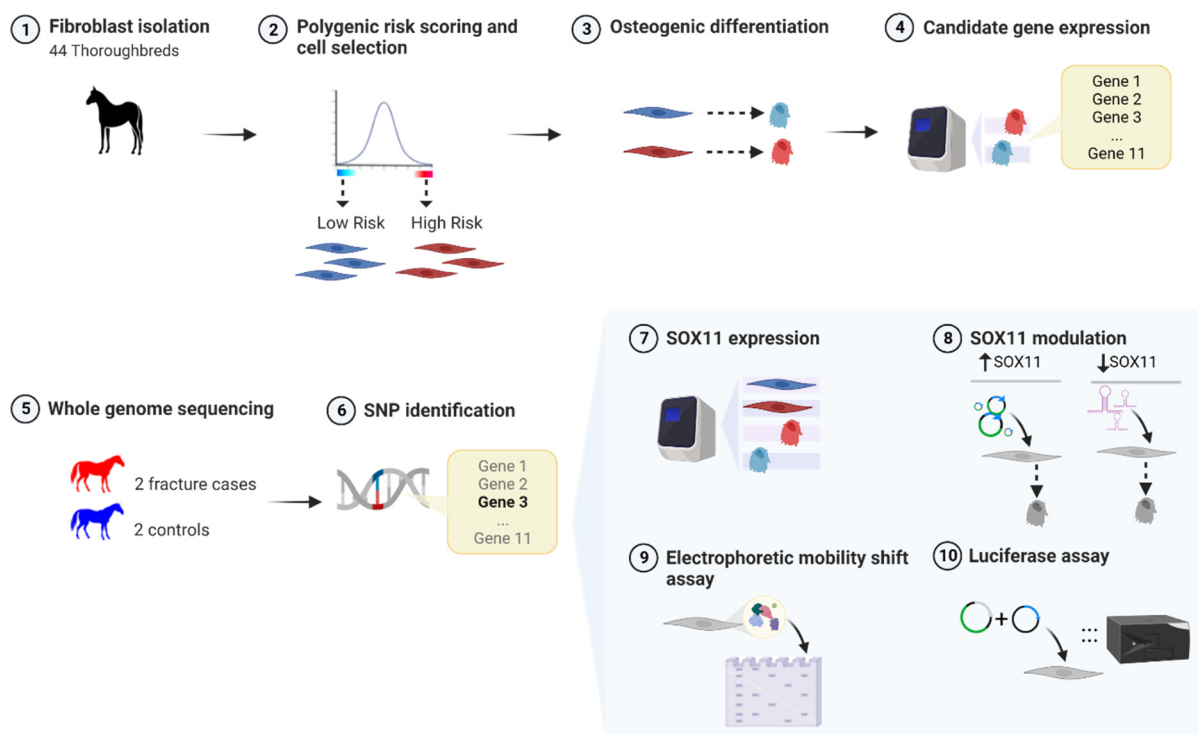

Figure 1. An overview of the experimental approach used demonstrating the order of the experiments. Red cells indicate cells from horses with a high polygenic risk score (PRS), and blue cells indicate cells from horses with a low PRS. Grey indicates cells from horses and ponies with either unknown or a range of risk scores. Created with BioRender.com (accessed on 6 June 2023).

2.2. Polygenic Risk Scoring of Thoroughbred Skin Fibroblasts

Genome-wide Complex Traits Analysis (GCTA) [42] was used on Equine SNP50 genotyping data from 269 cases and 253 controls from the previous study [9] ("the training data set") to derive BLUP (Best Linear Unbiased Prediction) allele effects for all individual SNPs across the genome (genomic BLUP, GBLUP). This is equivalent to deriving the heritability explained by each individual SNP, a prerequisite for polygenic risk score analysis [40]. DNA was extracted from skin fibroblasts using the QIAamp DNA mini kit (Qiagen, Manchester, UK) and genotyped on the Geneseek Genomic Profiler Equine Plus chip (based on Illumina's Equine SNP70 platform) (performed by Neogen, Ayr, UK), which includes over 70,000 evenly distributed SNPs, including those on the Illumina Equine SNP50 chip. Polygenic risk scores for the skin fibroblast samples were calculated by applying the PLINK (http://pngu.mgh.harvard.edu/purcell/plink/ (accessed on 12 August 2020) [43] score function to the fibroblast genotypes, in conjunction with the estimated SNP BLUP effects obtained from the training data set, applying the GBLUP protocol as described by Chung [41]. Additionally, we carried out genotyping and scoring of 382 DNA samples that had been collected from blood samples taken from Thoroughbreds (flat and National hunt) of unknown fracture status. These Thoroughbreds therefore represent the general population. Having risk scores from a large number of horses representing the general Thoroughbred population enabled us to then select the skin samples (from the separate bank of 44 samples that were collected) that represented the ends of the risk spectrum (see Figure 2). These

samples were from male horses. All DNA samples were used with the approval of the Royal Veterinary College Clinical Research Ethical Review Board (URN 2020 2004-2).
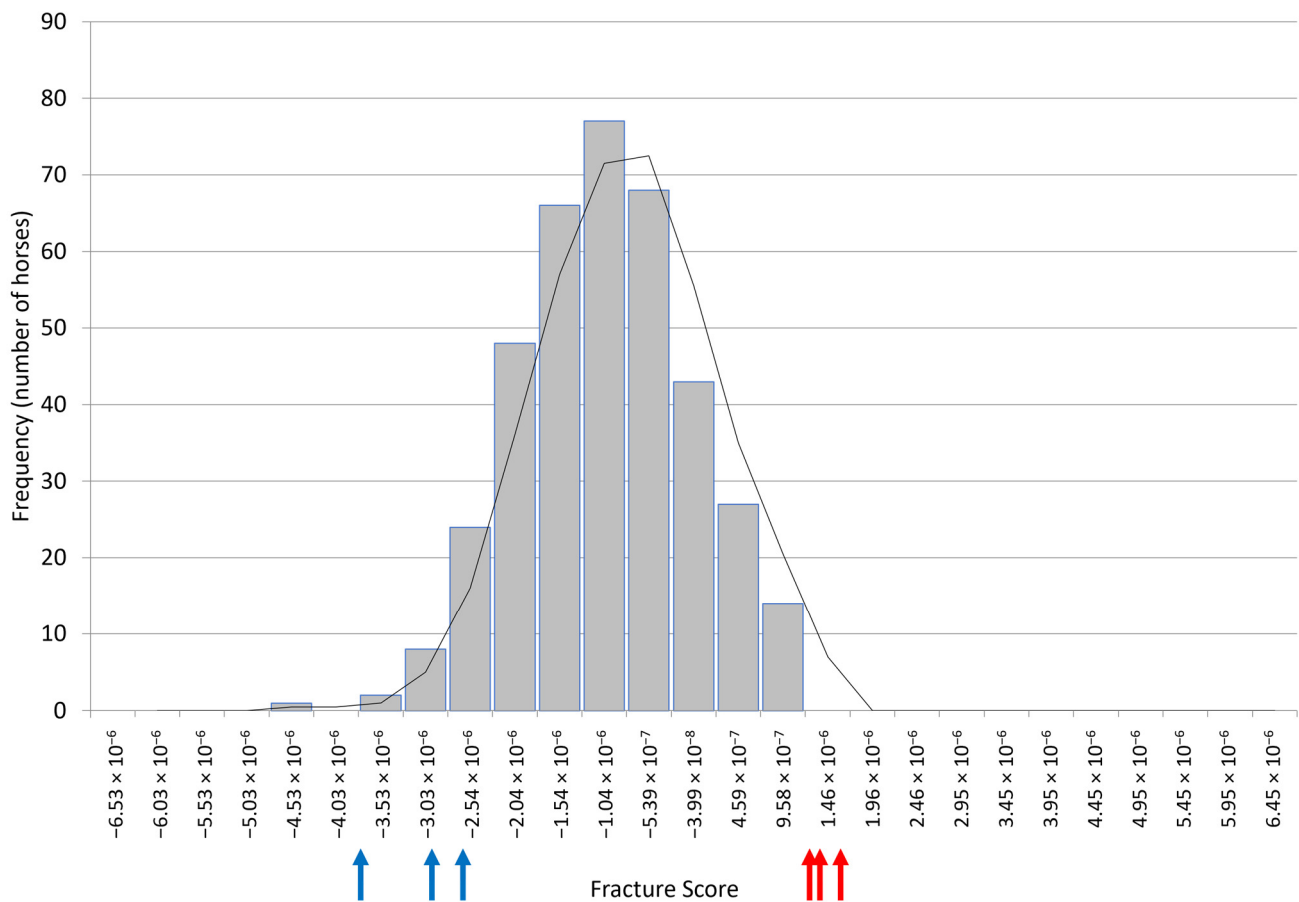

Figure 2. The polygenic risk scores of 382 Thoroughbreds of unknown fracture status (representing the general population) with arrows indicating the scores for the skin cells selected as the three highest (red) and three lowest (blue) risks.

2.3. Osteoblast Differentiation

To develop a relevant cell model, we promoted osteoblast-like cell differentiation of the primary equine fibroblast cells. Once cells reached approximately 70% confluency, they were cultured in osteogenic media (standard culture media plus 10 nM dexamethasone, 28 µM ascorbic acid, and either 2 mM or 10 mM β-glycerophosphate (all Sigma-Aldrich)) for 21 days at 37 °C, 5% CO_2, with the media replaced every 3–4 days.

2.4. Histological Staining

To determine matrix mineralisation following osteoblast-like cell differentiation, cells were stained with von Kossa (Abcam, Cambridgeshire, UK) according to the manufacturer's instructions. Calcium deposition was detected by incubating the cells with 2% Alizarin red S pH 4.2 for 5 min. Hydroxyapatite deposition was detected using the OsteoImage bone mineralisation assay (Lonza, Slough, UK).

2.5. RNA Extraction, cDNA Synthesis, and Quantitative PCR

RNA was extracted from skin cells pre and post osteogenic differentiation and following *SOX11* overexpression and knockdown. RNA was extracted using Tri-reagent (Sigma-Aldrich) and purified using an RNeasy mini kit (Qiagen) according to the manufacturer's instructions. Genomic DNA contamination was removed using Invitrogen™ DNA-free™ DNA Removal Kit (ThermoFisher). A total of 1 µg of RNA was used to

prepare cDNA using the Sensifast cDNA synthesis kit (Bioline, London, UK). In total, 2 μL aliquots of cDNA (corresponding to 20 ng of cDNA) were used in qPCR using SYBR green containing supermix (Bioline) on the Biorad CFX96 thermal cycler (Biorad, Hertfordshire, UK) in duplicate. Primers to equine-specific genes (note *SOX11* primers detect both the human and equine genes) were designed using NCBI primer blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) (accessed on 25 March 2019) and mfold (http://www.unafold.org/) (accessed on 25 March 2019) programs to produce amplicons of 50–150 bp devoid of secondary structures at Tm 60 °C and a melting temperature ™ of 58–62 °C and. Primer sequences are in Table S1. qPCR cycling parameters were 95 °C for 10 min, followed by 45 cycles of 95 °C for 15 s, 60 °C for 15 s, and 72 °C for 15 s. Gene expression was measured relative to the 18 s rRNA housekeeping gene using the $1/2^{\Delta\Delta CT}$ [44]. The 18 s rRNA was selected as a housekeeping gene based on its stability across the samples and experimental conditions being tested. Stabilities of 18 s rRNA, GAPDH and ACTB housekeeping genes were evaluated across the experiments using RefFinder [45] (https://www.heartcure.com.au/reffinder, (accessed on 12 February 2020) Supplementary Figure S1).

2.6. Immunocytochemistry

Skin fibroblasts pre and post osteogenic differentiation were cultured on gelatin-coated coverslips (Sigma-Aldrich), fixed in 3% paraformaldehyde for 20 min, and permeabilised for 1 h with 0.1% triton-X-100 prior to blocking in 2.5% normal horse serum (Vector Laboratories, Peterborough, UK) for 20 min at room temperature. Primary antibody incubations were performed at 4 °C overnight prior to incubation with the secondary antibody. All antibodies were used at optimised concentrations in blocking solution (Table S2). Coverslips were mounted onto glass sides using Vectashield Hardset mounting media with DAPI (Vector Laboratories) and imaged using a Nikon Eclipse Ti2 series microscope (Nikon, Surrey, UK). For RUNX2 and STAT1 staining, the intensity of nuclear staining was quantified using ImageJ [46].

