

Identification of a global gene expression signature associated with the genetic risk of catastrophic fracture in iPSC-derived osteoblasts from Thoroughbred horses

Esther Palomino Lago¹ | Amy K. C. Ross¹ | Alyce McClellan^{2,3} | Deborah J. Guest¹ 

¹Department of Clinical Sciences and Services, Centre for Vaccinology and Regenerative Medicine, The Royal Veterinary College, Hatfield, Herts, UK

²Animal Health Trust, Newmarket, UK

³Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

Correspondence

Deborah J. Guest, Department of Clinical Sciences and Services, Centre for Vaccinology and Regenerative Medicine, The Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Herts AL9 7TA, UK.
Email: djguest@rvc.ac.uk

Funding information

Horserace Betting Levy Board, Grant/Award Number: vet/prj/792; Alborada Trust

Abstract

Bone fractures are a significant problem in Thoroughbred racehorses. The risk of fracture is influenced by both genetic and environmental factors. To determine the biological processes that are affected in genetically susceptible horses, we utilised polygenic risk scoring to establish induced pluripotent stem cells (iPSCs) from horses at high and low genetic risk. RNA-sequencing on iPSC-derived osteoblasts revealed 112 genes that were significantly differentially expressed. Forty-three of these genes have known roles in bone, 27 are not yet annotated in the equine genome and 42 currently have no described role in bone. However, many of the proteins encoded by the known and unknown genes have reported interactions. Functional enrichment analyses revealed that the differentially expressed genes were overrepresented in processes regulating the extracellular matrix and pathways known to be involved in bone remodelling and bone diseases. Gene set enrichment analysis also detected numerous biological processes and pathways involved in glycolysis with the associated genes having a higher expression in the iPSC-osteoblasts from horses with low polygenic risk scores for fracture. Therefore, the differentially expressed genes may be relevant for maintaining bone homeostasis and contribute to fracture risk. A deeper understanding of the consequences of mis-regulation of these genes and the identification of the DNA variants which underpin their differential expression may reveal more about the molecular mechanisms which are involved in equine bone health and fracture risk.

KEY WORDS

bone, fracture, horse, induced pluripotent stem cells, osteoblasts, RNA sequencing, Thoroughbred

INTRODUCTION

Bone fractures in horses can occur owing to a single point traumatic event (such as a kick or fall), but in Thoroughbred racehorses, they often occur in response to high-intensity, repetitive loading that over time leads to structural failure (Stover, 2013). Fractures owing to bone overloading are the most common musculoskeletal

injury in training and racing accounting for 41% of all musculoskeletal injuries (Johnston, Riggs, et al., 2020). Non-fatal stress fractures lead to a significant loss of time in training, earnings and race starts (Johnston, Sidhu, et al., 2020). However, more complex fractures in horses can be difficult to treat because of the need for continuous weight-bearing in the injured limb during recovery. As such, fracture is the main reason for euthanasia on

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Animal Genetics* published by John Wiley & Sons Ltd on behalf of Stichting International Foundation for Animal Genetics.

the racecourse (McKee, 1995), accounting for approximately 75% of all UK racecourse fatalities (Rosanowski et al., 2018) and an average of 60 horses/year suffering a fatal, catastrophic distal limb fracture during racing in the UK (Parkin et al., 2004a). Fractures therefore have a large economical and welfare impact.

Fracture risk is influenced by both environmental risk factors (Anthenill et al., 2007; Georgopoulos & Parkin, 2017; Kristoffersen et al., 2010; Parkin et al., 2004b; Verheyen et al., 2006) and genetic risk factors. The heritability of fracture risk in Thoroughbreds is 0.21–0.37 (Welsh et al., 2014). A genome wide association study identified significant genetic variation associated with catastrophic fracture risk on chromosomes 9, 18, 22 and 31 and four significantly associated SNPs: three on ECA18 (horse chromosome 18) and one on ECA1 (Blott et al., 2014). Additional SNPs within the region on ECA18 were also found to be associated with fracture (Tozaki et al., 2020). However, the molecular processes affected in horses at high genetic risk and the interplay between genetic and environmental risk factors remain poorly understood. Studies on the pathology of fracture indicate stress-related damage to the bone prior to catastrophic injury (Stover, 2003). This may be related to abnormal bone repair, as horses euthanised following a fracture show changes in other bones including stress fractures and excessive bone remodelling (Anthenill et al., 2010; Parkin et al., 2006; Riggs et al., 1999; Stover et al., 1992). This suggests that the bone cells involved in tissue homeostasis may be affected in horses that suffer from fatal fractures. However, it has proven challenging to understand the genes and biological pathways that may be involved in fracture risk and studies using bone tissues are complicated by the confounding environmental risk factors.

We have recently developed a genome-wide polygenic risk score (PRS) for catastrophic fracture (Palomino Lago et al., 2024). Polygenic risk scores provide an estimate of the genetic risk of the phenotypic trait at the level of the individual (Choi et al., 2020) and their potential clinical utility is becoming increasingly evident (Lewis & Vassos, 2020). We previously utilised the PRS for fracture to establish an *in vitro* cell model to study bone gene regulation in cells from horses at high and low risk (Palomino Lago et al., 2024), in the absence of any environmental variability. In that study we used a candidate gene approach and identified that *COL3A1* was differentially expressed between high- and low-risk horses, in part owing to the presence of an upstream SNP that we subsequently demonstrated was significantly associated with fracture using genotyping of a cohort (Blott et al., 2014) of 91 phenotyped fracture cases and 86 fracture controls (Palomino Lago et al., 2024).

However, the candidate gene approach only provided insights into a very narrow range of genes and the use of primary cells in this model was associated with

technical limitations such as restricted cell growth. Induced pluripotent stem cells (iPSCs) may provide better cell models owing to their ability to differentiate into a wide range of cell types and unlimited growth *in vitro*. They have been used to study many inherited conditions in humans (Fear et al., 2023; Gorashi et al., 2023; Mae et al., 2023) and are of particular value when access to the affected cell types is limited and/or the cells do not expand well in culture (Pereira et al., 2023). More recently human iPSCs have been used in conjunction with polygenic risk scores (Coleman, 2022; Dobrindt et al., 2021; Yde Ohki et al., 2023) to study the mechanisms underlying complex genetic conditions. We have previously derived iPSCs from horses (Bavin et al., 2015; Palomino Lago et al., 2023) and established the methods to differentiate them into bone forming osteoblasts (Baird et al., 2018, 2019). Furthermore, we demonstrated that the genetic background of the iPSCs affects their expression of osteoblast-associated genes, whereas different iPSC-lines derived from the same horse had less variance (Baird et al., 2018).

The aim of this study was to use iPSC-osteoblasts derived from Thoroughbred horses with high and low PRS for catastrophic fracture and perform global gene expression analysis to identify the biological processes and pathways that are perturbed in osteoblasts from high-risk horses.

MATERIAL AND METHODS

Experimental design

An overview of the experimental design used in this study is shown in Figure 1. Frozen stocks of skin fibroblasts that had been derived and banked from six male Thoroughbred horses were used. The skin biopsies were taken at post-mortem from horses which 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). The samples had previously been genotyped and polygenic risk scores for catastrophic fracture had been calculated to select samples representing the high and low ends of the risk spectrum. Three samples were from horses at high genetic risk (HR) and three samples were from horses at low genetic risk (LR) (Palomino Lago et al., 2024).

iPSC generation

Induced pluripotent stem cell lines were generated previously from skin fibroblasts by retroviral transduction using methods previously described (Baird et al., 2018; Bavin et al., 2015). Briefly, phoenix gag-pol packaging

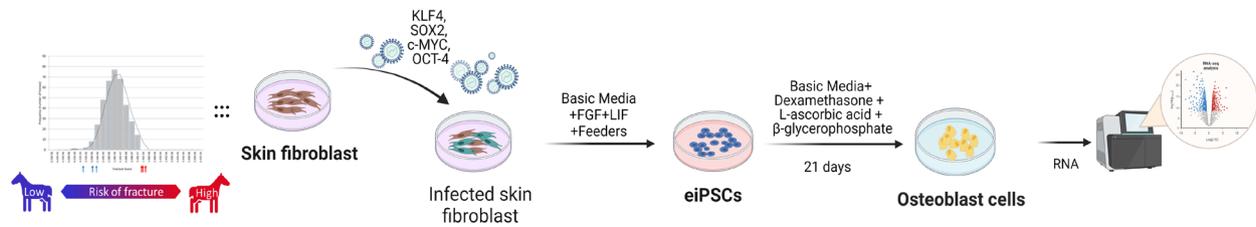


FIGURE 1 Overview of the experimental approach. Created with [Biorender.com](https://biorender.com). Skin fibroblasts from six Thoroughbred horses at low (blue arrows) or high (red arrows) genetic risk of fracture compared with the general population (382 Thoroughbreds of unknown fracture status) were selected as previously described (Palomino Lago et al., 2024). They were retrovirally transduced to overexpress pluripotency genes *KLF4*, *SOX2*, *c-MYC* and *OCT4* and clonal lines of induced pluripotent stem cells (iPSCs) were produced and maintained by culturing in the presence of feeders, FGF (fibroblast growth factor) and LIF (Leukaemia Inhibitory Factor). Equine iPSCs (eiPSCs) were induced to differentiate into osteoblasts by 21 days of culture in osteogenic media prior to RNA isolation and sequencing.

cells were transfected with 3 μg of pVPack-VSV-G (Agilent Technologies, UK) along with 3 μg of pMXs.hOct4 (Octamer-Binding Protein 4, RRID:Addgene_17217), pMXs.hSox2 (SRY-Box Transcription Factor 2, RRID:Addgene_17218), pMXs.hKlf4 (Kruppel Like Factor 4, RRID:Addgene_17219), pMXs.hc-Myc (MYC Proto-Oncogene, RRID:Addgene_17220) or pMX.GFP (Cell Biolabs, USA). Transfections were carried out using lipofectamine3000 and Opti-MEM media (both Invitrogen, Thermo Fisher, UK) according to the manufacturer's instructions. After 48 h culture supernatant containing the viral particles was pooled, filtered through a 0.45 μm filter (Nalgene, Thermo Fisher), supplemented with 10 $\mu\text{g}/\text{mL}$ polybrene (Sigma-Aldrich, UK) and used to infect equine skin fibroblasts which had been plated at a density of 1×10^4 the day before infection. After three rounds of viral infection, performed at 48 h intervals, infected cells were plated at a density of 5×10^3 cells per 10 cm plate pre-seeded with feeder cells (mitotically inactivated mouse embryonic fibroblasts). The media was replaced with basic media (DMEM/F12, 15% fetal bovine serum, 2 mM L-glutamine, 0.1 mM non-essential amino acids, 0.1 mM 2- β -mercaptoethanol, 1 mM sodium pyruvate (all from Thermo Fisher)) plus 1000 U/mL Leukaemia inhibitor factor (Preprotech, UK) and 10 ng/mL basic fibroblast growth factor (Peprotech). The iPSC media was replaced every other day until iPSC colonies began to appear and reached a large enough size to manually pick selected colonies. These colonies were used to establish clonal iPSC lines. For each skin fibroblast sample, two or three clonal lines of iPSCs were generated to provide a total of 14 iPSC lines. The iPSCs were mechanically passaged in the presence of 2 μm Thiazovivin (Miltenyi Biotec, UK). All of the equine iPSC clones were previously characterised for their ability to form embryoid bodies, express pluripotency markers and differentiate into derivatives of endoderm, ectoderm and mesoderm (Baird et al., 2018; Bavin et al., 2015). Furthermore, all of the iPSC lines used in this study were differentiated into osteoblasts capable of producing a mineralised matrix and expressing osteoblast associated genes (Baird et al., 2018, 2019; Palomino Lago et al., 2023).

iPSC differentiation into osteoblasts

Osteoblasts were generated as previously described (Baird et al., 2018). Small pieces of iPSC colonies were plated at a density of 7×10^4 cells per well of a 24-well OsteoAssay surface coated plate (Corning, Wiesbaden, Germany) without feeders and in iPSC media in the presence of 2 μm Thiazovivin for the first 24 h. The following day, the media was replaced with osteogenic media (DMEM/F12, supplemented with 15% fetal bovine serum, 2 mM L-glutamine, 1% non-essential amino acids, 1 mM sodium pyruvate, 0.1 mM 2-mercaptoethanol (all Invitrogen, Thermo Fisher), 10 mM β -glycerophosphate, 10 nM dexamethasone and 28 μM ascorbic acid (all Sigma-Aldrich)). Cells were cultured for 21 days with the media replaced every 2–3 days prior to RNA extraction.

RNA extraction

Following cell differentiation, RNA was collected using Tri-reagent (Sigma-Aldrich) and extracted using the RNeasy Mini Kit (Qiagen, UK) following the manufacturer's instructions. Purified RNA was treated with the DNA-free™ DNA removal kit (Invitrogen, Thermo Fisher) to remove genomic DNA contamination according to the manufacturer's instructions. RNA concentration was determined using a Qubit (Thermo Fisher Scientific) and the purity was determined using a DS-11 spectrophotometer to ensure 260/280 ratios ~ 2.0 . RNA integrity was measured using a TapeStation (2100 Bioanalyzer; Agilent, UK) and for all samples was confirmed to be >9.0 .

