ER procollagen storage defect without coupled unfolded protein response drives precocious arthritis

Running title: Proteostasis defects driving collagenopathies

Kathryn M. Yammine,[a] Sophia Mirda Abularach,[a] Seo-yeon Kim,[a] Agata A. Bikovtseva,[a] Jinia Lilianty,[b,c] Vincent L. Butty,[d] Richard P. Schiavoni,[e] John F. Bateman,[b,c] Shireen R. Lamandé,[b,c,f] Matthew D. Shoulders[a,e,g,]*

Affiliations:

[a]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, United States

[b]Murdoch Children's Research Institute, Parkville, Victoria, Australia

[c]Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia

[d]BioMicro Center, Massachusetts Institute of Technology, Cambridge, MA, United States

[e]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, United States

[f]The Novo Nordisk Foundation Center for Stem Cell Medicine (reNEW), Murdoch Children's Research Institute, Parkville, Victoria, Australia

[g]Broad Institute of MIT and Harvard, Cambridge, MA, United States

*To Whom Correspondence May Be Addressed:

Matthew D. Shoulders

Department of Chemistry

Massachusetts Institute of Technology

77 Massachusetts Avenue, 16-573A

Cambridge, MA 02139

Tel: (617) 452-3525

Email: mshoulde@mit.edu

Abstract

Collagenopathies are a group of clinically diverse disorders caused by defects in collagen folding and secretion. For example, mutations in the gene encoding collagen type-II, the primary collagen in cartilage, can lead to diverse chondrodysplasias. One example is the Gly1170Ser substitution in procollagen-II, which causes precocious osteoarthritis. Here, we biochemically and mechanistically characterize an induced pluripotent stem cell-based cartilage model of this disease, including both hetero- and homozygous genotypes. We show that Gly1170Ser procollagen-II is notably slow to fold and secrete. Instead, procollagen-II accumulates intracellularly, consistent with an endoplasmic reticulum (ER) storage disorder. Owing to unique features of the collagen triple helix, this accumulation is not recognized by the unfolded protein response. Gly1170Ser procollagen-II interacts to a greater extent than wild-type with specific proteostasis network components, consistent with its slow folding. These findings provide mechanistic elucidation into the etiology of this disease. Moreover, the cartilage model will enable rapid testing of therapeutic strategies to restore proteostasis in the collagenopathies.

Introduction

Collagens are the most abundant proteins in the human body and the molecular scaffold of animal life (Shoulders and Raines, 2009). Collagenopathies are a collection of genetic diseases caused by defects in the synthesis, folding, quality control, assembly, and/or function of members of the collagen protein family (Bateman et al., 2009). Autosomal dominant mutations in the 44 human genes that encode the 28 different types of collagen are the most common cause of collagenopathies, with mutations in 29 of these collagen genes already associated with disease (Hamosh et al., 2005; Jobling et al., 2014). As such, collagenopathies present with clinically diverse phenotypes that manifest most strongly in tissue. Currently, they have no cure. Moreover, effective treatments for the vast majority of collagenopathies are nearly universally lacking (Bateman et al., 2009; Wong and Shoulders, 2019).

The fibrillar collagen type-II (collagen-II) is the main proteinaceous component of cartilage, and is thus a central player in both joint function and endochondral bone formation (Eyre, 2002; Horton and Machado, 1988). Mutations in the gene encoding collagen-II (*COL2A1*) often lead to diseases termed type-II collagenopathies, or chondrodysplasias (Gregersen and Savarirayan, 1993; Spranger et al., 1994). These disorders encompass a broad range of clinical phenotypes, including skeletal dysplasia, ocular symptoms, and hearing impairment. The spectrum of severity ranges from relatively mild forms of premature arthritis to perinatal lethality. Type-II collagenopathies are overwhelmingly caused by mutations specifically within the procollagen-II triple helix-encoding portion of the *COL2A1* gene (UniProt, 2023). This 1014 amino acid-long domain is composed of Xaa-Yaa-Gly repeats, and is bookended by two smaller globular domains that are cleaved post-secretion (Bauer and Lanschuetzer, 2003). Mutations that encode substitutions of the conserved Gly residues found at every third position in the triple-helical domain are the most common cause of disease (UniProt, 2023).

One exemplary type-II collagenopathy is the chondrodysplasia caused by the autosomal-dominant c.3508 GGT > AGT mutation in exon 50 of *COL2A1,* leading to a p.Gly1170Ser substitution in the C-terminal region of

the procollagen-II triple-helical domain (**Fig. 1 A**) (Liu et al., 2005; Su et al., 2008). Patients harboring this mutation present with precocious osteoarthritis and Legg-Calvé-Perthes disease (a juvenile form of avascular necrosis of the femoral head), with symptoms reported as early as the first decade of life, and most often diagnosed in adulthood (Liu et al., 2005; Miyamoto et al., 2007; Su et al., 2008; Wang et al., 2014; Zhang et al., 2021). The disease is typically diagnosed via genetic screening engendered by the familial link. Treatment options are currently extremely limited, with total joint replacement as the standard-of-care.

A main challenge with studying collagenopathies in general, and type-II collagenopathies in particular, is the paucity of suitable systems to model these disorders. The disease phenotypes are most strongly manifested in tissue, making classical cell culture models insufficient. This challenge is especially acute for collagen-II and cartilage, as primary chondrocytes rapidly de-differentiate in classical culture conditions (Schnabel et al., 2002), and immortalized chondrocytes generally do not deposit an extensive extracellular matrix (ECM) (Goldring and Berenbaum, 1999). Moreover, mouse models of collagenopathies often fail to recapitulate human phenotypes, with heterozygous mice often showing little to no skeletal phenotypes, and homozygous mice presenting with much more severe pathology than that observed in patients.

Excitingly, recent advances in differentiating induced pluripotent stem cells (iPSC) towards chondrogenic lineages hold considerable promise for addressing this fundamental challenge, allowing us to begin to model type-II collagenopathies *in vitro*. Several protocols now exist for differentiating human pluripotent stem cells (hPSC) to the chondrogenic lineage through various growth factor treatments (Adkar et al., 2019; Craft et al., 2015; Koyama et al., 2013; Lamandé et al., 2023; Lee et al., 2015; Okada et al., 2015; Oldershaw et al., 2010). The inherent ability of stem cells to proliferate makes these systems highly expandable and thus much more amenable for biochemistry and drug discovery than mouse models. Methods for producing cartilage *in vitro* in the correct developmental context, combined with gene editing technology to create genetically matched cell lines incorporating disease-causing mutations (Hosseini Far et al., 2019; Howden et al., 2019; Kung et al., 2020; Lilianty et al., 2021; Lilianty et al., 2020; Yammine et al., 2023), provide an exceedingly promising

avenue for modeling, studying, and discovering potential therapies for genetic skeletal disorders, including the type-II collagenopathies.

Results

Cells expressing Gly1170Ser-substituted collagen-II deposit a cartilaginous ECM

To elucidate the etiology of disease caused by the Gly1170Ser substitution in collagen-II, in prior work we created and characterized genetically matched iPSC lines harboring the Gly1170Ser mutation in *COL2A1* in either one or both alleles using CRISPR/Cas9 gene editing (**Fig. 1 A**, and **Figs. S 1** and **S 2**) (Kung et al., 2020). These lines, along with the isogenic wild-type control, allow us to unambiguously assess the effect of the Gly1170Ser on collagen-II proteostasis – meaning its synthesis, folding, localization, and function – and how it affects the cells that express it and the matrix they deposit. We began by differentiating all three iPSC lines (wild-type, heterozygous Gly1170Ser, homozygous Gly1170Ser) to chondrocytes using the protocol recently described by Lamandé *et al*. (Lamandé et al., 2023). Briefly, we expanded iPSCs in feeder-free monolayer culture, then guided them through differentiation to sclerotome by growth factor treatment over the course of 4 days. To make individual spheroids, we transferred the cells to pellet culture for two weeks and finally to swirler culture, which promotes chondrocyte differentiation and deposition of cartilage ECM components (**Fig. 1 B**) (Lamandé et al., 2023). We repeated differentiations in three independent differentiations. After 44 days in culture, we harvested the chondronoids from all replicates and processed them for downstream assays.

To evaluate successful differentiation to chondrocytes, we interrogated differentiation-induced changes in the transcriptional profile of chondronoids after 44 days in culture using RNA-sequencing. The gene expression profile for all genotypes was consistent with that of chondrocytes, as evidenced by high expression of cartilage-specific markers such as collagens (*COL2A1*, *COL11A1*, *COL9A1*) and aggrecan (*ACAN*), but not of collagens associated with other tissue types, including *COL1A1*, the major component of bone (**Fig. 2 A**). Indeed, cartilage components were the most highly expressed core matrisome genes (Hynes and Naba, 2012).

To assess the bulk composition of the ECM deposited by chondrocytes expressing the different procollagen-II variants, we first looked for the most abundant cartilage ECM components, collagen-II and proteoglycans

(Eyre, 2002), in histology sections. We observed that the wild-type, clinically relevant heterozygote, and the more severe homozygote mutant abundantly deposited collagen-II in the ECM (**Fig. 2 B,** in green). Similarly, toluidine blue staining for proteoglycans revealed that, consistent with the wild-type, the matrices of the heterozygous and homozygous Gly1170Ser variants were also rich in proteoglycans (**Fig. 2 B,** in purple). There was no significant difference in the size of the chondronoids between the wild-type and clinically relevant Gly1170Ser heterozygote. However, chondronoids expressing homozygous Gly1170Ser collagen-II were smaller than the wild-type controls (**Fig. 2 C**), hinting at a possible defect in secretion of ECM components. Nevertheless, the abundant presence of collagen-II and proteoglycans indicates that, from a bulk composition perspective at least, chondrocytes expressing Gly1170Ser-subsituted procollagen-II deposit a cartilaginous ECM.

The matrix of disease variant chondronoids is deficient

We hypothesized that, despite the deposition of an ECM rich in collagen-II and proteoglycans, and given the osteoarthritic and avascular necrotic phenotype of patients, (Liu et al., 2005; Miyamoto et al., 2007; Su et al., 2008) expression of the disease-causing Gly1170Ser variant could lead to defects in the deposited ECM that were not apparent in the immunohistochemistry.

To assess the quality of the deposited matrix, we used transmission electron microscopy (TEM) to examine the chondronoid matrix at higher magnification and resolution. In the wild-type chondronoids, we observed an extensive network of connecting fibrils (**Fig. 3**), consistent with *in vivo* cartilage samples (Eyre, 2002). TEM imaging of the heterozygous Gly1170Ser chondronoids, in contrast, revealed a comparatively defective matrix (**Fig. 3**). There were fewer collagen-II fibrils in the ECM, leading overall to a deficient network. This reduced matrix could be caused by insufficient or improperly folded collagen-II being secreted from cells and deposited in the ECM, by defective crosslinking of the collagen that is secreted into the ECM, or by a combination of both. This phenotype observed in the clinically relevant heterozygous Gly1170Ser variant could help to explain the joint symptoms of patients (Liu et al., 2005; Miyamoto et al., 2007; Su et al., 2008). Such a deficient matrix

would likely lead to a defective coating of articular cartilage at joints, leaving the joint as a whole more susceptible to injury or the stress of repeated loading (Peters et al., 2018).

When compared to the wild-type chondronoids, and even the heterozygotes, homozygous Gly1170Ser matrices showed striking deficiencies via TEM (**Fig. 3**). The fibrils in chondronoids from this genotype were generally shorter, yielding a very sparse network. A healthy cartilage matrix is required not only for joint health, but also for proper skeletal development, as it is the tissue that precedes bone during the process of endochondral ossification.

Gly1170Ser-substituted collagen-II is slow to fold, a defect that is exacerbated in homozygotes

Since expression of Gly1170Ser-substituted collagen-II leads to a deficient cartilage matrix, we next sought to elucidate the underlying cause(s) of this deficiency. We hypothesized that the Gly→Ser substitution causes a delay or defect in the proper folding and assembly of the procollagen-II triple helix (Bateman et al., 2009; Wong and Shoulders, 2019), which in turn could propagate to slow secretion of the protein into the ECM.

As procollagen-II folds in the ER, it interacts with modifying enzymes that install post-translational modifications on the individual polypeptide chains, such as hydroxyl groups on proline and lysine residues, as well as *O*-glycans on hydroxylysine residues (Shoulders and Raines, 2009; Yamauchi and Sricholpech, 2012). Slower folding of the polypeptide chains leads to hyper accumulation of these modifications. Thus, a key signature of slow-folding collagen variants is reduced electrophoretic mobility on SDS-PAGE (Bateman et al., 1986; Bateman et al., 1984; Bonadio et al., 1985; Cabral et al., 2014; Godfrey and Hollister, 1988) with *O*-glycosylation of the excessively hydroxylated lysine residues contributing to the majority of the shift in molecular weight.

To test the hypothesis that the Gly1170Ser pathologic substitution causes a delay in collagen-II folding, we extracted collagen from the tissues by flash-freezing chondronoids and homogenizing the tissue, followed by a

pepsin digestion to extract the pepsin-resistant collagen fraction. After separation via SDS-PAGE, we observed that the heterozygous, pathological Gly1170Ser substitution did indeed lead to reduced electrophoretic mobility, indicating that the substituted protein was excessively post-translationally modified and, therefore, slow to fold within cells when compared to the wild-type protein (**Figs. 4** and **S 3**). The presence of a single broad, slow-migrating band as opposed to distinctive overmodified mutant versus normally modified wild-type strands is due to fact that the vast majority of trimers formed in heterozygotes (>85%) contain at least one Gly1170Ser strand that delays triple-helix folding.