2.7. Whole-Genome Sequencing

Whole-genome sequencing to identify novel DNA variants associated with fractures was performed on two Thoroughbred fracture cases and two Thoroughbred controls. The DNA samples were isolated as part of the previous study [9] used to generate the polygenic risk scores (i.e., part of the discovery cohort used to develop the "training data set" as described in Section 2.2). The fracture cases sustained distal limb fractures (one horse had a sesamoid fracture and one horse had a lateral condylar fracture) that required euthanasia while racing on a UK racecourse. The fracture controls were uninjured horses over four years of age with no history of fracture and had been racing during the same time period as the cases. Whole-genome sequencing was performed by Source Bioscience (Cambridge, UK) on 100 bp (base pair) PE (paired end) reads using an Illumina HiSeq 2000 to generate 30–36× coverage per sample. Reads were aligned to the reference genome (Equcab 3.0) and variant calls made using GATK (HaplotypeCaller) [47]. Base quality score recalibration, and indel realignment and duplicate removal were carried out and SNP discovery was performed according to GATK Best Practices [48]. Genomic Variant Call Format (VCF) files were combined into a single VCF file and Variant Effect Predictor (VEP) [49] used for cross-genome analysis. The effect of SNPs on transcription factor binding was predicted using TomTom [50]. The minor allele frequencies were calculated from publicly available datasets, including horses representing multiple breeds (http://gong_lab.hzau.edu.cn/Animal_SNPAtlas [51] (accessed on 4 June 2023)), 101 Japanese Thoroughbreds [52], and other individual breeds, including US/European Thoroughbreds [53]. The genotypes of five SNP heterozygotes and five SNP homozygotes were confirmed using the Integrative Genome Viewer [54]. Histone modifications in the region containing the SNP were confirmed using ENCODE data [55] and FANNGmine (version 1.3, https://faangmine.rnet.missouri.edu/ (accessed on 5 June 2023)).

2.8. Genotyping for the COL3A1 SNP

To determine the association of the *COL3A1* SNP with fracture allelic discrimination, assays were performed on 86 Thoroughbred fracture controls and 91 Thoroughbred fracture cases, isolated as part of the previous study [9] (i.e., part of the discovery cohort used to develop the "training data set"). Reactions were carried out in 8 μL volumes consisting of 2 μL (100 ng) genomic DNA in 0.2 μL of $40\times$ assay mix, 4 μL Luna universal qPCR Master Mix (New England Biolabs, Hitchin, UK), and 1.8 μL ultrapure water. Cycling parameters were 25 °C for 3 s and 95 °C for 3 min, followed by 35 cycles of 95 °C for 3 s and 60 °C for 10 s, with a final extension of 25 °C for 30 s. The primer and probe sequences were 5′-TCTTGAGGAATTGGTCACAGGA-3′ (forward primer); 5′-AGGACTTGGGCATATCTAACCA-3′ (reverse primer); CAAATATT**T**GCCTGTCAGGTCCCACC (VIC, 4,7,2′-trichloro-7′-phenyl-6-carboxyfluorescein); and CAAATATT**G**GCCTGTCAGGTCCCACC (FAM, 6-carboxyfluorescein). Allelic discrimination was performed using an Applied Biosystems StepOnePlus™ Real-Time PCR system (ThermoFisher) and results were analysed using Applied Biosystems StepOne Software v2.3.

2.9. Overexpression of SOX11

To determine the role of SOX11 in regulating *COL3A1*, SOX11 overexpression assays were performed. First, 1×10^6 skin fibroblasts (from one Thoroughbred and two Welsh Mountain ponies as this experiment is independent of breed) were suspended in 400 μL of standard culture media lacking penicillin/streptomycin and either 20 μg of pCMV6-AC-GFP human *SOX11* (NM_003108; Origene, Rockville, MD, USA) or pCAG-eGFP [56], which was a gift from Peter Andrews (University of Sheffield, Sheffield, UK). The human and equine SOX11 proteins are 90% identical. Electroporation was performed in a 4 mm gap cuvette with one 35 ms pulse of 170 V using a BTX EM830 (BTX, San Diego, CA, USA). Then, 48 h after electroporation, antibiotic resistance was used to select stably transfected cells using 1 mg /mL G418 or 3.5 μg/mL puromycin (both Sigma-Aldrich), respectively, for up to 14 days. Following selection, undifferentiated cells were harvested for analysis or differentiated into osteoblasts for 21 days as above prior to analysis.

2.10. SOX11 Knockdown

To determine the role of SOX11 in regulating *COL3A1*, *SOX11* knockdown assays were performed. Lentiviral vectors TRC1-pLK0.1 containing a nontarget, scrambled (NT) shRNA (5′-GCGATAGCGCTAATAATTT-3′ SHC202; Sigma-Aldrich) or a shRNA specific for human *SOX11* (100% identity to the equine SOX11 sequence) (5′-CTGGTGGATAAGGATTTGGAT-3′ clone NM_003108.3-1235s1c1; Sigma-Aldrich) were created by transfection into HEK293T packaging cells. As a positive control to visualise successful transfection and transduction, TRC2-pLKO.5-puro-CMV-TurboGFP plasmid was used (SHC 203; Sigma-Aldrich). HEK293T cells were plated at 1×10^5 cells per well of a 6-well plate and transfected with 1 μg of TRC2-pLK0.1 plasmid, 750 ng psPAX2, and 250 ng pMD2.G. pMD2.G and psPAX2 were a gift from Didier Trono (Addgene plasmids #12259 and 12260; http://n2t.net/addgene:12259 (accessed on 10 March 2023), http://n2t.net/addgene:12260 (accessed on 10 March 2023); RRID:Addgene_12259 and 12260). Packaging cell supernatant was collected after 48 h, filtered through a 0.45 μm filter (Millipore, Billerica, MA, USA), and used immediately. Skin fibroblasts (from three Thoroughbreds and one Welsh Mountain pony as the experiment is independent of breed) were seeded at 2×10^5 cells per well of a 6-well plate 24 h before infection. To achieve 80% of cells transduced (determined via GFP expression), up to three rounds of infection were performed in the presence of 10 μg/mL polybrene (Sigma-Aldrich). Next, 48 h after the last round of infection, puromycin (Sigma-Aldrich) selection was carried out as described above, and then cells were used for RNA extraction.

2.11. ELISA

To determine COL3A1 protein levels, whole-cell protein extract was prepared from undifferentiated skin cells (overexpressing SOX11 or control) using three rounds of freeze-thawing in extraction buffer containing 20 mM Hepes pH 7.9, 450 mM NaCl, 0.4 mM EDTA, 25% glycerol, 1 mM PMSF, and the supernatant collected by centrifugation. An equine-specific COL3A1 ELISA (EKU11662, Biomatik, DE, USA) was subsequently used to measure the amount of COL3A1 in 300 ng of total whole-cell extract. Measurements were made on a microplate absorbance reader (Mplex (Infinite M200); Tecan, Männedorf, Switzerland).

2.12. Electrophoretic Mobility Shift Assay (EMSA)

To determine if the DNA region containing the *COL3A1* SNP could bind to proteins nuclear protein extract (NPE) was extracted from skin fibroblasts derived from two different Welsh Mountain ponies (the experiment is independent of breed) that were overexpressing human SOX11 (as described above) using NE-PER Nuclear and Cytoplasmic Extraction Reagents (Thermofisher) according to the manufacturer's instructions. EMSAs were carried out in accordance with the protocol of LightShift chemiluminescent EMSA kit (Thermofisher). For this purpose, 20 fmol of biotinylated double-stranded oligonucleotides (equine sequences containing either the reference or alternative DNA variant), 5′-biotin-GAATCACTCAAATATT̲TGCCTGTCAGGTCC-3′ or 5′-biotin-GAATCACTCAAATATT̲GGCCTGTCAGGTCC-3′, were incubated in 20 μL of reaction mixture (10× binding buffer, 2.5% (v/v) glycerol, 5 mM MgCl$_2$, 50 ng/μL poly (dI·dC), 0.05% NP-40, 50 mM KCl, 10 mM EDTA) at room temperature for 20 min. For the shift assay, labelled probes were incubated with a total of 30 μg of NPE. For the competitor assay, the nuclear extract was preincubated with a 200-fold molar excess of unlabelled probes before adding the labelled probe. For the supershift assay, the nuclear extract was incubated for 1 h at room temperature with 5.8 μg of rabbit anti-SOX11 (sab4200450) (Sigma) after adding the labelled probe. A total of 5 μL of loading buffer was then added to stop the reaction. The samples were loaded onto Invitrogen™ DNA Retardation Gels (6%) (ThermoFisher) and run in 0.5X TBE for 1 h at 100 V. The gel was transferred to Invitrogen™ BrightStar™-Plus Positively Charged Nylon Membrane (Thermofisher) for 30 min at 380 mA. DNA membrane was cross-linked for 10–15 min on a transilluminator equipped with 312 nm UV bulbs. The detection was performed using LightShift chemiluminescent EMSA kit (Thermofisher) according to the manufacturer's instructions. Technical controls for the assay were carried out using Epstein-Barr Nuclear Antigen probes and protein extracts as provided in the kit.

2.13. Construction of the Luciferase Reporter Plasmids

To determine if the DNA region containing the *COL3A1* SNP had promoter activity, luciferase assays were performed. The luciferase reporter plasmid was generated using a region 5′ upstream of the *COL3A1* gene (ECA18:65,598,913–65,602,132 EquCab 3.0) from −3186 bp to +34 bp relative to the transcription start site (see Supplementary Figure S2). The sequence carrying the reference allele (TT) was PCR-amplified from genomic homozygous TT allele Thoroughbred DNA using the primers: Forward 5′-ctattggtaccGAGGAATTGGTCACAGGAATCAC-3′ and Reverse 5′-ctggaagcttGCAGTTCAAAGTAGCACCATC-3′, showing restriction sites for KpnI and HindIII, respectively. The alternative allele (GG) fragment was created from the reference allele (TT) PCR product using the Q5® Site-Directed Mutagenesis Kit (New England BioLabs, Ipswich, MA, USA). The sequence of primers carrying out the alternative allele, Forward 5′-CTCAAATATTGGCCTGTCAGG-3′ and Reverse 5′-TGATTCCTGTGACCAATTC-3′, were designed using online NEB primer design software, NEBaseChanger™ (https://nebasechanger.neb.com/ (accessed on 12 February 2021). pGL4.10[Luc2] (#E6651 Promega, Southampton, UK) and PCR products were digested with KpnI and HindIII, and column-purified. The KpnI- HindIII fragments with the reference or alternative allele were then cloned into the previously digested pGL4.10[Luc2] vector. In each step, PCRs were per-

formed by Q5 High-fidelity DNA Polymerase (New England Biolabs), and PCR products were verified by gel electrophoresis and sequencing (performed by DNA Sequencing and Service, University of Dundee, UK). In each case, ligation products were transformed in DH5α competent cells (ThermoFisher), which were then plated onto ampicillin-containing (100 µg/mL) LB agar (Sigma-Aldrich), and individual colonies were analysed by restriction enzyme digest and sequencing after 24 h of growth at 37 °C.

2.14. Luciferase Assays

Skin fibroblasts from six Thoroughbreds (0.3×10^5) with a range of overall risk scores were transfected in 24-well plates using Lipofectamine3000 DNA transfection reagent (ThermoFisher). The cells in each well were co-transfected with 1.4 µg of reference allele (TT) or alternative allele (GG) COL3A1-5′UTR firefly luciferase reporter vector and 0.14 µg of internal control Renilla luciferase vector, pGL4.74[hRluc/TK] (Promega), as the control for transfection efficiency. pGL4.10[Luc2] vector with no insertion was used as a promoter-less negative control. Then, 48 h after transfection, cells were harvested and assayed using Dual-Glo® luciferase assay system (Promega). The luciferase activities were measured using luminescence reader (Mplex (Infinite M200); Tecan). Promoter activity was expressed as relative firefly luciferase activity after normalization to Renilla luciferase activity.