RNA sequencing

RNA from 14 lines of iPSC-osteoblasts (passages 8–15) was used in RNA sequencing. mRNA library preparation and transcriptome sequencing were conducted by Novogene (Cambridge, UK) using an Illumina NovaSeq 6000 to generate 26.1–57.2 million 150 bp paired-end reads per sample. FASTQC (v.0.11.9) (Babraham Bioinformatics,

Cambridge, UK) and MULTIQC (v1.11) were used to establish raw sequencing quality. Raw RNA-sequencing (RNA-seq) reads were aligned to equine transcriptome obtained from the National Centre for Biotechnology Information (Sayers et al., 2022) EquCab 3.0 annotation release 103 (GCF_002863925.1_EquCab3.0, https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_002863925.1/, BioSample ID SAMN02953672) using the pseudoaligner SALMON (v.1.5.2) in Quasi-mapping-based mode with GC-bias correction (Patro et al., 2017). Then, transcript level mappings were imported into RSTUDIO (v.4.2.1) using TXIMPORT (v.1.20) (Soneson et al., 2015). Finally, read counts were normalised using DESEQ2's median of ratios method (Anders & Huber, 2010) and analysed for differential expression using DESEQ2 v.1.22.2 (Love et al., 2014). Genes with a \log_2 fold change (Log_2FC) of ± 1 and an adjusted p -value (p -adj) of ≤ 0.05 were considered as differentially expressed. p -Values were adjusted using the Benjamini and Hochberg method in DESEQ2.

Functional annotation analysis

To investigate the biological function related to the differentially expressed genes (DEGs), the PANTHER classification system (<https://www.pantherdb.org/>) v.17.0 was used to functionally annotate genes based on Gene Ontology (GO) terms (BP, biological process; MF, molecular function; CC, cellular component) and pathways. A false discovery rate (FDR) < 0.05 was considered to be a statistically significant enrichment. GENE ANALYTICS from the LifeMap's GeneCards Suite (<https://geneanalytics.genecards.org/>) was used to perform pathway analysis, with an entity score of > 6 being equivalent to a corrected p -value of ≤ 0.05 and therefore defined as significantly enriched. Pathway analysis was also performed using Qiagen's INGENUITY PATHWAY ANALYSIS. The results of the GO and pathway analyses were visualised using R (v.4.2.2).

Protein–protein interaction network construction

Network analysis was conducted using the STRING (v.10.5) protein network analyser plug-in of CYTOSCAPE (v3.9.1) (Shannon et al., 2003).

Gene set enrichment analysis

Functional class sorting of all expressed genes, regardless of whether they were significantly differentially expressed, was performed using GENE SET ENRICHMENT ANALYSIS (GSEA) software (v.4.3.2) based on H (Hallmark gene sets), C2 (curated gene sets): REACTOME, C2:WP, C2:KEGG, C5 (ontology gene sets) gene set collections (MSigDB v.2023.1) (Subramanian et al., 2005). GSEA first

ranked all expressed genes according to the significance of differential gene expression between the HR and LR of fracture groups. The enrichment score for each gene set is then calculated using the entire ranked list, which reflects how the genes for each set are distributed in the ranked list. Normalised enriched scores (NES) were determined for each gene set. The significant enrichment of gene set cut-off was $|\text{NES}| > 1$, nominal p -value ≤ 0.01 and FDR < 0.25 .

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA; Langfelder & Horvath, 2008) was performed on all expressed genes. The DESEQ2 (Love et al., 2014) package (v.1.44) was used to perform gene expression level normalisation with the 'varianceStabilizing Transformation' function. WGCNA package tool (v.1.75) within R software (v.4.4.1) was used to construct the gene co-expression networks (Langfelder & Horvath, 2008). An optimal soft threshold ($\beta = 14$) was selected based on the criterion of approximate scale-free topology. Module detection was performed using the blockwiseModules function. A heatmap was produced to demonstrate the modules most closely associated with fracture risk. Gene significance was used to measure the correlation between gene expression and fracture risk and module membership was used to measure the correlation between each module eigengene and gene expression.

cDNA synthesis and quantitative RT-PCR

To validate the RNAseq, quantitative PCR (qPCR) was used to measure a number of genes using the same samples as had been used in the RNAseq. cDNA was synthesised from 1 μg of RNA using the sensiFAST cDNA synthesis kit (Bioline, UK). Two microlitres of cDNA (corresponding to 20 ng) was used in qPCR. Equine specific primers were designed using PRIMERBLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>) and MFOLD (<http://www.unafold.org/>) to produce amplicons of 50–150 bp. Primer sequences can be found in Table S1. Quantitative PCR was performed in duplicate using SYBR Green containing supermix (Bioline) on a Bio-Rad C1000 Touch Thermal Cycler (Bio-Rad, UK). The PCR cycle parameters were as follows: 95°C (10 min), followed by 45 cycles of 95°C (15 s), 60°C (15 s) and 72°C (15 s). Following this, a melt curve was produced with readings taken every 1°C from 65 to 95°C. Relative gene expression levels were normalised with the housekeeping gene 18S rRNA using the $2^{-\Delta\Delta C_t}$ method (Livak & Schmittgen, 2001). Normality of the data was confirmed using a Shapiro–Wilk test and equal variance confirmed using Levene's test of homogeneity. An independent t -test was then used to compare the mean expression of

each gene between the LR and HR samples and $p < 0.05$ was considered statistically significant. All analysis was performed using SPSS (v.28.0; IBM, UK).

RESULTS

RNA sequencing reveals there are differentially expressed genes between iPSC-osteoblasts derived from horses with low and high genetic risk for catastrophic fracture

RNA sequencing was performed on 14 lines of iPSC-osteoblasts derived from six different horses (three with a low PRS (L1–L3) and three with a high PRS (H1–H3) for fracture) as detailed in Figure 1. Further information on the cells used in this study can be found in Table S2.

RNA sequencing revealed that, of the 29 197 mapped genes, there were 112 DEGs between the iPSC-derived osteoblasts from horses with a high PRS compared with those with a low PRS (Figure 2A and Appendix S2). To validate the RNA sequencing results,

qPCR was carried out to measure the expression of 17 genes (Figure 2B,C). These genes comprised four osteoblast markers (*ITGAV*, *COL1A1*, *SPARC*, *RUNX2*), six genes which showed no significant differences in expression in the RNA sequencing data (*CALCRL*, *GULP1*, *SLC40A1*, *STAT1*, *GLS* and *COL5A2*) and seven genes which did show differential expression in the RNA sequencing data (*FRMD4A*, *ROBO1*, *ITGA4*, *LOXL2*, *TRMP3*, *APOD* and *ANGPTL4*). We also measured the expression of *COL3A1*, which was not annotated in the reference transcriptome but was differentially expressed in our previous study (Palomino Lago et al., 2024). *FRMD4A*, *ROBO1*, *LOXL2* and *TRMP3* showed a significant difference in expression between HR and LR samples in the RNA sequencing data, but no significant difference in the qPCR. However, all genes showed the same trend for differential expression between the samples (Figure 2C). *RUNX2* was expressed at significantly higher levels in the HR horses in the qPCR (Figure 2C) and although it showed the same trend in the RNA sequencing data (average of 2095 normalised counts in HR

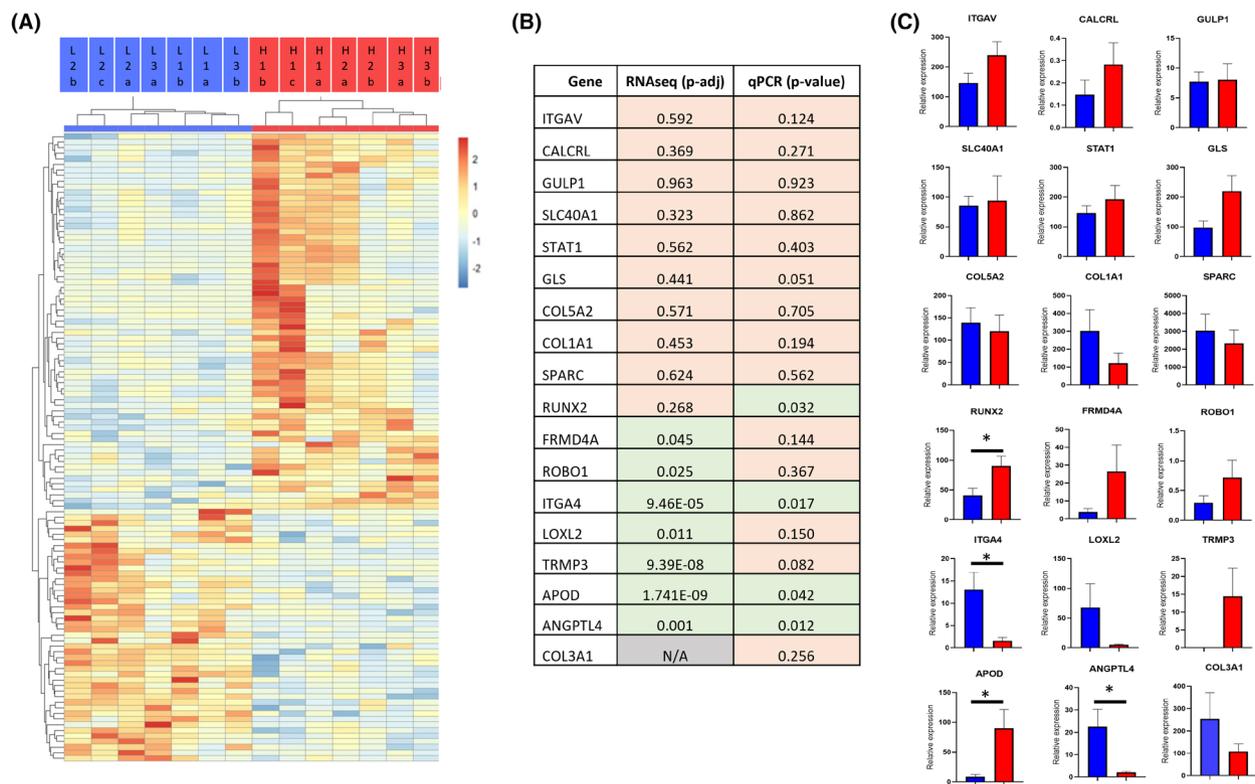


FIGURE 2 Differential gene expression in induced pluripotent stem cell (iPSC)-osteoblasts derived from horses at high (HR) and low risk (LR) of fracture. (A) Heatmap showing the 112 differentially expressed genes ($p\text{-adj} < 0.05$ and a \log_2 fold change (FC) of ± 1). The samples are shown in columns, three LR horses were used (L1–L3) and three HR horses (H1–H3) (blue=LR samples, red=HR samples). Two or three clonal lines of iPSCs were generated from each horse (a, b, c). Genes are clustered by row (see also Appendix S2). (B) Comparison of differentially expressed genes using RNA-sequencing (RNA-seq) and quantitative PCR (qPCR). Genes which are expressed at significantly different levels between HR and LR samples are highlighted in green (p -value; t -test from qPCR data, or p -adj; adjusted p -value from RNAseq data < 0.05); those that were not significantly different are highlighted in orange. N/A, Gene not annotated in the reference transcriptome. (C) qPCR data showing expression relative to the housekeeping gene. Blue bars, LR samples; red bars, HR samples. Error bars represent the SEM of seven replicates. * $p < 0.05$.

samples, vs. an average of 859 counts in the LR samples, Appendix S2), this was not significant. Overall, we demonstrated a 71% concordance between the RNA sequencing and qPCR, which is in accordance with other studies (Paterson et al., 2020). Two clusters of differentially expressed genes were observed: those that were significantly more highly expressed in the osteoblasts from HR horses (67 genes) and those that were expressed at significantly lower levels in the osteoblasts from HR horses (45 genes). With the exception of samples from the HR horse 3 (L3a and L3b), all of the samples derived from the same horse clustered together (Figure 2A).

Of the genes that were downregulated in the HR samples, LOC111769717 (predicted to be *PPWD1*) and LOC100067990 (predicted to be *GBA3*) had the greatest fold changes (\log_2FC -23.677 and -22.306 ; HR vs. LR). Of the genes that were upregulated in the HR samples, LOC111769076 (predicted to be *MUC2*), LOC111768280 (predicted to be *PRR20A*) and *PLA2G4D* had the greatest fold changes (\log_2FC 25.578 , 23.321 and 22.848 , respectively; HR vs. LR) (Figure 3).

A proportion of the differentially expressed genes have known roles in bone

Of the 112 DEGs, 27 are currently unannotated, 43 have published roles in bone or fracture and 42 have no

known role in bone formation of fracture (Table 1). For those genes with a published role in bone or fracture further information can be found in Table 2. The DEGs were distributed across the majority of the chromosomes (Figure S1).