In the homozygote, where all three procollagen-α1(II) chains contain the Gly1170Ser substitution, we observed that the delay in folding was further exacerbated (**Fig. 4**). That is, the homozygous Gly1170Ser collagen-II displayed even further reduced electrophoretic mobility on SDS-PAGE compared to the heterozygous variant.

Gly1170Ser-substituted procollagen-II is intracellularly retained and causes ER dilation

The deficient ECM combined with the slow folding kinetics strongly suggests that Gly1170Ser-substituted procollagen-II presents a challenge for cells to fold and secrete. We next sought to determine the fate of this variant in cells. We hypothesized that the slow folding of Gly1170Ser procollagen-II would lead to accumulation of the protein within cells and, specifically, within the ER, where this secreted protein folds. Intracellular retention is commonly observed in other diseases that impact protein folding and assembly (Teckman and Perlmutter, 2000).

To test this hypothesis, we visualized procollagen-II inside the chondrocytes. We observed that a significantly larger proportion of Gly1170Ser-expressing chondrocytes retained procollagen-II intracellularly as compared to the wild-type, as evidenced by collagen-II immunohistochemistry (IHC) (**Fig. 5 A**). We performed three independent, blinded scorings of multiple IHC images to quantify the fraction of cells with intracellularly retained procollagen-II in each genotype. In the clinically relevant heterozygous variant, 64% of cells showed signs of intracellular procollagen-II retention, as compared to just 34% of cells expressing wild-type

procollagen-II (**Fig. 5 B**). In the more severe homozygous variant, >80% of cells exhibited intracellular procollagen-II retention (**Fig. 5 B**).

To better localize the intracellular protein, we co-stained with antibodies against collagen-II and an ER marker, calreticulin. We found that much of the intracellularly retained procollagen-II co-localized with calreticulin, suggesting that a portion of the protein is retained within the ER. These observations, along with the slow folding kinetics (**Fig. 4**), indicate that this pathologic variant is challenging for cells to properly fold and secrete, presumably because the sterically bulky residue in place of the much smaller Gly residue hinders triple helix assembly and disrupts a key stabilizing interchain hydrogen bond within the helix (Shoulders and Raines, 2009). Instead, procollagen-II begins to strongly accumulate within the ER.

We next sought to ascertain whether this intracellular retention dysregulates ER homeostasis, thereby exacerbating pathology. In other protein folding disorders, including other collagenopathies, intracellular retention of misfolded or aggregated protein is often accompanied by dilation of the ER (Teckman and Perlmutter, 2000). Indeed, high magnification TEM revealed that expression of the pathologic Gly1170Ser variant did indeed lead to dilated ER in the chondrocytes (**Fig. 6**), a phenotype that was exacerbated in the homozygote. Consistent with ER storage diseases, this pathologic ER distention confirms the Gly1170Ser substitution leads to defects in procollagen-II proteostasis, which in patients would eventually propagate to tissue-level defects in cartilage and pathologic joints.

Cells do not robustly respond at the transcript level to the Gly1170Ser-substituted collagen-II ER storage defect

Our data indicate that Gly1170Ser-substituted procollagen-II is slow to fold, accumulates in the ER, and exhibits a secretion defect. That secretion defect produces pathologic ER phenotypes, including extensive dilation. We next asked whether this protein folding challenge was accompanied by a transcriptional stress response. The unfolded protein response (UPR) is generally responsible for maintaining ER proteostasis by

increasing levels of ER chaperones, quality control factors, and secretion machineries through signaling mediated by the transcription factors XBP1s and ATF6(f), and reducing non-essential translation via the PERK pathway (Walter and Ron, 2011). Upon sustained and unresolved UPR signaling, cellular apoptosis can also be induced by the PERK signaling pathway (Wong et al., 2018).

We tested for upregulation of UPR target genes (Phillips et al., 2018) in Gly1170Ser-expressing chondronoids versus wild-type using RNA-sequencing of the different genotypes at two different timepoints during chondronoid growth: day 34 and day 44 (**Fig. 7**). These timepoints were selected to reflect an early and a late stage of cartilage maturation, but with both timepoints harvested post-chondrogenesis so as not to interfere with the physiologic transient UPR activation that can be important in that process (Horiuchi et al., 2016). Notably, we observed that the UPR was not detectably induced in the clinically relevant heterozygote line at either timepoint. For the more severe homozygote line, there was no sign of UPR activation at the later timepoint. At the early timepoint, there was a very modest upregulation of UPR genes, which we discuss later in this section.

We next sought to determine if chondrocytes were instead raising any other organized transcriptional stress response when expressing Gly1170Ser procollagen. We looked for the non-canonical stress response to procollagen-I misfolding first described by Mirigian *et al.* (Gorrell et al., 2022; Mirigian et al., 2016), but we did not detect any significant upregulation of genes associated with this response (**Fig. 7**). We also scrutinized activation of the integrated stress response (ISR) (Gorrell et al., 2022; Grandjean et al., 2019) but observed no robust activation of this pathway for the clinically relevant heterozygote. There was, however, modest upregulation of this response in the more severe but clinically irrelevant homozygote variant specifically at the later timepoint. Similarly, we investigated genes related to oxidative stress and inflammation, but again did not observe any significant differences in expression levels in the Gly1170Ser-expressing cells compared to wild-type, at either the early or late timepoints (**Fig. 7**). These observations indicate that the chondrocytes were not raising such stress responses, at least when examined in the absence of paracrine signals from other cell types in the joint.

We next asked whether the challenge of folding Gly1170Ser procollagen-II in the ER causes apoptosis, despite no apparent induction of stress signaling at the transcriptional level. We performed a TUNEL assay to assess the number of apoptotic cells in chondronoids of the various genotypes. Consistent with our transcriptional data indicating a lack of any chronic UPR that could be pro-apoptotic, we did not observe an increase in apoptosis in the heterozygotes or the homozygotes (**Figs. 8** and **S 4**).

We then turned to gene set enrichment analysis (GSEA) as a more global approach to identify differentially expressed pathways between Gly1170Ser-expressing lines and the wild-type (**Figs. S 5** and **S 6**). GSEA similarly did not identify any differentially expressed stress pathways. It did, however, highlight subtle differences in a few cellular processes. For the clinically relevant heterozygote, the main differences were upregulation of gene sets related to cellular respiration, as well as a downregulation of gene sets associated with translation and mitosis at the late timepoint when compared to the wild-type. These differences could suggest a UPR-independent reshaping of energetic allocations for cellular functions as an adaptive response to the accumulation of slow-folding procollagen-II. Moreover, the downregulation of genes involved in translation could be related to an attempt to alleviate the protein folding challenge in the ER, though the lack of apparent UPR/ISR activation makes it unclear how this response is initiated. Leiken and co-workers have described a non-canonical response (Gorrell et al., 2022; Mirigian et al., 2016), but as noted above that response is not observed in our transcription data.

On the other hand, for the homozygote line, GSEA revealed upregulation of genes associated with collagen folding and extracellular matrix deposition (**Fig. S 6**). The implications of the observed differences in ECM-related processes will be interesting to follow-up on in future studies, albeit bearing in mind the reduced clinical relevance of this genotype. We also note that, possibly because cartilage-related pathways are generally poorly annotated, with few, if any, comprehensive gene sets, no cartilage-specific gene sets were identified as being differentially expressed in the Gly1170Ser-expressing cells. The most substantial transcriptional difference we observed, again present exclusively in the more severe and clinically irrelevant homozygous

variant, was downregulation of genes associated with chondrocyte hypertrophy (Lamandé et al., 2023), indicating that the homozygous chondrocytes were slower to mature when compared to wild-type (**Fig. 7**). Chondrocyte hypertrophy, characterized by enlarged chondrocytes expressing collagen-X, is a key step in long bone formation, when the cartilaginous, collagen-II rich matrix shifts to a mineralized, collagen-I bone matrix in a process known as endochondral ossification (Horton and Machado, 1988). The upregulation in the homozygous system of Matrix Gla Protein (MGP) (**Fig. 2 A**), which inhibits vascular calcification of the matrix *in vivo*, further supports the delay in hypertrophy, and could lead to differences in the biomechanical properties of the matrix. Such slow maturation could also explain the very modest upregulation of UPR targets observed exclusively at the early timepoint in the homozygous variant, as transient UPR signaling has been implicated in chondrocyte proliferation, differentiation, and hypertrophy (Horiuchi et al., 2016; Wang et al., 2009; Yuan et al., 2017).

Interestingly, the clinically relevant heterozygote showed the converse trend in hypertrophic gene expression, with certain genes for hypertrophy being upregulated compared to the wild-type ─ a sign of accelerated maturation ─ although this effect was more muted (**Fig. 7**). These differences in maturation speed likely have important implications for the overall functionality of the skeletal ECM deposited (Bateman et al., 2009). Further studies on the effect of maturation rate (specifically in the case of accelerated hypertrophy) are needed to better understand its role in pathogenesis of disease.

Quantitative comparative interactome analysis of Gly1170Ser procollagen-II variant reveals increased interactions with certain ER enzymes and folding clients

Given the absence of robust, well-organized transcriptional responses, we next asked whether the cellular proteostasis machinery can recognize Gly1170Ser-induced procollagen-II misfolding. We used quantitative mass spectrometry (MS) to elucidate and compare the interactomes of wild-type versus Gly1170Ser-substituted procollagen-II (**Fig. S 7**) (DiChiara et al., 2016). For this experiment, we transiently overexpressed epitope-tagged procollagen-II in HT-1080 cells to enable MS-grade immunoprecipitations. This cell line was

selected because it expresses the cellular machinery necessary for folding procollagen, but does not express endogenous procollagen-II which could confound our findings (Geddis and Prockop, 1993). **Figure S 7** outlines the experimental design and workflow.

Using tandem mass tag (TMT)-based MS, we quantitatively compared the interactomes of wild-type and Gly1170Ser procollagen-II to identify proteins that differentially interact with the bait (DiChiara et al., 2016; Doan et al., 2019; Doan et al., 2020; Plate et al., 2019). We identified multiple high confidence procollagen-II interactors (**Table 1**). After normalizing to bait, the fold interaction enrichment was calculated as $Fold\ Enrichment = \frac{\Sigma\, abundance\ in\ Gly1170Ser\ samples}{\Sigma\, abundance\ in\ wild-type\ samples}$.

The data reveal that the cellular proteostasis machinery does indeed differentially engage wild-type versus Gly1170Ser procollagen-II. Most notably, the pathogenic variant interacted to a greater extent with modifying enzymes known to engage with collagen in the ER (DiChiara et al., 2016). These interactors included procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 (*PLOD2*), protein disulfide-isomerase (*P4HB*), prolyl 3-hydroxylase 1 (*P3H1*), peptidyl-prolyl cis-trans isomerase (*FKBP10*), and peptidyl-prolyl cis-trans isomerase B (*PPIB*) (**Table 1**). Heightened interactions with procollagen-modifying enzymes are consistent with the substituted variant being hypermodified owing to slow folding.

Other differentially interacting proteins included known collagen chaperones calreticulin (*CALR*) and the collagen-specific molecular chaperone serpin H1 (*SERPINH1*) (**Table 1**). Increased interaction with these chaperones is consistent with Gly1170Ser being more challenging for cells to fold. Another notable interactor was the ER chaperone BiP (*HSPA5*) which interacted to the same extent with both wild-type and pathologic procollagen-II (presumably at their respective globular domains (Chessler and Byers, 1993; Doan et al., 2020; Lamandé et al., 1995)). This lack of differential interaction is consistent with the failure of Gly1170Ser procollagen-II to trigger the UPR, as seen in the RNA-sequencing data (**Fig. 7**).

Beyond modifying enzymes and chaperones, other interactors that differentially engaged Gly1170Ser procollagen-II were also proteins destined for secretion, such as cartilage-associated protein (*CRTAP*) and fibronectin (*FN1*) (**Table 1**). These secreted proteins are, like procollagen-II, folded in the ER, where they apparently develop more extensive interactions with the slow-folding Gly1170Ser procollagen-II variant than with wild-type procollagen-II. Intriguingly, other disease-causing variants of procollagen-II were previously shown to have increased interactions with fibronectin, a finding suggested to underly abnormal ECM formation (Ito et al., 2005).

Thus, the cellular proteostasis machinery can recognize slow-folding Gly1170Ser through increased interactions with certain ER proteostasis network components. These interactions do not alleviate the challenges associated with folding this variant, but they may present an opportunity for pharmacologic tuning of the ER proteostasis network to address the ER storage defect (Bateman et al., 2022; Sebastian and Shoulders, 2020; Wong and Shoulders, 2019).

Discussion

Since it was first described in 2005 (Liu et al., 2018; Liu et al., 2005; Miyamoto et al., 2007; Su et al., 2010; Su et al., 2008; Wang et al., 2014; Zhang et al., 2021), little progress has been made towards understanding the pathogenic mechanisms in the chondrodysplasia caused by the Gly1170Ser substitution in procollagen-II (Lian et al., 2019; Liang et al., 2014; Su et al., 2008; Yang et al., 2014). Here, we define the effects of the Gly1170Ser substitution in a robust human iPSC-based tissue model of cartilage and elucidate the proteostasis defects that arise from expression of this pathologic variant, in both the clinically relevant heterozygous model and the more severe but less clinically relevant homozygote. We find that, despite their ability to deposit a cartilage matrix rich in collagen-II and proteoglycans, cells expressing Gly1170Ser-substituted collagen-II show signs of pathology, in both the hetero- and homozygous models. Their matrix is deficient and lacks much of the structural collagen network. At the cellular level, Gly1170Ser procollagen-II is slow to fold and accumulates excessive post-translational modifications. Delayed triple-helix folding also results in a large amount of procollagen being retained within a dilated ER. This ER accumulation, however, is not accompanied by any substantive UPR.