2.15. Statistical Analysis

Statistical analysis of qPCR, immunocytochemical quantification, ELISA, and luciferase data was performed using SPSS (v29.0, IBM Corp, Armonk, NY, USA). Shapiro–Wilks normality testing was performed to confirm a normal distribution of the data (raw or log-transformed data), and Levene's test was used to confirm equal variance between the groups. For comparisons of two groups, Student's *t*-test (unpaired, two-tailed) was used, and for more than two groups, ANOVA was used with Tukey's post hoc test. A chi-square test was used to measure the association of the SNP with fractures. In all statistical tests, a significance threshold of $p < 0.05$ was set.

3. Results

3.1. Application of a Polygenic Risk Score to Select Thoroughbred Cells for Use

Skin fibroblasts were isolated from 44 Thoroughbred horses of unknown fracture status, DNA was extracted, and cells banked. Genotyping of the samples was performed to calculate their relative polygenic risk score for fracture (using their genotypes across the 50,000 SNPs in the Equine SNP50). We then selected the three samples with the lowest genetic risk and the three samples with the highest genetic risk for use in osteoblast differentiation. These samples were at the ends of the risk spectrum when compared with the scores from the 382 Thoroughbreds representing the general population (Figure 2).

3.2. Skin Fibroblasts Can Differentiate into Osteoblast-like Cells In Vitro

To generate a relevant cell type to study, we differentiated the skin fibroblasts into osteoblast-like cells. Osteoblast differentiation was carried out using two concentrations of β-glycerophosphate, 2 mM, and 10 mM. We demonstrated that after 21 days of culture, the skin fibroblasts expressed a range of different osteoblast-associated genes, and this was not significantly affected by the level of β-glycerophosphate (Figure 3A). Furthermore, both 2 mM and 10 mM β-glycerophosphate result in the production of a mineralised matrix (Figure 3B). No mineralised matrix was produced under normal culture conditions (Supplementary Figure S3).
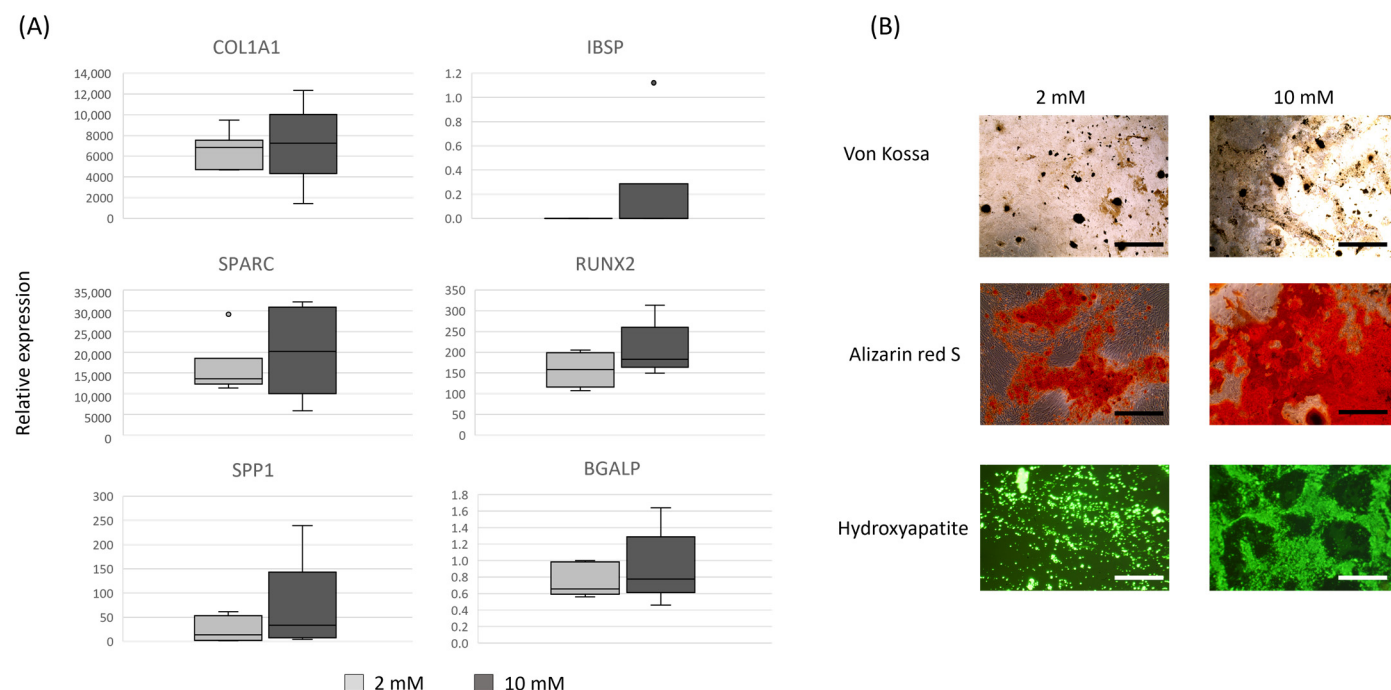
Figure 3. Skin fibroblasts can differentiate into mineralising osteoblasts: (**A**) Box and whisker plots showing osteoblast-associated gene expression in skin fibroblasts from six individual Thoroughbred horses. There are no significant differences in the expression of osteoblast associated genes following 21 days of differentiation in media with either 2 mM (light grey bars) or 10 mM β-glycerophosphate (dark grey bars). Expression is shown relative to the housekeeping gene. (**B**) A mineralised matrix is deposited following osteoblast differentiation in either 2 mM or 10 mM β-glycerophosphate. Positive staining for calcium deposition is shown in black for von Kossa and red for Alizarin Red S. Positive hydroxyapatite staining detected under a fluorescent light is shown in green. Scale bars = 100 μm. Images are representative of six biological replicates using cells at passages 5–6.

3.3. Candidate Gene Expression Analysis Reveals a Significant Difference in STAT1 and COL3A1 Gene Expression in Skin-Derived Osteoblasts from High- and Low-Risk Thoroughbred Horses

Previous work has demonstrated that a region on ECA18 was significantly associated with fracture risk [9]. This region contains 27 protein-coding genes with a known function. Of these, 11 have previously published associations with bone formation or fracture, as described in the introduction. We therefore measured the expression of these 11 candidate genes in skin-derived osteoblasts from the three horses at the highest and three horses at the lowest genetic risk of fracture (as shown in Figure 2). *STAT1* was expressed at significantly greater levels in cells derived from horses at high risk of fracture than in horses at low risk of fracture, and *COL3A1* was expressed at significantly greater levels in cells derived from horses at low risk of fracture (Figure 4). No expression of *ZNF804A* was detected, and no significant differences in the expression of the other genes between high- and low-risk horses were found.

STAT1 can sequester the essential regulator of osteoblast differentiation RUNX2 in the cytoplasm of cells. We therefore examined the cellular localisation of STAT1 and RUNX2 in either the cells derived from high- and low-risk horses, but no significant differences were observed (Figure 5). Although STAT1 is ubiquitously expressed in both pre- and postosteogenic differentiation, we have previously demonstrated the specificity of the STAT1 antibody to the equine protein [57,58].
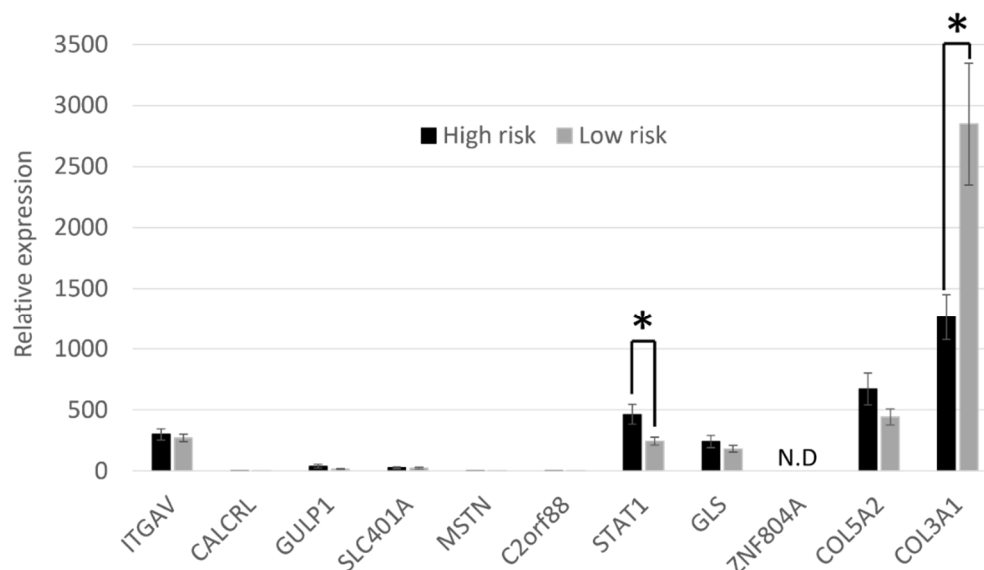
Figure 4. Expression of candidate genes from the fracture-associated region on ECA18 in skin-cell-derived osteoblasts. There is a significant difference in the expression of *STAT1* and *COL3A1* between cells derived from high- and low-risk Thoroughbred horses. Expression is shown relative to the housekeeping gene. Error bars represent the s.e.m of cells derived from three individual horses, each differentiated into osteoblasts in duplicate (once in 2 mM β-glycerophosphate and once in 10 mM β-glycerophosphate). * $p < 0.05$. N.D = expression not detected. Cells were at passages 5–6.

3.4. WGS Revealed a SNP in the Region Upstream of COL3A1 That Is Predicted to Result in the Loss of a SOX11 Binding Site and Is Significantly Associated with Fracture Risk

To identify novel SNPs that may be responsible for the differential expression of *STAT1* and *COL3A1*, we carried out preliminary whole-genome sequencing on two fracture cases and two controls. This did not reveal any exonic SNPs in the associated region on chromosome 18 that were predicted to affect protein structure or function. However, there were a large number of intronic/intergenic variants. One variant was located 3154 bp upstream of the transcriptional start site (TSS) of *COL3A1* in (position ECA18:65598945 T>G EquCab 3.0). Both cases were GG, one control was TT and one control was GT. This region is highly conserved between humans and horses, and ENCODE data demonstrate that the human region has markers of an enhancer/promoter (e.g., multiple transcription factor binding sites and histone H3 lysine 4 methylation, H3K4me). The Functional Annotation of Animal Genomes (FAANG) database also demonstrates the presence of H3K4me and H3K27ac (histone H3 lysine 27 acetylation) in the region covering the SNP in adipose tissue, ovaries, and skin.