However, network analysis revealed that there are many interactions between the proteins encoded by genes with known and unknown roles in bone formation and or fracture (Figure 4).

Functional enrichment of the differentially expressed genes

Gene ontology analysis of the differentially expressed genes demonstrated that they are involved in numerous biological processes including adhesion, development, morphogenesis, differentiation and extracellular matrix (ECM) organisation. Gene Ontology terms associated with development are generally the most significant and have the greatest number of genes associated with them. Those terms associated with the ECM have larger fold enrichments, but fewer genes associated and higher FDRs. However, many of the differentially expressed gene products were also located in the extracellular matrix and for the GO terms associated with cellular compartment the extracellular space, extracellular region and cell periphery were among the most significant and with the highest number of genes associated with them,

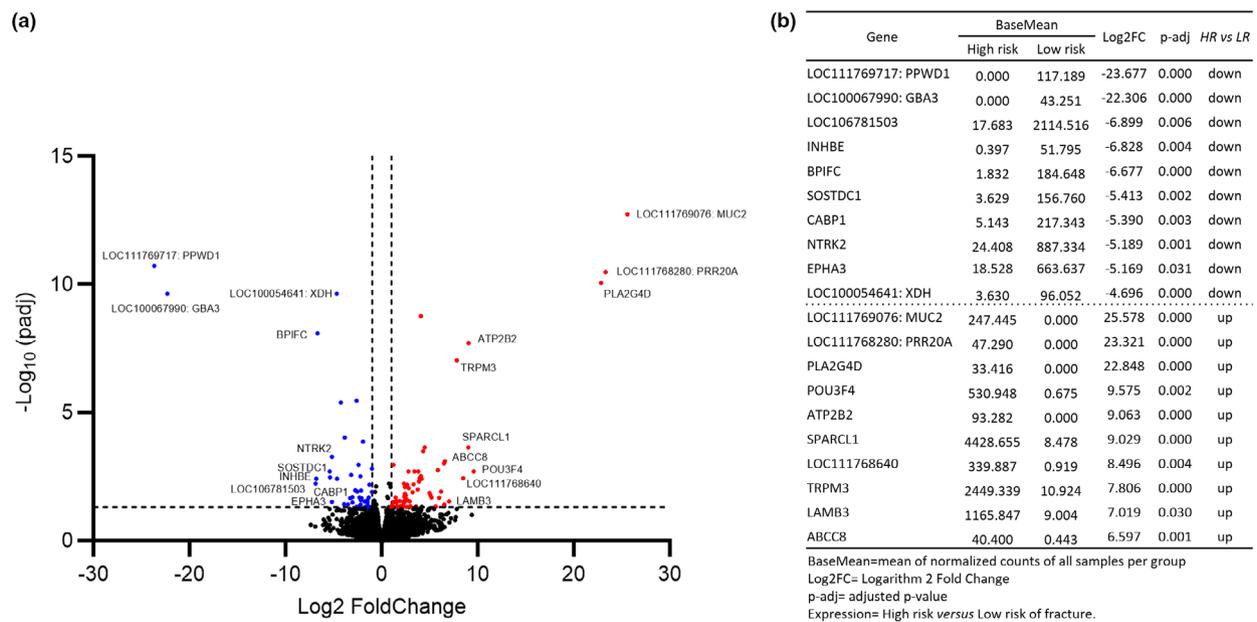


FIGURE 3 Expression of 29,197 genes in induced pluripotent stem cell (iPSC)-osteoblasts from high-risk (HR) compared with low-risk (LR) horses. (a) Volcano plot illustrating the size (\log_2 fold change, \log_2FC) and significance (p -adj) of the differentially expressed genes. Black dots represent genes that are not significantly different between the groups. Red dots represent genes that are significantly upregulated in the HR samples. Blue dots represent genes that are significantly downregulated in the HR samples (p -adj ≤ 0.05 and \log_2FC of ± 1). (b) The top 10 differentially expressed genes (DEGs) in the HR samples compared with the LR samples. For unannotated genes the predicted orthologues are provided after the identifier where possible (predictions from the National Centre for Biotechnology Information; Sayers et al., 2022).

TABLE 1 The differentially expressed genes grouped by their known roles in bone, bone formation and/or fracture. C16H3orf58=DIPK2A (National Centre for Biotechnology Information, NCBI; Sayers et al., 2022). For unannotated genes the predicted orthologues are provided after the identifier where possible (predictions from NCBI; Sayers et al., 2022).

27 Unannotated genes	43 Annotated genes			42 Annotated genes			
		Published role in bone			No published role in bone		
LOC100054641: <i>XHD</i>	LOC106781503	<i>ABCC5</i>	<i>GPM6B</i>	<i>OGN</i>	<i>ABCC8</i>	<i>ENO2</i>	<i>PLPPR3</i>
LOC100067768: <i>UGT3A2</i>	LOC106781645: <i>PCDHB4</i>	<i>ANGPTL4</i>	<i>HHIP</i>	<i>PARVB</i>	<i>ABCG4</i>	<i>FRMD4A</i>	<i>PNLIPRP3</i>
LOC100067990: <i>GBA3</i>	LOC106782354	<i>APOD</i>	<i>HOXD10</i>	<i>POU3F4</i>	<i>ABI3</i>	<i>GPHA2</i>	<i>PPP1R2</i>
LOC100069083: <i>ZNF658</i>	LOC111767614: <i>HS3ST4</i>	<i>ATP2B2</i>	<i>IL15</i>	<i>P2RY6</i>	<i>ADGRB1</i>	<i>INHBE</i>	<i>PTK7</i>
LOC100629234: <i>GTF2IRD1</i>	LOC111768280: <i>PRR20A</i>	<i>CHAC1</i>	<i>ITGA4</i>	<i>RBPM52</i>	<i>ADSSL1</i>	<i>KIZ</i>	<i>SCNNA1</i>
LOC102149222	LOC111768640	<i>CHADL</i>	<i>LAMB3</i>	<i>ROBO1</i>	<i>AMPD3</i>	<i>MATN4</i>	<i>SCRN2</i>
LOC102149317: <i>RALGDS</i>	LOC111769076: <i>MUC2</i>	<i>CNTNAP4</i>	<i>LAMC2</i>	<i>ROBO2</i>	<i>ASNS</i>	<i>MUC4</i>	<i>SNAP91</i>
LOC102149614	LOC111769201	<i>COBL</i>	<i>LGALS1</i>	<i>SENP6</i>	<i>BPIFC</i>	<i>NAPIL5</i>	<i>SPARCL1</i>
LOC102149688	LOC111769717: <i>PPWD1</i>	<i>COL11A2</i>	<i>LOXL2</i>	<i>SLC2A3</i>	<i>BSN</i>	<i>NKAIN1</i>	<i>SPRED3</i>
LOC102150273	LOC111771940	<i>COL2A1</i>	<i>LRRN3</i>	<i>SLC38A3</i>	<i>C16H3orf58</i>	<i>NXPH4</i>	<i>TDRKH</i>
LOC102150465	LOC111772113	<i>EFNB1</i>	<i>MME</i>	<i>SMPD3</i>	<i>CABP1</i>	<i>PCDH9</i>	<i>TMEM38A</i>
LOC106781042	LOC111774027	<i>EFNB2</i>	<i>NR2F1</i>	<i>SOSTDC1</i>	<i>CERKL</i>	<i>PEG10</i>	<i>UCP2</i>
LOC106781149	LOC111775969	<i>EPHA3</i>	<i>NR2F2</i>	<i>SOX10</i>	<i>CLHC1</i>	<i>PLA2G4D</i>	<i>WNK4</i>
LOC106781220: <i>ABCD2</i>		<i>EYA4</i>	<i>NTRK2</i>	<i>TRPM3</i>	<i>COTL1</i>	<i>PLCH2</i>	<i>ZSWIM5</i>
		<i>FBXO2</i>					

despite having lower fold enrichments than other terms. The GO term associated with molecular function was ATPase-coupled transmembrane transport activity, and although this was highly enriched, the number of genes associated with the term was small and the FDR was relatively high (Figure 5).

Pathway analysis was also performed to identify pathways that are over-represented by the differentially expressed genes (Figure 6). This revealed a range of affected pathways with extracellular signal-regulated kinase (ERK) signalling, transport of glucose and other sugars, phospholipase-C pathway degradation of the extracellular matrix and collagen chain trimerisation having the highest number of genes matched and high scores. Results of the pathway analysis were confirmed using Ingenuity Pathway Analysis (Figure S2).

Gene set enrichment analysis

As over-representation analysis was performed on the DEGs only, it excludes genes that did not pass the original cut-off ($p\text{-adj} < 0.05$, $\log_2FC \pm 1$) but that could have a relevant function in a complex disease. GSEA associates a disease phenotype (high risk of fracture) to a group of genes/proteins (Mootha et al., 2003; Subramanian et al., 2005). GSEA demonstrated that six GO-based gene sets, six KEGG (Kyoto Encyclopedia of Gene and Genomes)-based gene sets, 11 WP (WikiPathways)-based gene sets and four REACTOME-based gene sets, were significantly enriched (Table 3) (FDR < 0.05). Many of the enriched pathways and processes were associated with glycolysis, and the associated genes had a lower expression in the iPSC-osteoblasts from HR horses.

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis was performed on all expressed genes, irrespective of whether they were differentially expressed or not. After quality control, this consisted of 13 227 genes across all samples. We identified 32 modules containing highly related genes (Figure 7a,b). Five of these modules were significantly associated with fracture risk (Figure 7c) including the grey60 module (111 genes, correlation co-efficient (cor) = 0.69], p -value = 0.007), the saddlebrown module (30 genes, cor = 0.60], p -value = 0.02), the skyblue module (32 genes, cor = 0.58], p -value = 0.03), the steelblue module (25 genes cor = 0.55], p -value = 0.04) and the turquoise module (3879 genes, cor = 0.54], p -value = 0.04). The 232 genes that could not be included in any module were placed into the grey module and identified as non-co-expressed (p -value = 1) (Figure 7c). However, none of these modules demonstrated a strong correlation between module membership and gene significance (0.32–0.49) and hub genes could not be identified (data not shown).

DISCUSSION

In this study we identified 112 genes that were differentially expressed between iPSC-osteoblasts derived from horses at high and low genetic risk of fracture. Of these, 27 are unannotated, 43 have published roles in bone formation or fracture and 42 have no known role in bone formation or fracture. However, of these 42 genes, all but five of them have been reported to be expressed in equine bone tissue in publicly available datasets (Kemper et al., 2019; Kuemmerle et al., 2016). Furthermore, STRING analyses demonstrated that 18 of the 42 genes with no known role in bone encode

TABLE 2 Summary of the differentially expressed genes that have a published role in bone formation or fracture.

Gene	BaseMean	Log ₂ fold change HR vs. LR	p-adj	Chromosome	Highest expression	Role	Reference
<i>ABCC5</i>	2427.12	1.04	4.60×10^{-2}	19	HR	Involved in osteoclast formation	Kang et al. (2020), Mourskaia et al. (2012)
<i>ANGPTL4</i>	1453.10	-2.40	1.00×10^{-3}	7	LR	Stimulates bone resorption, osteoblast proliferation and differentiation	Knowles et al. (2010), Wilson et al. (2015)
<i>APOD</i>	1951.83	4.07	1.74×10^{-9}	19	HR	Deficiency associated with high bone turnover, low bone mass and impaired osteoblast function	Martineau et al. (2016)
<i>ATP2B2</i>	46.64	9.06	1.99×10^{-8}	16	HR	Upregulated during bone loading and part of a family of proteins involved in bone homeostasis	Francis et al. (2002), Kim et al. (2012), Mantila Roosa et al. (2011)
<i>CHAC1</i>	171.13	-3.20	2.76×10^{-3}	1	LR	<i>CHAC1</i> ^{-/-} mice have decreased osteoblast differentiation and decreased bone density	Crawford et al. (2016)
<i>CHADL</i>	10.31	6.47	1.00×10^{-3}	28	HR	Increased expression in OA bone; is an OA susceptibility gene	Tuerlings et al. (2021)
<i>CNTNAP4</i>	9.25	5.17	2.10×10^{-2}	3	HR	Involved in Nell-1 responsive osteogenesis	Li, Zheng, et al. (2018)
<i>COBL</i>	1404.20	2.64	7.00×10^{-3}	4	HR	Contains a fracture-associated locus	Trajanoska et al. (2018)
<i>COL11A2</i>	434.84	5.84	2.00×10^{-3}	20	HR	Expressed in osteoblasts during bone formation	Urabe et al. (2001)
<i>COL2A1</i>	820.86	6.53	4.50×10^{-2}	6	HR	Mutations associated with skeletal dysplasia with metaphyseal involvement. Bone impairment occurs prior to OA	Rolvien et al. (2020), Walter et al. (2007)
<i>EFNB1</i>	3228.79	-1.02	1.30×10^{-2}	X	LR	Required for bone formation	Cheng et al. (2013), Nguyen et al. (2016), Xing et al. (2010)
<i>EFNB2</i>	143.84	2.37	2.90×10^{-2}	17	HR	Required for normal bone formation	Arthur and Gronthos (2021)
<i>EPHA3</i>	341.08	-5.17	3.10×10^{-2}	26	LR	Part of a family of proteins involved in bone development and homeostasis	Arthur and Gronthos (2021)
<i>EY44</i>	105.62	4.20	5.00×10^{-2}	10	HR	Gene variants associated with upper limb and skull bone mineral density	Kemp et al. (2014)
<i>FBXO2</i>	323.25	3.88	5.00×10^{-3}	2	HR	Upregulates osteosarcoma cell proliferation via STAT3	Zhao et al. (2020)
<i>GPM6B</i>	1543.60	3.95	4.00×10^{-3}	X	HR	Silencing GPM6B leads to decreased bone differentiation of mesenchymal stromal cells (MSCs)	Drabek et al. (2011)
<i>HHIP</i>	300.51	-3.04	3.30×10^{-2}	2	LR	Increased HHIP inhibits osteogenic differentiation of MSCs	Yin et al. (2022)
<i>HOXD10</i>	173.82	2.48	2.30×10^{-2}	18	HR	Targeted disruption of HOXD10 leads to vertebral and hindlimb bone abnormalities	Carpenter et al. (1997)
<i>IL15</i>	705.42	-1.77	2.90×10^{-2}	2	LR	Pro osteoclastogenic cytokine	Ogata et al. (1999)