In a complementary approach, Liang *et al*. previously generated a knock-in mouse model of the Gly1170Ser substitution, and characterized wild-type, heterozygous, and homozygous mice at birth or earlier (Liang et al., 2014). Heterozygous pups showed no skeletal phenotype, while homozygous pups presented with severe skeletal malformations and died shortly after birth from respiratory distress. Consistent with the findings from our human cartilage model, they observed a sparse ECM deposited by transgenic mice, as well as ER dilation in both transgenic genotypes, and intracellular retention of collagen-II, albeit only in chondrocytes homozygote for the substitution. Conversely, they also observed activation of the UPR in homozygous chondrocytes by quantitative RT-PCR of six UPR-related genes (and upregulation of two of those genes in heterozygous chondrocytes). These differences highlight the intricacies of various model systems and underline the potential benefits of studying collagenopathies in models that most closely resemble the human context.

From a biophysical perspective, the pathologic Gly→Ser substitution disrupts the triple helix because it replaces the conserved, small glycine residue at the interior of the helix with a much larger serine, causing steric hindrance. Moreover, it disrupts a key interchain hydrogen bond that stabilizes the helix because the amide hydrogen of serine is not in the proper orientation to serve as a donor site (Shoulders and Raines, 2009). Such disruption slows triple-helix folding, causing the protein to accumulate posttranslational modifications, which we detected by reduced electrophoretic mobility. The hypermodification may itself be problematic, as we will discuss later, but it is the slow folding leading to procollagen-II accumulation within the ER that is the hallmark of an ER storage disease (Rutishauser and Spiess, 2002).

Notably, even as Gly1170Ser procollagen-II accumulates in the ER, it does not trigger a chronic UPR. Bypassing the UPR despite extensive ER accumulation of a misfolding protein can occur if the accumulating protein evades detection by UPR signaling pathways. The primary mechanism to activate the UPR relies on the ER chaperone BiP. Downstream UPR signaling is triggered when BiP is sequestered off the three UPR sensors, IRE1, ATF6 and PERK, to instead bind un- or misfolded proteins through hydrophobic interactions with exposed hydrophobic cores. Procollagen, however, is a long hydrophilic protein lacking such a core. Moreover, its triple helix – where the Gly1170Ser substitution is located – is adorned with many hydrophilic post-translational modifications (Eyre, 1987; Yamauchi and Sricholpech, 2012). In fact, there is currently no evidence that BiP binds the collagen triple helix: BiP does bind to procollagen's C-propeptide domain during folding, and interacts particularly strongly with misfolding-prone C-propeptide domain variants (Chessler and Byers, 1993; Doan et al., 2020; Lamandé et al., 1995; Li et al., 2021; Wong and Shoulders, 2019), but, again, there is no evidence for BiP binding to the triple-helical domain of even disease-causing triple-helical domain variants of collagen-I.

Our data suggest that an underlying cause of pathology is the slow folding and accumulation of Gly1170Ser procollagen-II intracellularly, which occurs in the absence of any UPR or other stress response. Indeed, for many of the common Gly→Ser substitutions that induce collagenopathies, UPR activation is, at most, inconsistently observed (Bateman et al., 2022). Our results add weight to the view that chronic, pathologic

UPR is by no means a universal mechanism underpinning the collagenopathies (Bateman et al., 2022; Wong and Shoulders, 2019). Relatedly, we do not detect any significant increase in apoptosis in cells expressing Gly1170Ser procollagen-II. This finding could be a product of the timepoint considered, but it is at least fully consistent with the absence of the type of chronic activation of the PERK arm of the UPR that can induce apoptosis over time.

From a therapeutic perspective, this finding that cells expressing the pathologic variant do not activate beneficial aspects of the UPR makes those pathways a compelling target for this, and likely many other, collagenopathies (Balch et al., 2008; Sebastian and Shoulders, 2020; Wong and Shoulders, 2019). Exogenous, branch-specific activation of the pro-folding IRE1 and ATF6 arms of the UPR could promote procollagen folding by upregulating chaperones and quality control factors, as well as clearing accumulated protein by upregulation of clearance machinery (Ryno et al., 2013; Shoulders et al., 2013). Even just one of these outcomes could greatly benefit cells expressing Gly1170Ser procollagen-II. This strategy has already been applied in other protein folding diseases (Ryno et al., 2013), and many small molecules are already known to specifically target different nodes within the UPR (Grandjean and Wiseman, 2020).

Further evidence that promoting proteostasis, whether via the UPR or more generally, may be a viable therapeutic strategy for this disease, derives from the location of the substitution within the collagen-II triple helix. Position 1170 does not lie in any known binding site for other proteins (Chen and Lin, 2019; Hamaia et al., 2012; Howes et al., 2014), and therefore likely does not impede important signaling or structural functions mediated by its interactors. Rather, it seems likely that pathology is due directly to proteostasis defects.

Generally, the nature and location of the Gly1170Ser substitution would suggest a much more severe pathology than that observed in both patients and in our cartilage model because (1) it involves a Gly substitution, which are generally more detrimental to procollagen folding than substitutions of Xaa- or Yaa-position residues (Bateman et al., 2009), and (2) it is C-terminal, and thus, since procollagen folds from the C- to N-terminus, would disrupt the folding of a much larger portion of the protein than if it were located near the

N-terminus (Bodian et al., 2008). The symptoms of this disease, however, fall on the milder side of the spectrum, which is likely attributed to the proline-rich region located directly N-terminal to the substitution. The propensity of peptides to assemble into triple helices increases with the number of prolines (Shoulders and Raines, 2009), and thus, the proline-rich region likely helps procollagen-II to more rapidly resume folding despite the presence of a large, interchain hydrogen bond-breaking amino acid.

Despite its “favorable” position, Gly1170Ser still causes pathology. The ER storage phenotype associated with expression of Gly1170Ser procollagen-II propagates to the tissue, as these cells deposit a deficient ECM, lacking many of the network-forming fibrils observed in the wild-type. In patients, such failure likely leads to cartilage degradation ultimately requiring total joint replacement. From a molecular perspective, multiple factors could be causing the deficient matrix that ultimately contributes to symptoms in patients. One possibility is that cells expressing the pathologic variant secrete less collagen into the extracellular space. Another is that the secreted collagen is structurally abnormal and unable to correctly assemble into fibrils. Alternatively, or in conjunction, the deficient matrix could be a product of a defect in crosslinking. Hypermodification of lysines in collagen is known to affect crosslinking, as the process requires both a lysine and a hydroxylysine devoid of *O*-glycans in order to form a covalent crosslink (Eyre, 1987). Our data indicate that Gly1170Ser-substituted procollagen-II is hypermodified due to slow folding, which likely leads to a higher proportion of *O*-glycosylated hydroxylysines, and therefore could disrupt crosslinking and overall matrix health.

Interestingly, while the slow folding of Gly1170Ser procollagen-II, intracellular accumulation of the protein, and deficient ECM phenotypes all display a stepwise difference in severity between the heterozygous and homozygous variants, our RNA-sequencing data reveal that the homozygous cells are not, at least from a transcriptome standpoint, simply a more severe version of the heterozygotes. Both heatmaps and GSEA analyses revealed that different genes and pathways were differentially expressed depending on the genotype. This finding has implications for mouse models where homozygotes are often used to model autosomal dominant disorders in humans, owing to a lack of disease phenotype in mice heterozygous for the mutation. Nonetheless, our finding that homozygous chondronoids displayed a delay in hypertrophy is consistent with the

severe skeletal dysplasia and disorganized growth plate observed in homozygous Gly1170Ser mice (Liang et al., 2014).

In this work, we focused on modeling growth plate cartilage to understand how the mutation impacts development of cartilage and the skeleton, and how those relate to pathology, in both the clinically relevant heterozygous variant and the more severe homozygote. In future work, it will be equally interesting to investigate the phenotype in articular cartilage, which can also be grown from iPSCs (Craft et al., 2015), to further understand how the mutation relates to the osteoarthritic phenotype in particular.

The comparison in a human model with isogenic background enabled by the chondronoid system allowed us to define the pathologic phenotype and elucidate the defects in procollagen-II proteostasis. Moreover, by virtue of being highly expandable and modeling tissue, chondronoids constitute an ideal platform for screening small molecules and genetic interventions to enhance procollagen-II proteostasis. The model is equally conducive to in depth biochemical analyses, and, with recent advances in iPSC-based modeling of the endochondral ossification process (Lamandé et al., 2023), could enable similar studies in collagenopathies involving bone disorders.

Methods

General Reagents

All media and cell culture reagents were obtained from Corning/Cellgro, unless otherwise noted. Restriction enzymes, ligases, and polymerases were obtained from New England BioLabs. HA-Agarose beads (A2095) were obtained from Millipore.

Plasmids

The plasmid encoding wild-type human *COL2A1* (variant IIB, consensus sequence) was obtained from Origene (#RG221644, NM_033150). The C-terminal GFP tag was removed and replaced by a stop codon. To the N-terminus, a pre-protrypsin signal sequence, HA epitope tag, and NotI site were inserted. The Gly1170Ser *COL2A1* plasmid was made using site directed mutagenesis of the wild-type plasmid.

Antibodies

Antibodies were obtained from the following suppliers, and used at the specified dilutions: anti-Collagen-II (Millipore MAB8887, 1:150), anti-Calreticulin (Invitrogen AB325990, 1:200), Goat Anti-Rabbit Alexa Fluor 488 (Abcam ab150077, 1:500), Goat Anti-Mouse Alexa Fluor 568 (Abcam ab175473, 1:500).

Induced pluripotent stem cell lines growth and maintenance

Generating the gene edited lines from ATCC fibroblast line 1502 is covered by the Human Research Ethics Committee of the Royal Children's Hospital, Victoria, Australia (HREC #33118) (Kung et al., 2020). iPSC lines MCRIi019-A (wild-type control) and MCRIi019-A-2 (*COL2A1* p.Gly1170Ser heterozygote) and the homozygote variant iPSCs were grown in a feeder-free manner on Matrigel (Corning)-coated 6-well plates in Essential 8 (E8) medium (ThermoFisher Scientific). Media was changed daily. Cells were routinely passaged (1:6 split) every 3 – 4 d and gently detached from plates using 0.5 mM EDTA in PBS.

Chondrocyte differentiation

Undifferentiated iPSCs at 70–90% confluency were dissociated using 0.5 mM EDTA in PBS and passaged into 6-well plates (~2×10^5 cells per well) pre-coated with Matrigel, and cultured for 48 h in E8 expansion medium. Differentiation to sclerotome was preformed as described (Lamandé et al., 2023), with APEL2 (StemCell Technologies) containing 5% Protein Free Hybridoma Medium (PFHM II; ThermoFisher Scientific). On day 0, medium was changed to anterior primitive streak inducing medium containing 30 ng/ml Activin A (R&D Systems), 4 μM CHIR99021 (Tocris), 20 ng/mL FGF2 (PeproTech) and 100 nM PIK90 (Merck Millipore). After 24 h this medium was replaced with medium containing 3 μM CHIR99021, 20 ng/mL FGF2, 1 μM A8301 (Tocris) and 0.25 μM LDN193189 (Cayman Chemical) to induce paraxial mesoderm for 24 h. Early somite development was then induced with medium containing 1 μM A8301 and 0.25 μM LDN193189, 3 μM C59 (Tocris) and 0.5 μM PD0325901 (Selleck Chemicals). After 24 h, sclerotome induction was initiated with 1 μM C59 and 2 μM purmorphamine (Sigma-Aldrich). After the first day of sclerotome induction (day 4) cells were dissociated from monolayer culture using Tryple Select (ThermoFisher Scientific), resuspended in sclerotome differentiation media, 300 μL aliquots containing 2×10^5 cells were dispensed into 96-well low-attachment, round-bottom plates (Corning), and pelleted by centrifugation at 400 × *g* for 3 min using a swing-out rotor. Pellets were incubated in sclerotome differentiation media for a further 48 h to complete sclerotome differentiation in the 96-well static culture format. Differentiations were performed in biological triplicate, where individual replicates originated from cells of separate plates differentiated with individually prepared differentiation media from different media bottles.

Chondrocyte maturation and extracellular matrix deposition

To promote chondrocyte differentiation, maturation, and ECM deposition, sclerotome cell pellets were cultured in 96-well round bottom plates in APEL2 medium supplemented with 5% PFHM II, 20 ng/mL FGF2 (PeproTech), 1× Penicillin-Streptomycin (ThermoFisher Scientific), and 200 μM freshly added sodium ascorbate (Amresco). On day 20, pellets were transferred to 6 cm non-adherent culture dishes (Greiner) with 15 – 20 pellets per dish in 5 mL of APEL2/PFHM II medium lacking FGF2 with orbital rotation at 60 rpm. Media was changed every 2 – 3 d for an additional 24 d of maturation. For harvesting, chondronoids were rinsed once in PBS and snap-frozen or processed for downstream assays. For each assay, chondronoids were harvested

from all three biological replicates. Chondronoids originating from the same differentiation were considered technical replicates. A micrometer was used for precise diameter measurement of each of 10 chondronoids per genotype.