The Genomic Evolutionary Rate Profiling (GERP) conservation score [59] of the SNP is −0.28. The minor allele frequency (MAF) was 0.32 in our 91 case horses and 0.22 in our 86 controls (i.e., 0.27 on average). Across 25 different breeds of horse [51], the MAF was 0.13. In many breeds, the MAF was either very low or absent (absent in Arabians, Belgians, Clydesdales, Icelandics, Shetlands, and Welsh ponies), whereas in US/European Thoroughbreds, it was 0.53, similar to Japanese Thoroughbreds (0.5). It was also very high in Quarter horses (0.66) (Supplementary Table S3). The genotype of ten Thoroughbreds with heterozygous or homozygous genotypes was confirmed to be accurate using the Integrative Genome Viewer.
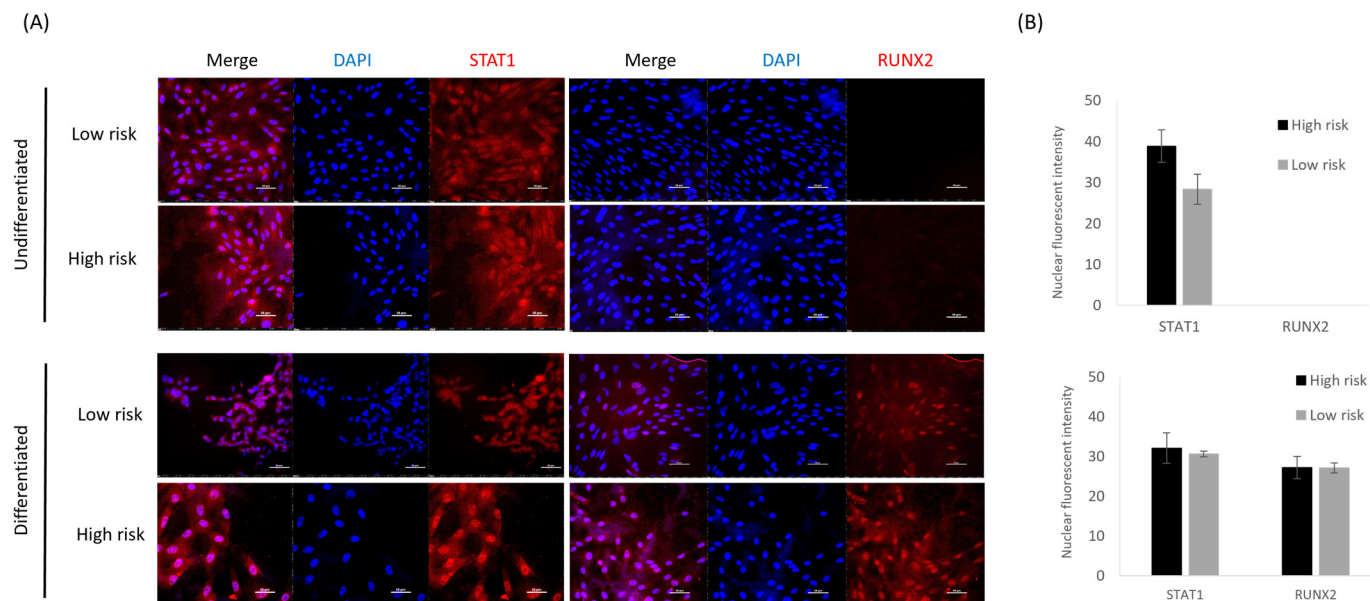
Figure 5. STAT1 and RUNX2 have similar cellular localisations in cells derived from high-risk and low-risk horses: (**A**) Immunocytochemistry demonstrates STAT1 (red) expression in undifferentiated Thoroughbred skin fibroblasts and following osteoblast differentiation, but RUNX2 (red) is only present following differentiation. DAPI staining of the nucleus is shown in blue. Scale bar = 50 μm. Images are representative of replicates using cells derived from three different donors. (**B**) Quantification of the intensity of the staining in the nuclei shows no significant differences between cells derived from high- and low-risk horses. Error bars represent the s.e.m from cells derived from three different donors.

The SNP is predicted to change a transcription factor binding motif from a SOX9/10/11 binding site to a KLF13 binding site (Figure 6A). Allelic discrimination assays were performed (Figure 6B,C) and demonstrated that the G variant is significantly associated with fractures ($p < 0.01$).

3.5. SOX11 Is Expressed in Undifferentiated and Osteoblast-Differentiated Skin Fibroblasts but Is Not Differentially Expressed between High- and Low-Risk Horses

SOX9 and 10 are both members of the SOXE group and have overlapping functions [60]. SOX9 is a master regulator of cartilage differentiation and is usually decreased in osteoblast differentiation [61]. In contrast, SOX11, a member of the SOXC group [62], is required for osteoblast precursor survival, proliferation, and differentiation [63]. We demonstrated that *SOX11* is expressed in skin fibroblasts before and after differentiation into osteoblast-like cells, but there is no significant difference in *SOX11* expression levels between high- and low-risk horses (Figure 7).
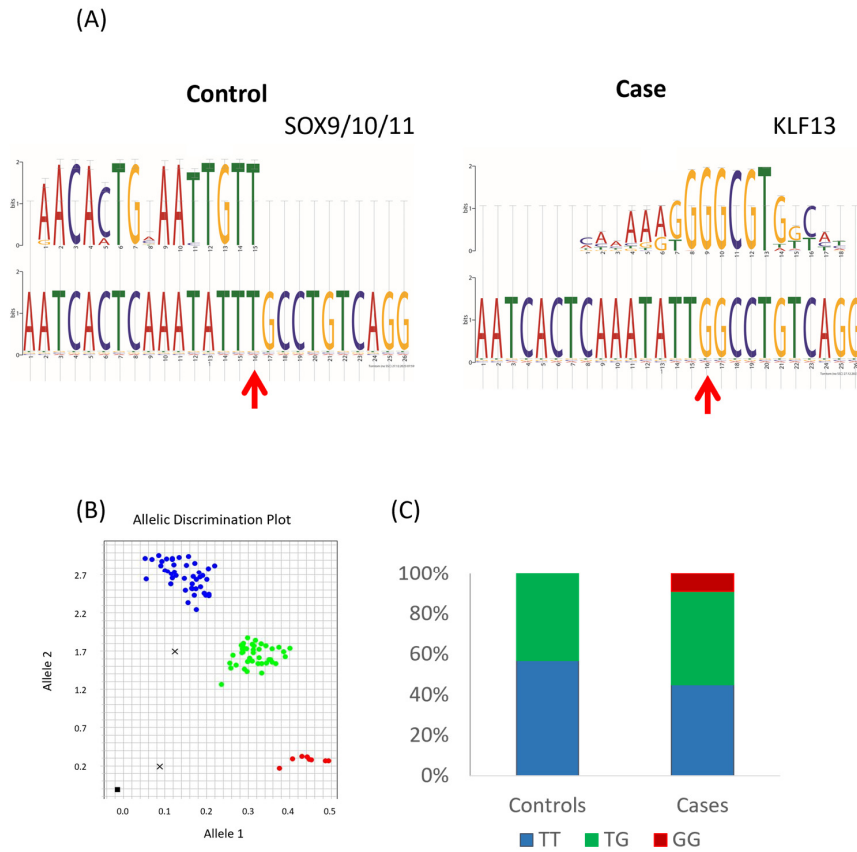
Figure 6. An associated SNP of interest lies upstream of *COL3A1*. (**A**) A SNP identified through preliminary whole-genome sequencing is predicted to cause the loss of a SOX9/10/11 binding site and the gain of a KLF13 binding site. Red arrow indicates the position of the SNP. Analysis output from TomTom. (**B**) Representative allelic discrimination plot of the *COL3A1* upstream SNP. Blue indicates TT homozygotes, green indicates TG heterozygotes, and red indicates GG homozygotes. (**C**) The percentage distribution of the genotypes in 86 Thoroughbred horses without a fracture (controls) and 91 Thoroughbred horses with catastrophic fractures (cases).
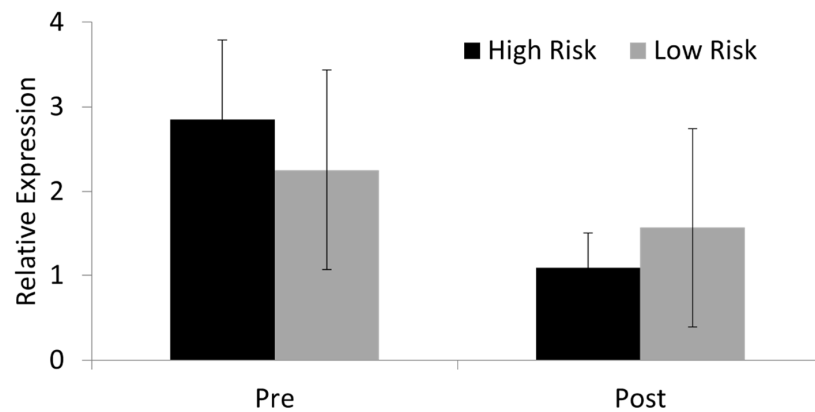

Figure 7. *SOX11* expression. *SOX11* is expressed in Thoroughbred skin fibroblasts pre and post differentiation into osteoblast-like cells. There are no significant differences in expression between cells derived from high- and low-risk horses. Expression is shown relative to the housekeeping gene. Error bars represent the s.e.m of cells derived from three individual horses. Cells were at passages 2–6.

3.6. SOX11 Modulation in Skin Fibroblasts Results in Significant Changes in the Expression of COL3A1

We next determined if SOX11 can regulate *COL3A1* expression by modulating its expression. Skin fibroblasts were stably transfected to overexpress the human *SOX11* gene, as the human and equine SOX11 proteins are 90% identical. In the undifferentiated skin fibroblasts, this resulted in a large, significant increase in *SOX11* expression (Figure 8A(i)) and a significant increase in *COL3A1* expression (Figure 8A(ii)). This suggests that SOX11 regulates *COL3A1*. Following osteoblast differentiation of the skin fibroblasts, there was no significant difference in the expression of *COL3A1* in cells overexpressing *SOX11* compared with the control cells. This is because the relative level of *SOX11* overexpression was reduced after differentiation, which may reflect a lower activity of the CMV promoter (used to drive overexpression) in osteoblasts (Figure 8A). No differences were observed in the total protein level or cellular localisation of COL3A1 in undifferentiated skin fibroblasts following SOX11 overexpression (Supplementary Figure S4). However, the lentiviral transfection of undifferentiated skin fibroblasts to express an shRNA to *SOX11* produced a significant decrease in *COL3A1* (Figure 8B). Taken together, these data demonstrate that SOX11 can regulate *COL3A1* expression.
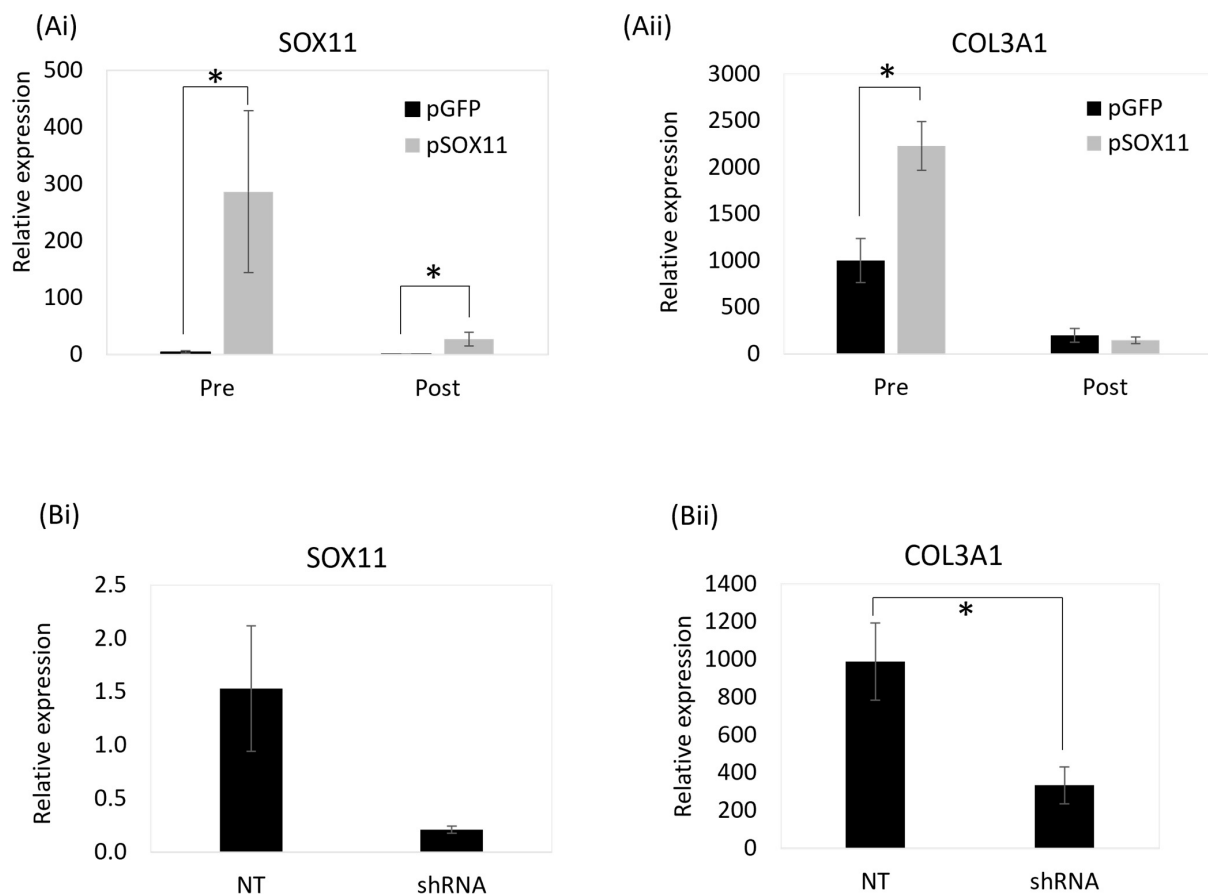

Figure 8. (**A**) SOX11 stable modulation results in significant changes in *COL3A1* expression: (**A(i)**) Stable transfection of a *SOX11* overexpression plasmid (pSOX11) results in a significant increase in *SOX11* expression in equine skin fibroblasts (one Thoroughbred, two ponies), pre and post osteogenic differentiation, compared with control cells (pGFP). (**A(ii)**) This results in a significant increase in *COL3A1* expression in the undifferentiated cells (pre osteogenic culture). (**B**) Lentiviral expression of a shRNA to *SOX11* reduces *SOX11* (**B(i)**) and *COL3A1* (**B(ii)**) expression in undifferentiated skin fibroblasts (three Thoroughbred, one pony). Expression is shown relative to the housekeeping gene. Error bars represent the s.e.m from cells derived from three or four different donors. * $p < 0.05$. Cells between passages 5 and 10 were used in these experiments.