TABLE 2 (Continued)

Gene	BaseMean	Log ₂ fold change HR vs. LR	p-adj	Chromosome	Highest expression	Role	Reference
<i>ITGA4</i>	210.42	-3.86	9.47×10^{-5}	18	LR	Involved in osteoblast differentiation and mineralisation and regulation of osteoclast activity	Mao et al. (2023)
<i>LAMB3</i>	587.43	7.02	3.00×10^{-2}	5	HR	Part of Laminino-332 which is produced by osteoblasts and inhibits osteoclast differentiation	Uehara et al. (2017)
<i>LAMC2</i>	1993.41	-2.17	3.90×10^{-2}	5	LR	Part of Laminino-332 which is produced by osteoblasts and inhibits osteoclast differentiation	Uehara et al. (2017)
<i>LGALS1</i>	9443.16	-1.48	3.40×10^{-2}	28	LR	K knockout mice have bone loss owing to a lack of bone formation. Promotes bone differentiation of mouse MSCs. Regulates osteoclast activity	Chen et al. (2022), Muller et al. (2019)
<i>LOXL2</i>	2208.11.	-7.74	1.10×10^{-2}	2	LR	Associated with the chondrogenic phase of fracture healing, weak LOXL2 staining found in osteoblasts	Iftikhar et al. (2011)
<i>LRRN3</i>	1134.25	1.69	5.00×10^{-2}	4	HR	K knockout mice have decreased bone mineral density and bone mineral content	Groza et al. (2022)
<i>MME</i>	556.80	-2.74	1.10×10^{-2}	2	LR	A marker of osteoblast differentiation of MSCs	Granéli et al. (2014)
<i>NR2F1</i>	776.04	2.99	4.65×10^{-2}	14	HR	Enhances osteogenic differentiation of human MSCs	Lee et al. (2022)
<i>NR2F2</i>	792.09	1.46	2.6×10^{-2}	1	HR	K knockout suppresses MSC osteogenic differentiation	Zhu et al. (2016)
<i>NTRK2</i>	455.87	-5.19	1.00×10^{-3}	23	LR	Involved in osteoblast differentiation through ERK signalling	Liu et al. (2018), Zhang et al. (2020)
<i>OGN</i>	3654.69	3.32	6.00×10^{-3}	23	HR	Increases osteoblast differentiation of MSCs through regulation of osteogenic genes; role in senile osteoporosis	Chen et al. (2017)
<i>PARVB</i>	119.36	-2.14	1.10×10^{-2}	28	LR	K knockout mice have increased bone mineral density and bone mineral content	Groza et al. (2022)
<i>POU3F4</i>	265.81	9.58	2.00×10^{-3}	X	HR	Regulates EFNB2 to control temporal bone development	Raft et al. (2014)
<i>P2RY6</i>	18.33	-4.67	3.88×10^{-3}	7	LR	Stimulates bone resorption by osteoclasts. Expressed in osteoblasts	Orriss et al. (2011)
<i>RBPMS2</i>	147.30	-1.29	5.00×10^{-2}	1	LR	K knockout mice have decreased bone mineral density and bone mineral content	Groza et al. (2022)
<i>ROBO1</i>	218.37	2.60	2.50×10^{-3}	26	HR	Slit/Robo signalling regulates bone formation and resorption	Jiang et al. (2022)
<i>ROBO2</i>	206.11	3.82	2.00×10^{-3}	26	HR	Slit/Robo signalling regulates bone formation and resorption	Jiang et al. (2022)
<i>SENP6</i>	6868.05	2.36	2.90×10^{-2}	10	HR	K knockout in osteochondroprogenitors impairs skeletal growth	Li, Lu, et al. (2018)
<i>SLC2A3</i>	3592.18	-1.36	2.08×10^{-2}	6	LR	Silencing in osteoblasts increases proliferation and contributes to glucose uptake	Arponen et al. (2022)
<i>SLC38A3</i>	29.53	4.99	1.41×10^{-2}	16	HR	Heterozygous knockout mice have increased bone mineral density and abnormal bone structure	Groza et al. (2022)

(Continues)

TABLE 2 (Continued)

Gene	BaseMean	Log ₂ fold change HR vs. LR	p-adj	Chromosome	Highest expression	Role	Reference
<i>SMPD3</i>	315.60	4.48	2.30×10^{-3}	3	HR	Deficiency in mice results in skeletal deformities and delayed extracellular matrix mineralisation. Role in mineralisation during fracture healing	Manickam et al. (2019)
<i>SOSTDC1</i>	80.20	-5.41	2.00×10^{-3}	4	LR	Involved in bone metabolism, maintenance of bone density and fracture repair	Collette et al. (2016), Tong et al. (2022)
<i>SOX10</i>	34.75	5.09	1.77×10^{-2}	28	HR	Involved in osteogenesis and bone regeneration of mandible	Stuepp et al. (2020)
<i>TRPM3</i>	1230.13	7.81	9.39×10^{-8}	23	HR	Regulates calcium mediated RANKL expression in osteoblasts	Son et al. (2018)

Note: BaseMean = average of the normalised count values, divided by size factors and taken over all samples.

Abbreviations: HR, induced pluripotent stem cell (iPSC)-osteoblasts derived from high-risk horses; LR, iPSC-osteoblasts derived from low-risk horses; OA, osteoarthritis.

proteins that interact (directly or indirectly) with proteins that have known roles in bone. A further four of the unknown genes encode proteins that interact with each other (SCANN1A with WNK4 and PEG10 with NAP1L5) but have no direct or indirect interactions with proteins that have a known role in bone. This suggests that at least a proportion of these genes are likely to have roles in bone which have not yet been discovered and should be followed up in future work.

Genes downregulated in osteoblasts from horses with a high polygenic risk score for fracture

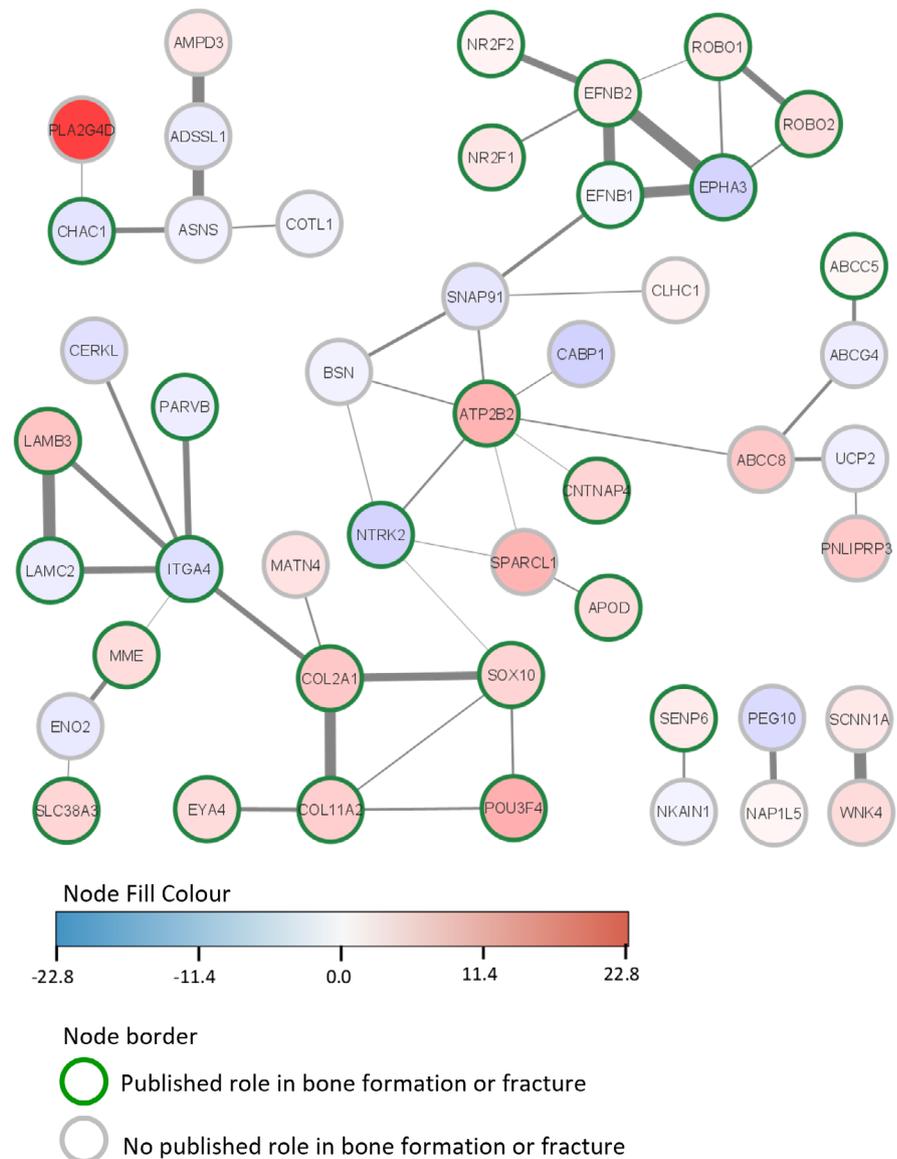
Of the top 10 genes with the largest differential expression (DE) that were downregulated in the HR samples, four of them are not yet annotated in the horse genome. However, LOC111769717 is predicted to be *PPWD1*, whose expression is associated with post-menopausal osteoporosis (Qian et al., 2019). Two of the other non-annotated transcripts also have predicted genes (*GBA3* and *XDH*), although a role in bone has not been reported. Three of the other top 10 downregulated genes have some previously published links with bone. For example, *SOSTDC1* has been shown to be involved in bone metabolism, the maintenance of bone density and fracture repair (Collette et al., 2016; Tong et al., 2022). *NTRK2* is involved in osteoblast differentiation of mesenchymal stromal cells (MSCs) (Zhang et al., 2020) and binding of *NTRK2* to its receptor results in the phosphorylation of ERK1/2, which drives the expression of the bone-promoting transcription factors *RUNX2* and *SP7* (Osterix) (Liu et al., 2018). *EPHA3* is a member of the ephrin receptor family and ephrin signalling is involved in bone development and homeostasis (Arthur & Gronthos, 2021).

Two of the genes, *INHBE* (a member of the TGF- β superfamily) and *BPIFC* (predicted to enable lipid binding) have no known role in bone formation or fracture, do not appear in the STRING analysis as associating at the protein level with the other differentially expressed encoded proteins, and are not in the gene sets of the enriched GO terms or pathway analysis. Therefore, it is not clear what their biological relevance might be. However, the expression of both genes has been detected in horse bone tissue, and/or cultured horse cells (Kuemmerle et al., 2016). *CABPI* is also in the top 10 and although it has no published role in bone, it appears in the main interaction node of the STRING analysis, is a calcium binding protein and is expressed in horse bone tissue (Kemper et al., 2019; Kuemmerle et al., 2016).