Histology and toluidine blue staining

Cultured chondronoids were fixed overnight at 4 °C in 10% neutral-buffered formalin (Millipore), washed in 70% ethanol, processed on a tissue processor (Sakura Tissue Tek VIP5) with a biopsy processing schedule as follows: 70% Ethanol – 10 mins, 95% Ethanol × 2 – 10 mins each, 100% Ethanol × 3 – 10 mins each, Xylene × 2 – 10 mins each, Paraffin wax × 4 – 10 mins each, and then embedded in “Paraplast X-tra”rrections paraffin wax using a Leica embedding station. Serial 4 µm sections were cut from the chondronoids using a Leica rotary microtome. Sections were floated onto a 45 °C water bath and mounted on Leica “X-tra” adhesive slides (Leica biosytems), drained, and dried at room temperature overnight. Before staining, sections were heated at 60 °C for 30 min. Sections were treated with xylene to remove the paraffin wax and then with an ethanol series comprising 100%, 90%, 70% ethanol then water to rehydrate the sections. Sections were stained with toluidine blue solution (1% toluidine blue, 0.2 M sodium phosphate pH 9, filtered) for 10 min to detect a cartilage proteoglycan matrix, then rinsed in distilled water. Microscope slides were cover slipped using ProLong Gold Antifade Reagent (Invitrogen) and sealed with nail polish. Images were captured using a Leica DM 2000 LED microscope with Leica Application Suite (LAS) software version 4.9.0. Image processing was performed using ImageJ.

Immunohistochemistry and confocal microscopy

Rehydrated histological sections underwent antigen retrieval by incubation with 1 mg/mL pepsin in 0.5 M acetic acid (Millipore) at 37 °C for 30 min. Samples were washed 3× with PBS, then incubated with 0.1% Triton X-100 for 30 min to permeabilize the cells. Samples were then incubated in a solution of 5% bovine serum albumin (BSA, Millipore) in PBS at RT for 1 h to block non-specific antibody binding. Samples were labeled simultaneously with anti-Col-II antibody (Millipore MAB8887, 1:150) and anti-Calr antibody (Invitrogen AB325990, 1:200) at in 1% BSA for 1 h at RT. Samples were washed 3 × with PBS. Both secondary antibodies, Alexa Fluor 568-conjugated anti-mouse and Alexa Fluor 488-conjugated anti-rabbit (Invitrogen;

1:500), were applied simultaneously in 1% BSA solution for 30 min – 1 h at RT. Samples were washed 3 × with PBS. Cover slips were mounted on the microscope slides using ProLong Gold Antifade Reagent with DAPI (Invitrogen) and sealed with nail polish. Images were acquired at the W. M. Keck Microscopy Facility at the Whitehead Institute on a Zeiss AxioVert200M microscope with a 10 ×, 63 ×, and 100 × oil immersion objective with a Yokogawa CSU-22 spinning disk confocal head with a Borealis modification (Spectral Applied Research/Andor) and a Hamamatsu ORCA-ER CCD camera. The MetaMorph software package (Molecular Devices) was used for image acquisition. The excitation lasers used to capture the images were 405 nm, 488 nm, and 561 nm. Image processing was performed using ImageJ.

TUNEL assay

Rehydrated histological sections were assayed for apoptosis using an *in situ* cell death detection kit (Millipore), according to the manufacturer's instructions. Five different chondronoids per genotype were used for analysis. Briefly, following cell permeabilization, DNA breaks were labeled directly by the enzyme terminal deoxynucleotidyl transferase (TdT) with fluorescein labeled dUTPs. Positive control sample was incubated with DNAse for 15 min at 37 °C prior to TdT labeling. Negative control sample did not include the TdT enzyme. Samples were visualized on a fluorescence microscope (ECHO Revolve 4). Signal quantification was performed on the ECHO Pro App by analyzing the pixel intensity over the highlighted region of interest. Significance was assessed using a student's *t*-test on at least 10 different sections per genotype.

Transmission electron microscopy

Samples were washed with PBS and fixed with 2.5% glutaraldehyde, 2.0% paraformaldehyde in 100 mM cacodylate buffer pH 7.2 for 2 h at 4 °C, then post-fixed with 1% osmium tetroxide in 1.25% potassium ferrocyanide. They were then en-bloc stained with 2% uranyl acetate in 0.05 M maleate buffer pH 5.2 overnight at RT, followed by serial dehydrations with ethanol, then embedded in resin at 60 °C for 48 h. 60 nm sections were obtained using a diamond knife on a Leica UC67 Ultramicrotome and observed at 120 kV on a T12 Spirit Transmission Electron Microscope (ThermoFisher Instruments). Micrographs were captured using a digital camera from Advanced Microscopy Techniques (AMT).

RNA-sequencing

For RNA-sequencing, individual chondronoids were snap frozen in liquid nitrogen and pulverized using a liquid nitrogen-cooled tissue grinder. RNA was extracted with TRIzol (Invitrogen), followed by purification using Direct-zol RNA Microprep kit spin columns (Zymo Research) according to manufacturer's instructions. RNA samples were run on a fraction analyzer at MIT's BioMicro Center for quality control. All RNA samples were processed at the same time using the same batch of reagents. RNA samples were quantified using an Advanced Analytical Fragment Analyzer. 10 ng of total RNA was used for library preparation on a Tecan ECO150. Custom primers were used to append a 6 bp barcode specific to each sample, as well as a unique molecular identifier (Integrated DNA technologies). Maxima H Minus Reverse Transcriptase (Thermo) was added per manufacturer's recommendations with a template-switching oligo, incubated at 42 °C for 90 min and inactivated by incubation at 80 °C for 10 min. Following the template switching reaction, cDNA from 24 wells containing unique well identifiers were pooled together and cleaned using RNA Ampure beads at 1.0×. cDNA was eluted with 17 µL of water, digested with Exonuclease-I (New England Biolabs) at 37 °C for 30 min, and inactivated by incubation at 80 °C for 20 min. Second strand synthesis and PCR amplification was done by adding Advantage 2 Polymerase Mix (Clontech) and the SINGV6 primer (Integrated DNA Technologies, 10 pmol, 5'-/5Biosg/ACACTCTTTCCCTACACGACGC-3') directly to half of the exonuclease reaction volume. Eight cycles of PCR were performed followed by clean up using regular SPRI beads at 0.6×, and elution with 20 µL of RSB. Successful amplification of cDNA was confirmed using the Fragment Analyzer. Illumina libraries were then produced using Nextera FLEX tagmentation substituting P5NEXTPT5-bmc primer (25 µM, Integrated DNA Technologies, (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG*A*T*C*T*-3' where * = phosphorothioate bonds.) in place of the normal N500 primer. Final libraries were cleaned using SPRI beads at 0.7 × and quantified using the Fragment Analyzer and qPCR before being loaded for paired-end sequencing using the Illumina NextSeq500 in high-output paired-end mode (20/57 nt reads). Analyses were performed using previously described tools and methods. Reads were aligned against hg19 (February 2009) using BWA mem v.0.7.12-r1039 (RRID:SCR_010910) with flags –t 16 –f, and mapping rates, fraction of multiply-mapping reads, number of unique 20-mers at the 5' end of the reads, insert size distributions and fraction of ribosomal

RNAs were calculated using BEDTools v. 2.25.0 (RRID:SCR_006646) (Quinlan and Hall, 2010). In addition, each resulting bam file was randomly down-sampled to a million reads, which were aligned against hg19, and read density across genomic features were estimated for RNA-Seq-specific quality control metrics. For mapping and quantitation, reads were scored against GRCh38/ENSEMBL 101 annotation using Salmon v.1.3 with flags quant -p 8 -l ISR –validateMappings (Patro et al., 2017). The resulting quant.sf files were imported into the R statistical environment using the tximport library (tximport function, option “salmon”), and gene-level counts and transcript per milllion (TPM) estimates were calculated for all genes (Soneson et al., 2015). Samples were clustered based on $\log_2$-transformed TPMs of protein-coding genes and obvious outliers were excluded from downstream analyses.

Differential expression was also analyzed in the R statistical environment (R v.3.5.1) using Bioconductor’s DESeq2 package (RRID:SCR_000154) (Love et al., 2014). Dataset parameters were estimated using the estimateSizeFactors(), and estimateDispersions() functions; read counts across conditions were modeled based on a negative binomial distribution, and a Wald test was used to test for differential expression (nbinomWaldtest(), all packaged into the DESeq() function), using the treatment type as a contrast. Shrunken $\log_2$ fold-changes were calculated using the lfcShrink function, based on a normal shrinkage estimator.(Love et al., 2014) Fold-changes, *p*-values and false-discovery rates were reported for each gene. In the heterozygous line, X-chromosome deactivation was eroded, as is a common occurrence in iPSC lines (Anguera et al., 2012; Fukuda et al., 2021; Mekhoubad et al., 2012), so hits caused by differential expression of X-linked genes were disregarded. All the gene sets considered were either previously published (Gorrell et al., 2022; Grandjean et al., 2019; Lamandé et al., 2023; Mirigian et al., 2016; Phillips et al., 2018), or curated from previously described gene ontologies.

Gene Set Enrichment Analysis (GSEA)

Differential expression results from DESeq2 were retrieved, and the “stat” column was used to pre-rank genes for GSEA analysis. These “stat” values reflect the Wald’s test performed on read counts as modeled by DESeq2 using the negative binomial distribution. Genes that were not expressed were excluded from the

analysis. GSEA (linux desktop version, v4.1) (Subramanian et al., 2005) was run in the pre-ranked mode against MSigDB 7.0 C5 (Gene Ontology) set, and ENSEMBL IDs were collapsed to gene symbols using the Human_ENSEMBL_Gene_ID_MSigDB.v7.4.chip (resulting in 25162 and 26199 genes for WT vs. mutant sample in early and late time-point comparisons, respectively). In addition, a weighted scoring scheme, meandiv normalization, cutoffs on MSigDB signatures sizes (between 5 and 2000 genes, resulting in 9,476 and 9,588 gene sets retained for early and late time-point wild-type vs. mutant comparisons, respectively) were applied and 5,000 permutations were run for *p*-value estimation.

Collagen extraction and SDS-PAGE analysis

Chondronoids of the same genotype were rinsed with PBS, blotted dry with a Kim wipe, pooled, snap frozen in liquid nitrogen, and pulverized using a liquid nitrogen-cooled tissue grinder. The powder obtained was added to 100 μL chondroitinase ABC solution (50 mM Tris HCl pH 8, 50 mM sodium acetate, 0.1 mU/μL chondroitinase ABC (Millipore)) and incubated in a thermomixer set to 37 °C and 1,500 rpm for 6 h. 900 μL of denaturing solution (50 mM sodium acetate, 10 mM EDTA, 4 M guanidine hydrochloride, 65 mM dithiothreitol) was added, and samples were rotated end-over-end at 4 °C overnight. The next day, samples were centrifuged at 13,000 rpm at 4 °C for 10 min. The supernatant was removed and the pellet resuspended in 200 μL of cold pepsin solution (0.1 mg/mL pepsin (Millipore) in 0.5 M acetic acid) and rotated end-over-end at 4 °C overnight. The next day, samples were centrifuged at 13,000 rpm at 4 °C for 10 min. The supernatant was saved, neutralized with 10 M NaOH, and lyophilized. Lyophilized samples were reconstituted in loading buffer for SDS-PAGE (50 mM Tris HCl pH 6.8, 2 M urea, 0.1% SDS, 0.1% sucrose, 0.01% bromophenol blue (Millipore)) and boiled for 15 min. Samples were separated on homemade 4/8% tris SDS-PAGE gels. Gels were stained in Coomassie Blue solution (0.1% Coomassie Blue (Millipore), 10% acetic acid, 50% methanol, 40% water) and destained in the same solution without Coomassie Blue. Gels were imaged on an Epson scanner.

Intracellular collagen-II quantification

A survey of confocal images of immunohistochemistry samples stained for collagen-II was used to quantify the percent of cells with intracellularly retained collagen-II. 75 images of the different genotypes at different

magnifications (40 – 100 ×) were arranged in a random order. Each genotype set consisted of at least 4 individual chondronoids. Three independent and blinded participants were asked to count the total number of cells and the number of cells with intracellular staining in each image. The proportion of positive cells was averaged between all participants for each genotype, and the error bars represent the standard error of the mean. Statistical significance between genotypes was calculated using a heteroscedastic student's *t*-test for each participant, and the largest of the three *p*-values obtained is displayed on the chart.

Proteomics

Cell Culture, transfections, and immunoprecipitation. HT-1080 cells (ATCC) were cultured in complete DMEM supplemented with 10% FBS. Transfections of wild-type and Gly1170Ser *COL2A1*-encoding plasmids were performed using Transit-2020 transfection reagent (Mirus) according to manufacturer's instructions. The next day, media was replaced with fresh complete DMEM supplemented with 50 μM sodium ascorbate (Amresco). Two days post-transfection, cells were harvested and washed three times with 1× phosphate buffered saline (PBS) and crosslinked in plate with 0.2 mM dithiobis(succinimidyl propionate) (DSP; Lomant's Reagent, Thermo) in PBS at rt for 30 min. Crosslinking was quenched by addition of 0.1 M Tris at pH 8.0 (final concentration) for 15 min at rt. Cells were scraped from plates in PBS using a cell scraper, then pelleted and resuspended in RIPA: 150 mM sodium chloride, 50 mM Tris-HCl, pH 7.5, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), protease inhibitor tablets and 1.5 mM phenylmethylsulfonyl fluoride (PMSF; Amresco). Cells were lysed for 30 min on ice, then centrifuged at 21,100 × g for 15 min at 4 °C. Supernatants were collected, quantified for protein content (BCA assay, Pierce), normalized, and incubated for 16 h with the HA antibody conjugated beads, rotating end-over-end at 4 °C. Samples were washed with RIPA 3 × and eluted by boiling for 10 min in 300 mM Tris at pH 6.8 with 6% SDS and 600 mM dithiothreitol (DTT, Millipore) to release crosslinks. Eluates were prepared for mass spectrometry analysis.