3.7. The Region Containing the SNP Binds to Nuclear Proteins from Equine Cells

To determine if the region of DNA containing the novel SNP could bind to nuclear proteins, EMSA was performed using nuclear extracts from undifferentiated skin fibroblasts overexpressing SOX11. We observed a gel shift indicating protein binding to the probe, whether the probe contained the TT control allele or the fracture-associated GG allele. However, the addition of a SOX11 antibody did not result in a supershift (Figure 9).
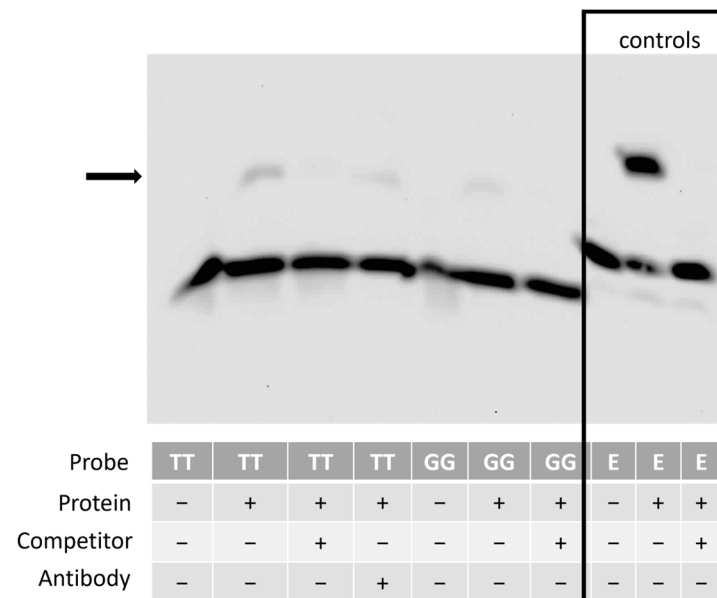

Probe	TT	TT	TT	TT	GG	GG	GG	E	E	E
Protein	−	+	+	+	−	+	+	−	+	+
Competitor	−	−	+	−	−	−	+	−	−	+
Antibody	−	−	−	+	−	−	−	−	−	−

Figure 9. The region upstream of *COL3A1* containing the SNP can bind to a nuclear protein. EMSA revealed that a probe containing either the wild-type (TT) or variant (GG) allele was capable of binding to a nuclear protein as seen by the shift (arrow). The use of an unlabelled competitor probe prevented this shift from occurring. The addition of an antibody to SOX11 did not result in a detectable supershift. Image representative of results observed using nuclear extracts from fibroblasts isolated from two different ponies. The boxed area represents a technical control for the assay, carried out using a biotin-labelled probe for Epstein–Barr Nuclear Antigen (EBNA, labelled E in the figure), an unlabelled probe, and an EBNA protein extract. The image is of the entire full-size blot and is not cropped.

3.8. The Region Containing the SNP Has Promoter Activity, and the SNP Affects Reporter Gene Expression in a Genetic-Background-Dependent Manner

To determine if the region of DNA containing the novel SNP had any promoter activity, we carried out luciferase assays in undifferentiated skin fibroblasts. We demonstrated that the region upstream of the equine *COL3A1* gene, including the SNP, does have promoter activity (Figure 10). However, the effect of the SNP is variable and depends on the genetic background of the individual. In skin fibroblasts derived from some horses, the presence of the SNP caused a reduction in promoter activity (Figure 10A,B). In contrast, in other horses, the presence of the SNP caused an increase in promoter activity (Figure 10C,D) and, in some horses, the presence of the SNP had no significant effect on promoter activity (Figure 10E,F). There was no correlation between the result and the overall risk score of the horse from which the cells were derived. Each assay was carried out on three to six technical replicates performed simultaneously. However, additional replicates were also performed using the same donor horse but on replicates set up at three different times. For the same donor cell line, the results are consistent between these independent experiments (Supplementary Figure S5).
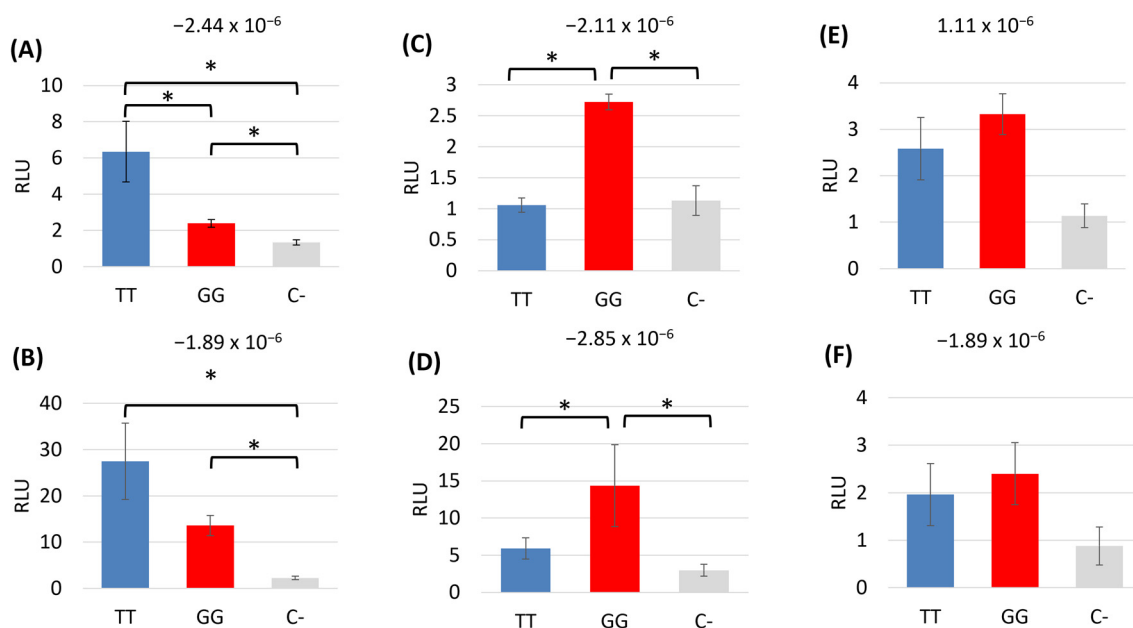
Figure 10. The region upstream of *COL3A1* containing the SNP has promoter activity, and the effect of the SNP varies with the donor cells used. Luciferase assays showing the relative light unit (RLU) for the 3220 bp region upstream of *COL3A1* containing either the reference allele (TT) or the alternative allele (GG) compared with a promoter-less control (C−). Error bars represent the s.e.m of three to six technical replicates. Graphs (**A**–**F**) represent the assay being performed in skin fibroblasts from six different donor Thoroughbred horses with a range of polygenic risk scores. The risk score for each horse is displayed above each graph. * $p < 0.05$.

We were not able to perform the luciferase assays in the skin cells following osteogenic differentiation as the cells cannot be transfected.

4. Discussion

In this study, we employed polygenic risk scores using genome-wide information to facilitate the establishment of an in vitro system to study bone gene regulation in horses at high and low genetic risks of fracture. These samples were not part of the original cohort of cases and controls used to derive the risk score. The validation of the polygenic risk score in a new cohort of cases and controls is currently ongoing.

Skin fibroblasts were selected as the starting cell type due to their accessibility and ease of culture, which allowed us to create a bank of cells from a large number of horses representing the general Thoroughbred population and select cells which were at very high or very low genetic risk. Primary osteoblasts could not be used, as they have very limited proliferative capacity in vitro [64]. Skin fibroblasts have previously been shown to share properties with MSCs in vitro and can be driven to differentiate into mineralising osteoblasts [65,66]. We confirm this finding using horse skin fibroblasts. In future studies, it may be possible to use other cells as the starting population, such as primary MSCs [67] or induced pluripotent stem cells (iPSCs) [68–70]. We selected skin fibroblasts only from male Thoroughbreds to ensure we captured samples at the ends of our risk spectrum and because some previous studies have shown an association of catastrophic fracture with sex [3,6,71], but future studies using larger numbers of female and male cells would be warranted. Our cell-based system was established to allow us to control for environmental variability, but it is well known that other factors can influence cell behaviour in vitro, e.g., donor age. This study only used cells from young Thoroughbred horses between the ages of two and four years old, but we cannot rule out that other biases may have existed in the cells which affected the results.

To differentiate the skin fibroblasts into osteoblasts, we tested two concentrations of β-glycerophosphate. An amount of 10 mM of β-glycerophosphate has been commonly applied to equine-induced pluripotent stem cells (iPSCs) [72], equine MSCs [73–75], and human and mouse skin fibroblasts [65,66]. However, there have been reports that levels of β-glycerophosphate over 2 mM can lead to nonphysiological mineral deposition [76]. In this study, we saw no significant effect of β-glycerophosphate concentration on gene expression or matrix mineralisation, but additional biological replicates may enable the detection of smaller differences in differentiation.

We used the skin-derived osteoblasts to carry out a candidate gene expression analysis of genes that lie within a region on ECA18 that has previously been shown to be associated with catastrophic fracture [9]. All of the candidate genes were expressed in the skin-derived osteoblasts, with the exception of *ZNF804A*, which was previously suggested to be a candidate gene for fracture risk [9] but was not expressed in our experimental system. Of the candidate genes examined, *COL3A1* and *STAT1* were differentially expressed in the high- and low-risk cells. In the future, analyses using RNA sequencing to examine global gene expression profiles would be beneficial. As we had a limited number of samples (three high-risk and three low-risk), future work using larger cohorts may detect other differentially expressed genes. However, we were limited in the samples we had available as we wanted to ensure that we selected cells that were at the ends of the risk spectrum [38].

STAT1$^{-/-}$ mice have a significant increase in bone mineral density, mineral content, and bone growth [30]. This is due to increased osteoblast differentiation [28]. STAT1 acts to sequester the essential regulator of osteoblast differentiation RUNX2 in the cytoplasm of cells. Therefore, in the absence of STAT1, more RUNX2 enters the nucleus to activate osteoblast-associated gene transcription [28]. Furthermore, it has been shown in mice that STAT1 inhibition accelerates fracture healing [29]. RUNX2 expression and activity need to be tightly controlled, as overexpression of *Runx2* leads to mice that have osteopenia and suffer from multiple fractures [77]. Therefore, in our study, the higher level of *STAT1* observed in the high-risk cells may contribute to the misregulation of other bone-specific genes and predisposition to fracture. However, we were unable to demonstrate any differences in the cellular localisation of STAT1 or RUNX2 in the high- or low-risk samples. Furthermore, our WGS was not able to detect any putative, functional SNPs in the vicinity of the *STAT1* gene. This may reflect the small number of WGS samples used. In the future, it would be beneficial to perform WGS on more samples to increase our power to detect SNPs and/or other variants (e.g., structural variants). However, our results may also suggest that other factors are regulating the expression of *STAT1*, such as long-range enhancers. Further work to determine the mechanisms behind the differential expression of *STAT1* and identify any downstream effects is therefore required.