Genes upregulated in osteoblasts from horses with a high polygenic risk score for fracture

Of the top 10 genes that were upregulated in the HR samples, three were not annotated and, where

FIGURE 4 Functional network interactions of the proteins encoded by the 112 differentially expressed genes (DEGs) conducted with STRING. Proteins encoded by genes which are upregulated in the high-risk (HR) horses are shown in shades of red; proteins which are downregulated in HR horses are shown in blue. Proteins with a known role in bone formation/fracture have a green border; proteins with no known role in bone formation/fracture have a grey border. A thicker connecting line indicates a stronger interaction score. The DEGs with zero nodes are not plotted.



available, their predicted genes (*MUC2* and *PRR20A*) have no links to bone or fractures. Three of the genes, *PLA2G4D*, *SPARCL* and *ABCC8* have no known role in bone formation or fracture. However, all three genes are expressed in horse bone tissue and/or cultured horse cells (Kuemmerle et al., 2016) suggesting that they may have novel roles in bone which have yet to be defined. *ABCC8* functions as a modulator of ATP-sensitive potassium channels and insulin release and it is in the enriched gene set for the overrepresented pathway transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds. Furthermore, it interacts with *ATB2BP2*, which is also significantly upregulated in the HR samples. *SPARCL* also interacts with *ATB2BP2*. *SPARCL*, which is predicted to enable calcium ion, collagen and extracellular matrix binding activity, is a paralogue of *SPARC*, which plays a critical role in bone remodelling (Rosset & Bradshaw, 2016). However,

the role of *SPARCL* in bone is so far unknown. *PLA2G4D* also appears in the STRING network and interacts with *CHAC1*. Knock-out of *CHAC1* in mice leads to decreased osteoblast differentiation and bone density (Crawford et al., 2016). *PLA2G4D* is part of the phospholipase A2 family, and it is present in the enriched gene set for the overrepresented pathway phospholipase-C pathway.

The remaining five genes of the top 10 that were upregulated in the HR samples include *TRPM3*, which is involved in bone remodelling through the regulation of *RANKL* expression in osteoblasts (Son et al., 2018). Similarly, *LAMB3* is also involved in bone remodelling. *LAMB3* forms part of laminin-332, which is produced by osteoblasts and suppresses osteoclast differentiation (Uehara et al., 2017). *ATP2B2* is involved in maintaining calcium homeostasis. It is upregulated during bone loading (Mantila Roosa et al., 2011), is expressed by osteoblasts (Francis et al., 2002) and is part of a family of proteins which are required for

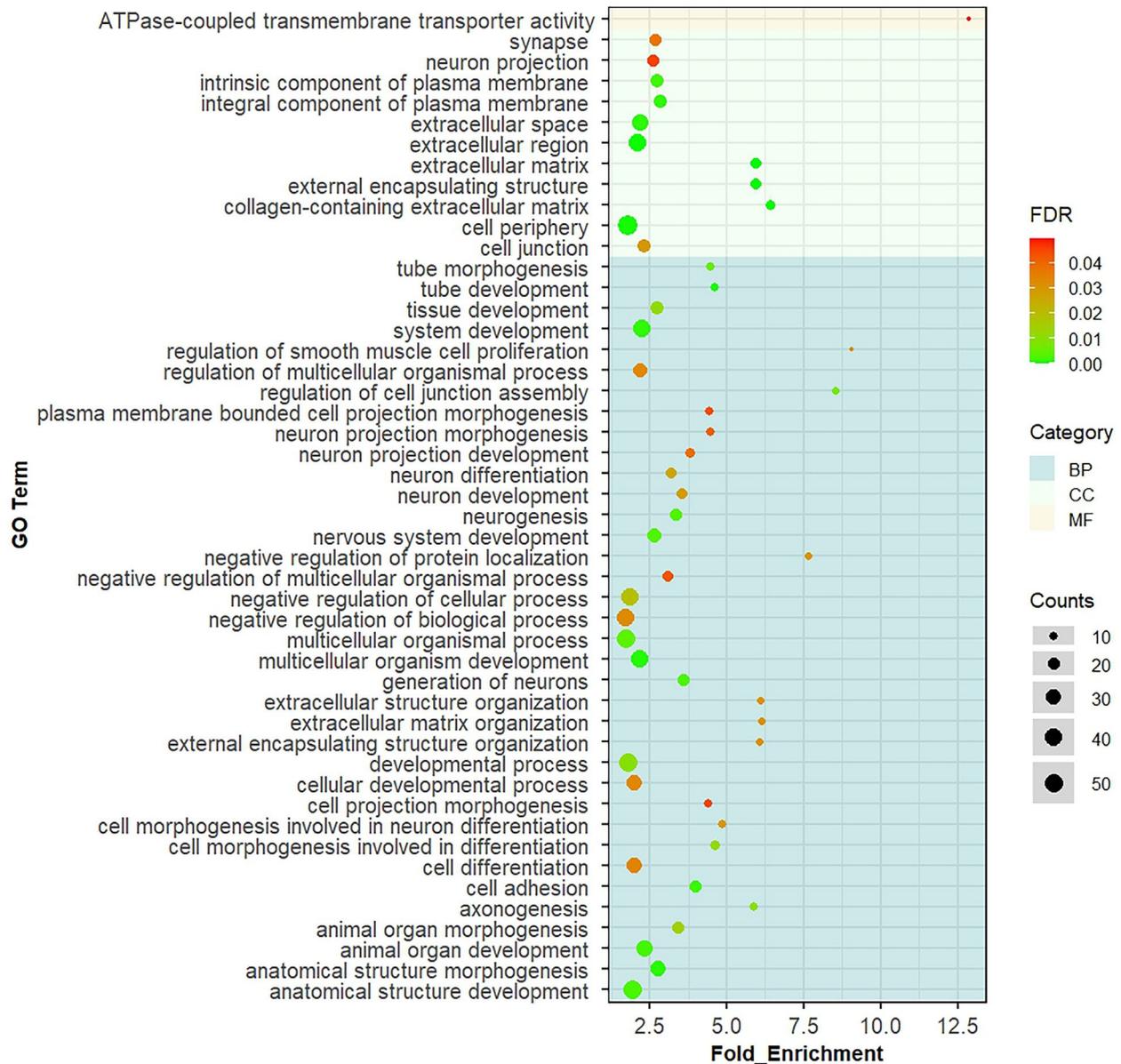


FIGURE 5 Gene Ontology (GO) terms significantly over-represented by the differentially expressed genes. The background colour corresponds with biological processes (BP, dark blue), cellular component (CC, light blue) and molecular function (MF, light orange). The colour of the dot shows the significance of the GO terms (colour scale bar indicating the range from 0 [green] to 0.04 [red] false discovery rate). The size of the dot represents the number of differentially expressed genes (DEGs) in that specific GO term.

regulating bone mass (Kim et al., 2012). *POU3F4* regulates *EFNB2* to control temporal bone development (Raft et al., 2014). Interestingly, *EFNB2* itself was also significantly upregulated in the HR samples and the ephrin B reverse signalling pathway was significantly over-represented by the differentially expressed genes.

Therefore, the known functions of the top 10 differentially expressed genes suggest that the regulation of bone remodelling and calcium signalling may be altered in bone cells from horses at high risk of fracture. However, it is possible that the PRS of the iPSCs affects their efficiency of differentiation into osteoblasts and that this accounts for some of the differences in gene expression. Therefore,

functional *in vitro* studies are required to confirm the role of the differentially expressed genes in bone and fracture risk and provide biological validation of our results.

Function of the differentially expressed genes

Gene Ontology analysis of all the differentially expressed genes revealed that the molecular function 'ATPase-coupled transmembrane transporter activity' was over-represented, although despite a high fold enrichment, the number of genes associated with this term was low (four genes: *ABCC5*, *ABCC4*, *ATP2B2* and *ABCC8*)

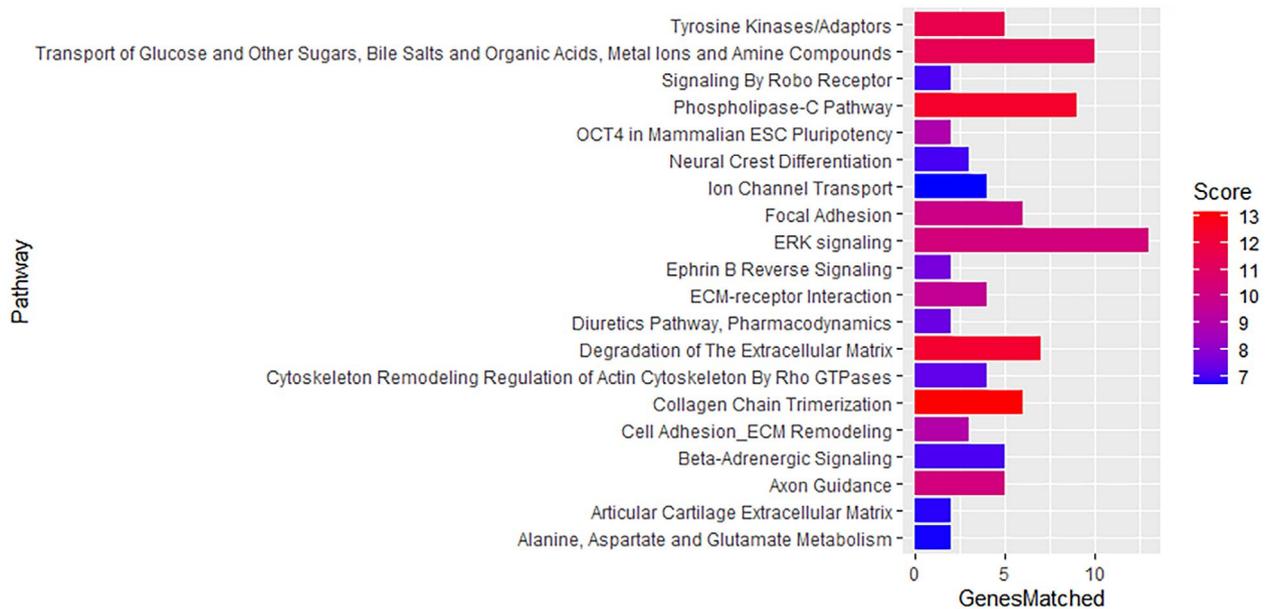


FIGURE 6 Pathways that are significantly over-represented by the differentially expressed genes (DEGs). The score is a transformation of the binomial distribution p -value and the score range is divided into three quality levels, according to the p -value they are derived from (after correction for multiple comparisons), represented from lower (blue) to higher (red) score. GenesMatched = number of genes over-represented in that specific pathway.

and the FDR is relatively high. However, ATPase activity is crucial for regulating bone formation and resorption (Francis et al., 2002) and *ABCC5* and *ATP2B2* are involved in osteoclast formation and bone homeostasis respectively (Francis et al., 2002; Kang et al., 2020; Kim et al., 2012; Mantila Roosa et al., 2011; Mourskaia et al., 2012). Furthermore, the proteins encoded by all four of these genes interact and therefore further investigation of their role in fracture risk is warranted. Gene Ontology analysis further revealed that the extracellular matrix/region was most often over-represented by the differentially expressed genes. For the biological processes, in addition to the many processes involved in differentiation and development, there were a number of processes related to the ECM, although fewer genes were associated with these terms and they had higher FDR. Nevertheless, for the GO terms associated with cellular compartment, the extracellular space and region had high numbers of associated genes and were among the most significant. Together this may suggest that the differentially expressed genes are ultimately affecting bone homeostasis and ECM components. It is noted that many neuronal GO processes are also over-represented by the DEGs. The cross-talk between brain and bone health is becoming increasingly apparent (Otto et al., 2020) and it is possible that many of the DEGs have to date been better studied in neuronal cell types than osteogenic cells, influencing the GO processes they are known to be involved in. For example, Slit/Robo signalling was initially found to be essential during nerve development, before its role in regulating bone formation was discovered (Jiang et al., 2022). Therefore, *ROBO1/ROBO2* appear in many of the gene

lists associated with neuronal biological processes and pathways.

Pathway analysis revealed multiple pathways involved in bone. For example, the ERK signalling pathway had the greatest number of DEGs and this pathway is involved in fracture repair (Chen & Luan, 2019). Collagen chain trimerisation was one of the most significantly over-represented pathways and this is affected in numerous bone diseases (Bourhis et al., 2012). Similar to the GO analysis, pathways relating to the degradation and remodelling of the ECM were also over-represented. The phospholipase C pathway and transport of glucose also had high numbers of genes matched and high scores. The phospholipase C pathway has previously been shown to activate osteoclastogenesis and is involved in fracture repair (Li, Yuan, et al., 2018). As fracture is a complex genetic disease, we also performed GSEA and WGCNA on all genes, irrespective of whether they were differentially expressed between HR and LR samples. GSEA revealed biological processes and numerous pathways involved in glycolysis. Human patients with diabetes are at higher risk of bone fractures (Valderrábano & Linares, 2018) and altered glucose metabolism may be important in maintaining bone homeostasis (Karner & Long, 2018). Weighted gene co-expression network analysis identified five modules that were significantly associated with fracture risk with each module consisting of 25–3879 genes. However, none of these modules demonstrated a strong correlation between module membership and gene significance and hub genes could not be identified. We therefore did not perform

TABLE 3 GENE SET ENRICHMENT ANALYSIS (GSEA) showing the significantly enriched gene ontology terms and pathways.