Mass spectrometry and interactome characterization. Samples were reduced, alkylated, and digested with sequencing-grade pepsin (Promega) using an S-Trap Micro (Protifi) following standard protocol. After

digestion, the samples were labeled with TMT 10plex™ reagents (Thermo) according to manufacturer's instructions. Briefly, the dried tryptic peptides were resuspended in 100 mM triethylammmonium bicarbonate (TEAB), vortexed. Acetonitrile was added to each of the 10-plex TMT label reagents, vortexed, and briefly centrifuged. Peptides and labeling reagents were then combined in a 1:2 ratio and incubated for 1 h at rt, after which the reaction was quenched with 5% hydroxylamine for 15 min. Equal amounts of each channel were combined, dried by speed-vacuum, and resuspended in TEAB. Samples were then fractionated with Pierce High pH Reversed-Phase peptide Fractionation Kits (cat#84868) using standard protocols, then dried and resuspended in 0.2% formic acid for injection on LC-MS/MS. Labeled tryptic peptides were separated by reverse phase HPLC (Thermo Ultimate 3000) using a Thermo PepMap RSLC C_{18} column (2 µm tip, 75 µm × 50 cm PN# ES903) over a gradient before nano-electrospray injection onto an Orbitrap Exploris 480 mass spectrometer (Thermo).

The mass spectrometer was operated in a data-dependent mode. The parameters for the full scan MS were: resolution of 120,000 across 375-1600 *m/z* and maximum IT of 25 ms. The full MS scan was followed by MS/MS for as many precursor ions in a three second cycle with a NCE of 32, dynamic exclusion of 30 s and resolution of 45,000. Raw mass spectral data files (.raw) were searched using Sequest HT in Proteome Discoverer (Thermo). Sequest search parameters were: 10 ppm mass tolerance for precursor ions; 0.05 Da for fragment ion mass tolerance; 2 missed cleavages of trypsin; fixed modification were carbamidomethylation of cysteine and TMT modification on the lysines and peptide N-termini; variable modifications were methionine oxidation, methionine loss at the N-terminus of the protein, acetylation of the N-terminus of the protein, methylation of histidine, arginine, and glutamine, hydroxylation of proline, and methionine-loss plus acetylation of the protein N-terminus; total PTM per peptide was set to 8. The false discovery rate was set at 0.01. Data was searched against a human database and a common MS contaminant database made in house. Human Database: https://www.uniprot.org/proteomes/UP000005640

Raw abundance values were used for quantification. First, abundance of each protein in the negative control was determined as the average abundance of the three no transfection control conditions. High confidence

interactors were identified as any protein detected with an abundance of at least 2 × that of the negative control samples across all three replicates. Next, for each experimental sample, the abundance of each protein was normalized to the abundance of bait (collagen-II) to obtain a normalized abundance value. The fold enrichment was calculated as the ratio of the average of normalized abundance values in the Gly1170Ser conditions to the average of normalized abundance values in the wild-type conditions (Plate et al., 2019). Significance was evaluated using a homoscedastic student's *t*-test across the three biological replicates, as well as the two-stage step-up method of Benjamini, Krieger, and Yekutieli with a desired FDR set to 10%.

Acknowledgements

We extend our gratitude to Dr. Louise Kung, Lisa Sampurno, and Lynn Rowley at Murdoch Children's Research Institute for generously sharing their expertise and technical knowledge in iPSC culture, differentiation, and analysis. We thank Margaret Bisher, David Mankus, and Dr. Abigail Lytton-Jean at the Koch Institute's Robert A. Swanson Biotechnology Center Nanotechnology Materials Facility for their expertise and technical assistance with electron microscopy, as well as Caroline Atkinson at the MIT Division of Comparative Medicine for technical support and advice. We thank Brandyn Braswell and others at the W.M. Keck Microscopy Facility at the Whitehead Institute for technical support with confocal microscopy. We thank Noelani Kamelamela, Samuel Mildrum, and Dr. Stuart Levine at the MIT BioMicro Center for their assistance with RNA-sequencing. We thank Jimin Yoon, Jess Patrick, and Sorin Srinivasa for generously offering their time to quantify intracellular collagen. We also thank Andrea Yammine for contributions to graphic design.

Funding

This work was supported by the National Institutes of Health (Grant 1R01AR071443 to M.D.S.), a Research Grant from the G. Harold and Leila Y. Mathers Foundation (to M.D.S., S.R.L., and J.F.B.), the Australia National Health & Medical Research Council (GNT2003393 and GNT1146952 both to S.R.L. and J.F.B., and GNT1146902 to J.F.B.), and the Victorian Government's Operational Infrastructure Support Program. K.M.Y. was supported by an NIH Ruth L. Kirschstein Predoctoral Fellowship (F31AR079263). S.K. was supported by a Kwanjeong Fellowship. This work was also supported in part by the National Cancer Institute (Koch Institute core) Grant P30-CA14051 and the National Institute of Environmental Health Sciences (core) Grant P30-ES002109. Work in the Novo Nordisk Foundation Center for Stem Cell Medicine is supported by a Novo Nordisk Foundation grant (NNF21CC0073729).

Conflict of Interest

The authors declare that they have no conflicts of interest.

References

Adkar, S.S., C.L. Wu, V.P. Willard, A. Dicks, A. Ettyreddy, N. Steward, N. Bhutani, C.A. Gersbach, and F. Guilak. 2019. Step-wise chondrogenesis of human induced pluripotent stem cells and purification via a reporter allele generated by CRISPR-Cas9 genome editing. *Stem Cells*. 37:65-76.

Anguera, M.C., R. Sadreyev, Z. Zhang, A. Szanto, B. Payer, S.D. Sheridan, S. Kwok, S.J. Haggarty, M. Sur, J. Alvarez, A. Gimelbrant, M. Mitalipova, J.E. Kirby, and J.T. Lee. 2012. Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. *Cell Stem Cell*. 11:75-90.

Balch, W.E., R.I. Morimoto, A. Dillin, and J.W. Kelly. 2008. Adapting proteostasis for disease intervention. *Science*. 319:916-919.

Bateman, J.F., R.P. Boot-Handford, and S.R. Lamandé. 2009. Genetic diseases of connective tissues: cellular and extracellular effects of ECM mutations. *Nat Rev Genet*. 10:173-183.

Bateman, J.F., D. Chan, T. Mascara, J.G. Rogers, and W.G. Cole. 1986. Collagen defects in lethal perinatal osteogenesis imperfecta. *Biochem J*. 240:699-708.

Bateman, J.F., T. Mascara, D. Chan, and W.G. Cole. 1984. Abnormal type I collagen metabolism by cultured fibroblasts in lethal perinatal osteogenesis imperfecta. *Biochem J*. 217:103-115.

Bateman, J.F., M.D. Shoulders, and S.R. Lamandé. 2022. Collagen misfolding mutations: the contribution of the unfolded protein response to the molecular pathology. *Connect Tissue Res*. 63:210-227.

Bauer, J.W., and C. Lanschuetzer. 2003. Type XVII collagen gene mutations in junctional epidermolysis bullosa and prospects for gene therapy. *Clin Exp Dermatol*. 28:53-60.

Bodian, D.L., B. Madhan, B. Brodsky, and T.E. Klein. 2008. Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations. *Biochemistry*. 47:5424-5432.

Bonadio, J., K.A. Holbrook, R.E. Gelinas, J. Jacob, and P.H. Byers. 1985. Altered triple helical structure of type I procollagen in lethal perinatal osteogenesis imperfecta. *J Biol Chem*. 260:1734-1742.

Cabral, W.A., I. Perdivara, M. Weis, M. Terajima, A.R. Blissett, W. Chang, J.E. Perosky, E.N. Makareeva, E.L. Mertz, S. Leikin, K.B. Tomer, K.M. Kozloff, D.R. Eyre, M. Yamauchi, and J.C. Marini. 2014. Abnormal type I collagen post-translational modification and crosslinking in a cyclophilin B KO mouse model of recessive osteogenesis imperfecta. *PLoS Genet*. 10:e1004465.

Chen, E.A., and Y.S. Lin. 2019. Using synthetic peptides and recombinant collagen to understand DDR-collagen interactions. *Biochim Biophys Acta Mol Cell Res*. 1866:118458.

Chessler, S.D., and P.H. Byers. 1993. BiP binds type I procollagen pro alpha chains with mutations in the carboxyl-terminal propeptide synthesized by cells from patients with osteogenesis imperfecta. *J Biol Chem*. 268:18226-18233.

Craft, A.M., J.S. Rockel, Y. Nartiss, R.A. Kandel, B.A. Alman, and G.M. Keller. 2015. Generation of articular chondrocytes from human pluripotent stem cells. *Nat Biotechnol*. 33:638-645.

DiChiara, A.S., R.J. Taylor, M.Y. Wong, N.D. Doan, A.M. Rosario, and M.D. Shoulders. 2016. Mapping and exploring the collagen-I proteostasis network. *ACS Chem Biol*. 11:1408-1421.

Doan, N.D., A.S. DiChiara, A.M. Del Rosario, R.P. Schiavoni, and M.D. Shoulders. 2019. Mass spectrometry-based proteomics to define intracellular collagen interactomes. *Methods Mol Biol*. 1944:95-114.

Doan, N.D., A.S. Hosseini, A.A. Bikovtseva, M.S. Huang, A.S. DiChiara, L.J. Papa, 3rd, A. Koller, and M.D. Shoulders. 2020. Elucidation of proteostasis defects caused by osteogenesis imperfecta mutations in the collagen-alpha2(I) C-propeptide domain. *J Biol Chem*. 295:9959-9973.

Eyre, D. 1987. Collagen cross-linking amino acids. *Methods Enzymol*. 144:115-139.

Eyre, D. 2002. Collagen of articular cartilage. *Arthritis Res*. 4:30-35.

Fukuda, A., D.Z. Hazelbaker, N. Motosugi, J. Hao, F. Limone, A. Beccard, P. Mazzucato, A. Messana, C. Okada, I.G. San Juan, M. Qian, A. Umezawa, H. Akutsu, L.E. Barrett, and K. Eggan. 2021. De novo DNA methyltransferases DNMT3A and DNMT3B are essential for XIST silencing for erosion of dosage compensation in pluripotent stem cells. *Stem Cell Reports*. 16:2138-2148.

Geddis, A.E., and D.J. Prockop. 1993. Expression of human *COL1A1* gene in stably transfected HT1080 cells: the production of a thermostable homotrimer of type I collagen in a recombinant system. *Matrix*. 13:399-405.

Godfrey, M., and D.W. Hollister. 1988. Type II achondrogenesis-hypochondrogenesis: identification of abnormal type II collagen. *Am J Hum Genet*. 43:904-913.

Goldring, M.B., and F. Berenbaum. 1999. Human chondrocyte culture models for studying cyclooxygenase expression and prostaglandin regulation of collagen gene expression. *Osteoarthritis Cartilage*. 7:386-388.

Gorrell, L., E. Makareeva, S. Omari, S. Otsuru, and S. Leikin. 2022. ER, mitochondria, and ISR regulation by mt-HSP70 and ATF5 upon procollagen misfolding in osteoblasts. *Adv Sci (Weinh)*. 9:e2201273.

Grandjean, J.M.D., L. Plate, R.I. Morimoto, M.J. Bollong, E.T. Powers, and R.L. Wiseman. 2019. Deconvoluting stress-responsive proteostasis signaling pathways for pharmacologic activation using targeted RNA sequencing. *ACS Chem Biol*. 14:784-795.

Grandjean, J.M.D., and R.L. Wiseman. 2020. Small molecule strategies to harness the unfolded protein response: where do we go from here? *J Biol Chem*. 295:15692-15711.

Gregersen, P.A., and R. Savarirayan. 1993. Type II collagen disorders overview. *In* GeneReviews((R)). M.P. Adam, G.M. Mirzaa, R.A. Pagon, S.E. Wallace, L.J.H. Bean, K.W. Gripp, and A. Amemiya, editors, Seattle (WA).

Hamaia, S.W., N. Pugh, N. Raynal, B. Nemoz, R. Stone, D. Gullberg, D. Bihan, and R.W. Farndale. 2012. Mapping of potent and specific binding motifs, GLOGEN and GVOGEA, for integrin alpha1beta1 using collagen toolkits II and III. *J Biol Chem*. 287:26019-26028.

Hamosh, A., A.F. Scott, J.S. Amberger, C.A. Bocchini, and V.A. McKusick. 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 33:D514-517.

Horiuchi, K., T. Tohmonda, and H. Morioka. 2016. The unfolded protein response in skeletal development and homeostasis. *Cell Mol Life Sci*. 73:2851-2869.

Horton, W.A., and M.M. Machado. 1988. Extracellular matrix alterations during endochondral ossification in humans. *J Orthop Res*. 6:793-803.

Hosseini Far, H., Y.N. Patria, A. Motazedian, A.G. Elefanty, E.G. Stanley, S.R. Lamandé, and J.F. Bateman. 2019. Generation of a heterozygous *COL1A1* (c.3969_3970insT) osteogenesis imperfecta mutation human iPSC line, MCRIi001-A-1, using CRISPR/Cas9 editing. *Stem Cell Res*. 37:101449.

Howden, S., H. Hosseini Far, A. Motazedian, A.G. Elefanty, E.G. Stanley, S.R. Lamandé, and J.F. Bateman. 2019. The use of simultaneous reprogramming and gene correction to generate an osteogenesis imperfecta patient *COL1A1* c. 3936 G>T iPSC line and an isogenic control iPSC line. *Stem Cell Res*. 38:101453.

Howes, J.M., D. Bihan, D.A. Slatter, S.W. Hamaia, L.C. Packman, V. Knauper, R. Visse, and R.W. Farndale. 2014. The recognition of collagen and triple-helical toolkit peptides by MMP-13: sequence specificity for binding and cleavage. *J Biol Chem*. 289:24091-24101.