COL3A1$^{-/-}$ mice are not viable. However, COL3A1$^{-/+}$ mice have reduced bone volume, bone density, and trabecular thickness, and COL3A1$^{-/-}$ MSCs have reduced osteoblast differentiation compared with wild-type MSCs [22]. In humans, the loss of the function mutations of *COL3A1* results in Ehlers–Danlos Syndrome, and these patients have a tendency to fracture, low bone mass, and abnormal bone structures [78]. Therefore, in our study, the lower levels of *COL3A1* in the cells derived from horses at high risk of fracture may contribute to their risk. *COL3A1* regulation is not well-defined, but our WGS data revealed an SNP 3154 bp upstream of the *COL3A1* transcriptional start site. As our WGS was performed on only two cases and two controls, we carried out allelic discrimination genotyping in a further 86 controls and 91 cases and found that the SNP was significantly associated with fracture.

The SNP exists in a region which contains histone modifications in multiple equine tissues, suggesting promoter/enhancer function. No histone modifications were found in the FAANG database for bone tissue, but these tissues had lower-quality metrics, and the data were more limited than for other tissues [79]. Although its GERP score does not indicate high levels of evolutionary conservation, there are limitations to using this approach to identify deleterious mutations, particularly in noncoding sites [80]. The MAF

of the SNP varied between breeds, with most breeds showing very low frequencies and Thoroughbreds and Quarter horses showing much higher frequencies. The incidence of catastrophic fracture is high in both of these breeds [81]. Interestingly, the MAF appeared to be lower in our UK Thoroughbreds compared with variant catalogues of international Thoroughbreds, which may reflect some population differences. The higher allele frequencies in some Thoroughbred populations and the small difference in the cases versus the controls in the UK population (0.32 versus 0.22) suggest it is unlikely to be a variant of high effect. However, complex diseases are most likely explained by multiple variants of small-to-moderate effect size, and while minor alleles are often the risk alleles [82], the major allele within a breed or population can sometimes be the risk allele [83,84].

Bioinformatics tools suggested that the SNP may decrease SOX11 binding and promote KLF13 binding. KLF13 is a transcription factor involved in erythropoiesis [85]. It has also been shown to act as a tumour suppressor in colorectal cancer cells [86] and to be a novel regulator of heart development [87]. It is not clear if KLF13 has any role in bone formation or function, and we did not measure KLF13 expression in this study. However, KLF13 has been shown to be upregulated by dexamethasone during osteoblast differentiation [88] and is expressed in developing mouse skeletons [89]. The potential effect of SNPs on KLF13 binding and KLF13 regulation of *COL3A1* therefore warrants future investigation.

Although SOX11 is involved in bone differentiation [63,90], it has not previously been shown to regulate *COL3A1*. In skin fibroblasts, the stable knockdown of *SOX11* through an integrated shRNA resulted in a significant decrease in *COL3A1* expression. Likewise, the overexpression of *SOX11* produced a significant increase in *COL3A1* gene expression. This was not reflected at the protein level; however, protein and RNA levels do not always correlate, particularly in response to perturbation [91]. This may suggest that COL3A1 is also regulated at the level of translation. Following osteogenic differentiation, no significant effect of *SOX11* overexpression was observed, but this likely reflects the lower levels of *SOX11* overexpression following differentiation. Promoter activity has been correlated with cell type, and CMV in particular shows variable expression in different cells [92]. Future work to test the activity of different promoters in equine osteoblast cells is required to overcome this limitation.

The promoter region of the equine *COL3A1* gene has not previously been defined. EMSA using equine nuclear extracts from *SOX11* overexpressing cells demonstrated protein binding to the region of DNA containing the SNP. Protein binding was observed using DNA containing both the control (TT) and alternative (GG) allele. Fainter binding was observed in the presence of the GG allele, but we cannot rule out the possibility that this reflects technical variation in the assay rather than a true reduction in binding. We were unable to detect a supershift using an antibody to SOX11 and are therefore unable to confirm the identity of the binding proteins. This may reflect a genuine lack of SOX11 binding to our region of interest, or it may reflect a lack of detection. SOX11 has previously been reported to contain a conserved domain that inhibits DNA binding in vitro and prevents detectable binding using EMSAs [62,93], although other studies have detected SOX11 binding using in vitro translated proteins [94]. We used nuclear extracts from SOX11 overexpressing cells to ensure that robust levels of SOX11 were present. However, the SOX11 protein is tagged with GFP. This may interfere with SOX11 antibody binding and may account for the lack of supershift. Alternatively, it may reflect the weak binding of SOX11 under the conditions used in this in vitro assay. The GG allele is predicted to result in both the loss of the SOX11 binding site and the gain of a KLF13 binding site. As both proteins are of a similar size, it is not possible to determine if the binding proteins to the TT probe are the same as those binding to the GG probe without further analysis. Chromatin immunoprecipitation is required to determine if SOX11 can bind to the *COL3A1* region within cells.

The SOX11 knockdown and overexpression assays were performed using cells from both Thoroughbreds and Welsh Mountain ponies and produced similar results. The EMSA was performed using cells from Welsh Mountain ponies. Welsh Mountain ponies do not carry the *COL3A1* variant (Supplementary Table S3) and do not commonly have

fractures due to bone overloading and are therefore may be at low genetic risk of fracture, or this may reflect the disciplines they are used in (i.e., not racing). In this study, pony cells were only used in experiments that were independent of the breed of horse, as they do not determine the effect of genetic variants but only demonstrate that SOX11 can regulate *COL3A1* expression and that the region of interest can bind to nuclear proteins. Such experiments are usually performed in either immortalized/cancerous cell lines or in laboratory animal cells to understand human biology [63,90]. This is because gene functions are well-conserved between species and the outcomes are unlikely to be breed-specific.

Luciferase assays in undifferentiated skin cells demonstrated that the proximal region (from −3186 to +34) of *COL3A1* containing the SNP of interest does have promoter activity. We attempted to perform the luciferase assays in an osteoblast cell type, but it was not possible to transfect the skin cells following osteoblast differentiation, possibly due to the presence of the mineralised matrix they produce. To our knowledge, there are no reports of a successful transfection of cells following osteogenic culture and mineralisation. There are also very few reports on the successful culture of primary equine osteoblasts, and we did not have access to suitable tissue samples to derive these cells. Attempts to use primary equine osteoblasts for these experiments were therefore outside the scope of this paper.

Furthermore, the luciferase assays demonstrated that the effect of the SNP was dependent on the genetic background of the horse, with different effects observed using cells from different horses. This reflects the complexity of studying complex genetic conditions, even in an in vitro system where environmental effects are minimised. The results suggest that multiple genetic modifications may interplay to regulate *COL3A1* expression, and further work to identify other pathways that may converge on *COL3A1* is required. In addition, the findings demonstrate that although the SNP was significantly associated with fractures, it is unlikely to be a better predictor of fracture risk than the use of a polygenic risk score that utilizes a genome-wide approach.

The major limitation of this study was the small sample sizes, as described. In addition, the luciferase assays, EMSAs, and SOX11 modulation experiments were only successfully performed in the undifferentiated fibroblasts due to the technical limitations described and the lack of an equine osteoblast cell line. However, this work is the first demonstration that SOX11 is capable of regulating *COL3A1* and demonstrates that a specific region upstream of *COL3A1* has promoter activity in vitro. Furthermore, the fibroblasts used have osteogenic differentiation potential. Bone remodelling and fracture repair involve the activities of both mature osteoblasts and progenitor cells [95], and future work to understand the regulation and role of COL3A1 in both cell types during bone remodelling is warranted.

5. Conclusions

Increasing our understanding of the genetic factors and biological processes involved in fracture risk may lead to novel methods to reduce risk in the future. Here, we established a cell-based system to measure the expression of candidate genes present in a fracture-associated region on ECA18 [9,11]. We demonstrated that one of the genes within this region, *COL3A1*, is differentially expressed in cells from horses at high and low genetic risk of fracture and identified a novel SNP upstream of *COL3A1* that lies in a promoter region. The SNP may affect SOX11 binding, and we demonstrate that SOX11 modulates *COL3A1* expression. As a complex disease, this SNP is likely to play only a small role in fracture risk, but the findings demonstrate the feasibility of using cell models to identify novel DNA variants and biological processes associated with equine diseases.

Figure S4: COL3A1 protein levels do not change in response to SOX11 overexpression; Figure S5: Luciferase assays showing the relative light unit (RLU) for the 3220 bp region upstream of COL3A1 containing either the reference allele (TT) or the alternative allele (GG) compared to a promoterless control (C−).

Institutional Review Board Statement: Equine skin fibroblasts were taken from horses that had been euthanised for reasons unrelated to this study and with the consent of the Animal Health Trust Ethical Review Committee (AHT_02_2012) and Royal Veterinary College Clinical Research Ethical Review Board (URN 2021 2035-2).

Informed Consent Statement: Not applicable.

Data Availability Statement: All data generated and analysed during this study are included in this published article and its supplementary information files. The WGS data have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB60529. (https://www.ebi.ac.uk/ena/browser/view/PRJEB60529 (accessed on 2 May 2023).

Conflicts of Interest: E. Palomino Lago and D.J. Guest are affiliated with The Royal Veterinary College, which holds patent WO 2015/019097 "Predictive Method for Bone Fracture Risk in Horses" in relation to this work. This patent claims a method of predicting fracture risk in horses using one or more genetic variations from within the associated region on ECA18. None of the other authors have any other competing interests to declare.

References

1. Parkin, T.D.; Clegg, P.D.; French, N.P.; Proudman, C.J.; Riggs, C.M.; Singer, E.R.; Webbon, P.M.; Morgan, K.L. Risk of fatal distal limb fractures among Thoroughbreds involved in the five types of racing in the United Kingdom. *Vet. Rec.* **2004**, *154*, 493–497. [CrossRef] [PubMed]
2. McKee, S.L. An update on racing fatalities in the UK. *Equine Vet. Ed.* **1995**, *7*, 202–204. [CrossRef]
3. Georgopoulos, S.P.; Parkin, T.D. Risk factors for equine fractures in Thoroughbred flat racing in North America. *Prev. Vet. Med.* **2017**, *139*, 99–104. [CrossRef] [PubMed]
4. Parkin, T.D.; Clegg, P.D.; French, N.P.; Proudman, C.J.; Riggs, C.M.; Singer, E.R.; Webbon, P.M.; Morgan, K.L. Race- and course-level risk factors for fatal distal limb fracture in racing Thoroughbreds. *Equine Vet. J.* **2004**, *36*, 521–526. [CrossRef] [PubMed]
5. Verheyen, K.; Price, J.; Lanyon, L.; Wood, J.L.N. Exercise distance and speed affect the risk of fracture in racehorses. *Bone* **2006**, *39*, 1322–1330. [CrossRef] [PubMed]
6. Anthenill, L.A.; Stover, S.M.; Gardner, I.A.; Hill, A.E. Risk factors for proximal sesamoid bone fractures associated with exercise history and horseshoe characteristics in Thoroughbred racehorses. *AJVR* **2007**, *68*, 760–771. [CrossRef] [PubMed]
7. Kristoffersen, M.; Parkin, T.D.; Singer, E.R. Catastrophic biaxial proximal sesamoid bone fractures in UK Thoroughbred races (1994–2004): Horse characteristics and racing history. *Equine Vet. J.* **2010**, *45*, 420–424. [CrossRef]
8. Welsh, C.E.; Lewis, T.W.; Blott, S.C.; Mellor, D.J.; Stirk, A.J.; Parkin, T.D. Estimates of genetic parameters of distal limb fracture and superficial digital flexor tendon injury in UK Thoroughbred racehorses. *Vet. J.* **2014**, *200*, 253–256. [CrossRef]
9. Blott, S.C.; Swinburne, J.E.; Sibbons, C.; Fox-Clipsham, L.Y.; Helwegen, M.; Hillyer, L.; Parkin, T.D.; Newton, J.R.; Vaudin, M. A genome-wide association study demonstrates significant genetic variation for fracture risk in Thoroughbred racehorses. *BMC Genom.* **2014**, *15*, 147. [CrossRef]
10. Blott, S.C.; Swinburne, J.E. Predictive Method for Bone Fracture Risk in Horses and Humans. 2015. Available online: https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2015019097 (accessed on 4 May 2023).
11. Tozaki, T.; Kusano, K.; Ishikawa, Y.; Kushiro, A.; Nomura, M.; Kikuchi, M.; Kakoi, H.; Hirota, K.; Miyake, T.; Hill, E.W.; et al. A candidate-SNP retrospective cohort study for fracture risk in Japanese Thoroughbred racehorses. *Anim. Genet.* **2020**, *51*, 43–50. [CrossRef]