Database collections	Gene set	NES	FDR	Higher expression in
<i>GO-based list (C5: BP, CC, MP)</i>				
GOBP_PROTEIN_EXIT_FROM_ENDOPLASMIC_RETICULUM	GO:0032527	2.30	0.037	LR
GOBP_GLYCOLYTIC_PROCESS_THROUGH_FRUCTOSE_6_PHOSPHATE	GO:0061615	2.27	0.033	LR
GOBP_ENDOPLASMIC_RETICULUM_TO_CYTOSOL_TRANSPORT	GO:1903513	2.27	0.026	LR
GOCC_ENDOPLASMIC_RETICULUM_PROTEIN_CONTAINING_COMPLEX	GO:0140534	2.23	0.024	LR
GOBP_NADH_REGENERATION	GO:0006735	2.22	0.027	LR
GOBP_NEGATIVE_REGULATION_OF_NUCLEOTIDE_METABOLIC_PROCESS	GO:0045980	2.20	0.031	LR
<i>Pathways-based list (C2)</i>				
REACTOME_IRE1ALPHA_ACTIVATES_CHAPERONES	R-HSA-381070	2.30	0.007	LR
REACTOME_COHESIN_LOADING_ONTO_CHROMATIN	R-HSA-2470946	-2.27	0.003	HR
REACTOME_ESTABLISHMENT_OF_SISTER_CHROMATID_COHESION	R-HSA-2468052	-2.17	0.012	HR
REACTOME_MITOTIC_TELOPHASE_CYTOKINESIS	R-HSA-68884	-2.14	0.012	HR
WP_CORI_CYCLE	WP1946	2.28	0.005	LR
WP_PHOTODYNAMIC_THERAPYINDUCED_HIF1_SURVIVAL_SIGNALING	WP3614	2.27	0.003	LR
WP_GLYCOLYSIS_IN_SENESCENCE	WP5049	2.14	0.011	LR
WP_AEROBIC_GLYCOLYSIS	WP4629	2.13	0.010	LR
WP_GLYCOLYSIS_AND_GLUONEOGENESIS	WP534	2.05	0.020	LR
WP_BASE_EXCISION_REPAIR	WP4752	2.03	0.024	LR
WP_CLEAR_CELL_RENAL_CELL_CARCINOMA_PATHWAYS	WP4018	1.96	0.041	LR
WP_PHOTODYNAMIC_THERAPYINDUCED_UNFOLDED_PROTEIN_RESPONSE	WP3613	1.96	0.037	LR
WP_ONCOSTATIN_M_SIGNALING_PATHWAY	WP2374	1.95	0.036	LR
WP_MRNA_PROTEIN_AND_METABOLITE_INDUCATION_PATHWAY_BY_CYCLOSPORIN_A	WP3953	1.93	0.044	LR
WP_NIPBL_ROLE_IN_DNA_DAMAGE_CORNELIA_DE_LANGE_SYNDROME	WP5119	-2.03	0.038	HR
KEGG_GLYCOLYSIS_GLUONEOGENESIS	hsa00010	0.52	0.005	LR
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	hsa04672	0.55	0.017	LR
KEGG_BASE_EXCISION_REPAIR	hsa03410	0.51	0.012	LR
KEGG_GALACTOSE_METABOLISM	hsa00052	0.54	0.023	LR
KEGG_FRUCTOSE_AND_MANNOSE_METABOLISM	hsa00051	0.49	0.020	LR
KEGG_LEISHMANIA_INFECTON	hsa05140	0.42	0.047	LR

Abbreviations: BP, biological process; CC, cellular component; FDR, false discovery rate; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; MP, molecular process; NES, normalised enrichment score; WP, WikiPathways.

subsequent gene ontology analysis on the module genes. This probably reflects the small sample size used in the study as it has been suggested that over 20 samples are required for robust WGCNA analysis (Ballouz et al., 2015).

Limitations of the study

This study had a number of limitations. The polygenic risk score to select cells for use in this study was established previously (Palomino Lago et al., 2024) and it has

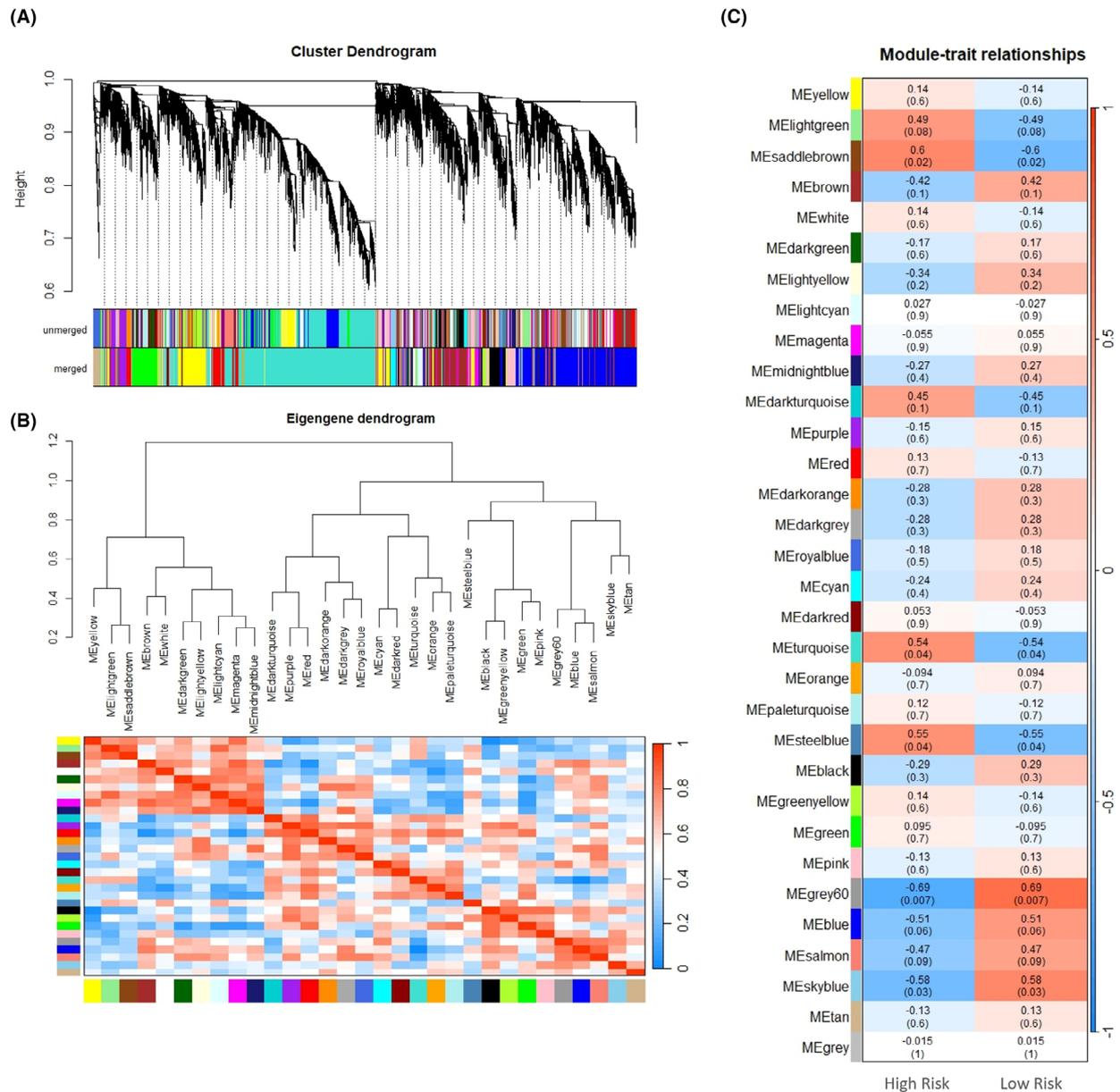


FIGURE 7 Construction of weighted gene co-expression network identification of modules. (a) Clustered dendrogram of genes and modules eigengenes; unmerged = preliminary module construction; merged = integrated module. The branches and colour bands represent the assigned module. The tips of the branches represent genes. The distance between two genes is shown as the height on the y-axis. (b) Clustering tree based on modules' eigengenes with the heatmap plot of the adjacencies of modules. Red represents high adjacency (positive correlation) and blue represents low adjacency (negative correlation). (c) Relationships of consensus module eigengenes and risk of fracture. The module name is shown on the left side of each cell. Numbers in the table report the correlations of the corresponding module eigengenes and traits, with the *p*-values printed below the correlations in parentheses. The table is colour-coded by correlation according to the colour legend. The intensity and direction of correlations are indicated on the right side of the heatmap (red, positively correlated; blue, negatively correlated).

not yet been validated in a new cohort of fracture cases and controls. Methods of RNA sequencing analysis are undergoing constant improvements, and it is possible that newer approaches to normalisation and differential expression analyses would generate different results (Smid et al., 2018). Furthermore, osteoblasts were only derived from a small number of horses. This may have resulted in a lack of power to detect additional significant DEGs and gene co-expression networks. However, owing to the samples we had available (Palomino Lago

et al., 2024), we were not able to increase the sample size while selecting samples that were at the most extreme ends of the polygenic risk score. Following guidelines on the use of iPSCs for disease modelling, we utilised multiple clones per horse. The number of donors and clones used here is similar to those in other studies of complex diseases in humans (Gorashi et al., 2023; Yde Ohki et al., 2023) and reflects the time and expense involved in deriving and validating an iPSC line. Interestingly a study using iPSCs from three patients with diabetes and

three healthy patients (one or two clonal lines from each) found a similar number (245) of differentially expressed genes to us (Gorashi et al., 2023). However, the power to detect statistically significant differences in gene expression is likely to increase if sample size could be increased and recent studies have shown that this will be of more benefit than including more clonal lines from the same individual (Beekhuis-Hoekstra et al., 2021; Hoffman et al., 2019). Our restricted cohort from which we could select samples also meant that it was not possible to validate the RNA sequencing with qPCR in a larger number of samples. The qPCR performed showed good corroboration of the sequencing results and where differences were observed between the techniques, this was at the level of significance with the same trends in expression patterns being shown for all of the genes examined. Further functional *in vitro* studies are therefore critical to determine gene function and validate our findings. Finally, the original samples were collected anonymously, and the ages and history of the horses were not known. Likewise, for the majority of samples it was not known if the horses were gelded or entire (although two of the HR horses were known to be gelded). During the generation of human iPSCs, the epigenetic clock is reset and is close to zero (Horvath, 2013). This has yet to be confirmed for equine iPSCs; however, it is likely that any variation in the age or work history of the donors does not impact the results.

In our previous study (Palomino Lago et al., 2024), we demonstrated that fibroblast cells differentiated directly into osteoblast-like cells, with a significant difference in the expression of *COL3A1* and *STAT1* between samples isolated from HR and LR horses (Palomino Lago et al., 2024). In this study, while we saw the same trend for *STAT1*, the difference was not significant (Figure 2 and Appendix S2). The reference transcriptome used did not have *COL3A1* annotated, and therefore it was not analysed in the RNA sequencing. However, qPCR analysis also revealed that while it had the same trend as our previous study (expressed 2.3 fold lower in the HR samples), this difference was not significant owing to the variability between samples. This may reflect the different starting populations of cells and the fact that the osteoblast populations we derived in both studies are likely to be heterogeneous. However, both studies were consistent in demonstrating that *ITGAV*, *CALCRL*, *GULP1*, *SLC401A*, *GLS* and *COL5A2* show no significant differences in expression between the HR and LR osteoblasts. Both studies also demonstrate little or no expression of *MSTN* and *ZNF804A* in the samples. The *COL3A1* and *STAT1* genes lie within a fracture associated region on ECA18 (Blott et al., 2014; Tozaki et al., 2020). In this study, we found only three of the differentially expressed genes to be located on ECA18 (*HOXD10*, *ITGA4*, *CERKL*). However, none of them lie within the associated region. It is not clear if our study lacked the power to detect smaller differences in the expression of any of

these genes from this region, or if DNA variants within the region are regulating more distant genes.

Finally, in our model, we directed the iPSCs to differentiate into osteoblasts. However, it is likely that other cell types (e.g. osteoclasts) also contribute to fracture risk. Human iPSCs have successfully been differentiated into osteoclasts (Chen, 2020), but to date this not yet been reported for equine iPSCs and it would be of benefit to measure gene expression profiles in other cell types in the future. Similarly, we did not have access to bone tissue from fracture cases and controls and so were unable to confirm if these differences in expression also occur *in vivo* at either the gene or protein level.

CONCLUSION

In conclusion, we have demonstrated that iPSC-osteoblasts derived from horses with high and low polygenic risk scores for catastrophic fracture, have many differently expressed genes that are overrepresented in various pathways and processes that have relevance to bone homeostasis and fracture. A deeper understanding of the consequences of mis-regulation of these genes and the identification of the DNA variants which underpin their differential expression may reveal more about the molecular mechanisms that are involved in equine bone health and fracture risk. Future work should now perform functional *in vitro* studies to determine the specific roles of these genes in bone and fracture risk.

AUTHOR CONTRIBUTIONS

Esther Palomino Lago: Data curation; formal analysis; investigation; methodology; validation; visualization; writing – original draft; writing – review and editing. **Amy K. C. Ross:** Data curation; formal analysis; writing – review and editing. **Alyce McClellan:** Investigation; writing – review and editing. **Deborah J. Guest:** Conceptualization; funding acquisition; methodology; project administration; resources; supervision; writing – original draft; writing – review and editing.