Hynes, R.O., and A. Naba. 2012. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. *Cold Spring Harb Perspect Biol*. 4:a004903.

Ito, H., E. Rucker, A. Steplewski, E. McAdams, R.J. Brittingham, T. Alabyeva, and A. Fertala. 2005. Guilty by association: some collagen II mutants alter the formation of ECM as a result of atypical interaction with fibronectin. *J Mol Biol*. 352:382-395.

Jobling, R., R. D'Souza, N. Baker, I. Lara-Corrales, R. Mendoza-Londono, L. Dupuis, R. Savarirayan, L. Ala-Kokko, and P. Kannu. 2014. The collagenopathies: review of clinical phenotypes and molecular correlations. *Curr Rheumatol Rep*. 16:394.

Koyama, N., M. Miura, K. Nakao, E. Kondo, T. Fujii, D. Taura, N. Kanamoto, M. Sone, A. Yasoda, H. Arai, K. Bessho, and K. Nakao. 2013. Human induced pluripotent stem cells differentiated into chondrogenic lineage via generation of mesenchymal progenitor cells. *Stem Cells Dev*. 22:102-113.

Kung, L.H.W., L. Sampurno, K.M. Yammine, A. Graham, P. McDonald, J.F. Bateman, M.D. Shoulders, and S.R. Lamandé. 2020. CRISPR/Cas9 editing to generate a heterozygous *COL2A1* p.G1170S human chondrodysplasia iPSC line, MCRIi019-A-2, in a control iPSC line, MCRIi019-A. *Stem Cell Res.* 48:e101962.

Lamandé, S.R., S.D. Chessler, S.B. Golub, P.H. Byers, D. Chan, W.G. Cole, D.O. Sillence, and J.F. Bateman. 1995. Endoplasmic reticulum-mediated quality control of type I collagen production by cells from osteogenesis imperfecta patients with mutations in the pro alpha 1 (I) chain carboxyl-terminal propeptide which impair subunit assembly. *J Biol Chem*. 270:8642-8649.

Lamandé, S.R., E.S. Ng, T.L. Cameron, L.H.W. Kung, L. Sampurno, L. Rowley, J. Lilianty, Y.N. Patria, T. Stenta, E. Hanssen, K.M. Bell, R. Saxena, K.S. Stok, E.G. Stanley, A.G. Elefanty, and J.F. Bateman.

2023. Modeling human skeletal development using human pluripotent stem cells. *Proc Natl Acad Sci U S A*. 120:e2211510120.

Lee, J., S.E. Taylor, P. Smeriglio, J. Lai, W.J. Maloney, F. Yang, and N. Bhutani. 2015. Early induction of a prechondrogenic population allows efficient generation of stable chondrocytes from human induced pluripotent stem cells. *FASEB J*. 29:3399-3410.

Li, R.C., M.Y. Wong, A.S. DiChiara, A.S. Hosseini, and M.D. Shoulders. 2021. Collagen's enigmatic, highly conserved *N*-glycan has an essential proteostatic function. *Proc Natl Acad Sci U S A*. 118.

Lian, C., X. Wang, X. Qiu, Z. Wu, B. Gao, L. Liu, G. Liang, H. Zhou, X. Yang, Y. Peng, A. Liang, C. Xu, D. Huang, and P. Su. 2019. Collagen type II suppresses articular chondrocyte hypertrophy and osteoarthritis progression by promoting integrin beta1-SMAD1 interaction. *Bone Res*. 7:8.

Liang, G., C. Lian, D. Huang, W. Gao, A. Liang, Y. Peng, W. Ye, Z. Wu, P. Su, and D. Huang. 2014. Endoplasmic reticulum stress-unfolding protein response-apoptosis cascade causes chondrodysplasia in a *COL2A1* p.Gly1170Ser mutated mouse model. *PLoS One*. 9:e86894.

Lilianty, J., J.F. Bateman, and S.R. Lamandé. 2021. Generation of a heterozygous *COL2A1* (p.G1113C) hypochondrogenesis mutation iPSC line, MCRIi019-A-7, using CRISPR/Cas9 gene editing. *Stem Cell Res*. 56:102515.

Lilianty, J., Y. Nur Patria, E.G. Stanley, A.G. Elefanty, J.F. Bateman, and S.R. Lamandé. 2020. Generation of a heterozygous *COL2A1* (p.R989C) spondyloepiphyseal dysplasia congenita mutation iPSC line, MCRIi001-B, using CRISPR/Cas9 gene editing. *Stem Cell Res*. 45:101843.

Liu, F., Z. Xiong, Q. Liu, J. Hu, W. Li, and N. Zhang. 2018. *COL2A1* mutation (c.3508G>A) leads to avascular necrosis of the femoral head in a Chinese family: A case report. *Mol Med Rep*. 18:254-260.

Liu, Y.F., W.M. Chen, Y.F. Lin, R.C. Yang, M.W. Lin, L.H. Li, Y.H. Chang, Y.S. Jou, P.Y. Lin, J.S. Su, S.F. Huang, K.J. Hsiao, C.S. Fann, H.W. Hwang, Y.T. Chen, and S.F. Tsai. 2005. Type II collagen gene variants and inherited osteonecrosis of the femoral head. *N Engl J Med*. 352:2294-2301.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550.

Mekhoubad, S., C. Bock, A.S. de Boer, E. Kiskinis, A. Meissner, and K. Eggan. 2012. Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell*. 10:595-609.

Mirigian, L.S., E. Makareeva, E.L. Mertz, S. Omari, A.M. Roberts-Pilgrim, A.K. Oestreich, C.L. Phillips, and S. Leikin. 2016. Osteoblast malfunction caused by cell stress response to procollagen misfolding in alpha2(I)-G610C mouse model of osteogenesis imperfecta. *J Bone Miner Res*. 31:1608-1616.

Miyamoto, Y., T. Matsuda, H. Kitoh, N. Haga, H. Ohashi, G. Nishimura, and S. Ikegawa. 2007. A recurrent mutation in type II collagen gene causes Legg-Calvé-Perthes disease in a Japanese family. *Hum Genet*. 121:625-629.

Okada, M., S. Ikegawa, M. Morioka, A. Yamashita, A. Saito, H. Sawai, J. Murotsuki, H. Ohashi, T. Okamoto, G. Nishimura, K. Imaizumi, and N. Tsumaki. 2015. Modeling type II collagenopathy skeletal dysplasia by directed conversion and induced pluripotent stem cells. *Hum Mol Genet*. 24:299-313.

Oldershaw, R.A., M.A. Baxter, E.T. Lowe, N. Bates, L.M. Grady, F. Soncin, D.R. Brison, T.E. Hardingham, and S.J. Kimber. 2010. Directed differentiation of human embryonic stem cells toward chondrocytes. *Nat Biotechnol*. 28:1187-1194.

Patro, R., G. Duggal, M.I. Love, R.A. Irizarry, and C. Kingsford. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 14:417-419.

Peters, A.E., R. Akhtar, E.J. Comerford, and K.T. Bates. 2018. The effect of ageing and osteoarthritis on the mechanical properties of cartilage and bone in the human knee joint. *Sci Rep*. 8:5931.

Phillips, A.M., M.B. Doud, L.O. Gonzalez, V.L. Butty, Y.S. Lin, J.D. Bloom, and M.D. Shoulders. 2018. Enhanced ER proteostasis and temperature differentially impact the mutational tolerance of influenza hemagglutinin. *Elife*. 7.

Plate, L., B. Rius, B. Nguyen, J.C. Genereux, J.W. Kelly, and R.L. Wiseman. 2019. Quantitative interactome proteomics reveals a molecular basis for ATF6-dependent regulation of a destabilized amyloidogenic protein. *Cell Chem Biol*. 26:913-925 e914.

Quinlan, A.R., and I.M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26:841-842.

Rutishauser, J., and M. Spiess. 2002. Endoplasmic reticulum storage diseases. *Swiss Med Wkly*. 132:211-222.

Ryno, L.M., R.L. Wiseman, and J.W. Kelly. 2013. Targeting unfolded protein response signaling pathways to ameliorate protein misfolding diseases. *Curr. Op. Chem. Biol.* 17:346-352.

Schnabel, M., S. Marlovits, G. Eckhoff, I. Fichtel, L. Gotzen, V. Vecsei, and J. Schlegel. 2002. Dedifferentiation-associated changes in morphology and gene expression in primary human articular chondrocytes in cell culture. *Osteoarthritis Cartilage*. 10:62-70.

Sebastian, R.M., and M.D. Shoulders. 2020. Chemical biology framework to illuminate proteostasis. *Annu Rev Biochem*. 89:529-555.

Shoulders, M.D., and R.T. Raines. 2009. Collagen structure and stability. *Annu Rev Biochem*. 78:929-958.

Shoulders, M.D., L.M. Ryno, J.C. Genereux, J. Moresco, P.G. Tu, C.L. Wu, J.R.I. Yates, A.I. Su, J.W. Kelly, and R.L. Wiseman. 2013. Stress-independent activation of XBP1s and/or ATF6 reveals three functionally diverse ER proteostasis environments. *Cell Rep.* 3:1279-1292.

Soneson, C., M.I. Love, and M.D. Robinson. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 4:1521.

Spranger, J., A. Winterpacht, and B. Zabel. 1994. The type II collagenopathies: a spectrum of chondrodysplasias. *Eur J Pediatr*. 153:56-65.

Su, P., L. Zhang, Y. Peng, A. Liang, K. Du, and D. Huang. 2010. A histological and ultrastructural study of femoral head cartilage in a new type II collagenopathy. *Int Orthop*. 34:1333-1339.

Su, P.Q., R. Li, S.L. Liu, Y. Zhou, X.G. Wang, N. Patil, C.S. Mow, J.C. Mason, D.S. Huang, and Y.M. Wang. 2008. Age at onset-dependent presentations of premature hip osteoarthritis, avascular necrosis of the femoral head, or Legg-Calve-Perthes disease in a single family, consequent upon a p.Gly1170Ser mutation of *COL2A1*. *Arthritis Rheum-Us*. 58:1701-1706.

Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, and J.P. Mesirov. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 102:15545-15550.

Teckman, J.H., and D.H. Perlmutter. 2000. Retention of mutant alpha(1)-antitrypsin Z in endoplasmic reticulum is associated with an autophagic response. *Am J Physiol Gastrointest Liver Physiol*. 279:G961-974.

UniProt, C. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res*. 51:D523-D531.

Walter, P., and D. Ron. 2011. The unfolded protein response: from stress pathway to homeostatic regulation. *Science*. 334:1081-1086.

Wang, L., H. Pan, and Z.A. Zhu. 2014. A genetic pedigree analysis to identify gene mutations involved in femoral head necrosis. *Mol Med Rep*. 10:1835-1838.

Wang, W., N. Lian, L. Li, H.E. Moss, W. Wang, D.S. Perrien, F. Elefteriou, and X. Yang. 2009. Atf4 regulates chondrocyte proliferation and differentiation during endochondral ossification by activating Ihh transcription. *Development*. 136:4143-4153.

Wong, M.Y., A.S. DiChiara, P.H. Suen, K. Chen, N.D. Doan, and M.D. Shoulders. 2018. Adapting secretory proteostasis and function through the unfolded protein response. *Curr Top Microbiol Immunol*. 414:1-25.

Wong, M.Y., and M.D. Shoulders. 2019. Targeting defective proteostasis in the collagenopathies. *Curr Opin Chem Biol*. 50:80-88.

Yamauchi, M., and M. Sricholpech. 2012. Lysine post-translational modifications of collagen. *Essays Biochem*. 52:113-133.

Yammine, K.M., S. Mirda Abularach, L. Sampurno, J.F. Bateman, S.R. Lamandé, and M.D. Shoulders. 2023. Using CRISPR/Cas9 to generate a heterozygous *COL2A1* p.R719C iPSC line (MCRIi019-A-6) model of human precocious osteoarthritis. *Stem Cell Res*. 67:103020.

Yang, R.C., M.H. Chen, P.Y. Chen, C.Y. Chen, S.F. Tsai, C.K. Cheng, and J.S. Sun. 2014. A mutation of the *Col2A1* gene (G1170S) alters the transgenic murine phenotype and cartilage matrix homeostasis. *J Formos Med Assoc*. 113:803-812.

Yuan, X., H. Liu, L. Li, H. Liu, J. Yang, W. Shi, Y. Feng, H. Huang, and L. Wu. 2017. The roles of endoplasmic reticulum stress in the pathophysiological development of cartilage and chondrocytes. *Curr Pharm Des*. 23:1693-1704.

Zhang, Z., K. Zhu, H. Dai, Q. Wang, C. Zhang, and Z. Zhang. 2021. A novel mutation of *COL2A1* in a large Chinese family with avascular necrosis of the femoral head. *BMC Med Genomics*. 14:147.