12. Bulathsinhala, L.; Hughes, J.M.; McKinnon, C.J.; Kardouni, J.R.; Guerriere, K.I.; Popp, K.L.; Matheny, R.W., Jr.; Bouxsein, M.L. Risk of Stress Fracture Varies by Race/Ethnic Origin in a Cohort Study of 1.3 Million US Army Soldiers. *J. Bone Miner. Res.* **2017**, *32*, 1546–1553. [CrossRef] [PubMed]

13. Korvala, J.; Hartikka, H.; Pihlajamaki, H.; Solovieva, S.; Ruohola, J.P.; Sahi, T.; Barral, S.; Ott, J.; Ala-Kokko, L.; Mannikko, M. Genetic predisposition for femoral neck stress fractures in military conscripts. *BMC Genet.* **2010**, *11*, 95. [CrossRef] [PubMed]

14. Zhao, L.; Chang, Q.; Huang, T.; Huang, C. Prospective cohort study of the risk factors for stress fractures in Chinese male infantry recruits. *J. Int. Med. Res.* **2016**, *44*, 787–795. [CrossRef] [PubMed]

15. Varley, I.; Greeves, J.P.; Sale, C.; Friedman, E.; Moran, D.S.; Yanovich, R.; Wilson, P.J.; Gartland, A.; Hughes, D.C.; Stellingwerff, T.; et al. Functional polymorphisms in the P2X7 receptor gene are associated with stress fracture injury. *Purinergic Signal* **2016**, *12*, 103–113. [CrossRef] [PubMed]

16. Warden, S.J.; Burr, D.B.; Brukner, P.D. Stress fractures: Pathophysiology, epidemiology, and risk factors. *Curr. Osteoporos. Rep.* **2006**, *4*, 103–109. [CrossRef] [PubMed]

17. Warden, S.J.; Creaby, M.W.; Bryant, A.L.; Crossley, K.M. Stress fracture risk factors in female football players and their clinical implications. *Br. J. Sports Med.* **2007**, *41* (Suppl. 1), i38–i43. [CrossRef] [PubMed]

18. Long, P.; Samnakay, P.; Jenner, P.; Rose, S. A yeast two-hybrid screen reveals that osteopontin associates with MAP1A and MAP1B in addition to other proteins linked to microtubule stability, apoptosis and protein degradation in the human brain. *Eur. J. Neurosci.* **2012**, *36*, 2733–2742. [CrossRef]

19. Logan, N.J.; Camman, M.; Williams, G.; Higgins, C.A. Demethylation of ITGAV accelerates osteogenic differentiation in a blast-induced heterotopic ossification in vitro cell culture model. *Bone* **2018**, *117*, 149–160. [CrossRef]

20. Wang, L.; Shi, X.; Zhao, R.; Halloran, B.P.; Clark, D.J.; Jacobs, C.R.; Kingery, W.S. Calcitonin-gene-related peptide stimulates stromal cell osteogenic differentiation and inhibits RANKL induced NF-kappaB activation, osteoclastogenesis and bone resorption. *Bone* **2010**, *46*, 1369–1379. [CrossRef]

21. Park, G.; Park, S.-Y.; Lee, E.-H.; Lee, Y.-J.; Kim, S.-Y.; Kim, S.-H.; Kim, Y.-H.; Kim, I.-S.; Kim, J.-E. High trabecular bone mass induced by reduced function of osteoclasts in GULP1-deficient mice. In Proceedings of the European Calcified Tissue Society Conference ECTS, Rome, Italy, 14–17 May 2016.

22. Volk, S.W.; Shah, S.R.; Cohen, A.J.; Wang, Y.; Brisson, B.K.; Vogel, L.K.; Hankenson, K.D.; Adams, S.L. Type III collagen regulates osteoblastogenesis and the quantity of trabecular bone. *Calcif. Tissue Int.* **2014**, *94*, 621–631. [CrossRef]

23. Hong, D.; Chen, H.X.; Yu, H.Q.; Liang, Y.; Wang, C.; Lian, Q.Q.; Deng, H.T.; Ge, R.S. Morphological and proteomic analysis of early stage of osteoblast differentiation in osteoblastic progenitor cells. *Exp. Cell Res.* **2010**, *316*, 2291–2300. [CrossRef]

24. Pereira, M.; Ko, J.-H.; Logan, J.; Protheroe, H.; Kim, K.-B.; Tan, A.L.M.; Croucher, P.I.; Park, K.-S.; Rotival, M.; Petretto, E.; et al. A trans-eQTL network regulates osteoclast multinucleation and bone mass. *eLife* **2020**, *9*, e55549. [CrossRef]

25. Wu, L.-F.; Zhu, D.-C.; Wang, B.-H.; Lu, Y.-H.; He, P.; Zhang, Y.-H.; Gao, H.-Q.; Zhu, X.-W.; Xia, W.; Zhu, H.; et al. Relative abundance of mature myostatin rather than total myostatin is negatively associated with bone mineral density in Chinese. *J. Cell. Mol. Med.* **2018**, *22*, 1329–1336. [CrossRef]

26. Bialek, P.; Parkington, J.; Li, X.; Gavin, D.; Wallace, C.; Zhang, J.; Root, A.; Yan, G.; Warner, L.; Seeherman, H.J.; et al. A myostatin and activin decoy receptor enhances bone formation in mice. *Bone* **2014**, *60*, 162–171. [CrossRef]

27. Burger, M.G.; Steinitz, A.; Geurts, J.; Pippenger, B.E.; Schaefer, D.J.; Martin, I.; Barbero, A.; Pelttari, K. Ascorbic Acid Attenuates Senescence of Human Osteoarthritic Osteoblasts. *Int. J. Mol. Sci.* **2017**, *18*, 2517. [CrossRef]

28. Kim, S.; Koga, T.; Isobe, M.; Kern, B.E.; Yokochi, T.; Chin, Y.E.; Karsenty, G.; Taniguchi, T.; Takayanagi, H. Stat1 functions as a cytoplasmic attenuator of Runx2 in the transcriptional program of osteoblast differentiation. *Genes Dev.* **2003**, *17*, 1979–1991. [CrossRef]

29. Tajima, K.; Takaishi, H.; Takito, J.; Tohmonda, T.; Yoda, M.; Ota, N.; Kosaki, N.; Matsumoto, M.; Ikegami, H.; Nakamura, T.; et al. Inhibition of STAT1 accelerates bone fracture healing. *J. Orthop. Res.* **2010**, *28*, 937–941. [CrossRef]

30. Xiao, L.; Naganawa, T.; Obugunde, E.; Gronowicz, G.; Ornitz, D.M.; Coffin, J.D.; Hurley, M.M. Stat1 controls postnatal bone formation by regulating fibroblast growth factor signaling in osteoblasts. *J. Biol. Chem.* **2004**, *279*, 27743–27752. [CrossRef]

31. Yu, Y.; Newman, H.; Shen, L.; Sharma, D.; Hu, G.; Mirando, A.J.; Zhang, H.; Knudsen, E.; Zhang, G.-F.; Hilton, M.J.; et al. Glutamine Metabolism Regulates Proliferation and Lineage Allocation in Skeletal Stem Cells. *Cell Metab.* **2019**, *29*, 966–978.e964. [CrossRef]

32. Groza, T.; Gomez, F.L.; Mashhadi, H.H.; Muñoz-Fuentes, V.; Gunes, O.; Wilson, R.; Cacheiro, P.; Frost, A.; Keskivali-Bond, P.; Vardal, B.; et al. The International Mouse Phenotyping Consortium: Comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res.* **2022**, *51*, D1038–D1045. [CrossRef]

33. Flint, J.; Mackay, T.F. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.* **2009**, *19*, 723–733. [CrossRef]

34. Li, Z.-k.; Zhou, Q. Cellular models for disease exploring and drug screening. *Protein Cell* **2010**, *1*, 355–362. [CrossRef]

35. Park, I.-H.; Arora, N.; Huo, H.; Maherali, N.; Ahfeldt, T.; Shimamura, A.; Lensch, M.W.; Cowan, C.; Hochedlinger, K.; Daley, G.Q. Disease-specific induced pluripotent stem (iPS) cells. *Cell* **2008**, *134*, 877–886. [CrossRef]

36. Barral, S.; Kurian, M.A. Utility of Induced Pluripotent Stem Cells for the Study and Treatment of Genetic Diseases: Focus on Childhood Neurological Disorders. *Front. Mol. Neurosci.* **2016**, *9*, 78. [CrossRef]

37. Chamberlain, S.J. Disease modelling using human iPSCs. *Hum. Mol. Genet.* **2016**, *25*, R173–R181. [CrossRef]

38. Dobrindt, K.; Zhang, H.; Das, D.; Abdollahi, S.; Prorok, T.; Ghosh, S.; Weintraub, S.; Genovese, G.; Powell, S.K.; Lund, A.; et al. Publicly Available hiPSC Lines with Extreme Polygenic Risk Scores for Modeling Schizophrenia. *Complex. Psychiatry* **2021**, *6*, 68–82. [CrossRef]

39. Coleman, J.R.I. Feasibility and application of polygenic score analysis to the morphology of human-induced pluripotent stem cells. *Mol. Genet. Genom.* **2022**, *297*, 1111–1122. [CrossRef]

40. Choi, S.W.; Mak, T.S.-H.; O'Reilly, P.F. Tutorial: A guide to performing polygenic risk score analyses. *Nat. Protoc.* **2020**, *15*, 2759–2772. [CrossRef]

41. Chung, W. Statistical models and computational tools for predicting complex traits and diseases. *Genom. Inform.* **2021**, *19*, e36. [CrossRef]

42. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [CrossRef]

43. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]

44. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real time quantitative PCR and the $2^{-deltadeltaCT}$ Method. *Methods* **2001**, *25*, 402–408. [CrossRef]

45. Xie, F.; Xiao, P.; Chen, D.; Xu, L.; Zhang, B. miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. *Plant Mol. Biol.* **2012**, *80*, 75–84. [CrossRef]

46. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675. [CrossRef]

47. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytsky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [CrossRef]

48. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.11–11.10.33. [CrossRef]

49. McLaren, W.; Gil, L.; Hunt, S.E.; Riat, H.S.; Ritchie, G.R.S.; Thormann, A.; Flicek, P.; Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **2016**, *17*, 122. [CrossRef]

50. Gupta, S.; Stamatoyannopoulos, J.A.; Bailey, T.L.; Noble, W.S. Quantifying similarity between motifs. *Genome Biol.* **2007**, *8*, R24. [CrossRef]

51. Jagannathan, V.; Gerber, V.; Rieder, S.; Tetens, J.; Thaller, G.; Drögemüller, C.; Leeb, T. Comprehensive characterization of horse genome variation by whole-genome sequencing of 88 horses. *Anim. Genet.* **2019**, *50*, 74–77. [CrossRef]

52. Tozaki, T.; Ohnuma, A.; Kikuchi, M.; Ishige, T.; Kakoi, H.; Hirota, K.-I.; Kusano, K.; Nagata, S.-I. Rare and common variant discovery by whole-genome sequencing of 101 Thoroughbred racehorses. *Sci. Rep.* **2021**, *11*, 16057. [CrossRef]

53. Durward-Akhurst, S.A.; Schaefer, R.J.; Grantham, B.; Carey, W.K.; Mickelson, J.R.; McCue, M.E. Genetic Variation and the Distribution of Variant Types in the Horse. *Front. Genet.* **2021**, *12*, 758366. [CrossRef]

54. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef]

55. Luo, Y.; Hitz, B.C.; Gabdank, I.; Hilton, J.A.; Kagda, M.S.; Lam, B.; Myers, Z.; Sud, P.; Jou, J.; Lin, K.; et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **2020**, *48*, D882–D889. [CrossRef]

56. Liew, C.-G.; Draper, J.S.; Walsh, J.; Moore, H.; Andrews, P.W. Transient and Stable Transgene Expression in Human Embryonic Stem Cells. *Stem Cells* **2007**, *25*, 1521–1528. [CrossRef]