ACKNOWLEDGEMENTS

This study was kindly funded by the Horserace Betting Levy Board (vet/prj/792) and the Alborada Trust Fund A.K.C.R.

CONFLICT OF INTEREST STATEMENT

E. Palomino Lago, A.K.C. Ross 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. A. McClellan has no competing interests to declare.

DATA AVAILABILITY STATEMENT

The RNA sequencing datasets are available in the National Centre for Biotechnology Information Gene Expression Omnibus repository (www.ncbi.nlm.nih.gov/geo) under accession number GSE255417. The differentially expressed genes and normalised counts data are included in Appendix S2.

ORCID

Deborah J. Guest  <https://orcid.org/0000-0002-0034-3332>

REFERENCES

- Anders, S. & Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biology*, 11, R106.
- Anthenill, L.A., Gardner, I.A., Pool, R.R., Garcia, T.C. & Stover, S.M. (2010) Comparison of macrostructural and microstructural bone features in thoroughbred racehorses with and without mid-body fracture of the proximal sesamoid bone. *American Journal of Veterinary Research*, 71, 755–765.
- Anthenill, L.A., Stover, S.M., Gardner, I.A. & Hill, A.E. (2007) Risk factors for proximal sesamoid bone fractures associated with exercise history and horseshoe characteristics in thoroughbred racehorses. *American Journal of Veterinary Research*, 68, 760–771.
- Arponen, M., Jalava, N., Widjaja, N. & Ivaska, K.K. (2022) Glucose transporters GLUT1, GLUT3, and GLUT4 have different effects on osteoblast proliferation and metabolism. *Frontiers in Physiology*, 13, 1035516.
- Arthur, A. & Gronthos, S. (2021) Eph-ephrin signaling mediates cross-talk within the bone microenvironment. *Frontiers in Cell and Development Biology*, 9, 598612.
- Baird, A., Dominguez Falcon, N., Saeed, A. & Guest, D.J. (2019) Biocompatible three-dimensional printed thermoplastic scaffold for osteoblast differentiation of equine induced pluripotent stem cells. *Tissue Engineering. Part C, Methods*, 25, 253–261.
- Baird, A., Lindsay, T., Everett, A., Iyemere, V., Paterson, Y.Z., McClellan, A. et al. (2018) Osteoblast differentiation of equine induced pluripotent stem cells. *Biology Open*, 7, bio033514.
- Ballouz, S., Verleyen, W. & Gillis, J. (2015) Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics*, 31, 2123–2130.
- Bavin, E.P., Smith, O., Baird, A.E., Smith, L.C. & Guest, D.J. (2015) Equine induced pluripotent stem cells have a reduced tendon differentiation capacity compared to embryonic stem cells. *Frontiers in Veterinary Science*, 2, 55.
- Beekhuis-Hoekstra, S.D., Watanabe, K., Werme, J., de Leeuw, C.A., Paliukhovich, I., Li, K.W. et al. (2021) Systematic assessment of variability in the proteome of iPSC derivatives. *Stem Cell Research*, 56, 102512.
- Blott, S.C., Swinburne, J.E., Sibbons, C., Fox-Clipsham, L.Y., Helwegen, M., Hillyer, L. et al. (2014) A genome-wide association study demonstrates significant genetic variation for fracture risk in thoroughbred racehorses. *BMC Genomics*, 15, 147.
- Bourhis, J.-M., Mariano, N., Zhao, Y., Harlos, K., Exposito, J.-Y., Jones, E.Y. et al. (2012) Structural basis of fibrillar collagen trimerization and related genetic disorders. *Nature Structural & Molecular Biology*, 19, 1031–1036.
- Carpenter, E.M., Goddard, J.M., Davis, A.P., Nguyen, T.P. & Capocchi, M.R. (1997) Targeted disruption of Hoxd-10 affects mouse hindlimb development. *Development*, 124, 4505–4514.
- Chen, I.P. (2020) Differentiation of human induced pluripotent stem cells (hiPSCs) into osteoclasts. *Bio-Protocol*, 10, e3854.
- Chen, M.-Q. & Luan, J.-J. (2019) HMGB1 promotes bone fracture healing through activation of ERK signaling pathway in a rat tibial fracture model. *The Kaohsiung Journal of Medical Sciences*, 35, 550–558.
- Chen, X., Chen, J., Xu, D., Zhao, S., Song, H. & Peng, Y. (2017) Effects of osteoglycin (OGN) on treating senile osteoporosis by regulating MSCs. *BMC Musculoskeletal Disorders*, 18, 423.
- Chen, Y., Liu, Y., Zhang, Y., Yu, J. & Tang, L. (2022) Galectin-1 deletion in mice causes bone loss via impaired osteogenic differentiation potential of BMSCs. *The FASEB Journal*, 36, e22516.
- Cheng, S., Kesavan, C., Mohan, S., Qin, X., Alarcon, C.M., Wergedal, J. et al. (2013) Transgenic overexpression of ephrin B1 in bone cells promotes bone formation and an anabolic response to mechanical loading in mice. *PLoS One*, 8, e69051.
- Choi, S.W., Mak, T.S.-H. & O'Reilly, P.F. (2020) Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, 15, 2759–2772.
- Coleman, J.R.I. (2022) Feasibility and application of polygenic score analysis to the morphology of human-induced pluripotent stem cells. *Molecular Genetics and Genomics*, 297, 1111–1122.
- Collette, N.M., Yee, C.S., Hum, N.R., Murugesu, D.K., Christiansen, B.A., Xie, L. et al. (2016) Sostdcl deficiency accelerates fracture healing by promoting the expansion of periosteal mesenchymal stem cells. *Bone*, 88, 20–30.
- Crawford, R.R., Prescott, E.T. & Mungrue, I.N. (2016) Genetic inhibition of Chac1 leads to dysregulation of body composition. *The FASEB Journal*, 30, 717.3.
- Dobrindt, K., Zhang, H., Das, D., Abdollahi, S., Prorok, T., Ghosh, S. et al. (2021) Publicly available hiPSC lines with extreme polygenic risk scores for modeling schizophrenia. *Complex Psychiatry*, 6, 68–82.
- Drabek, K., van de Peppel, J., Eijken, M. & van Leeuwen, J.P. (2011) GPM6B regulates osteoblast function and induction of mineralization by controlling cytoskeleton and matrix vesicle release. *Journal of Bone and Mineral Research*, 26, 2045–2051.
- Fear, V.S., Forbes, C.A., Shaw, N.C., Farley, K.O., Mantegna, J.L., Htun, J.P. et al. (2023) Gene editing and cardiac disease modelling for the interpretation of genetic variants of uncertain significance in congenital heart disease. *Stem Cell Research & Therapy*, 14, 345.
- Francis, M.J.O., Lees, R.L., Trujillo, E., Martín-Vasallo, P., Heersche, J.N.M. & Mobasher, A. (2002) ATPase pumps in osteoclasts and osteoblasts. *The International Journal of Biochemistry & Cell Biology*, 34, 459–476.
- Georgopoulos, S.P. & Parkin, T.D. (2017) Risk factors for equine fractures in thoroughbred flat racing in North America. *Preventive Veterinary Medicine*, 139, 99–104.
- Gorashi, R., Rivera-Bolanos, N., Dang, C., Chai, C., Kovacs, B., Alharbi, S. et al. (2023) Modeling diabetic endothelial dysfunction with patient-specific induced pluripotent stem cells. *Bioengineering & Translational Medicine*, 8, e10592.
- Granéli, C., Thorfvé, A., Ruetschi, U., Brisby, H., Thomsen, P., Lindahl, A. et al. (2014) Novel markers of osteogenic and adipogenic differentiation of human bone marrow stromal cells identified using a quantitative proteomics approach. *Stem Cell Research*, 12, 153–165.
- Groza, T., Gomez, F.L., Mashhadi, H.H., Muñoz-Fuentes, V., Gunes, O., Wilson, R. et al. (2022) The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Research*, 51, D1038–D1045.
- Hoffman, G.E., Schrode, N., Flaherty, E. & Brennand, K.J. (2019) New considerations for hiPSC-based models of neuropsychiatric disorders. *Molecular Psychiatry*, 24, 49–66.
- Horvath, S. (2013) DNA methylation age of human tissues and cell types. *Genome Biology*, 14, R115.
- Iftikhar, M., Hurtado, P., Bais, M.V., Wigner, N., Stephens, D.N., Gerstenfeld, L.C. et al. (2011) Lysyl oxidase-like-2 (LOXL2) is a major isoform in chondrocytes and is critically required for differentiation. *Journal of Biological Chemistry*, 286, 909–918.

- Jiang, L., Sun, J. & Huang, D. (2022) Role of slit/Robo signaling pathway in bone metabolism. *International Journal of Biological Sciences*, 18, 1303–1312.
- Johnston, A.S., Riggs, C.M., Cogger, N., Benschop, J., Rogers, C.W. & Rosanowski, S.M. (2020) Using time-series analysis techniques to enhance the understanding of musculoskeletal injury in thoroughbred racehorses. *Equine Veterinary Journal*, 52, 699–708.
- Johnston, A.S., Sidhu, A.B.S., Riggs, C.M., Verheyen, K.L.P. & Rosanowski, S.M. (2020) The effect of stress fracture occurring within the first 12 months of training on subsequent race performance in thoroughbreds in Hong Kong. *Equine Veterinary Journal*, 53, 460–468.
- Kang, J.-H., Ko, H.-M., Han, G.-D., Lee, S.-Y., Moon, J.-S., Kim, M.-S. et al. (2020) Dual role of phosphatidylserine and its receptors in osteoclastogenesis. *Cell Death & Disease*, 11, 497.
- Karner, C.M. & Long, F. (2018) Glucose metabolism in bone. *Bone*, 115, 2–7.
- Kemp, J.P., Medina-Gomez, C., Estrada, K., St Pourcain, B., Heppe, D.H.M., Warrington, N.M. et al. (2014) Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. *PLoS Genetics*, 10, e1004423.
- Kemper, A.M., Drnevich, J., McCue, M.E. & McCoy, A.M. (2019) Differential gene expression in articular cartilage and subchondral bone of neonatal and adult horses. *Genes*, 10, 745.
- Kim, H.J., Prasad, V., Hyung, S.W., Lee, Z.H., Lee, S.W., Bhargava, A. et al. (2012) Plasma membrane calcium ATPase regulates bone mass by fine-tuning osteoclast differentiation and survival. *The Journal of Cell Biology*, 199, 1145–1158.
- Knowles, H.J., Cleton-Jansen, A.M., Korsching, E. & Athanasou, N.A. (2010) Hypoxia-inducible factor regulates osteoclast-mediated bone resorption: role of angiopoietin-like 4. *The FASEB Journal*, 24, 4648–4659.
- Kristoffersen, M., Parkin, T.D. & Singer, E.R. (2010) Catastrophic biaxial proximal sesamoid bone fractures in UK thoroughbred races (1994–2004): horse characteristics and racing history. *Equine Veterinary Journal*, 45, 420–424.
- Kuemmerle, J.M., Theiss, F., Okoniewski, M.J., Weber, F.A., Hemmi, S., Mirsaiidi, A. et al. (2016) Identification of novel equine (*Equus caballus*) tendon markers using RNA sequencing. *Genes*, 7, 97.
- Langfelder, P. & Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.
- Lee, E., Park, S.-Y., Moon, J.-Y., Ko, J.-Y., Kim, T.K. & Im, G.-I. (2022) Metabolic switch under glucose deprivation leading to discovery of NR2F1 as a stimulus of osteoblast differentiation. *Journal of Bone and Mineral Research*, 37, 1382–1399.
- Lewis, C.M. & Vassos, E. (2020) Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*, 12, 44.
- Li, C., Zheng, Z., Ha, P., Chen, X., Jiang, W., Sun, S. et al. (2018) Neurexin superfamily cell membrane receptor contactin-associated protein like-4 (Cntnap4) is involved in neural EGFL-like 1 (Nell-1)-responsive osteogenesis. *Journal of Bone and Mineral Research*, 33, 1813–1825.
- Li, J., Lu, D., Dou, H., Liu, H., Weaver, K., Wang, W. et al. (2018) Desumoylase SENP6 maintains osteochondroprogenitor homeostasis by suppressing the p53 pathway. *Nature Communications*, 9, 143.
- Li, W., Yuan, L., Tong, G., He, Y., Meng, Y., Hao, S. et al. (2018) Phospholipase C signaling activated by parathyroid hormone mediates the rapid osteoclastogenesis in the fracture healing of orchietomized mice. *BMC Musculoskeletal Disorders*, 19, 311.
- Liu, Q., Lei, L., Yu, T., Jiang, T. & Kang, Y. (2018) Effect of brain-derived neurotrophic factor on the neurogenesis and osteogenesis in bone engineering. *Tissue Engineering. Part A*, 24, 1283–1292.
- Livak, K.J. & Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, 25, 402–408. Available from: <https://doi.org/10.1006/meth.2001.1262>
- Love, M.I., Huber, W. & Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.
- Mae, S.I., Hattanda, F., Morita, H., Nozaki, A., Katagiri, N., Ogawa, H. et al. (2023) Human iPSC-derived renal collecting duct organoid model cystogenesis in ADPKD. *Cell Reports*, 42, 113431.
- Manickam, G., Moffatt, P. & Murshed, M. (2019) Role of SMPD3 during bone fracture healing and regulation of its expression. *Molecular and Cellular Biology*, 39, e00370-18.
- Mantila Roosa, S.M., Liu, Y. & Turner, C.H. (2011) Gene expression patterns in bone following mechanical loading. *Journal of Bone and Mineral Research*, 26, 100–112.
- Mao, L., Wang, L., Xu, J. & Zou, J. (2023) The role of integrin family in bone metabolism and tumor bone metastasis. *Cell Death Discovery*, 9, 119.
- Martineau, C., Najyb, O., Signor, C., Rassart, É. & Moreau, R. (2016) Apolipoprotein D deficiency is associated to high bone turnover, low bone mass and impaired osteoblastic function in aged female mice. *Metabolism*, 65, 1247–1258.
- McKee, S.L. (1995) An update on racing fatalities in the UK. *Equine Veterinary Education*, 7, 202–204.
- Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J. et al. (2003) PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics*, 34, 267–273.
- Mourskaia, A.A., Amir, E., Dong, Z., Tiedemann, K., Cory, S., Omeroglu, A. et al. (2012) ABCC5 supports osteoclast formation and promotes breast cancer metastasis to bone. *Breast Cancer Research*, 14, R149.
- Muller, J., Duray, E., Lejeune, M., Dubois, S., Plougouven, E., Léonard, A. et al. (2019) Loss of stromal galectin-1 enhances multiple myeloma development: emphasis on a role in osteoclasts. *Cancers (Basel)*, 11, 261.
- Nguyen, T.M., Arthur, A., Paton, S., Hemming, S., Panagopoulos, R., Codrington, J. et al. (2016) Loss of ephrinB1 in osteogenic progenitor cells impedes endochondral ossification and compromises bone strength integrity during skeletal development. *Bone*, 93, 12–21.
- Ogata, Y., Kukita, A., Kukita, T., Komine, M., Miyahara, A., Miyazaki, S. et al. (1999) A novel role of IL-15 in the development of osteoclasts: inability to replace its activity with IL-2. *Journal of Immunology*, 162, 2754–2760.
- Orriss, I.R., Wang, N., Burnstock, G., Arnett, T.R., Gartland, A., Robaye, B. et al. (2011) The P2Y(6) receptor stimulates bone resorption by osteoclasts. *Endocrinology*, 152, 3706–3716.
- Otto, E., Knapstein, P.-R., Jahn, D., Appelt, J., Frosch, K.-H., Tsitsilonis, S. et al. (2020) Crosstalk of brain and bone—clinical observations and their molecular bases. *International Journal of Molecular Sciences*, 21, 4946.
- Palomino Lago, E., Baird, A., Blott, S.C., McPhail, R.E., Ross, A.C., Durward-Akhurst, S.A. et al. (2024) A functional single-nucleotide polymorphism upstream of the collagen type III gene is associated with catastrophic fracture risk in thoroughbred horses. *Animals*, 14, 116.
- Palomino Lago, E., Jelbert, E.R., Baird, A., Lam, P.Y. & Guest, D.J. (2023) Equine induced pluripotent stem cells are responsive to inflammatory cytokines before and after differentiation into musculoskeletal cell types. *In Vitro Cellular & Developmental Biology—Animal*, 59, 514–527.
- Parkin, T.D., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R. et al. (2004a) Horse-level risk factors for fatal distal limb fractures in racing thoroughbreds in the UK. *Equine Veterinary Journal*, 36, 513–519.
- Parkin, T.D., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R. et al. (2004b) Race- and course-level risk factors