Table 1. High-confidence interactors differentially engage with wild-type and Gly1170Ser procollagen-II across three biological replicates.[a]

Protein (*Gene*)	**Fold-enrichment**	***p*-value**[c]	***q*-value**[d]
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 (*PLOD2*)	1.95	0.0014	0.0181
Protein disulfide-isomerase (*P4HB*)	1.67	0.0004	0.0101
Prolyl 3-hydroxylase 1 (*P3H1*)	1.55	0.0352	0.1486
Cartilage-associated protein (*CRTAP*)	1.53	0.0328	0.1486
Fibronectin (*FN1*)	1.48	0.0397	0.1486
Peptidyl-prolyl cis-trans isomerase FKBP10 (*FKBP10*)	1.45	0.0450	0.1486
Serpin H1 (*SERPINH1*)	1.45	0.1294	0.2847
Peptidyl-prolyl cis-trans isomerase B (*PPIB*)	1.41	0.0324	0.1486
Prolyl 4-hydroxylase subunit alpha-1 (*P4HA1*)	1.41	0.0947	0.2777
Protein canopy homolog 2 (*CNPY2*)	1.36	0.3609	0.4793
Peptidyl-prolyl cis-trans isomerase FKBP9 (*FKBP9*)	1.36	0.1174	0.2847
Glucosidase 2 subunit beta (*PRKCSH*)	1.32	0.1252	0.2847
Prolyl 4-hydroxylase subunit alpha-2 (*P4HA2*)	1.31	0.1772	0.3340

Calumenin (*CALU*)	1.31	0.1904	0.3352
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 (*PLOD1*)	1.29	0.2534	0.3935
Calreticulin (*CALR*)	1.28	0.0116	0.1021
Multifunctional procollagen lysine hydroxylase and glycosyltransferase LH3 (*PLOD3*)	1.24	0.3631	0.4793
Procollagen galactosyltransferase 1 (*COLGALT1*)	1.24	0.3873	0.4869
Cytoskeleton-associated protein 4 (*CKAP4*)	1.17	0.4509	0.5411
Thioredoxin domain-containing protein 5 (*TXNDC5*)	1.14	0.6372	0.6887
Endoplasmic reticulum chaperone BiP (*HSPA5*)	1.13	0.1698	0.3340
Protein disulfide-isomerase A3 (*PDIA3*)	1.13	0.2887	0.4233
Endoplasmin (*HSP90B1*)	1.13	0.2434	0.3935
Protein ERGIC-53 (*LMAN1*)	1.11	0.6354	0.6887
Protein disulfide-isomerase A4 (*PDIA4*)	1.09	0.6521	0.6887
Protein disulfide-isomerase A6 (*PDIA6*)	1.01	0.9746	0.9896

[a]Proteins displaying a >2-fold increase in signal compared to negative control conditions in at least four out of the six experimental samples. See S3 data for complete, unfiltered data set.
[b]Fold-enrichment was calculated as the ratio of bait-normalized peptides detected in Gly1170Ser to the bait-normalized peptides detected in wild-type.
[c]*p*-value was calculated using a homoscedastic student's *t*-test across the three biological replicates.
[d]*q*-value was calculated using a two-stage step-up method of Benjamini, Krieger, and Yekutieli with a desired false discovery rate set to 10%.

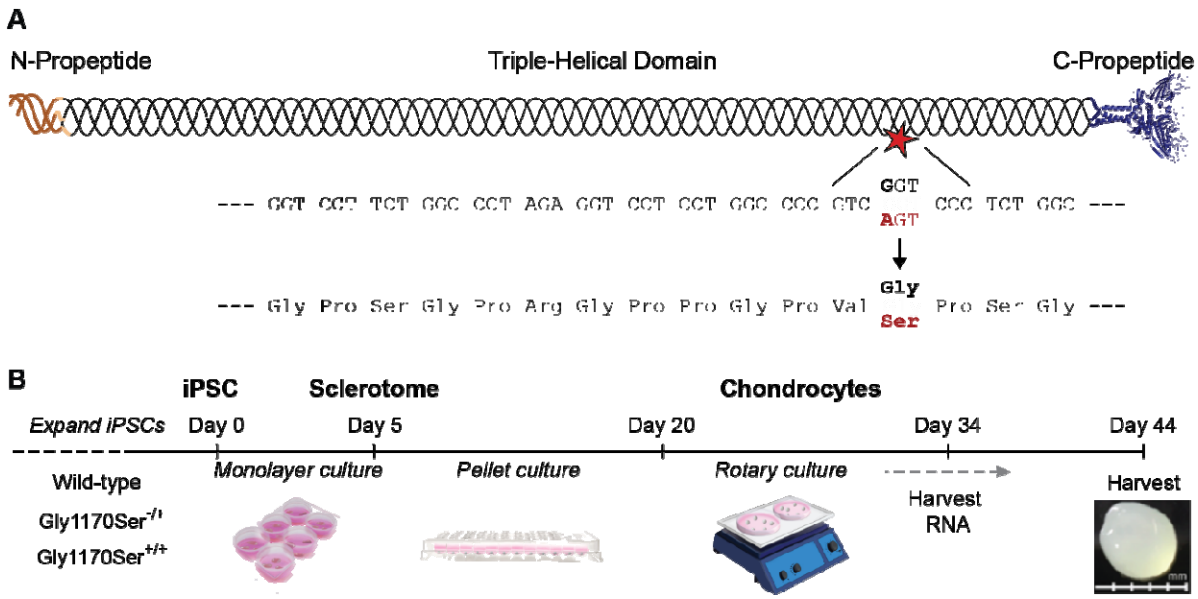

Fig. 1. A. Procollagen-II is composed of a lengthy (1,014 amino acid) triple-helical domain sandwiched by two smaller propeptide domains. The c.3508 GGT > AGT mutation in exon 50 results in a Gly→Ser substitution at position 1170 in the procollagen-II amino acid sequence (shown in red). **B**. Differentiation scheme from iPSCs to chondronoids.

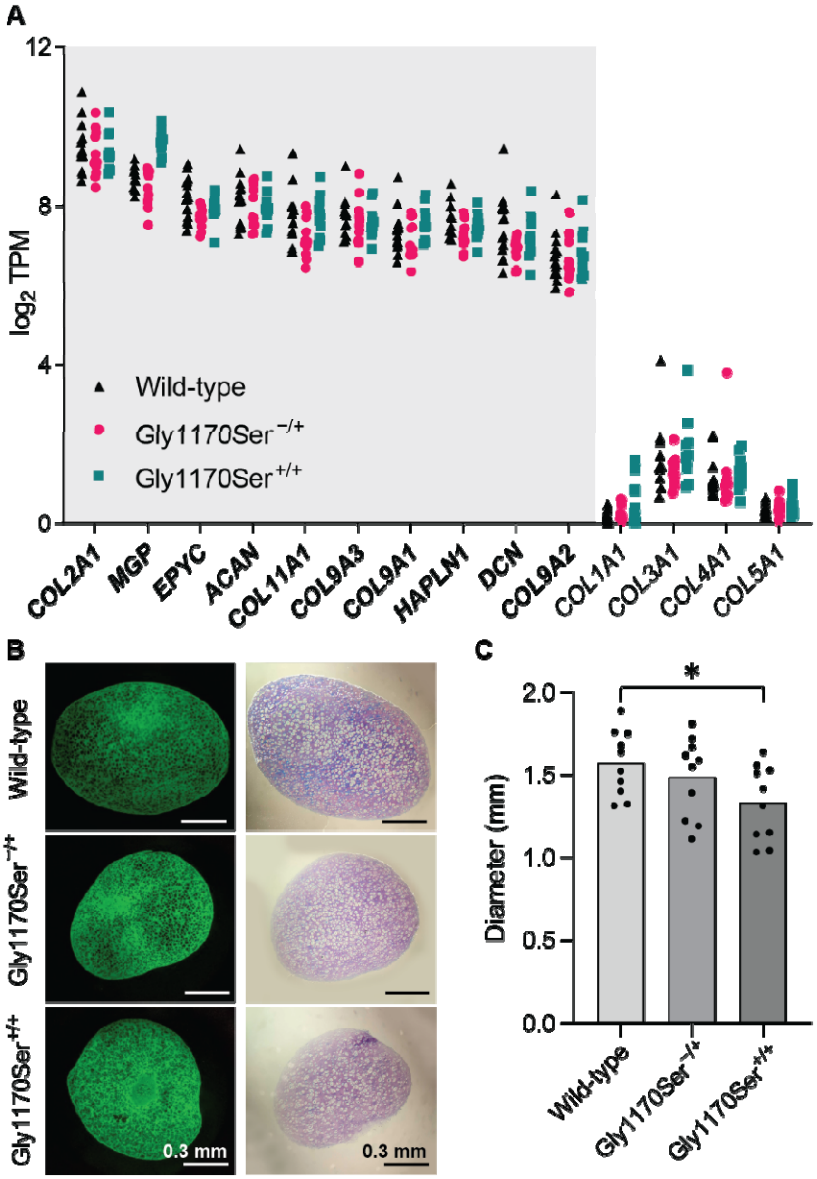

Fig. 2. A. Ten most highly expressed core matrisome genes (bolded) demonstrating high expression of cartilage-specific markers. Low to no expression of collagens of other tissue types highlights the faithful differentiation to chondrocytes. RNA samples were harvested 44 days after initiation of the chondrocyte differentiation. TPM: transcripts per million. **B.** Immunohistochemistry of collagen-II on histological sections of chondronoids shows that all genotypes deposit a matrix rich in collagen-II (green). Toluidine blue staining (purple) reveals an extensive proteoglycan matrix in all genotypes. **C**. Size distribution of chondronoids by genotype. Homozygous Gly1170Ser chondronoids are significantly smaller than wild-type ($p < 0.05$).

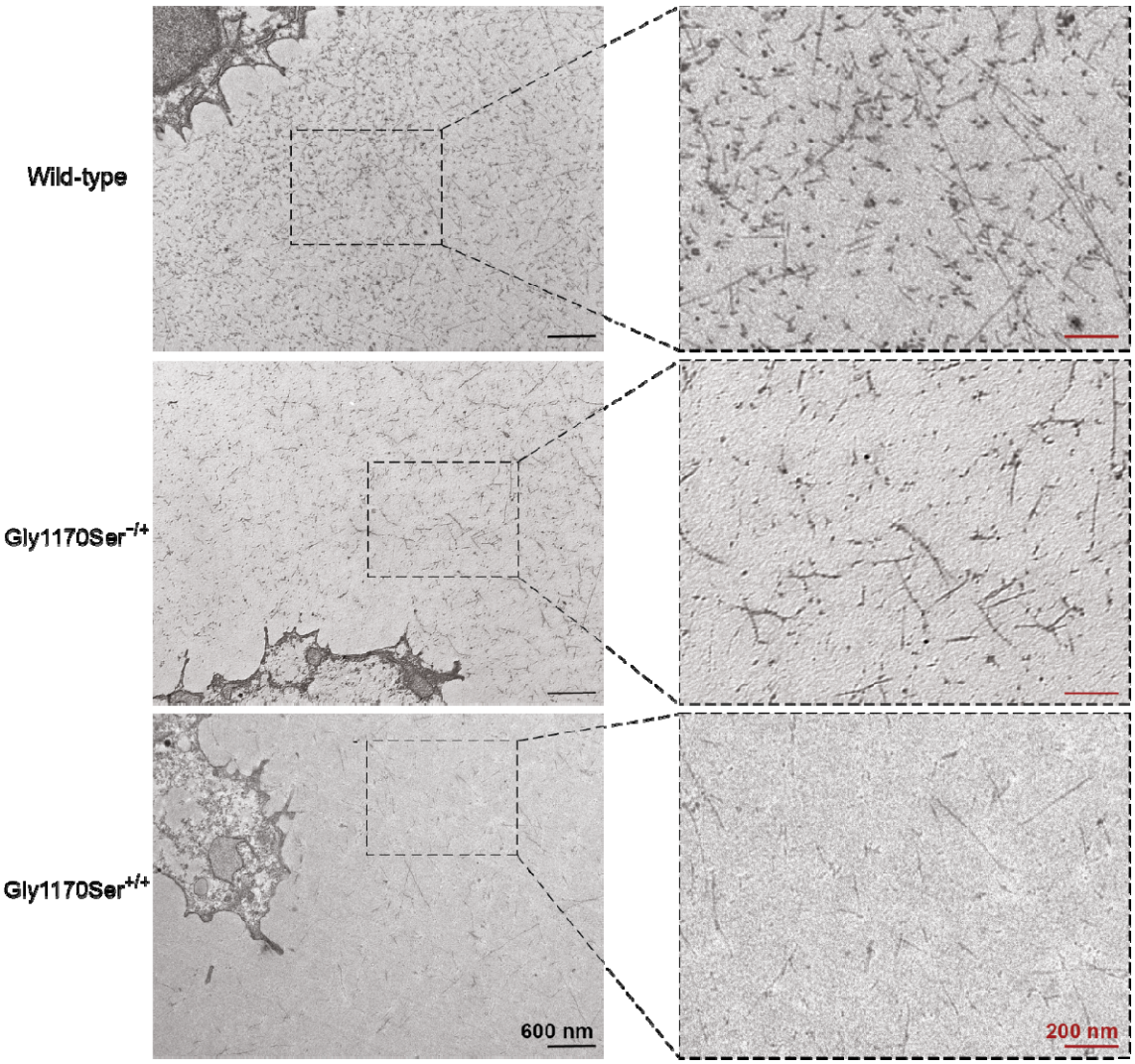

Fig. 3. TEM imaging of chondronoid ECMs. While the wild-type matrix is rich in network-forming collagen fibers, the clinically relevant heterozygous matrix is comparatively deficient, a phenotype that is further exacerbated in the homozygous chondronoids.

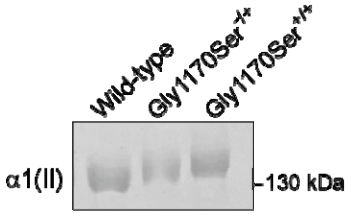

Fig. 4. SDS-PAGE separation of chondronoid-extracted collagen showing reduced electrophoretic mobility for the heterozygous and homozygous variants compared to wild-type collagen-II.