57. Palomino Lago, E.; Jelbert, E.R.; Baird, A.; Lam, P.Y.; Guest, D.J. Equine induced pluripotent stem cells are responsive to inflammatory cytokines before and after differentiation into musculoskeletal cell types. *In Vitr. Cell. Dev. Biol. Anim.* **2023**, *59*, 514–527. [CrossRef]

58. Smith, E.J.; Beaumont, R.E.; McClellan, A.; Sze, C.; Palomino Lago, E.; Hazelgrove, L.; Dudhia, J.; Smith, R.K.W.; Guest, D.J. Tumour necrosis factor alpha, interleukin 1 beta and interferon gamma have detrimental effects on equine tenocytes that cannot be rescued by IL-1RA or mesenchymal stromal cell-derived factors. *Cell Tissue Res.* **2023**, *391*, 523–544. [CrossRef]

59. Cooper, G.M.; Stone, E.A.; Asimenos, G.; Green, E.D.; Batzoglou, S.; Sidow, A. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **2005**, *15*, 901–913. [CrossRef]

60. Haseeb, A.; Lefebvre, V. The SOXE transcription factors-SOX8, SOX9 and SOX10-share a bi-partite transactivation mechanism. *Nucleic Acids Res.* **2019**, *47*, 6917–6931. [CrossRef]

61. Haseeb, A.; Kc, R.; Angelozzi, M.; de Charleroy, C.; Rux, D.; Tower, R.J.; Yao, L.; Pellegrino da Silva, R.; Pacifici, M.; Qin, L.; et al. SOX9 keeps growth plates and articular cartilage healthy by inhibiting chondrocyte dedifferentiation/osteoblastic redifferentiation. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2019152118. [CrossRef]

62. Dy, P.; Penzo-Méndez, A.; Wang, H.; Pedraza, C.E.; Macklin, W.B.; Lefebvre, V. The three SoxC proteins--Sox4, Sox11 and Sox12--exhibit overlapping expression patterns and molecular properties. *Nucleic Acids Res.* **2008**, *36*, 3101–3117. [CrossRef]

63. Gadi, J.; Jung, S.H.; Lee, M.J.; Jami, A.; Ruthala, K.; Kim, K.M.; Cho, N.H.; Jung, H.S.; Kim, C.H.; Lim, S.K. The transcription factor protein Sox11 enhances early osteoblast differentiation by facilitating proliferation and the survival of mesenchymal and osteoblast progenitors. *J. Biol. Chem.* **2013**, *288*, 25400–25413. [CrossRef] [PubMed]

64. Kassem, M.; Ankersen, L.; Eriksen, E.F.; Clark, B.F.C.; Rattan, S.I.S. Demonstration of cellular aging and senescence in serially passaged long-term cultures of human trabecular osteoblasts. *Osteoporos. Int.* **1997**, *7*, 514–524. [CrossRef]

65. Lorenz, K.; Sicker, M.; Schmelzer, E.; Rupf, T.; Salvetter, J.; Schulz-Siegmund, M.; Bader, A. Multilineage differentiation potential of human dermal skin-derived fibroblasts. *Exp. Dermatol.* **2008**, *17*, 925–932. [CrossRef]

66. Halcsik, E.; Forni, M.F.; Fujita, A.; Verano-Braga, T.; Jensen, O.N.; Sogayar, M.C. New insights in osteogenic differentiation revealed by mass spectrometric assessment of phosphorylated substrates in murine skin mesenchymal cells. *BMC Cell Biol.* **2013**, *14*, 47. [CrossRef]

67. Glynn, E.R.; Londono, A.S.; Zinn, S.A.; Hoagland, T.A.; Govoni, K.E. Culture conditions for equine bone marrow mesenchymal stem cells and expression of key transcription factors during their differentiation into osteoblasts. *J. Anim. Sci. Biotechnol.* **2013**, *4*, 40. [CrossRef]

68. Nagy, K.; Sung, H.-K.; Zhang, P.; Laflamme, S.; Vincent, P.; Agha-Mohammadi, S.; Woltjen, K.; Monetti, C.; Michael, I.P.; Smith, L.C.; et al. Induced pluripotent stem cell lines derived from equine fibroblasts. *Stem Cell Rev. Rep.* **2011**, *7*, 693–702. [CrossRef]

69. Bavin, E.P.; Smith, O.; Baird, A.E.; Smith, L.C.; Guest, D.J. Equine Induced Pluripotent Stem Cells have a Reduced Tendon Differentiation Capacity Compared to Embryonic Stem Cells. *Front. Vet. Sci.* **2015**, *2*, 55. [CrossRef]

70. Sharma, R.; Livesey, M.R.; Wyllie, D.J.; Proudfoot, C.; Whitelaw, B.; Hay, D.C.; Donadeu, X. Generation of functional neurons from feeder-free, keratinocyte-derived equine induced pluripotent stem cells. *Stem Cells Dev.* **2014**, *23*, 1524–1534. [CrossRef]

71. Sun, T.C.; Riggs, C.M.; Cogger, N.; Wright, J.; Al-Alawneh, J.I. Noncatastrophic and catastrophic fractures in racing Thoroughbreds at the Hong Kong Jockey Club. *Equine Vet. J.* **2019**, *51*, 77–82. [CrossRef]

72. Baird, A.; Lindsay, T.; Everett, A.; Iyemere, V.; Paterson, Y.Z.; McClellan, A.; Henson, F.M.D.; Guest, D.J. Osteoblast differentiation of equine induced pluripotent stem cells. *Biol. Open* **2018**, *7*, bio033514. [CrossRef]

73. Guest, D.J.; Ousey, J.C.; Smith, M.R.W. Defining the expression of marker genes in equine mesenchymal stromal cells. *Stem Cells Cloning Adv. Appl.* **2008**, *1*, 1–9. [CrossRef] [PubMed]

74. Burk, J.; Ribitsch, I.; Gittel, C.; Juelke, H.; Kasper, C.; Staszyk, C.; Brehm, W. Growth and differentiation characteristics of equine mesenchymal stromal cells derived from different sources. *Vet. J.* **2013**, *195*, 98–106. [CrossRef] [PubMed]

75. Radcliffe, C.H.; Flaminio, J.B.F.; Fortier, L.A. Temporal analysis of equine bone marrow aspirate during establishment of putative mesenchymal progenitor cell populations. *Stem Cells Dev.* **2010**, *19*, 269–282. [CrossRef] [PubMed]

76. Chung, C.-H.; Golub, E.E.; Forbes, E.; Tokuoka, T.; Shapiro, I.M. Mechanism of action of β-glycerophosphate on bone cell mineralization. *Calcif. Tissue Int.* **1992**, *51*, 305–311. [CrossRef] [PubMed]

77. Liu, W.; Toyosawa, S.; Furuichi, T.; Kanatani, N.; Yoshida, C.; Liu, Y.; Himeno, M.; Narai, S.; Yamaguchi, A.; Komori, T. Overexpression of Cbfa1 in osteoblasts inhibits osteoblast maturation and causes osteopenia with multiple fractures. *J. Cell Biol.* **2001**, *155*, 157–166. [CrossRef] [PubMed]

78. Yen, J.L.; Lin, S.P.; Chen, M.R.; Niu, D.M. Clinical features of Ehlers-Danlos syndrome. *J. Formos. Med. Assoc. Taiwan Yi Zhi* **2006**, *105*, 475–480. [CrossRef] [PubMed]

79. Kingsley, N.B.; Hamilton, N.A.; Lindgren, G.; Orlando, L.; Bailey, E.; Brooks, S.; McCue, M.; Kalbfleisch, T.S.; MacLeod, J.N.; Petersen, J.L.; et al. "Adopt-a-Tissue" Initiative Advances Efforts to Identify Tissue-Specific Histone Marks in the Mare. *Front. Genet.* **2021**, *12*, 649959. [CrossRef]

80. Huber, C.D.; Kim, B.Y.; Lohmueller, K.E. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PLoS Genet.* **2020**, *16*, e1008827. [CrossRef]

81. Beisser, A.L.; McClure, S.; Wang, C.; Soring, K.; Garrison, R.; Peckham, B. Evaluation of catastrophic musculoskeletal injuries in Thoroughbreds and Quarter Horses at three Midwestern racetracks. *J. Am. Vet. Med. Assoc.* **2011**, *239*, 1236–1241. [CrossRef]

82. Kido, T.; Sikora-Wohlfeld, W.; Kawashima, M.; Kikuchi, S.; Kamatani, N.; Patwardhan, A.; Chen, R.; Sirota, M.; Kodama, K.; Hadley, D.; et al. Are minor alleles more likely to be risk alleles? *BMC Med. Genom.* **2018**, *11*, 3. [CrossRef]

83. Pritchard, J.K.; Cox, N.J. The allelic architecture of human disease genes: Common disease-common variant...or not? *Hum. Mol. Genet.* **2002**, *11*, 2417–2423. [CrossRef] [PubMed]

84. Pritchard, J.K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet.* **2001**, *69*, 124–137. [CrossRef] [PubMed]

85. Gordon, A.R.; Outram, S.V.; Keramatipour, M.; Goddard, C.A.; Colledge, W.H.; Metcalfe, J.C.; Hager-Theodorides, A.L.; Crompton, T.; Kemp, P.R. Splenomegaly and modified erythropoiesis in KLF13-/-mice. *J. Biol. Chem.* **2008**, *283*, 11897–11904. [CrossRef] [PubMed]

86. Yao, W.; Jiao, Y.; Zhou, Y.; Luo, X. KLF13 suppresses the proliferation and growth of colorectal cancer cells through transcriptionally inhibiting HMGCS1-mediated cholesterol biosynthesis. *Cell Biosci.* **2020**, *10*, 76. [CrossRef]

87. Lavallée, G.; Andelfinger, G.; Nadeau, M.; Lefebvre, C.; Nemer, G.; Horb, M.E.; Nemer, M. The Kruppel-like transcription factor KLF13 is a novel regulator of heart development. *EMBO J.* **2006**, *25*, 5201–5213. [CrossRef] [PubMed]

88. Leclerc, N.; Luppen, C.A.; Ho, V.V.; Nagpal, S.; Hacia, J.G.; Smith, E.; Frenkel, B. Gene expression profiling of glucocorticoid-inhibited osteoblasts. *J. Mol. Endocrinol.* **2004**, *33*, 175–193. [CrossRef] [PubMed]

89. Martin, K.M.; Metcalfe, J.C.; Kemp, P.R. Expression of Klf9 and Klf13 in mouse development. *Mech. Dev.* **2001**, *103*, 149–151. [CrossRef]

90. Xu, L.; Huang, S.; Hou, Y.; Liu, Y.; Ni, M.; Meng, F.; Wang, K.; Rui, Y.; Jiang, X.; Li, G. Sox11-modified mesenchymal stem cells (MSCs) accelerate bone fracture healing: Sox11 regulates differentiation and migration of MSCs. *FASEB J.* **2015**, *29*, 1143–1152. [CrossRef]

91. Vogel, C.; Marcotte, E.M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **2012**, *13*, 227–232. [CrossRef]

92. Qin, J.Y.; Zhang, L.; Clift, K.L.; Hulur, I.; Xiang, A.P.; Ren, B.-Z.; Lahn, B.T. Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLoS ONE* **2010**, *5*, e10611. [CrossRef]

93. Wiebe, M.S.; Nowling, T.K.; Rizzino, A. Identification of novel domains within Sox-2 and Sox-11 involved in autoinhibition of DNA binding and partnership specificity. *J. Biol. Chem.* **2003**, *278*, 17901–17911. [CrossRef] [PubMed]

94. Kan, A.; Ikeda, T.; Fukai, A.; Nakagawa, T.; Nakamura, K.; Chung, U.-I.; Kawaguchi, H.; Tabin, C.J. SOX11 contributes to the regulation of GDF5 in joint maintenance. *BMC Dev. Biol.* **2013**, *13*, 4. [CrossRef] [PubMed]

95. Bragdon, B.C.; Bahney, C.S. Origin of Reparative Stem Cells in Fracture Healing. *Curr. Osteoporos. Rep.* **2018**, *16*, 490–503. [CrossRef]