- for fatal distal limb fracture in racing thoroughbreds. *Equine Veterinary Journal*, 36, 521–526.
- Parkin, T.D., Clegg, P.D., French, N.P., Proudman, C.J., Riggs, C.M., Singer, E.R. et al. (2006) Catastrophic fracture of the lateral condyle of the third metacarpus/metatarsus in UK racehorses – fracture descriptions and pre-existing pathology. *The Veterinary Journal*, 171, 157–165.
- Paterson, Y.Z., Cribbs, A., Espenel, M., Smith, E.J., Henson, F.M.D. & Guest, D.J. (2020) Genome-wide transcriptome analysis reveals equine embryonic stem cell-derived tenocytes resemble fetal, not adult tenocytes. *Stem Cell Research & Therapy*, 11, 184.
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14, 417–419.
- Pereira, I., Lopez-Martinez, M.J. & Samitier, J. (2023) Advances in current in vitro models on neurodegenerative diseases. *Frontiers in Bioengineering and Biotechnology*, 11, 1260397.
- Qian, G.F., Yuan, L.S., Chen, M., Ye, D., Chen, G.P., Zhang, Z. et al. (2019) PPWD1 is associated with the occurrence of postmenopausal osteoporosis as determined by weighted gene co-expression network analysis. *Molecular Medicine Reports*, 20, 3202–3214.
- Raft, S., Coate, T.M., Kelley, M.W., Crenshaw, E.B., 3rd & Wu, D.K. (2014) Pou3f4-mediated regulation of ephrin-b2 controls temporal bone development in the mouse. *PLoS One*, 9, e109043.
- Riggs, C.M., Whitehouse, G.H. & Boyde, A. (1999) Pathology of the distal condyles of the third metacarpal and third metatarsal bones of the horse. *Equine Veterinary Journal*, 31, 140–148.
- Rolvien, T., Yorgan, T.A., Kornak, U., Hermans-Borgmeyer, I., Mundlos, S., Schmidt, T. et al. (2020) Skeletal deterioration in COL2A1-related spondyloepiphyseal dysplasia occurs prior to osteoarthritis. *Osteoarthritis and Cartilage*, 28, 334–343.
- Rosanowski, S.M., Chang, Y.-M., Stirk, A.J. & Verheyen, K.L.P. (2018) Risk factors for race-day fatality in flat racing thoroughbreds in Great Britain (2000 to 2013). *PLoS One*, 13, e0194299.
- Rosset, E.M. & Bradshaw, A.D. (2016) SPARC/osteonectin in mineralized tissue. *Matrix Biology*, 52–54, 78–87.
- Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C. et al. (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 50, D20–D26.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2498–2504.
- Smid, M., Coebergh van den Braak, R.R.J., van de Werken, H.J.G., van Riet, J., van Galen, A., de Weerd, V. et al. (2018) Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics*, 19, 236.
- Son, A., Kang, N., Kang, J.Y., Kim, K.W., Yang, Y.M. & Shin, D.M. (2018) TRPM3/TRPV4 regulates Ca²⁺-mediated RANKL/NFATc1 expression in osteoblasts. *Journal of Molecular Endocrinology*, 61, 207–218.
- Soneson, C., Love, M.I. & Robinson, M.D. (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, 4, 1521.
- Stover, S.M. (2003) The epidemiology of thoroughbred racehorse injuries. *Clinical Techniques in Equine Practice*, 2, 312–322.
- Stover, S.M. (2013) Diagnostic workup of upper-limb stress fractures and proximal sesamoid bone stress remodeling. *AAEP Proceedings; In-Depth: Racing-Related Lameness*, 59, 427–435.
- Stover, S.M., Johnson, B.J., Daft, B.M., Read, D.H., Anderson, M., Barr, B.C. et al. (1992) An association between complete and incomplete stress fractures of the humerus in racehorses. *Equine Veterinary Journal*, 24, 260–263.
- Stuepp, R.T., Modolo, F., Trentin, A.G., Garcez, R.C. & Biz, M.T. (2020) HNK1 and Sox10 are present during repair of mandibular bone defects. *Biotechnic & Histochemistry*, 95, 619–625.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 15545–15550.
- Tong, X., Zhu, C., Liu, L., Huang, M., Xu, J., Chen, X. et al. (2022) Role of Sostdcl in skeletal biology and cancer. *Frontiers in Physiology*, 13, 1029646.
- Tozaki, T., Kusano, K., Ishikawa, Y., Kushiro, A., Nomura, M., Kikuchi, M. et al. (2020) A candidate-SNP retrospective cohort study for fracture risk in Japanese thoroughbred racehorses. *Animal Genetics*, 51, 43–50.
- Trajanoska, K., Morris, J.A., Oei, L., Zheng, H.-F., Evans, D.M., Kiel, D.P. et al. (2018) Assessment of the genetic and clinical determinants of fracture risk: genome wide association and Mendelian randomisation study. *BMJ*, 362, k3225.
- Tuerlings, M., van Hoolwerff, M., Houtman, E., Suchiman, E.H.E.D., Lakenberg, N., Mei, H. et al. (2021) RNA sequencing reveals interacting key determinants of osteoarthritis acting in subchondral bone and articular cartilage: identification of IL11 and CHADL as attractive treatment targets. *Arthritis and Rheumatology*, 73, 789–799.
- Uehara, N., Kukita, A., Kyumoto-Nakamura, Y., Yamaza, T., Yasuda, H. & Kukita, T. (2017) Osteoblast-derived Laminin-332 is a novel negative regulator of osteoclastogenesis in bone micro-environments. *Laboratory Investigation*, 97, 1235–1244.
- Urabe, K., Jingushi, S., Ikenoue, T., Okazaki, K., Sakai, H., Li, C. et al. (2001) Immature osteoblastic cells express the pro-alpha2(XI) collagen gene during bone formation in vitro and in vivo. *Journal of Orthopaedic Research*, 19, 1013–1020.
- Valderrábano, R.J. & Linares, M.I. (2018) Diabetes mellitus and bone health: epidemiology, etiology and implications for fracture risk stratification. *Clinical Diabetes and Endocrinology*, 4, 9.
- Verheyen, K., Price, J., Lanyon, L. & Wood, J.L.N. (2006) Exercise distance and speed affect the risk of fracture in racehorses. *Bone*, 39, 1322–1330.
- Walter, K., Tansek, M., Tobias, E.S., Ikegawa, S., Coucke, P., Hyland, J. et al. (2007) COL2A1-related skeletal dysplasias with predominant metaphyseal involvement. *American Journal of Medical Genetics. Part A*, 143a, 161–167.
- Welsh, C.E., Lewis, T.W., Blott, S.C., Mellor, D.J., Stirk, A.J. & Parkin, T.D. (2014) Estimates of genetic parameters of distal limb fracture and superficial digital flexor tendon injury in UK thoroughbred racehorses. *Veterinary Journal*, 200, 253–256.
- Wilson, S.S., Wong, A., Toupadakis, C.A. & Yellowley, C.E. (2015) Expression of angiopoietin-like protein 4 at the fracture site: regulation by hypoxia and osteoblastic differentiation. *Journal of Orthopaedic Research*, 33, 1364–1373.
- Xing, W., Kim, J., Wergedal, J., Chen, S.-T. & Mohan, S. (2010) Ephrin B1 regulates bone marrow stromal cell differentiation and bone formation by influencing TAZ transactivation via complex formation with NHERF1. *Molecular and Cellular Biology*, 30, 711–721.
- Yde Ohki, C.M., Walter, N.M., Bender, A., Rickli, M., Ruhstaller, S., Walitza, S. et al. (2023) Growth rates of human induced pluripotent stem cells and neural stem cells from attention-deficit hyperactivity disorder patients: a preliminary study. *Journal of Neural Transmission (Vienna)*, 130, 243–252.
- Yin, X.H., Wang, X.Y., Liu, S.C., Yan, L., He, B.R., Hao, D.J. et al. (2022) lncRNA HHIP-AS1/HHIP modulates osteogenic differentiation of BM-MSCs by regulating hedgehog signaling pathway. *Aging (Albany NY)*, 14, 8839–8855.
- Zhang, W., Zhu, Y., Chen, J., Wang, J., Yao, C. & Chen, C. (2020) Mechanisms of miR-128-3p in inhibiting osteoblast

differentiation from bone marrow-derived mesenchymal stromal cells. *Molecular Medicine Reports*, 22, 5041–5052.

Zhao, X., Guo, W., Zou, L. & Hu, B. (2020) FBXO2 modulates STAT3 signaling to regulate proliferation and tumorigenicity of osteosarcoma cells. *Cancer Cell International*, 20, 245.

Zhu, N., Wang, H., Wang, B., Wei, J., Shan, W., Feng, J. et al. (2016) A member of the nuclear receptor superfamily, designated as NR2F2, supports the self-renewal capacity and pluripotency of human bone marrow-derived mesenchymal stem cells. *Stem Cells International*, 2016, 5687589.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Palomino Lago, E., Ross, A.K.C., McClellan, A. & Guest, D.J. (2025) Identification of a global gene expression signature associated with the genetic risk of catastrophic fracture in iPSC-derived osteoblasts from Thoroughbred horses. *Animal Genetics*, 56, e13504. Available from: <https://doi.org/10.1111/age.13504>