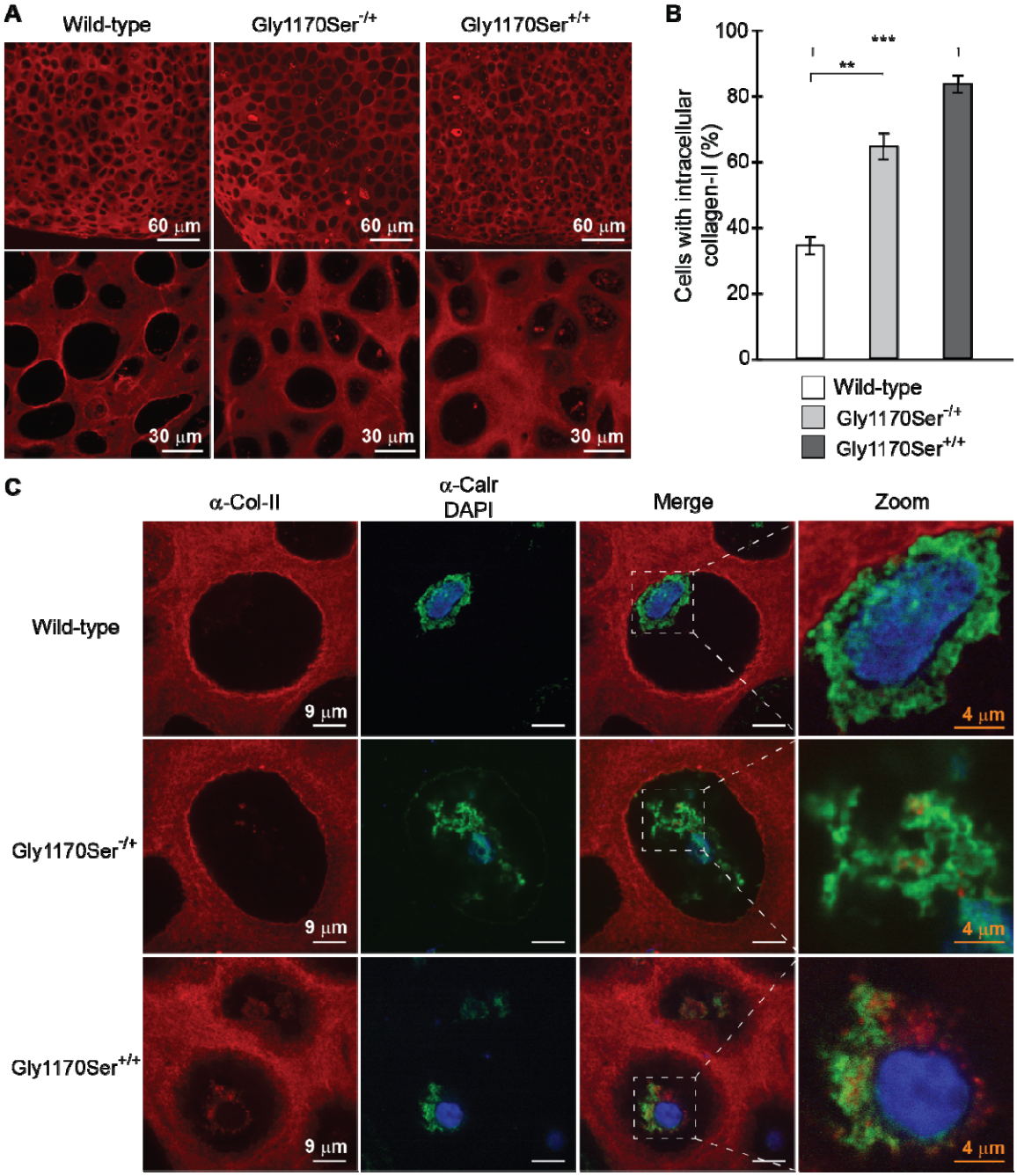

Fig. 5. A. Confocal images of collagen-II IHC at 10× and 63× magnification show an increase in intracellular procollagen-II retention across genotypes. **B.** Blinded quantification of the extent of procollagen-II intracellular retention. 34% of wild-type cells have intracellular procollagen-II, 64% of heterozygous cells, and 83% of homozygous cells. **C.** Confocal images of chondronoids stained for collagen-II, the ER marker calreticulin, and the nuclear stain DAPI reveal that much of the intracellularly retained procollagen-II in disease variants is localized in the ER.

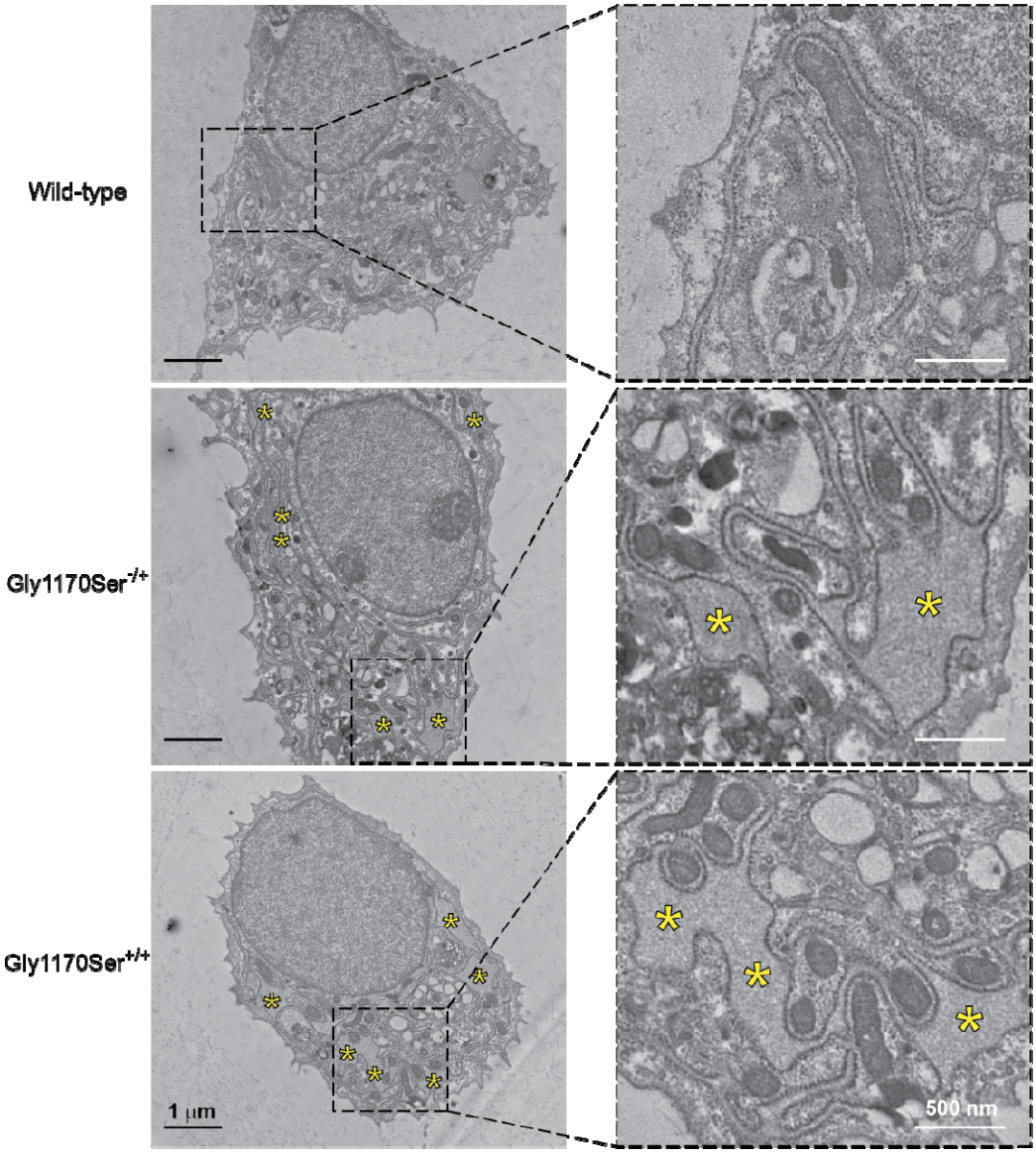

Fig. 6. Representative TEM images reveal the presence of distended ERs (marked with yellow stars) in cells expressing Gly1170Ser-substituted procollagen-II. The frequency and extent of ER dilation is further increased in homozygous cells compared to heterozygotes.

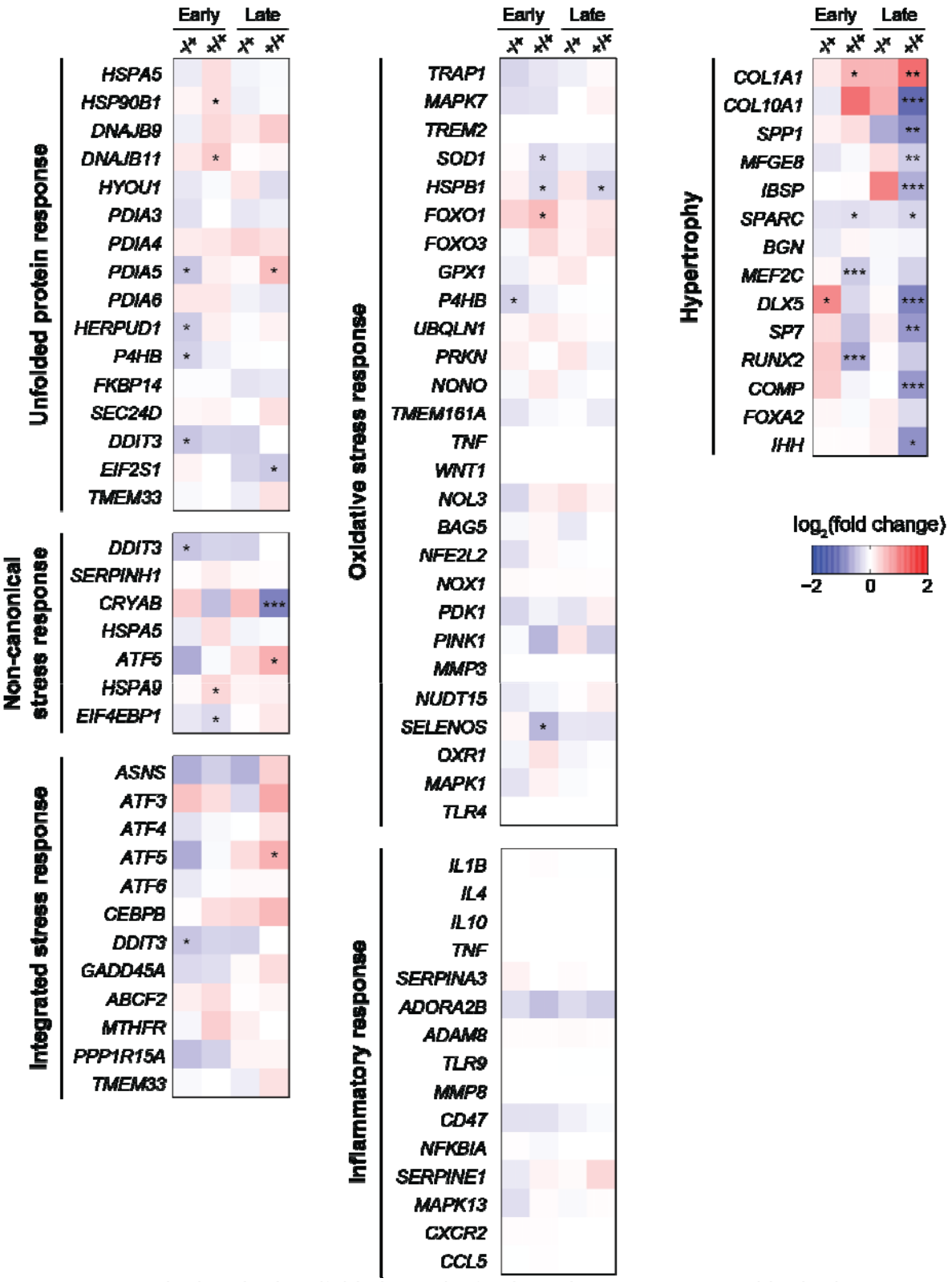

Fig. 7. Heatmap plotting the $\log_2$(fold change) of selected genes expressed in the heterozygous (−/+) and homozygous (+/+) Gly1170Ser chondronoids compared to wild-type chondronoids, at the early and late time points. * *p adjusted* < 0.05; ** *p adjusted* < 0.01; *** *p adjusted* < 0.001.

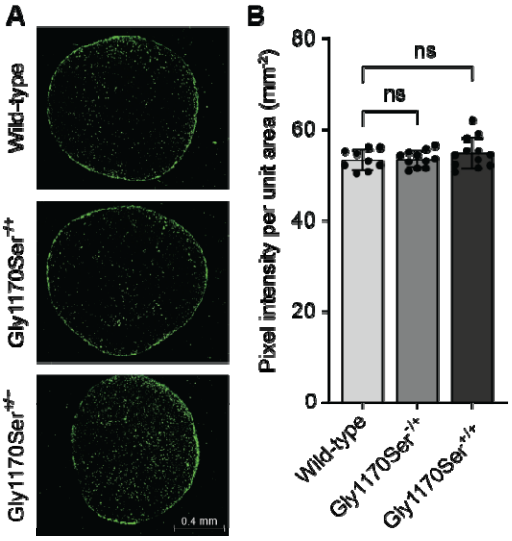

Fig. 8. A. Representative images of TUNEL-stained chondronoid sections for each genotype. Breaks in DNA are labeled with fluorescent nucleotides (fluorescein, visualized in green). Apoptotic cells and residual sheared DNA appear TUNEL-positive. **B.** Quantification of TUNEL signal (as pixel intensity per area) reveals no significant difference in apoptosis between genotypes.